The role of short RNAs in *Canna indica* leaf colour formation

Ohood Khaled Hesseki

100181588

A thesis submitted for the degree of Doctor of Philosophy (PhD)

School

School of Biological Sciences

University of East Anglia

Norwich

United Kingdom

June 2023

Table of Contents

Abstract

MicroRNAs (miRNA) are non-coding single stranded RNAs that plays crucial role in regulating gene expression. The objective of this project is to study the colour formation on the leaves of the *Canna indica (C. indica)* plant using two different approaches. The first the hypothesis I wanted to test was that the red and green sectors of *C. indica* cv Cleopatra are caused by the absence or presence of transposons. Around 500 differentially expressed genes between red and green sectors of *C. indica* cv Cleopatra have been identified and several of these were Myeloblastosis (MYBs) and Basic Helix-Loop-Helix (bHLHs) transcription factors. Five MYBs genes and CHS were analysed by PCR but none of them contained transposons, therefore a different approach was taken to find transposons in genes that may interfere with the anthocyanin pathway. Genomic DNA was extracted and the whole genome was sequenced by Pacific Bioscience (PacBio) as there was no available genome sequence for *C. indica* cv Cleopatra. The bioinformatic analysis identified a potential transposon in an intronic region of a MYB gene that was differentially expressed between green and red sectors. The validation of this transposon was not completed due to a lack of time but is being further investigated by the research group. The second objective was, to identify small RNAs differentially expressed in yellow striped *C. indica* cv Pretoria and pink striped of *C. indica* cv Durban compared to plain green and red leaves, respectively. The hypothesis was the sRNAs target enzymes involved in the anthocyanin biosynthetic pathway and silencing these genes resulting in the pinks strip. I also hypothesised that in the green leaves, sRNAs target enzymes involved in the chlorophyl biosynthetic pathway and silencing these genes lead to yellow strips. We found four miRNAs, miR166, miR529, miR530 and miR6300 that were differentially expressed in the different coloured leaves. I identified targets for some of these miRNAs but due to the lack of complete transcriptome sequence we were not able to predict targets for all four miRNAs.

Since the submission of my thesis, the research group took advantage the recently generated *Canna* genome sequence and found some potential targets for all differentially expressed miRNAs. Validation these targets will shed light on the effect of miRNAs in colour formation in *Canna* leaves.

Acknowledgements

After thanking the lord, I am grateful to Saudi Arabia government for giving me a full scholarship to study my PhD at East Anglia University. Thanks to the Saudi Arabia government for their huge support during Covid 19 pandemic.

Thanks to the main supervisor Prof Tamas Dalmay, who was kind and never stressed me out when experiments did not work. He continued to support, motivate, advise, and guide me throughout my PhD study. I extend my sincere gratitude to him. Thanks to the second supervisor, Dr Ben Miller for his support during my PhD study. I also would like to thank David Prince and other lab members, Rocky, Mamdouh and Firas, for their help with experiments.

My huge final thank goes to my family. I give my first warm thank to my parent, Rafa and Khaled, who supported and encouraged me to pursue my PhD after I completed my master's degree in environmental science at the University of East Anglia. I give my second worm thank to my lovely husband, Abdullah, who did not only keep supporting and encouraging me throughout my PhD studies, but also, cheer me up and solve any problems I faced during my PhD study. Third warm thank goes to my two lovely sons, Faris and Fahid, for being nice at home and patient with me during my PhD journey in the UK.

List of Figures

List of Tables

Abbreviations

ARP - Asymmetric Leaves1/Rough Sheath 2/Phantastica

ARF - Auxin Response Factors

AMV - Avian Myeloblastosis Virus

BHLH - Basic Helix Loop Helix

BLAST - Basic Local Alignment Search Tool

C. indica - *Canna indica*

CHI - Chalcone Isomerase

CHS - Chalcone Synthase

CLV-WUS - CLAVATA-WUSCHEL

CUC2 - CUP-SHAPED COTYLEDON2

SE - C2H2 zinc-finger protein SERRATE (SE)

DNA - Deoxyribonucleic Acid

DCL1- DICER Like protein 1

D-bodies - Dicing bodies

DMSO - Dimethyl Sulfoxide

HEN1- HUA Enhancer 1

GDNA - Genomic DNA

GSPs - Gene-specific primers

GRF - Growth-Regulating Factor

HST - HASTY protein

HD-ZIP - Homeodomain-leucine Zipper

HYL1- Hyponastic Leaves 1

OH - Hydroxyl

KAN - KANADI

KNOX1- KNOTTED 1-like homeobox genes

LCR - Leave Curling Responsiveness

MYB - Myeloblastosis

Nt - Nucleotide

PacBio - Pacific Bioscience

PHB - PHABOLUSA

PHV - PHAVOLUTA

PHB - PHABULOSA

Pri-miRNA - Primary miRNA

PAZ - PIWI/Argonaute/Zwille

REV - REVOLUTA/INTERFASCICULAR FIBERLESS1

RNA - Ribonucleic Acid

RIN - Ripening Inhibitor

RISC - RNA-Induced Silencing Complex

QC - Quality Control

SAM - Shoot Apical Meristem

TasiRNAs - Trans-acting short-interfering RNAs

TF - Transcription Factor

TE - Transposable Element

THC - Tetrahydroxy Chalcone

UV - Ultraviolet

5’-RACE - 5’ Rapid Amplification of CDNA Ends

Chapter 1: Introduction

1.1 MiRNA

A major macromolecule of life, ribonucleic acids (RNA) play a crucial role in gene expression. Ribosomal RNAs and tRNAs are key factors in protein synthesis, while messenger RNAs convey the genetic information encoded by the DNA to the sites of protein synthesis. In addition, RNA can also regulate the expression of genes. There are two types of RNAs: coding RNAs and non-coding RNAs. The most abundant small non-coding RNAs in plants are the 24 nucleotide (nt) small interfering RNAs and the 21-nt microRNAs (miRNAs). In plants, siRNAs were first discovered in relation to their contribution to post-transcriptional gene silencing and transcriptional gene silencing (1–3). Later, it was shown that they can cause specific downregulation of mammalian genes through the RNA Interfering pathway (4). MiRNA was first discovered in nematodes (*Caenorhabditis elegans*) by Victor Ambros and Gary Ruvkun who showed that a specific miRNA (lin-4) regulates nematode larval development (5). Many hundreds of miRNAs have been reported since then in animals and plants (6)

It has further been found that 22-nt miRNA can cut mRNA targets in plants, and the cleavage products can then be turned into dsRNA by RNA-dependent RNA polymerase 6. DICER like protein 4 subsequently generates secondary 21-nt siRNAs by cleaving the dsRNA. These siRNAs have been named phased siRNAs because the dsRNA is not cleaved randomly but in a phased manner, producing the same 21mer siRNAs (7). There are two types of phasiRNA: trans-acting siRNA and cis-acting siRNA (8).

The miRNAs that occur in plants are 21-24 nt, single-stranded RNAs that do not code for proteins (9). MiRNA genes are small, and they are located either between or within the protein-coding regions. MiRNA genes are transcribed into primary miRNA (pri-miRNA) by RNA polymerase II (10). Pri-miRNA has an imperfect double-stranded region that forms stem-loop secondary structures from which mature miRNAs are produced. In the nucleus, Pri-miRNAs can be stabilized through a nuclear-localised, forkhead-associated domain of a-protein called DAWDLE (11). After transcription, pri-miRNA is cleaved into hairpins to create pre-miRNA by the enzyme DICER like 1 protein (DCL1), with the help of the double-stranded RNA binding proteins Hyponastic Leaves 1 (HYL1) and C2H2 zinc-finger protein SERRATE (SE) (12,13).

These proteins are also necessary for recognising and directing the pri-miRNA, and for helping DCL1 to bind in an appropriate position. In the nucleus, the HYL1 and SE proteins react with DCL1 in dicing bodies (D-bodies)- or SmD3/SmB-bodies, leading to pri-miRNA stabilisation (14,15). DCL1 then cleaves the pre-miRNA to produce miRNA/miRNA* duplexes, which contain a 2-nt overhang- at the 3'-ends. These then get methylated by the 3' methyltransferase Hua Enhancer 1 (HEN1) by adding methyl groups to the 2'-hydroxyl (OH) at the 3'-terminal nucleotides of each strand (16). The methylation of the 2-nt overhangs is crucial in preventing the miRNA from being degraded and uridylated by the exonuclease activity of the small RNA degrading nuclease family (17–19). The methylated miRNA/miRNA* is transported from the nucleus into the cytoplasm through HASTY protein (HST) (20,21). After separating miRNA duplexes in the cytoplasm, one strand is selected and loaded into the RNA-induced silencing complex (RISC), which contains Argonaute family proteins, while the other strand is degraded. The strand that has the 5' end at the weaker side of the miRNA/miRNA* duplex is

preferentially loaded into the RISC (22). Figure 1 illustrates miRNA biogenesis (23).

In plants, Argonaute 1 is important for miRNA function (24). The Argonaute-protein has-two main domains: the N-terminal PIWI/Argonaute/Zwille (PAZ) and the PIWI domains. The PAZ domain can bind to the 3'end of single-stranded RNA, while PIWI domain is likely to bind to the 5'end of RNA through its carboxy-terminus. Each Argonaute family member is related with multiple silencing pathways (25–27). Mature miRNA recognises a single target location in the coding region, and a high level of complementarity between the miRNA and its target leads to successful target mRNA cleavage (10). The function of mature miRNA is to provide negative regulation by targeting mRNA (28).

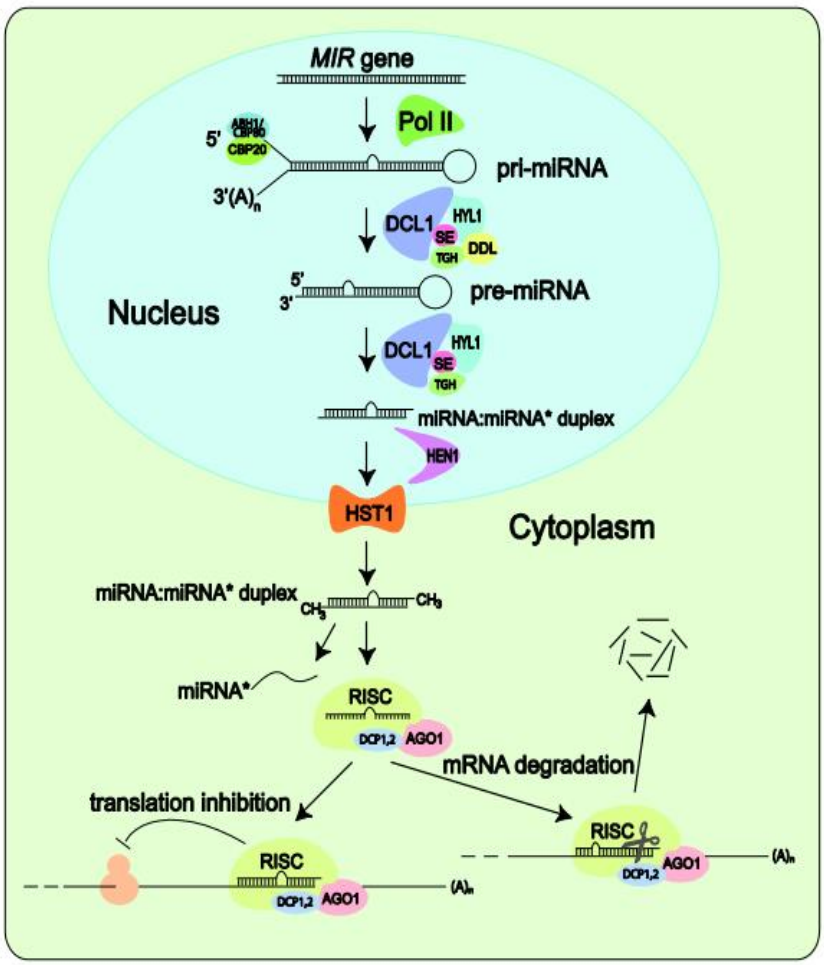


Figure 1: Plant miRNA biogenesis schematic diagram (23).

1.2 MiRNA and Leaf Development

Both genetic information and environmental factors play a role in regulating plant growth and development, and the expression of miRNA is an essential part of this (29). A significant role played by miRNAs is the regulation of gene expression by targeting specific mRNAs. Several miRNAs function by interfering with the level of plant hormones. One of the plant hormones is auxin, which plays an important role in leaf development and its level is regulated by miRNAs (30). Throughout the lifecycle of a plant, new organs can be produced continuously. The embryonic apical meristem contains stem cells that can divide into more stem cells by self-renewing and eventually differentiate into specific tissues. It is the

shoot apical meristem (SAM) that forms and develops the aboveground organs of the plant. Meristem activity is regulated via the *SHOOT MERISTEMLESS-WUSCHEL CLAVATA* pathway (31).

In leaf development, leaf primordium is differentiated from the SAM and develops to leaf blades. There are a variety of regulatory factors at play in these processes. In the SAM, auxin distribution and polar transport are responsible for organogenesis (32). MiR160 regulates auxin responses by targeting the Auxin Response Factors (ARF10, ARF16, and ARF17). For example, in *Arabidopsis thaliana* (*A. thaliana*), the leaves of mutants of ARF10 and ARF17, which are resistant to miR160, are serrated and curl upward, and the cotyledons are abnormally numbered (33).

Further, MYB domain protein transcription factor (TF) is expressed in leaf primordium that promotes leaf genesis. In order to promote growth and differentiation, the gene Asymmetric Leaves1/Rough Sheath 2/Phantastica (ARP) functions as a transcription suppressor to inhibit KNOTTED 1-like homeobox genes (KNOX1) expression. As plant leaves establish dorsal-ventral polarity, the expression of KANADI (KAN), ARF3 and ARF4 determine the dorsal axis, while homeodomain-leucine zipper (HD-ZIP) III determines the ventral axis. Leaf dorsal development is determined by the YABBY gene, which acts downstream of KAN. Due to the inhibition of HD-ZIP III family members on the abaxial side by miR165/166, HD-ZIP III expression is maintained only on the adaxial side (34,35).

For miR165/166 to regulate and restrict PHABOLUSA (PHB) to the adaxial side, AGO1 is necessary to target it at the HD-ZIP III transcripts in leaves (36). Similar

to AGO1, AGO10 is localised on the adaxial side of the leaf, where it inhibits miR165/166 activity and maintains HD-ZIP III mRNA accumulation (37). In addition, miR390 and AGO7 are critical for leaf polarisation. *ARF3* and *ARF4* are inhibited by TAS3 Trans-acting short-interfering RNAs (tasiRNAs) on the abaxial side of leaves, thus determining the adaxial side (Figure 2) (38)**.**

It is possible for miR165/166 to move from the adaxial surface to the abaxial side in leaves, thereby regulating HD-ZIP III gene expression patterns and forming leaf polarity. The SAM miR394 is a mobile signal produced by L1 layer cells that moves toward L2 and L3 layers in order to suppress Leave Curling Responsiveness (LCR) expression. Stem cell pluripotency can be maintained by affecting the CLAVATA-WUSCHEL (CLV-WUS) loop by suppressing the LCR signal in the lower stem cells (Figure - 2) (39).

Leaf size and structure are also regulated by miRNA. Leaf cell number and meristem size are ultimately determined by the miR396 and growth-regulating factor (GRF) balance (40). Furthermore, miR396 targets basic helix loop helix (bHLH) 74 to regulate leaf size (41) while CUP-SHAPED COTYLEDON2 (CUC2) helps forming the organ primordial boundary. By degrading the mRNA of the TCP-like TF family that regulates *CUC2*, miR319 regulates the growth and development of the leaves of *A. thaliana* (42). The expression of *CUC2* is also regulated by the repressor miR164 (43). During the evolution of composite leaves, the *CUC2*-miR164 system is critical (44).

When the level of miR171 decreases in plants, it can result in developmental defects, short vegetative phytomers, and delayed flowering. In addition, when the silencing suppressor HC-Pro gene is overexpressed in *A. thaliana*, there is an

increase in the level of miR171 target mRNAs, possibly leading to virus-mediated symptoms (45).

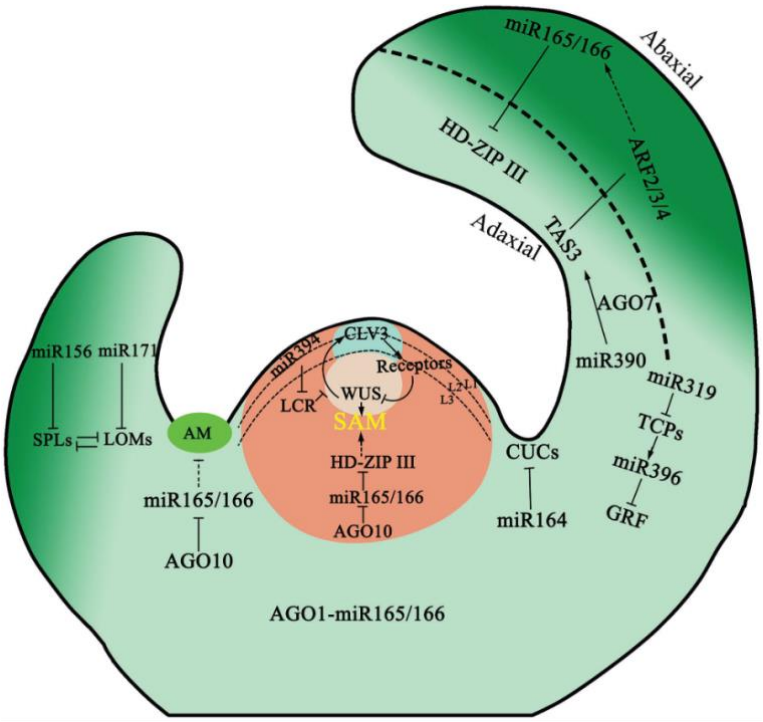


Figure 2: Functions of miRNAs in shoot apex (29).

As a result of vascular plants' xylem and phloem, water and nutrients are transported upward by xylem and carbohydrate is transported downward by phloem. Three tissues make up the vascular bundle: xylem, procambium, and phloem (29). There is high expression of HD-ZIP III genes in vascular bundles of roots, stems, and leaves of *A. thaliana.* In *A thaliana*, overexpression of miR165 reduces transcription levels of all members of HD-ZIP III family, thereby influencing vascular tissue differentiation and plant morphogenesis (46).

1.3 The structure and function of Transposons

Transposons were discovered by Barbara McClintock in maize (*Zea mays*) while conducting research on chromosome breakage for which she won a Nobel Prize. Her findings showed that these elements caused chromosomal translocations, deletions, and insertions. Because she noticed that the maize transposons could change the expression of genes close to or at the site of TE insertion, McClintock originally gave them the name "controlling elements." (47). A transposable element (TE) is a Deoxyribonucleic Acid (DNA) sequence. It can replicate, move, and integrate into new regions of the genome. It can copy itself and insert itself elsewhere in the genome (48–50). There are two major categories of transposons based on their movement (50). Retrotransposons (Class I) are able to spread their copy sequences throughout the host genome through the reverse transcription of an intermediate RNA molecule using a copy-and-paste mechanism. Class II DNA transposons are typically cut and pasted into a different region of the genome after being removed from their original location. Plants are particularly rich in retrotransposons, which make up a large portion of their genome. In maize, retrotransposons constitute 80% of the genome (51). Class I, Retrotransposons can be classified into two types based on their DNA sequence structure: The first type of retrotransposons is called long terminal repeat (LTR)-type retrotransposons because they have long terminal repeats at either end of their DNA sequence. A second type of retrotransposons is non long terminal repeat (non-LTR) retrotransposons, consisting of short interspersed nuclear elements and long interspersed nuclear elements. Both kinds of retrotransposons are found in plants in large numbers (52).

TE insertions can affect phenotypes in a variety of ways. First, insertion can increase genome size, which could slow down development and negatively affect fitness in some environments (53–55). A second issue is that TE insertions can

disrupt gene or regulatory sequences and impair gene function. Mendel, for example, studied wrinkled pea mutations caused by TE insertions as loss-of-function mutations (56). However, it is also possible that transposons are inserted into promoter regions of protein coding genes that may lead to either increased or reduced expression of the gene.

1.4 Pigment Flavonoids and Anthocyanin

Flavonoids are widespread in the plant kingdom and are classified under a group of multi-functional compounds. They are important for regulating auxin transport, determining male fertility, and helping plants defend against biotic and abiotic stresses. They are also important as pharmacological, medical, and nutritional compounds (57). Flavonoids are essential to a plant's secondary metabolism, which is related to their colouration mechanisms, chemistry, biochemistry, genetics, and molecular biology. Flavonoids, such as cyanidin glucosides, are also used in food colouring (58) and are the source of the various colours of *Cannas*, with the different flavonoids distinguishing which flowers will be red over yellow, etc. (9). Flavonoids are phenolics with a skeleton of diphenylpropane. The basic structure of flavonoids is C6-C3-C6. They are classified under several groups, such as flavones, flavanols, anthocyanins and chalcones (58,59).

1.4.1 Biological Synthesis of Flavonoids

The major pathways of flavonoid biosynthesis (60) are illustrated in (Figure 3). These pathways appear in all flowering plants and are well understood. Flavonoids are synthesized in the cytosol. The biosynthesis enzymes form a super molecular complex (i.e., metabolon) via protein–protein interactions and are

anchored to the membrane of the endoplasmic reticulum (60). Plants recruited enzymes from pre-existing metabolic pathways, such as 2-oxoglutarate-dependent dioxygenases, cytochromes P450 (P450) and glucosyl transferases (58).

Among the committed enzymes in this pathway, chalcone synthase (CHS) is responsible for catalysing the synthesis of tetrahydroxy chalcone (THC) from one coumaroyl CoA-4 molecules and three malonyl CoA molecules. Chalcone isomerase (CHI) rapidly and stereospecifically converts THC into colourless (2S)-naringenin (58).

1.4.2 Structure and Biosynthesis Pathways of Anthocyanin in Plants

There are 19 known types of anthocyanidins, which are the precursors of anthocyanins (Figure 4), although there are six main types are known: peonidin, delphinidin, petunidin, cyanidin, pelargonidin, and malvidin. The difference in colours is related to how many hydroxyl groups there are on the B-ring. For example, a large number of hydroxyls lead to a blue colour, while a red colour is due to the O-methylation of anthocyanins (58). While many enzymes are responsible for synthesizing the various anthocyanin compounds, TFs regulate the expression of these genes and therefore play an important part in colour formation (61). Combinations of R2R3 MYB, bHLH and WD40-type TFs are responsible for the spatial and temporal expression of structural genes during anthocyanin biosynthesis (58).

The structures and biosynthesis pathways of anthocyanins and flavonoids are summarised in Figure 3. The flavanone, flavonol, and flavone hydroxylation

process is catalysed by the F3'H and F3'5'H groups. *Rose hybrida, Chrysanthemum X morifolium, Dianthus caryophyllus,* and many other floricultural crops lack varieties of the violet or blue colour due to the absence of delphinidin. It is believed that this is due to their lack of the F3'5'H gene, which may have been lost during their evolution (58). Blue/violet carnations and roses have been instead developed by expressing the F3'5'H gene heterologously (62).

Anthocyanins are non-photosynthetic pigments produced in the cytoplasm and stored in the vacuoles of epidermal cells (63)**.** In fruit, anthocyanins are commonly in the outer cell layers. Some strawberry cultivars contain as much as 40% of their total phenol in anthocyanins in the outer layers of the fruit's cells (64). In plants, anthocyanins occur in flowers, leaves, fruit, grains, tubers, and roots (65). Anthocyanins are water-soluble, nontoxic pigments. These pigments are pH-sensitive and their colour changes when their pH of their environment changes (66)**.** At pH 3 and low, anthocyanins become more red and more stable as the quinonoidal base forms, while at pH 6 and above, they become more blue and more unstable (58).

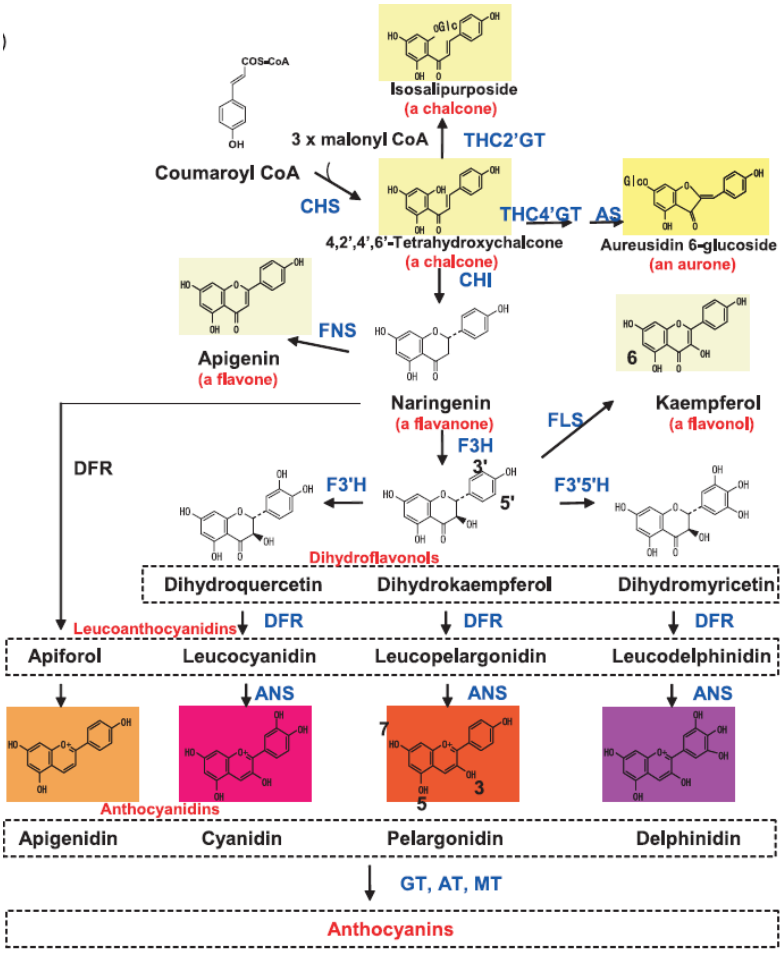


Figure 3: The structures and biosynthesis pathways of anthocyanins and flavonoids (58).

1.4.3 Producing and Consuming Fruits with Anthocyanins

Many agricultural products, including apples, grapes, and berry fruits, have a higher market value due to their anthocyanin-mediated fruit colouration (67). Anthocyanins possess powerful antioxidant and anti-inflammatory properties, which contribute to their health-promoting effects in humans (68). A substantial amount of evidence has demonstrated that anthocyanin-rich fruits can lower the risk of diabetes and cardiovascular diseases (69). Furthermore, anthocyanin consumption has been associated with positive effects on brain function (70). Accordingly, adding these bioactive compounds to fruits and vegetables enhances

their health properties and increases their consumer appreciation and market value. It is also for this reason that food producers use anthocyanins to add colour to their products (71), and as artificial foods become more prevalent, anthocyanins derived from fruit are in greater demand (71).

1.5 Transcription Factors regulating the anthocyanin pathway

1.5.1 MYB Genes

Specialized metabolites in plants include many different compounds, including flavonoids, phenolic acids, stilbenes, and monolignols. A variety of functions of these compounds have been identified, including acting as cell wall components (i.e., because of their mechanical strength), pigments, signalling molecules, phytoalexins, and ultraviolet (UV) light protectors. MYB TFs make up one of the largest families of TFs identified in plants. These MYB TFs are involved in the regulation of specialized metabolite biosynthesis. In almost all eukaryotes, the MYB TFs have a conserved MYB domain at their N-terminus (72).

The MYB structure is divided into two regions: the N-terminal region, which binds the domain with DNA, and the C- terminal region, which is responsible for regulating the activity of proteins. The MYB domain contains 4 imperfect amino acids sequence repeats (R) among its 52 amino acids (73). The MYB family is divided into four classes: 1R, R2R3, 3R and 4R. R2R3 is related to plants and is widespread in plants (73).

The first MYB TF to be identified in the Avian Myeloblastosis Virus (AMV) was v-MYB (74). In the MYB TF family, the number of imperfect repetitions varies from one to four (75). There are several classes of MYB TFs based on the number

of adjacent MYB repeats: 1R-MYB, R2R3-MYB, 3R-MYB, and 4R-MYB. In plants, the *COLOURED1* gene has been identified as the first gene encoding a MYB domain protein; in addition, it participates in the synthesis of anthocyanin in the aleurone layer of the kernels of Zea mays (76). MYB TFs are involved in many biological processes, including the defence and stress response, the circadian rhythm, seed and floral development, cell fate and identity, and the regulation of the primary and specialized metabolism (77). MYB TFs play a particularly important role in regulating the specialized metabolites of medicinal plants (78), and many factors affect the specialized metabolisms of these plants (79). For example, lignin syntheses can be inhibited by *Musa nana* MYB TFs (MusaMYB31) (80).

The evolution of plants is affected by their interaction with their environment. In turn, this results in specialized metabolites that can accumulate in the plant, protect against pathogenic microorganisms, and play a crucial role in plant metabolism. In studies of this process, MYB TFs were found to regulate the specialized metabolism in *Chrysanthemum X morifolium* (CmMYB1), *Panicum virgatum* (PvMYB4a), *Eucalyptus gunnii* (EgMYB1), *Leucaena leucocephala* (LlMYB1), and *Z. mays* (ZmMYB31 and ZmMYB42), which can inhibit lignin synthesis (81,82). In *Malus domestica*, MdMYBA, MdMYB3 and MdMYB1 are responsible for the red-pigmented anthocyanin biosynthesis in the peel (83,84). In the fruit cortex, MdMYB110a contributes to the red colouration during the late stages of fruit maturity, while MdMYB10 contributes to anthocyanin production in the foliage, peel and flesh (85,86).

R2R3 MYB is involved in the phenylpropanoid metabolism, biotic and abiotic stress, defence, differentiation, the hormonal response and cell shape (73). For

example, grapes are white or red based on the presence/absence of activators or repressors of anthocyanin. In grapes, four types of MYB genes (VvMYBA1 or VvMYBA2 or both or none) were found to be responsible for the accumulation of anthocyanin. Two of these genes are mutated in white grapes, while the active forms of the two genes lead to red grapes (87).

1.5.2 Structure and Function of bHLH TFs

The structure of the bHLH family protein involves the helix-loop-helix (HLH) (88). The bHLH gene family has the second highest number of TF genes (89), and bHLH proteins play a critical role in many physiological and developmental processes in plants, including anthocyanin biosynthesis (90), stress defence (91), and growth and development (92). There are usually 60 amino acids in a typical bHLH domain and 2 functional areas: helix and loop. The basic region consists of 10–15 amino acids at the N- terminus and a spiral region (i.e., HLH region) contains 40 amino acids (93).

The bHLH superfamily has an increasing number of members. Various plant species have been functionally characterised since bHLH in *Z. mays* was first identified as playing a significant role in anthocyanin biosynthesis (94). It is encoded by the ZmR and ZmB genes (95) and Roth et al (1991). Demonstrated that ZmC1 and ZmR interact in maize to enhance activity of ZmUFGT, a gene that is critical for the synthesis of anthocyanins (96).

Arabidopsis also contains AtEGL3, AtGL3, and AtTT8 bHLH TFs, which are also related to anthocyanin biosynthesis (97). The VvMYCA1 gene is essential for anthocyanin synthesis in grapes (98), while MdbHLH3 and MdbHLH33 are

essential for the synthesis of anthocyanin in apples. Anthocyanin synthesis can be promoted by MdBHLH3 and MdBHLH33 through interactions with MdMYB10 that stimulate the MdDFR promoter (85). A ternary complex (MBW) is formed between the bHLH, MYB, and WD40 (MBW) TFs. This complex regulates structural gene activity and flavonoid biosynthesis (99). It has been found that VcbHLHL1 interacts with VcMYBL1 and VcWDL2 to promote the accumulation of anthocyanins and the development of colour in blueberries (100).

Five MabHLH genes were found to be involved in cooling tolerance induced by methyl jasmonate (MeJA) in banana fruit (101). Specifically, as part of the bHLH gene family, MaMYC2 binds to MaICE1 to regulate induced cooling tolerance in fruits (102). MabHLH6 transcripts and proteins are both increased in bananas, and they are positively regulated by starch degradation during fruit ripening (103). The ripening of bananas is delayed from ripening by MaMYB3, which negatively regulates MabHLH6 (104).

1.5.3 Homeodomain-Leucine Zipper (HD-ZIP)

A few years after homeodomains (HDs) were first discovered in animals (105) they were also discovered in plants. Specifically, the first plant TF with an HD was discovered in maize (106). In the same year, TFs unique to plants, HD-Zip TFs, were discovered (107). There are two types of dimerization motifs in these TFs: the HD, which binds DNA, and the leucine zippers, which contain leucine in every seven residues (107). In plants, HD-Zip TFs have been associated with environmental responses (107), including those related to drought, salinity, flooding, and UV radiation (108).

There are four subfamilies of HD-ZIP TFs, which are classified according to their evolutionary relationships and proteins structures: HD-ZIP I, HD-ZIP II, HD-ZIP III, and HD-ZIP IV (109). HD-Zip class I members are generally involved in responses to abiotic stress, such as light and water (110). Research has revealed that water-limiting conditions and abscisic acid (ABA) application upregulate *A. thaliana* homeobox 12 (ATHB12) and ATHB7 expression (111). Expression and transformation studies have revealed that class II proteins play a major role in phototropism and auxin responses (112). *Arabidopsis* ATHB-2 regulates shade avoidance responses in three distinct phytochromes. In *Arabidopsis* seedlings overexpressing ATHB-2, hypocotyl elongation has been included at a level that is similar to what observed in wild type plants when grown under far red light (113).

