Transcriptome sequencing and digital expression analyses of flower morphogenesis genes in *Clivia miniata*

by

Palesa Naidoo

Submitted in fulfilment of the requirements for the degree *Magister Scientiae* in the Department of Genetics, Faculty of Natural and Agricultural Sciences at the University of the Free State.

April 2023

Supervisor: Dr. Mathabatha F. Maleka

Declaration

I, Palesa Naidoo, hereby declare that the work contained in this dissertation that I herewith submit for the master's degree in Genetics at the University of the Free State, is my independent work and that I have not previously submitted it for a qualification at another institution of higher education. Furthermore, I relinquish my rights as author in favour of the University of the Free State.

Palesa Naidoo

April 2023

Dedication

I dedicate this dissertation to my family, specifically to my late uncle Corporal TC Tutubala. Thank you for the encouragement and continuous support throughout my studies.

Acknowledgements

My most sincere gratitude to my supervisor, Dr Mathabatha Maleka for allowing me the opportunity to pursue this degree. Sir, there are no words to express my gratitude and appreciation to you. I genuinely want to take a moment to thank you for your consistent support, guidance, patience and motivation during the initiation, development and completion of this study. For believing in me, I thank you.

To my mother, Mpho Liphapang, I love you as a flower loves water and sunshine. You have helped me grow, prosper, and bloom into the flower that I am today. Thank you for always directing me towards the light, even in your darkest eras. Your love and support are immensely appreciated.

To Emmanuel Naidoo, thank you for being an amazing husband and father. I am truly grateful for the effort you put into bringing joy to my heart during the most difficult times of this study.

To Tsholofelo Modise, thank you for being a friend, the laughs we shared in the lab, always being a phone call away and encouraging me when I was demotivated.

To Stephanus Riekert (High Performance Computing unit, University of the Free State, Bloemfontein), thank you for your help with the Linux cluster. Your willingness to always help is highly appreciated.

I am also grateful for the financial support from the National Research Foundation (i.e., NRF Postgraduate Scholarship) and the University of the Free State (i.e., UFS tuition fee bursary).

Thank you to the Johannesburg Clivia Society president Glynn Middlewick for allowing me to collect flower samples.

In conclusion, I would also like to thank my Heavenly Father God for the strength he has given me to complete this study. By his grace…

Table of Contents

List of tables

List of figures

List of abbreviations

Abbreviation	Meaning
%ID	Percentage identity
AG	AGAMOUS
AGL	AGAMOUS-LIKE
AP	APETALA
BP	Biological process
Bp	Base pairs
BUSCO	Benchmarking Universal Single-Copy Orthologs
CAL	Cauliflower
CC	Cellular component
CLV	Clavata
CO	CONSTANS
DEG	Differentially expressed genes
FDR	False discovery rate
FLC	Flowering locus C
FLOR-ID	Flowering Interactive Database
FT	Flowering Locus T
GO	Gene ontology
IB	Intermediate buds
IUCN	International Union for Conservation of Nature
KEGG	Kyoto Encyclopaedia of Genes and Genomes
KNU	KNUCKLES
LEU	LEUNIG

LFY	LEAFY
LMI	LATE MERISEM IDENTITY
MB	Miniata buds
MF	Molecular function
NGS	Next Generation Sequencing
ORF	Open reading frames
PCA	Principal Component Analysis
PI	PISTILATA
PPI	Protein-protein interaction
SEP	SEPALLATA
SOC1	Suppressor of Overexpression of Constans
STM	SHOOT MERISTEMLESS
SUP	SUPERMAN
TB	Tiny buds
TF	Transcription factor
TFL	TERMINAL FLOWER
USA	United States of America
ULT	ULTRAPETALA
WUS	WUSCHEL

Abstract

Flower morphology is an important aspect in horticulture as it can be used to create new varieties, it plays an important role in pollination and it influences the visual appeal of a plant. The genus *Clivia* is important in horticulture due to the plant's ornamental value, adaptability, low maintenance requirements, easy propagation and showy flowers. The genus comprises plants that exhibit two distinct flower morphologies, namely, the pendulous (drooping) and non-pendulous (upright) forms. The aims of the study were to sequence the transcriptome of the two *Clivia* flower shapes; identify flower morphogenesis genes and identify differentially expressed morphogenesis genes between the pendulous and non-pendulous *Clivia* flowers. Illumina sequencing and analyses generated a 200 MB *Clivia* flower transcriptome. A total of 36 flower development-related genes from the Flower Interactive database (FLOR-ID) were identified, including *AGAMOUS, APETALA, SEPALLATA, PISTILATA, TERMINAL FLOWER1* and *WUSCHEL*. DE analysis resulted in 7,303 DEGs which were annotated using the SwissProt/Uniprot, KEGG, GO, PANTHER, Pfam, TMHMM, SignalP, EggNog and RNAmmer databases. The top three species with sequence homologies to the annotated *Clivia* transcriptome were *Arabidopsis thaliana*, *Zea mays* and *Oryza sativa*. Several transcription factors were found to be unique to each flower form, including a MADS-box gene likely to be responsible for the different morphologies. Our results contribute towards understanding the molecular mechanisms of flowering, flower development and morphogenesis in *Clivia*.

Keywords: *Clivia*, differential expression, flower development, MADS-box, morphogenesis, next-generation sequencing, transcriptome analysis.

General introduction

Clivia is a genus from the Amaryllidaceae family. The genus is distributed along the east coast of South Africa and Swaziland (Ran et al. 2001). The genus consists of herbaceous, perennial monocotyledonous angiosperms with green strap-like leaves and bright flowers of varying colours (Spies et al. 2011). The plant also produces berries that are somewhat spherical and contain ivory-white coloured seeds. The bush lily, as it is sometimes called, occurs naturally in habitats including riverbanks, swamps, dunes and forests (Swanevelder 2003). Specimens usually flourish in shaded habitats with well-drained soil. *Clivia* species are cultivated in most parts of the world including Australia, China, Japan and the United States of America (USA). Moreover, clivias produce alkaloids with medicinal properties (Hirasawa et al. 2022). Thus, it is a horticulturally and medically important genus. The genus comprises non-model plant species and therefore developmental studies and gene discovery is not as extensive as reported for model plants.

Flowers are an important part of nature for several reasons, besides horticulture. One reason, possibly the most important of all, is that flowers have a responsibility in plant reproduction. The morphology of a flower (which serves as the reproductive organ of a plant) is very important for complete pollination to occur. Flower morphology needs to allow the suitable pollinator to access it successfully and easily i.e., tubular/trumpet shaped flowers for bird pollinators or funnel-shaped flowers for butterfly pollinators. The mismatch of flower morphology and the type of pollinator will hinder pollination and thus halt reproduction. To ensure reproductive success, angiosperms have diversified immensely with regards to floral colour, scent, size, morphology and flowering time. Research on the evolutionary and developmental genetics of floral display and diversification has created a foundation for insights into the genetic basis of flower morphogenesis and formation (Theissen and Melzer 2007), and thus, the molecular regulatory interactions that could possibly control morphogenesis of *Clivia* flowers.

Several regulatory transcription factor (TF) genes have been identified through forward genetics and linked to the developmental control of floral shape. The molecular basis of floral organ

development has been studied in model plants such as *Arabidopsis thaliana* and *Antirrhinum majus* (Bowman et al. 1989; Schwarz-Sommer et al. 1990), leading to the development of the ABCDE model (Almeida et al. 1997). Flower architecture is controlled by interactions among regulatory genes involved in the ABCDE model of flower development. Interestingly, the functions of the ABCDE model genes appear to be conserved in angiosperms (Wozniak and Sicard 2018), providing us with a foundation to study the process in any plant species. The discovery of the MADS-box gene family as well as the ABCDE model of flower development created a foundation that genes involved in the initiation, morphogenesis and flowering are conserved among angiosperms (Shore and Sharrocks 1995). This implied that phenotypic variation between closely-related species may be due to differentially expressed regulatory genes, which are either upregulated or downregulated (Chen 2007). Fortunately, advances in next-generation sequencing technolgies(NGS), as well as comparative transcriptome analysis methods have made it possible to identify regulatory genes that may be responsible for morphological differences between particular flowers (Eveland et al. 2010; Schneeberger 2014; Bian et al. 2019).

The overall aims of the study were to:

- Use NGS to sequence and *de novo* assemble transcriptome
- Identify flower morphogenesis genes.
- Identify differentially expressed morphogenesis genes occurring between pendulous and non-pendulous *Clivia* flowers.

The objectives of the study included:

- Plant sampling and RNA extraction
- Next-generation sequencing of the RNA-derived from pendulous and non-pendulous flowers of *Clivia.*

- Assemble and characterize the pendulous and non-pendulous and combined transcriptome.
- Identify flower development and morphogenesis genes as well as possible gene interactions that control flower morphogenesis in the two flower forms.
- Identify differentially expressed genes (DEGs) that may be responsible for pendulous and non-pendulous flower forms.

This dissertation is presented in the following manner: **Chapter One** will comprise an extended review of the literature. The first section will describe the morphology, climate and distribution of *Clivia* species. Then, the second section will cover the genetics of flower development, and the final section will explain NGS as a technique and its applications in gene discovery and expression analysis

Chapter Two will describe the materials and methods used to address the objectives of this study. These will include defining the selected plant material, the RNA extraction protocol, various bioinformatic techniques and software. **Chapter Three** will describe the findings obtained after performing different experiments outlined in Chapter Two; **Chapter Four** will discuss the meaning of the findings. **Chapter Five** will comprise general concluding remarks that seek to address the objectives relating to the aims of the study. The final chapter, **Chapter Six**, will contain all references cited in the document.

Chapter One - Literature review

The genetics of flower development and its implications in the non-model horticultural crop *Clivia miniata*.

1.1. Introduction

Flowers are important plant parts that ensure the successful reproduction of angiosperms (flowering plants), providing food for birds, humans and insects as well as serving as medical remedies. Flowers serve as the reproductive organs of all angiosperms and they show variations in their colour, size, scent, structure/shape as well as flowering time (Sauquet et al. 2017). Angiosperms have evolved and diversified in many ways. For example, plants show off their flowers to ensure reproductive success (Ma et al. 2017). Flowers have male and female reproductive parts (carpels and stamens, respectively) that partake in plant reproduction (Stegemann et al. 2019). A standard flower typically has four main organs, namely, sepals, petals, carpels and stamens (Wessinger and Hileman 2016). Sepals are the outermost organs that protect the flower bud before anthesis and maturation. Petals are usually perceived to be the beautiful and brightly-coloured organs that attract pollinators to the plant (Bercu et al. 2017). The arrangement of floral organs can influence the overall morphology of the flower.

Clivia is a genus of monocotyledonous angiosperms within the family Amaryllidaceae (Ran et al. 2001). Commonly referred to as ‘clivia’, the plants are herbaceous with green, strap-like leaves and brightly-coloured flowers (Spies and Spies 2018). Flowers of *Clivia* species typically have three sepals and three petals that are collectively referred to as ‘tepals’. This is mainly due to the organs’ similarity in shape and colour (Yue et al. 2018). Members of the genus *Clivia* occur naturally in southern Africa, albeit the plant has been cultivated in most parts of the world including Australia, Belgium, China, Japan, New Zealand and USA (Liu et al. 2020). *Clivia* is of high interest to enthusiastic breeders and collectors (Biggs 2021). Being a non-model plant means that there are many under-explored research aspects regarding this plant, such as the molecular basis and regulation of flower development. In-depth knowledge of the genetic regulatory network of flower development would be advantageous to plant breeders and horticulturists as they would be able to better plan and produce desirable crops.

Studies on the molecular basis of flower development in the model plant *Arabidopsis thaliana* revealed many genes relating to organogenesis and identity (Huang et al. 2020). The mutations and phenotypes in *A. thaliana* led to the development of the ABCDE model, where each letter refers to a gene(s) from the MADS-box-containing gene family (Almeida et al. 1997). NGS technologies are essential tools that have made it possible to study many gene families in different flowering plants, including non-model species (Ward 2012). Such technologies include the popular Illumina (https://www.illumina.com; Shendure and Ji 2008) and Ion Torrent (https://www.thermofisher.com /za/en/home/brands/ion-torrent.html; Rothberg et al. 2011) platforms that permit fast, accurate, efficient and cost-effective sequencing of nucleic acids (Sasaki et al. 2017). Increasing evidence from flower development studies suggested that the ABCDE model is universally conserved in angiosperms, and RNA sequencing experiments (an NGS-based research strategy) have been instrumental in the discovery and expression profiling of many flower development genes (Li et al. 2017; Wozniak and Sicard 2018).

The objectives of this literature review were to 1) describe the climate, distribution, relationships and morphological differences of species in the genus *Clivia*; 2) report on the species' importance for horticulture and for medicinal purposes; 3) discuss the genetics of flower development, including the endogenous and exogenous pathways of flowering, the ABCDE model of flower development as well as the architecture of gene regulatory networks that control flower development in angiosperms; and 4) describe the background of NGS techniques and applications of such techniques in gene discovery and expression analysis studies.

1.2. Clivia – defining the plant.

1.2.1. The morphology of a typical *Clivia* flower.

In *Clivia*, petals are usually wider than sepals, but because they are so similar, they are jointly referred to as 'tepals'. Each flower consists of six tepals that are fused at the base in a tube-like manner. Individual flowers typically possess six stamens that have long filaments and end with anthers (Fig. 1.1). Stigmas (the sticky knob at the top of the pistil) are positioned at the end of the style (slender stalk within the plant) and can usually be seen protruding past the length of tepals. Although the basic structure remains the same, flowers from certain species are upright and open, whereas others are narrow and hang downwards (Duncan 2004, Koopowitz 2002; Swanevelder and Fisher 2009). These are referred to as non-pendulous (open, upright) and pendulous (narrow, drooping) flowers, respectively.

1.2.2. Flower morphology in the genus Clivia

Clivia are evergreen, rhizomatous plants that have stiff, strap-like distichous leaves. The inflorescences have several flower stalks/pedicels (pseudo-umbel) that spread from a common point and grow on a solid, compressed peduncle, that produces predominantly red sub-globose berries. The berries have yellowish or ivory-white seeds buried in a yellow pulp (Koopowitz 2002).

The most common and commercialised species in the genus *Clivia* is *C. miniata* (Fig. 1.2F). *Clivia miniata* typically produces 10 – 40, upright, trumpet-shaped, open flowers (Kiepiel and Johnson 2021). Although similar, the pendulous flower forms differ in minor ways including the number of flowers on a plant, tepal length, recurving (curving inwards) tepal tips and green tipped flowers. The most remarkable difference in flower shape remains that of the non-pendulous flower form *Clivia robusta* (Fig. 1.2A), for example, is regarded as the tallest of all the *Clivia* species and usually has 30 – 55 mm long falcate (curved like a sickle) flowers with a flaring apex. A typical *Clivia nobilis* (Fig. 1.2B) specimen has about 28 – 40 flowers per inflorescence with slightly curved, drooping green pedicels and a compact umbel (Dixon 2011). The flowers are tubular and

linear with straight inner petals that range from 20 to 40 mm in length. *Clivia mirabilis* (Fig. 1.2C) typically produces 20 – 48 drooping falcate flowers (about 35 to 50 mm in length) with increased flaring towards the apex of the flower. Their drooping pedicle is red or orange during flowering and green during fruiting (Rourke 2002). *Clivia caulescens* (Fig. 1.2D) has a tight umbel that is flattened on one side. The pedicle is erect and drooping towards the flower. Flowers are typically 30 – 35 mm long, tubular and curved with inner petals being recurved (deflecting downwards) (Abel and Abel 2003). *Clivia gardenii* (Fig. 1.2E) has flowers that are tubular and falcate toward the tips with recurved inner petals hanging on stiff red or orange pedicles. This species typically has 10 – 20 flowers that are usually 40 – 52 mm long. The leaves are recurved and have loose inflorescences flattened on one side and rounded on the other (Swanevelder 2003).

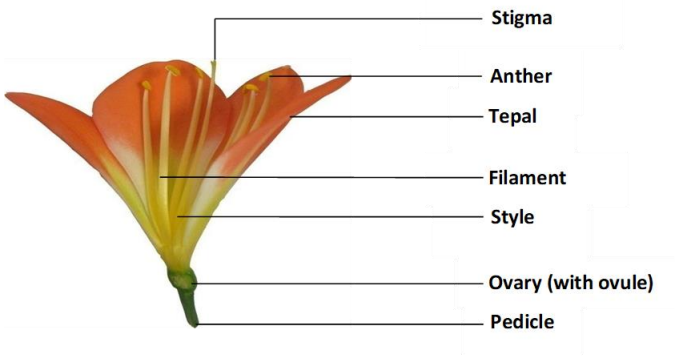

Figure 1.1: A diagram showing different parts of the flower in *Clivia miniata*. The anthers and filaments are collectively termed the stamen and the carpel consists of the ovary, stigma and style. Diagram obtained from Snyman (2010).

1.2.3. The genus *Clivia* and its climatic preferences.

Clivia species have been cultivated as ornament plants worldwide and are highly important in horticulture. The genus *Clivia* was first described in 1828 by Lindley and Hooker independently. However, in many publications, Hooker is omitted. The genus Clivia comprises sex recognised species (Fig. 1.2). The first species to be described was *Clivia nobilis* (1828). Then, *Clivia gardenii* was described in 1856 by Hooker. In 1864, Regel described *Clivia miniata*. It was followed by *Clivia caulescens*, which was the first species to be scientifically described in South Africa by Dyer in 1943 (Duncan 2004; Koopowitz 2002). *Clivia mirabilis* was described by Rourke 2002. The last species to date, *Clivia robusta* was identified by Murray et al. (2004). Initially, *C. robusta* was grouped together with *C. gardenii* and the morphological differences between the two species were attributed to natural variation (Murray et al. 2004). It was only in 2004 that *C. robusta* was recognised as a separate species. The species *C. robusta* is still commonly referred to as the 'robust form' of *C. gardenii*.

Clivias are evergreen, resilient plants that can grow in any type of well-drained soil and shaded habitat (Spies et al. 2011). They thrive in a wide range of terrains along the east coast of southern Africa, including forests, swamps, dunes and riverbanks (Duncan 1999). Specimens prefer cool areas with summer rainfall. They can sometimes exist as epiphytes – plants that grow on other plants, commonly referred to as "air plants" because they are not anchored in the soil (Honiball and Robbertse 2001). *Clivia* specimens desire growth positions under trees, shady garden corners or in large pots on porches (Duncan and Du Plessis 1989). Overall, climatic preferences of the different *Clivia* species affect the natural distribution of different *Clivia* species.

So, *C. mirabilis* prefers the drier climate prevalent in the Northern Cape Province of South Africa (Murray et al. 2011). In fact, the anatomy of *C. mirabilis* and *C. nobilis* show features that have developed to survive in arid environments (Matimati 2017). Such features include the semi-succulent, stiff and leathery leaves that contain cutan, a biopolymer usually found in plants that are

adapted to survive during prolonged periods of drought (Cookey et al. 2018). The roots of these two species are also adapted to aerated leaf litter (not soil), and have strong absorbent and water storing properties (Rourke 2002). All *Clivia* specimens may survive and grow in any type of soil, but then they will flower poorly. Rather, these plants are distinctly adapted to wet and dry periods on rocks or in leaf litter. The roots require a lot of oxygen around them and do not prefer being submerged in water (Gagiano 2006). For this reason, thick soils like clay soils are not ideal for plant growth.

Clivia plants do not prefer direct sunlight or scorching hot temperatures, even during daytime. They prefer temperatures between 5℃ and 32℃ (Honiball and Robbertse 2001). The natural habitat of *Clivia* specimens is mostly forests. Forests are crowded with trees that provide shade to make this a suitable environment (Shawky 2016). Fallen leaves act as mulch and therefore protect the plants from frost in extremely cold conditions.

The flowering times of different *Clivia* species are mirrored by habitat conditions such as rainfall peaks where they prefer bimodal rain. Bimodal rain (a rainfall period during two wet seasons separated by a dry season) in thicket climates peak in early autumn and late spring, but copious rain may fall at any time of the year (Dixon 2011). *Clivia mirabilis*, *C. nobilis* and *C. caulescens* usually flower around October to November (Voster 1994; Duncan 2004), whereas *C. gardenii* and *C. robusta* flower from April to June (Murray et al. 2004; Papenfuss et al. 2014). *Clivia miniata*, in contrast, usually flowers around August to October, depending on where it is located (de Smedt et al. 1996; Kiepiel and Johnson 2021). The areas of highest plant endemism in South Africa, which happen to be the natural range of *C. robusta*, *C. gardenii* and *C. caulescens*, are also the areas with the highest rainfall (Dixon 2011).

1.2.4. The natural distribution of Clivia species.

Clivia, which is also referred to as a ‘bush lily’ is endemic to southern Africa where it occurs from the east coast of South Africa up to eSwatini, formerly Swaziland (Winter 2000) (Fig. 1.4). The most popular species of *Clivia*, *C. miniata*, occurs in eSwatini as well as the Eastern Cape,

KwaZulu-Natal (KZN) and Mpumalanga Provinces of South Africa (Swanevelder 2003). *Clivia caulescens*, on the other hand, occurs in Swaziland, Mpumalanga and Limpopo (Abel 2003). *Clivia robusta* occurs in the Eastern Cape and KwaZulu-Natal, whereas *C. gardenii* is only found in Kwa-Zulu Natal (Murray 2004). *C. mirabilis* only grows in the Northern Cape (Rourke 2002) and only in the Eastern Cape will you find *C. nobilis* (Voster 1994).

1.2.5. Relationships among *Clivia* species

In the field of plant science, 'form' can refer to any plant or group of plants showing differences in a specific character or trait. For example, yellow-flowering *Clivia* specimens within a population of commonly orange-flowering plants represents a different form of clivia plants (Conrad et al. 2003). Considering flower shape within the genus *Clivia*, the most remarkable difference is the pendulous versus non-pendulous flower form (Swanevelder 2003; Conrad 2008; Dixon 2011); while there are no major discernible differences between the pendulous species (Fig. 1.2). Therefore, it becomes interesting as to why different species exist among the pendulous flower forms.

Phylogenetic studies have revealed that *C. miniata* is more closely related to *C. gardenii* and *C. robusta* (Fig. 1.3). Ran et al. (2001) used sequences of the nuclear ribosomal 5S non-transcribed spacer region and the 45S rDNA internal transcribed spacer (ITS) region to karyotype and determine phylogenetic relationships among *Clivia* species. Analysis of the ITS region revealed indel variations between the species, whereby *C. nobilis* had two nucleotides less than *C. miniata*, *C. gardenii* and *C. robusta* and one nucleotide less than *C. caulescens.* According to Spies et al. (2011) in a *Clivia* phylogenetic analysis study, *C. miniata* is, however, indistinguishable from its pendulous counterparts on a molecular level. As such, this makes the current study more valuable and interesting.

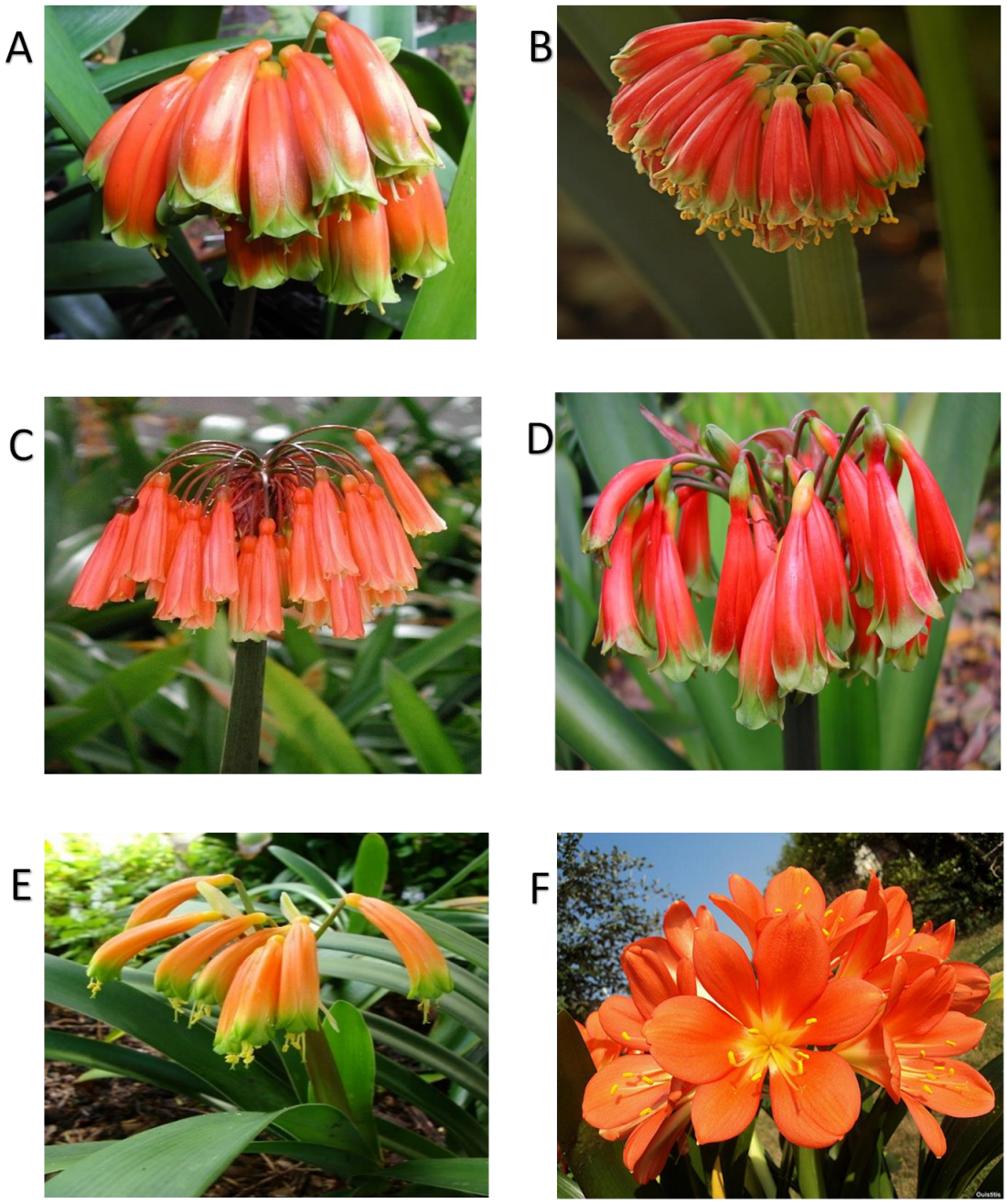

Figure 1.2: Flower morphology in species of *Clivia*. A- *Clivia robusta*; B- *Clivia nobilis*; C- *Clivia mirabilis*; D- *Clivia caulescens*; E- *Clivia gardenii* and F- *Clivia miniata*. Images obtained from https://www.cliviape.co.za/listings/cliva-species-information and https://www.cliviape.co.za/galleries/all.

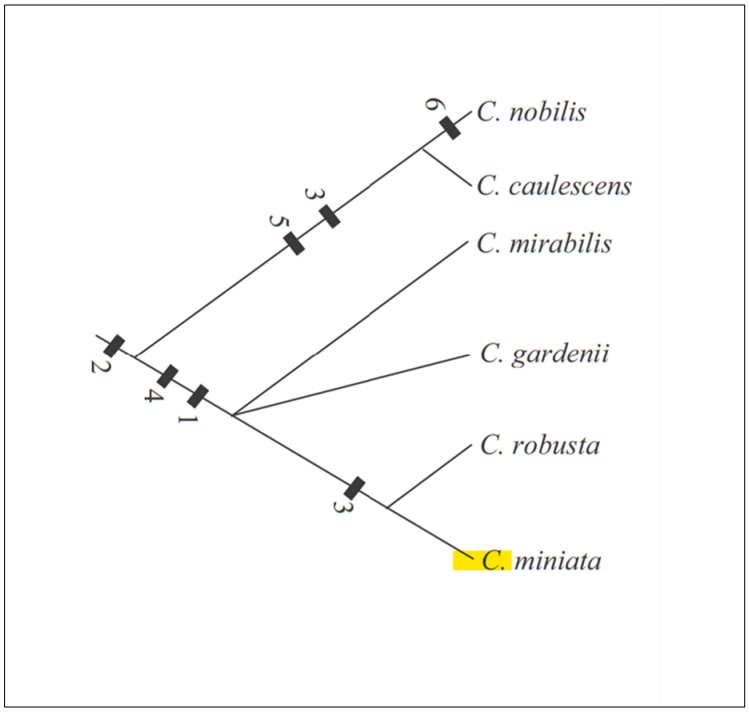

Figure 1.3: Phylogenetic analysis of the genus *Clivia*. The maximum likelihood tree is from Ran et al. (2001).

1.2.6. The relationship between the flower morphology and pollination in *Clivia*

One of the few studies done on flower morphology in the genus *Clivia* was by Conrad et al. (2003). Essentially, the study showed that the open, upright flowers of *C. miniata* are a derived morphological characteristic. Later, Kiepiel and Johnson (2014) suggested that floral modifications during plant diversification can explain pollinator shifts. They found that flower morphology was responsible for the type of pollinator the plant attracts: in *Clivia*, all but one species (*C. miniata*) is bird pollinated, whereas the remaining species are butterfly pollinated (Fig. 1.4). The difference in pollinators is attributed to the evolution of upright trumpet-shaped flowers in *C. miniata* (Kiepel and Johnson 2014). The morphology of pendulous flowers in *Clivia* species allows sunbirds to reach the deeper flower parts (and nectar) while perched on the umbel stem. In the non-pendulous *C. miniata*, however, it will be challenging for a bird to do this as it cannot be perched on the inflorescence stem – something that butterflies can do with ease.

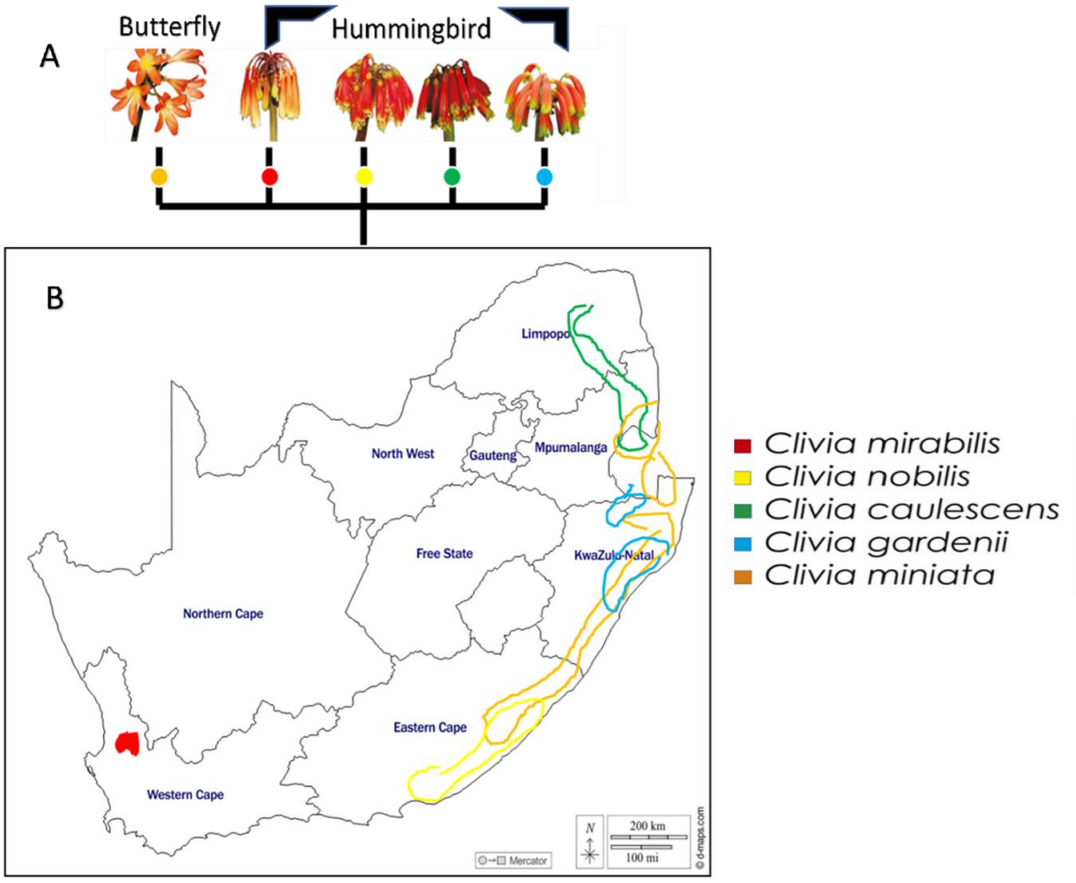

Figure 1.4: The relationship among pollinators, flower shape and *Clivia* species distribution in South Africa. Part A shows different *Clivia* species (named in part B) and their respective pollinator. Part B depicts a map of South Africa showing the distribution and localities of each *Clivia* species corresponding to the colour-coded labelled key. Figures obtained from Kiepiel and Johnson (2014).

Other physiological and morphological traits unique to *C. miniata* include the production of small nectar volumes and scent (Manning 2005). *Clivia miniata* emits volatile compounds that are dominated by benzaldehyde, benzyl alcohol and benzyl benzoate. These compounds produce scent that is 50 times greater than that of other *Clivia* species. (Manning 2005) and discernible by a human nose. In ecological terms, floral scent is an important feature because pollinators (e.g., butterflies) use scent as a guide towards finding the flower. Omura et al. (1999) discussed the relationship between scent and colour combinations in stimulating the butterfly *Pieris rapae* to visit

flowers of *Brassica rapa*. Anderson (2003), on the other hand, explained the importance of scent and flower colour to butterfly pollinators. Overall, scent and colour may attract many pollinators to flowers, but then, flower morphology will filter out the most suitable pollinator.

1.2.7. Horticultural importance of *Clivia* species

Flower morphology, colour and patterning are some of the most sought after characteristics in the South African horticulture market (Middleton 2015). *Clivia miniata* has received more attention for its flower colour variation than the other *Clivia* species. However, interspecific hybrids have become popular among breeders in countries such as China, with the pendulous *Clivia* species being particularly sought after (Swanevelder 2003). Increasing interest in interspecific plants has encouraged breeders to hybridize different species to produce new varieties in an industry where new, different and unique equates to highly valuable and most wanted (Williamson 2008). The sale of ornamental plants provides an avenue for innovation where growers and retailers are constantly introducing new varieties of plants, keeping the industry dynamic and exciting.

In South Africa, a Clivia Club was founded in 1992 where public awareness on the plants was heightened through shows and newsletters (Duncan 1999). This allowed collectors of rare cultivars to show off their plant collection and to some degree set a sale price. Factors that contributed to the commercial success of *Clivia* plants was, apart from the showy flowers, the simplicity of its growth and hybridization (Swanevelder 2003). So far, commercially sold *Clivia* hybrids were largely produced in South Africa (Rourke 2002), and they have reached prices in the region of R100, 000 for a single plant (Naicker 2019). The sale of ornamental plants plays a significant role in the growth and sustainability of the constantly evolving horticulture industry in South Africa.

1.2.8. Medicinal properties and importance of *Clivia*

Clivia species, specifically *C. miniata*, is not only important to South Africa's floriculture industry, but also the informal traditional medicine fraternity. It has been shown that plant extracts from *C. miniata* can be used as medicine to treat human illnesses including fevers, coughs, malaria, pneumonia and influenza (Growers 2017). Some local and rural communities in southern Africa use the plant to induce or facilitate child labour whereas others even use it as an antidote for snakebites (Musara et al. 2021). Hirasawa et al. (2022) identified many alkaloids that may be responsible for the plant's antiviral, antimicrobial and antidiabetic properties. These properties make the plant highly sought after and the lack of legislation and protection has made *C. miniata* a threatened species in the wild due to uncontrolled harvesting by community members, traditional healers, and plant collectors (Williams et al. 2013). The loss of habitat caused by agriculture and urbanization is also a contributing factor that could drive the wild population into extinction (Aubrey 2001). Altogether, *Clivia* is considered "vulnerable" on the International Union for Conservation of Nature (IUCN) red list of threatened species due to its medicinal properties, overharvesting and habitat disruption.

1.3. The genetics of flower development

Flowering is arguably one of the most important phases during plant development. It is essential that a small flower bud develops and matures into a flower. Flowering is particularly important for animal and insect-pollinated plant species because the more visually appealing the flower is, the higher the chance of it attracting pollinators and being pollinated (Reid 2005). Reproduction is one of the key measures of plant fitness and flowering is essential for reproduction to take place (Vimont et al. 2019). Reproduction varies a lot in response to environmental cues. However, the genetic determinism of reproduction, and more specifically flowering, is not well understood in most non-model plants, such as *Clivia*. For a plant species to survive and thrive, flowering needs to be regulated timely (Funnel 2016). Weather conditions, pollinators, flowering

time, as well as endogenous (autonomous and gibberellin pathways) and exogenous cues (vernalization and photoperiodic pathways) need to be concise such that flowering can result in successful reproduction (Bernier and Perilleux 2005). In the model plant *Arabidopsis thaliana*, four pathways relating to flowering have been described thus far (Cheng et al. 2017). These include the autonomous, gibberellin, photoperiod and vernalization pathways, and they will be described below.

1.3.1. Endogenous and exogenous pathways leading to flowering and floral development.

One of the main environmental cues toward flowering is prolonged periods of cold weather, such as during winter or in artificially induced cold conditions by refrigeration. This process of cold exposure is known as vernalization, and it prepares plants for flowering (Sheldon et al. 1999; Austin et al. 2005). However, additional factors may be required for flowering to initiate (Fig 1.5). Typically, the expression of a gene called *FLOWERING LOCUS C (FLC)* will prevent flowering by repressing the expression of flowering promotion genes, including the *SUPPRESSOR of OVEREXPRESSION of CONSTANS (SOC1)* and *FLOWERING LOCUS T (FT)* (Michaels and Amasino 1999). In the vernalization pathway, cold conditions reduce the expression levels of *FLC* and, in turn, increase the expression levels of *SOC1* and *FT*. This then initiates flowering. Further, *FLC* can also alter flower morphology via the repression of *SEPALLATA (SEP),* a floral morphology regulator (Deng et al. 2011). The importance of the vernalization pathway lies in its ability to regulate the timing of flowering, ensuring that plants do not flower too early or too late as this can impact plant growth and reproduction.

The photoperiodic pathway, which is regulated by photoperiodic signals, can induce flowering by inducing floral promotion genes *SOC1* and *FT* (Helliwell et al. 2006). Floral meristem identity genes *LEAFY (LFY)* and *APETALA1* (*AP1*) are activated by the floral integrators (*SOC1* and *FT*) along with gibberellins and a product of *FLOWERING LOCUS D (FLD)* to effect a

developmental change from the vegetative to reproductive phases. *FLC* expression is maintained by a complex involving *FRIGIDA* (*FRI*), Polymerase-Associated Factor (PAF) and SWI2/SNF2-related 1 chromatin Remodelling Complex (SWR1) during the vegetative phase (Alexandre and Hennig 2008). Plants with recessive *FLC* alleles will tend to flower too early due to *SOC1* and *FT* not being repressed (Searle et al. 2006). Once flowering is initiated, several MADS-box transcription factors (TFs) known as organ identity genes (including *AP1*, *PISTILLATA* (*PI*), *AGAMOUS* (*AG*) and the paralogs *SEP1-4*), will control the differentiation process (Villar et al. 2020). These TFs will interact in an assembly known as the 'ABCDE model of flower development', which ensures that correct genes are expressed in a temporally coordinated manner, and ultimately, flower organs develop in corresponding whorls during development (Fernandez et al. 2014).

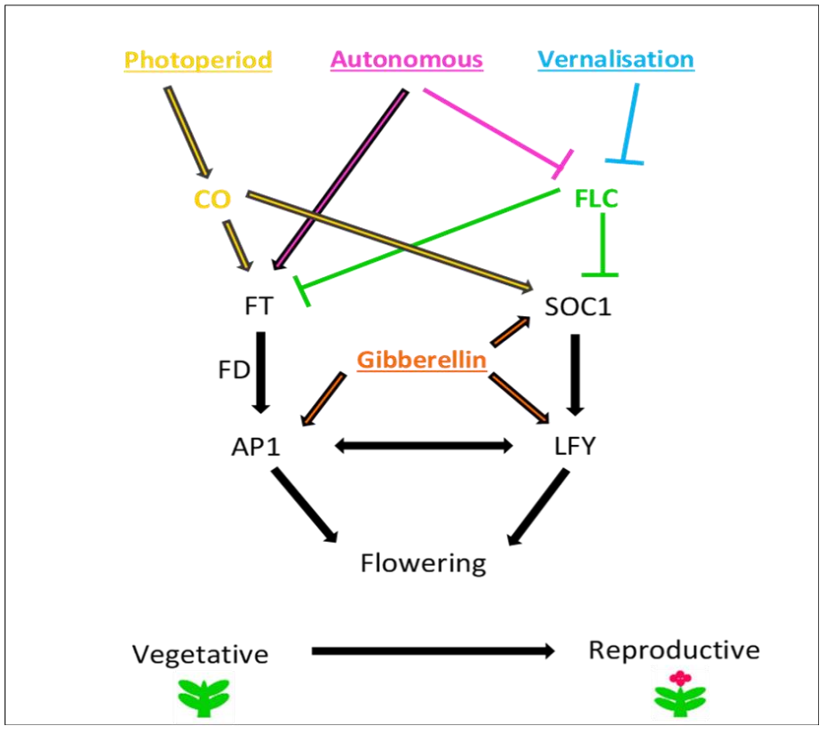

Figure 1.5: The four main endogenous and exogenous pathways of flowering in *Arabidopsis thaliana* (adapted from Williamson 2008). Photoperiod pathway (yellow), autonomous pathway (pink), vernalization pathway (blue) and gibberellin pathway (orange).

Studies have shown that flowering genes are conserved in flowering plants (Ng and Yanofsky 2001) and this provides an interesting starting point for research in non-model species. Tzeng et al. (2003) determined that *SEPALLATA-LIKE* genes are highly expressed in petals, carpels, stamens and sepals of mature *Arabidopsis thaliana* flowers in comparison to buds. As expected, these genes were not expressed at all in leaves. Zhao et al. (2021) found high expression levels of *AP1* during the final stage of dormancy in flower buds of *Populus tomentosa*. In another study, Vimont et al. (2019) defined and studied the five stages of flower bud development in sweet cherry plants and found the expression of *AG* and *AP3* to peak during eco-dormancy (stage 5 bud, the stage they described as the final bud stage). Therefore, this means that the eco-dormancy stage may be ideal for studying gene expression during flower development in angiosperms.

1.3.2. *Arabidopsis* as a model plant in flower development studies

The mystery of flower formation has been pondered for centuries (Mandel and Yanofsky 1995). Remarkable progress over the years using model plants such as *Arabidopsis thaliana* and *Antirrhinum majus* has led to a simple model that describes the genetic control of flower formation (Almeida et al. 1997). As a model plant, *A. thaliana* plays an important role in plant genetics and research (Meyerowitz 2001) as it has many characteristics of a good model organism. First, the plant matures very rapidly and has a very short lifecycle (six weeks) compared to other angiosperms (Dinneny et al. 2005). Second, the plant is phenotypically small and occupies little space (Koorneef and Somerville 2002). This means many individual plants can be grown in a small space such as a laboratory. Third, the genome sequence has been determined completely. In fact, *A. thaliana* was the first plant to have its whole genome sequenced (Kaul et al. 2000). Last, the plant is easily manipulated, i.e., genetic transformation of the plant is generally not a tedious task and gene function studies can be done easily (Alonso et al. 2003). Bowman (1997) was interested in these characteristics and therefore studied the influence of genes on a flower's architecture and used this to describe a model (ABC model) of flower development.

1.3.3. The ABC model of flower development.

A typical flower is comprised of four whorls/rings (Fig. 1.6). The first (outermost) ring is represented by the sepals; the second ring comprises the petals. The third and fourth rings consists of the stamens and carpels (innermost), respectively (Alvarez-Buylla et al. 2010). The letters A, B and C in the model refer to broad classes of genes involved in flower development. Class A genes include *AP1*, *AP2* and *LEUNIG* (LEU). Class B consists of *AP3* and *PI* and class C contains the gene *AG* (Wang et al. 2020). Class A genes are typically active in the first and second ring (Colombo et al. 1996), whereas class B genes are active in the second and third rings (Irish 2017) – meaning, their activity overlaps with that of class A genes in the second ring. *AG*, the class C gene, is usually active in rings three and four, and thereby, overlapping with class B genes in the third ring (Coen and Meyerowitz 1991). Importantly, the activities of class A and C genes do not overlap. In fact, these two classes seemingly antagonise each other (Kinoshita et al. 2020). Activities of the ABC class genes will direct the meristem cells to develop into carpels, stamens, petals or sepals (Zhao et al. 2020). Despite the diversity in flower morphology among different species, the ABC model is integral to the development and organization of the floral organs (Thomson and Wellmer 2019). *Clivia* species are likely to be no exception to other angiosperms and we can, therefore, use the ABC model as a guide to study flower morphogenesis in the genus.

1.3.4. Mutations in flower development genes

Mutations in one of the genes in the three classes may affect the morphology and structure of flowers (Shi et al. 2020). Interestingly, when class A genes are mutated, class C genes may take over their roles (McCarthy et al. 2015). Thus, stamens may develop in the place of petals and carpels may develop instead of sepals (Fig. 1.6). Mutations in the class C gene may replace carpels with sepals and stamens with petals (Duttke et al. 2012).

The variation in flower structure typically relates to differences in organ numbers and sizes between plants (Wang et al. 2012; Sornchai et al. 2020). In *A. thaliana*, a regulatory assembly that

is not part of the ABCDE model, involving *WUSCHEL (WUS)*, *CLAVATA (CLV)* and additional genes including *ULTRAPETALA (ULT)* controls flower size (Moreau et al. 2016). Mutations in *CLV* may result in larger floral organs developing, whereas mutations in WUSCHEL (*WUS)* produce flowers with smaller organs, and thus, smaller floral sizes (Moyroud and Glover 2017).