HD-Zip class III members contribute to morphogenesis. During embryonic development, REVOLUTA/INTERFASCICULAR FIBERLESS1 (REV), PHABULOSA (PHB), and PHAVOLUTA (PHV) were found to control the pattern of apical formation (114). Molecular analyses also revealed the importance of REV and PHB in maintaining and initiating SAMs (115). Researchers determined the mechanism of action of REV, PHB, PHV and KANADI in controlling lateral organ patterning (116). In plant organs, HD-Zip class IV is often expressed in outer cells where it regulates epidermal fate, anthocyanin accumulation and trichome formation (115).

1.5.3.1 Structure and function of HD-Zip class III TFs

The HD-Zip III subfamily consists of five members: PHB/ATHB14, ATHB8, PHV/ATHB9, CORONA/ATHB15, and REV/IFL1. Between the HD and zipper-loop-zipper (ZLZ) domains of this subfamily, there are four additional conserved amino acids, and they engage in steroidogenic acute regulatory protein related

lipid transfer (START), which is followed by a conserved adjacent region called the START-adjacent domain (SAD). It has been well established that animal START-containing proteins also contain lipid ligands, but plants have not yet been found to contain any of these ligands (117). It is likely that this motif plays a significant role in activity regulation due to its high conservation throughout evolution. Each member of the superfamily has a conserved domain at the C terminus, known as the MEKHLA domain, which was named after the goddess of lightning, water, and rain. Several proteins throughout all kingdoms of life contain a similar domain known as the Per-Arnt-Sim (PAS) domain, which is involved in sensing light, oxygen, and redox potentials (118).

1.5.3.2 Expression and Function of HD_ZIP Class III

Subfamily III TFs have been well characterised in terms of their roles in apical meristem development, vascular bundle development, auxin transport, and adaxial organ development (i.e., through REV/FL1, PHB/ATHB9 and PHV/ATHB14) and regulating vascular development (i.e., through ATHB8 and ATHB15) (114). This subfamily only contains one single mutant with a phenotype: *Arabidopsis* REV. Five members of this subfamily share functional overlap and there are defects in the development of shoots, leaves, stem cells, and vascular systems in the REV phenotype (114,119). Mutants of ATHB8 do not exhibit any phenotypic defects; however, constitutive expression causes xylem overproduction (120). In addition, ATHB8 regulates procambial development at an early stage (121). Transcriptional and posttranscriptional regulation are both present in HD-Zip III genes, for example both miR165 and miR166 target HD-Zip III mRNAs (116,122).

A primary cell wall is composed of xyloglucans, crystalline cellulose microfibers and pectic polysaccharides. These cells can withstand mechanical forces and secure space in which to grow because of this remarkable combination of strength and pliancy (123). Secondary cell walls contain mechanical strength and rigidity by combining cellulose, hemicellulose, and lignin, enabling vascular plants to grow tall and compete for light (124). Studies have shown that miRNA165 and miRNA166 regulate HD-Zip III members (125,126). An 859 amino acid protein is encoded by OsHox32. It has been identified as a member of the HD-Zip III family (127), and it is thought that OsHox32 inhibits OsCAD2 and OsCESA7 expression by binding to their respective promoters, thus reducing the amount of lignin and cellulose (128).

Researchers have found that knocking down miR166 in rice affects the stem xylem development and reduces the hydraulic conductivity; due to this, the plant's leaf rolls and becomes drought-resistant (129). However, a study found that plants with an overexpression of OsHox32 have narrow leaves and a shorter height (130). Furthermore, this study also demonstrated reduced plant height and rolled leaves in OEHox32 and STTM166b transgenic rice plants, supporting the findings (128). OEHox32 and STTM166b transgenic plants were also observed to have drooping leaves and brittle stems. In addition to modulating plant growth and development, OsmiR166b-OsHox32 also enhances mechanical strength, which indicates that it plays an important role to play in both (128). Similarly, in bananas miR166 targets HOX32 and regulates plant height and pinnacle length (131).

1.6 *Canna Lilly*

1.6.1 *Canna* History

There are 10 species in the *Canna* genus, all native to tropical and subtropical America, with the majority coming from South America. Since the mid-19th century, *Canna* has primarily been known as a starchy crop. The rhizomes of *C. edulis* are still harvested as food throughout the tropics and have been cultivated in Peru for at least 4,500 years. The seeds of *C. indica* are hard and black, giving it the nickname 'Indian shot'(132). *Canna* is also called the C. lilly (133).

Canna grows widely in muddy soil (9). *Canna* has also evolved to flourish in Europe, especially in Italy, under very different climatic conditions than those found in West Indies (134). It forms small or large monotype stems, and it can be found in various areas, such as in forests, near villages, and along roads (135). It is also known as an ornamental plant for gardens because of its colourful flowers and foliage (133).

C. indica can tolerate any type of soil, such as tropical latosols, weathered soils and acidic soils. It prefers deep heavy soil that is fertile, well-aerated, well-drained and rich in humus. The rhizomes grow slowly in compacted clay. *C. indica* prefers warm weather and temperatures up to 32 °C. It can further grow in forest areas, wetlands, river zones, water courses and wastelands (136).

C. indica is said to have been named indica by Charles de l'Ecluse, who described and sketched it (137). During the Victorian era, *Canna* became very popular as an ornamental plant after it was first hybridized. The RHS trials in 1906 included more than 270 *Canna* entries. During the 20th century, their popularity declined

but in the 21st century they became more popular again. The majority of cultivars are complex hybrids that originated from *Canna indica, Canna warscewiczii, Canna ridiflora, Canna glauca, and Canna flaccida*. The first ornamental *Canna* plants were bred in France in 1848 after Thré Année brought back plants that he had collected in South America. Année's plants were chiefly foliage subjects, but twenty years later, a fellow French breeder named Antoine Crozy, successfully developed cultivars that have larger flowers (132).

The books *Hortus Cliffortianus* and *Viridiarum Cliffortianum* were published by Carl Linnaeus in 1737 and contained a list of the plants grown in the garden of George Clifford. In this garden, Linnaeus noticed two kinds of plants:

1. A type of *C. indica* which originated in the warmer region of Asia, Africa and America and grows in the wild and it has in four varieties.
    - *A naturalis allegata planta, flore luteo*.
    - *A Cannacorus flore luteo punctate* Tournef.
    - *A Cannacorus amplissimo folio, rutile flore* Tournef.
    - *A Cannacorus flore coccineo splendente* Tournef.

2. A type of *Canna glauca* which originated in America's hot and humid regions (138).

By arranging these descriptive sentences Linnaeus was later able to distinguish between *Canna indica* and *Canna glauca*, the two species of *Canna* he would later name. When Linnaeus published his book 'Species Plantarum' in 1753, the official taxonomy of Cannaceae began. Using his new binominal system, Linnaeus arranged the plants in this book, listing three species of *Canna*:

- *Canna indica*.

- *Canna glauca*.
- *Canna angustifolia* (139).

In 1754, Miller continues to call the plant Cannacorus by the generic name in his abbreviated 4th edition. Six species were consisted of the genus, each described using over than two Latin words, with the first being *C. latifolius vulgaris* which is *Canna indica* (140).

The first plant list compiled by William Aiton grown at Kew in the Royal Botanic Gardens in 1789, called Hortus Kewensis. It included *Canna glauca* and *Canna indica*, with Aiton dividing *Canna indica* into four varieties based on its leaves' shape flower colour and whether the flowering parts were erect or reflexed:

- *Canna indica* β *lutea* Aiton.
- *Canna indica* α *rubra* Aiton.
- *Canna indica* δ *patens* Aiton.
- *Canna indica* γ *coccinea* Aiton (141).

In 1789, A.L. de Jussieu created the family Cannaceae. The genus *Canna* and a number of monocotyledonous genera belonging to the Costaceae, Zingiberaceae, and Marantaceae familiesm (142). The Cannae and the Scitamineae were split by Roscoe in 1807 according to Linnaeus's definition of Cannae having one stamen and one style. Based on John Lindley's classification in 1853, Marantaceae, which includes Cannaceae, was classified as Anomales, which contains plants with unsymmetrical flowers and parallel leaves(143). Canneae, Maranteae, Museae, and Zingibereae were considered tribes of Scitamineae according to the Bentham & Hooker system, as reported in *Genera Plantarum* in 1883. The definition of Cannaceae was revised by Otto Petersen in 1888 (144).

In most taxa of *Canna*, chromosome numbers are 2n = 18, but most hybrid species are triploid 3n = 27 (145). The explanation for 3x = 27 is that a fertilised cell has three sets of chromosomes (3n), totalling 27 (146) and *Canna indica* has been reported to be triploid (147).

1.6.2 *Canna* Uses

The leaves of *C. indica* have historically been used for wrapping food, and the seeds have been used for making necklaces and rosaries or as shot in children's pop guns (135). *C. indica* can also treat industrial wastewater in structured wetlands by removing contaminants from the soil (148). Further, it has been used as a traditional medicine in some tropical and sub-tropical regions, as it has beneficial effects in the treatment of several diseases, such as hepatitis and rheumatism (136). *Canna* leaves are fibrous, and they can be used to make paper. The red flower of *C. indica* can be used as a food colourant and a natural antioxidant (136). As *Canna* is highly tolerant of contaminants, it can be also used to remove many undesirable pollutants in wetlands (149), such as organic pollutants, phosphorus, nitrogen, and heavy metals (150). Spray made from *Canna* stem or leaf has also been used to kill insects (133). Last, fresh leaves can be used as fodder, containing approximately 90% water, 1% protein, 0.2% fat, 7% carbohydrates, and 1.4% ash; their digestibility is about 20% (133). *Canna* seeds are usually black with a hard-coating and can be used to make beads or rosaries. Poultices made from pounded seeds are used to relieve headaches in Java, while rhizome juice is used to treat diarrhoea. The hard, impenetrable seed covering on *Canna* makes it the only Liliopsida member (i.e., in the monocot group) that is known to hibernate seeds (151).

These plants have been used to heal humans since the dawn of time. Globally, all cultures have a deep understanding of herbal medicine, and *C. indica* has been used in traditional medicine for the treatment of a variety of ailments. It contains several phytochemicals including carbohydrates, alkaloids, proteins, flavonoids, cardiac glycosides, terpenoids, steroids, oils, saponins, tannins, phlobatannins and anthocyanin pigment. Pharmacological studies have shown that these plants have anti-viral, anthelmintic, exerted antibacterial, anti-inflammatory, antioxidant, analgesic, hepatoprotective, anti-diarrheal and many other properties (152).

Specifically, *C. indica* has been used in the treatment of malaria, dysentery, diarrhoea, cuts and bruises, fevers and dropsy (153); it is also a diuretic and, diaphoretic (154). *Canna* leaves are also used for treating rheumatic pain, malaria, arthritis, and they have homeostatic properties as stop clotting and bleeding (155).

Flower decoctions have been used for external wound healing and treatment of eye diseases. The anti-inflammatory, anti-diabetic and anti-viral properties of *Canna*, have been demonstrated in numerous pharmacological experiments along with its effect on the enzyme acetylcholinesterase and treating anti-brain disorders (156,157). There are also various bioactive molecules in *Canna*, including flavonoids, tannins, cardiac glycosides, saponins, steroids, alkaloids and terpenoids (158). Extracts of the whole plant of *C. indica* have been shown to be anti-inflammatory, potentially radical scavenging and neuroprotective (157).

Several chemical compounds are present in *C. indica*, including henicosdine, tarixer-14-en-3-one, tetracosane, and tricosane (159). Mouse studies have shown that *C. indica* reduces clotting, abdominal capillary permeability, and bleeding time (152). According to histological studies, a carbon ionic liquid electrode (CILE) changes the morphology of tissue samples to some extent, demonstrating some insignificant changes in tissue sections. Methanolic extracts of *C. indica*, however, have been shown to have protective effects on tissues (160). The results of that study also showed that rat livers are protected from CCl4-induced liver damage by aerial methanol extracts of *C. indica* (160).

In *C. indica*, the leaf epidermis is very thin, and the lower epidermis is narrower, more irregular and has more costal bands than does the upper epidermis, but both epidermis layers are rectangular (161). There is only one layer of hypodermis beneath the epidermis, and the hypodermal cells of the upper layer serve as expansion cells (162).

1.6.3 *Canna* origin and phenotype

There are huge variations in flower colour, staminodes length, leaf number, and leaf shape among *C. indica* cultivars (163). There are many cultivars, most of which are unnamed. South America is known for two cultivars: ‘Verdes’ and ‘Morados’, which have dirty white 'corms' and vivid violet foliage. *C. indica* stems have a thin, cutinised epidermis, elongated cells, and somewhat thickened walls, according to Noe (164). In parenchyma, one to two layers of small cells lie beneath colourless hypodermal cells, which separating them it from the epidermis. In addition, chlorenchyma patches remain attached to sclerenchyma patches as well. A few thick-walled parenchyma bundles are sometimes found adjacent to xylem and phloem, but vascular bundles are scattered irregularly

(164). There are two layers of parenchyma in the leaves, palisade and spongy mesophyll cells- which have characteristically oblique round cells with oblique ends. In 1961, researchers found no differentiation between palisade and spongy parenchyma in mesophyll. Occasionally, stomata are found on the upper surface of leaves, and the type of stomata is unknown. The height of the *Canna* is less than 0.75 m to 2.4 m. Its flowers have a variety of colours, including yellow, cream, pink, red and orange (9).

The leaf of *Canna* is lanceolate or ovate and its leaves measures 10 cm to 30 cm long by 10 cm to 20 cm wide. The laminae can reach a length of 60 cm, and the inflorescence has waxy-glucose peduncles that are about 30 cm long. There are dark green leaves with purple veins and margins, around the stem, and the petioles of the oblong leaves extend downwards as sheathing. Veins are parallelly arranged throughout the lamina. Smooth, wavy margins with acute apexes appear on leaf margins (152). A typical cultivar of the plant has green leaves, but some varieties have glaucous, maroon, brownish, or variegated leaves (134). It grows from an underground, large, thick rootstock. Like banana plants, it has large leaves, but they are smaller than those of banana plants (152).

1.6.4 *Canna* Flower

Canna flowers are asymmetrical and consist of three sepals and three petals; they are small and inconspicuous, as they are hidden beneath extravagant stamens. Staminodes, which appear to be petals, are actually highly modified stamens. Inflorescences of flowers can be spikes or panicles (i.e., thyrses), and are usually red, orange, or yellow (or any combination of these). Despite their odd appearance, the flowers were designed by nature to attract insects that collect nectar and pollen, such as sunbirds, bees, hummingbirds, and bats. These

pollination mechanisms are conspicuously specialised. Pollination is done by insects picking up pollen from flattened styles. They are pollinated by hummingbirds in their homeland (165). It blooms between August and October (166).

1.6.5 *Canna* Rhizomes

Genetically identical populations result from vegetative propagation. Pieces of harvested root or tuber crop are used for planting, including aroids, *Canna*, yams and Irish potatoes. One use of the rhizome is to produce genetically identical progenies for the horticultural industry (133). The *Canna* species is important to the world's economy and medicine. *Canna* has the largest starch grains of any plant, and these grains grow from swollen underground stems, called rhizomes, which store starch (134). Rhizomes can be eaten raw or cooked, tasting similar to cassava, potatoes or taro. As such they are an important source of starch. When rhizomes are cooked, they become sweet, translucent and mucilaginous. They are used in different ways in many countries. For example, in Thailand the local people value the starch in boiled rhizomes. In Vietnam, a piece of starch is used in making cakes, candies and rice paper. *Canna* starch can also be used to make noodles, soup and fried dishes. In Vietnam, it appears in a unique dish that is served during celebrations. In China, people make alcohol from *Canna* starch. In Indonesia, *Canna* starch is frequently sold on the street, simply because it is delicious. In Hawaii, people use it to make a dessert, while in Colombia, people use it to make salted crackers. It is sometimes, mixed with cheese, either both at home or in factories, for commercial distribution (136).

Acute hepatitis is treated in Hong Kong with a decoction of fresh *Canna* rhizomes. Traditional medicine in Indonesia and China also uses crushed fresh rhizomes topically to treat traumatic injuries. Rhizome decoctions are used diuretics, and macerated rhizomes are used to treat nose bleeds in the Philippines (133). For all these and other uses, 2,000 to 4,000 tons per year of *Canna* starch is produced in and exported from Australia (133).

Cattle feed can also be made from both the rhizomes and leaves of *Canna* (133). Rhizomes from *C. indica* have approximately the following properties per 100g edible portion: water (75 g), protein (1 g), calcium (21 mg), fat (0.1 g), iron (20 mg), carbohydrates (22.6 g), phosphorus (70 mg), vitamin C (10 mg) and vitamin B (0.1 mg) (133). They contain more than 90% starch, in the form of carbohydrates and less than 10% sugar (i.e., sucrose and glucose) (133).

1.6.6 Leaf phenotype of *Canna Indica* cv Cleopatra

C. indica cv Cleopatra is a stunning cultivar also known as 'Yellow King Humbert' and 'Queen of Italy'. The cultivar can grow up to 1.4 m tall, with yellow flowers dotted with orange with randomly appearing red sectors. *C. indica* cv 'King Humbert' from which Cleopatra is originated has scarlet flowers and grows large (167). Its leaves are normally green, but bronze sectors appear along the veins on either or both the dorsal and ventral sides (Figure 4, A). It is the red pigment anthocyanin in the epidermal cells that gives the bronze colour, while the rest of the tissues is green. In addition to the bronze sections on the leaves, bronze can also be seen on inflorescence stalks and bracts. The stripes always run parallel to the long axis and vary in width (Figure 4, B) (134). *C. indica* cv Cleopatra. leaf is illustrated in Figure 5 (134). The leaves of the *C. indica* cv Cleopatra both sectors red and green had shown under the microscope. To

understand that the leaves colour in *C. indica* cv Cleopatra. *C. indica* cv Cleopatra red and green sectors from behind surface of the leaf (Figure 6, A). *C. indica* cv Cleopatra transverses red and green boundary of the leaf (Figure 6, B). *C. indica* cv Cleopatra transverses red both sides of the leaf (Figure 6, C).

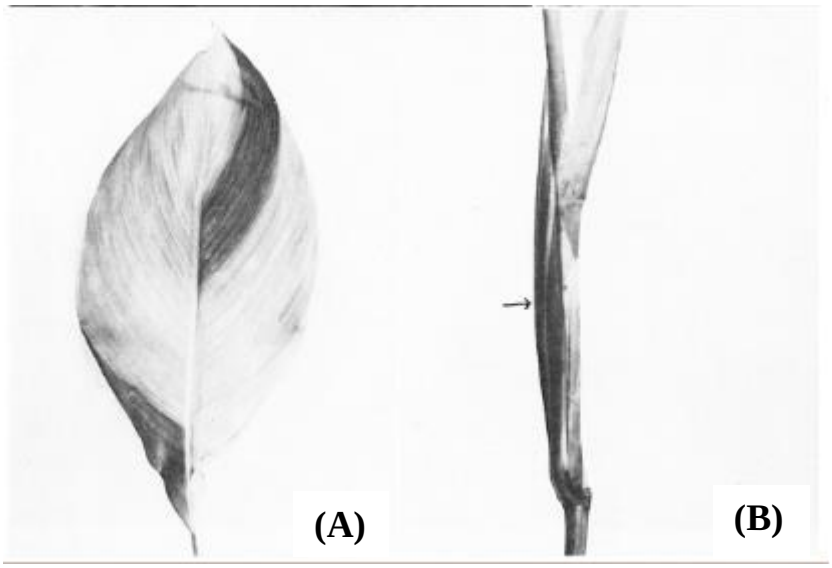


Figure 4: Display bronze sectors on the leaves and inflorescence stalks of the *C. indica* cv Cleopatra 'Queen of Italy' (134).

A: The bronze sections on the leaves. **B:** The bronze sections on the inflorescence stalks.

Figure 5: *C. indica* cv Cleopatra leaf.

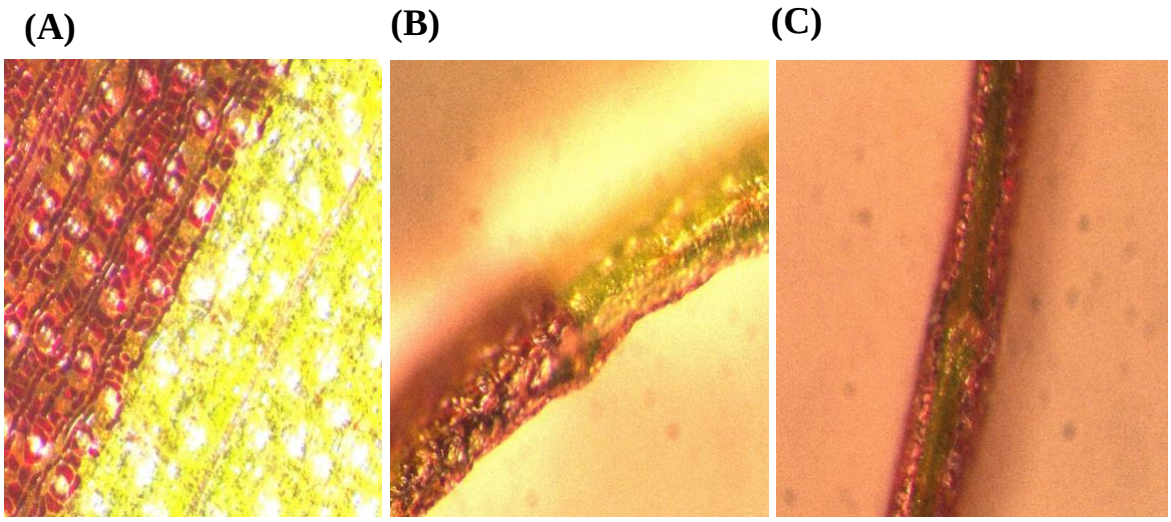


Figure 6: *C. indica* cv Cleopatra leaf red and green sectors under the microscope.

A: Cleopatra red and green sectors from behind surface. **B:** Cleopatra transverses red and green boundary. **C:** Cleopatra transverses red both sides.

Transverse sections of the staminodia of *C. indica* cv Cleopatra and 'King Humbert' flowers that comes from different leaves sectors (i.e., green, yellow-bronze, and bronze areas), were cut to examine the distribution of red and yellow pigments. The red colour is produced by the vacuoles, while the yellow is produced by yellow plastids. The epidermis of *C. indica* cv Cleopatra flowers is mostly yellow, though there are small areas (i.e., 1-5 cells in width) with both yellow plastids and red pigment. In normal flowers, these areas are scattered throughout the epidermis, indicating red stripes and spots. In some cases, such red spots and stripes are confined to the upper surface. In both surfaces, the subepidermal layers are yellow, but the tissue between the epidermis is also dotted with yellow ceils (Figure 7-B). Red pigment and yellow plastids are found in the epidermis of the staminodia of periclinal chimeras producing totally red flowers from bronze sectors in inflorescence stalks (Figure 7-C). There are also very small stray cells in the two epidermis surfaces that contain only yellow plastids and a lack of red colour. In addition to yellow plastids on both the upper and lower surfaces, stray yellow cells can also be found in the middle parenchymatous layer. This also applies to 'King Humbert' flowers (134).

The same distribution of pigments is found in the chimeral staminodia of both types of flowers described above. A pure yellow flower does not contain red pigment, but there are yellow plastids found in the surface epidermis, the sub-epidermis and the parenchymal cells between them (Figure 7-A). A single inflorescence of this cultivar is capable of transitioning from pure yellow to different degrees of red spotting and finally, to pure red flowers. The mechanism of this is not understood but we hypothesise that a transposon is responsible for the over-expression of the anthocyanin pathway in red sectors. When this transposon jumps out in a cell at the L2 layer of the meristem, all cells derived from that cell display lower expression of anthocyanin therefore the leaf tissue is

green and flower tissue is yellow in the absence of high level of anthocyanin (134).

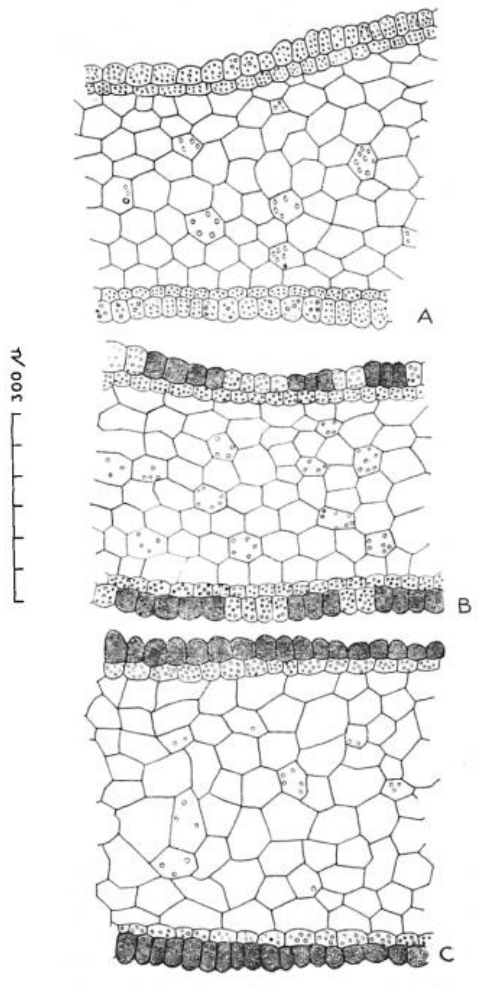


Figure 7: A transverse section of staminodia from *C. indica* cv Cleopatra (134).

Showing red pigment distribution (shaded ceilings) and yellow plastid distribution (hollow circles) **A:** Completely yellow. **B:** Red with yellow stripes and spots. **C:** Completely red flowers

1.6.7 Leaf phenotype of *Canna Indica* cv Durban

The leaves of *C. indica* cv Durban have multicolour leaves, which is largely bronze with pink, red, carmine, green, yellow, burgundy, gray-purple, orange, and deep grey stripes. The cultivar is also known as ‘Phasion’ and *Canna Tropicanna* and is one of the most popular *Canna* varieties because of its attractive leaves. Its flowers are large and orange in colour and its average height is 1.5m (166).

The botanical name of Durban is *Canna* hybrida and also known as "Phasion.". Jan H. Potgieter, discovered the new cultivar in Bethal, South Africa, by open-pollinating *Canna* cv ‘Wyoming’. Potgieter then selected the new cultivar from its progeny primarily based on its unique leaf colour. Through successive generations of asexual reproduction, horticultural study and comparative trials have shown that the characteristics of “Durban” will remain firmly fixed (168) The leaf phenotype vary significantly with changes in the environment including temperature, light intensity and day length without any changes in the genotype. The leaves contrast spectacularly with its orange flowers. Furthermore, *Durban* has shorter flower stems and shorter leaves than does Wyoming, along with a slower growth rate and smaller size (168). *C. indica* cv Durban leaf is illustrated in Figure 8.

Figure 8: *C. indica* cv Durban leaf.

1.6.8 Leaf phenotype of *Canna Indica* cv Pretoria

Canna 'Pretoria', also known as *Canna* 'Bengal Tiger' is a variegated *Canna* cultivar from the Italian group. The average height of the plant is 190 cm and of the foliage is 140 cm. It has ovoid foliage with variegated yellow veins on a green background; its stems and leaves grow upright, with gentle spreading. An edge of maroon surrounds the leaf, while the staminodes, labellum, and stamens are orange, occasionally blushing reddish, with yellow edges. The stigma is a deep red-orange colour; the petals are strongly flushed red. The flowers look crumpled and have a low fertility level, as do the seeds and pollen, while the rhizomes are thick, being up to 3 cm in diameter (132). *C. indica* cv Pretoria leaf is illustrated in Figure 9.

Figure 9: *C. indica* cv Pretoria Leaf.

1.7 Aim of the Project

The aim of the project is to study colour formation in the leaves of *C. indica* plants by two approaches. First, to test the hypothesis that the red-green leaf colour is determined by the absence/presence of a transposon in a key anthocyanin pathway related gene. The main observation that supports this hypothesis is that the boundary between green and red sectors is very sharp, suggesting that it is not caused by the diffusion of a compound but by a difference between the genomes of green and red cells. The second supporting observation is that there are green and red lines on the stem and leaves/flowers that develop in the path of green lines appear green/yellow, respectively while red leaves/flowers develop in the path of red lines. The translocation of a TE in the L2 layer of the shoot meristem could explain these observations, because the red colour appears in the epidermis

cells, which develops from the L2 layer. Since it is extremely unlikely that a TE would always jump into the same gene or a gene that regulates the anthocyanin pathway, our hypothesis is that the jumping out that causes the phenotype. It is known that TEs can either suppress or activate gene expression depending on the location of the TE, therefore we have to account for two possible scenarios: (1) the TE suppresses the expression of a gene required for anthocyanin production and in this scenario it would be present in/near to a gene in the green epidermal cells or (2) the TE enhances the expression of a gene required for anthocyanin production and in this scenario it would be present in/near to a gene in the red epidermal cells. To test our hypothesis, I will study the green and red sectors of *C. indica* cv Cleopatra leaves. Total RNA will be extracted from red and green sectors and the expression of all mRNAs will be measured by RNAseq. The bioinformatic analysis will be done by Simon Moxon (UEA, Norwich, UK) and the differentially expressed mRNAs will be further investigated by BLAST. Because of the uncertainty in the function of the TE (either suppression or enhancement), it is not possible to have an expectation about the nature of expression pattern, other than the gens should be involved in anthocyanin production and differently expressed between red and green sectors.

The second part of the project aims to identify the role of small non-coding RNA molecules in colour formation. To achieve this, cDNA libraries of small RNA will be generated from green and red sectors of *C. indica* cv Cleopatra leaves and also from leaves of *C. indica* cv Durban and *C. indica* cv Pretoria. The libraries will be sequenced by next generation sequencing. The bioinformatic analysis of the sequence data will be carried out by Danial Richardson who will search for differentially expressed small RNAs between the different samples. Differential expression of small RNAs will be validated by Northern blot. If differentially expressed miRNAs will be identified, we will predict target mRNAs for those

miRNAs. Then 5'RACE analysis will be used for validating the predicted target mRNAs.

1.8 Objectives

- Searching for transposon sequences in green and red sectors. Experimental validation by PCR of any possible gene disruption by transposons. This is described in chapter 3.
- Purify sufficient quantity and quality of gDNA from green and red sector of *C. indica cv Cleopatra* for PacBio sequencing. This is described in chapter 4.
- Small RNA libraries will be sequenced from green, red, green with yellow stripes and red with pink stripes leaves. Differentially expressed microRNAs will be identified by bioinformatics analysis and their targets will be predicted and experimentally validated. This is described in chapter 5.

Chapter 2: Methods and materials

2.1 Plant Material and Growth Conditions

2.1.1 *Canna indica* Plants

Canna indica plants were purchased from Urban Jungle (Norfolk, UK) and grown in a growth room (CER) for a few weeks at 22°C under long day conditions (8h dark /16h light). The *Canna* plants were watered twice per week during the winter and autumn and three times per week during summer and spring. Three cultivars of *C. indica* were used in this study: *C. indica* cv Cleopatra, *C. indica* cv Pretoria and *C. indica* cv Durban.

2.2 Buffers

All buffers are given at final concentrations

2.2.1 DNA Extraction Buffer

Genomic DNA (gDNA) extraction buffer for PCR was 200 mM Tris-HCl at pH 7.5, 250 mM NaCl, 25 mM Ethylenediaminetetraacetic acid (EDTA), and 0.5% sodium dodecyl sulphate (SDS).

2.2.2 Sorbitol Wash Buffer

Sorbitol wash buffer used in the extraction of high molecular weight DNA from *C. indica* cv Cleopatra for Pacific BioSciences (PacBio) sequencing was 5 mM EDTA at pH 8.0, 100 mM Tris-HCl at pH 8.0, 0.35 M sorbitol (Fisher Bioreagents), and 1% (w/v) polyvinylpyrrolidone (average molecular weight 40,000; PVP40). Prior to use in DNA extraction, 2-mercaptoethanol (1% v/v) was added.

2.3 Media

2.3.1 Lysogeny Broth (LB) Media

LB media was prepared by mixing 1 L distilled water, 10 g tryptone (Melford), 5 g yeast extract (Invitrogen) and 5 g sodium chloride (Sigma-Aldrich) in a 2 L flask, which was then autoclaved to sterilise the media.

2.3.2 LB Agar

LB agar was prepared by mixing 200 mL of LB media, 4 g of agar in a 500 mL flask, which was then sterilised and cooled to solidify the agar solution.

2.4 RNA Methods

2.4.1 Total RNA Extraction

The leaf was weighted after being cut by scalpel and placed in a falcon tube 50 mL before being submerged in liquid nitrogen. A total of 2 grams of *C. indica* cv Cleopatra was used per green and red sectors of leaves to extract the total RNA by using 20 mL of Tri Reagent Solution (Thermo Fisher Scientific, AM9738) as directed by the manufacturer. The nucleoprotein complexes were dissociated by incubation for 5 minutes in a fume hood. Each reaction was divided into 12 of the 2 mL tubes in 800 µL aliquots. In the next step, 200 µL chloroform was added to each tube to separate the RNA from the DNA and proteins. The RNA was recovered from the upper aqueous phase to a new 2 mL tube, and 400 µL of isopropanol was added to precipitate the RNA for two hours at -80 °C. The RNA was washed two times with 950 µL 80% ethanol and air dried before dissolving in 10 µL distilled water. The final step was combining the content of all tubes into one tube (keeping the samples from red and green sectors apart from each other). A NanoDrop 8000 spectrophotometer (Thermo Fisher Scientific) was used to

measure the RNA concentration. Gel electrophoresis was used to test the integrity of the RNA by separating them on a 1.5% (w/v) agarose gel. One µg of RNA was run in 0.5 X Tris-Borate-EDTA (TBE) buffer. The gels were stained in 10 mg/mL ethidium bromide solution (Fisher Scientific). A Typhoon FLA 9500 scanner was used for image capture. The total RNA was submitted to the Earlham Institute (Norwich, UK) for RNA-seq analysis.

2.4.2 RNA- seq Data Analysis

Five µg total RNA was sequenced using next generation sequencing technologies at Earlham Institute to determine differentially expressed mRNAs and Simon Moxon (UEA, Norwich, UK) analysed the data. Basic Local Alignment Search Tool (BLAST) analysis was then conducted to determine whether MYB and bHLH TFs were differentially expressed between red and green sectors. The read length for the sequencing was Illumina 150 base pair, paired-end sequencing.

2.4.3 Construction and Sequencing of Small RNA Libraries

2.4.3.1 RNA Extraction for sRNA Library

A miRVana miRNA isolation kit (AM1560, Ambion) was used to clean the total RNA samples to prepare the sRNA libraries. First, the total RNA samples were diluted to 50 µL. Five volumes of lysis/binding buffer were added to the RNA, and a 1/10th volume of miRNA homogenate additive was then added. After mixing the samples, they were placed on ice for 10 minutes.

After 10 minutes, 1.25 volume of 100% ethanol was added and thoroughly mixed. After centrifuging at 10,000 g for 30 seconds, the ethanol and lysate mixture was filtered through a miRVana filter cartridge. After removing the flow-through, 700

µL of miRNA wash solution 1 was added to the column. The column was washed twice with miRNA wash solution 2 and 3 using 500 µL each time. The empty column was spun for 1 minute at 10,000 rpm to remove all wash buffers. The columns were transferred to new collection tubes and 50 µL of 95°C elution solution, was added. The column was soaked for 2 minutes and centrifuged at full speed for 30 seconds. The final elute volume was 100 µL after adding another 50 µL of preheated elution solution. Afterwards, 100 µL of purified RNA was mixed with 10 µL of 3M NaOAc (S2889, Sigma-Aldrich) and three volumes of 100% ethanol and precipitated overnight at -80°C.

The precipitated RNA was pelleted by centrifuging at 4°C at full speed for 30 minutes. After removing the supernatant, the pellets were washed with 80 % ethanol. The pellets were air dried after one final centrifugation at 10,000 rcf to remove the remaining ethanol. The pellets were resuspended in 15 µL of nuclease-free water and then was quantified using a NanoDrop 8000 (Thermo-Fisher Scientific).

2.4.3.2 3’ Adapter Adenylation

The protocol for preparing sRNA libraries was published by Xu et al,. 2015 (169). The adenylation of the 3’ HD adapter was performed using a New England Biolabs (E2610L, NEB) 5’ DNA Adenylation kit, according to the manufacturer’s directions. Oligo Clean and Concentrator™ (D4061, Zymo-Research) was used to clean up this reaction then a NanoDrop 8000 spectrophotometer (Thermo Fisher Scientific) was used to measure concentration. The nuclease-free water was added to the eluted adapter to achieve a concentration of 10 µM. Adenylated and non-adenylated adapters were separated on a 16% urea polyacrylamide gel to verify adenylation. A successful

adenylation occurs if the adenylated sample runs slower than the adapter sample, due to increased molecular weight.

2.4.3.3 Ligation of Small RNA Adapters and PCR

A total of 2 µg of miRVana purified total RNA was ligated to the adenylated 3' HD adapter followed by a RecJ exonuclease (M0264S, New England Biolabs) treatment. Adapter/adapter ligation products are reduced by RecJ exonuclease, as mentioned previously (169). T4 RNA ligase (T4Rnl1, NEB) was used to ligate the RNA-3' adapter product with the 5' HD adapter followed by a cleaning step using Oligo Clean and Concentrator™ (D4061, Zymo-Research). Generate the cDNA strand. Several PCR programs were used in this protocol to amplify cDNA, varying in the number of cycles used. These libraries were amplified with Phusion High-Fidelity DNA Polymerase using 12, 14 and 16 cycles of Phusion DNA Polymerase (F530L, Thermo Scientific™).

2.4.3.4 Selection of Library Size

After the DNA was amplified by PCR and run on 8% polyacrylamide gels, the band size of the library was selected as described by Xu et al., (2015) (169).

2.4.3.5 Library Normalisation for sRNAs

Each library was prepared for loading into the sequencer by normalising it and loading the same quantities. The Novex Hi-Density TBE Sample Buffer (LC6678, Invitrogen) was mixed with 1 µL of each completed library and run on an 8% polyacrylamide gel. A Typhoon FLA 9500 (GE Healthcare Life Sciences) imaging system was used after the gels were stained with SYBR Gold nucleic acid gel stain (S11494, Invitrogen). The library intensity was quantified with

Image Quant (GE Healthcare) and normalised to one library. As a result of the value produced, the 1 µL loaded was adjusted and then combined with the Novex sample buffer for each library. Following the adjustment of the volumes, the libraries were run on a further 8% polyacrylamide gel and imaged again, then pooled and sent for sequencing if the libraries were similar in intensity.

2.4.3.6 Small RNA Library Sequencing

Illumina NextSeq550 was used to sequence the normalised sRNA libraries at approximately 33 million single end reads each.

2.5 DNA Methods

2.5.1 Genomic DNA Extraction for PCR Using DNA Extraction Buffer

One g of the *C. indica* cv Cleopatra leaves (from green or red sectors) were placed in sterile 1.5 mL tubes and immerse the tubes in liquid nitrogen then ground with the pellet pestles. Followed by the addition of 500 µL of DNA extraction buffer (200 mM Tris-HCl at pH 7.5, 250 mM NaCl, 25 mM EDTA, 0.5% SDS) and then vortexed for 5 seconds. The ground plant tissues were incubated in a water bath at 60°C for 30 minutes, with mixing by vortexing halfway through. Chloroform (Fisher Scientific) was added to an equal volume and centrifuged for 5 minutes at 15,000 rcf at 4°C when phase separation was clearly visible.

The aqueous supernatant was transferred to new 1.5 mL tubes, equal volume of isopropanol (I9516, Sigma-Aldrich) was added to the volume of the aqueous phase. The DNA sample was incubated for 30 minutes at −20°C to precipitate the DNA and centrifuged for 5 minutes at 15,000 rcf at 4°C. A solid pellet was formed at the bottom of the tube.

Eighty percent ethanol was used to wash the pellet after removing the supernatant. Autoclaved distilled water was used to resuspend the sample after the ethanol was removed. To remove starch and other polysaccharides, DNA from each sample was first centrifuged for 5 minutes at 15,000 rcf and then incubated on ice for 2 minutes. The supernatant was transferred to sterile 1.5 mL tubes and measured using NanoDrop 8000 spectrophotometer (Thermo Fisher Scientific).

DNA was cleaned using the Genomic DNA Clean & Concentrator kit, following the manufacturer's protocol (Zymo Research). A NanoDrop spectrophotometer was used to measure the DNA concentration (Thermo Fisher Scientific).

2.5.2 Genomic DNA Extraction for MYB PCR Using DNeasy Plant Mini Kit

The DNA was extracted from *C. indica* cv Cleopatra red and green sectors by using DNeasy Plant Mini Kit (Qiagen) according to the manufacturer's instructions. A total of 0.02 g of *C. indica* cv Cleopatra for each red and green sector was used. The pellet was eluted in 100 µL buffer AE.

2.5.3 Unsuccessful Protocols for Extraction Genomic DNA for PacBio

2.5.3.1 Treatment the DNA with the Nucleon Phytopure kit

The initial attempt to extract the gDNA from *C. indica* cv Cleopatra red sector by using Nucleon phytopure kit was carried out by following the manufacturer's instructions. The DNA pellet was eluted with 50 µL TE buffer. The DNA was cleaned by the gDNA Clean & Concentrator kit in accordance with the manufacturer's instructions with a minor change. Pre-heating the elution buffer at 70°C and 10 µL elution buffer was added and incubated 2 minutes at room temperature before centrifuging. This step was repeated one more time.

2.5.3.2 Treatment of the DNA with RNAse A and Proteinase K

The Nucleon Phytopure kit was used for extracting gDNA from *C. indica* cv Cleopatra red sector, accordance of the manufacturer's instructions. The DNA was treated with 10 µL RNAse A (Thermo Scientific) during the extraction, which was after adding 4.6 mL Reagent 1, followed by incubation at 37°C for 30 minutes. Then, 75 µL proteinase K (600 U/mL) (Thermo Scientific) was added after Reagent 2 and incubated at 65°C for 10 minutes.

2.5.3.3 Treatment of the DNA with the Phenol:chloroform:isoamyl Alcohol

Other attempts used the same previous method with slight modifications. After RNAse A (Thermo Scientific) treatment, 1 volume of the Phenol:chloroform:isoamyl alcohol (25:24:1) (Sigma-Aldrich) was added and mixed by vortexing. This was followed by 5 minutes centrifugation at high speed and then the DNA was purified using the Genomic DNA Clean & Concentrator kit, following the manufacturer's protocol (Zymo Research).

2.5.3.4 Treatment of the DNA with the Sorbitol Wash Buffer

After breaking the cell wall, DNA was washed three times in 25 mL sorbitol wash buffer containing 250 µL 2-mercaptoethanol to improve gDNA quality. After that, the Nucleon Phytopure kit was used. RNA was degraded by adding 10 µL RNAse A (Thermo Scientific) and incubating at 37°C for 30 minutes. 1.5 mL of reagent 2 was added to the solution and mixed with 75 µL proteinase K (600 U/mL) (Thermo Scientific) and incubated at 65°C for 10 minutes. 1 volume of the Phenol:chloroform:isoamyl alcohol (25:24:1) (Sigma-Aldrich) was added and the DNA was purified as described in 2.5.3.3.

2.5.3.5 Reducing Starch Content in Leaves

C. indica cv Cleopatra leaves were covered with a black plastic bag with holes in it (Figure 10). treatment is crucial for reducing the production of carbohydrates, polysaccharides, and polyphenolic compounds (170). After a few days (3-5 days) DNA was extracted by the Nucleon Phytopure kit. The DNA was washed three times in 25 mL sorbitol wash buffer mixed with 250 µL 2-mercaptoethanol to improve gDNA quality. RNAse A (Thermo Scientific) and proteinase K (600 U/mL) (Thermo Scientific) were added to the solution. 1 volume of the Phenol:chloroform:isoamyl alcohol (25:24:1) (Sigma-Aldrich) was added. In the DNA precipitation step, the DNA was washed three times with 1 mL of 70% ethanol chilled. After resuspending the DNA pellets with 60 µL TE buffer, they were incubated overnight at room temperature in the shaker.

Figure 10: Various sized leaves of *C. indica* cv Cleopatra are shown in both pictures wrapped in black plastic bags with some holes to reduce the accumulation of polysaccharides.

2.5.4 Extraction Genomic DNA for PacBio Sequencing

In order to break the plant cell wall, 1g *C. indica* (fresh weight) leaf tissue was placed into a mortar, liquid nitrogen was poured into it and the leaves were left in the mortar until all liquid nitrogen evaporated. The frozen leaves were ground consistently for 90 seconds with a pestle. That step was repeated 6 times. The powder was transferred to a 15 mL centrifuge tube.

To aid the removal of polyphenols and polysaccharides the samples were washed with sorbitol wash buffer. 25 mL Sorbitol wash buffer (as described above in the buffer section) and 2-mercaptoethanol (1% v/v) were added to the ground powder, inverted and vortexed to mix thoroughly. The tube was centrifuged at

4,000 rcf for 5 minutes at 4°C. The supernatants were carefully discarded. Two more washes were performed in the same manner.

Cell lysis was carried out using a Nucleon PhytoPure DNA Extraction kit (GE Healthcare) following the manufacturer's protocol with minor changes. 4.6 mL of reagent 1 was added, followed by 10 µL RNase A (Thermo Scientific), and then incubated for 30 minutes at 37°C. 1.5 mL of reagent 2 and then 75 µL of Proteinase K (600 U/mL) (Thermo Scientific) were added, followed by incubation at 65°C for 10 minutes. The sample was then chilled on ice for 20 minutes.

300 µL of PhytoPure DNA Extraction Resin was added to the previous solution. One volume of phenol:chloroform:isoamyl alcohol (25:24:1) (Sigma-Aldrich) was added to the solution. The tube was mixed at 4°C for 10 minutes on a tube rotator. The sample was centrifuged at 3,000 rcf for 10 minutes. The upper phase was placed into fresh 2 mL tubes.

One volume of chilled isopropanol (Fisher Scientific) was added to the solution, and the mixture was inverted and incubated at 20°C for 15 minutes. The solution was centrifuged for 5 minutes at 40,000 rcf. Three washes with 1 mL of chilled 70% ethanol were performed. To remove residual phenol (and ethanol/isopropanol) from the tube, it was placed in a fume hood for 30-60 minutes. The DNA from the tubes was combined into one tube. The quality of the DNA was analysed using a NanoDrop (Thermo Fisher Scientific), and the quantity of the DNA was assessed using a Qubit (Thermo Fisher Scientific).

2.5.5 Genomic DNA Data Analysis

This work had done by David Prince (UEA, Norwich, UK). The *C. indica* cv Cleopatra PacBio reads were *de novo* assembled using Canu to produce a draft genome (171). The RNA-seq reads were mapped to the genome using HISAT2 and a reference-guided transcriptome produced using StringTie2 (172,173). Kallisto program was used to pseudoalign the RNA-seq reads to the transcriptome and differentially expressed genes determined using DESeq2 in R (174,175). MYBs were annotated using MYB_annotator (176). TEs were predicted for the draft *C. indica* cv Cleopatra genome using EDTA software (177).

2.6 PCR

MYB gene disruption by transposons was experimentally tested using PCR. PCR primers were designed for the MYB genes along with CHS and SDN2-1 as control primer of *A. thaliana* and they are listed in the Table 1.

Table 1: MYB, CHS Gene Primer Sequences and the control primers of *A. thaliana.*

Primer Name	Sequence (5`>3`)
***Canna* 1FW** ***Canna* 1R**	CTTACATGCAGTCGACATTGGGA CGAGACACCAAACTATTCAGTTCATACAAA
***Canna* 2FW** ***Canna* 2R**	TACAATGAGGAGCCCCTGCTG ATGCTACGTACTTAACTGAAAACTAGTTTACATTGTAAAA
***Canna* 3FW** ***Canna* 3R**	AGCTTACAAAGGAAGATGTTGGAAGAAAA ACTTTATATTATGAGGATTTTGCTGTTCGCT
***Canna* 4FW** ***Canna* 4R**	TGTCTCGCTTCCCTGCCTTG GACCTAATACTATATTAAGTTTGGTTCATCATCATCAC
***Canna* 5FW** ***Canna* 5R**	CCCCAACTCTCTTACTCCCCTTCT ATCAGTTAGTAACACACAGAATCACATTAAGATCAC
CHS FW **CHS R**	AGCCAGTTCCCCACGCACACG TAGGTGCAATCGCACGCCATTAAC
SDN2-1 FW **SDN2-1 R**	ATTTCGATCTTTGTCGGTTCC CATAAACTCTGTCTTTGGCGC

2.6.1 PCR Amplification of Differentially Expressed MYB Genes

Several PCR mixtures and programmes were used to detect potential transposons that may be present in the MYB genes differently expressed in the red and green sectors. PCR was performed using a 25 µL reaction volume. The 25 µL reaction contained the following: 17.8 µL of H_2O, 3 µL of 5X Phusion HF buffer, 0.5 µL of dNTPs (10 mM each), 1 µL of 10 µM Forward Primer, 1 µL of 10 µM Reverse Primer, 0.75 µL of Dimethyl Sulfoxide (DMSO), 0.7 µL of DNA (10 ng/ µL) and 0.25 µL of Phusion DNA Polymerase (2.5 U/ µL, New England Biolabs). The PCR program was the following: 94°C for 2 minutes; 30 cycles of 94°C for 30 seconds, 58°C for 30 seconds and 72°C for 5 minutes; and a final cycle of 72°C for 10 minutes. The PCR reactions were repeated at different annealing

temperatures including 50°C, 53°C, 58°C and 60°C. For each PCR, the products were separated on 1.5 % agarose gels for 60 minutes.

Furthermore, multiple PCR attempts were conducted. PCR master mixes were prepared using the previous PCR component with minor changes. Various amounts of DNA were used: 10 µL, 8 µL and 6 µL and the PCR reactions were run at different annealing temperatures: 62°C, 67°C, 68°C, and 71°C.

In addition, PCR master mix was made as the previous one but without DMSO. The PCR was performed using a 22.8 µL reaction volume. The 22.8 µL reaction contained the following: 10.95 µL of H_2O, 4.3 µL of 5X Phusion HF buffer, 0.4 µL of dNTPs (10 mM each), 1 µL of 10 µM Forward Primer, 1 µL of 10 µM Reverse Primer, 5 µL of DNA (10 ng/ µL) and 0.2 µL of Phusion DNA Polymerase (2.5 U/ µL, New England Biolabs). Using the previous PCR components, a new master mix was prepared, and the volume of DNA was changed to 2 µL. The PCR annealing temperature was performed at two different temperatures in each time, 62°C and 67°C.

Each Phusion PCR reaction consisted of 50 µL. Five serial dilutions (1:1) of the red leave sector of the DNA template with water were used. 3 µL of H_2O, 10 µL of 5× Phusion HF Buffer (Thermo Scientific), 1 µL of 10 mM dNTPs, 1 µL of 10 µM Forward Primer, 1 µL of 10 µM Reverse Primer, 0.5 µL of Phusion DNA Polymerase (F530L, Thermo Scientific) and 33.5 µL of template DNA. The PCR program for primer number 1,2,5 and 6 were 94°C for 2 minutes; 30 cycles of 94°C for 30 seconds, 62°C for 30 seconds and 72°C for 5 minutes; and a final cycle at 72°C for 10 minutes. While the PCR program of the primers number 3 and 4 were the same as the previous one with annealing temperature 58°C. Afterward, the reactions were run for 1 hour at 120 V on a 1.5% (w/v) agarose

gel. Typhoon FLA 9500 images (GE Healthcare Life Sciences) were captured after the gel was stained with ethidium bromide (Fisher Scientific).

2.7 Small RNA Northern Blots

2.7.1 Urea Polyacrylamide Gel Electrophoresis for Separating RNA

The RNA samples were treated with one volume of Gel Loading Buffer II (AM8546G, Thermo Fisher Scientific) prior to electrophoresis, followed by 2 minutes of heating at 70°C to denature the secondary structures. The gel was prepared with 15% polyacrylamide containing urea. A gel was made by mixing 2.1 g of urea (10142740, Fisher Scientific) with 1.25 mL of deionised water and 0.5 mL of 5X TBE and then dissolving it in a microwave for 20 s.

After cooling, the gel was mixed with 19:1 acrylamide/bis solution (1610144, Bio-Rad) and then with 2.5 µL tetramethylethylenediamine (TEMED) (110-18-9, Sigma-Aldrich), followed by 50 µL 10% ammonium persulfate solution (17874, Thermo-Fisher). Glass plates with 1mm spacer (1651824 by Bio-Rad) were used to pour the gel. A 1 mm comb (1653359, Bio-Rad) was inserted into the gel and allowed to set.

A Mini-Protean Tetra Cell (185-8000, BioRad) was used to run the denatured RNA on a 15% urea polyacrylamide gel until the dye front reached the bottom. To run the gel, 0.5X TBE was used as a buffer. The gels were stained for 5 minutes with ethidium bromide (239-45-8, Fisher Scientific) and then scanned using a Typhoon FLA 9500 scanner (GE Healthcare Life Sciences).

2.7.2 RNA Transfer to Nylon Membranes

A 9 cm × 7 cm square of Whatman filter paper (3030-335, Thermo Fisher Scientific) was pre-soaked in water. A 9cm × 7cm square of Amersham Hybond-NX membrane was cut for the gel (GE Healthcare Life Sciences) and pre-soaked in water. On the base plate of a semidry apparatus (Fisher), three pieces of soaked Whattman filter paper were stacked, and the air bubbles were removed with a sterile serological pipette. This was followed by placing the soaked Hybond-NX membrane on top. The urea polyacrylamide gel was then applied. On top of the gel, also three pieces of soaked Whattman filter paper were added, and the pipette was rolled again to remove any air bubbles.

A constant voltage of 20 V was applied to the semidry apparatus for 90 minutes after the top plate was placed on the apparatus. When this period elapsed, the apparatus was disassembled, and a Typhoon FLA 9500 was used to re-image the urea polyacrylamide gel. The transfer was considered successful if no RNA was visible on the gel.

2.7.3 Chemical Cross Linking

A 50 mL falcon tube was then filled with 10 mL of dH_2O, 10 mL of concentrated HCl (H1758, Sigma-Aldrich) and 122.5 mL of 12.5 M 1-methylimidazole (M50834, Sigma) for the crosslinking reaction. Then, 0.373 g of 1-ethyl-3-(dimethylaminopropyl) carbodiimide hydrochloride (EDC) (22980, Thermo-Fisher) dissolve with 10 mL from the previous rection was in a 15 mL falcon tube. Afterwards, H_2O was added until 12 mL is reached and shake to dissolve the EDC.

A piece of Whattman filter paper cut to 9 cm × 7 cm was placed on a sheet of saran wrap and soaked with 5 mL of cross-linking solution. The soaked membrane was covered with the saran wrap, then placed this in the oven (Thermo) at 60°C for 90 minutes with RNA side facing up.

After the cross-linking procedure, the saran wrap was taken off. The top left corner was cut to mark the orientation of the Hybond-NX membrane. The Hybond-NX membranes were washed with dH_2O in a petri dish and then shaken for 10 minutes.

2.7.4 Hybridisation

In each hybridisation, 2 µL of γCTP-32P (Perkin Elmer), 2 µL of 10 µM sRNA specific DNA oligo (Sigma) without 5' phosphate and T4 polynucleotide kinase (PNK) (M0201S, New England Biolabs) were added and measured according to the manufacturer's recommendations. The first initial hybridisation, the U6 probe was added first on the membrane and then miRNA probe. The second initial hybridisation, the miRNA probe was added first and then U6 probe to avoid having a strong signal from U6. The reaction was carried out at 37°C for 60 minutes. A list of all probes used is presented in Table 2.

Table 2:The probes of sRNA libraries

Candidates - MiRNAs	Sequences
MIR529- FIRST STRAND	AGAAGAGAGAGAGTACAGCCT
MIR529 – SECOND STRAND	GCTGTACCCTCTCTCTTCTTC
MIR156/MIR157	GCTCTCTATGCTTCTGTCATC
MIR6300	GTCGTTGTAGTATAGTGGTGAGTATTCCC
MIR530	TGCATTTGTACCTGCACCTAA
MIR528	CCTGTGCCTGCCTCTTCCACG
MIR171	CGAGCCGAACCAATATCACTC
MIR408	TGCACTGCCTCTTCCCTGGCT
MIR397	TTGAGTGCAGCGTTGATGAGA
MIR166	CGGACCAGGCTTCATTCCCC
MIR162	TCGATAAACCTCTGCATCCG

About 5 mL of ULTRAhyb-Oligo buffer (AM8663, Thermo Fisher) was added to the cross-linked membranes at 37°C, and the membranes were rotated in the hybridisation oven (Thermo). In the tube, the RNA side of the membrane was positioned facing inward. To create a volume of the 50 µL labelled probe solution, 30 µL of dH_2O was added. Then, 50 mL of this solution was added to the membranes and tubes containing a prehybridisation solution. Great care was taken not to apply any probe solution directly to the membranes. Incubation was performed overnight at 37°C with spinning.

To wash the membranes, a Saline Sodium Citrate (SSC) buffer and 0.1% Sodium Dodecyl Sulphate (SDS) were prepared. A 50 mL wash buffer was used to wash the membranes four times. The membranes were incubated at 37°C for 20 minutes each time. After hybridizing the membranes, they were placed on a Fujifilm plate (Fujifilm) at 4°C for at least 30 minutes. Following exposure, a Typhoon FLA 9500 was used to scan the Fujifilm plate.

2.7.5 Membrane Stripping

A 0.1% SDS strip solution was prepared to remove the probes from the membranes. Incubation was carried out for 60-120 minutes at 85°C, with spinning of the membranes to be stripped. Following incubation, Geiger counters were run across the membranes to ensure that there was no detectable radioactivity above the background level to determine if they had successfully been stripped.

2.7.6 Data Interpretation

Using Image Quant (GE Healthcare) software, the pixels of each band was quantified after imaging. Normalisation was performed by selecting one sample (usually the control) and setting its band intensity to 1. The intensity of all remaining bands was normalised to this and calculated as a ratio of the normalising sample. The loading control was performed using a U6 probe. Similarly, Image Quant was used to measure the intensity of U6 on each membrane, giving the same value of 1 to the same sample as in the previous hybridisation. The first hybridisation intensity values were divided by the U6 intensity values for each band. Each northern blot Figure presents these values.

2.8 5’ Rapid Amplification of CDNA Ends (5’-RACE)

2.8.1 Isolation of MRNA

To determine whether DN31218 and DN40098 mRNAs were cleaved by miR166 and whether DN43985 was cleaved by miR530, I performed a modified 5’-RACE experiment. Total RNA was extracted from 2 g green *C. indica* cv Cleopatra leaf using the methods previously described in “Total RNA Extraction”. The mRNA

was isolated from the total RNA using a Dynabeads mRNA purification kit (Invitrogen, 61006), following the manufacturer's instructions.

2.8.2 Ligating the Adapter to the 5' End of Cleaved MRNAs

The 5' ends of accessible mRNA fragments were ligated to the GeneRacer RNA Oligo. 6.5 µL of mRNA was mixed with 1.25 µL of GeneRacer RNA Oligo. The mixture was incubated at 65°C for 5 minutes to relax the RNA secondary structures, followed by cooling on ice for 2 minutes. A total of 1 µL of 10× ligase buffer, 0.6 µL of RNaseOut (Invitrogen, 50 µ/ µL) and 1 µL of T4 RNA ligase (Ambion, 5 µ/ µL) were added to the RNA, mixed and incubated at 37°C for 1 hour. Afterward, mRNA was cleaned with an RNA cleaning and concentrator kit (Zymo Research). The RNA was eluted with 12 µL of distilled water.

2.8.3 Reverse Transcription-PCR

SuperScript II (Invitrogen) reverse transcriptase was used following the manufacturer's protocol with the following minor changes. The concentration of the mRNA was 888.6 ng/ µ l and 10 µL ligated mRNA, 1 µL of GeneRacer Oligo dT primer (50 µM) and 1 µL of dNTPs (10 mM) were used. These reactions were incubated at 65°C for 5 minutes. Then, 4 µL 5× first strand buffer, 1 µL of 1 M DTT, 1 µL of RNaseOut (40 µ/ µL), and 1 µL (200 units) of SuperScript II (Invitrogen) were added and incubated at 50°C for 1 hour and deactivated at 70°C for 15 minutes.

2.8.4 Touch-Down PCR

Gene-specific primers (GSPs) and nested GSPs for the candidate target genes were designed using Primer-BLAST software (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). The predicted PCR product of the nested GSP for a candidate target gene was designed to be at least 50 base pairs shorter than the predicted product of the GSP. The primer sequences specific for the candidate target genes and GeneRacer Oligo are listed in Table 3. For each candidate target genes, two PCR reactions were set up. Using the GeneRacer Oligo, the fragmented transcripts of interest were amplified specifically with touch-down PCR.

Using a thermocycler, the following reactions were carried out. The PCR program included the following: 98°C for 2 minutes; 5 cycles of 98°C for 30 second, 73°C for 40 seconds; 5 cycles of 98°C for 30 second, 71°C for 40 second; 5 cycles of 98°C for 30 second, 69°C for 30 second; and 25 cycles of 98°C for 30 second, 68°C for the 30 second, 72°C for 1 minute; final step of 72°C for 10 minute.

To visualise the amplification products 19 µL of the gene-specific PCR was loaded in a 1.5% (w/v) agarose gel and scanned using a Typhoon FLA 9500 (GE Healthcare Life Sciences). The bands of interest were cut from the gel using a sterile scalpel, and the gel was recovered using the Zymoclean Gel DNA Recovery kit (D4002) according to the manufacturer's instructions. The volume of DNA elution buffer was 8 µL. The DNA was measured with a NanoDrop 8000 spectrophotometer (Thermo Fisher Scientific). The fragments were cloned using a pJET1.2/blunt cloning vector (50 ng/ µL) and transformed into *Escherichia* coli (DH5α).

2.8.5 Colony PCR

Colony PCR was used to identify whether the colonies contained an insert in the pJET1.2 vector by using GoTaq PCR master mix. The PCR was performed using a 50 µL reaction volume included: 10 µL 5X GREEN Go Taq Flexi Buffer, 2 µL $MgCl_2$ (25 mM), 1 µL dNTP (10 mM), 2.5 µL pJET1.2 Forward Sequencing Primer (10 µM), 2.5 µL pJET1.2 Reverse Sequencing Primer (10 mM), 0.25 µL Go Taq DNA polymerase (5u/µl), 31.75 µL Nuclease-Free water. The PCR process included the following: 95°C for 2 minutes; 35 cycles of 95°C for 30 seconds, 60°C for 1minute and 72°C for 1 minute; and a final cycle of 72°C for 5 minutes. In order to purify DNA, colonies were grown in ampicillin (100 ug/ml) and LB-broth (100 ug/ml) for 16 hours in the shaking at 37°C.

2.8.6 Clone Sequencing Analysis

For each clone, 4 mL of overnight culture was pelleted by centrifuging at top speed for 3 minutes at room temperature. The plasmid DNA was extracted using the QIAprep Spin Miniprep Kit (Qiagen), following the manufacturer's instructions. Each plasmid was sequenced using the forward and reverse primer. The plasmid DNA was sent to Eurofins MWG Operon (Germany) for sequencing.

Table 3: Primers used for 5'RACE.

Primer name	MicroRNA	Candide target gene	Sequence (5' > 3')
MiR166_DN31218_GSP	MiR166	DN31218	ACAAATGGTTGGGGCATAAACCTGC
MiR166_DN31218_nested_GSP	MiR166	DN31218	TCCATTATTTGCAGTTGGCAGCAC
MiR166-DN40098_GSP **MiR166_DN40098_nested_GSP**	MiR166 MiR166	DN40098 DN40098	GACCACCAGTTGCAGGAGTCAGC CCCTCCACTACCGGTGGGAAT
MiR530_ DN43985_GSP **MiR530_DN43985_nested_GSP**	MiR530 MiR530	DN43985 DN43985	ACCATGCAAAAGAGGTCGGCTGTG TCTCTGCTGGTACTACACTCACCAC
GeneRacer™ 5′ Primer **GeneRacer™ 5′ Nested Primer**	N/A	N/A	CGACTGGAGCACGAGGACACTGA GGACACTGACATGGACTGAAGGAGT
Forward pJET 1.2 sequencing primer	N/A	N/A	CGACTCACTATAGGGAGAGCGGC
Reverse pJET 1.2 sequencing primer	N/A	N/A	AAGAACATCGATTTTCCATGGCAG

Chapter 3: Identification of Differentially Expressed Genes Between Red and Green Sectors of *C. indica* cv Cleopatra Leaves

3.1 Introduction

Our hypothesis is that the red-green leaf colour in *C. indica* cv Cleopatra is determined by the absence and presence of a transposon in a TF gene. The hypothesis is based on the observation that in some red leaves of *C. indica* cv Cleopatra green sector appears randomly, sometimes resulting in mostly green leaves with red sectors. The transposons are most likely responsible for that as the boundary between red and green is extremely sharp (Figure 6B). The purpose of this chapter is to identify the gene(s) that is/are responsible for the red colour of the leaf. According to the hypothesis, in the red sector, the genes in the anthocyanin pathway, that are responsible for the red colour, are transcribed to mRNAs, which then translate to proteins. In the green sectors, one of those gene genes may contain a transposon and therefore the mRNA may not be stable or even if it is stable, it would be translated into a different protein, which would not be functional, therefore the leaf remains green. Alternatively, the hypothesised transposon may be present in the gene(s) in the red sectors and absent in the gene(s) in the green sectors if the function of the gene is to suppress the anthocyanin pathway (Figure 11). The hypothesis is that the transposon was in a MYB or bHLH TFs, because both these classes of TFs have been shown to regulate the anthocyanin biosynthetic pathway (178). At the time that the project started there was no *Canna* genome available. Therefore, to identify differentially expressed genes, RNA was extracted, and the cDNA libraries were sequenced and analysed from red and green sectors of the leaf. We identified differentially expressed genes and some of them have been predicted to be MYB and bHLH TF genes based on homology to known MYB and bHLH genes. Additionally, we

also investigated some of the gene's coding for enzymes in the anthocyanin biosynthetic pathway, including CHS, which is the first and a key enzyme in this pathway (58). DNA was extracted using multiple methods to produce DNA of sufficient quality for PCR and PCR was performed to determine if there were differences in bands size produced between DNA extracted from red and green sectors using primers specific to the differentially expressed genes.