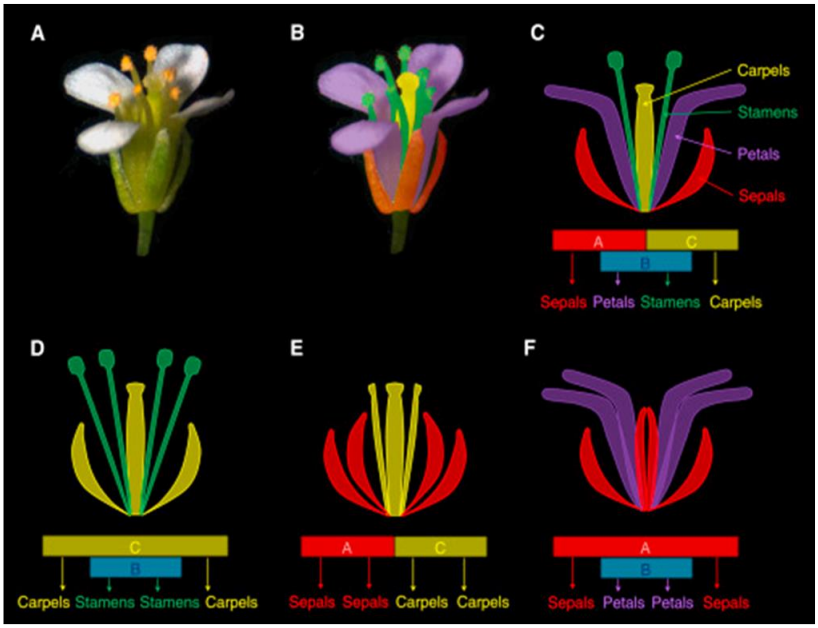

Figure 1.6: The genetic basis of flower morphology in *Arabidopsis thaliana*. Part A shows the morphology of a wild-type flower in *A. thaliana*, B- color-coded version of *Arabidopsis* flower (sepals; red, petals; purple, stamens; green and carpels; yellow). C- activity of ABC genes for the development of each organ. D, E and F- mutant phenotypes of *A. thaliana* (obtained from Irish 2017).

Mutations in other genes result in some interesting floral outcomes. For instance, a mutation in the TF protein-encoding gene *LEAFY* results in the development of leaves instead of flowers (Krishnamurphy and Bahadur 2015). Also, a mutation that affects the *ETTIN* (*ETT*)/auxin interaction causes larger flowers with extra sepals and carpels as well as less stamens (Waki et al. 2018). The *ETT*/ADP-ribosylation factor 1 (*ARF1*) complex works independently of *CLV* and *WUS*

for the formation of larger flowers. Therefore, this shows the genetic complexity of flower development and morphology.

Another interesting finding was that mutations in all ABC gene classes at the same time will produce leaves instead of flowers (Pelaz et al. 2001). According to Dornelas and Dornelas (2005), Von Goethe reported flower organs as modified leaves back in 1790 and this has now been observed in some plants. The ABC class genes are important for the formation of different flower organs (Refahi et al. 2021). However, reverse genetics studies on flower development genes have yielded questionable results. For instance, when class A and B genes were activated in leaves, such leaves did not develop into petals as expected (Pelaz et al. 2000). This led to the ABC model being revised into the ABC-DE model (Ghaemizadeh et al. 2019). In essence, it was discovered that there are two other gene classes (classes D and E) and their activities are also responsible for the formation of sepals, petals, stamens and carpels (Himani et al. 2019) (Fig. 1.7). Class D consists of the genes *SEEDSTICK* (*STK*) and *SHATTERPROOF1* (*SHP1*) (Favaro et al. 2003) and class E comprises of *SEP1*, *SEP2*, *SEP3* and *SEP4* (Ditta et al. 2004). In short, the activities of class A and E genes produce green sepals (Moyroud and Glover 2017). When class B genes are active alongside class A and E genes, they will together result in the petals formation. When class B genes are combined with class A and E genes, they will result in the attractive petals. The interactions among class B, C and E genes on the other hand, will produce stamens (male reproductive part), whereas the activities of class C and E genes will yield carpels (female reproductive part). Class D genes require the activity of class C and E genes to produce the ovules (Krishnamurphy and Bahadur 2015). Ultimately, all these organ identity genes in *Arabidopsis* are believed to be activated by *LFY* and *AP1* (Wellmer and Riechmann 2010).

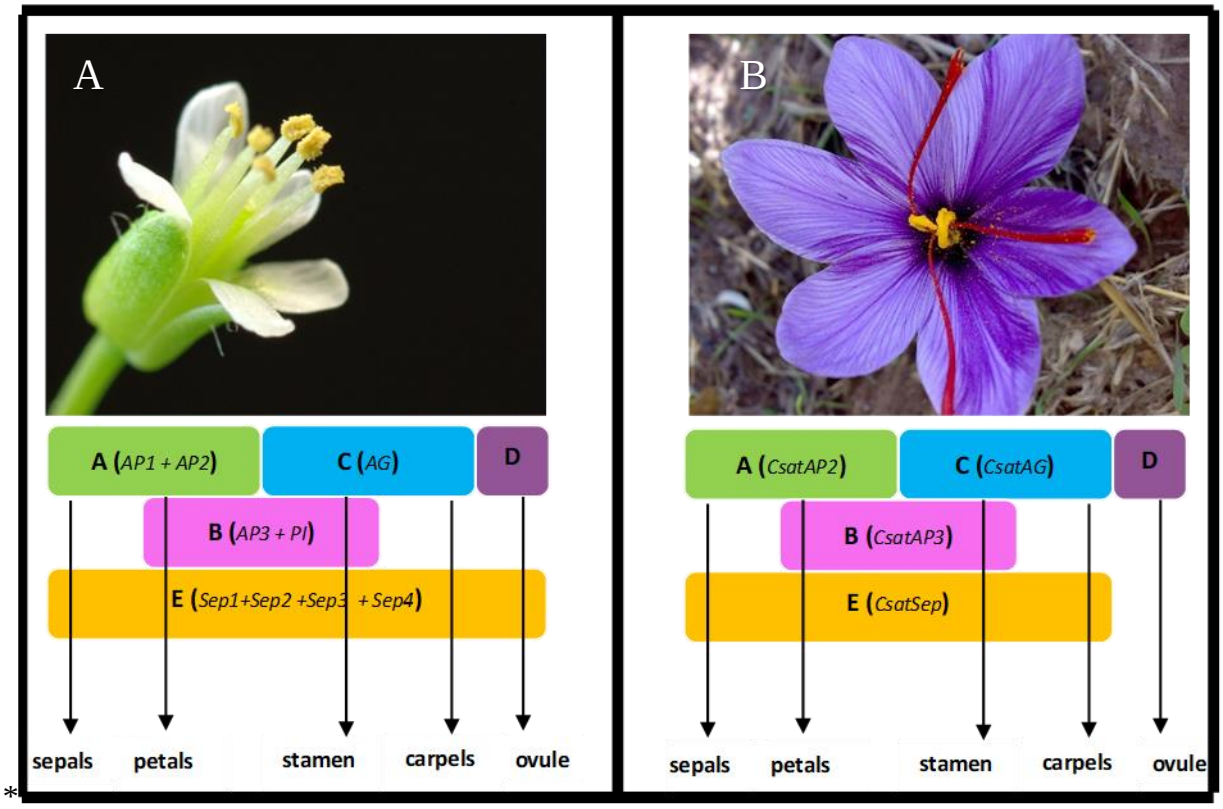

Figure 1.7: The revised ABCDE model showing genes in each gene class in *Arabidopsis thaliana* (https://za.pinterest.com/saimas/science/) and B; *Crocus sativus* (https://www.amazon.com/Crystal-Crocus-Sativus-Saffron-Bulbs/dp/B08XTVNN71?dchild=1).

The revised ABCDE model has since been confirmed in several plant species. Tsaftaris et al. (2011) cloned and characterized cDNA sequences encoding different MADS-box TFs that constitute the ABCDE gene classes in saffron (*Crocus sativus*). Expression analyses of the isolated genes showed that saffron supported the ABCDE model hypothesis by showing homologies as follows: the class A gene is *CsatAP2*, class B *CsatAP3/CsatPI*, class C *CsatAG* and class E *CsatSep*. Elsewhere, Yamade et al. (2009) established that expression of the class D gene *STK* was responsible for the formation of ovules in alloplasmic wheat lines (*Triticum aestivum* L.). Similarly, studies in different cereals (Ali et al. 2019) also found evidence supporting the ABCDE model. In rice (*Oryza sativa*), the class B and C genes' functions are dependent on the activity of class E

genes to specify stamen and petal development (Fornara et al. 2003). Therefore, the ABCDE model is essential for understanding the development and interconversion of different flower organs.

1.3.5. Polypetalous flowers

Some changes in phenotypes caused by mutations result in desirable traits. A polytepalous clivia is surely of high horticultural value. A ‘double flower’ is a showy bloom that is characterised by 17 to 25 petals revolving around the centre of the flower (Nashima et al. 2021). It is sometimes referred to as a flower within a flower (Mandel and Yanofsky 1995). Examples of plants exhibiting double flowers include hybrid tea rose, carnations and camellias (Meyerowitz et al. 1989). Another contributor to flower structure is the type of organs the flower is made of (Lin et al. 2021). Roses, for example, evolved from just five petals to having over 40 petals in modern-day specimens due to stamens being repeatedly converted into petals (Dubois et al. 2010). Mutations have significant influences on floral phenotype.

The double flower phenotype has been attributed to mutations in the class C gene, *AG* (Johnson and Lenhard 2011). In *A. thaliana,* a class C gene mutation converted stamens into petals and carpels into a new flower, and thus, producing a double flower phenotype (Waki et al. 2018). Similarly, in petunia, when class C genes named Floral binding protein6 (*FBP6*) and Petunia Mads box gene3 (*PMADS3*) have reduced expression, double flowers form (Gattolin et al. 2020). This suggest that the *AG* mutations directly impact flower numbers.

The interactions between *AG* and *WUS,* a gene that promotes the formation and maintenance of stem cells, interestingly terminates the maintenance of stem cells (Lenhard et al. 2001; Lohmann et al. 2001). Normally, *AG* is required for the development of stamens and carpels. However, in *ag* mutants of *Arabidopsis*, these organs are missing (Jibran et al. 2017). Seemingly, *AG* regulates *WUS* negatively, thus, preventing the indeterminate growth of the flower meristem (Cruz et al. 2021). The expression of *WUS* in *ag* mutants is turned off in the centre of the flower, allowing the

stem cells to differentiate and grow into new flowers (Mohrholz et al. 2019), forming the double-flower phenotype.

Another example of the *ag* mutant causing a double-flower phenotype was observed in camellia (*Camellia japonica*), which is highly priced in the floriculture/horticulture market due to its unique characteristics (Sun et al. 2014). The mutation resulted in the lack of reproductive structures (stamens and carpels) in the flower, meaning, the plants produce an endless array of petals creating a flower within a flower (Jibran et al. 2017; Ma et al. 2018). This phenotype can also be predicted using the ABCDE model, whereby a new flower forms within the original flower (in place of carpels) due to a mutation in the class C gene (Wollmann et al. 2010). Other examples are found in the transcriptomes of yellow horn (*Xanthoceras sorbifolium*) and Japanese rose (*Kerria japonica*) whereby the mutated *AG* gene (*Xsag* and *Kjag* respectively) resulted in double flowers (Ma et al. 2018; Wang et al. 2023). NGS can be used to expand knowledge of the relationship between class C genes and the double flower formation in addition to identifying further gene-phenotype correlations in other gene classes and more.

1.4. Next-Generation Sequencing (NGS)

NGS has revolutionized molecular biology research including plant biology (Brautpigam and Gowic 2010). Before the first application of NGS based on Roche's 454 technology in 2005 (Varadaraju et al. 2021), genome sequencing studies were conducted using the ever-popular Sanger sequencing method (Sanger et al. 1977). For example, the *Arabidopsis thaliana* genome was successfully sequenced using the Sanger method and was completed in 2004 (Kaul et al. 2000). Since then, *A. thaliana* has become one of the most important plants for plant molecular studies (Hamilton and Robin 2012). Advances in plant genome sequencing provide more accurate results in a much shorter period and at lower costs (Marks et al. 2021).

With NGS, identifying candidate genes involved in producing structural differences in flowers has become easier. This can be done using either forward or reverse genetics (Ma et al.

2017). For example, a specific phenotype is observed in a population and subsequently an attempt is made to identify the genes and sequence variants/mutations responsible for that phenotype. This is known as forward genetics. Reverse genetics is when a study aims to predict the phenotype that will be produced from a certain sequence or specific sequence variant. NGS, being a fast, efficient, cost-effective and high-throughput approach, provides the opportunity to fulfil the abovementioned scenarios.

1.4.1. Transcriptome research

NGS experiments typically follow a basic workflow that consists of three parts, namely DNA preparation, sequencing, and data analysis/bioinformatics (Fig. 1.8) (Hess et al. 2020). Initially, the biological sample is collected and identified. Then, from this sample, cDNA is synthesised, which is followed by library preparation (fragmentation, size selection and the addition of linkers) and the sequencing of cDNA templates. Lastly, the raw data (in the form of short sequence reads) produced from the NGS platform is analysed (Griffith 2015).

As an example, the Illumina protocol is described. The process of Illumina sequencing starts with the isolation and purification of RNA (Cheng et al. 2017). Transposomes then fragment the mRNA and tag it with adapters. Each fragment will be amplified in lanes, on a glass slide, that have different oligonucleotides complementary to the two adapters (Fig. 1.9). The fragment will bind to one complementary oligonucleotide, then DNA polymerase synthesises the second strand of the fragment. The now double-stranded molecule is melted and the template strand is discarded, leaving behind the strand that has just been synthesized. This strand folds over and aligns with the oligonucleotide complementary to the alternate adaptor (opposite end of the fragment) and the strand is used as a template, now in the opposite direction. After DNA synthesis, the two strands of the ds DNA are separated by melting and will bend bind to open oligonucleotides on the glass slide. This method of DNA amplification is called bridge amplification. This happens in all the clusters. After the removal of all the reverse strands, only forward strands are left for sequencing.

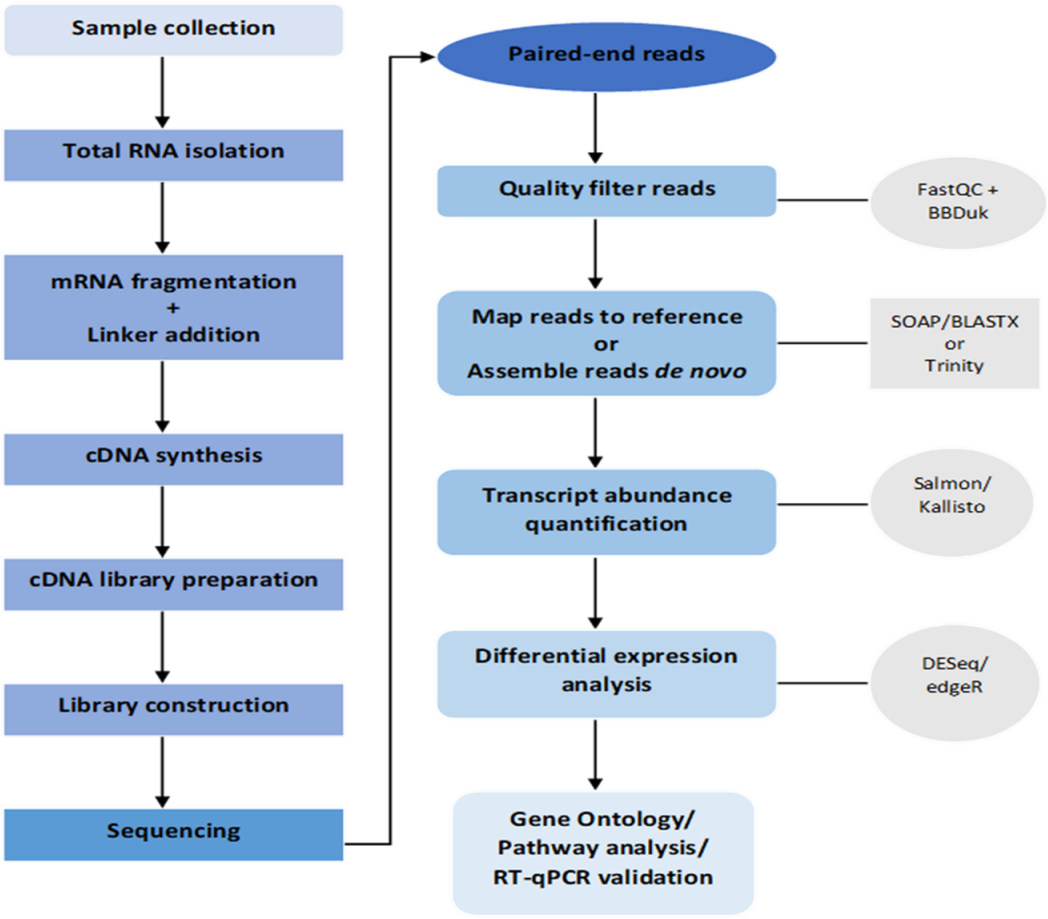

Figure 1.8: A typical RNA-Seq experiment workflow from sample preparation to data analyses (compiled using information from Singh and Jain 2014 and Griffith 2015).

The process of sequencing by synthesis of DNA fragments based on Illumina process has been reviewed many times before (Bentley et al. 2008; Liu et al. 2012; Sucker et al. 2012; Knief 2014; Taheri et al. 2018; Layton et al. 2019; Bjornson et al. 2020; El-Bairi et al. 2020; Nelson and Stewart 2020; Schilbert et al. 2020). When a nucleotide is incorporated on the newly synthesized strand, the clusters on the glass slide are excited and emit light because each nucleotide has a fluorescently labelled tag. This is done in a massively parallel way, whereby all the fragments that emit the same colour are sequenced simultaneously. This happens for the reverse fragments as well,

and then, the forward and reverse sequences are paired to create contiguous sequences. When aligned, these sequences are termed as paired-end (PE) reads. The raw PE reads can then be used for part three of the process, data analysis.

After sequencing, the data files containing raw sequencing reads are usually in the fastq format. The software FastQC (https//bioinformatics.babraham.ac.uk/projects/fastqc/), for example, can be used to open and visualise the data (Huang et al. 2020). The data will then undergo quality control analysis including adapter removal and length trimming followed by conversion to fasta format if necessary (Montero-Vargas et al. 2020). The quality trimmed reads can be mapped against a reference genome (if available) or assembled *de novo.* Having a reference genome sequence should simplify sequence assembly as well as the mining of many candidate genes. Also, a reference genome can serve as a vital resource for investigating gene functions (Vlk and Repkova 2017).

After read mapping, information about the number of reads that mapped to the genome and putative genes in the genome will be available. Further, read mapping data can also be used to study gene expression within and between samples. An R-based programme such as edgeR (http://bioconductor.org/packages/release/bioc/html/edgeR.html; Robinson et al. 2010) is better suited for scenarios where less than 12 biological replicates are being studied, whereas DESeq2 (http://www.bioconductor.org/packages/release/bioc/html/DESeq2.html; Love et al. 2014) is ideal where more than 12 replicates are available (Pessoa et al. 2021). The identification of differentially expressed genes (DEGs) can help to discover candidate genes involved in different phenotypical or morphological scenarios

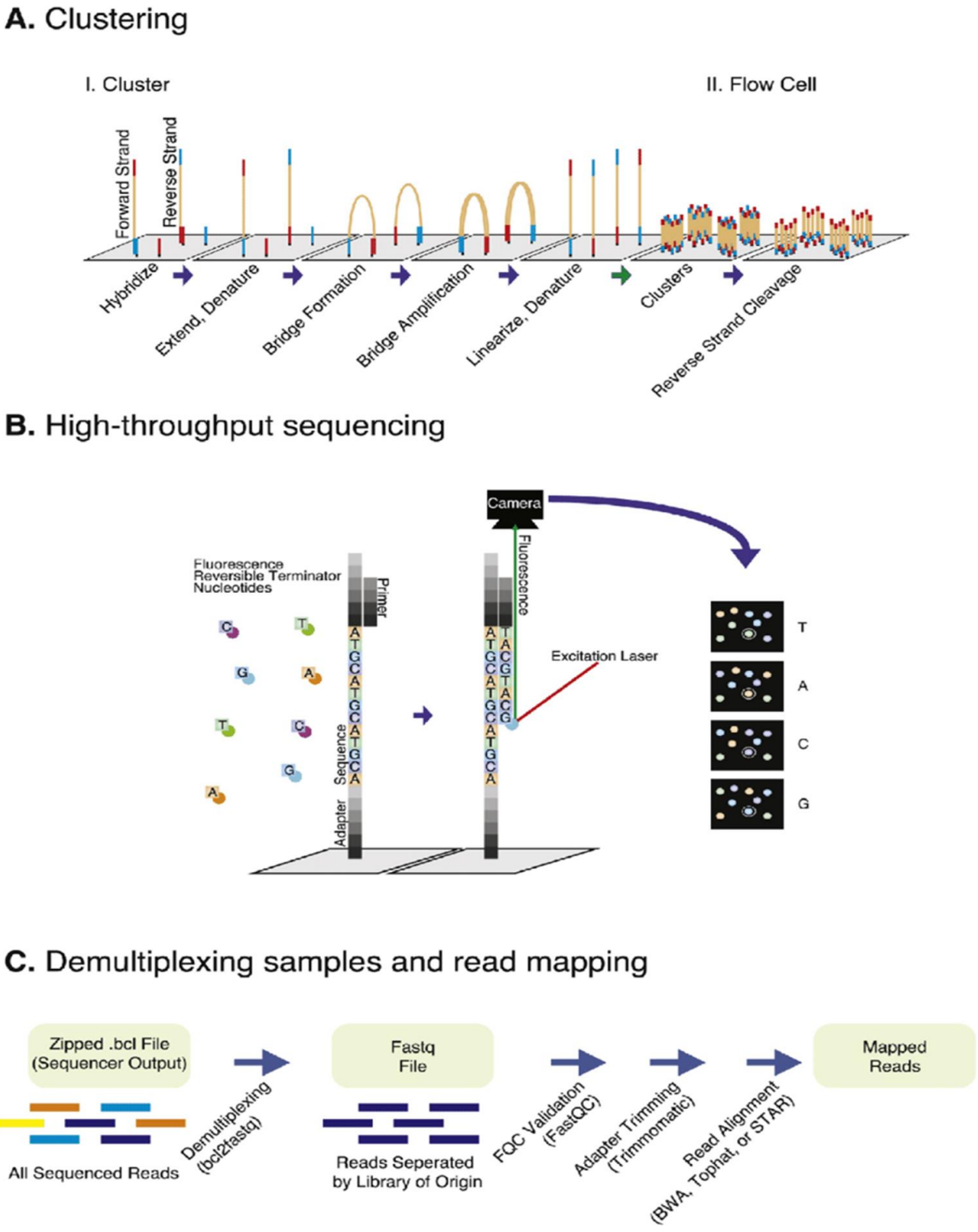

Figure 1.9: Illumina's sequencing-by-synthesis procedure (image obtained from Chaitankar et al. 2016).

.