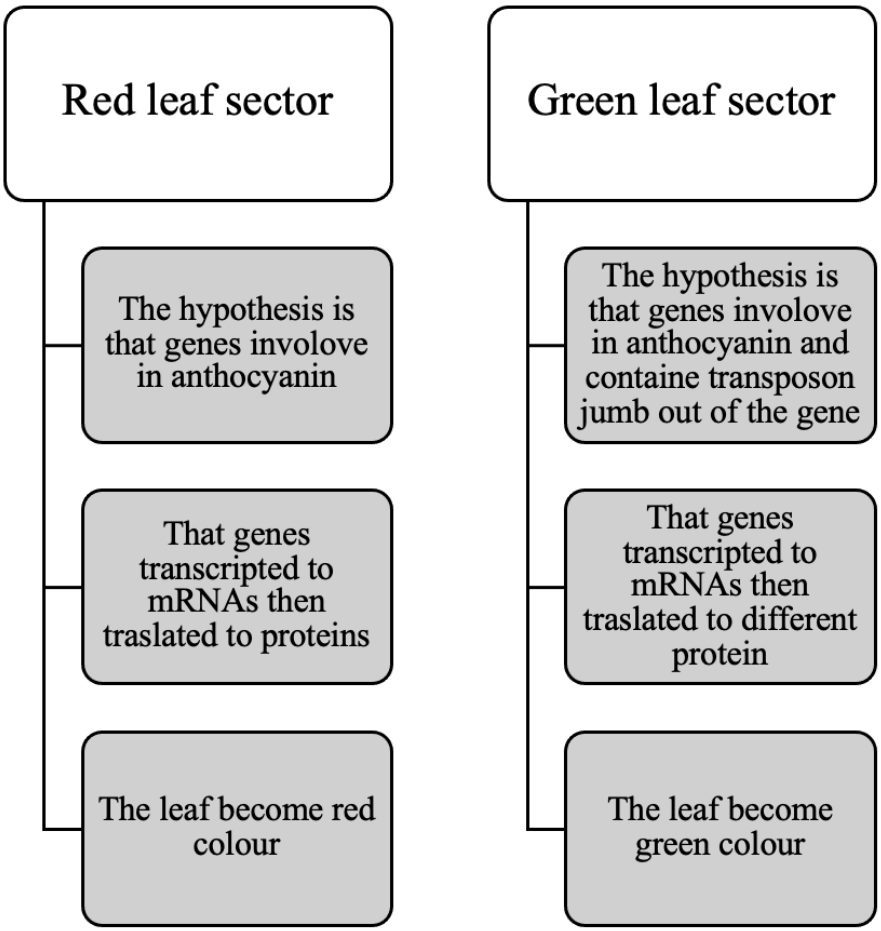


Figure 11: Schematic Diagram of the hypothesis for Red and Green Sectors of *C. indica* cv Cleopatra Leaves.

3.2 Result

3.2.1 RNA Extraction for RNA-seq

Total RNA was extracted from three biological replicates of green and red sectors of *C. indica* cv Cleopatra leaves. It was found that 2 g of leaf material was sufficient to produce the required quantity of RNA for RNA-seq. Total RNAs were sequenced by next generation sequencing technologies at Earlham Institute by RNA-seq using the paired-end Illumina sequencing kit. The results were analysed by Simon Moxon (UEA, Norwich, UK). Simon made a *de novo* transcriptome from the RNA-seq data, as there was no genome or annotation available at the time and generated a list of differentially expressed mRNAs. Using a two-fold difference threshold, he found about 500 differentially expressed mRNA. I then used BLAST to determine if any of the differentially expressed genes were similar to MYBs and bHLHs genes. An example of a MYB gene is shown in (Figure 12, A) and an example of a bHLH gene is shown in (Figure 12, B). Overall, there were five MYBs and two bHLHs genes amongst the differentially expressed genes.

(A)

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
PREDICTED: Musa acuminata subsp. malaccensis myb-related pr...	Musa acu...	640	640	28%	4e-178	86.50%	2315	XM_009382799.2
PREDICTED: Musa acuminata subsp. malaccensis myb-related pr...	Musa acu...	640	640	28%	4e-178	86.50%	2322	XM_009382798.2
PREDICTED: Musa acuminata subsp. malaccensis myb-related pr...	Musa acu...	564	564	29%	3e-155	83.87%	2173	XM_018829292.1
PREDICTED: Musa acuminata subsp. malaccensis myb-related pr...	Musa acu...	564	564	29%	3e-155	83.87%	2257	XM_018829291.1
PREDICTED: Musa acuminata subsp. malaccensis myb-related pr...	Musa acu...	551	551	21%	2e-151	89.19%	2079	XR_001978371.1
PREDICTED: Musa acuminata subsp. malaccensis myb-related pr...	Musa acu...	551	551	21%	2e-151	89.19%	2048	XM_018827108.1

(B)

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
PREDICTED: Musa acuminata subsp. malaccensis transcription ...	Musa acu...	662	662	67%	0.0	82.52%	1088	XM_009411475.2
PREDICTED: Musa acuminata subsp. malaccensis transcription ...	Musa acu...	597	597	66%	1e-165	81.30%	991	XM_009382455.2
PREDICTED: Zingiber officinale transcription factor bHLH35-like ...	Zingiber o...	475	475	65%	7e-129	78.97%	995	XM_042574711.1
PREDICTED: Zingiber officinale transcription factor bHLH35-like ...	Zingiber o...	475	475	65%	7e-129	79.00%	919	XM_042566816.1
PREDICTED: Zingiber officinale transcription factor bHLH35-like ...	Zingiber o...	464	464	63%	1e-125	79.23%	1040	XM_042566811.1
PREDICTED: Zingiber officinale transcription factor bHLH35-like ...	Zingiber o...	444	444	60%	2e-119	79.14%	923	XM_042566815.1

Figure 12: The BLAST result for two differentially expressed genes.

A: The BLAST result of a putative MYB gene. **B:** The BLAST result of a putative bHLH gene.

3.2.2 PCR of Differentially Expressed MYB and bHLH Genes

To amplify the putative MYB and bHLH TFs and assess the presence/absence of a transposon, DNA was extracted from red and green sectors, and PCRs were performed using gene specific primers that were supposed to amplify the entire open reading frames. Two methods were tried for extracting DNA from both red and green leaf sectors of *C. indica* cv Cleopatra, in order to get the cleanest samples possible for performing the PCR. The first DNA approach used the DNA

extraction buffer technique without any commercial kit. The DNA pellet from red sectors were very viscous as shown in Figure 13. The DNA pellet ratio of 260/280 was below 1.7 because the DNA was contaminated with salts and proteins due to high levels of secondary metabolites and polysaccharides. The good DNA pellet 260/280 ratio is expected to be above 1.7. I cleaned DNA with the Genomic DNA Clean & Concentrator kit, which improved the DNA pellet quality as the ratio of 260/280 was above 1.7. Table 4 showed the result of the DNA quality before and after cleaning the DNA with the Genomic DNA Clean & Concentrator kit based on the result of the nanodrop.

Table 4: DNA quality before and after cleaning by the Genomic DNA Clean & Concentrator kit based on the nanodrop

DNA	**A260**	**A280**	**A260/280**	**A260/230**	**Ng/µL**
DNA before cleaning	4.169	4.079	1.02	0.50	208.4
DNA after cleaning by the Genomic DNA Clean & Concentrator kit	2.361	1.237	1.91	1.31	118

Figure 13: Pellet of DNA extracted from *C. indica* cv Cleopatra red sector.

Several amounts of the DNA concentrated 118 ng/ µL were used (82.6 ng, 708 ng, 944 ng and 1180 ng) and several annealing temperatures (50°C, 53°C, 58°C, 60, 62°C, 67°C, 68°C, and 71°C) were used for PCR amplification of the putative MYB and bHLH genes, but no bands were seen. I used *A. thaliana* DNA with DNS2 primer as a positive control that produced the expected band, suggesting that the problem was the quality of the *Canna* DNA or the primers specific to the *Canna* gene. Due to the negative result of the PCR I tried to improve the DNA quality by the using DNeasy Plant Mini Kit.

The purity of the DNA pellet obtained using the DNeasy Plant Mini Kit was still low based on the NanoDrop reading as an example shown in Table 5, because the 260/280 ratio was still below 1.7. Then, several PCR was performed at several annealing temperatures (58°C, 60°C and 65°C) together with the *A. thaliana* as a positive control with DNS2 primer. There were no bands for the MYB and CHS genes in either red or green sectors. As there was no band produced by the PCR from this DNA method, we returned to the first method of extracting DNA and did some modifications for the PCR master mix.

Table 5: Nanodrop reading of the DNA extracted using the DNeasy Plant.

DNA	A260	A280	A260/280	A260/230	Ng/µL
DNA using the DNeasy Plant Mini Kit	2.974	3.574	0.83	0.46	148.7

Although the first and second extraction methods for the DNA and the subsequent PCRs, did not allow for the amplification of the genes of interest, we extracted DNA again by using the first method of the DNA extraction buffer. The master mix of the PCR was slightly modified as follows: 5 µL of the DNA template of the green leaf sectors were incorporated into the master mix and the DMSO was not added to the PCR master mix. The PCR containing primer pair 6 was run with an annealing temperature of 66 °C, whereas the PCR for primer pairs 1 to 5 were run with an annealing temperature of 62°C. The 6 PCR reactions were run on a 1.5% (w/v) agarose gel for 1 hour at 120 V. The DNA bands of the green leaf sectors were observed in the gel documentation system (Figure 14).

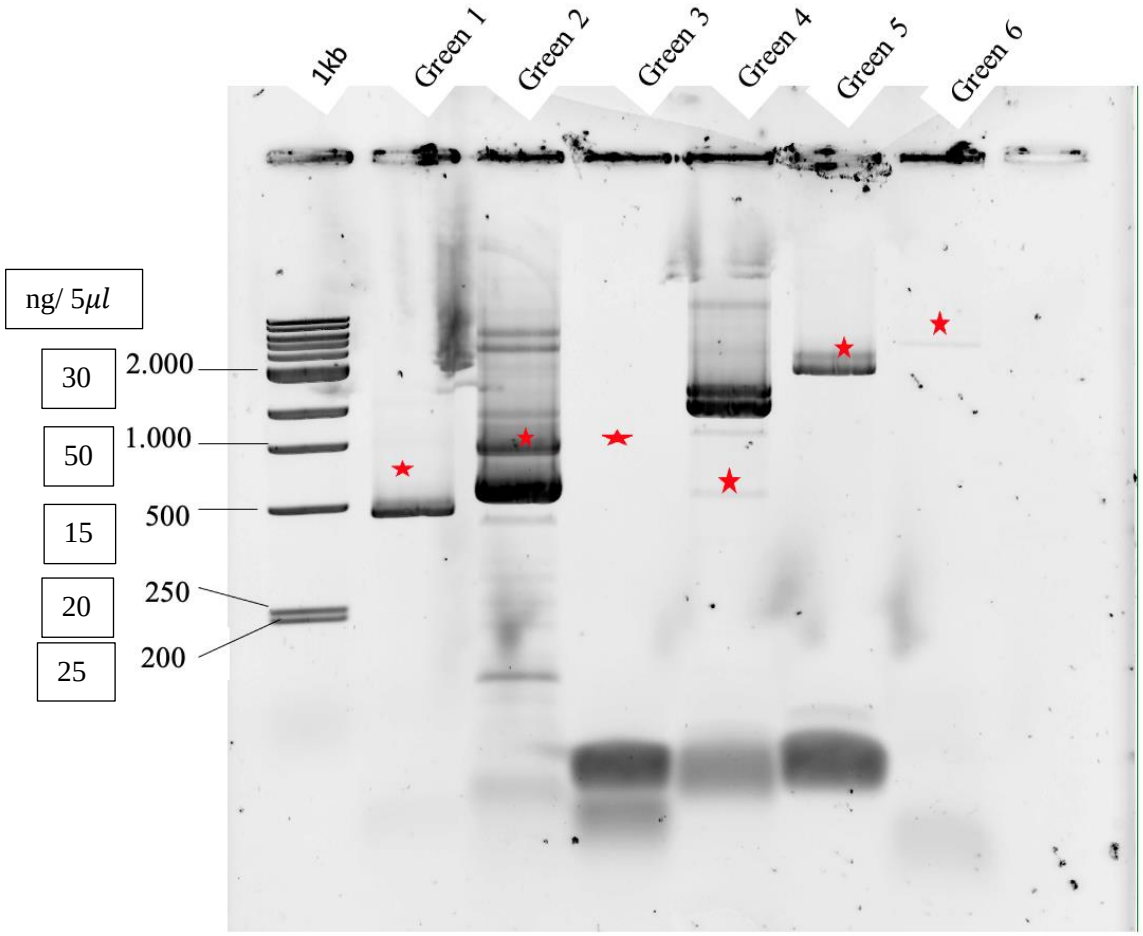


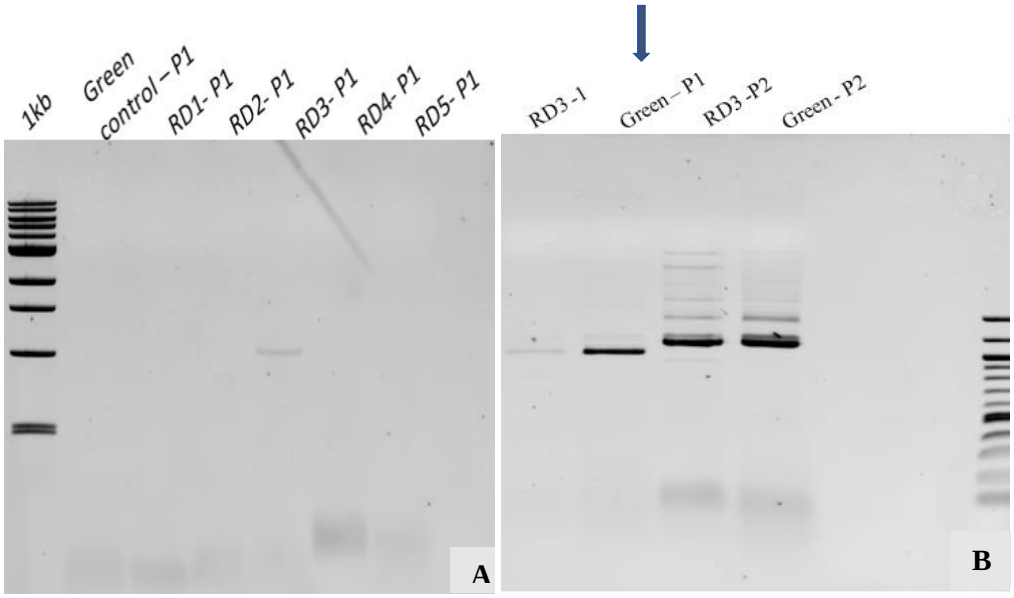
Figure 14: PCR analysis of DNA extracted from green sectors of *C. indica* cv Cleopatra leaves with primer pairs 1 to 6.

Green 1: green leaf sector of the *C. indica* cv Cleopatra with primer forward and reverse number 1. Green 2: green leaf sector of the *C. indica* cv Cleopatra with primer forward and reverse number 2. Green 3: green leaf sector of the *C. indica* cv Cleopatra with primer forward and reverse number 3. Green 4: green leaf sector of the *C. indica* cv Cleopatra with primer forward and reverse number 4. Green 5: green leaf sector of the *C. indica* cv Cleopatra with primer forward and reverse number 5. Green 6: green leaf sector of the *C. indica* cv Cleopatra with primer forward and reverse number 6. Red stars indicate the size of the expected bands.

To optimise the PCR for DNA extracted from red sectors of *C. indica* cv Cleopatra leaves, five serial dilutions (1:1) of the DNA samples were used. The

PCR reactions containing primers pairs 1, 2, 5 and 6 were at annealing temperature of 62°C, while reactions with primers pairs 3 and 4 were run at annealing temperature of 58°C.

The diluted DNA samples from red sectors were used for PCR with each primer pairs. The DNA from green sectors was used as a positive control. The PCR reaction containing primer pair number 1, DNA bands were observed for dilution number 3 (Figure 15, A). In addition, for primer pair number 2, DNA bands were observed for dilution number 3 (Figure 15, B). For primer pair number 3, DNA bands were observed for dilutions number 4 and 5 (Figure 15, C). For primer pair number 4, DNA bands were observed for dilution number 4 (Figure 15, D). For primer pair number 5 and 6, DNA bands were observed for dilutions number 4 and 5 (Figure 15, E - F).



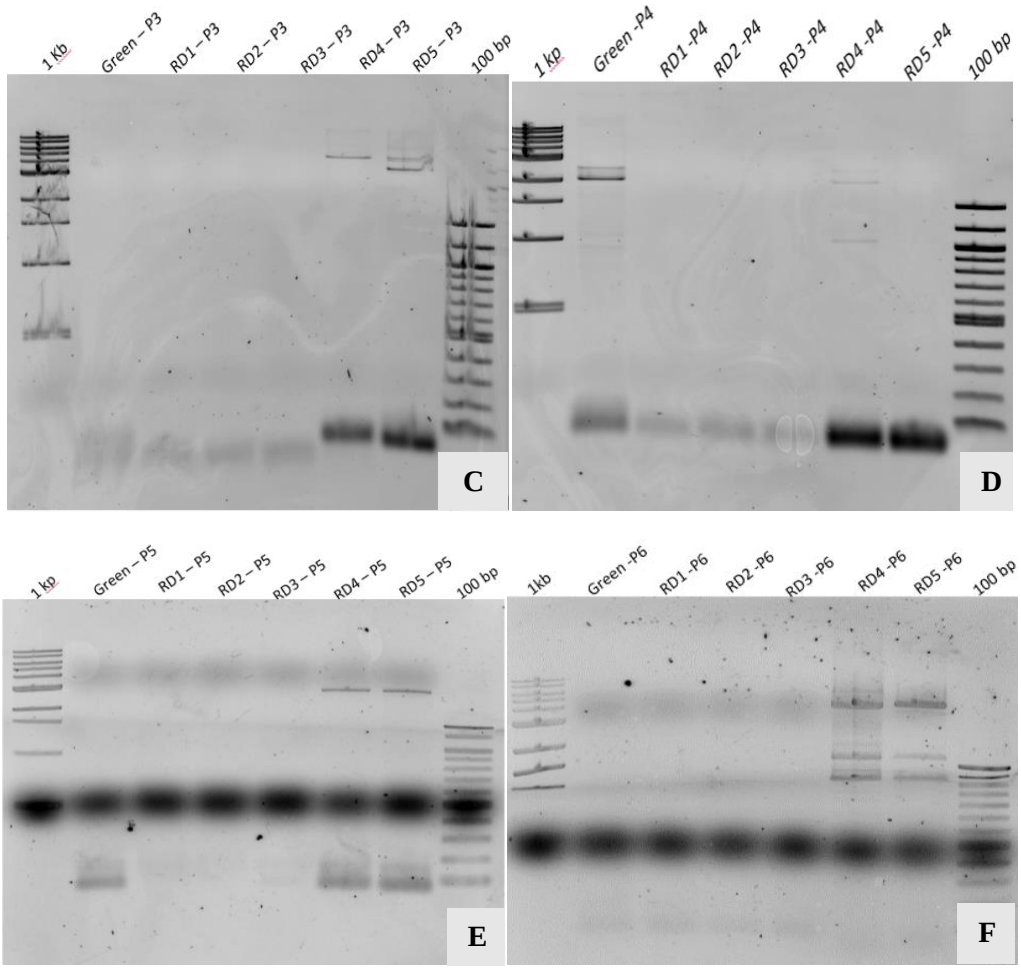


Figure 15: PCR analysis of candidate genes.

A: *C. indica* cv Cleopatra genomic DNA from red (RD) sectors was diluted and the target gene was amplified with primer pair 1. **B**: *C. indica* cv Cleopatra genomic DNA from red (RD) sectors was diluted and the target gene was amplified with primer pair 2. **C:** *C. indica* cv Cleopatra genomic DNA from red (RD) sectors was diluted and the target gene was amplified with primer pair 3. **D:** *C. indica* cv Cleopatra genomic DNA from red (RD) sectors was diluted and the target gene was amplified with primer pair 4. **E:** *C. indica* cv Cleopatra genomic DNA from red (RD) sectors was diluted and the target gene was amplified with primer pair 5. **F:** *C. indica* cv Cleopatra genomic DNA from red (RD) sectors was diluted and the target gene was amplified with primer pair 6.

Finally, DNA samples that produced bands for both red and green samples were analysed by PCR with each primer pair and the results combined on a single agarose gel. The band size with each primer pairs was the same with no

differences between green and red sectors. Therefore, we concluded that there was no transposon for the tested MYB and bHLH genes differently expressed between red and green sectors (Figure 16), or at least they were not detectable with the primers I used. It is impossible to test all genes by PCR because there were about 500 differentially expressed genes. The strategy therefore changed in chapter 4, where we sequenced the entire genome of *C. indica* cv Cleopatra to try to identify the hypothesised transposon.

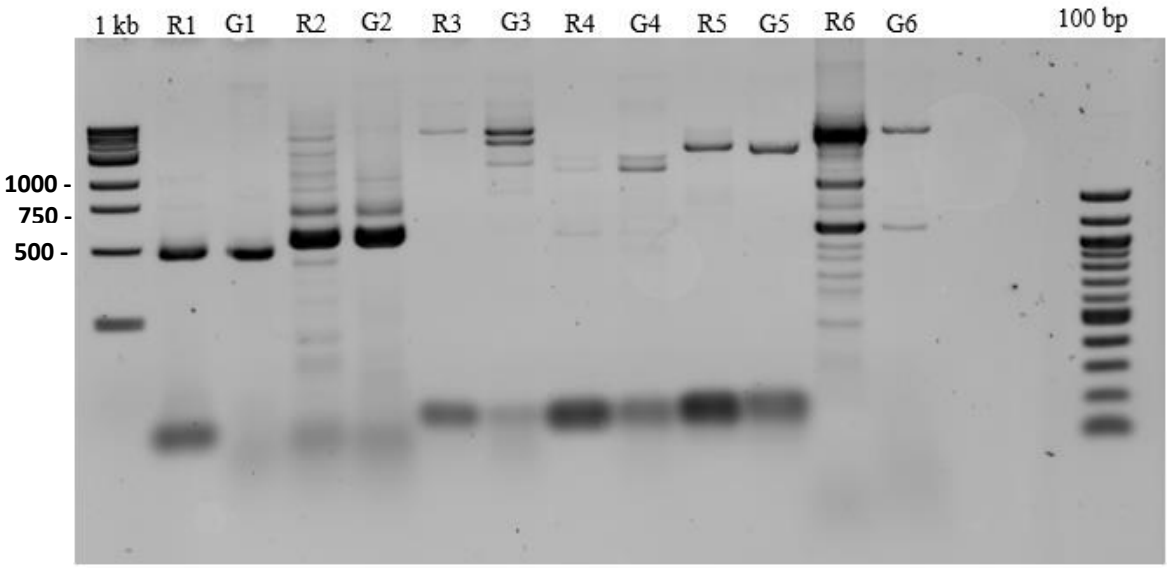


Figure 16: The DNA bands of *C. indica* cv Cleopatra both red and green sectors.

* The Numbers above the bands refer to primer pair numbers, **R:** red and **G:** green.

3.3 Discussion

In this chapter, we tried to identify the genes responsible for the red colour of *C. indica* cv Cleopatra and determine whether they contain a transposon that causes the green leaf phenotype. The gene expression and relationships between biological processes can be revealed by RNA-seq technology under a variety of

conditions, identifying key genes (179). To achieve this, RNA was sequenced from both green and red sectors of *C. indica* cv Cleopatra leaves. We identified MYB and bHLH genes that may be responsible for the red colour of the leaf. Five MYBs and two bHLHs TFs were found among the differentially expressed genes using BLAST. To further investigate these MYB and bHLH genes, along with CHS, we extracted DNA and attempted to PCR amplify these genes to reveal any potential size difference due to the hypothesised transposon insertion. Due to the poor quality of the DNA first we changed the DNA extraction method and used the DNeasy Plant Mini Kit (Qiagen). The DNA pellet of the *Dimorphandre mollis* leaf also had a similar issue, being very sticky and highly viscous, and the brownish pellets indicated phenol contamination (180). Therefore, poor quality of extracted genomic DNA may be a common problem for several species with high starch content.

The PCR did not produce any bands from either the first or the second DNA extractions. This indicated that the chemical compounds in the leaves may be responsible for the absence of bands (180) Previous work attributed the absence of PCR product to high content of chemical compounds, including flavonoid which binds to the DNA upon lysis that oxidized the DNA by phenolic metabolites (180,181).

Amplification can be enhanced by adding DMSO to PCR. Secondary structures can be reduced with DMSO but can be problematic when dealing with GC-rich templates (182). In our work conducting PCR, the DMSO may be responsible for absence of the bands, therefore we tried to run the PCR reactions for DNA from green sectors in the absence of DMSO. Since this approach worked, we kept that procedure for all following PCRs.

However, the bands of the red leaf sector PCR were not present even when DMSO was removed from the PCR mix. We hypothesised that flavonoid contamination was higher in DNA extracted from red leaves and therefore diluting the DNA would reduce the concentration of flavonoids, but the samples would still contain sufficient amount of DNA as it was shown before (183). After diluting the DNA samples from red sectors, bands for both red and green sectors were clearly visible. To compare the size of bands between red and green leaf sectors, we visualised the amplification products on an agarose gel. The result of the PCR showed no difference between the DNA from the red and green sectors therefore we could not find a transposon in the selected genes. However, the hypothesis could still be right, but the transposon could be in other gene that was not identified by the RNAseq or the transposon may be outside of the coding regions. Therefore, we developed a new alternative approach based on genome sequencing, which is described in chapter 4.

Anthocyanin synthesis pathways in plants are controlled by MYB genes, which often act synergistically with structural genes (184). It was found that TFs bHLH and MYB were detected to be upregulated among Red President flowers of the *Canna* (185). That study might be similar for what we found in the *C. indica* cv Cleopatra leaf. Both studies found MYB and bHLH genes were higher in the red flower and leaf of *Canna*. In this case, MYB genes might be responsible for the red colour by contributing together with the bHLH. At the time, no studies have been conducted on the colouring of *C. indica* leaf.

Chapter 4: Whole Genome Sequencing of *C. indica* cv Cleopatra

4.1 Introduction

In the previous chapter (Chapter 3), there was no evidence of a transposon in the differentially expressed MYB and bHLH genes identified, or in the CHS gene. Therefore, we developed an alternative approach, which was made possible by PacBio sequencing. There was no available gDNA sequencing of *Canna* species at the time, therefore we aimed to sequence the entire genome *C. indica* cv Cleopatra with PacBio sequencing to identify the gene that contain transposon that leads to an upregulation of the anthocyanin pathway. The technology produces read sequences longer than 200k and is suitable for sequencing the whole genome of *C. indica*. Establishing the genome of *C. indica* is helpful for the search for the transposon that causes the red leaf color. To accomplish this, we needed to produce sufficient quantities and quality of gDNA from red sector of *C. indica* cv Cleopatra leaves for PacBio sequencing.

4.2 Results

The gDNA sample to be sequenced with PacBio sequencing at the Earlham Institute (Norwich, UK) had to meet specific requirements as provided by colleagues at the Earlham Institute. Table 5 shows the details of the requirement.

Table 6: Quality requirements of gDNA samples for PacBio sequencing.

Criteria	Requirement
Quantity of DNA (µg)	20
Minimum concentration (ng/ µL)	50
Volume (µL)	50 - 400
A 260/280	1.6 - 2.0
A 260/230	1.8 - 2.4
Material size (<40Kb)	< 20%

4.2.1 Treatment of the DNA with the Nucleon Phytopure kit

The DNA was extracted using the Nucleon Phytopure kit following the manufacturer's instruction. The pellet was resuspended in 50 µL of TE buffer and it was viscous and had contamination based on the Nanodrop reading. This DNA was cleaned using the gDNA Clean & Concentrator kit, following the manufacturer's protocol. The DNA was resuspended with preheated 20 µL elution buffer and quantified using a Nanodrop. The DNA was sent to the Earlham Institute for quality control (QC). The sample failed the QC check because the DNA contained high level of RNA, polysaccharides and polyphenol. Table 7 shows the ratio of the DNA based on the result of the Nanodrop before and after cleaning the DNA by the gDNA Clean & Concentrator kit.

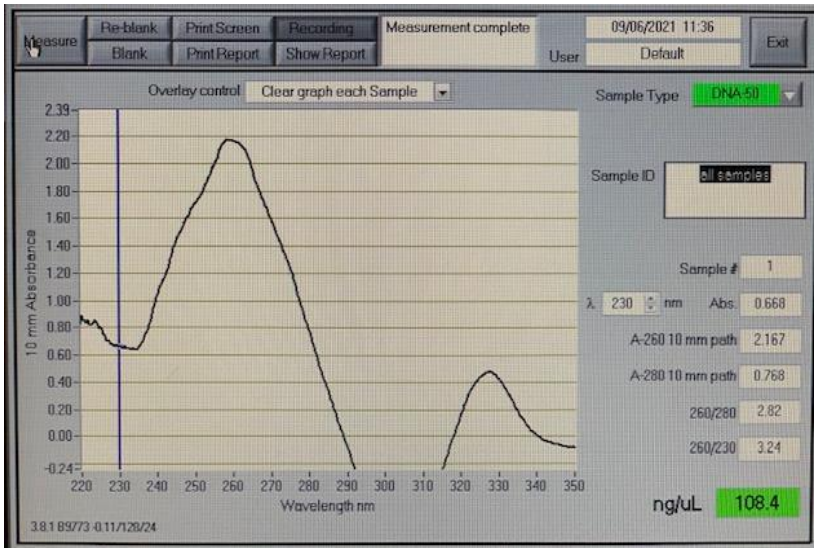
Table 7: DNA ratio based on the Nanodrop.

DNA	A260	A280	A260/280	A260/230	Ng/µL
DNA before cleaning	0.961	0.396	2.42	3.53	48
DNA after cleaning	6.205	2.919	2.13	2.52	310.2

4.2.2 Treatment of the DNA with RNAse A, Proteinase K and phenol:chloroform:isoamyl alcohol

DNA was extracted from *C. indica* cv Cleopatra red leaves using the Nucleon Phytopure kit as before and following the manufacturer's protocol with some changes to try to improve the quality. The sample was treated with RNAse A and proteinase K to remove RNA and protein contamination, then it was cleaned with phenol:chloroform:isoamyl alcohol. The phenol/chloroform traces were removed by using the gDNA Clean & Concentrator kit, and the concentration of DNA measured by Nanodrop, (Figure 17, A). Furthermore, the DNA was measured by Qubit 4 reading (Figure 17, B). This gDNA sample also failed the Earlham Institute QC check as the DNA was still contaminated of polysaccharides and polyphenol, and 72% of the DNA fragment size were less than 40 kb.

(A)



(B)

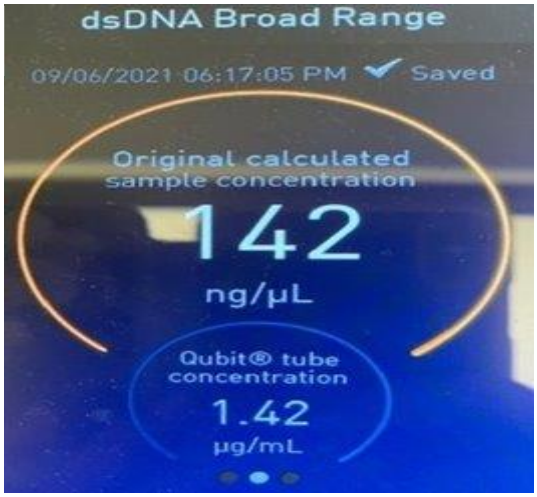


Figure 17: Quality of DNA extracted using RNAse A, proteinase K, and phenol:chloroform:isoamyl alcohol treatment

A: Concentration of the DNA quantified by Nanodrop. **B:** Concentration of the DNA quantified qubit 4.

4.2.2 Treatment of Leaf Samples with Sorbitol Wash Buffer

Next, a new DNA extraction was performed including three washes of sorbitol wash buffer to remove the polyphenols and polysaccharides, before continuing with the Nucleon Phytopure Kit, RNAse A, proteinase K and phenol:chloroform:isoamayl alcohol protocol used previously. The DNA was

measured by Nanodrop. Table 8 shows the quality of the DNA before and after treatment. In addition, the extracted DNA was also measured by Qubit 4 (Figure 18). There was an improvement in the sample's purity as the ratios of 260/280 and 160/230 was below 2 and better than the previous sample. The DNA was sent to the Earlham Institute to check the quality. The sample did not meet the QC requirements as the gDNA was contaminated with polyphenols and polysaccharides.

Table 8: The DNA quality with sorbitol wash buffer treatment based on the Nanodrop.

DNA	A260	A280	A260/280	A260/230	Ng/µL
DNA before treatment	5.671	3.266	1.74	0.99	283.5
DNA after treatment	1.665	0.920	1.81	1.84	83.3

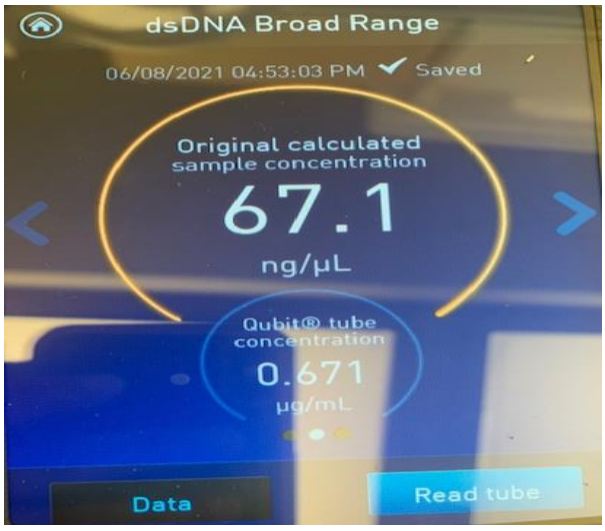


Figure 18: Concentration of the DNA quantified by using qubit 4

4.2.3 Dark Treatment to Improve DNA Quality

The red leaves of *C. indica* cv Cleopatra were covered with a black plastic bag with some air holes for 3-5 days to reduce the accumulation of polysaccharides (170). A sample was extracted by the Nucleon Phytopure Kit, sorbitol wash buffer, RNAse A, proteinase K, phenol:chloroform:isoamayl alcohol treatment. The DNA quality was much improved with covering the leaves based on the Nanodrop Table 9.

Table 9: The result of the DNA with dark treatment based on the Nanodrop.