1.4.2. Using NGS for gene discovery and gene expression analyses

Every cell in the multicellular organisms contain an identical genetic makeup but differ in the genes that are expressed, giving rise to a variety of cell types (Unamba et al. 2015). RNA sequencing (RNA-seq) is an NGS approach used to analyse the expression level of mRNA and enable the identification of genes that are expressed at any given time (Hong et al. 2021). The main limitation of RNA-seq as a genotyping tool specifically is that several genes in certain tissues may be lowly expressed or not expressed at all. (Onda and Mochida 2016). The advantages, however, outweigh the limitations because RNA-seq provides an enormous amount of information on expression in the form of either global or local gene expression profiles that reveal a specific section of the entire gene set (Wang et al. 2009; Liu et al. 2012; Stickler et al. 2012; Taheri et al. 2018). Overall, RNA-seq can be used to perform gene discovery and expression analyses, particularly in non-model organisms. The applications of NGS in gene discovery and expression analyses provide a basis for deeper understanding of processes such as flower development.

1.4.2.1. Applications for in gene discovery

NGS can be used to discover many candidate genes simultaneously (Xu et al. 2021). For example, Sasaki et al. (2017) confirmed that almost all flower development TF gene families associated with flower morphology reported in *A. thaliana* were also present in *Chrysanthemum morifolium* including *AP3* and *PI*. However, they also identified new MADS-box TFs such as the *TM6* gene identified for the first time in *C. morifolium* that is putatively involved in diverse floral trait formations such as floral architecture, petal shape and patterning (Sasaki et al. 2017). Liu et al. (2017) detected a total of 12,664 DEGs between a normal and a malformed flower of a sugar-apple (*Annona squamosa*), of which 701 were flower development-associated that, in turn were mostly downregulated in the malformed form. The downregulated DEGs included *AGAMOUS-LIKE 9* (*AGL9), SEP1, AP1-LIKE, AP3-LIKE* and *PI*. Gene discovery using NGS is not only limited to developmental genes. For example, Illumina sequencing was used to identify drought resistance-

induced TFs including dehydrins (DHN), Rubisco-binding protein (RBP) and calpains (ClpA and ClpP) in red clover (*Trifolium pratense*) (Vaseva et al. 2011). Also, Garnica et al. (2013) applied the RNA-seq method to search for genes that may aid the breeding of wheat lines that are resistant to the damaging pathogen, *Puccinia striiformis*. Overall, NGS techniques can be used for the discovery of candidate and novel genes involved in specific plant functions.

1.4.2.2. Applications for gene expression analysis in flower development

Differential gene expression analysis, in the context of NGS, refers to the use of statistical tools to compare quantitative changes in expression levels between two or more experimental sample groups (Salimonti et al. 2021). DEGs can provide valuable information on a wide range of processes such as flower development and morphogenesis. There are many statistical packages that can be used for expression analyses because, as NGS methods advance, so do the packages. Yang et al. (2020) identified 7,395 floral development DEGs in daylily (*Hemerocallis citrina*) using the Cufflinks software (http://cole-trapnell-lab.github.io/cufflinks/manual/; Trapnell et al. 2010).

Most of the DEGs were related to the regulation of the two processes (flower opening and senescence), which are the most import processes during daylily flower development (Nitta et al. 2010). NGS expression analysis is not limited to flower morphogenesis. Zeng et al. (2017) used RNA-seq to investigate gene expression in three developmental stages of crested wheatgrass, namely, the stem elongation stage, boot stage and anthesis. A total of 1,544 DEGs were identified between the stem elongation and boot stages, whereas 1,671 DEGs were detected between the boot and anthesis stages. Among the many DEGs was a circadian clock gene that showed higher expression in the stem elongation stage than in the boot stage, indicating its role in the transition to flowering (Zeng et al 2017).

Singh and Jain (2014) also used RNA-seq analysis to profile the expression of transcripts during flower development in chickpea (*Cicer arietinum*). The expression of 1,572 genes were

differential in floral tissues as compared to vegetative tissues. Among these, 1304 genes were upregulated and 269 genes were downregulated in floral tissues. Hu et al. (2021) used fertile and sterile Chinese cabbage (*Brassica rapa*) flowers before and after stamen (male reproductive organs) maturity. The sterile and fertile flowers had 5,645 DEGs in common, which suggested that these genes are important for stamen development. The fertile line had 6779 unique DEGs possibly indicating that these genes may be important for normal stamen growth. Genes unique to particular phenotypes can be studied further to uncover their involvement in morphogenesis.

1.5. Concluding remarks

The genus *Clivia* is naturally distributed along the east coast of South Africa and some parts of Swaziland. The plant prefers shady areas with little to no direct sunlight, especially in hot climates. *Clivia* consists of six species, five with pendulous flowers and one with non-pendulous flowers. The two morphologies are distinct and have different pollinators. This genus is of medical and horticultural importance due to its medicinal properties and showiness, making it a good genus to unravel further for morphology studies. Moreover, the differences in flower morphology within the genus presents an opportunity to dissect the genetics of flower development in *Clivia*.

Previously, analysis of developmental processes and gene interactions responsible for floral architecture in angiosperms revolved around model plants such as *Arabidopsis thaliana* and *Antirrhinum majus*. The ABCDE model of flower development, which is the most basic and widely accepted model that explains the genetic control of flower organ arrangement and morphology, along with the endogenous and exogenous pathways of flowering, presents a starting point for studying floral identity genes and transcription factors. Model species play an important role because the activities and gene expression profiles of well-understood floral regulators can now be analysed and compared in different species.

Numerous studies have clarified that flower development genes are to some degree conserved in all angiosperms. There are many genes that have been shown to be involved in flower

and organ development. Genes within these classes can be compared within a species or between different plant species. Gene homologs can be identified in other species, gene functions can be clearly outlined, and clearer developmental pathways can be established using NGS technologies. The advances and efficiency of NGS can be used in transcriptome analyses to identify morphogenesis genes and their functions in the anatomy of angiosperms, including the *Clivia* genus.

Chapter Two: Materials and Methods

2.1. Plant material and collection

A total of nine *Clivia* flower buds were collected and used for RNA extraction. These included three tiny bud (TB) samples collected from three independent pendulous plants, three intermediate (medium) buds (IB) from another three independent pendulous plants (*Clivia nobilis)* and three intermediate buds were collected from three independent non-pendulous plants (*Clivia miniata*; MB) (Appendix A; Fig. S1). The plants originated from local landscape gardens in Bloemfontein (Free State Province) and the private collection of an enthusiast in North Cliff (Gauteng Province). All samples were snap-frozen in liquid nitrogen immediately after picking and transported to the laboratory on dry ice. The samples were then stored in a -80°C freezer until required.

2.2. Total RNA extraction and NGS analysis

Frozen bud tissue samples were powdered with a mortar and pestle using liquid nitrogen. Then, total RNA was extracted from each sample using the TRIzol® reagent (Thermo Fisher Scientific, Waltham, MA, USA), based on the manufacturer's recommendations. Specifically, approximately 0.1 g of tissue was homogenised in 500 µl of TRIzol, followed by incubation at room temperature for 5 minutes. Next, 100 µl of chloroform was added per sample, shaken vigorously and incubated for another 3 minutes. Centrifugation was done at 12 000 x g for 15 minutes (4°C) to effect phase separation. The aqueous phase (top layer) was transferred to a new microcentrifuge tube and RNA was precipitated using 250 µl isopropyl alcohol. The mixture was incubated for 10 minutes at 35 °C and centrifuged as before. The RNA pellets were purified with 500 µl cold 75% ethanol, shaken vigorously, and centrifuged at 7 500 x g for 5 minutes (4°C). After discarding the supernatant, the pellet was air-dried, resuspended in RNAse-free water and incubated for 10 minutes at 58°C and chilled on ice before spectrophotometry. The RNA's quality was measured using a Nanodrop Spectrophotometer (Thermo Fisher Sci.), whereas quality was measured by gel electrophoresis (1.2% (w/v) TAE agarose gels). The extracted total RNA was sent to the ARC Biotechnology

Platform (https://www.arc.agric.za/pages/btp.aspx) for NGS on an Illumina HiSeq2500 platform. As further validation, the ARC BP also analysed the RNA samples for quality and quantity.

2.3. Read filtering based on quality and length

Upon receiving sequence data from the ARC, the raw reads were filtered for base call quality and length. Read quality was inspected using the software FastQC version 0.11.5 (http:/bioinformatics.babraham.ac.uk/projects/fastq). To filter the reads based on quality, sequences were freed of adaptors and trimmed using the BBDuk shellscript from the BBtools package (http://jgi.doe.gov/data-and-tools/bb-tools/) in a two-step process. The first step entailed trimming the Illumina adapters from reads and the second step is where the clean reads were obtained by removing low quality reads. Subsequent analyses were completed using the high-quality cleaned data.

2.4. *De novo* sequence assembly and assessment

The softwares Trinity-v2.10.0 (Grabherr et al. 2011) and SPAdes-v3.14.1 (Bankevich et. al 2012) were used to assemble different flower bud transcriptomes (*de novo*) into contigs and scaffolds. Specifically, an assembly was created for each of the three separate bud sample groups (hereafter, referred to as the 'Tiny buds or TB', 'Intermediate buds or IM' and 'Miniata buds or MB') as well as for all bud samples combined (the 'Combined assembly', hereafter referred to as the '*Clivia* transcriptome'). After assembly, transcripts in each individual assembly were clustered based on 0.95 and 0.99 sequence similarity using the CD-HIT-EST package (Li and Godzik 2006; Fu et al. 2012). The clustered assembly was assessed for quality through the N50 and ExN50 statistics as integrated into the Trinity software. N50 represents the length of the shortest scaffold that covers 50% of the total assembled sequences solely based on length whereas ExN50 (exact N50) takes both length and accuracy into consideration and excludes scaffolds with errors. Assembly completeness was analysed by mapping the reads back to the assembly (read representation) and

BUSCO analyses as follows. The software package Bowtie2 (Langmead and Salzberg 2012) was used to index the *Clivia* transcriptome into a reference and aligning the reads against the reference sequences. A good assembly should have no less than 80% of the reads aligning back to the reference (https://github.com/trinityrnaseq/trinityrnaseq/wiki/Transcriptome-Assembly-Quality-Assessment). The assembled transcripts (in the *Clivia* transcriptome) were compared to a reference dataset of orthologous genes from various species using the software BUSCO v5.4.3 (Manni et al. 2021). For the current study, the viridiplantea, eudicots and liliopsida lineage datasets were used for analysis.

2.5. Quantifying transcript abundance

The alignment-free quantification method of Salmon v1.10.1 (Patro et al. 2015) was used to measure the abundance of transcripts in each sample, and to build transcript and gene expression matrices. A principal component analysis (PCA) was performed to assess correlations among the replicates within each tissue sample and across all samples. All these analyses were done as implemented in the Trinity package.

2.6. Functional annotation and pathway mapping analysis

2.6.1. Functional annotation by homology searching

Functional annotation and gene ontology (GO) assignments of sequences in the *Clivia* transcriptome were performed using the Trinotate v3.2.2 pipeline (Bryant et al. 2017). For these analyses, the assembled transcripts were first annotated to protein sequences (open reading frames, ORFs) using the software TransDecoder v2.0.1 (Bryant et al. 2017), and then the proteins were searched against several databases including the Swissprot/Uniprot database (release 2022_01), KEGG (Kyoto Encyclopedia of Genes and Genomes) v105.0 and Gene Ontology (GO)

2.6.2. Enrichment analysis of GO and KEGG annotations

Enriched GO terms in the *Clivia* transcriptome were visualized using the PANTHER database (http://www.pantherdb.org/; Mi et al. 2013) and the software ShinyGO (Ge et al. 2020). The PANTHER database was also used to characterize protein classes enriched in the *Clivia* transcriptome. Pathway enrichment analysis based on KEGG annotations were performed at the DAVID database (Database for Annotation, Visualization and Intergrated Discovery (https://david.ncifcrf.gov/tools.jsp) using UniProt accessions obtained after analysis with Trinotate.

2.6.3. Flower development genes

Genes related to flower development were identified based on the list of *Arabidopsis thaliana* sequences available at the Flowering-Interactive Database (FLOR-ID; http://www.phytosystems.ulg.ac.be/florid/; Bouche et al. 2016). The *A. thaliana* gene sequences were obtained from TAIR (https://www.arabidopsis.org/) and converted into a local BLAST-searchable database to enable the identification of flower development genes. The transcript sequences in the *Clivia* transcriptome were searched against the database using the TBLASTN algorithms (Altschul et al. 1990). For these searches, the parameter max_target_seq was set to 10, whereas percentage ID and sequence coverage were set at ≥ 50%. The e-value cut-off was set to 10^{-5}.

2.7. Analysis and identification of DEGs

2.7.1. Differential expression analysis

The transcript abundance count matrices were used to determine differential expression (DE) between sample pairs based on the software DESeq2 version 1.36.0 (Love et al. 2014). A false discovery rate (FDR) of less than 0.005 (FDR < 0.005) and a log fold-change of 2 were used as minimum parameters to identify genes that show significant DE between pairs of samples. The DEGs were reported via a heatmap that showed sample clusters across all nine samples. Further, the transcripts showing DE between sample pair comparisons (MB vs TB, TB vs IB and IB vs MB)

were presented via Venn diagrams to illustrate transcripts shared between conditions. Venn diagrams were constructed online at this website (https://bioinformatics.psb.ugent.be/webtools/Venn/). Functional annotation of identified DEGs was carried out using the PANTHER DB and ShinyGO as before.

2.7.2. Candidate genes linked to the pendulous and non-pendulous flower morphologies

Genes showing significant DE between the pendulous and non-pendulous flower samples were screened to identify putative tissue-specific expression patterns. Identified genes were subsequently annotated against the Swissprot/UniProt database using BLASTx searches (e-value cut-off = 10^{-5}). The BLAST hits were manually inspected to identify flower development-related genes. Further, protein-protein interaction (PPI) network analysis was carried out using the STRING database v11.5 (https://string-db.org) to identify proteins that interact or show co-expression with identified genes of interest.

Chapter Three: Results

3.1. Plant material and RNA extraction

After the extraction of total RNA from the nine flower samples, the isolates were tested for quality and quantity prior to NGS. The analysis of RNA quantity via spectrophotometry revealed that the pendulous IB had the highest concentrations (Appendix A; Table S1). The mean A260/A280 and A260/A230 absorbance ratios were 2.01 and 2.2 respectively. As expected, agarose gel electrophoresis showed clear and distinct bands for all nine samples (Fig. 3.1). The top band representing the 28S RNA band appeared to be twice as bright as the lower 18S RNA band. The RNA isolates with the highest concentrations and distinct bands were submitted for NGS.

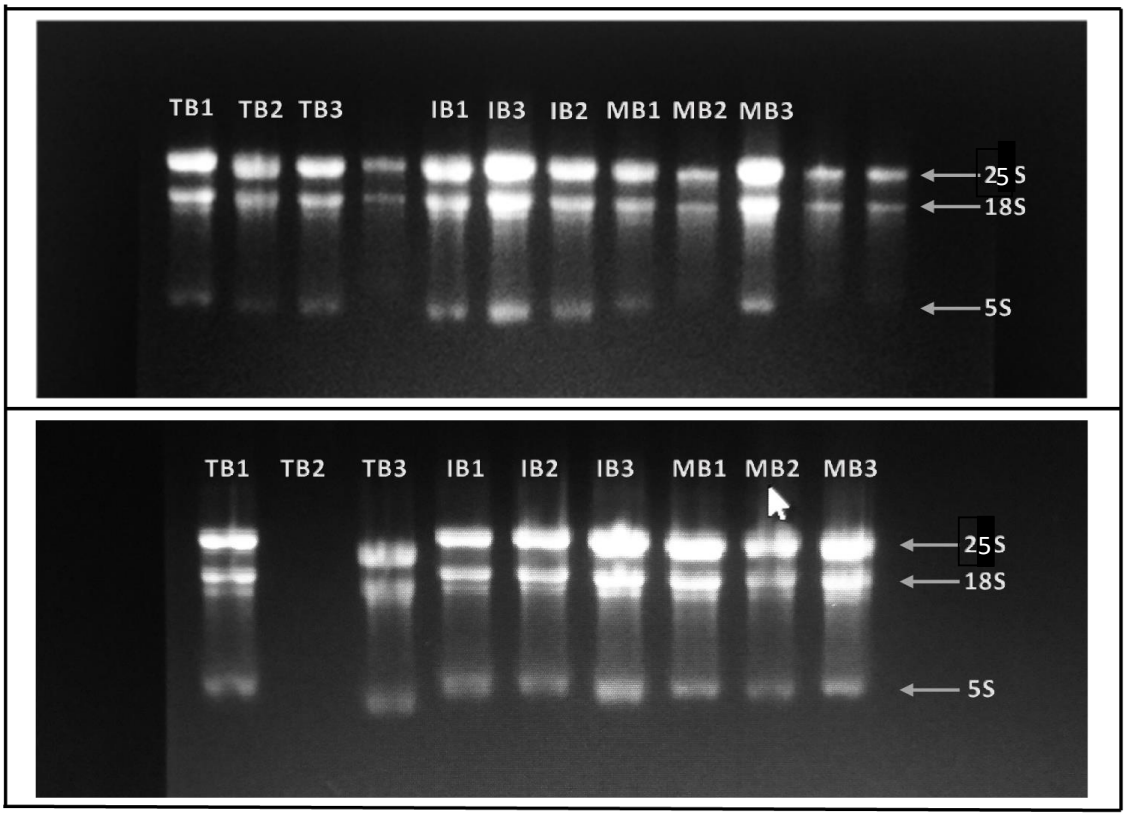

Figure 3.1: Agarose gel electrophoreses of RNA isolates from different flower buds in *Clivia*. The samples named 'Tiny buds' (TB) and 'Intermediate buds' (IB) originate from the pendulous flower of *Clivia nobilis* x *Clivia caulescens*, whereas the Miniata buds (MB) originated from the non-pendulous flowers of *Clivia miniata*. The bottom section of the figure represents the samples sent for sequencing.

3.2. Quality control and filtering of the reads

Illumina sequencing of the nine samples yielded 149,860,954 raw reads in total (Table 3.1). The raw reads were 151 base pairs (bp) long and the GC-content in different samples ranged between 45% and 47%. On average, 97% of the read sequences were retained after removing adapters and poor-quality reads. IB3 had the highest number of trimmed sequences and the lowest percentage of retained reads at 91.38%. Sequence length after trimming ranged from 10 to 121 bp.

Table 3.1: Illumina sequencing results before and after quality control and trimming. Inter buds refers to Intermediate bud samples.

Sample ID	Total sequences (raw reads)		Trimmed/retained reads	
	Before trimming	After trimming	No. trimmed seq.	% Reads retained
Inter Bud 1	17,700,307	17,430,943	269,364	98.48
Inter Bud 2	16,632,383	16,283,468	348,915	97.90
Inter Bud 3	17,017,710	15,550,267	1,467,443	91.38
Tiny Bud 1	17,053,333	16,664,428	388,905	97.72
Tiny Bud 2	15,505,110	15,107,990	397,120	97.44
Tiny Bud 3	16,896,296	16,658,013	238,283	98.59
Miniata Bud 1	16,682,778	16,508,029	174,749	98.95
Miniata Bud 2	16,746,195	16,401,516	344,679	97.94
Miniata Bud 3	15,626,842	15,237,717	389,125	97.51
Total	**149,860,954**	**145,842,371**	**4,018,583**	**97%**

3.3. *De novo* sequence assembly

The number of scaffolds assembled by Trinity varied for each assembly. The IB assembly contained a total of 398,674 scaffolds compared to 309,449 and 309,956 scaffolds in the MB and TB

assemblies, respectively (Appendix A; Table S2). The MB assembly, however, contained more scaffolds that are longer than 5 KB compared to the other two assemblies, whereas the TB assembly had more scaffolds longer than 2.5 KB compared to the other two assemblies. The SPAdes assembler provided the somewhat better results. The IB assembly contained a total of 343,957 scaffolds compared to 266,075 and 263,648 scaffolds in the MB and TB assemblies, respectively (Table 3.2). The MB assembly, however, contained more and longer scaffolds in the categories 1, 2.5, 5 and 10 KB compared to the other two assemblies, whereas the IB assembly was the only one with a scaffold longer than 25 KB. Based on these statistics, subsequent analyses were performed using assemblies generated with the software SPAdes.

The reads from three independent transcriptomes were further combined to create one flower transcriptome representing flower development in *Clivia*. The *Clivia* transcriptome contained 583,766 scaffolds that totalled 429.540 MB (megabites) (Table 3.3a). After clustering, 125 scaffolds of 10 KB or longer were clustered into the other scaffolds to yield a final number of 154 scaffolds longer than 10 KB (Table 3.3b). The transcriptome then contained 531,668 scaffolds totalling 353.907 MB. Overall, the combined assembly comprised the longest and highest number of scaffolds compared to the three individual assemblies, even after clustering the assembly.

Table 3.2: Scaffold length (bp) distribution in the *de novo* assembled bud samples.

A: Intermediate buds		
Minimum scaffold length	No. of scaffolds	Total scaffold length
100	343,957	193,093,350
250	235,373	169,063,016
500	97,361	122,943,268
1 KB	47,523	88,088,416
2.5 KB	7,754	26,975,727
5 KB	646	4,238,499
10 KB	41	531,791
25 KB	1	31,235
B: Miniata buds		
Minimum scaffold length	No. of scaffolds	Total scaffold length
100	266,075	180,420,442
250	192,425	164,134,954
500	94,183	130,983,734
1 KB	50,682	100,315,344
2.5 KB	10,394	37,373,980
5 KB	1145	7,421,485
10 KB	49	566,224
C: Tiny buds		
Minimum scaffold length	No. of scaffolds	Total scaffold length
100	263,648	162,846,836
250	183,834	145,035,243
500	84,911	112,093,002
1 KB	44,744	83,746,273
2.5 KB	7,613	26,180,429
5 KB	642	4,022,498
10 KB	30	365,473

Table 3.3: Scaffold length (bp) distribution in the *de novo* assembled combined transcriptome. The combined assembly comprises of all the bud samples (**a**) before and (**b**) after clustering the transcriptome.

a. Before clustering		
Minimum scaffold length	No. of scaffolds	Total scaffold length
100	583,766	429,540,032
250	424,087	394,319,989
500	210,848	322,304,751
1 KB	118,752	257,659,911
2.5 KB	31,144	119,407,327
5 KB	4,736	31,733,819
10 KB	279	3,481,970
25 KB	1	25,078
b. Post clustering		
Minimum scaffold length	No. of scaffolds	Total scaffold length
100	531,668	353,907,434
250	380,878	320,591,866
500	175,900	251,368,059
1 KB	92,311	193,136,342
2.5 KB	22,370	83,701,445
5 KB	2,983	19,718,394
10 KB	154	1,917,065
25 KB	1	25,078

3.4. Completeness of the assembled *Clivia* transcriptome

3.4.1. Read representation in the assembly as well as N50 and ExN50 statistics

Read representation in the *Clivia* transcriptome was assessed and determined using Bowtie2. Overall, 97.23% of the reads aligned back to the transcriptome (Appendix A; Fig. S2). Assemblies of the three sample groups (IB, MB and TB) and the combined assembly all showed a GC-content averaging 39.5% (Table 3.4). The MB assembly showed better assembly statists compared to the other assemblies, as it had the highest average contig length, median contig length and N50 value at 678.08, 344 and 1192 bp, respectively. Half of the assembled bases were found in contigs with a minimum length of 875, 1192, 1046 and 1176 bp for IB, MB, TB and combined assembly, respectively. The combined assembly had the highest total assembled bases, as expected. The ExN50 statistic (Appendix A; Table S3), displayed the highest Ex value (peak) at 91 and, therefore,

implying that an ExN50 value of 2,754 bp is representative for the transcripts of this assembly (Fig. 3.2). Ex value represents the top most highly expressed transcripts that represent X% of the total normalized expression data (Geniza and Jaiswal 2017).

Table 3.4: Summary of the scaffold and N50 statistics in the three assemblies.

Parameter	Intermediate Bud	Miniata Bud	Tiny Bud	Combined
Total assembled bases	193,093,418	180,420,442	162,846,836	353,907,434
Median contig length	309	344	323	334
Average contig length	561.39	678.08	617.67	665.65
%GC	39.45	39.29	40.10	39.01
N50 (bp)	857	1192	1046	1176

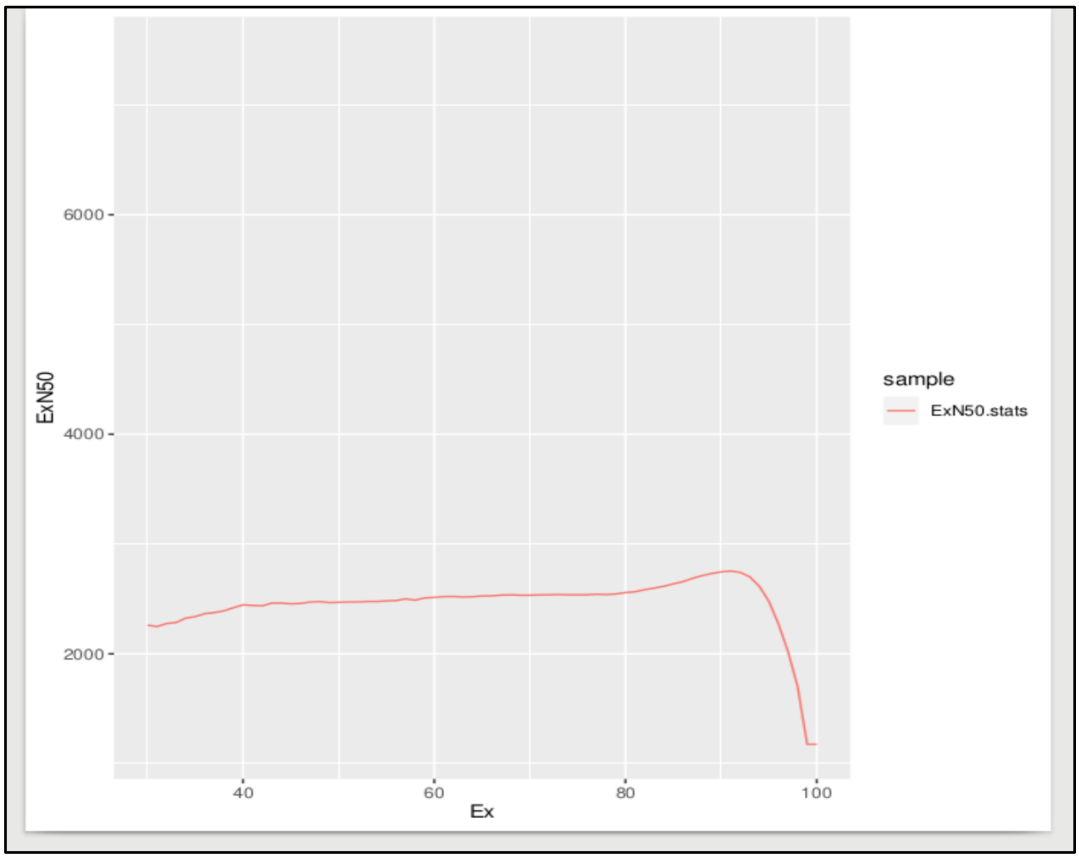

Figure 3.2: Plot of the Ex-value against ExN50 statistics of the main assembly.

3.4.2. BUSCO (Benchmarking Universal Single-Copy Orthologs)

For the analyses of BUSCOs, the Viridiplantae dataset contained the most references, followed by Eudicots and the monocot-specific lineage Liliopsida. As expected, most of the transcripts in the *Clivia* transcriptome matched sequences within the lineage Liliopsida (Table 3.5). Specifically, 90.5% of the BUSCOs were complete and only 2.9% were missing. The lineage Viridiplantae revealed the least total BUSCOs matching the *Clivia* sequences, but there were no missing BUSCOs. The Eudicots lineage only had 2.1% fragmented BUSCOs and 6.7% of the BUSCOs were missing.

Table 3.5: The assessment of assembly completeness through the analysis of BUSCOs.

BUSCOs	Viridiplantae	Eudicots	Liliopsida
Total datasets	425	2326	3236
Complete	421	2121	2930
Complete and single copy	186	930	1318
Complete and duplicated	235	1191	1612
Fragmented	4	50	215
Missing	0	155	91

3.5 Quantification of gene and transcript abundance.

3.5.1. Biological replicates and Principal Component Analysis (PCA)

A PCA was constructed to determine the relationships between and among the nine flower bud samples (Appendix A; Figure S3). Overall, it revealed that there is a good correlation among the three replicate sets for each sample group (Appendix A; Fig. S3). Sample correlation was confirmed by the display of more differentially expressed transcripts (red dots) between replicate 3 and the

other replicates for all three sample groups. Overall, PCA (Fig. 3.3) shows that the sample replicates grouped together away from the other sample groups.

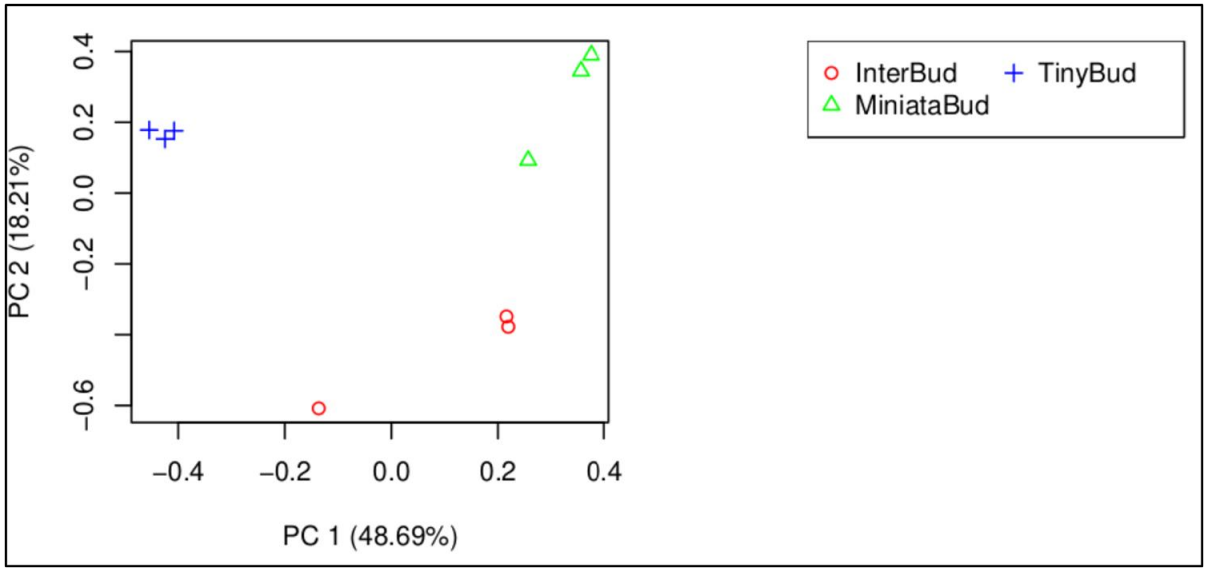

Figure 3.3: Principal Component analysis (PCA) showing correlations among replicates within each bud sample.

3.6. Functional annotation and gene/transcript enrichment analyses

3.6.1. Sequence homology searches and functional annotation

Functional annotation was done through sequence homology searches against the Swissprot/Uniprot and NR databases where sequence similarity searches were performed against the databases using the BLASTx algorithms (E-value cut-off set to $\leq 1e^{-5}$). A total of 62,580 and 145,080 annotations were identified against the Swissprot and NR databases, respectively. Overall, the results from both tests were similar with minor deviations excluding species distribution (Fig. 3.4). Approximately 4% and 5% of the mapped sequences were exact matches (E-value = 0) for Swissprot and NR, respectively. Further, 30% of the query sequences revealed similarities > 80% in both tests. Regarding the species hits at the Swissprot database, more than half of the annotated sequences matched subjects from *Arabidopsis thaliana* (51%). The other species with significant matches

included *Oryza sativa* (17%), *Nicotiana tabacum* (3%), *Solanum nigrum* (1%) and *Zea mays* (1%). The species hits from NR were different with almost half of the annotated sequences matched subjects from *Asparagus officinalis* (45%), whereas other species included *Iris pallida* (11%), *Elaeis guineensis* (6%), *Ananas comosus* (3%) and *Dendrobium catenatum* (2%). Interestingly, 46 queries matched sequences from *Clivia* species including *C. miniata, C. gardenii, C. caulescens* and *C. robusta* (data not shown). Overall, many sequences in the *Clivia* transcriptome hit transcripts from other plant species, specifically other monocots.

3.6.2. Enrichment analysis: GO and KEGG

3.6.2.1. GO annotation

GO is an internationally standardized functional classification system for genes, aimed at describing the properties of genes and their product within an organism using a dynamic-updated controlled vocabulary. It comprises three main categories, namely, molecular function (MF), cellular component (CC) and biological process (BP). GO analysis of the *Clivia* transcriptome revealed that the most popular annotations under the BP category were the “primary metabolic process”, “biological regulation”, “organic substance metabolic process”, “cellular metabolic process” and “developmental process” (Fig. 3.5). Under the scheme MF, the popular terms included “catalytic activity”, “ion binding”, “transferase activity” and “organic cyclic compound binding”. The predominant ontologies under CC included “cellular anatomical entity”, “intracellular, organelle, membrane-bound organelle” and “cytoplasm”. Most of the observed subcategories are relevant in flower development.

Figure 3.4: Characterization of the *Clivia* transcripts against the SwissProt database. A) Statistics of BLASTx searches against the SwissProt database including the distribution of E-values, sequence similarity and top species hits. B) Statistics of BLASTp searches against the SwissProt database including the distribution of E-values, sequence similarity and top species hits.

The *Clivia* transcripts were further aligned to the *Arabidopsis thaliana* proteome using ontologies at the PANTHER database (Fig. 3.6). This yielded 23 different classes of protein functions, of which the majority were “metabolite interconversion enzymes” (30.5%) (Fig. 3.6). Other significant protein functions included “protein modifying enzymes” (14.4%), “transporter proteins” (11%), “RNA metabolism proteins” (7.9%) and “gene-specific transcriptional regulators” (7.8%). These protein class categories show metabolites and metabolism as highly active protein functions.

3.6.2.2. KEGG

To better understand biochemical pathways in the *Clivia* transcriptome, all transcripts were mapped against the KEGG database to identify transcripts involved in specific metabolic pathways. As a result, 112,335 protein sequences inferred from the *Clivia* transcripts were assigned to 436 biological pathways. These include “metabolic pathways” (K:01100) with 1053 proteins, “biosynthesis of secondary metabolites” (K:01110) with 475 proteins, “microbial metabolism in diverse environments” (K:01120) with 163 proteins, 133 proteins for “biosynthesis of cofactor” and 101 proteins (K:01200) under “carbon metabolism”. Among these pathways, the circadian rhythm-plant pathway (K:04712) was analysed further due to its connection to flowering and flower development. Overall, 24 sequences from the *Clivia* transcriptome (Fig. 3.7) were identified that are potentially related to the circadian rhythm pathway. These included sequences encoding the proteins GI, cryptochrome (CRY), phytochrome A (PHYA), several MYB containing TFs, LHY, CCA1 and TOC1. Photoperiodic genes were also identified, and they included *PIF3*, *ELF3*, *FT* and *CO*.

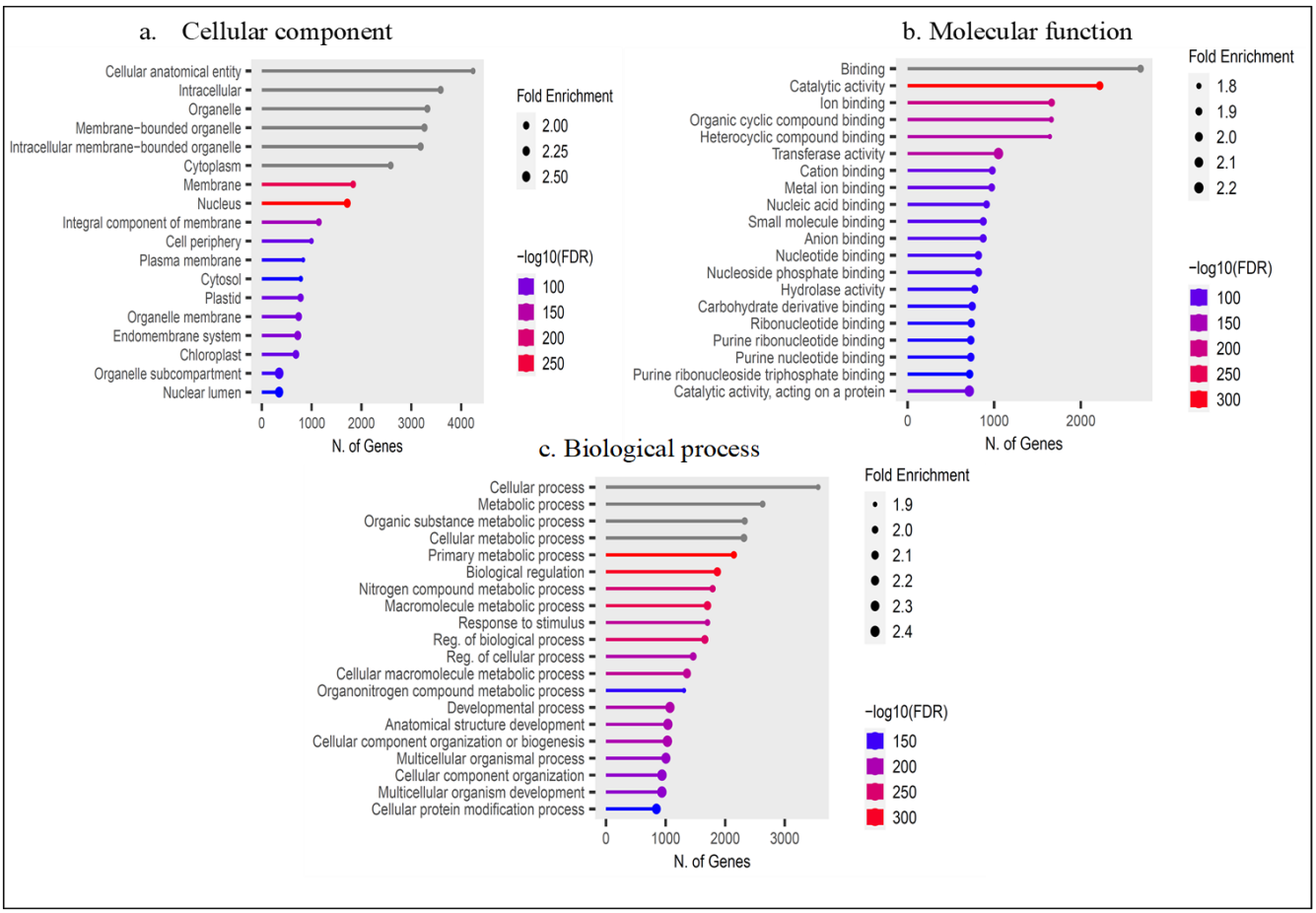

Figure 3.5: Gene ontology enrichment analysis of the *Clivia* transcriptome. The 20 top-most significantly enriched genes in enrichment categories (a) cellular component, (b) molecular function and (c) biological process are depicted. The false discovery rate (FDR) cut-off was set at 0.05.

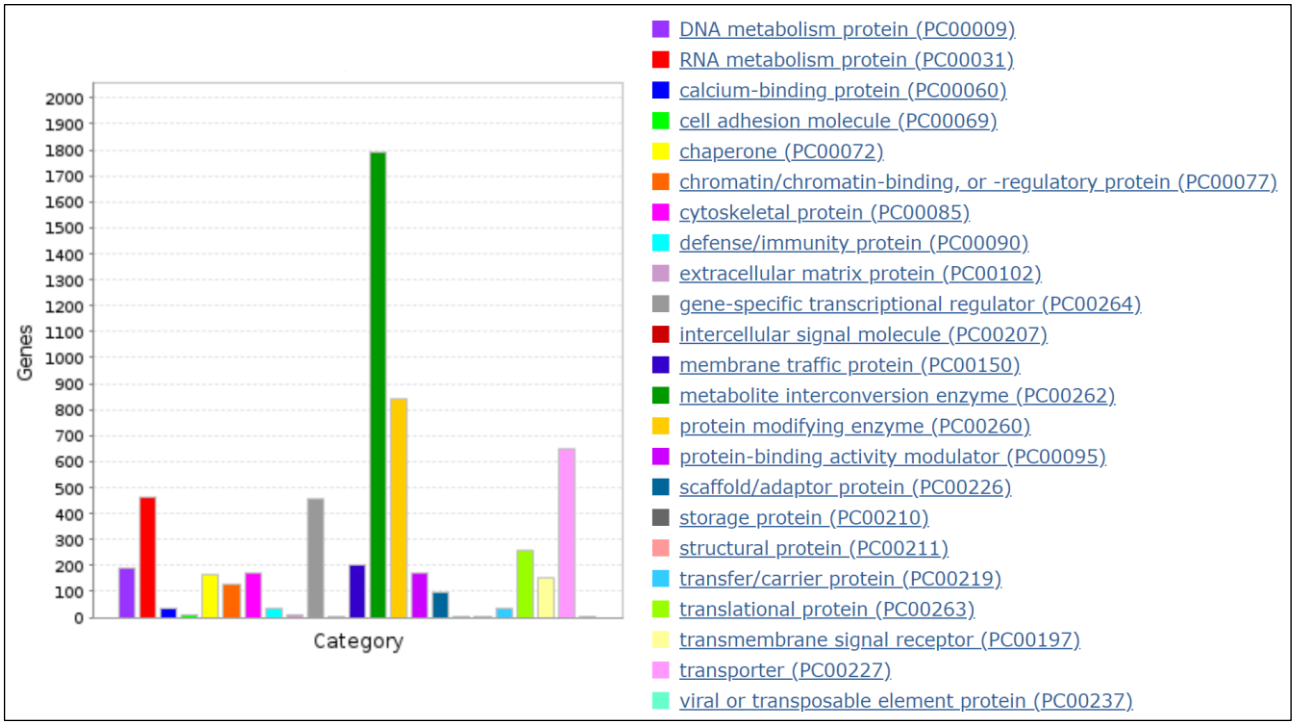

Figure 3.6: PANTHER protein class categories of the *Clivia* transcriptome assembly. Bar graph depicting the broad biological functions associated with the transcriptome's proteins. The colour legend specifies each category.

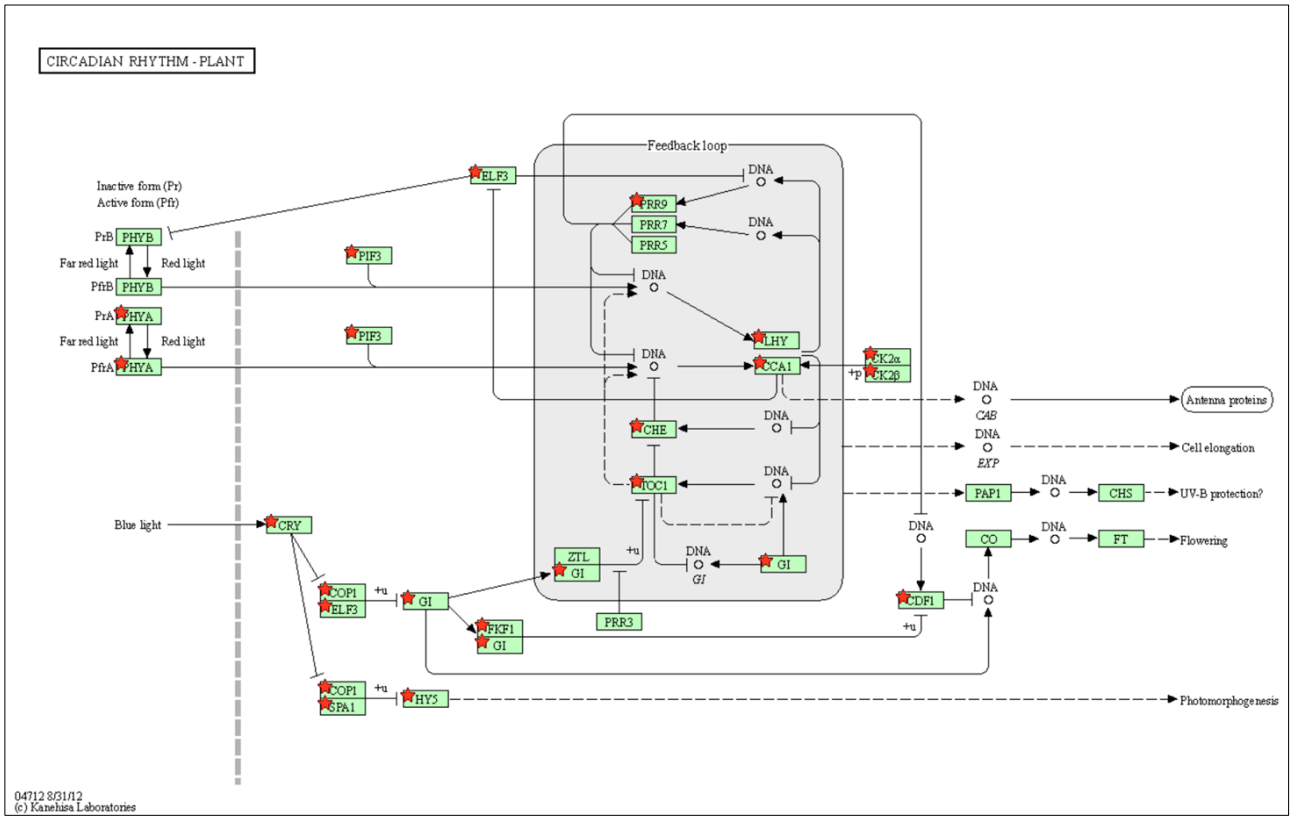

Figure 3.7: KEGG pathway enrichment analysis of genes involved in the circadian rhythm pathway. The red stars indicate genes contained in the *Clivia* transcriptome.

3.6.3. Identification and characterization of flower development genes.

Flower development genes of *Arabidopsis* (http://www.phytosystems.ulg.ac.be/florid/) were used as reference sequences to identify the corresponding potential orthologs in the *Clivia* transcriptome (Fig. 3.8). Overall, 46 flower development genes were selected from the FLOR-ID database as reference genes. Sequence homology searches revealed 70 sequences from the *Clivia* transcriptome having matched the 46 reference genes. These hits (Table 3.6) represented 17 different flowering genes, including *AG*, *AP1*, *AP2*, *CAL*, *SEP* and *UTL*. Some of these genes are the main candidates that constitute the ABC model of flower development (Fig. 3.8).