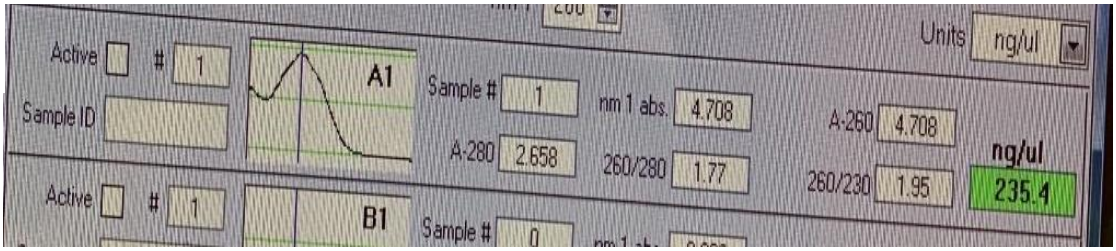
DNA	A260	A280	A260/280	A260/230	Ng/µL
DNA with dark treatment	4.533	2.494	1.82	1.91	226.7

4.2.4 Increasing the Length of the Sample Grinding Time of the gDNA Extraction

After consulting with colleagues at the Earlham Institute, the method used above was further optimised by using small leaves for the leaf samples and increasing the time for grinding the samples from 2 minutes to 20 minutes. The longer the period of grinding frozen leaves, the more likely the cell wall will be broken successfully. The DNA concentration and purity were measured by Nanodrop and Qubit 4 (Figure 19, A). In addition, the extracted DNA were measured Qubit 4, (Figure 19, B). The sample was clean, and both 260/230 - 260/280 ratios were in the optimal range. The DNA was sent to the Earlham Institute to assess the quality and finally the sample passed the QC check (Figure 19. C, Table 6). The

Earlham Institute carried out the PacBio sequencing, and the data was analysed by David Prince (UEA, Norwich, UK).

(A)



(B)

(C)

Sample ID					DNA Qubit	Nanodrop			FEMTO Pulse				
Sample	Sample ID	Sample Name	Customer Ref	Well ID	ng/µl	ng/µl	260/280	260/230	% 0-39,999bp	% 40,000 - 400,000bp	Vol. of sample (µl)	Total Amount (µg)	OVERALL Pass/Fail
1	A64328	R0588-S0009	20220113canna	N/A	188.0	245.9	1.80	1.93	17.1	83.1	151	28.4	Passed

Figure 19: Concentration of the DNA quantified by using Nanodrop

A: Concentration of the DNA quantified by using Nanodrop. **B:** Concentration of the DNA quantified by qubit. **C:** The result of the gDNA from Earlham Institute which was passed.

4.2.5 Analysis of the gDNA Sequencing Data

As analysing PacBio sequencing data requires specialised bioinformatic skills, this part of the project was done by David Prince (UEA, Norwich, UK). The *C. indica* cv Cleopatra PacBio reads were *de novo* assembled using Canu to produce a draft genome. The Canu tool is specialized in assembling PacBio's long read, produce highly continuous assemblies and more accuracy (171). The RNA-seq reads were mapped to the genome using HISAT2 and a reference-guided transcriptome produced using StringTie2. Kallisto was used to pseudoalign the RNA-seq reads to the transcriptome and differentially expressed genes determined using DESeq2 in R. In StringTie2, transcripts are assembled from genome-aligned RNA-seq reads (186). MYBs were annotated using MYB_annotator. Two of the differentially expressed genes were annotated as MYBs. TEs were predicted for the draft *C. indica* cv Cleopatra genome using EDTA software. One of the two MYB differentially expressed genes was predicted to contain an intact TIR in an intron. Figure 20 illustrate the analysis process for gDNA sequencing data. The transposon position in the gene (Figure 21).

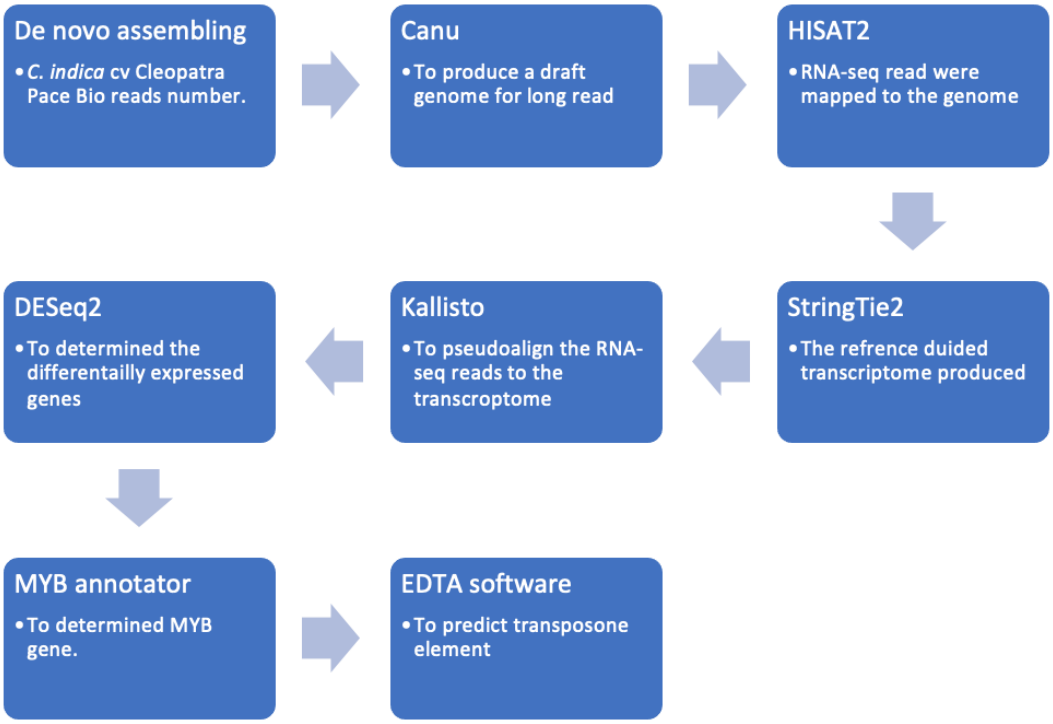


Figure 20: Schematic diagram of the analysis of the gDNA sequencing data.

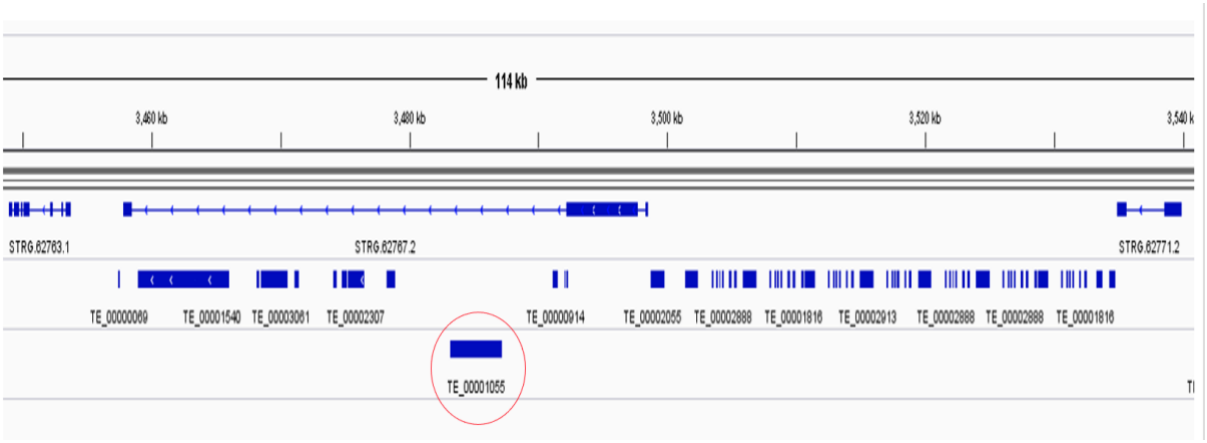


Figure 21: The Position of the transposon in the gene.

The top set of blue boxes/lines represent genes. STRG.62767.2 is a potential MYB gene. Blue boxes are exons, linked by blue lines are introns. The middle set of blue boxes are predicted TEs on the same strand of the genome as the MYB gene. The bottom line of blue boxes (just the one, TE_00001055) are predicted intact TEs. The transposon in the intron that we are interested in is circled with red.

4.3 Discussion

C. indica leaves vary greatly in their leaf colour for reasons that remain unclear (187). The whole genome of *C. indica* cv Cleopatra was sequenced, as at that time there was no genome sequence available to aid in identifying genes with transposons. PacBio sequencing for the whole *C. indica* genome was appropriate to determine the genes that are responsible for red colour sector because long sequencing read is possible with this technology (188). Furthermore, the long-read sequencing is good to identify the transposons' repeats in the genome rather than the short-read sequencing (189). Therefore, the approach we used was to identify a transposon in a gene involved in the anthocyanin pathway, most likely one of the MYB genes, as they are the strongest regulators of this pathway. The transposon could either enhance the expression of a gene in this case the transposon would be present in the red sectors or disrupt the expression of a gene in this case the transposon would be present in the green sectors. If a transposon had been inserted into a gene involved in colour formation, it would always have to be inserted into the same gene, which is very unlikely as transposons are inserted randomly into the genome. Therefore, it is much more likely that the transposon exerts its effect by jumping out of a gene, which would result in a colour change (190).

The biggest challenge was to purify gDNA that is pure and long enough for PacBio sequencing. Typically, DNA with high quality has a 260/280 ratio between 1.8 and 2.0 and does not contain polysaccharides or phenols as contaminants (191). It was previously reported that extracting DNA from *Canna* leaves was difficult because the leaves contain high level of fibres, polysaccharide and phenolic compounds (192). In order to get pure DNA, we tried a variety of approaches.

Using the gDNA Clean & Concentrator kit helped to improve the DNA quality, but it fragmented the DNA, and it is important for PacBio sequencing that the fragments are sufficiently long. Therefore, I stopped using the gDNA Clean & Concentrator kit and used the sorbitol wash buffer instead.

The dark treatment was an effective method to reduce the carbohydrates, polysaccharides, and polyphenolic compounds in *C. indica* cv Cleopatra leaves (Table 9) as it was reported previously (170). Furthermore, by grinding the tissue with liquid nitrogen for a much longer period of time and using small leaves, the walls of the cells were broken down better.

Plant leaf samples usually contain oils, sugars, and other endogenous chemicals that can make the DNA extraction difficult. However, several plants produce secondary metabolites which are rich in polysaccharides and polyphenols, that make this problem even worse (193). Furthermore, after cell lysis, polyphenols can bind irreversibly to DNA, whilst polysaccharides can form co-precipitates with DNA. Using sorbitol before cell lysis makes the extracted DNA purer as it is an osmotically active sugar alcohol (193). It means that the sugar alcohol in sorbitol was capable of removing polyphenols and polysaccharides from DNA before cell lysis. The washing step was repeated three times with sorbitol wash buffer due to the supernatant was viscous and brown colour. After the third wash the supernatant was light yellow in colour indicating that the polysaccharides and polyphenols had been reduced. Proteinase K and RNAse A were used to remove residual RNA and protein (194). Indeed, these two enzymes improved the purity of the DNA extracted from *C. indica* cv Cleopatra red leaf.

In order to separate DNA from other organic compound such as chloroform and resin of the Nucleon PhytoPure DNA extraction buffer, the phenol:chloroform:isoamyl alcohol (25:24:1) extraction was sufficient to improve the DNA quality (Figure 17 , A). In conclusion, the Nucleon PhytoPure DNA extraction kit protocol was modified with additional steps in order to generate purified DNA matching the requirement for sequencing the entire genome of *C. indica* cv Cleopatra using PacBio sequencing at the Earlham Institute (Figure 19. C, Table 6).

Although, the sequence analysis was not my work but carried out by David Prince, for completeness sake, I discuss the results here. He identified a transposon in the intron of a MYB gene. *C. indica* cv Cleopatra's MYB genes were compared with MYB genes annotated in banana (*Musa acuminate*), and the MYB gene containing the transposon in an intron was most similar to banana MYB clade 26 Figure (22) (195). While the function in banana of clade 26 MYBs is unknown, their close relationship to flavonoid biosynthesis regulators indicates they may function similarly to R2R3-MYB proanthocyanidin regulatory. Further investigation is needed into whether clade 26 MYBs regulate Zingiberales-specific biosynthesis pathways (195).

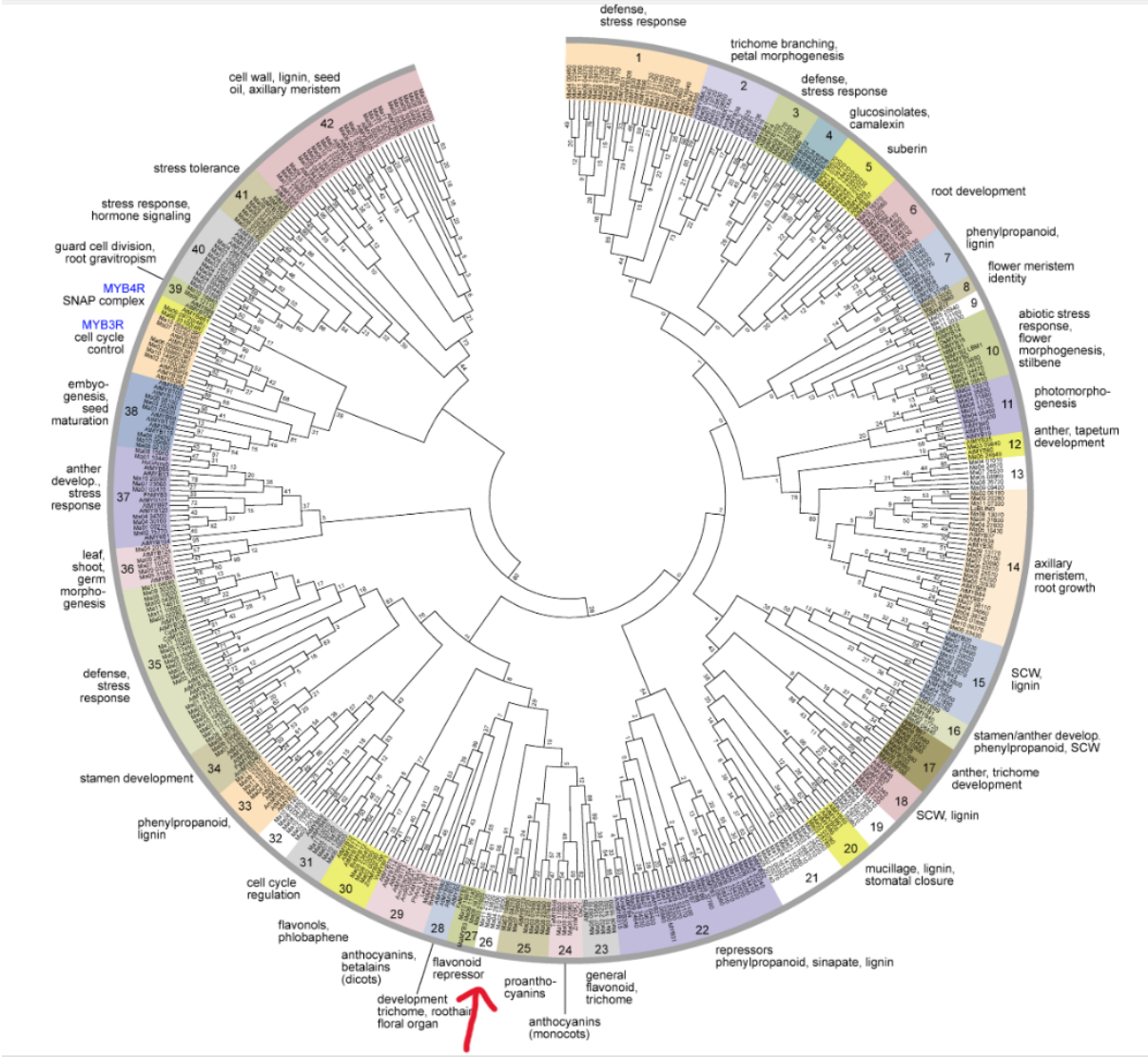


Figure 22: Banana MYB clade 26 that is similar to the Canna MYB TF

In peaches depending on whether their flesh and skin contain colour pigments, they are yellow- or white-fleshed and red- or pale-skinned (196). Among plant pigments, anthocyanin plays a significant role in peach flesh and skin colour (197). Peach skin is red due to the MYB transcription factor (PpMYB10). It was discovered that the red flesh around the stone phenotype was associated with a 487 bp deletion affecting the PpMYB10.1 promoter region (198).

Our results suggest that the presence of a transposon in an intronic region of a MYB gene leads to a higher level of expression in the red sectors of *C. indica* cv Cleopatra leaves. Whereas, in the green sectors of *C. indica* cv Cleopatra, the transposon would jump out of the MYB gene, and the reduced expression level make the leaves green.

Transposons and fragments inserted into the promoter region alter gene expression through different mechanisms, activating or inactivating R2R3-MYB activators. Occasionally, insertion of a transposon inactivates target genes by changing their promoter structure, preventing the TF from binding and thereby preventing transcription (199). In carrots, root pigmentation is caused by the high expression of DcMYB7, not coding region variations. In non-purple carrot roots, a tandem duplication and two transposon insertions in the promoter inactivate DcMYB7 (200). Strawberry fruit naturally vary in color due to different expression levels of FaMYB10. A transposon (FaEnSpm-2) that enhances the expression of FaMYB10-2 in red-fleshed strawberry lines is responsible for the enhanced expression of FaMYB10-2 (194). Transcriptional activity has been enhanced by R2R3-MYB activators, for example MdMYB10 promoter tandem repeats are associated with anthocyanin accumulation in red-fleshed apple varieties (85). These examples align well with our results, as we also identified a potential TE in the intronic region of a MYB TF, that could potentially upregulate its expression therefore leading to a higher anthocyanin production.

Our results suggest that the presence of a transposon in an intronic region of a MYB gene leads to a higher level of expression in the red sectors of *C. indica* cv Cleopatra leaves. Whereas, in the green sectors of *C. indica* cv Cleopatra, the

transposon would jump out of the MYB gene, and the reduced expression level leads to a loss of anthocyanin production.

Chapter 5: The Role of Small Non-coding RNAs in Colour Leaf Formation

5.1 Introduction

This part of the projects addresses the role of small non-coding RNAs in leaf colour formation. We hypothesised that small RNAs are responsible for yellow and pink stripes. It is hypothesised that in green leaves, the small RNAs target enzymes involved in the chlorophyll biosynthetic pathway and silencing these genes result in yellow stripes (201). It is also hypothesised that in red leaves, the small RNAs target enzymes involved in the anthocyanin biosynthetic pathway and silencing these genes result in pink stripes (202). These hypothesis is supported by the observation that the boundaries around the edges of the stripes not sharp but gradual, suggesting the diffusion of a compound, such as small RNAs. In addition, we also explored the potential role of miRNAs by looking for differentially expressed miRNAs in red sectors of *C. indica* cv Cleopatra and pink striped leaves of *C. indica* cv Durban, and in green sectors of *C. indica* cv Cleopatra and yellow striped leaves of *C. indica* cv Pretoria.

RNA was extracted from both red and green leaf sectors of *C. indica* cv Cleopatra, and from leaves of *C. indica* cvs Durban and Pretoria. Small RNA libraries have been generated from these RNA and sequenced by next generation sequencing. A bioinformatics analysis predicted differentially expressed microRNAs.

Differentially expressed microRNAs candidates were validated through northern blot analysis, miRNAs with interesting expression patterns were selected and 5'RACE performed on predicted targets. The identity of validated miRNA target was further investigated with BLAST.

5.2 Results

5.2.1 RNA Extraction for Small RNA Library

We grew *C. indica* cv Cleopatra, *C. indica* cv Pretoria and *C. indica* cv Durban with different leaf coloration (green and red sectors, green with yellow stripes and red with pink stripes respectively). Three biological replicates of leaf tissue were obtained from each cultivar. We used miRVana miRNA isolation kit to isolate sRNA from the tissue, following the manufacturer's instructions. A Nanodrop 8000 spectrophotometer was used to measure RNA quality and quantity (Table 10) and RNA integrity assessed using 1.5% (w/v) agarose gel electrophoresis (Figure 23). Although these cultivars of *Canna* contain pigments in their leaves which may affect RNA quality, the extraction protocol used was sufficient to extract clean RNA.

Table 10: Nanodrop quality and quantity assessment of RNA samples

Canna cultivar	Leaf tissue colour	Biological replicate	Concentration (ng/µl)	A 260-280 ratio	A 260-230 ratio	Amount in µg used for sRNA library
***C. indica* cv Cleopatra**	Red	1	1792.2	2.15	2.19	2
***C. indica* cv Cleopatra**	Red	2	445.7	2.10	2.17	1.9
***C. indica* cv Cleopatra**	Red	3	1714	2.13	2.28	1.9
***C. indica* cv Cleopatra**	Green	1	1575.7	2.12	2.17	1.9
***C. indica* cv Cleopatra**	Green	2	2101.5	2.15	2.26	1.9
***C. indica* cv Cleopatra**	Green	3	872.8	2.10	2.02	1.9
***C. indica* cv Pretoria**	Yellow and green stripes	1	2212.3	2.14	2.22	2.2
***C. indica* cv Pretoria**	Yellow and green stripes	2	1457.4	2.09	2.16	2
***C. indica* cv Pretoria**	Yellow and green stripes	3	2970.5	2.03	2.21	2
***C. indica* cv Durban**	Red with pink stripes	1	1442	2.11	2.05	2.2
***C. indica* cv Durban**	Red with pink stripes	2	2623.2	2.09	2.07	2
***C. indica* cv Durban**	Red with pink stripes	3	2229.3	2.06	1.88	2

⁎ **A 260-280 ratio determine protein contamination.**

⁎ **A 260-230 ratio determine organic contamination.**

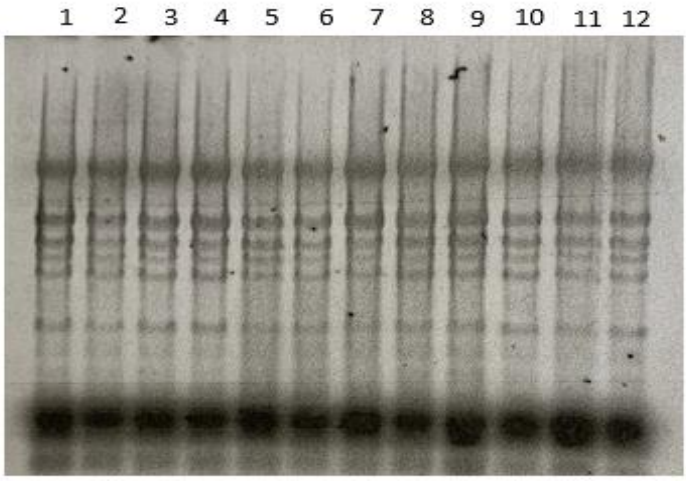


Figure 23: RNA integrity assessed from *C. indica* cv Cleopatra, Pretoria and Durban using 1.5% (w/v) agarose gel electrophoresis.

The well number **1** refer to *C. indica* cv Cleopatra Red sector replicate 1. **2:** *C. indica* cv Cleopatra Red sector replicate 2. **3:** *C. indica* cv Cleopatra Red sector replicate 3. **4:** *C. indica* cv Cleopatra Green sector replicate 1. **5:** *C. indica* cv Cleopatra Green sector replicate 2. **6:** *C. indica* cv Cleopatra Green sector replicate 3. **7:** *C. indica* cv Durban replicate 1. **8:** *C. indica* cv Durban replicate 2. **9:** *C. indica* cv Durban replicate 3. **10:** *C. indica* cv Pretoria replicate 1. **11:** *C. indica* cv Pretoria replicate 2. **12:** *C. indica* cv Pretoria replicate 3.

5.2.2 Small RNA Library Preparation

The extracted RNA was used for preparing sRNAs libraries, with three biological replicates for each *C. indica* cultivar. To start the sRNA library, we used 1.9 -2.2 µg for each sample. The adenylated 3'HD adapter efficiently was ligated to the sRNAs followed by a purification where RecJ Exonuclease was used to remove the excess of 3'HD adapter. Then, the 5' HD adapter was ligated to the sRNAs by T4 RNA ligase followed by a purification step. Next, reverse transcriptase enzyme was used to make a cDNA and then a PCR was run by primers that can anneal to the 5' and 3' adapters. The PCR products were run on a 8% polyacrylamide gel to purify the libraries that are between 145 and 150 bp (169) *C. indica* cv Durban cDNA library is shown as an example of the gel before and

after the bands were cut (Figure 24, A-B). All the libraries were loaded on 8% polyacrylamide gel to normalise them and scanned. Then pooled the rest of the library in one tube and sequenced them to find miRNA differently expressed for the cultivars. Several of the bands appeared light, and some were strong, which might be due to the differences in efficiency of purifying the PCR products from the gel (Figure 25).

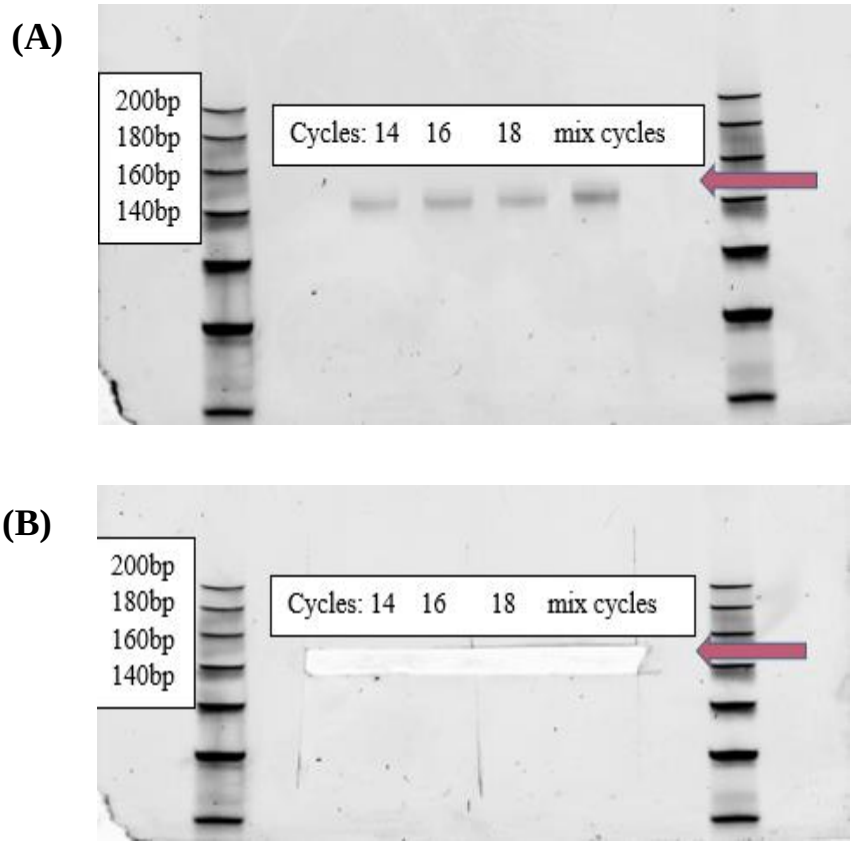


Figure 24: Gel extraction of the DNA.

A: cDNA library of *C. indica* cv Durban before the bands were cut. **B**: cDNA library of *C. indica* cv Durban after the bands were cut. Mix cycles means after running the three different PCR cycles which are 14,16 and 18, 5µl from each PCR cycles was added in a new PCR tube and run on the gel to cut the band for the three different PCR cycles and mixing sample.

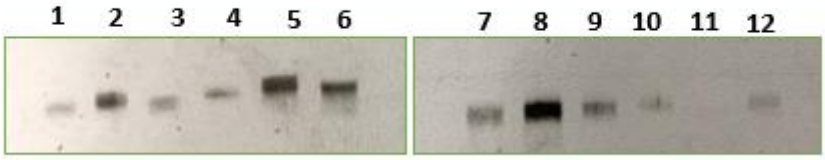


Figure 25: The gels of the cDNA library quantification by using the GE Healthcare Life Sciences, Typhoon FLA 9500.

1: *C. indica* cv Durban replicate 3. **2:** *C. indica* cv Durban replicate 2. **3:** *C. indica* cv Durban replicate 1. **4:** *C. indica* cv Pretoria replicate 3. **5:** *C. indica* cv Pretoria replicate 2. **6:** *C. indica* cv Pretoria replicate 1. **7:** *C. indica* cv Cleopatra red leaf sector replicates 1. **8:** *C. indica* cv Cleopatra red leaf sector replicate 2. **9:** *C. indica* cv Cleopatra red leaf sector replicate 3. **10:** *C. indica* cv Cleopatra green leaf sector replicate 1. **11:** *C. indica* cv Cleopatra green leaf sector replicate 2. **12:** *C. indica* cv Cleopatra green leaf sector replicate 3. ImageQuant was used to quantify library intensity and normalize it to one library.

5.2.3 Sequencing Data of Small RNA Library

Small RNA libraries were sequenced on an Illumina NextSeq 550 (203). The analysis of the sequenced libraries was conducted by Daniel Richardson (UEA, Norwich, UK) and he identified 10 microRNAs that were potentially differentially expressed between the different kinds of leaves. The read numbers of the miRNAs differentially expressed in *Canna* are shown in Table 11.

Table 11: The read numbers for the differentially expressed set of *C. indica* miRNAs. Green Cleopatra: *C. indica* cv Cleopatra green leaves sectors, Red Cleopatra: *C. indica* cv Cleopatra red leaves sectors, Pretoria: *C. indica* cv Pretoria, Durban: *C. indica* cv Durban.

Candidate miRNA	Sequences	Mean count Green Cleopatra	Mean count Red Cleopatra	Mean count Pretoria	Mean count Durban
MiR529 – First strand	AGAAGAGAGAGAGTACAGCCT	1700	10957	44	163
MiR529 – Second strand	GCTGTACCCTCTCTCTTCTTC	1620	6420	26	67
MiR156/miR157	GCTCTCTATGCTTCTGTCATC	93	1376	17	348
MiR6300	GTCGTTGTAGTATAGTGGTGAGTATTCCC	4253	436	2936	650
MiR530	TGCATTTGTACCTGCACCTAA	706	2569	27	209
MiR528	CCTGTGCCTGCCTCTTCCACG	359	1120	4	108
MiR171	CGAGCCGAACCAATATCACTC	122	376	5	13
MiR408	TGCACTGCCTCTTCCCTGGCT	1174	3615	23	226
MiR397	TTGAGTGCAGCGTTGATGAGA	140	639	1	12
MiR166	CGGACCAGGCTTCATTCCCC	11654	2108	70	508
MiR162	TCGATAAACCTCTGCATCCG	109	2012	31	308

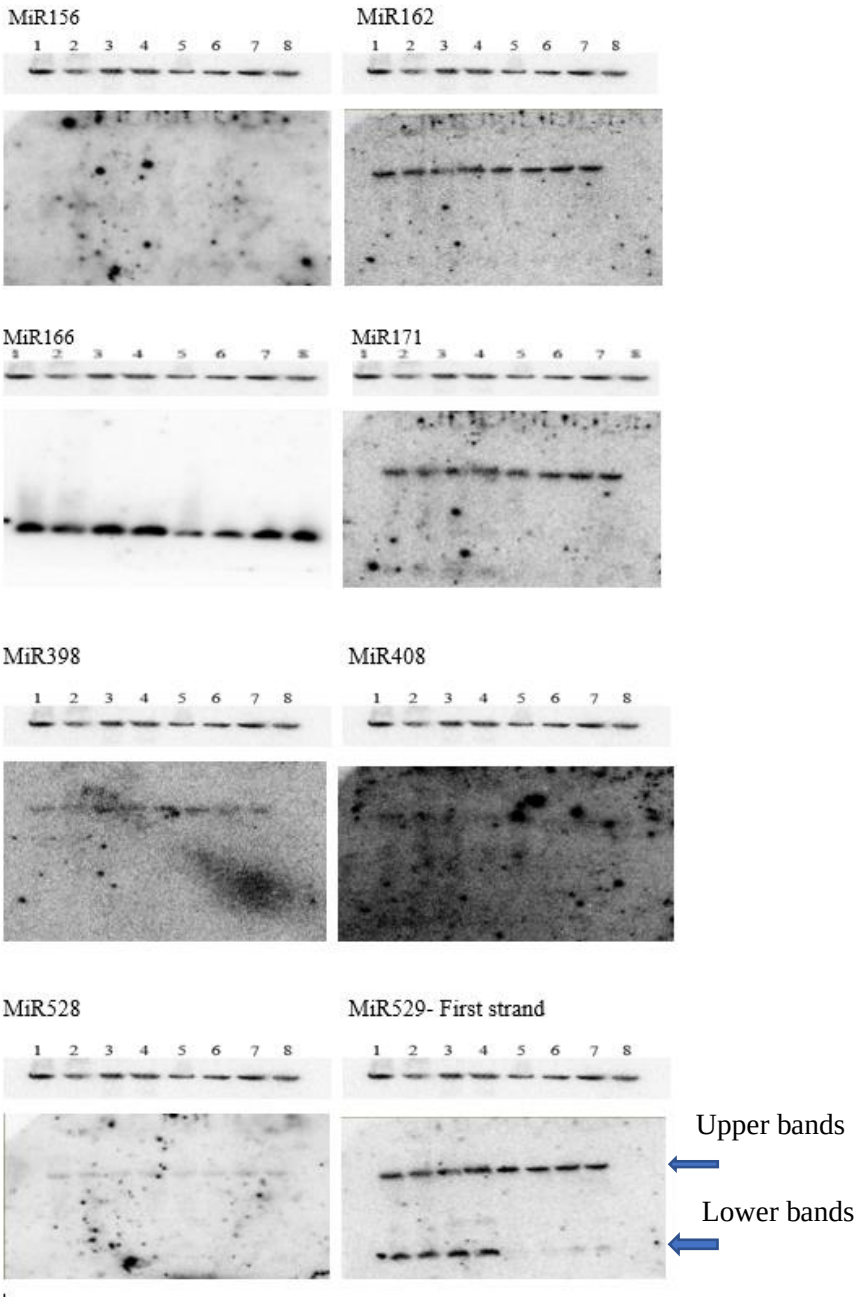
5.2.4 Northern Blot of MiRNA Candidates

After identifying 10 microRNAs that were potentially differentially expressed between green and red and between the plain colour and striped leaves, the miRNA candidates were examined by northern blot analysis in order to validate

their differential expression. For each miRNA and *Canna* cultivar two biological replicates were analysed. A denaturing polyacrylamide-urea gel was used to separate RNA samples, followed by transfer to nylon membranes. The miRNA band intensity was quantified using ImageQuant Figure 25.