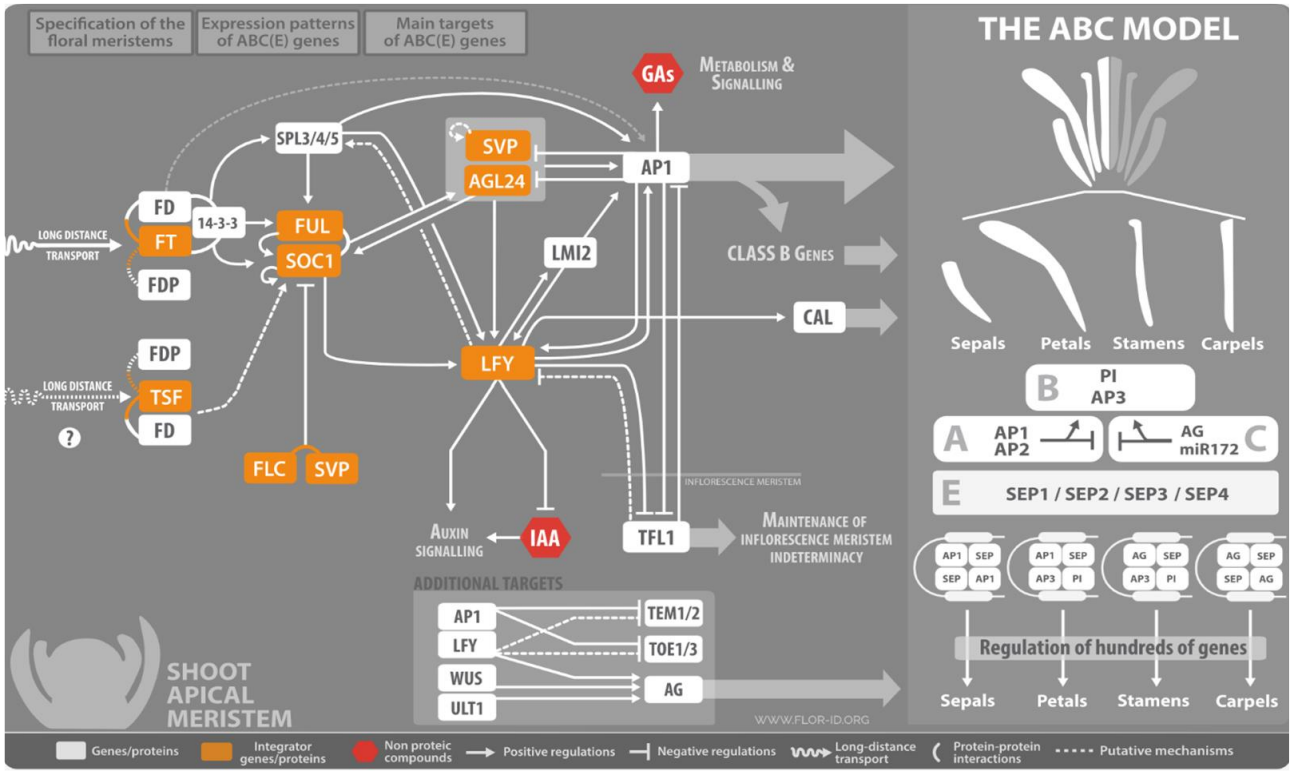

Figure 3.8: Flower development pathway and the ABC model of flower development based on the FLOR-ID database. A total of 46 genes featured in the models were used as references for the identification of flower development genes in *Clivia*.

Table 3.6: Identification of flowering genes in the *Clivia* transcriptome based on sequence homology to the genes in *Arabidopsis thaliana*. The number of transcript hits (obtained for each flower development gene from the FLOR-ID database) alongside the average percentage similarity, average *p*-value and the function of the identified *Arabidopsis* gene during flower development.

TAIR_ID	Gene name	No. of transcript hits	Mean %ID	Mean p-value	Functional annotation
AT4G18960	AGAMOUS (AG)	3	65.2	4.8E-95	Specifies floral meristem and carpel and stamen identity
AT1G69120	APETALA 1 (AP1)	5	63.4	4.36E-70	Specifies floral meristem and sepal identity
AT4G36920	APETALA 2 (AP2)	2	63.8	1.65E-124	Specification of floral organ identity, establishment of floral meristem identity, suppression of floral meristem indeterminacy, and development of the ovule
AT3G57130	BLADE ON PETIOLE 1 (BOP1)	3	68.5	5.61E-211	Nectary development and formation of normal abscission zones.
AT2G40360	BLADE ON PETIOLE 2 (BOP2)	7	65.6	1.53E-264	Essential role in cell growth
AT1G26310	CAULIFLOWER (CAL)	5	63.79	4.35E-60	Specification of flower meristem
AT4G38130	HISTONE DEACETYLASE 1 (HDA1, HDA19)	6	75.9	1.48E-251	Plays a role in embryogenesis
AT3G02310	SEPALLATA 2 (SEP2)	5	63.9	3.55E-78	Ensures proper development of petals, stamens and carpels
AT2G03710	SEPALLATA 4 (SEP4)	1	69.9	1.55E-62	Plays a central role in the determination of flower meristem and organ identity
AT5G15800	SEPALLATA1 (SEP1)	5	63.8	3.55E-79	Involved flower and ovule development
AT2G42830	SHATTERPROOF 2 (SHP2)	1	61.8	1.79E-50	Involved in fruit development

AT1G62360	SHOOT MERISTEMLESS (STM)	2	72.45	1.88E-126	Prevents incorporation of cells in the meristem centre into differentiating organ primordia
AT2G45640	SIN3 ASSOCIATED POLYPEPTID P18 (SAP18)	7	70.29	1.98E-58	Involved in the regulation of salt stress
AT2G25170	SUPPRESSOR OF SLR 2 (SSL2)	8	60.49	1.23E-201	Involved in post-germination repression of embryonic development
AT5G03840	TERMINAL FLOWER 1 (TFL1)	4	69.4	2.76E-86	Controls inflorescence meristem identity
AT4G28190	ULTRAPETALA 1 (ULT1)	1	68.8	2.19E-118	Production of extra flowers and floral organs
AT2G20825	ULTRAPETALA 2 (ULT2)	1	64.9	6.97E-109	Negative regulator of cell accumulation in shoot and floral meristems.

3.7. Analysis and identification of DE transcripts

3.7.1. Differential expression of transcripts

Gene expression of transcripts differed between the pairs of sample groups. The MA (M for log-ratio and A for log-average on the axes) plot and Volcano plots show the number of significant DE transcripts between sample groups (Appendix A; Fig. S4). As visulaised on the heat map (Fig. 3.9) there are two clusters of DEGs, the top (**A**) and bottom (**B**) clusters. Cluster A is further divided into 2 sub-clusters whereby in the top sub-cluster, most transcripts in the tissue IB seem to be highly expressed (the yellow segments), while transcripts in MB show low expression levels and TB revealed both high and low expressed transcripts. In the bottom sub-cluster, IB and MB mostly still contains highly expressed transcripts, whereas TB shows low transcript expression levels (the purple segment). Cluster B shows low expression for IB and MB compared to the highly expressed transcripts of TB.

Expression analysis using DESeq2 (p-value < 0.05) yielded 3,026, 14,928 and 7,411 DEGs for the Intermediate/Miniata, Miniata/Tiny and Intermediate/Tiny comparisons, respectively (Fig. 3.10a). Only 119 DEGs were common to all three tissue comparisons. Among the DEGs, 2414 were upregulated and 612 downregulated in IB/MB, 4107 were upregulated and 3304 downregulated in IB/TB and 7031 upregulated and 7915 downregulated) in MB/TB (Fig. 3.10b). The IB/MB and IB/TB tissue comparisons revealed only 902 DEGs. The majority of DEGs occurred between the MB/TB bud tissue comparisons. Overall, 9,201 DE transcripts were unique to the MB/TB bud tissue comparison, whereas 749 occurred in the Inter/Miniata tissues and 2038 in the Inter/Tiny tissues.

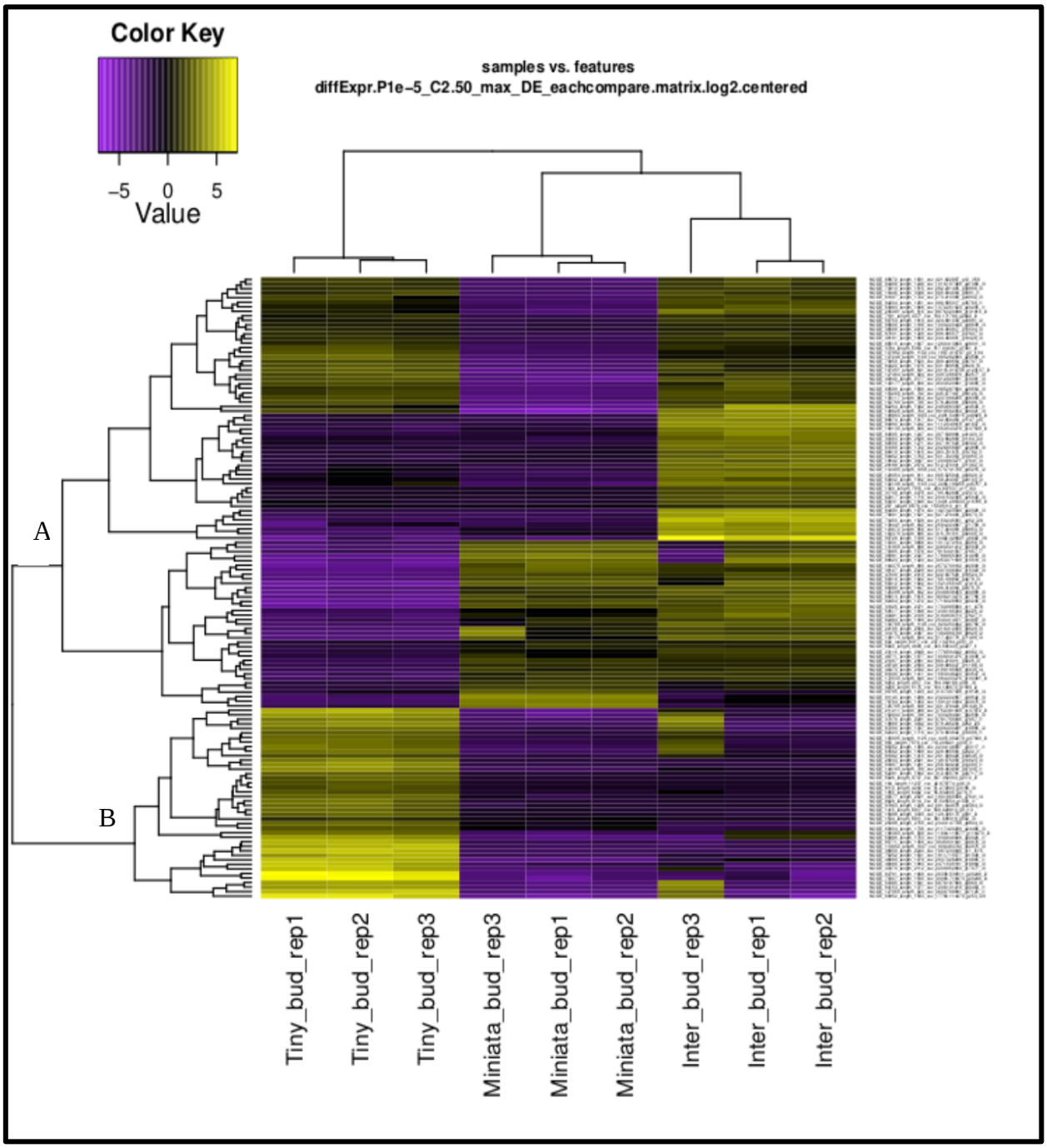

Figure 3.9: Heatmap displaying differential expression of the top 50 genes within pairwise comparisons. Significantly expressed transcripts (FDR < 0.001) with 2 fold differential expression are shown in normalized TPM value. Colour key indicates the intensity associated with expression. The yellow colour indicates highly expressed or upregulated transcripts whereas the purple transcripts are downregulated or have low expression levels.

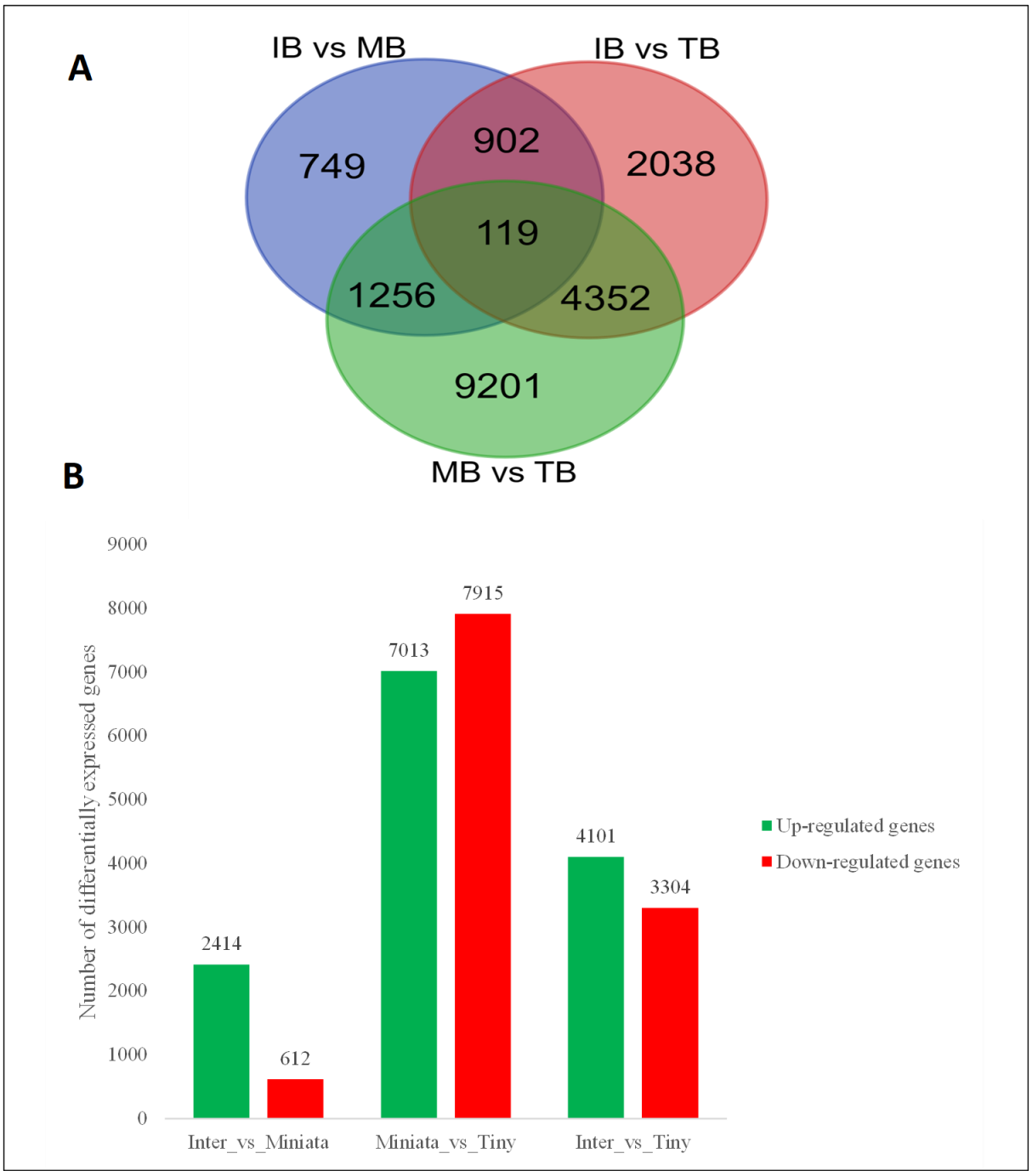

Figure 3.10: General profile of differentially expressed genes among three pairwise comparisons. (**a**) Number of differentially expressed transcripts pairwise comparisons of the three bud tissues. (**b**) Bar graph showing the number of significantly upregulated (green) and down regulated (red) genes.

3.7.2. Annotation of DE transcripts

GO enrichment analyses of the DE transcripts were performed using the PANTHER database. According to these analyses, 28 ontologies were enriched in the DE transcripts (Fig. 3.11). Under the scheme BP, the DE transcripts were involved in processes including “cellular process” (GO:0009987), “metabolic process” (GO:0008152), “biological regulation” (GO:0065007), “localization” (GO:0051179) and “developmental process” (GO:0032502). MF ontologies included involvement of the DEGs in “catalytic activity” (GO:0003824) and “binding” (GO:0005488), followed by “transporter activity” (GO:0005215) and “ATP-dependent activity” (GO:0140657). The CC shows enrichment in only two processes, namely, “cellular anatomical entity” (GO:0110165) and “protein-containing complex” (GO:0032991).

3.7.3. Transcripts unique to the pendulous or non-pendulous flower morphology

Searching the 46 flowering genes of *Arabidopsis* against the *Clivia* transcriptome revealed several transcripts, some of which encode TFs. These included transcripts encoding *GI*, the TF named ABA Ethylene-responsive (ERF) as well as several bHLH (bHLH140 and bHLH74) and MYB (MYB2, MYB8 and MYBS3) TFs that were specific to the pendulous flower samples. A few other proteins, including *FRI*, and the TFs bHLH130, MYB17 and TCP3 were uniquely detected in the non-pendulous flower morphology. The most interesting transcript was the *AGAMOUS* homology 1, which encodes a MADS1-containing domain - flower development-related gene (Roy-Choudhury et al. 2012). Analysis of the *AG* protein sequence at the STRING database revealed associations between *AG* and other flower developmental proteins (Fig. 3.12). Overall, the interaction network showed that *AG* directly interacts with *SEP3, AP3* and *AGL2*. Other proteins that seemingly interact with *AG* include *WUS, AP2, LFY, SUPERMAN (SUP), BEL1* and *LEU*. Transcripts encoding some of these TFs were identified in the *Clivia* transcriptome and may possibly explain the morphological differences between pendulous and non-pendulous flowers in *Clivia* species.

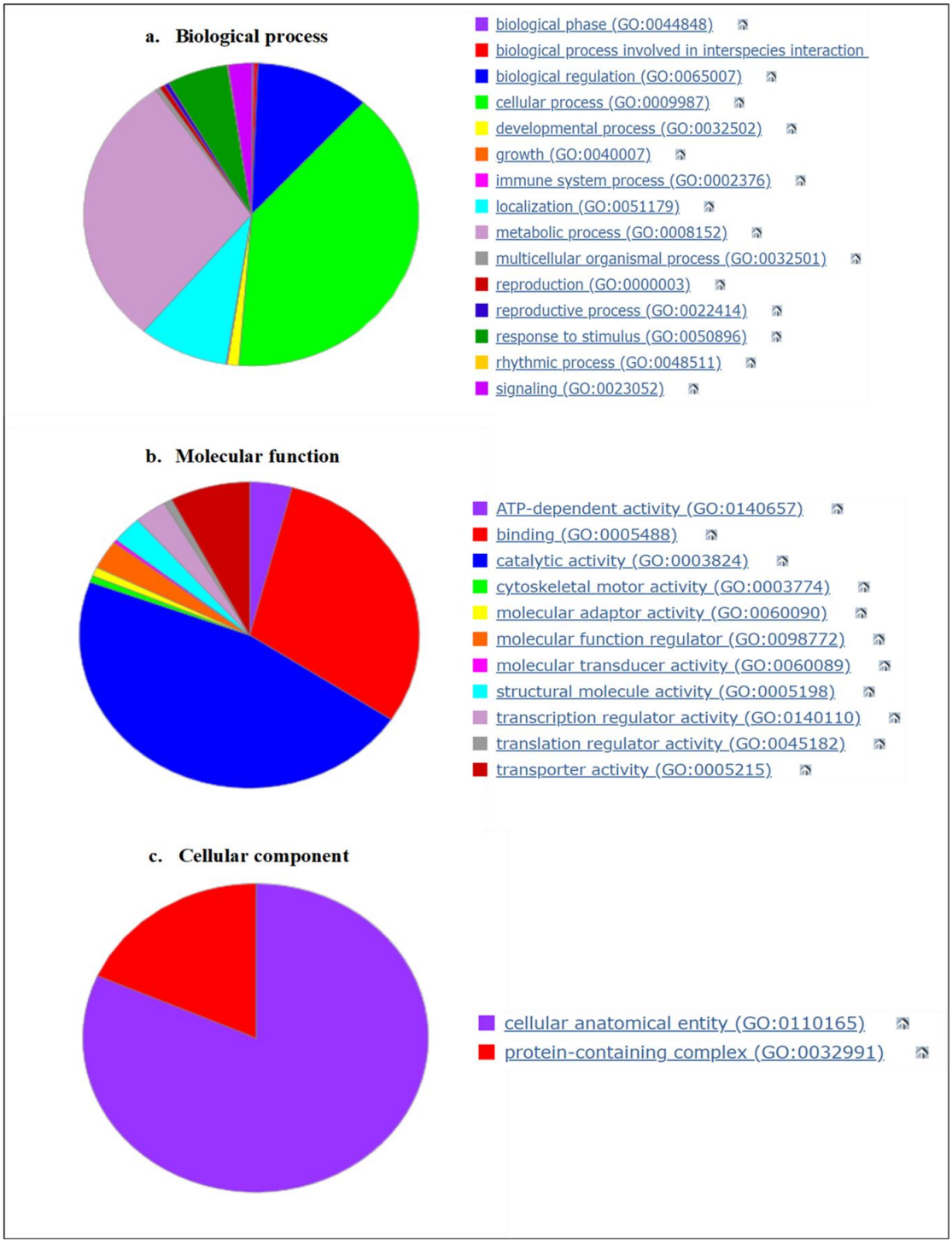

Figure 3.11: GO annotation of the DE transcripts. The colour legend indicates the relative enriched GO terms. The genes are distributed among three categories: (a) biological process, (b) molecular function and (c) cellular component.