5.2.5 Validation of MiRNA Candidates by Northern Blots

There were bands at a position where the pri-miRNAs could be expected on some of the northern blots (Figure 26), and we wanted to determine if they were pre-miRNAs. These upper bands are well above to 21nts, which is larger than would be expected for mature miRNAs. We examined two miRNAs conserved across plant species to investigate this (miR162 and miR171) in *C. indica* cv Cleopatra as well as in *Arabidopsis* and wheat. We extracted RNA from both sectors of *C. indica* cv Cleopatra, *A. thaliana* and wheat cv Pavon with two replicates for each cultivar. Northern blots were performed to validate these pre-miRNAs (Figure 27). However, in these experiments we did not find the upper bands in any of the species. After troubleshooting, it turned out that the upper bands were due to probing with U6 first and not washing it off sufficiently.


MiR156
1 2 3 4 5 6 7 8
MiR162
1 2 3 4 5 6 7 8
MiR166
MiR171
MiR398
1 2 3 4 5 6 7 8
MiR408
1 2 3 4 5 6 7 8
MiR528
1 2 3 4 5 6 7 8
MiR529- First strand
1 2 3 4 5 6 7 8
Upper bands
Lower bands

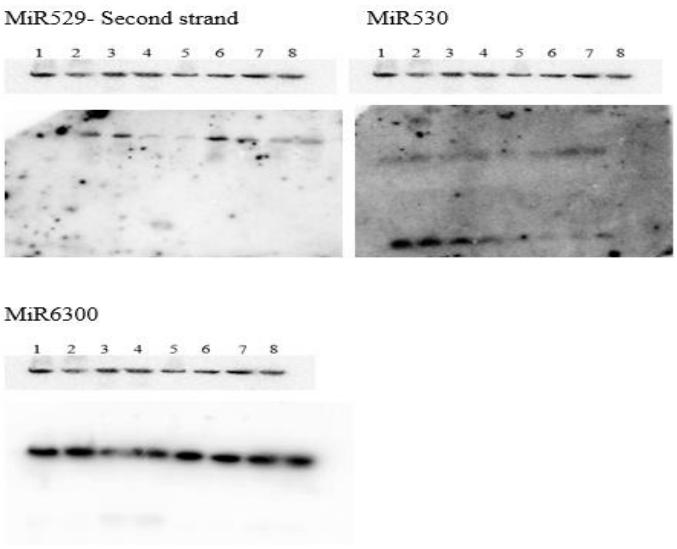


Figure 26: Whole membranes of the northern blots for of the miRNA candidates.

Total RNA was separated on denaturing polyacrylamide gels and blotted to membranes. The membranes were hybridised with miRNA specific probes indicated above the membranes. The lanes contained samples of **1**: *C. indica* cv Cleopatra red sector replicate 1. **2**: *C. indica* cv Cleopatra red sector replicate 2. **3**: *C. indica* cv Cleopatra green sector replicate 1. **4**: *C. indica* cv Cleopatra green sector replicate 2. **5**: *C. indica* cv Pretoria replicate 1. **6**: *C. indica* cv Pretoria replicate 2. **7:** *C. indica* cv Durban replicate 1. **8**: *C. indica* cv Durban replicate 2. An RNA loading control was performed using a U6 specific probe, that are shown above each panel.

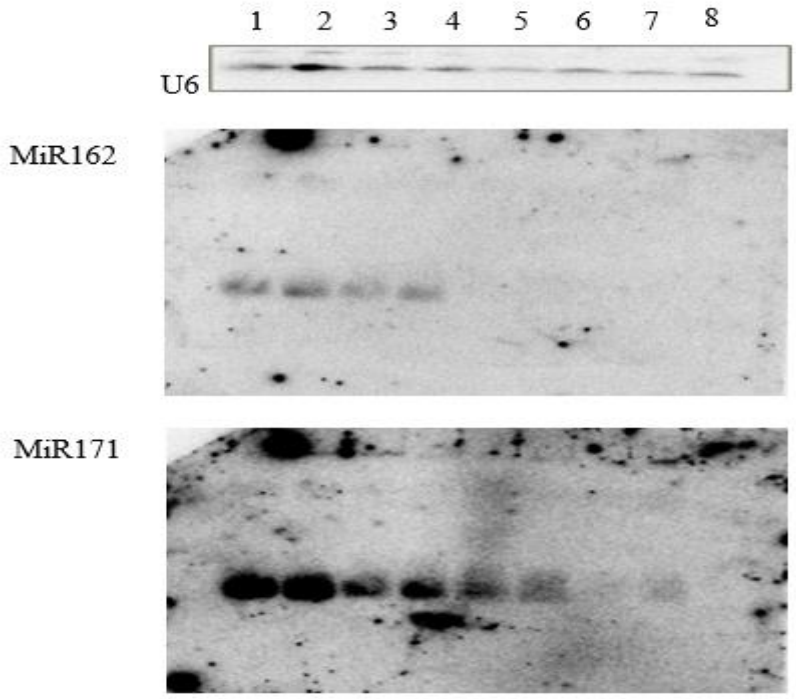


Figure 27: Whole membranes of northern blots of *C. indica* cv Cleopatra, *A.thaliana* and wheat cv Pavon.

Tested upper band (pre-miRNA). **1**: *C. indica* cv Cleopatra red sector replicate 1. **2**: *C. indica* cv Cleopatra red sector replicate 2. **3**: *C. indica* cv Cleopatra green sector replicate 1. **4**: *C. indica* cv Cleopatra green sector replicate 2. **5**: *A. thaliana* replicate 1. **6**: *A. thaliana* replicate 2. **7**: Wheat cv Pavon replicant 1. **8**: Wheat cv Pavon replicate 2. An RNA loading control was performed using a U6 specific probe.

The validation results for the 10 differently expressed miRNAs are shown in Figure 28. MiR166, miR529, miR530 and miR6300 showed the most interesting patterns. The expression of miR166 is much higher in green sectors of *C. indica* cv Cleopatra and Durban than in Red sectors of *C. indica* cv Cleopatra and Pretoria. One strand of miR529 –showed higher expression pattern in the green and red sectors of *C. indica* cv Cleopatra compared to the striped samples (Figure 31). The expression of miR530 is much higher in the red and green sectors of *C. indica* cv Cleopatra's compared to the striped samples of *C. indica* cv Pretoria

and Durban. MiR6300 showed high expression in green sectors of *C. indica* cv Cleopatra and a very similar low expression level in the other three samples. MiR528 had visible bands in all samples but no difference in expression between them. Therefore, the northern blot did not match the pattern of the read numbers, which showed high read numbers in the red sector of *C. indica* cv Cleopatra and lower read numbers in *C. indica* cv Pretoria. Therefore, this miRNA was not investigated further. Moreover, miRNA 156/157, miR162, miR171, miR397, miR408 and miR529 – Second strand showed no pattern of differential expression in the northern blots and were not investigated further.

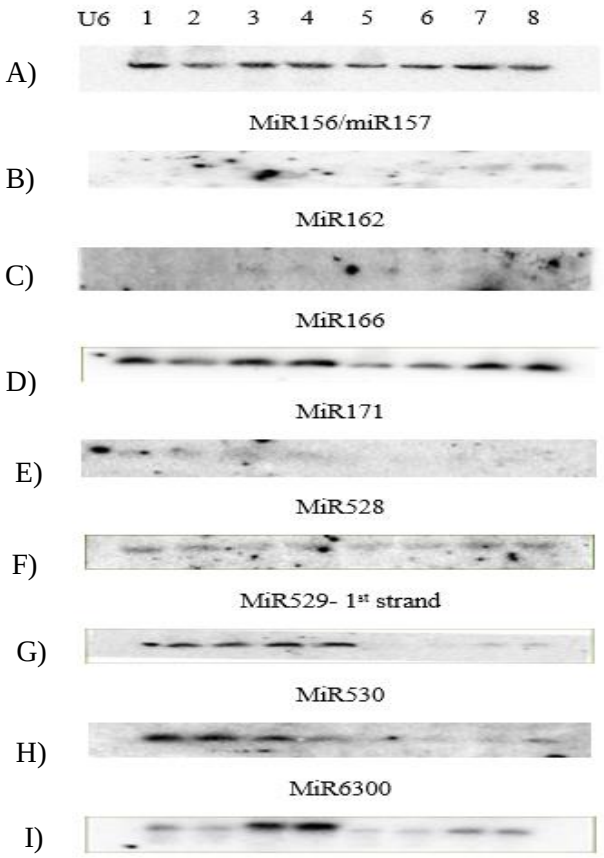


Figure 28: Validation of differentially expressed miRNA candidates using northern blot.

Total RNA was separated on denaturing polyacrylamide gels and blotted to membranes. The membranes were hybridised with miRNA specific probes indicated above the membranes. The lanes contained samples of **1**: *C. indica* cv

Cleopatra red sector replicate 1. **2**: *C. indica* cv Cleopatra red sector replicate 2. **3**: *C. indica* cv Cleopatra green sector replicate 1.**4**: *C. indica* cv Cleopatra green sector replicate 2. **5**: *C. indica* cv Pretoria replicate 1. **6**: *C. indica* cv Pretoria replicate 2. **7**: *C. indica* cv Durban replicate 1. **8**: *C. indica* cv Durban replicate 2. A) An RNA loading control was performed using a U6 specific probe.

5.2.6 The Analyses of Validation of MiRNA Candidates

The intensity of bands on the northern blots were quantified and the level of miRNAs was normalised based on the intensity of the U6 loading control. After normalisation, the pattern of miR530 (Figure 29, A) matched the expected pattern predicted by the read numbers (Table 11) as red sector showed the highest expression followed by the green sector and then, *C. indica* cv Durban and *C. indica* cv Pretoria were lower. The expression of miR530 was different between the plain leaves of *C. indica* cv Cleopatra red, green sectors and the striped leaves of the *C. indica* cvs Durban and Pretoria.

However, the bioinformatics analysis of the sRNA libraries for miR166 indicated that the highest read number of *C. indica* cvs were in the green leaf sector of Cleopatra, followed by red leaf sector of Cleopatra, Durban and Pretoria, respectively. The pattern (Figure 29, B) of the normalized band intensity (expression) of *C. indica* cvs was highest in green leaf sector of Cleopatra, followed by Durban, red leaf sector of Cleopatra and Pretoria, respectively. The expression is not much different between the plain leaves of *C. indica* cv Cleopatra red, green sectors and the striped leaves of the *C. indica* cvs Durban and Pretoria.

The miR529 pattern is broadly matched the expected pattern predicted by the read numbers (Table 11) as the red and green sectors have much higher expression than the Pretoria and Durban. The difference is that the read numbers were much higher for red Cleopatra than green Cleopatra, whereas the northern blot shows a slightly higher level in green compared to red (Figure 29, C). The expression is much different between the plain leaves of *C. indica* cv Cleopatra red and green sectors, and the striped leaves of the *C. indica* cvs Durban and Pretoria.

The bioinformatics analysis of the read numbers of MiR6300 showed highest expression in the green sectors of *C. indica* cv Cleopatra, with slightly lower read number in Pretoria, and much lower in Durban and red sectors of *C. indica* cv Cleopatra. That is slightly different from the northern blot results, which showed the highest accumulation of miR6300 in the green sectors of *C. indica* cv Cleopatra, and much lower in the other three (Figure 29, D). However, the very high expression in green sector was confirmed by northern blot, which makes this miRNA a very interesting candidate to follow up. The northern blot result does not always match the sequencing result, but the northern blot is more reliable and therefor it is important to validate the sequencing result by northern blot. The adapter ligation step during small RNA library preparation is biased, as it favours small RNAs that can anneal to the adapters used in the reactions, while the intensity of the bands on northern blots is purely based on the amount of RNA on the membrane.

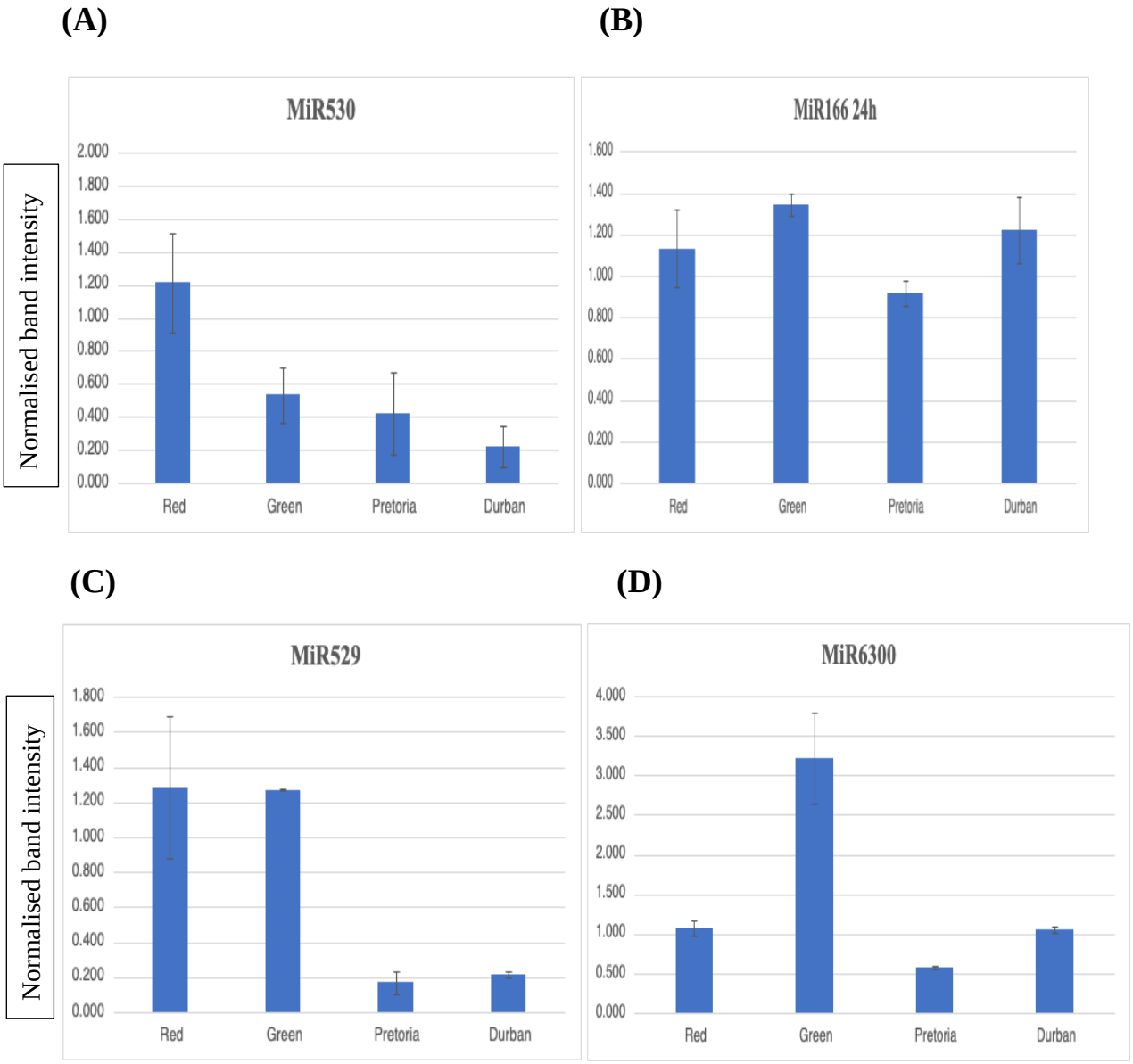


Figure 29: Quantification of the most interesting miRNA expression patterns obtained by the northern blot.

Each bar made up of the two biological replicates**. Red** = *C. indica* cv Cleopatra red leave sector, **Green** = *C. indica* cv Cleopatra green leave sector. **Pretoria** = *C. indica* cv Pretoria, **Durban** = *C. indica* cv Durban. The number in the left is normalized band intensity. **A:** MiR530. **B:** MiR166. **C:** MiR529. **D:** MiR6300. Error Bars represent the standard deviations of miRNA expression in *Canna* tissues.

5.2.7 Validating Predicted miRNA Targets Using 5'RACE

Candidate targets for the differentially expressed miRNAs validated by northern blot were predicted computationally by Simon Moxon (UEA, Norwich, UK)

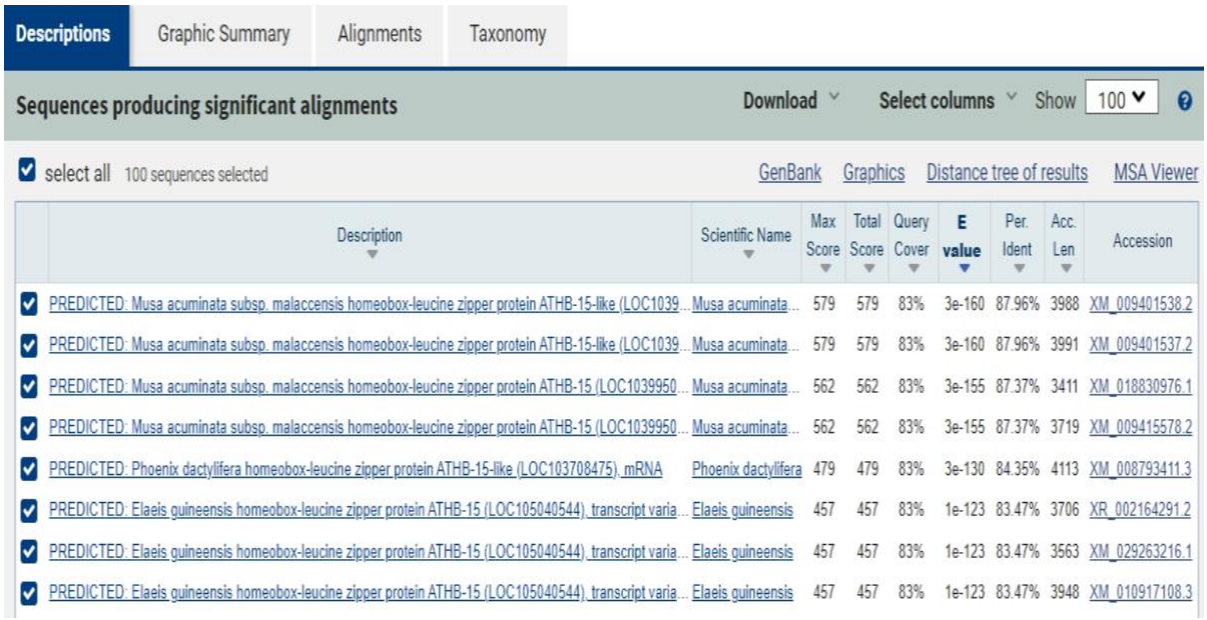
(Table 12). He was able to predict target mRNAs for miR166 and miR530 but did not find any potential targets for miR529 and miR6300. BLAST analysis of the three predicted target genes revealed that the miR166 target DN31218 is similar to known ATHB-15 genes belonging to the HD-ZIP III TF family (Figure 30, A) and the miR166 target DN40098 is similar to known OSHOX32 genes (Figure 30, B). However, we were not able to find any known gene similar to the miR530 target DN43985 (Figure 30, C).

A 5'RACE analysis was used to validate these candidate target mRNAs (204). Validation of predicted target mRNA is important because there are always false positive predictions. 5'RACE is an appropriate method due to two reasons. First, the cleavage of target mRNAs happens at a specific position between the 10^{th} and 11^{th} position of the miRNA. Second, the downstream cleavage product is relatively stable, therefore it is possible to sequence its 5' end. mRNAs can be fragmented but that happens randomly, therefore if a specific fragment can be identified that consistently starts at the predicted cleavage position, that can prove that the mRNA is not randomly degraded but targeted by the miRNA. Total mRNA from green leaf tissue of *C. indica* cv Cleopatra was used for the analysis as both miRNAs being investigated were expressed in this tissue. Gene-specific PCR and nested gene-specific PCR were performed (Figure 31) and the PCR products were extracted from the gel and cloned into the pJET1.2 vector. The band size of the miR166 targeted DN31218 is 217 bp, the miR166 targeted DN40098 is 217 bp and the miR530 targeted DN43985 is 243 bp.

Table 12: MiRNAs target and their gene targets.

MiRNA	*De novo* transcriptome accession number of predicted targets	Potential homolog based on BLAST
MiR166	DN31218	ATHB-15
MiR166	DN40098	OSHOX32
MiR530	DN43985	N/A

(A)



(B)

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
PREDICTED: Musa acuminata subsp. malaccensis homeobox-leucine zipper prote...	Musa acumi...	1284	1284	80%	0.0	88.90%	3195	XM_009410276.2
PREDICTED: Musa acuminata subsp. malaccensis homeobox-leucine zipper prote...	Musa acumi...	1280	1280	80%	0.0	88.73%	3000	XM_009400770.2
PREDICTED: Musa acuminata subsp. malaccensis homeobox-leucine zipper prote...	Musa acumi...	1266	1266	80%	0.0	88.54%	3078	XM_009390689.2
PREDICTED: Musa acuminata subsp. malaccensis homeobox-leucine zipper prote...	Musa acumi...	1253	1253	80%	0.0	88.26%	3072	XM_009399116.2
PREDICTED: Musa acuminata subsp. malaccensis homeobox-leucine zipper prote...	Musa acumi...	1242	1242	80%	0.0	88.15%	3343	XM_009399442.2
PREDICTED: Elaeis guineensis homeobox-leucine zipper protein HOX32 (LOC105...	Elaeis guine...	1195	1195	80%	0.0	87.28%	3245	XM_010907119.3
PREDICTED: Elaeis guineensis homeobox-leucine zipper protein HOX32 (LOC105...	Elaeis guine...	1195	1195	80%	0.0	87.28%	3248	XM_010907118.3
PREDICTED: Phoenix dactylifera homeobox-leucine zipper protein HOX32-like (L...	Phoenix dac...	1190	1190	80%	0.0	87.19%	3205	XM_039114591.1
PREDICTED: Phoenix dactylifera homeobox-leucine zipper protein HOX32 (LOC1...	Phoenix dac...	1173	1173	80%	0.0	86.90%	3172	XM_008796230.4
PREDICTED: Elaeis guineensis homeobox-leucine zipper protein HOX32 (LOC105...	Elaeis guine...	1168	1168	80%	0.0	86.81%	3025	XM_010925376.2

(C)

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
PREDICTED: Mercurialis annua uncharacterized LOC126668819 (LOC12666...	Mercurialis ...	60.2	60.2	4%	6e-04	92.68%	3477	XM_050362016.1
PREDICTED: Mercurialis annua uncharacterized LOC126668819 (LOC12666...	Mercurialis ...	60.2	60.2	4%	6e-04	92.68%	3394	XM_050362015.1
PREDICTED: Mercurialis annua uncharacterized LOC126668819 (LOC12666...	Mercurialis ...	60.2	60.2	4%	6e-04	92.68%	3556	XM_050362013.1
PREDICTED: Mercurialis annua uncharacterized LOC126668819 (LOC12666...	Mercurialis ...	60.2	60.2	4%	6e-04	92.68%	3455	XM_050362012.1
PREDICTED: Mercurialis annua uncharacterized LOC126668819 (LOC12666...	Mercurialis ...	60.2	60.2	4%	6e-04	92.68%	3460	XM_050362011.1
PREDICTED: Mercurialis annua uncharacterized LOC126668819 (LOC12666...	Mercurialis ...	60.2	60.2	4%	6e-04	92.68%	3464	XM_050362010.1

Figure 30: Homology search for the predicted miRNA targets

A: MiR166 target DN31218. **B:** MiR166 target DN40098. **C:** MiR530 target DN43985. The sequences picked up by the miR530 target is not accepted as similar because of the low coverage.

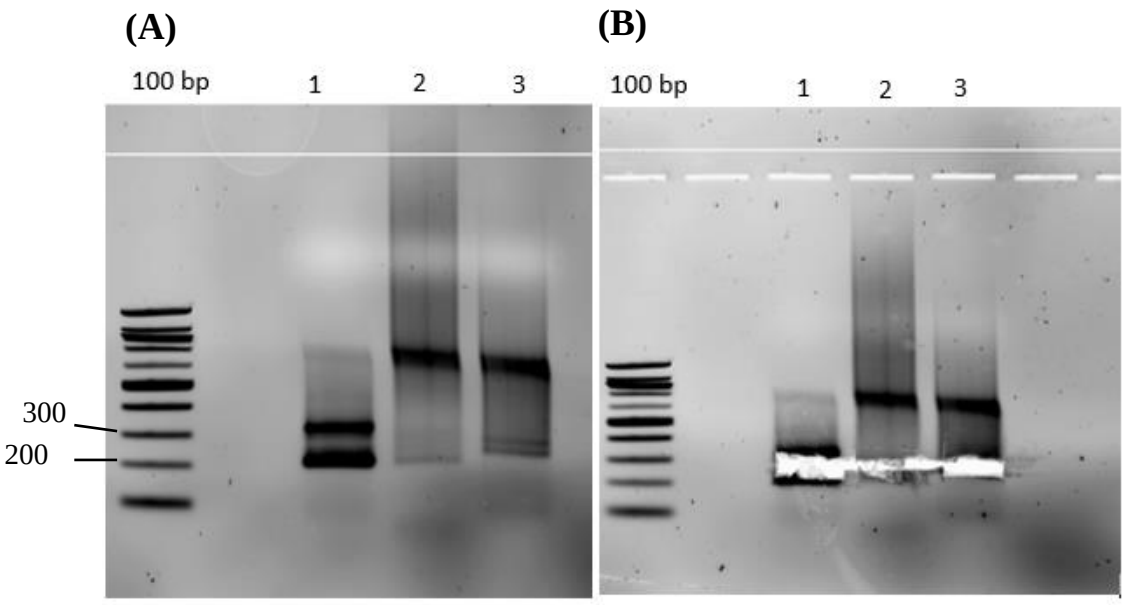


Figure 31: Result of the nested PCR of 5'RACE.

5'RACE was carried out using three different primer sets specific to MiR166 candidate target DN31218 (lane **1**), MiR166 candidate target DN40098 (lane **2**) and MiR530 candidate target DN43985 (lane **3**). The final PCR products were separated on agarose gels (panel **A**) and then the bands with the expected size around 200bp were cut out from the gel (panel **B**).

For candidate miR166 target DN31218, 23 clones were sequenced and 16 contained the amplification product of the transcript precisely cleaved between nucleotides complementary to the 9th and 10th nucleotide of the miR166, therefore the miRNA166 mediated cleavage of the DN31218 transcript was clearly demonstrated (Figure 32, A).

For candidate miR166 target DN40098, 18 clones were sequenced and 13 contained the amplification product of the transcript precisely cleaved between nucleotides complementary to the 9th and 10th nucleotide of miR166, therefore the DN40098 transcript has been validated as a miRNA166 target (Figure 32, B).

For candidate miR530 target DN43985, 12 clones were sequenced and 9 contained the amplification product of the transcript precisely cleaved between nucleotides complementary to the 10th and 11th nucleotide of miR530, therefore miR530 cleaves the DN43985 transcript (Figure 32, C).

Figure 32: 5'RACE target validation results for predicated miR166 andMiR530 targets.

The arrows indicate the position where the majority of the clones mapped following cloning and sequencing of the purified PCR fragments obtained by 5'RACE. The upper lines show the sequence of the predicted targets, and the lower lines are the sequences of the miRNAs. **A**: Predicted target of miR166 - DN31218 between the 10th and 11th nucleotide and 16/23 refer to number of colonies. **B**: Predicted target of miR166-DN40098 between the 10th and 11th nucleotide and 13/18 refer to number of colonies. **C**: Predicted target of miR530-DN43985 between the 10th and 11th nucleotide and 9/12 refer to number of colonies.

5.3 Discussion

5.3.1 Studying the Role of Small Non-coding RNA Molecules in Colour Formation

C. indica leaves have interesting different colour patterns. The aim of this work was to identify the small RNAs that are responsible for red and green leaf of *C. indica* cv Cleopatra and yellow and pink striped leaf of *C. indica* cv Pretoria and Durban. The role of the sRNAs may play in colour formation in *C. indica* leaves has not been identified in previous studies. To study the role of sRNAs, RNA was extracted for sRNA libraries from *C. indica* cvs Cleopatra red and green leaf sectors, Durban and Pretoria. Some of the *C. indica* cultivars produced strong bands and others had faint bands (Figure 25). It might be due to the total RNA that was utilized, or concentration of the total RNA sample was not high (169). An Illumina NextSeq 550 was used for sequencing the libraries. It is one of the most widely used NGS platforms for sequencing sRNA libraries because it is capable of massively parallel sequencing of 50-100bp DNA (169). A total of ten microRNAs were identified as potentially differentially expressed between red and green sectors of *C. indica* cvs Cleopatra, Pretoria and Durban and validation of these miRNA candidates was conducted using northern blot analysis. The advantage of the northern blot is that it can provide important information about size, abundance, and identity of RNA. Whereas the disadvantage of the northern

blotting is that it requires significant amount of RNA, and it can examine only one gene at a time (205).

Initially, northern blots were performed hybridising a U6 positive control to the membrane first, followed by scanning, stripping and re-probing with miRNA specific probes and scanning. Some membranes showed an upper band which could be pre-miRNA (Figure 26). Investigating these upper bands required further study. To explore the upper bands of the *C. indica* cv Cleopatra for both red and green sectors and compare them to other plants we selected *A. thaliana* and wheat cv Pavon. We performed northern blots with RNA extracted from those plant to validate miRNA162 and miR171 patterns. These miRNAs were selected because they are well conserved and had been investigated in previous studies. MiR162-guided cleavage is known to regulate the DCL1 mRNA in *A. thaliana* (206). Sequenced libraries contain miR162 in wheat (207) and miR162 was also identified in wheat in another study (208). Figure 27 shows the result of the northern blots for the miRNAs in those plants species. It appeared that no pre-miRNAs (upper bands) were visible on those two miRNA blots. Even though, miR162 was not expressed in *A. thaliana* and wheat cv Pavon (Figure 27). I recognized that miRNA probes should be used first before the U6 control, as the expression of U6 is very strong and sometimes it is difficult to strip off the membrane. Therefore, subsequently the miRNA specific probes were applied to the blot before the U6 probe. After the northern blot experiments, we concluded that four miRNAs are differentially expressed across the *Canna* cultivars: miRNA166, miRNA529, miRNA530 and miR6300 Figure 29. The normalized bands intensity of the miR166 was highest in the green sector, followed by Durban, red sector and Pretoria. The miR529 northern blot expression showed high expression in the plain red and green sectors of *C. indica* cv Cleopatra compared to the striped leaves in Pretoria and Durban. The normalized bands

intensity of the miR530 was higher in plain red and green leaves compared to the striped leaves of Pretoria and Durban. In addition, the normalized bands intensity of the miR6300 was high in the green sector, and low in Durban, red sector, and Pretoria.

It does appear that miR529 participates in *Anthurium andraeanum* spathe coloration by interacting with miR156 (209). Furthermore, in *Canna* cv Red president the red flowers expressed miR529 at a higher level than the yellow flowers of Tropicana president cultivar (9), although it was not validated by northern blot. In our experiment the sequencing also predicted an almost ten times higher level of miRNA529 in the red sectors compared to the green sectors, but the northern blot showed a very similar level of miRNA529 in green and red sectors. This contradiction should be investigated further. It is possible that the sequencing result is not reliable, as it is well known that northern blots do not always confirm the expression patterns predicted by sequencing (as several other miRNAs were predicted to be differentially expressed by our sequencing result that could not be validated by northern blot) (205). Also, the other strand of miR529 could not be detected by northern blot, although it was identified by sequencing.

MiR6300 showed an interesting expression pattern with high level of expression in *C. indica* cv Cleopatra green sectors but very low level in the *C. indica* cv Cleopatra red sectors, *C. indica* cv Durban and *C. indica* cv Pretoria, respectively. A previous study found that miR6300 could regulate defence against Fusarium in cucumber (210). Furthermore, *Acer pictum subsp. mono* miR6300 is largely responsible for stable anthocyanin accumulation (211). Red leaves of *Acer pictum subsp. mono* have significantly higher levels of cyanidin-3-O-glucoside than

green leaves due to the upregulation of *ApUFGT* and the downregulation of miR6300, as *ApUFGT* was targeted by miR6300 (211). Since the expression of miR6300 was low in red sectors and high in green, it is possible that miR6300 regulates the expression of the anthocyanin biosynthesis pathway also in *Canna* similarly to *Acer*. Unfortunately, we were not able to predict a target gene for miRNA6300, most likely because at the time no full transcriptome was available. It will be worth repeating the target prediction when a full transcriptome is available, or specifically try to find a miRNA6300 target site in *ApUFGT* homologs.

Several of the miRNAs that were predicted to be differently expressed by the sequencing results, miR156/157, miR162, miR171 and miR528 did not show interesting expression patterns by northern blots (Figure 28), whereas miR166 and miR530 did show interesting patterns. The following paragraphs will discuss these two miRNAs and their gene targets.