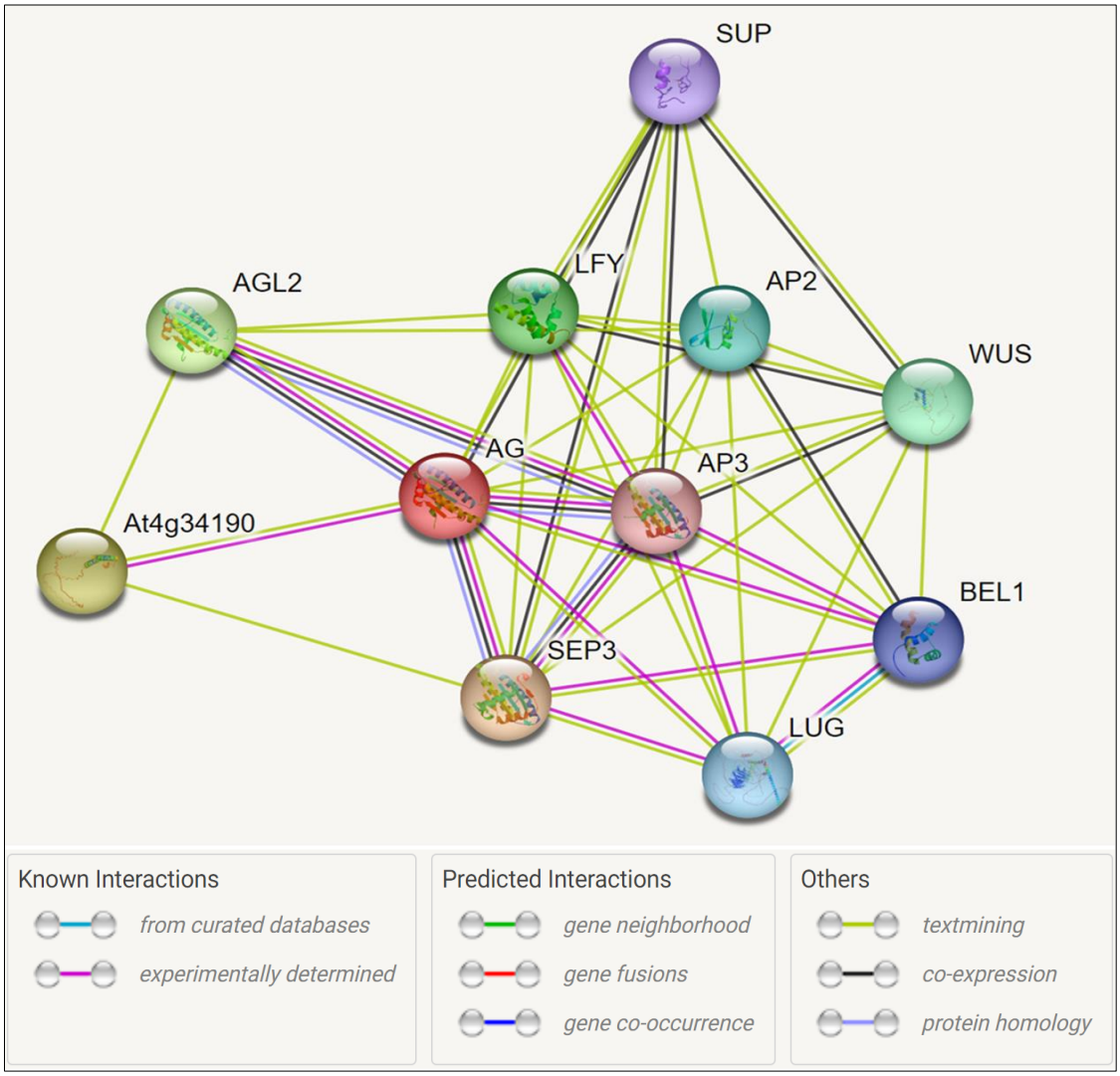

Figure 3.12: Protein interaction involving the protein AGAMOUS (red node) obtained from the STRING database.

Chapter Four: Discussion

Flowering is defined as the transition of plant tissue from the vegetative to reproductive phase. This process has been used many times to study the regulatory gene networks involved in floral initiation and morphogenesis in different plants (Dadpour et al. 2011; Mauracher and Wagner 2021; Rouet et al. 2022). *Clivia* is a genus of monocots that is regarded as horticulturally and medicinally important. The genus is endemic to southern Africa and comprises six species, but only one is popular (*Clivia miniata*) throughout the world. Five of the six species have flowers with a drooping (pendulous) morphology, whereas *C. miniata* displays open and upright (non-pendulous) flowers. However, flower architecture in *Clivia* species has not received much attention and this knowledge gap should be addressed to boost the use of clivia in the horticulture market. The aims of this study were to sequence, assemble and characterize the *Clivia* transcriptome, identify candidate genes responsible for the variation in *Clivia* flower shape (pendulous and non-pendulous); and lastly, identify morphogenesis genes that are differentially expressed between the different flower forms. As a non-model crop with limited genetic resources, the *Clivia* transcriptome was first sequenced and assembled *de novo*. Then, after validating the quality and completeness of the assembly, the transcripts were functionally annotated and analysed for global and differential expression in the pendulous and non-pendulous flower samples. Overall, many transcripts were identified, including several MADS-box genes that may possibly explain the variation in flower shape of *Clivia* species.

It is essential that necessary measures are taken during RNA extraction procedures to ensure good quality RNA isolation. This is because compromised RNA quality could effectuate unreliable findings in gene expression studies (Vermeulen et al. 2011). The Trizol reagent method is an inexpensive method that does not require complicated tasks. Wang et al. (2022) compared four RNA extraction methods using lilies (*Lilium davidii*) and concluded that the Trizol method was the most effective compared to other reagents. To improve RNA extraction, Roy et al. (2020) added an extra ethanol purification step that resulted in consistent yield and purity. Dutta et al. (2020) used a

modified Trizol method to combat co-precipitation of carbohydrates in starch-rich maize. Overall, Trizol reagent offers the better method for RNA extraction because it maintains RNA integrity by efficiently hindering RNase activity during homogenization, even in samples with high secondary metabolites such as passion fruit (*Passiflora edulis*) leaves (Luz et al. 2016). In this study, RNA extraction was conducted on nine RNA samples using Trizol and agarose gel electrophoresis. The samples with the best absorbance ratios and high concentrations were used (Fig. 3.1). The 260/280 and 260/230 absorbance ratios averaged 2.01 and 2.2 respectively. These results indicate that the RNA was pure and relatively free from protein contamination or reagent carry-overs. Gel electrophoresis revealed clear, intact bands corresponding to 18S and 25S leading to the conclusion that the RNA had no obvious degradation and was, therefore, good quality suitable for subsequent analyses.

Illumina sequencing in the current study yielded a total of ~150 million raw reads. Only a small portion of the reads were of bad quality. For most studies, the bad quality reads would account for 4% to 33% of the total raw reads (Singh and Jain 2014; Lee et al. 2016; Yue et al. 2018; Gul et al. 2019; Xu et al. 2022). That only 3% of the Illumina sequencing results from the current study were flagged as poor quality is most likely due to the good quality RNA obtained from the *Clivia* flower tissue. After read trimming, the read lengths ranged between 10 - 121 bp. Longer reads are necessary as they permit better sequence overlap and mapping with reduced repetition but, may not necessarily improve sequence assembly (Ekblom and Wolf 2014). According to Chang et al. (2014) and Chhangawala et al. (2015), the length of PE reads does not affect the completeness of an assembly or expression analyses once a threshold of about >50 bp has been reached. Therefore we can be confident in the findings or the current study.

De novo transcriptome assembly is a process used to construct an entire transcriptome from short read sequences without a reference genome (Grabherr et al. 2011). The transcriptome was assembled *de novo* using the Trinity package and the SPAdes assembler since the predicted 19 Gbp *Clivia* genome (Maleka 2020) is not yet available. Similarly, Wang et al. (2018) also assembled the

Clivia leaf transcriptome *de novo* due to the lack of the genome sequence. In general, the Trinity assembly algorithms tend to perform better than many other *de novo* assemblers (Cabau 2017). However, in the current study, SPAdes yielded better assembly results. This is due to the use of two longer kmers (43 and 63) in the assembly process compared to one kmer (25) used by Trinity. Unlike Trinity that uses a default kmer of 25 (Haas et. al 2013), SPAdes selects suitable multi-kmers based on input read length (Bankevich et. al 2012). Indeed, it is important to use different software to compare the statistics of different parameters as this allows for evaluation and selection of the best outcomes (Gurevich et al. 2013). Sequence assembly herein was done to represent each bud sample as well as a large assembly that corresponded to all bud tissue samples. The large assembly was subsequently clustered based on similar sequences to eliminate duplication which reduced the total number of scaffolds to 531,668. Sequence clustering is known to aid the identification of genes active in similar biological processes (Roberts and Roalson 2017). Understandably, combining the assemblies should optimize transcript representation in the transcriptome (Chen et al. 2015; Sadat-Hosseini et al. 2020). The individual assemblies comprised 266,075 (MB), 353,907 (TB) and 343957 (IB) scaffolds, but the combined assembly had 531,668 scaffolds with a total length of 353,907,434 bp. The large numbers of scaffolds indicate that there are still many disjointed sequences in the assembly and these can be reduced to a realistic number by further sequencing.

The final *Clivia* transcriptome was analysed in different ways, and each method showed that the dataset is comprehensive and biologically meaningful. An overall 97.23% reads mapped back to the transcriptome. When duplicated alignments were discounted, the alignment rate was 89.25%, which is still quite high and acceptable. A number of studies (Zhang et al. 2014, Zeng et al. 2017, Paya-Milans et al. 2019, Alhinho et al. 2021) found that a range of 63 – 90% of reads could be mapped back to the genome uniquely. The Nx statistic was used to assess assembly quality using transcript (contig) lengths. Half of the assembled bases (N50 statistic) in the current study ranged between 857 and 1192. However, considering that the typical transcript length of an average gene in

a plant genome is ~1500 bp (Maleka 2020), then these values represent about 50 – 80% of the envisaged transcript lengths. To supplement the N50 statistic, Trinity applies the ExN50 statistic that takes transcript expression data into consideration (Hass et al. 2013). The ExN50 statistic of the *Clivia* transcriptome (Fig 4.2 and Table S1) peaked at Ex=91. In the current study, Ex=83 and transcript length of 2,599 bp showed results similar to those from Chen et al. (2019) where they predicted 23,473 transcripts in the transcriptome of *Clivia nobilis* leaves. In *A. thaliana*, 27,000 transcripts were identified (Arabidopsis Genome Initiative 2000). Therefore, the ExN50 of the combined assembly suggests that the current study may have detected transcripts corresponding to most genes in the *Clivia* genome.

BUSCO analyses revealed that the *Clivia* flower transcriptome contains most of the single-copy genes present in the Liliopsida lineage (Only 2.8% were missing). A high number of duplicated sequences were detected even after sequence clustering at a threshold level of 95% sequence identity. While these may relate to the limited sequence coverage leading to an incomplete transcriptome, it is also believed that *Clivia* may have been an allopolyploid (with four sets of chromosomes) that may have since lost some of its duplication (Prof. JJ Spies, pers. comm.). This genomic feature could attribute to the high duplication levels observed in the *Clivia* genome but awaits confirmation through genomic sequencing. Malli et al. (2019) conducted a BUSCO search on their *de novo* assembled *Lavandula angustifolia* transcriptome. The results revealed that 53% of the BUSCOs were duplicated and 78% of these duplicated BUSCOs represented real duplication events in the genome and not assembly errors. *Lanandula angustifolia* is reported to be a polyploid (Upson et al. 2004), and this scenario could also apply in *Clivia.*

Correlation matrices (Fig. S3: Appendix A) and PCA (Fig. 4.4) confirmed the biological relationships among the studied *Clivia* samples. Each sample within a tissue sample group appears to be a biological replicate. However, IB replicate three appears to be somewhat separated from the other two. This means that the sample may be genetically different from the other two replicates. BLASTx against the Swissprot and NR databases generally revealed plant subjects. The species

distribution results obtained from the Swissprot database are similar to those from Lu et al. (2014) where the transcriptome of *Litchi chinesis* was assembled *de novo and* Bai et al. (2011) who focused on the ash tree (*Fraxinus spp*.). The top hits resulted from plant genomes of either monocots or dicots. The NR database contains much more data compared to the Swissprot database, and thus, hits represented different species. These species, however, were mainly monocots including *Asparagus officinalis, Iris pallida, Elaeis guineensis, Ananas comosus* and *Dendrobium catenatum*. In the current study, the species *A. officinalis* matched the most *Clivia* sequence hits from the NR database. This is no surprise considering that both *Asparagus* and *Clivia* are perennial monocots. Annotation challenges included that, of the 560,000 sequences, about 62,000 and ~145,000 had hits against Swissprot and NR, respectively. Further, that *Clivia* is not a model plant means there is no reference genome available. Therefore, the clivia-specific sequences may not be annotated properly at this point. However, in the NR annotation, 46 *Clivia* hits were matched. Also, many sequences in the transcriptome were short (<100 bp) and could have been excluded during searches on the protein database. Nevertheless, the NR database was more informative.

Functional annotation of the Clivia proteins using GO and the PANTHER database revealed that the *Clivia* transcriptome encodes proteins that perform a number of different functions. Similar to the current study, Wang et al. (2021) found that in the floral transcriptome of *Solanum lycopersicum*, the terms “cell process”, “metabolic process”, “response to stimulation”, and “biological regulation” were significantly enriched in the category biological process. Liu at al. (2016) identified the GO terms “flower development” and “stamen development” in the category biological process of *Chrysanthemum morifolium*, GO terms that we also expected to identify in the current study. Analysis of the *Clivia* protein class categories at the PANTHER database revealed “RNA metabolism protein”, “RNA binding protein”, “transcriptional regulator” and “transporters” as the most popular. RNA metabolism plays an important role in several developmental process in *A thaliana*, including the induction of flowering (Kuhn et al. 2007). RNA-binding proteins, on the other hand, are known for their involvement in metabolism during plant development and growth

(Lee and Kang 2016). Transcriptional regulation of flower development genes is mostly controlled by TFs from the bHLH, TCP and MYB families that are involved at different developmental stages (Chen et al. 2018; Zhou et al. 2019). Transporters were among the top five enriched protein classes, and some are involved in development. As an example, Gill (2021) found that membrane transporters regulate plant architecture. In conclusion, the enriched protein classes play important roles in initiating flowering, flower development and architecture.

Transcripts in the *Clivia* transcriptome mapped to 436 biological pathways at the KEGG database, more pathways in comparison to most of the similar studies. Unigenes from *Chrysanthemum morifolium* mapped into 280 pathways (Liu et al. 2016), Liu et al. (2017) mapped transcripts of *Annona squamosa* into 124 pathways, whereas Zhang et al. (2014) predicted 123 pathways in *Litchi chinensis* and Lee et al. (2016) predicted 129 pathways in *Rafflesia cantleyi*. The circadian rhythm pathway was identified as the only one relating to plant flowering, and therefore, considered for further analysis in this study. Circadian rhythms are oscillations that regulate various metabolic and developmental processes within a 24-hour period (Marshall et al. 2023). The circadian pathway is related to the photoperiod pathway of flowering (Bunning 1936; Niwa et al. 2007; Liang et al. 2021). About 73% of the circadian pathway genes in *A. thaliana* were detected in the transcriptome of *Clivia*, including *CCA1*, *PPR7*, *TOC1*, *LHY*, *PIF3*, *GI*, *CO* and *FT*. This suggests that the photoperiod pathway is active during flowering in *Clivia* and provides a basis for further studies to investigate the photoperiod pathway and other flowering processes in *Clivia*. The circadian pathway is indirectly involved in flower development though the photoperiod pathway as the circadian clock controls *CO* expression through *GI* (*GI* regulates the expression of *CO* to hasten flowering). *CO* activates *FT*, a TF that activates *AP1* expression, and thus, promoting floral meristem genes (Williamson 2008). Abdul-Awal et al. (2020) reported that, in sorghum, *gi* mutants showed delayed flowering compared to the wild-type genotypes under both short day and long day photoperiod conditions. Mutations in *GI* and *CO* may cause a delay in flowering, and thus, an extended vegetative growth phase, which, in turn, may affect morphology (e.g., increased node

number) (Suarez-Lopez et al. 2001; Abdul-Awal et al. 2020). Further studies may look into the effects of delayed flowering on flower morphology in *Clivia*.

The *Clivia* transcriptome assembly contained transcript sequences that matchd 17 of the 46 FLOR-ID genes (Bouche et al. 2016). The identified genes included *AP1,* which is known as a meristem identity and floral induction gene (Gregis et al. 2009). Once flower development begins, MADS-box TFs commonly known as "organ identity genes" will control the tissue differentiation process (Villar et al. 2020). The *Clivia* transcriptome had several MADS-box TFs including *AP2, AP3, AG, PI, LEU* and *SEP1-4*. These genes are also referred to as the ABC-model of flower development genes (Almeida et al. 1997). *KNUCKLES* (*KNU)* is activated by *AG* to repress the expression of *WUS* (a stem cell maintenance gene) as soon as flower patterning is complete. *UTL1* is involved in the formation of extra petals (Fletcher 2001). Overall, the current study seems to have addressed the aims accordingly.

In order to acquire better understanding of the genes expressed in *Clivia* tissue, we analysed the expression patterns through three pairwise comparisons namely, IB vs TB, IB vs MB and MB vs TB. The heatmap (Fig. 3.9) presented two main clusters whereby MB and TB have opposite trends in terms of transcript expression. Most DEGs were found between the MB (upregulated) vs TB (downregulated) tissue comparison. This is likely due to the different flower morphologies that they are supposed to produce and the physical sizes of the studies samples in relation to flower development. Unexpectedly, the tissues IB and TB revealed many DEGs, possibly due to the differences that occurred in IB replicate three. It could be possible that IB replicate three may be genetically different from the replicates one and two. One of the most important enriched functions in this study was the "transcription regulator activity" and it correlates with the presence of TF families linked to flower development (Li et al. 2020). These include MYB, bHLH, ERF and MADS-box flower development-related TF families. This indicates that flower development TF families are active in the *Clivia* transcriptome to direct flower development.

The differential expression of genes between small buds versus mature buds that are preparing to flower involves various candidates that associate with different aspects of flower development and maturation (Singh et al. 2013). *SEPALLATA* genes, for example, encode TFs that are essential for the development of floral organs (Pelaz et al. 2000). Expression levels may differ in TB versus IB buds as preparation for flower development initiates (Smyth et al. 1990). In the current study three of the five *SEP* transcripts showed higher expression levels in IB than in TB. Specifically, transcript NODE_100197_length_1184_cov_1053.818912_g37053_i4 was highly expressed in IB (115.9) compared to TB (10.2) tissues. Another transcript (NODE_107884_length_1103_cov_104.919231_g49287_i0) showed expression levels of 221.5 in IB and 0.3 in TB, whereas NODE_129682_length_906_cov_52.002372_g64197_i0 showed 58.6 (IB) versus 37.4 (TB). The other transcripts had similar expression levels in both sizes. Regarding *APETALA* 1, all transcripts showed higher expression levels in IB compared to TB, excluding NODE_108500_length_1097_cov_46.030948_g49686_i0 (IB=33.3; TB=35). *AP1* is one of the floral organ identity genes and species stamen identity and thus is expected to be expressed more in larger buds (Villar et al. 2020). Another important gene is *AGAMOUS*, whose expression may be more pronounced in larger buds to indicate the formation of carpels and/or stamen within the flower (Jibran et al. 2017). In the current study, the *AGAMOUS* transcript (NODE_108375_length_1098_cov_206.778744_g3219_i3) was expressed higher in IB than in TB. However, the difference in expression levels may not be significant (IB=293.3; TB=257.6). It is possible that IB was still in the process of maturing, meaning, the expression levels may increase further.

One flower-related gene that stood out as being differentially expressed between the pendulous and non-pendulous flower shape is *AGAMOUS*. The organ identity gene, *AG,* is involved in the formation of reproductive organs (stamens and carpels) in the third whorl (Alvarez-Buylla et al. 2010). The *Clivia* transcriptome was screened for tissue specific genes that may be involved in the development of pendulous and non-pendulous forms. Sequence homology searches on the genes

showed very few flowering-related gene, with an *Arabidopsis* ortholog of the flower development-related protein MADS1/AGAMOUS homology 1 being the most noticeable. This TF was expressed in the pendulous flower sample (39.5, 0.000637, DESeq2) but not detected in the non-pendulous flower samples (0). Quantitative PCR (qPCR) analysis of this transcript should be able to clarify its differential expression and that of many other DEGs in the *Clivia* flower tissues.

Protein-protein interaction analysis showed that *AG* associates with some of the genes present in the ABC model of flower development from the FLOR-ID database (Fig. 3.12) and *Clivia* flower development genes (Table 3.6). These include *SEP3, AP3, AGL2, WUS, AP2, LFY, SUP, BEL1* and *LEU*. Zhang et al. (2022) studied grape morphology and found that MADS-box TFs were amongst the candidate genes potentially responsible for differences in floral morphology. This suggests that the same could be true for other flowering plants, including *Clivia*, whereby variations in flower morphology between pendulous and non-pendulous flowers could be attributed to the MADS1 proteinas. Overall, future studies should focus on the sequencing of this gene to confirm its influence on morphology in *Clivia* species.

Chapter Five: Conclusions and future recommendations

Molecular genetic studies have identified many genes and gene regulatory networks involved in flower development. Included in these networks are transcription factors that regulate the expression of target genes, signaling molecules and other regulatory components. While there is still much to learn about the underlying regulatory and genetic mechanism of flower morphogenesis, significant advances have been made in the last two decades. Floral regulatory genes and gene networks have been characterized in crops such as maize (Bolduc et al. 2012), model plants such as *A. thaliana* (Kanrar et al. 2008) and even in some non-model plants such as *Akebia trifoliata* (Liu et al. 2010). Different developmental pathways such as cell division, expansion and differentiation are activated at different stages of flower morphogenesis as genetic and molecular signals propagate through the system and looking into the transcriptome can provide information on these happenings in the cell.

Members of the genus *Clivia* are increasing in popularity among breeders, who target and increase the quantity and quality of desirable traits in these plants. Despite the wide attention towards the genus, very little is known regarding the molecular basis of flower development and morphogenesis. Elucidating these processes may offer insights towards altering flower morphology in *Clivia* for the benefit of consumers in the horticulture market.

The study was conducted to assemble and characterize the *Clivia* transcriptome as well as to determine candidate genes responsible for the different flower forms within the genus by identifying specifically differentially expressed morphogenesis genes between pendulous and non-pendulous morphologies. Flower development has been looked at from two perspectives by Creane (2018). One can look at modifications over time that give rise to variations in flower morphology. Alternatively, one can study how morphology is affected by physical and internal factors such as genes and the environment. We focused on the second perspective and carried out next-generation sequencing (NGS) and transcriptome analysis to try and unravel the effect of flower morphogenesis genes on the phenotype of a clivia flower. We successfully identified genes known to be involved in flower development in the *Clivia* transcriptome. These included genes such as *AP1-3, SEP1-4, AG,*

PI, LEU, LATE MERISTEM IDENTITY (LMI1), BRC1, KNU, WUS, ULT1-2, TFl1 and *STM*. Of these genes, MADS1 (an AG homolog) was upregulated in to the pendulous flower samples but not the non-pendulous samples, permitting the conclusion that this gene may be involved in the observed phenotypic variations.

Flower morphology is not only important in plant reproduction, species identification and evolution, it is an important factor for the ongoing success of horticulture and floriculture industries where anything new or unique is desirable. The floriculture industry has been identified by Coetzee and Hoffman (2017) as one of the industries with the highest opportunity to grow and contribute to the economy of South Africa. The applications of NGS and transcriptomics in flower morphogenesis also offers new possibilities for the accurate design of breeding strategies to produce desired traits and to improve horticultural value. An attempt to determine the genetic basis of an observed phenotype is known as forward genetics. This approach requires phenotypic variation (natural or experimental mutants) to be observable such as the pendulous and non-pendulous flower forms of *Clivia* specimens.

It came as a surprise that only a few flowering-related genes were detected and even less flower morphogenesis genes. This could be either due to the small sample size, and thus for, further studies, we recommend using more samples or due to the strict p-value and adjusted p-value cut-off that were set to ≤ 0.005. Further studies could use genome-wide functional genomics techniques combined with novel genome editing tools such as CRISPR/Cas9, which could conclude with certainty that the MADS-box gene *AGAMOUS-LIKE* is responsible for the morphological differences between pendulous and non-pendulous *Clivia* species. This can be extended to other plant species, including crop plants.