5.3.2 MiR166 Targets ATHB-15

By using 5' RACE, it is possible to validate cleaved target mRNAs because the 3' end products are relatively stable, and cleavage takes place at a specific position in between the miRNA's 10^{th} and 11^{th} nucleotides (212,213). Two miR166 candidate targets belonging to the HD-ZIP III family were successfully validated through 5'RACE (Figure 32). The first one is ATHB-16 and the second one is Hox32. MiR166 is one of the most conserved miRNA families (214) and HD ZIP-III TFs are known to be targeted by miR166 in other species (214). Moreover, there is only one nucleotide difference between mature miRNA sequences of miR165 and miRNA166 (215). It has been well established that miR165 and miR166 target the same genes, HD-ZIP III family TFs, which are

primarily responsible for embryonic patterning, building and maintaining SAMs, and determining lateral organ polarity (215).

A study of miR165/miR166 reported high expression in leaves in *A. thaliana* (216,217) This may be consistent with expression of the miR166 in *C. indica* cvs which also showed high expression in the leaves. Many plants, including *Arabidopsis*, *Medicago*, and soybeans all contain miR166, which plays an important role in a variety of developmental processes, including the development of seeds, roots, organ polarity, SAMs, as well as vascular development of shoots (218,219). It was reported that microRNA 165/166 may regulate the formation of secondary cell walls by targeting HD-ZIP III TF (46). Furthermore, it was mentioned that miR165 is necessary for thickening of the pith wall (216). In *Canna* flowers floral development as well as the transition from the vegetative to reproductive phase of meristems are critically influenced by miR166 (9).

AtHB15/CORONA which is one of the five classes of III HD-ZIP III TFs, plays critical roles in xylem tissue formation, meristem formation and organ polarity (220). However, as a result of transgenic analysis and the characterization of knockout mutants, it has been established that AtHB15 is responsible for the development of the pith secondary wall. The *stp–2d* mutant repressed AtHB15 and the phenotype of the leaf changes to curled leaves in *Arabidopsis* (216). The ATHB15 regulates vascular cell growth and phloem cell development in *A. thaliana* (125). *Canna* cultivars expressing miRNA166 might be in a similar situation. Curled leaves were observed on *C. indica* cultivars in their initial growth, and the example of a curled leaf in *C. indica* cv Durban is shown in the introduction chapter 1 (Figure 8). It might be ATHB15 more expressed when the

leaf is curled and becomes less expressed when the leaf is open. This needs to be further investigated by development of the gene silencing protocol in *Canna* and silencing the *AtHB15* to see if there is a phenotypic change of the leaves such as curling of the leaves.

5.3.3 MiR166 Targets Hox32

Another validated target for *Canna* miRN166 was a Hox32 homolog. Hox32 expression in rice was high in young panicles and stem apical meristems but was low in mature panicles and old leaves (221). Hox32 regulates rice leaf morphology through regulation of the *YABBY* gene. Upon overexpression of Hox32, leaves became narrow and rolled upward. In addition, Hox32 overexpression caused rolled and filamentous leaves (130). In contrast, *Canna* shows a different result. The morphology of *Canna* leaves is similar to banana leaves. Some leaves appear rolled inside for a few days since initiation growing, possibly because they are expressing the Hox32 gene.

Plants' mechanical strength is influenced by the OsmiR166b-OsHox32 module that regulates the expression levels of genes involved in cell wall formation, resulting in altered leaf shape and rigidity (128). However, overexpression of Hox32 reduced plant height and narrowed leaves (130). It will be interesting to study further the biological relevance of the miR166/Hox32 regulation.

5.3.4 MiR530

Northern blot analysis of miR530 revealed an interesting pattern across *Canna* cultivars (Figure 29) and its predicted target gene was successfully validated through 5'RACE (Figure 32). BLAST analysis did not reveal any sequence

similarity of this target to other gene. MiR530 was first reported in Helianthus and the putative target is the 70-kDa heat shock protein (222). 70-kDa heat shock protein function is important in the tolerance of the biotic and abiotic stress (223). A study showed that miR530 regulated rice growth period and blocking miR530 led to earlier flowering and seed maturity due to its involvement with rice resistance and yield (224). MiR530 is also expressed in banana fruit and 14-3-3-like protein D is a target gene (225). A wide range of processes are regulated by 14-3-3-like protein D, including metabolism, growth, development, transport, and stress response (226). Although, we validated a target gene for miR530 in *Canna*, unfortunately that gene does not show homology to any known gene, therefore it does not help understanding the biological role of miR530 in *Canna* leaves. The level of miR530 is similarly high in both green and red sectors of *C. indica* cv Cleopatra and very low in the striped *C. indica* cvs Pretoria and Durban leaves. It is unlikely it is involved in anthocyanin production, since its level is similar in red and green sectors and very low in the red leaf Durban cultivar, therefore it does not correlate with the level of anthocyanin. Further study of miR530 biological relevance will be interesting. Figure 33 shows the summarizing of miRNAs results and the pathways.

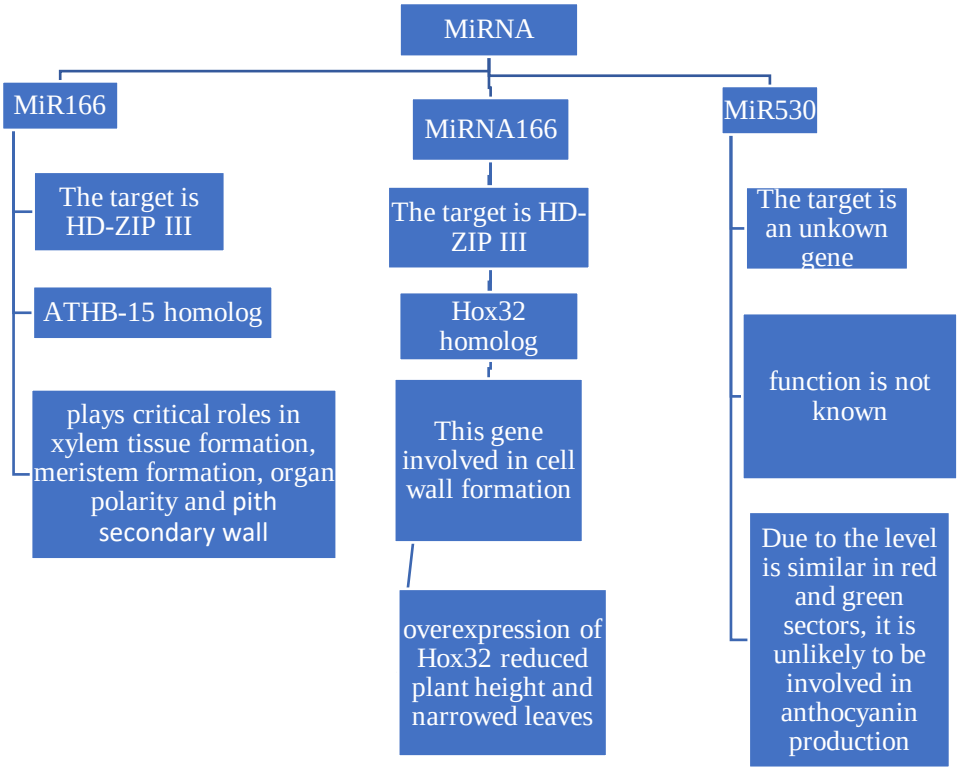


Figure 33: The summarizing of the miRNAs targets and pathways involved.

Chapter 6: General Discussion and Conclusion

The purpose of this project is to investigate the molecular mechanism of colour formation in the leaves of *C. indica* cvs plants by using two different approaches. The first approach of this project (chapter 3) investigated whether the presence or absence of transposon was responsible for the red-green sectors on *C. indica* cv Cleopatra leaves. If there were transposons, the transposons either disrupt or increase the expression of genes responsible for the red colour of leaves due to the sharp boundary between red and green sectors (Figure 7B).

The RNA was extracted from red and green sectors of *C. indica* cv Cleopatra. RNAseq data were used to create a *de novo* transcriptome since there were no genomes or annotations available at the time, and differentially expressed mRNAs were identified by Simon Moxon. About 500 differentially expressed mRNAs were identified. BLAST was then used to determine if any of the differentially expressed genes were similar to MYBs and bHLHs transcription factors. Among the differentially expressed genes, five MYB and two bHLH genes were found. MYBs and bHLHs regulate the anthocyanin biosynthetic pathway (58). MYB TFs constitute one of the largest gene families of plant TFs (72). There were several highly expressed MYB and bHLH genes in the flowers of Red President of *Canna*, which is a cultivar with red flower, suggesting that those genes may be involved in anthocyanin synthesis (185). In addition to TFs, CHS is a key gene of the anthocyanin biosynthetic pathway (58). Therefore, primers specific to MYBs and CHS had been designed to investigate whether there are transposons in these genes. PCR reactions were performed for both MYBs and Chalcone synthase genes, but we did not find transposon in these candidates (Figure 16). There is still the possibility that the hypothesis is right,

but the transposon may be located in another gene not detected by the RNAseq, or the transposon may be located outside the coding region.

Chapter 4 described an alternative approach using whole genome sequencing of the *C. indica* cv Cleopatra as at the time there was no genome sequence available for *C. indica*. High quality high molecular weight genomic DNA was purified from the red and green sectors of *C. indica* cv Cleopatra. The genome sequencing was done by PacBio sequencing. With this technology, sequences reads longer than 200k can be produced, which allows the whole genome of *C. indica* to be sequenced. The analysis of the PacBio sequencing was done by David Prince (UEA, Norwich, UK). A MYB gene was identified that contained a predicted transposon element in the intron region (Figure 21). This MYB gene was compared to the MYB genes annotated in banana and the most closely related banana gene was the MYB clade 26 (176). The potential transposon could regulate the expression of the MYB, which would lead to the higher expression of anthocyanin, like in red apple fruit that is associated with MYB10 promoter tandem repeats (227). When transposon insertions occur, the structure of the promoter can be altered, preventing TF binding and thereby inactivating target genes. The purple root cultivars of carrots are characterized by high levels of DcMYB7 expression. In nonpurple carrot roots, DcMYB7 transcriptional inactivation is caused by a non-functional tandem duplication and two transposon insertions in the promoter (200). In strawberry fruit colour variation is primarily controlled by FaMYB10. An insertion of a CACTA-like transposon into the FaMYB10-2 promoter enhances the expression of FaMYB10-2 in red-fleshed strawberry (194). However, the presence of a retrotransposon in the grape MYB-related VvmybA1 TF leads to white grape cultivars unable to produce anthocyanins (87)

The second project (Chapter 5) aims to determine the role of small non-coding RNA molecules in the formation of leaf colour. The hypothesis is that green leaves have yellow striped due to small RNAs that target enzymes involved in chlorophyll biosynthesis and the yellow stripes are the result from the silencing of these genes. Also, the red leaves have pink stripes due to small RNAs that target enzymes involved in the anthocyanin biosynthesis pathway. Moreover, we investigated whether any miRNAs are differentially expressed in leaves with different colours.

To accomplish that, RNA from *C. indica* Durban, and *C. indica* cv Pretoria leaves, as well as from red and green sectors of *C. indica* cv Cleopatra, were used for generating small RNA cDNA libraries. Next-generation sequencing was used to sequence the libraries. Danial Richardson (UEA, Norwich, UK) analysed the sequence data using bioinformatic techniques to search for differentially expressed small RNAs between the samples. It was not feasible to identify differentially expressed sRNAs (that were not miRNAs) that may target chlorophyll or anthocyanin pathway genes, as the genome sequence was not available at the time. However, ten miRNAs were predicted to be differentially expressed based on the read numbers (Table 11). These ten candidate miRNAs were further analysed by northern blots, as this technique is useful for identifying size, type, abundance, and identity of RNA (205). Furthermore, housekeeping gene loading controls ensure reliable results (205). Whereases the negative aspect of the northern blot is that only one gene could be analysed at a time. As a large-scale analysis of many genes requires a lot of RNA and reagents, northern blotting is time-consuming and expensive (205).

The northern blot validation was initially performed by hybridizing a positive control U6 probe to the membrane first. Then, the membrane was scanned, stripped, re-hybridised with miRNA-specific probes, and then scanned, stripped and re-probed again. Several membranes showed a band at a significantly higher position than the mature miRNA, which required further investigation. MiRNA162 and miRNA171 probes, that produced these upper bands, were examined on northern blots including *Canna, A. thaliana* and wheat RNA. These two miRNAs were selected due to previous studies in *Arabidopsis* and wheat that only produced bands of the mature miRNA without any upper bands. MiRNA162 and miRNA171 probes were applied to the membrane before the U6 probe (Figure 27). I discovered that miRNA probe should be used prior to U6 control as the U6 signal is very strong and sometimes the stripping is not complete. Therefore, the observed upper bands were not pre-or pri-miRNAs but U6 due to the U6 probe remained bound to the membrane after stripping. This confirmed previous finding that miR162 and miR171 probes detected only the mature form of miRNA (207).

After completing the northern blot analysis of all ten miRNA candidates, we concluded that only four of them showed a detectable differential expression. These were miR166, miR529, miR530 and miR6300. However, not all four differentially expressed miRNAs' expression pattern obtained by northern blot matched perfectly the predicted expressed pattern based on sequencing read numbers.

The normalized band intensities of the northern blots for miR166s, miR530, miR529 and miR6300 showed interesting patterns (Figure 28). The expression pattern of miR530 matched the expected pattern predicted by the read numbers.

It was higher in plain red and green leaves compared to the striped leaves of Pretoria and Durban.

The read numbers of miR166 did not match the normalized band intensities obtained through northern blot. The read numbers were high in the red and green sector of *C. indica* cv Cleopatra compared to the striped leaves in Pretoria and Durban. Whereas, in the normalized bands intensity was highest in the green sector, followed by Durban, red sector and Pretoria.

The read numbers of miR529 matched the expression pattern obtained by the northern blot. The read numbers showed high expression in the plain red and green sectors of *C. indica* cv Cleopatra compared to the striped leaves in Pretoria and Durban and the same result was obtained by northern blot. This should be further investigated because the red flower of *Canna* cv Red president showed higher expression of miR529 than the yellow flowers of Tropicana president in a previous study (9).

The read numbers of miR6300 did not match with the expression pattern obtained by northern blot, either. The read numbers were high in the green sectors and Pretoria compared to the Durban and red sector. Whereas the bands intensity was high in the green sector, and low in Durban, red sector, and Pretoria. It was found that miR6300 was a key miRNA that regulated anthocyanin accumulation in *Acer. pictum. subsp. mono* leaves (211). Northern blot is more reliable than high throughput sequencing therefore it is important to validate small RNA sequencing data by northern blot. Our results also showed that read numbers did not always match the expression pattern obtained by northern blot, therefore read numbers only should not be fully trusted without validation.

Target mRNAs were predicated for the four differentially expressed miRNAs by Simon Moxon and potential targets were identified for miR166 and miR530. However, he could not find potential targets for miR529 or miR6300. The most likely reason for this is that we did not have the full transcriptome for *Canna*, therefore if a potential target site is located in a region that is not present in our transcriptome database, it cannot be found. The predicated targets for miRNA166s and miRNA530 were confirmed through 5' RACE (Figure 32). MiR166 targets ATHB-15 (DN31218), HD-ZIP III family member and also Hox32 (DN40098) another HD-ZIP III family member. Whereas the validated miR530 target, DN43985, did not show homology to any known genes.

It has been reported that miRNA166 targets HD-ZIP III transcription factors in other species to regulate the formation of secondary cell walls (46). The ability of the shoot apical meristem to function normally depends on the proper regulation of class HD-ZIP III gene activity (126). MiR166 plays a crucial role in floral development and the transition of meristems from vegetative to reproductive phases of *Canna* flowers (9). MiR166 was expressed in all the *C. indica* cvs Cleopatra red and green sectors, Durban, Pretoria (Figure 29, B). It potentially participates in secondary cell wall formation and responsible of transition from vegetative to reproductive phases in *C. indica* leaves, also.

The AtHB15 TFs plays an important role in the formation of xylem tissue, the formation of meristems, and the formation of organs (220). Furthermore, it was found that AtHB15 is responsible for the development of the pith secondary wall in *Arabidopsis*. *Arabidopsis'* stp–2d mutation repressed AtHB15, resulting in curled leaves (216). In Figure 9, *C. indica* cv Durban shows the example of curled leaves after initial growth and open over time.

Hox32 gene is targeted by miR166 in rice, and it was described that Hox32 overexpression led to rolled and filamentous leaves (130). *Canna* plants have large leaves resembling those of banana plants, but it is smaller (152). It appears that some leaves of *Canna* are rolled after a few days, and then they open, which could be due to miR166 activity in the *Canna* leaves.

MiR530 is predicted target of DN43985 but did not show similarity to any known gene (Figure 30). Banana has the closest genome sequence to *Canna* by taxonomic classification (9). MiR530 targets class III chitinase in immature banana fruits, possibly responsible for ripening-associated protein synthesis (228). In addition, miR530 target 14-3-3-like protein D in banana fruits and may be responsible for regulatory processes of banana fruit ripening (229). MiR530 was expressed at a higher level in the plain red and green sectors of *C. indica* cv Cleopatra leaves compared to striped leaves of Durban and Pretoria. This would be interesting for further study.

Taking the validate targets into account, it is unlikely that miR166 plays a role in colour formation of *Canna* leaves, as none of the validated targets seems to be related to the anthocyanin pathway. MiR530 may play a role in colour formation but at the moment it is not possible to make a conclusion either way because the validated target is an unknown gene, and it is not clear what its function is. It is also not possible to make a clear conclusion on the potential role miR529 and miR6300 may play in colour formation because we were not able to predict targets for these two miRNAs due to the incomplete transcriptome sequence. However, the most likely candidate is miR6300 as it shows a significantly higher accumulation in green sectors compared to red sectors. Indeed, miR6300 has been shown to regulate anthocyanin production in Acer pictum subsp. mono miR6300

(211). Green leaves of Acer pictum subsp. mono have significantly lower levels of cyanidin-3-O-glucoside than red leaves due to the downregulation of ApUFGT, which is targeted by miR6300 (211). Green leaves of Acer pictum subsp. mono contain much higher level of miR6300 than the red leaves, which is also the case for *Canna*, therefore it is likely that miR6300 regulates the expression of the anthocyanin biosynthesis pathway in both species.

In conclusion, we found transposon in the intron region of a MYB gene through the analysis of the genome sequence of *C. indica* cv Cleopatra. This transposon could be responsible for the colour of the red sectors by enhancing the expression of the MYB gene and when it jumps out of the MYB gene in the shoot meristem, the cell lineages generated from those cells appear green due to much lower anthocyanin production. This hypothesis needs to be further investigated. In addition, the second part of this project identified differentially expressed miRNAs in green sectors of *C. indica* cv Cleopatra and yellow striped leaves of *C. indica* cv Pretoria, and in red sectors of *C. indica* cv Cleopatra and pink striped leaves of *C. indica* cv Durban. We found different miRNA expression patterns, but none of them appear to target chlorophyl biosynthesis. Potentially, there are small RNAs which do not come from miRNA genes, but from long inverted repeats caused by gene duplications. Therefore, if there are long inverted repeats, small RNAs are produced and target genes responsible for yellow stripes of the leaf. The full genome sequence of *C. indica* cv Pretoria and Durban were not available at the time, so we were not able to identify long inverted repeats and small RNA target genes. *Antirrhinum majus* genome contained a long-inverted repeat that produced small RNAs and targeted the Am4'CGT gene, which made the flower yellow (230). It could be a similar situation of *C. indica* cv Pretoria and *C. indica* cv Durban.

Future Work

There are two important findings in my thesis. The first is the identification of a potential TE in the intronic region of a MYB TF that may enhance the expression of the anthocyanin pathway. The other is the identification of a few differentially expressed miRNAs in differently coloured Canna leaves. Bothe of these findings need to be followed up by carryon out future experiments.

The presence of the TE in the intronic region of a MYB TF should be validated by sequencing PCR products generated by primers that amplify the insertion sites of the TE at each end. One difficulty with this experiment is that at the moment it is not clear what the genotype of the mesophyll cells. The red colour is only visible in the single epidermal cell layer, and it is not known at the moment whether it is because the genes involved in anthocyanin expression are not expressed in the mesophyll cells or because the genome of the mesophyll cells are different from the epidermal cells. A series of PCR reactions using primers specific to the transposon sequence identified in the intron region of the MYB gene, as well as the flanking regions of that transposon insertion site, on DNA samples extracted from either red or green sectors could shed light on this question. There are different potential scenarios, which would inform future experiments. For example, if the green sectors contain genetically identical cells without the transposon, then DNA from green sector can be extracted and amplification and sequencing the transposon insertion site would show the genome sequence after the transposon jumped out. However, the green sector may be a chimera where the mesophyll cells still contain the transposon (but are green because the anthocyanin pathway genes are not expressed) and the transposon jumped out in a shoot meristem cell that gives rise to the epidermal cell layer. In that case, primers flanking the transposition site would amplify the insertion site without the transposon from the DNA of the epidermal cells and

primer pairs where one primer is on the flanking region and the other is in the transposon would amplify the junction regions between the transposon and the host genome from the DNA of the mesophyll cells.

The other line of future work is to better understand the role of small RNAs in colour formation in *Canna* leaves. One aspect is the role of miRNAs and another is the potential role of siRNAs. We were only able to predict targets for two miRNAs due to the lack of complete genome and therefore transcriptome sequences. However, since submission of my thesis, a more complete genome sequence has been published and therefore it should be possible to predict and target further miRNA targets. In addition, sequencing the whole genome of *C. indica* cv Pretoria and *C. indica* cv Durban, would facilitate to identification of long-inverted repeats that may produce small RNAs targeting chlorophyll and/or anthocyanin biosynthetic genes, that could cause the stripes in these cultivars.

References

1. Hamilton AJ, Baulcombe DC. A species of small antisense RNA in posttranscriptional gene silencing in plants. Science (1979). 1999;286(5441).
2. Pal-Bhadra M, Bhadra U, Birchler JA. RNAi related mechanisms affect both transcriptional and posttranscriptional transgene silencing in Drosophila. Mol Cell. 2002;9(2).
3. Sijen T, Vijn I, Rebocho A, van Blokland R, Roelofs D, Mol JNM, et al. Transcriptional and posttranscriptional gene silencing are mechanistically related. Current Biology. 2001;11(6).
4. Elbashir SM, Harborth J, Lendeckel W, Yalcin A, Weber K, Tuschl T. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature. 2001;411(6836).
5. Lee RC, Feinbaum RL, Ambros V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell. 1993;75(5).
6. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res. 2006;34(Database issue).
7. Borges F, Martienssen RA. The expanding world of small RNAs in plants. Vol. 16, Nature Reviews Molecular Cell Biology. 2015.

8. Chen HM, Chen LT, Patel K, Li YH, Baulcombe DC, Wu SH. 22-Nucleotide RNAs trigger secondary siRNA biogenesis in plants. Proc Natl Acad Sci U S A. 2010;107(34).

9. Roy S, Tripathi AM, Yadav A, Mishra P, Nautiyal CS. Identification and expression analyses of miRNAs from two contrasting flower color cultivars of canna by deep sequencing. PLoS One. 2016;11(1).

10. Eldem V, Okay S, Ünver T. Plant microrna: New players in functional genomics. Vol. 37, Turkish Journal of Agriculture and Forestry. 2013.

11. Yu B, Bi L, Zheng B, Ji L, Chevalier D, Agarwal M, et al. The FHA domain proteins DAWDLE in Arabidopsis and SNIP1 in humans act in small RNA biogenesis. Proc Natl Acad Sci U S A. 2008;105(29).

12. Hiraguri A, Itoh R, Kondo N, Nomura Y, Aizawa D, Murai Y, et al. Specific interactions between Dicer-like proteins and HYL1/DRB-family dsRNA-binding proteins in Arabidopsis thaliana. Plant Mol Biol. 2005;57(2).

13. Park W, Li J, Song R, Messing J, Chen X. CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in Arabidopsis thaliana. Current Biology. 2002;12(17).

14. Fujioka Y, Utsumi M, Ohba Y, Watanabe Y. Location of a possible miRNA processing site in SmD3/SmB nuclear bodies in arabidopsis. Plant Cell Physiol. 2007;48(9).

15. Fang Y, Spector DL. Identification of Nuclear Dicing Bodies Containing Proteins for MicroRNA Biogenesis in Living Arabidopsis Plants. Current Biology. 2007;17(9).

16. Yang L, Liu Z, Lu F, Dong A, Huang H. SERRATE is a novel nuclear regulator in primary microRNA processing in Arabidopsis. Plant Journal. 2006;47(6).

17. Ramachandran V, Chen X. Degradation of microRNAs by a family of exoribonucleases in Arabidopsis. Science (1979). 2008;321(5895).

18. Yu B, Yang Z, Li J, Minakhina S, Yang M, Padgett RW, et al. Methylation as a crucial step in plant microRNA biogenesis. Science (1979). 2005;307(5711).

19. Li J, Yang Z, Yu B, Liu J, Chen X. Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in Arabidopsis. Current Biology. 2005;15(16).

20. Mee YP, Wu G, Gonzalez-Sulser A, Vaucheret H, Poethig RS. Nuclear processing and export of microRNAs in Arabidopsis. Proc Natl Acad Sci U S A. 2005;102(10).

21. Bollman KM, Aukerman MJ, Park MY, Hunter C, Berardini TZ, Scott Poethig R. HASTY, the Arabidopsis ortholog of exportin 5/MSN5, regulates phase change and morphogenesis. Vol. 130, Development. 2003.

22. Baumberger N, Baulcombe DC. Arabidopsis ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. Proc Natl Acad Sci U S A. 2005;102(33).

23. Yang X, Li L. Analyzing the microRNA transcriptome in plants using deep sequencing data. Vol. 1, Biology. 2012.

24. Vaucheret H, Vazquez F, Crété P, Bartel DP. The action of ARGONAUTE1 in the miRNA pathway and its regulation by the miRNA pathway are crucial for plant development. Genes Dev. 2004;18(10).

25. Parker JS, Roe SM, Barford D. Crystal structure of a PIWI protein suggests mechanisms for siRNA recognition and slicer activity. EMBO Journal. 2004;23(24).

26. Liu J, Carmell MA, Rivas F v., Marsden CG, Thomson JM, Song JJ, et al. Argonaute2 is the catalytic engine of mammalian RNAi. Science (1979). 2004;305(5689).

27. Cerutti L, Mian N, Bateman A. Domains in gene silencing and cell differentiation proteins: tTe novel PAZ domain and redefinition of the Piwi domain. Trends Biochem Sci. 2000;25(10).

28. Carrington JC, Ambros V. Role of microRNAs in plant and animal development. Vol. 301, Science. 2003.

29. Dong Q, Hu B, Zhang C. microRNAs and Their Roles in Plant Development. Vol. 13, Frontiers in Plant Science. 2022.

30. Jodder J. miRNA-mediated regulation of auxin signaling pathway during plant development and stress responses. Vol. 45, Journal of Biosciences. 2020.

31. Schoof H, Lenhard M, Haecker A, Mayer KFX, Jürgens G, Laux T. The stem cell population of Arabidopsis shoot meristems is maintained by a regulatory loop between the CLAVATA and WUSCHEL genes. Cell. 2000;100(6).

32. Veit B. Hormone mediated regulation of the shoot apical meristem. Vol. 69, Plant Molecular Biology. 2009.

33. Liu PP, Montgomery TA, Fahlgren N, Kasschau KD, Nonogaki H, Carrington JC. Repression of AUXIN RESPONSE FACTOR10 by microRNA160 is critical for seed germination and post-germination stages. Plant Journal. 2007;52(1).

34. Piazza P, Jasinski S, Tsiantis M. Evolution of leaf developmental mechanisms. Vol. 167, New Phytologist. 2005.

35. Hay A, Craft J, Tsiantis M. Plant hormones and homeoboxes: Bridging the gap? Vol. 26, BioEssays. 2004.

36. Kidner CA, Martienssen RA. Spatially restricted microRNA directs leaf polarity through ARGONAUTE1. Nature. 2004;428(6978).

37. Liu Q, Yao X, Pi L, Wang H, Cui X, Huang H. The ARGONAUTE10 gene modulates shoot apical meristem maintenance and establishment of leaf polarity by repressing miR165/166 in Arabidopsis. Plant Journal. 2009;58(1).

38. Chitwood DH, Nogueira FTS, Howell MD, Montgomery TA, Carrington JC, Timmermans MCP. Pattern formation via small RNA mobility. Genes Dev. 2009;23(5).

39. Knauer S, Holt AL, Rubio-Somoza I, Tucker EJ, Hinze A, Pisch M, et al. A Protodermal miR394 Signal Defines a Region of Stem Cell Competence in the Arabidopsis Shoot Meristem. Dev Cell. 2013;24(2).

40. Kim JH, Choi D, Kende H. The AtGRF family of putative transcription factors is involved in leaf and cotyledon growth in Arabidopsis. Plant Journal. 2003;36(1).

41. Debernardi JM, Rodriguez RE, Mecchia MA, Palatnik JF. Functional specialization of the plant miR396 regulatory network through distinct microRNA-target interactions. PLoS Genet. 2012;8(1).

42. Palatnik JF, Wollmann H, Schommer C, Schwab R, Boisbouvier J, Rodriguez R, et al. Sequence and Expression Differences Underlie Functional Specialization of Arabidopsis MicroRNAs miR159 and miR319. Dev Cell. 2007;13(1).

43. Koyama T, Mitsuda N, Seki M, Shinozaki K, Ohme-Takagi M. TCP transcription factors regulate the activities of ASYMMETRIC LEAVES1 and miR164, as well as the auxin response, during differentiation of leaves in Arabidopsis. Plant Cell. 2010;22(11).

44. Blein T, Pulido A, Vialette-Guiraud A, Nikovics K, Morin H, Hay A, et al. A conserved molecular framework for compound leaf development. Science (1979). 2008;322(5909).

45. Kasschau KD, Xie Z, Allen E, Llave C, Chapman EJ, Krizan KA, et al. P1/HC-Pro, a viral suppressor of RNA silencing, interferes with Arabidopsis development and miRNA function. Dev Cell. 2003;4(2).

46. Du Q, Wang H. The role of HD-ZIP III transcription factors and miR165/166 in vascular development and secondary cell wall formation. Plant Signal Behav. 2015;10(10).

47. MCCLINTOCK B. Controlling elements and the gene. Cold Spring Harb Symp Quant Biol. 1956;21.

48. Bennetzen JL, Wang H. The contributions of transposable elements to the structure, function, and evolution of plant genomes. Vol. 65, Annual Review of Plant Biology. 2014.

49. Gregory TR, Nicol JA, Tamm H, Kullman B, Kullman K, Leitch IJ, et al. Eukaryotic genome size databases. Nucleic Acids Res. 2007;35(SUPPL. 1).

50. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, et al. A unified classification system for eukaryotic transposable elements. Vol. 8, Nature Reviews Genetics. 2007.

51. Jiao Y, Peluso P, Shi J, Liang T, Stitzer MC, Wang B, et al. Improved maize reference genome with single-molecule technologies. Nature. 2017;546(7659).

52. Sahebi M, Hanafi MM, van Wijnen AJ, Rice D, Rafii MY, Azizi P, et al. Contribution of transposable elements in the plant's genome. Vol. 665, Gene. 2018.

53. Qiu F, Baack EJ, Whitney KD, Bock DG, Tetreault HM, Rieseberg LH, et al. Phylogenetic trends and environmental correlates of nuclear genome size variation in Helianthus sunflowers. New Phytologist. 2019;221(3).

54. Bilinski P, Albert PS, Berg JJ, Birchler JA, Grote MN, Lorant A, et al. Parallel altitudinal clines reveal trends in adaptive evolution of genome size in Zea mays. PLoS Genet. 2018;14(5).

55. Ågren JA, Wright SI. Co-evolution between transposable elements and their hosts: A major factor in genome size evolution? Chromosome Research. 2011;19(6).

56. Bhattacharyya MK, Smith AM, Ellis THN, Hedley C, Martin C. The wrinkled-seed character of pea described by Mendel is caused by a transposon-like insertion in a gene encoding starch-branching enzyme. Cell. 1990;60(1).

57. Liu J, Osbourn A, Ma P. MYB transcription factors as regulators of phenylpropanoid metabolism in plants. Vol. 8, Molecular Plant. 2015.

58. Tanaka Y, Sasaki N, Ohmiya A. Biosynthesis of plant pigments: Anthocyanins, betalains and carotenoids. Vol. 54, Plant Journal. 2008.

59. Barnaba C, Medina-Meza IG. Flavonoids Ability to Disrupt Inflammation Mediated by Lipid and Cholesterol Oxidation. In: Advances in Experimental Medicine and Biology. 2019.

60. Grotewold E. The genetics and biochemistry of floral pigments. Vol. 57, Annual Review of Plant Biology. 2006.

61. Liu J, Liu M, Wang J, Zhang J, Miao H, Wang Z, et al. Transcription factor MaMADS36 plays a central role in regulating banana fruit ripening. J Exp Bot. 2021 Oct 26;72(20):7078–91.

62. Chandler S, Tanaka Y. Genetic modification in floriculture. Vol. 26, Critical Reviews in Plant Sciences. 2007.

63. Andersen ØM, Markham KR. Flavonoids: Chemistry, biochemistry and applications. Flavonoids: Chemistry, Biochemistry and Applications. 2005.

64. Aaby K, Mazur S, Nes A, Skrede G. Phenolic compounds in strawberry (Fragaria x ananassa Duch.) fruits: Composition in 27 cultivars and changes during ripening. Food Chem. 2012;132(1).

65. Ayash A, Al-Tameemi K, Nassour R. Anthocyanin pigments: Structure and biological importance. Article in Journal of Chemical and Pharmaceutical Sciences. 2020;13(4).

66. Srivastava J, Vankar PS. Carotenoids: As natural food colorant from Canna flowers. Pigment and Resin Technology. 2015;44(1).

67. Honda C, Bessho H, Murai M, Iwanami H, Moriya S, Abe K, et al. Effect of temperature on anthocyanin synthesis and ethylene production in the fruit of earlyand medium-maturing apple cultivars during ripening stages. HortScience. 2014;49(12).

68. Seeram NP. Berry fruits: Compositional elements, biochemical activities, and the impact of their intake on human health, performance, and disease. In: Journal of Agricultural and Food Chemistry. 2008.

69. Hidalgo M, Martin-Santamaria S, Recio I, Sanchez-Moreno C, de Pascual-Teresa B, Rimbach G, et al. Potential anti-inflammatory, anti-adhesive, anti/estrogenic, and angiotensin-converting enzyme inhibitory activities of anthocyanins and their gut metabolites. Genes Nutr. 2012;7(2).

70. Bowtell JL, Aboo-Bakkar Z, Conway ME, Adlam ALR, Fulford J. Enhanced task-related brain activation and resting perfusion in healthy older adults after chronic blueberry supplementation. Applied Physiology, Nutrition and Metabolism. 2017;42(7).

71. Deroles S. Anthocyanin Biosynthesis in Plant Cell Cultures: A Potential Source of Natural Colourants. In: Anthocyanins. 2008.

72. Martin C, Paz-Ares J. MYB transcription factors in plants. Vol. 13, Trends in Genetics. 1997.

73. Ambawat S, Sharma P, Yadav NR, Yadav RC. MYB transcription factor genes as regulators for plant responses: An overview. Vol. 19, Physiology and Molecular Biology of Plants. 2013.

74. Weston K, Bishop JM. Transcriptional activation by the v-myb oncogene and its cellular progenitor, c-myb. Cell. 1989;58(1).

75. Cao Y, Han Y, Li D, Lin Y, Cai Y. MYB transcription factors in chinese pear (Pyrus bretschneideri rehd.): Genome-wide identification, classification, and expression profiling during fruit development. Front Plant Sci. 2016;7(APR2016).

76. Paz-Ares J, Ghosal D, Wienand U, Peterson PA, Saedler H. The regulatory c1 locus of Zea mays encodes a protein with homology to myb proto-oncogene products and with structural similarities to transcriptional activators. EMBO J. 1987;6(12).

77. Jin H, Martin C. Multifunctionality and diversity within the plant MYB-gene family. Vol. 41, Plant Molecular Biology. 1999.

78. Cao Y, Meng D, Han Y, Chen T, Jiao C, Chen Y, et al. Comparative analysis of B-BOX genes and their expression pattern analysis under various treatments in Dendrobium officinale. BMC Plant Biol. 2019;19(1).

79. Ramya M, Kwon OK, An HR, Park PM, Baek YS, Park PH. Floral scent: Regulation and role of MYB transcription factors. Vol. 19, Phytochemistry Letters. 2017.

80. Tak H, Negi S, Ganapathi TR. Overexpression of MusaMYB31, a R2R3 type MYB transcription factor gene indicate its role as a negative regulator of lignin biosynthesis in banana. PLoS One. 2017;12(2).

81. Sonbol FM, Fornalé S, Capellades M, Encina A, Touriño S, Torres JL, et al. The maize ZmMYB42 represses the phenylpropanoid pathway and affects the cell wall structure, composition and degradability in Arabidopsis thaliana. Plant Mol Biol. 2009;70(3).

82. Zhu L, Shan H, Chen S, Jiang J, Gu C, Zhou G, et al. The Heterologous Expression of the Chrysanthemum R2R3-MYB Transcription Factor CmMYB1 Alters Lignin Composition and Represses Flavonoid Synthesis in Arabidopsis thaliana. PLoS One. 2013;8(6).

83. Takos AM, Jaffé FW, Jacob SR, Bogs J, Robinson SP, Walker AR. Light-induced expression of a MYB gene regulates anthocyanin biosynthesis in red apples. Plant Physiol. 2006;142(3).

84. Vimolmangkang S, Han Y, Wei G, Korban SS. An apple MYB transcription factor, MdMYB3, is involved in regulation of anthocyanin biosynthesis and flower development. BMC Plant Biol. 2013;13(1).

85. Espley R V., Hellens RP, Putterill J, Stevenson DE, Kutty-Amma S, Allan AC. Red colouration in apple fruit is due to the activity of the MYB transcription factor, MdMYB10. Plant Journal. 2007;49(3).

86. Chagné D, Lin-Wang K, Espley R v., Volz RK, How NM, Rouse S, et al. An ancient duplication of apple MYB transcription factors is responsible for novel red fruit-flesh phenotyp. Plant Physiol. 2013;161(1).

87. Kobayashi S, Goto-Yamamoto N, Hirochika H. Retrotransposon-Induced Mutations in Grape Skin Color. Science (1979). 2004;304(5673).

88. Kavas M, Baloğlu MC, Atabay ES, Ziplar UT, Daşgan HY, Ünver T. Genome-wide characterization and expression analysis of common bean bHLH transcription factors in response to excess salt concentration. Molecular Genetics and Genomics. 2016;291(1).

89. Feller A, MacHemer K, Braun EL, Grotewold E. Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. Plant Journal. 2011;66(1).

90. Lim SH, Kim DH, Kim JK, Lee JY, Ha SH. A radish basic helix-loop-helix transcription factor, rstt8 acts a positive regulator for anthocyanin biosynthesis. Front Plant Sci. 2017;8.

91. Ji X, Nie X, Liu Y, Zheng L, Zhao H, Zhang B, et al. A bHLH gene from Tamarix hispida improves abiotic stress tolerance by enhancing osmotic potential and decreasing reactive oxygen species accumulation. Tree Physiol. 2015;36(2).

92. Pires N, Dolan L. Origin and diversification of basic-helix-loop-helix proteins in plants. Mol Biol Evol. 2010;27(4).

93. Heim MA, Jakoby M, Werber M, Martin C, Weisshaar B, Bailey PC. The basic helix-loop-helix transcription factor family in plants: A genome-wide study of protein structure and functional diversity. Mol Biol Evol. 2003;20(5).

94. Ludwig SR, Habera LF, Dellaporta SL, Wessler SR. Lc, a member of the maize R gene family responsible for tissue-specific anthocyanin production, encodes a protein similar to transcriptional activators and contains the myc-homology region. Proc Natl Acad Sci U S A. 1989;86(18).

95. Chandler VL, Radicella JP, Robbins TP, Chen J, Turks D. Two regulatory genes of the maize anthocyanin pathway are homologous: isolation of B utilizing R genomic sequences. Plant Cell. 1989;1(12).

96. Roth BA, Goff SA, Klein TM, Fromm ME. C1- and R-dependent expression of the maize Bz1 gene requires sequences with homology to mammalian myb and myc binding sites. Plant Cell. 1991;3(3).

97. Nesi N, Debeaujon I, Jond C, Pelletier G, Caboche M, Lepiniec L. The TT8 gene encodes a basic helix-loop-helix domain protein required for expression of DFR and BAN genes in Arabidopsis siliques. Plant Cell. 2000;12(10).

98. Matus JT, Poupin MJ, Cañón P, Bordeu E, Alcalde JA, Arce-Johnson P. Isolation of WDR and bHLH genes related to flavonoid synthesis in grapevine (Vitis vinifera L.). Plant Mol Biol. 2010;72(6).

99. Zhou H, Lin-Wang K, Wang H, Gu C, Dare AP, Espley R v., et al. Molecular genetics of blood-fleshed peach reveals activation of anthocyanin biosynthesis by NAC transcription factors. Plant Journal. 2015;82(1).

100. Zhao M, Li J, Zhu L, Chang P, Li L, Zhang L. Identification and characterization of MYB-bHLH-WD40 regulatory complex members controlling anthocyanidin biosynthesis in blueberry fruits development. Genes (Basel). 2019;10(7).

101. Peng H huan, Shan W, Kuang J fei, Lu W jin, Chen J ye. Molecular characterization of cold-responsive basic helix-loop-helix transcription factors MabHLHs that interact with MaICE1 in banana fruit. Planta. 2013;238(5).

102. Zhao ML, Wang JN, Shan W, Fan JG, Kuang JF, Wu KQ, et al. Induction of jasmonate signalling regulators MaMYC2s and their physical interactions with MaICE1 in methyl jasmonate-induced chilling tolerance in banana fruit. Plant Cell Environ. 2013;36(1).

103. Xiao YY, Kuang JF, Qi XN, Ye YJ, Wu ZX, Chen JY, et al. A comprehensive investigation of starch degradation process and identification of a transcriptional activator MabHLH6 during banana fruit ripening. Plant Biotechnol J. 2018;16(1).

104. Fan ZQ, Ba LJ, Shan W, Xiao YY, Lu WJ, Kuang JF, et al. A banana R2R3-MYB transcription factor MaMYB3 is involved in fruit ripening through modulation of starch degradation by repressing starch degradation-related genes and MabHLH6. Plant Journal. 2018;96(6).

105. Gehring WJ. Homeo boxes in the study of development. Science (1979). 1987;236(4805).

106. Vollbrecht E, Veit B, Sinha N, Hake S. The developmental gene Knotted-1 is a member of a maize homeobox gene family. Nature. 1991;350(6315).

107. Ruberti I, Sessa G, Lucchetti S, Morelli G. A novel class of plant proteins containing a homeodomain with a closely linked leucine zipper motif. EMBO Journal. 1991;10(7).

108. Mukherjee K, Brocchieri L, Bürglin TR. A comprehensive classification and evolutionary analysis of plant homeobox genes. Mol Biol Evol. 2009;26(12).

109. Ariel FD, Manavella PA, Dezar CA, Chan RL. The true story of the HD-Zip family. Vol. 12, Trends in Plant Science. 2007.

110. Gago GM, Almoguera C, Jordano J, Gonzalez DH, Chan RL. Hahb-4, a homeobox-leucine zipper gene potentially involved in abscisic acid-dependent responses to water stress in sunflower. Plant Cell Environ. 2002;25(5).

111. Olsson ASB, Engström P, Söderman E. The homeobox genes ATHB12 and ATHB7 encode potential regulators of growth in response to water deficit in Arabidopsis. Plant Mol Biol. 2004;55(5).

112. Sessa G, Carabelli M, Sassi M, Ciolfi A, Possenti M, Mittempergher F, et al. A dynamic balance between gene activation and repression regulates the shade avoidance response in Arabidopsis. Genes Dev. 2005;19(23).

113. Morelli G, Ruberti I. Light and shade in the photocontrol of Arabidopsis growth. Vol. 7, Trends in Plant Science. 2002.

114. Prigge MJ, Otsuga D, Alonso JM, Ecker JR, Drews GN, Clark SE. Class III homeodomain-leucine zipper gene family members have overlapping, antagonistic, and distinct roles in Arabidopsis development. Plant Cell. 2005;17(1).

115. Otsuga D, DeGuzman B, Prigge MJ, Drews GN, Clark SE. REVOLUTA regulates meristem initiation at lateral positions. Plant Journal. 2001;25(2).

116. Emery JF, Floyd SK, Alvarez J, Eshed Y, Hawker NP, Izhaki A, et al. Radial Patterning of Arabidopsis Shoots by Class III HD-ZIP and KANADI Genes. Current Biology. 2003;13(20).

117. Schrick K, Nguyen D, Karlowski WM, Mayer KFX. START lipid/sterol-binding domains are amplified in plants and are predominantly associated with homeodomain transcription factors. Genome Biol. 2004;5(6).

118. Mukherjee K, Bürglin TR. MEKHLA, a novel domain with similarity to PAS domains, is fused to plant homeodomain-leucine zipper III proteins. Plant Physiol. 2006;140(4).

119. Prigge MJ, Clark SE. Evolution of the class III HD-Zip gene family in land plants. Evol Dev. 2006;8(4).

120. Baima S, Possenti M, Matteucci A, Wisman E, Altamura MM, Ruberti I, et al. The arabidopsis ATHB-8 HD-zip protein acts as a differentiation-promoting transcription factor of the vascular meristems. Plant Physiol. 2001;126(2).

121. Kang J, Dengler N. Cell cycling frequency and expression of the homeobox gene ATHB-8 during leaf vein development in Arabidopsis. Planta. 2002;216(2).

122. Mallory AC, Reinhart BJ, Jones-Rhoades MW, Tang G, Zamore PD, Barton MK, et al. MicroRNA control of PHABULOSA in leaf development: Importance of pairing to the microRNA 5' region. EMBO Journal. 2004;23(16).

123. Cosgrove DJ, Jarvis MC. Comparative structure and biomechanics of plant primary and secondary cell walls. Vol. 3, Frontiers in Plant Science. 2012.

124. Taylor-Teeples M, Lin L, de Lucas M, Turco G, Toal TW, Gaudinier A, et al. An Arabidopsis gene regulatory network for secondary cell wall synthesis. Nature. 2015;517(7536).

125. Kim J, Jung JH, Reyes JL, Kim YS, Kim SY, Chung KS, et al. microRNA-directed cleavage of ATHB15 mRNA regulates vascular development in Arabidopsis inflorescence stems. Plant Journal. 2005;42(1).

126. Williams L, Grigg SP, Xie M, Christensen S, Fletcher JC. Regulation of Arabidopsis shoot apical meristem and lateral organ formation by microRNA miR166g and its AtHD-ZIP target genes. Development. 2005;132(16).

127. Agalou A, Purwantomo S, Övernäs E, Johannesson H, Zhu X, Estiati A, et al. A genome-wide survey of HD-Zip genes in rice and analysis of drought-responsive family members. Plant Mol Biol. 2008;66(1–2).

128. Chen H, Fang R, Deng R, Li J. The OsmiRNA166b-OsHox32 pair regulates mechanical strength of rice plants by modulating cell wall biosynthesis. Plant Biotechnol J. 2021;19(7).

129. Zhang J, Zhang H, Srivastava AK, Pan Y, Bai J, Fang J, et al. Knockdown of rice microRNA166 confers drought resistance by causing leaf rolling and altering stem xylem development. Plant Physiol. 2018;176(3).

130. Li Y ying, Shen A, Xiong W, Sun Q lin, Luo Q, Song T, et al. Overexpression of OsHox32 Results in Pleiotropic Effects on Plant Type Architecture and Leaf Development in Rice. Rice. 2016;9(1).

131. Bhakta S, Tak H, Ganapathi TR. Exploring diverse roles of micro RNAs in banana: Current status and future prospective. Physiol Plant. 2021;173(4).

132. Gray J. RHS Plant Trials and Awards RHS Plant Trials and Awards Canna. 2003; Available from: www.rhs.org.ukwww.rhs.org.uk

133. Plant Resources of South-East Asia.

134. Khoshoo TN, Mukherjee I. Genetic-evolutionary studies on cultivated cannas - VI. Origin and evolution of ornamental taxa. Theoretical and Applied Genetics. 1970;40(5).

135. ADEOGUN OO, ADEKUNLE AA, ADONGBEDE EM. Effects of Leaf Extracts of Selected Plants on Quality of Stored Citrus sinensis (Sweet Orange) Juice. Not Sci Biol. 2017;9(2).

136. Lim TK. Canna indica. In: Edible Medicinal and Non-Medicinal Plants [Internet]. Dordrecht: Springer Netherlands; 2016. p. 69–81. Available from: http://link.springer.com/10.1007/978-94-017-7276-1_3

137. Indrayan AK, Bhojak NK, Kumar N, Shatru A, Gaur A. Chemical composition and antimicrobial activity of the essential oil from the rhizome of Canna indica Linn. Indian Journal of Chemistry - Section B Organic and Medicinal Chemistry. 2011;50(8).

138. Linnaeus C. Hortus Cliffortianus. Amsterdam. 1738.

139. Nannfeld JA. SPECIES PLANTARUM. Taxon. 1953;2(3).

140. Miller P. The gardeners dictionary: containing the methods of cultivating and improving the kitchen, fruit and flower garden, as also the physick garden, wilderness, conservatory, and vineyard. The gardeners dictionary: containing the methods of cultivating and improving the kitchen, fruit and flower garden, as also the physick garden, wilderness, conservatory, and vineyard. 2011.

141. Aiton W, Bauer FA, Ehret GD, Nicol George, Sowerby J. Hortus Kewensis, or, A catalogue of the plants cultivated in the Royal Botanic Garden at Kew /by William Aiton ... Hortus Kewensis, or, A catalogue of the plants cultivated in the Royal Botanic Garden at Kew /by William Aiton ... 2011.

142. Jussieu AL de. Antonii Laurentii de Jussieu Genera plantarum :secundum ordines naturales disposita, juxta methodum in Horto regio parisiensi exaratam, anno M.DCC.LXXIV. Antonii Laurentii de Jussieu Genera plantarum :secundum ordines naturales disposita, juxta methodum in Horto regio parisiensi exaratam, anno M.DCC.LXXIV. 2011.

143. Lindley J. The vegetable kingdom : or, The structure, classification, and uses of plants, illustrated upon the natural system / by John Lindley. The vegetable kingdom : or, The structure, classification, and uses of plants, illustrated upon the natural system / by John Lindley. 2011.

144. Bentham G, Hooker JD. Genera plantarum :ad exemplaria imprimis in Herberiis Kewensibus servata definita /auctoribus G. Bentham et J.D. Hooker. Genera plantarum :ad exemplaria imprimis in Herberiis Kewensibus servata definita /auctoribus G. Bentham et J.D. Hooker. 2011.

145. Tanaka N, Uchiyama H, Matoba H, Koyama T. Karyological analysis of the genus Canna (Cannaceae). Plant Systematics and Evolution. 2009;280(1–2).

146. Venkatasubban KR. A preliminary survey of chromosome numbers in scitamineÆ of Bentham and Hooker. Proceedings of the Indian Academy of Sciences - Section B. 1946;23(6).

147. Belling J. Chromosomes of canna and of hemerocallis. Journal of Heredity. 1925;16(12).

148. Choudhary AK, Kumar S, Sharma C, Kumar P. Performance of constructed wetland for the treatment of pulp and paper mill wastewater. In: World Environmental and Water Resources Congress 2011: Bearing Knowledge for Sustainability - Proceedings of the 2011 World Environmental and Water Resources Congress. 2011.

149. Ayaz SÇ, Akça L. Treatment of wastewater by natural systems. Environ Int. 2001;26(3).

150. Neralla S, Weaver RW, Varvel TW, Lesikar BJ. Phytoremediation and on-site treatment of septic effluents in sub-surface flow constructed wetlands. Environmental Technology (United Kingdom). 1999;20(11).

151. Grootjen CJ, Bouman F. Seed structure in cannaceae: Taxonomic and ecological implications. Ann Bot. 1988;61(3).

152. Al-Snafi AE. Bioactive components and pharmacological effects of Canna indica-An Overview. International Journal of Pharmacology and toxicology. 2015;5(2).

153. Odugbemi TO, Akinsulire OR, Aibinu IE, Fabeku PO. Medicinal plants useful for malaria therapy in Okeigbo, Ondo State, Southwest Nigeria. African Journal of Traditional, Complementary and Alternative Medicines. 2007;4(2).

154. Thepouyporn A, Yoosook C, Chuakul W, Thirapanmethee K, Napaswad C, Wiwat C. Purification and characterization of anti-HIV-1 protein from Canna indica L. leaves. Southeast Asian Journal of Tropical Medicine and Public Health. 2012;43(5).

155. Lin FR, Ferrucci L, Metter EJ, An Y, Zonderman AB, Resnick SM. Hearing Loss and Cognition in the Baltimore Longitudinal Study of Aging. Neuropsychology. 2011;25(6).

156. Talluri MR, Killari KN, Viswanadha Murthy Manepalli NVS, Konduri P, Bandaru KK. Protective effect of Canna indica on cerebral ischemia-reperfusion injury in rats. Agriculture and Natural Resources. 2017;51(6).

157. Chigurupati S, Abdul Rahman Alharbi N, Sharma AK, Alhowail A, Vardharajula VR, Vijayabalan S, et al. Pharmacological and pharmacognostical valuation of Canna indica leaves extract by quantifying safety profile and neuroprotective potential. Saudi J Biol Sci. 2021;28(10).

158. Bolade OP, Akinsiku AA, Adeyemi AO, Jolayemi GE, Williams AB, Benson NU. Qualitative analysis, total phenolic content, FT-IR and GC-MS characterisation of Canna indica: Bioreducing agent for nanoparticles synthesis. In: Journal of Physics: Conference Series. 2019.

159. Bachheti RK, Rawat GS, Joshi A, Pandey DP. Phytochemical investigation of aerial parts of Canna indica collected from Uttarakhand India. Int J Pharmtech Res. 2013;5(2).

160. Joshi YM, Kadam VJ, Patil Y v, Kaldhone PR. Investigation of Hepatoprotective activity of Aerial Parts of Canna indica L . on carbon tetrachloride treated rats. J Pharm Res. 2009;2(12).

161. TOMLINSON PB. The anatomy of Canna. Journal of the Linnean Society of London, Botany. 1961;56(368).

162. Löv L. Zur Kenntnis der Entfaltungszellen monokotyler Blätter. Flora oder Allgemeine Botanische Zeitung. 1926;120(4).

163. Miao MZ, Liu HF, Kuang YF, Zou P, Liao JP. Floral vasculature and ontogeny in Canna indica. Nord J Bot. 2014;32(4).

164. Noe AC. Anatomy of Monocotyledons Systematische Anatomie der Monokotyledonen. Vol. IV. Farinosae. H. Solereder , F. J. Meyer . Botanical Gazette. 1930;89(1).

165. Glinos E, Cocucci AA. Pollination biology of Canna indica (Cannaceae) with particular reference to the functional morphology of the style. Plant Systematics and Evolution. 2011;291(1).

166. Retail Plant Focus Canna [Internet]. Available from: www.besMDlants.co.uk/trade

167. Taylor's encyclopedia of garden plants. Choice Reviews Online. 2004;41(07).

168. Phasion_Durban_patent.

169. Xu P, Billmeier M, Mohorianu I, Green D, Fraser WD, Dalmay T. An improved protocol for small RNA library construction using High Definition adapters. Methods in Next Generation Sequencing. 2015;2(1).

170. Li Z, Parris S, Saski CA. A simple plant high-molecular-weight DNA extraction method suitable for single-molecule technologies. Plant Methods. 2020;16(1).

171. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: Scalable and accurate long-read assembly via adaptive κ-mer weighting and repeat separation. Genome Res. 2017;27(5).

172. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019;37(8).

173. Kovaka S, Zimin A V., Pertea GM, Razaghi R, Salzberg SL, Pertea M. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. Genome Biol. 2019;20(1).

174. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol. 2016;34(5).

175. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12).

176. Pucker B. Automatic identification and annotation of MYB gene family members in plants. BMC Genomics. 2022;23(1).

177. Ou S, Su W, Liao Y, Chougule K, Agda JRA, Hellinga AJ, et al. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. Genome Biol. 2019;20(1).

178. Xu W, Dubos C, Lepiniec L. Transcriptional control of flavonoid biosynthesis by MYB-bHLH-WDR complexes. Vol. 20, Trends in Plant Science. 2015.

179. Vilperte V, Lucaciu CR, Halbwirth H, Boehm R, Rattei T, Debener T. Hybrid de novo transcriptome assembly of poinsettia (Euphorbia pulcherrima Willd. Ex Klotsch) bracts. BMC Genomics. 2019;20(1).

180. Moreira PA, Oliveira DA. Leaf age affects the quality of DNA extracted from Dimorphandra mollis (Fabaceae), a tropical tree species from the Cerrado region of Brazil. Genet Mol Res. 2011;10(1).

181. John ME. An efficient method for isolation of RNA and DNA from plants containing polyphenolics. Nucleic Acids Res. 1992;20(9).

182. Kang J, Myung SL, Gorenstein DG. The enhancement of PCR amplification of a random sequence DNA library by DMSO and betaine: Application to in vitro combinatorial selection of aptamers. J Biochem Biophys Methods. 2005;64(2).

183. Castle SC, Song Z, Gohl DM, Gutknecht JLM, Rosen CJ, Sadowsky MJ, et al. DNA template dilution impacts amplicon sequencing-based estimates of soil fungal diversity. Phytobiomes J. 2018;2(2).

184. Allan AC, Hellens RP, Laing WA. MYB transcription factors that colour our fruit. Vol. 13, Trends in Plant Science. 2008.

185. Tripathi AM, Niranjan A, Roy S. Global gene expression and pigment analysis of two contrasting flower color cultivars of Canna. Plant Physiology and Biochemistry. 2018;127.

186. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 2015;33(3).

187. Fu Y, Jiang S, Zou M, Xiao J, Yang L, Luo C, et al. High-quality reference genome sequences of two Cannaceae species provide insights into the evolution of Cannaceae. Front Plant Sci. 2022 Jul 28;13.

188. Rhoads A, Au KF. PacBio Sequencing and Its Applications. Vol. 13, Genomics, Proteomics and Bioinformatics. 2015.

189. Yasir M, Turner AK, Lott M, Rudder S, Baker D, Bastkowski S, et al. Long-read sequencing for identification of insertion sites in large transposon mutant libraries. Sci Rep. 2022;12(1).

190. Pray L, Zhaurova K. Barbara McClintock and the Discovery of Jumping Genes (Transposons). Nature Education. 2008;1(1).

191. Kasem S, Rice N, Henry RJ. DNA extraction from plant tissue. In: Plant Genotyping II: SNP Technology. 2008.

192. Chauhan RP, Wijayasekara D, Webb MA, Verchot J. A reliable and rapid multiplex RT-PCR assay for detection of two potyviruses and a pararetrovirus infecting canna plants. Plant Dis. 2015 Dec 1;99(12):1695–703.

193. Jones A, Schwessinger B. Sorbitol washing complex homogenate for improved DNA extractions. ProtocolsIo. 2020;16(7).

194. Castillejo C, Waurich V, Wagner H, Ramos R, Oiza N, Muñoz P, et al. Allelic variation of MYB10 Is the major force controlling natural variation in skin and flesh color in strawberry (Fragaria spp.) fruit. Plant Cell. 2020;32(12).

195. Pucker B, Pandey A, Weisshaar B, Stracke R. The R2R3-MYB gene family in banana (Musa acuminata): Genome-wide identification, classification and expression patterns. PLoS One. 2020;15(10 October).

196. Veerappan K, Natarajan S, Chung H, Park J. Molecular insights of fruit quality traits in peaches, prunus persica. Vol. 10, Plants. 2021.

197. Tuan PA, Bai S, Yaegaki H, Tamura T, Hihara S, Moriguchi T, et al. The crucial role of PpMYB10.1 in anthocyanin accumulation in peach and relationships between its allelic type and skin color phenotype. BMC Plant Biol. 2015;15(1).

198. Guo J, Cao K, Deng C, Li Y, Zhu G, Fang W, et al. An integrated peach genome structural variation map uncovers genes associated with fruit traits. Genome Biol. 2020;21(1).

199. Yang J, Chen Y, Xiao Z, Shen H, Li Y, Wang Y. Multilevel regulation of anthocyanin-promoting R2R3-MYB transcription factors in plants. Vol. 13, Frontiers in Plant Science. 2022.

200. Xu ZS, Yang QQ, Feng K, Xiong AS. Changing carrot color: Insertions in DcMYB7 alter the regulation of anthocyanin biosynthesis and modification. Plant Physiol. 2019;181(1).

201. Shimura H, Pantaleo V, Ishihara T, Myojo N, Inaba J ichi, Sueda K, et al. A Viral Satellite RNA Induces Yellow Symptoms on Tobacco By Targeting a Gene Involved in Chlorophyll Biosynthesis Using the RNA Silencing Machinery. PLoS Pathog. 2011;7(5).

202. Aharoni A, De Vos CHR, Wein M, Sun Z, Greco R, Kroon A, et al. The strawberry FaMYB1 transcription factor suppresses anthocyanin and flavonol accumulation in transgenic tobacco. Plant Journal. 2001;28(3).

203. Xu P, Billmeier M, Mohorianu I, Green D, Fraser WD, Dalmay T. An improved protocol for small RNA library construction using High Definition adapters. Methods in Next Generation Sequencing. 2015;2(1).

204. Adamopoulos PG, Tsiakanikas P, Stolidi I, Scorilas A. A versatile 5' RACE-Seq methodology for the accurate identification of the 5' termini of mRNAs. BMC Genomics. 2022;23(1).

205. Moustafa K, Cross JM. Genetic approaches to study plant responses to environmental stresses: An overview. Vol. 5, Biology. 2016.

206. Hirsch J, Lefort V, Vankersschaver M, Bouaiem A, Lucas A, Thermes C, et al. Characterization of 43 non-protein-coding mRNA genes in arabidopsis, including the MIR162a-derived transcripts. Plant Physiol. 2006;140(4).

207. Li YF, Zheng Y, Jagadeeswaran G, Sunkar R. Characterization of small RNAs and their target genes in wheat seedlings using sequencing-based approaches. Plant Science. 2013;203–204.

208. Kurtoglu KY, Kantar M, Budak H. New wheat microRNA using whole-genome sequence. Funct Integr Genomics. 2014;14(2).

209. Lin F, Chen SP, Lin KH, Chen C, Yao F, Zhong L, et al. Integrated small RNA profiling and degradome analysis of Anthurium andraeanum cultivars with different-colored spathes. J Plant Res. 2022 Jul 1;135(4):609–26.

210. Xu J, Xian Q, Zhang N, Wang K, Zhou X, Li Y, et al. Identification of mirna-target gene pairs responsive to fusarium wilt of cucumber via an integrated analysis of mirna and transcriptome profiles. Biomolecules. 2021;11(11).

211. Lin B, Ma H, Zhang K, Cui J. Regulatory mechanisms and metabolic changes of miRNA during leaf color change in the bud mutation branches of Acer pictum subsp. mono. Front Plant Sci. 2023 Jan 12;13.

212. Llave C, Xie Z, Kasschau KD, Carrington JC. Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. Science (1979). 2002;297(5589).

213. Ronemus M, Vaughn MW, Martienssen RA. MicroRNA-targeted and small interfering RNA-mediated mRNA degradation is regulated by argonaute, dicer, and RNA-dependent RNA polymerase in Arabidopsis. Plant Cell. 2006;18(7).

214. Yadav A, Kumar S, Verma R, Lata C, Sanyal I, Rai SP. microRNA 166: an evolutionarily conserved stress biomarker in land plants targeting HD-ZIP family. Vol. 27, Physiology and Molecular Biology of Plants. 2021.

215. Zhu H, Hu F, Wang R, Zhou X, Sze SH, Liou LW, et al. Arabidopsis argonaute10 specifically sequesters miR166/165 to regulate shoot apical meristem development. Cell. 2011;145(2).

216. Du Q, Avci U, Li S, Gallego-Giraldo L, Pattathil S, Qi L, et al. Activation of miR165b represses AtHB15 expression and induces pith secondary wall development in Arabidopsis. Plant Journal. 2015;83(3).

217. Jung JH, Park CM. MIR166/165 genes exhibit dynamic expression patterns in regulating shoot apical meristem and floral development in Arabidopsis. Planta. 2007;225(6).

218. Kruszka K, Pieczynski M, Windels D, Bielewicz D, Jarmolowski A, Szweykowska-Kulinska Z, et al. Role of microRNAs and other sRNAs of plants in their changing environments. J Plant Physiol. 2012;169(16).

219. Li X, Xie X, Li J, Cui Y, Hou Y, Zhai L, et al. Conservation and diversification of the miR166 family in soybean and potential roles of newly identified miR166s. BMC Plant Biol. 2017;17(1).

220. Schuetz M, Smith R, Ellis B. Xylem tissue specification, patterning, and differentiation mechanisms. Vol. 64, Journal of Experimental Botany. 2013.

221. Luan WJ, Shen A, Jin ZP, Song SS, Li ZL, Sha AH. Knockdown of OsHox33, a member of the class III homeodomain-leucine zipper gene family, accelerates leaf senescence in rice. Sci China Life Sci. 2013;56(12).

222. Barozai MYK, Baloch IA, Din M. Identification of MicroRNAs and their targets in Helianthus. Mol Biol Rep. 2012;39(3).

223. Ul Haq S, Khan A, Ali M, Khattak AM, Gai WX, Zhang HX, et al. Heat shock proteins: Dynamic biomolecules to counter plant biotic and abiotic stresses. Vol. 20, International Journal of Molecular Sciences. 2019.

224. Li Y, Wang LF, Bhutto SH, He XR, Yang XM, Zhou XH, et al. Blocking miR530 Improves Rice Resistance, Yield, and Maturity. Front Plant Sci. 2021;12.

225. Bi F, Meng X, Ma C, Yi G. Identification of miRNAs involved in fruit ripening in Cavendish bananas by deep sequencing. BMC Genomics. 2015;16(1).

226. Camoni L, Visconti S, Aducci P, Marra M. 14-3-3 proteins in plant hormone signaling: Doing several things at once. Vol. 9, Frontiers in Plant Science. 2018.

227. Espley R V., Brendolise C, Chagné D, Kutty-Amma S, Green S, Volz R, et al. Multiple repeats of a promoter segment causes transcription factor autoregulation in red apples. Plant Cell. 2009;21(1).

228. Peumans WJ, Proost P, Swennen RL, Van Damme EJM. The abundant class III chitinase homolog in young developing banana fruits behaves as a transient vegetative storage protein and most probably serves as an important supply of amino acids for the synthesis of ripening-associated proteins. Plant Physiol. 2002;130(2).

229. Li MY, Xu BY, Liu JH, Yang XL, Zhang J Bin, Jia CH, et al. Identification and expression analysis of four 14-3-3 genes during fruit ripening in banana (Musa acuminata L. AAA group, cv. Brazilian). Plant Cell Rep. 2012;31(2).

230. Bradley D, Xu P, Mohorianu II, Whibley A, Field D, Tavares H, et al. Evolution of flower color pattern through selection on regulatory small RNAs. Science (1979). 2017;358(6365).