Chapter Six: References

Abdul-Awal S, Chen J, Xin Z and Harmon F. 2020. A sorghum gigantea mutant attenuates florigen gene expression and delays flowering time. Plant Direct 4(11): e00281.

Alexandre C and Hennig L. 2008. FLC or not FLC: The other side of vernalization. Journal of Experimental Botany 59: 1127-1135.

Alhinho A, Ramos M, Alves S, Rocheta M, Morais-Cecílio L, Gomes-Laranjo J, Sobral R and Costa M. 2021. The dynamics of flower development in *Castanea sativa* Mill. Plants 10(8): 1538.

Ali Z, Raza Q, Atif R, Aslam U, Ajmal M and Chung G. 2019. Genetic and molecular control of floral organ identity in cereal. International Joural of Molecular Sciences 20(11): 2743.

Almeida J, Rocheta M and Galego L. 1997. Genetic control of flower shape in *Antirrhinum majus.* Development 124: 1387-1392.

Alonso J, Stepanova A, Leisse T, Kim C, Chen H, Shinn P, Stevenson D, Zimmerman J, Barajas P, Cheuk R and Gadrinab C. 2003. Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. Science 301(5633): 653-657.

Altschul S, Gish W, Miller W, Myers E and Lipman D. 1990. Basic local alignment search tool. Journal of Molecular Biology 215(3): 403-410.

Alvarez-Buylla E, Azpeitia E, Barrio R, Benítez M and Padilla-Longoria P. 2010. From ABC genes to regulatory networks, epigenetic landscapes and flower morphogenesis: Making biological sense of theoretical approaches. Seminars in Cell and Developmental Biology 21(1): 108-117.

Anderson S. 2003. Foraging responses in the butterflies *Inachisio, Aglais urticae* (Nymphalidae), and *Gonepteryx rhamni* (Pieridae) to floral scents. Chemoecology 13: 1-11.

Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408(6814): 796-815.

Aubrey A. 2001. *Clivia miniata* [Lindl.] Regel. South African National Biodiversity Institute 1-5.

Austin I, Alonso-Blanco C and Mertinez-Zapater C. 2005. Environmental regulation of flowering. International Journal of Development Biology 49: 689-705.

Bai X, Rivera-Vega L, Mamidala P, Bonello P, Herms D and Mittapalli O. 2011. Transcriptomic signatures of ash (Fraxinus spp.) phloem. PloS ONE 6(1): e16368.

Bankevich A, Nurk S, Antipov D, Gurevich A, Dvorkin M, Kulikov A, Lesin V, Nikolenko S, Pham S, Prjibelski A and Pyshkin A. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Journal of Computational Biology 19(5): 455-477.

Bentley D, Balasubramanian S, Swerdlow H, Smith G, Milton J, Brown C, Hall K, Evers D, Barnes C, Bignell H and Boutell J. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456(7218): 53-59.

Bercu R, Bavaru A and Popoviciu R. 2017. Contribution to the anatomy of some angiosperms flower parts with therapeutic value. Annals of the Romanian Society for Cell Biology 21(2): 51-58.

Bernier G and Perilleux C. 2005. A physiological overview of the genetics of flowering time control. Plant Biotechnology Journal 3: 3-16.

Bian F, Yang X, Ou Z, Luo J, Tan B, Yuan M, Chen T and Yang R. 2019. Morphological characteristics and comparative transcriptome analysis of three different phenotypes of *Pristella maxillaris*. Frontiers In Genetics 10: 698.

Biggs M. 2021. Madame Ganna Walska. In Lessons from the Great Gardeners. University of Chicago Press 116-121.

Bjornson M, Kajala K, Zipfel C and Ding P. 2020. Low-cost and high-throughput RNA-seq library preparation for Illumina sequencing from plant tissue. Bio-protocol 10(20): 3799.

Bokulich N, Subramanian S, Faith J, Gevers D, Gordon J, Knight R, Mills D and Caporaso J. 2013. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nature Methods 10(1): 57-59.

Bolduc N, Yilmaz A, Mejia-Guerra MK, Morohashi K, O'Connor D, Grotewold E and Hake S. 2012. Unravelling the KNOTTED1 regulatory network in maize meristems. Genes and Development 15: 1685-1690.

Bouché F, Lobet G, Tocquin P and Périlleux C. 2016. FLOR-ID: an interactive database of flowering-time gene networks in *Arabidopsis thaliana*. Nucleic Acids Research 44(D1): 1167-1171.

Bowman J. 1997. Evolutionary conservation of angiosperm flower development at the molecular and genetic levels. Journal of Biosciences 22(4): 515-527.

Bowman J, Smyth D and Meyerowitz E. 1989. Genes directing flower development in *Arabidopsis*. The Plant Cell 1: 37-52.

Bryant D, Johnson K, DiTommaso T, Tickle T, Couger M, Payzin-Dogru D, Lee T, Leigh N, Kuo T, Davis F and Bateman J. 2017. A tissue-mapped axolotl de novo transcriptome enables identification of limb regeneration factors. Cell reports 18(3): 762-776.

Brautpigam A and Gowik U. 2010. What can next generation sequencing do for you? Next generation sequencing as a valuable tool in plant research. Plant Biology 12(6): 831-841.

Buning E. 1936. Die endogene Tagesrhythmik als Grundlage der photoperiodischen Reaktion. Berichte der Deutscher Botanischen. Gesellschaft 4: 590-607.

Cabau C, Escudié F, Djari A, Guiguen Y, Bobe J and Klopp C. 2017. Compacting and correcting Trinity and Oases RNA-Seq *de novo* assemblies. Peer-reviewed Journal 5: e2988.

Chaitankar V, Karakülah G, Ratnapriya R, Giuste F, Brooks M and Swaroop A. 2016. Next generation sequencing technology and genome wide data analysis: Perspectives for retinal research. Progress in Retinal and Eye Research 55: 1-31.

Chang Z, Wang Z and Li G. 2014. The impacts of read length and transcriptome complexity for *de novo* assembly: a simulation study. PloS One 9(4): e94825.

Chhangawala S, Rudy G, Mason C and Rosenfeld J. 2015. The impact of read length on quantification of differentially expressed genes and splice junction detection. Genome Biology 16(1): 1-10.

Chen D, Yan W, Fu L and Kaufmann K. 2018. Architecture of gene regulatory networks controlling flower development in *Arabidopsis thaliana*. Nature Communications 9(1): 4534.

Chen S, McElroy J, Dane F and Peatman E. 2015. Optimizing transcriptome assemblies for *Eleusine indica* leaf and seedling by combining multiple assemblies from three *de novo* assemblers. The Plant Genome 8(1): 10.

Chen Z. 2007. Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. Annual Review of Plant Biology 58: 377-406.

Cheng J, Zhou Y, Lv T, Xie C and Tian C. 2017. Research progress on the autonomous flowering time pathway in *Arabidopsis*. Physiology and Molecular Biology of Plants 3: 477-485.

Clivia websites accessed on 12 April 2021. https://www.cliviape.co.za/listings/cliva-species-information and https://www.cliviape.co.za/galleries/all

Coen E and Meyerowitz E. 1991. The war of the whorls: genetic interactions controlling flower development. Nature 353(6339): 7-31.

Coetzee J and Hoffman E. 2017. Social and environmental responsibility of the floriculture industry in South Africa. InVII International Conference on Managing Quality in Chains (MQUIC2017) and II International Symposium on Ornamentals. 479-484.

Colombo L, Busscher M, Franken J and Angenent G. 1996. MADS box genes controlling flower development in *Petunia*. Flowering Newsletter 22: 13-19.

Conrad F, Reeves G and Rourke J. 2003. Phylogenetic relationships of the recently discovered species- *Clivia mirabilis*. South African Journal of Botany 69: 204-206.

Conrad F. 2008. Molecular systematics, biogeography and dating of the tribe Haemantheae (Amaryllidaceae) and the phylogeography of *Clivia*, University of Cape Town, Cape Town, South Africa.

Cookey G, Tambari B and Iboroma D. 2018. Evaluation of the corrosion inhibition potentials of green-tip forest lily (*Clivia nobilis*) leaves extract on mild steel in acid media. Journal of Applied Sciences and Environmental Management 22(1): 4-90.

Craene R. 2018. Understanding the role of floral development in the evolution of angiosperm flowers: clarifications from a historical and physio-dynamic perspective. Journal of Plant Research 131(3): 367-393.

Cruz J, San M, Lubini G, Strini E, Sobral R, Pinoti V, Ferreira P, Thome V, Quiapim A, Dornelas M, Pranchevicius M, Madueno F, Costa M and Goldman M. 2021. SCI1 is a direct target of AGAMOUS and WUSCHEL and is specifically expressed in the floral meristematic cells. Frontiers in Plant Science 12: 399.

Dadpour M, Movafeghi A, Grigorian W and Omidi Y. 2011. Determination of floral initiation in *Malus domestica*: a novel morphogenetic approach. Biologia Plantarum 55: 243-252.

De Smedt V, van Huylenbroeck J and Debergh P. 1996. Influence of temperature and supplementary lighting on growth and flower initiation of *Clivia miniata* Regel. Scientia Horticulturae 65: 65-72.

Deng W, Ying H, Helliwell C, Taylor J, Peacock W and Dennis E. 2011. Flowering Locus C (FLC) regulates development pathways throughout the life cycle of *Arabidopsis*. Plant Biology 108(16): 6680-6685.

Dinneny J, Weigel D and Yanofsky M. 2005. A genetic framework for fruit patterning in *Arabidopsis thaliana*. Development 132: 4687-4696.

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

Dixon R. 2011. Relationships in *Clivia*. Researchgate.net 63-81.

Dornelas M and Dornelas O. 2005. From leaf to flower: revising Goethe's concepts on "The metamorphosis of plants" (1790). Brazilian Journal of Plant Physiology 17: 4.

Dubois A, Raymond O, Maene M, Baudino S and Langlade N. 2010. Tinkering with the C-Function: A molecular frame for the selection of double flowers in cultivated roses. PLoS One 5(2): e9288.

Duncan G and Du Plessis N. 1989. Bulbous plants of southern Africa. Tafelberg Publishers 130-133.

Duncan G. 1999. Grow Clivias. Kirstenbosch Gardening Services National Botanical Institute. Cape Town.

Duncan G. 2004. Growing the miracle *Clivia*: feature. Veld and Flora 90(4): 9-148.

Dutta S, Muthusamy V, Chhabra R, Zunjare R and Hossain F. 2020. Two-step method for isolation of high-quality RNA from stored seeds of maize rich in starch. 3 Biotech 10:433.

Duttke S, Zoulias N and Kim M. 2012. Mutant flower morphologies in the wind orchid, a novel orchid model species. Plant Physiology 158(4): 1542-1547.

Dyer R. 1943. *Clivia caulescens*. The flowering plants of South Africa 23: 891.

Ekblom R and Wolf J. 2014. A field guide to whole-genome sequencing, assembly and annotation. Evolutionary applications 7(9): 1026-1042.

El-Bairi K, Azzam F and Amrani M. 2020. The arrival of Next-Generation Sequencing: An overview of current technologies. Illuminating Colorectal Cancer Genomics by Next-Generation Sequencing 73-89.

Eveland A, Satoh-Nagasawa N, Goldshmidt A, Meyer S, Beatty M, Sakai H, Ware D and Jackson D. 2010. Digital gene expression signatures for maize development. Plant Physiology 154(3): 1024-1039.

Favaro R, Pinyopich A, Battaglia R, Kooiker M, Borghi L and Ditta G. 2003. MADS-Box protein complexes control carpel and ovule development in *Arabidopsis.* Plant Cell 15: 2603-2611.

Fernandez D, Wang C, Zheng Y, Adamczyk B, Singhal R and Hall P. 2014. The MADS-Domain factors AGAMOUS-LIKE15 and AGAMOUS-LIKE18, along with SHORT VEGETATIVE PHASE and AGAMOUS-LIKE24, are necessary to block floral gene expression during the vegetative phase. Plant Physiology 165(4): 1591-1603.

Fletcher J. 2001. The ULTRAPETALA gene controls shoot and floral meristem size in *Arabidopsis*. Development 8(8): 1323-1333.

Fornara F, Marziani G, Mizzi L, Kater M and Colombo L. 2003. MADS-box genes controlling flower development in rice. Plant Biology 1:16-22.

Fu L, Niu B, Zhu Z, Wu S and Li W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28(23): 3150-3152.

Funnell K. 2016. Scheduling flowering in *Clivia miniata* Regel for different markets. International Symposium on Flower Bulbs and Herbaceous Perennials 1171: 39-46.

Gagiano A, 2006. Genetic variation in *Clivia miniata* var. *citrina*. MSc. Thesis. University of the Free State, Bloemfontein.

Garnica D, Upadhyaya N, Dodds P and Rathjen J. 2013. Strategies for wheat stripe rust pathogenicity identified by transcriptome sequencing. PLoS One 8: 67150.

Gattolin S, Cirilli M, Chessa S, Stella A, Bassi D and Laura-Rossini L, 2020. Mutations in orthologous PETALOSA TOE-type genes cause a dominant double-flower phenotype in phylogenetically distant eudicots. Journal of Experimental Botany 71(9): 2585-2595.

Ge S, Jung D and Yao R. 2020. ShinyGO: a graphical gene-set enrichment tool for animals and plants. Bioinformatics 36(8): 2628-2629.

Geniza M and Jaiswal P. Tools for building de novo transcriptome assembly. Current Plant Biology 11: 41-45.

Ghaemizadeh F, Dashti F and Shafeinia A. 2019. Expression pattern of ABCDE model genes in floral organs of bolting garlic clone. Gene Expression Patterns 34: 119059.

Gill R, Ahmar S, Ali B, Saleem M, Khan M, Zhou W and Liu S. 2021. The role of membrane transporters in plant growth and development, and abiotic stress tolerance. Internal Journal of Molecular Science 22(23): 12792.

Github website accessed 30 April 2022 on https://github.com/trinityrnaseq/trinityrnaseq/wiki/Transcriptome-Assembly-Quality-Assessment.

Grabherr M, Haas B, Yassour M, Levin J, Thompson D, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q and Chen Z. 2011. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nature biotechnology 29(7): 644.

Growers C. 2017. Trees, shrubs, aloes, grasses and ground covers. KwaZulu-Natal, South Africa. Accessed on 21 Sep 2022 at https://cjmgrowers. co.za/clivia-miniata/

Gul H, Tong Z, Han X, Nawaz I, Wahocho SA, Khan S, Zhang C, Tian Y, Cong P and Zhang L. 2019. Comparative transcriptome analysis between ornamental apple species provides insights into mechanism of double flowering. Agronomy 9(3): 112.

Gurevich A, Saveliev V, Vyahhi N and Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 9(8): 1072-1075.

Hass B, Papanicolaou A, Yassour M, Grabherr M, Blood P and Bowden J. 2013. *De novo* transcript sequence reconstruction from RNA Seq: reference generation and analysis with Trinity. Nature Protocol 8(8): 1494-1512.

Hamilton J and Robin B. 2012. Advances in plant genome sequencing. The Plant Journal 70: 177-190.

Helliwell C, Wood C, Robertson, Peacock W and Dennis E. 2006. The *Arabidopsis* FLC protein interacts directly *in vivo* with SOC1 and FT chromatin and is part of a high-molecular-weight protein complex. The Plant Journal 46: 183-192.

Hess J, Kohl T, Kotrova M, Ronsch K, Paprotka T, Mohr V, Hutzenlaub T, Bruggemann M, Zengerle R, Niemann S and Paust N. 2020. Library preparation for next generation sequencing: A review of automation strategies. Biotechnology Advances 41: 107537.

Himani A, Ramkumar T, Tyagi S, Sharma H, Upadhyay S and Sembi J. 2019. Tracing the footprints of the ABCDE model of flowering in *Phalaenopsis equestris* (Schauer) (Orchidaceae). Journal of Plant Biotechnology 46: 255-273.

Hirasawa Y, Tanaka T, Hirasawa S, Wong CP, Uchiyama N, Kaneda T, Goda Y and Morita H. 2022. Cliniatines A–C, new Amaryllidaceae alkaloids from *Clivia miniata*, inhibiting Acetylcholinesterase. Journal of Natural Medicines 76: 7-171.

Hong M, Tao S, Zhang L and Huang X. 2020. RNA sequencing: new technologies and applications in cancer research. Journal of Haematology and Oncology 13: 166.

Honiball C and Robbertse P. 2001. Low temperature treatment advances flowering in *Clivia miniata* (Lindley) Regel. South African Journal of Plant and Soil 18(4): 70-169.

Hooker W. 1856. *Clivia gardenii*. Curtis's Botanical Magazine series III 12: 4895.

Hu J, Chang X, Zhang Y, Yu X, Qin Y, Sun Y and Zhang L. 2021. The pineapple MADS-box gene family and the evolution of early monocot flower. Scientific reports 11: 1-4.

Hu J, Lan M, Xu X, Yang H, Zhang L, Lv F, Yang H, Yang D, Li C and He J. 2021. Transcriptome profiling reveals molecular changes during flower development between male sterile and fertile Chinese cabbage (*Brassica rapa* ssp. pekinensis) lines. Life 11(6): 525.

Huang B, Rong H, Ye Y, Ni Z, Xu M, Zhang W and Xu L. 2020. Transcriptomic analysis of flower colour variation in the ornamental crabapple (*Malus spp.)* half-sib family through Illumina and PacBio Sequel sequencing. Plant Physiology and Biochemistry 149: 27-35.

Huang N, Tien H and Yu T. 2020. The movement of a leaf derived mobile AGL24 mRNA specifies floral organ identity in *Arabidopsis*. bioRxiv.

Irish V. 2017. The ABC model of floral development. Current Biology 27(17): 887-890.

Jibran R, Tahir J, Cooney J, Hunter D and Dijkwel P. 2017. *Arabidopsis* AGAMOUS regulates sepal senescence by driving jasmonate production. Frontiers in Plant Science 8: 2101.

Johnson K and Lenhard M. 2011. Genetic control of plant organ growth. New Phytologist 191: 319-333.

Kanrar S, Bhattacharya M, Arthur B, Courtier J and Smith H. 2008. Regulatory networks that function to specify flower meristems require the function of homeobox genes PENNYWISE and POUND-FOOLISH in *Arabidopsis*. The Plant Journal 54(5): 924-937.

Kaul S, Koo H, Jenkins J, Rizzo M, Rooney T, Tallon L, Feldblyum T, Nierman W, Benito M, Lin X and Town C. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408(6814): 796-815.

Kiepiel I and Johnson S. 2014. Shift from bird to butterfly pollination in *Clivia* (Amaryllidaceae). American Journal of Botany 101: 190-200.

Kiepiel I and Johnson S. 2021. Responses of butterflies to visual and olfactory signals of flowers of the bush lily *Clivia miniata*. Arthropod-Plant Interactions 15: 253-263.

Kinoshita A, Vayssières A, Richter R, Sang Q, Roggen A, Van Driel A, Smith R and Coupland G. 2020. Regulation of shoot meristem shape by photoperiodic signalling and phytohormones during floral induction of *Arabidopsis*. eLife 9: 1-29.

Knief C. 2014. Analysis of plant microbe interactions in the era of next generation sequencing technologies. Frontiers in Plant Science 5: 216.

Koopowitz H. 2002. Clivias. Timber Press.

Koorneef M and Somerville C. 2002. A fortunate choice: the history of *Arabidopsis* as a model plant. Nature Reviews Genetics 3: 883-889.

Krishnamurphy K and Bahadur B. 2015. Genetics of flower development. Plant biology and biotechnology: Plant Diversity, Organization, Function and Improvement 385-407.

Kuhn J, Breton G and Schroeder J. 2007. mRNA metabolism of flowering-time regulators in wild-type *Arabidopsis* revealed by a nuclear cap binding protein mutant, abh1. Plant Journal 50(6): 1049-1062.

Langmead B and Salzberg S. 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods 9(4): 357-359.

Layton C, McMahon P and Greenleaf W. 2019. Large scale, quantitative protein assays on a high-throughput DNA sequencing chip. Molecular Cell 73(5): 1075-1082.

Lee K and Kang H. 2016. Emerging roles of RNA-binding proteins in plant growth, development, and stress responses. Molecular Cells 39(3): 179-185.

Lee X, Mat-Isa M, Mohd-Elias N, Aizat-Juhari M, Goh H, Dear P, Chow K, Haji-Adam J, Mohamed R, Firdaus-Raih M and Wan K. 2016. Perigone lobe transcriptome analysis provides insights into *Rafflesia cantleyi* flower development. PloS ONE 11(12): e0167958.

Lenhard M, Bohnert A, Jurgens G and Laux T. 2001. Termination of stem cell maintenance in *Arabidopsis* floral meristems by interactions between WUSCHEL and AGAMOUS. Cell 105: 805-814.

Li M, Tan S, Tan G, Luo Y, Sun B, Zhang Y, Chen Q, Wang Y, Zhang F, Zhang Y and Lin Y. 2020. Transcriptome analysis reveals important transcription factor families and reproductive biological processes of flower development in Celery (*Apium graveolens* L.). Agronomy 10(5): 653.

Li W and Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22(13): 1658-1659.

Li W, Zhang L, Ding Z, Wang G, Zhang Y, Gong H, Chang T and Zhang Y. 2017. *De novo* sequencing and comparative transcriptome analysis of the male and hermaphroditic flowers provide insights into the regulation of flower formation in andromonoecious *Taihangia rupestris.* BMC Plant Biology 17: 54.

Li Y, Gao R, Zhang J, Wang Y, Kong P, Lu K, Adnan A, Liu M, Ao F, Zhao C and Wang L. 2022. The biochemical and molecular investigation of flower colour and scent sheds lights on further genetic modification of ornamental traits in *Clivia miniata.* Horticulture Research.

Liang L, Zhang Z, Cheng N, Liu H, Song S, Hu Y, Zhou X, Zhang J and Xing Y. 2021. The transcriptional repressor OsPRR73 links circadian clock and photoperiod pathway to control heading date in rice. Plant, Cell and Environment 44(3):842-855.

Lin W, Gupta S, Arazi T and Spitzer-Rimon B. 2021. MIR172d is required for floral organ identity and number in tomato. International Journal of Molecular Sciences 22(9): 4659.

Lindley J. 1828. *Clivia nobilis*. Edwards's botanical register 14: 1182.

Liu C, Zhang J, Zhang N, Shan H, Su K, Zhang J, Meng Z, Kong H and Chen Z. 2010. Interactions among proteins of floral MADS-box genes in basal eudicots: implications for evolution of the regulatory network for flower development. Molecular Biology and Evolution 27(7): 1598-1611.

Liu H, Sun M, Du D, Pan H, Cheng T, Wang J, Zhang Q and Gao Y. 2016. Whole-transcriptome analysis of differentially expressed genes in the ray florets and disc florets of *Chrysanthemum morifolium*. BMC Genomics 17(1): 1-7.

Liu K, Li H, Li W, Zhong J, Chen, Shen C and Yuan C. 2017. Comparative transcriptomic analyses of normal and malformed flowers in sugar apple (*Annona squamosa*) to identify the differential expressed genes between normal and malformed flowers. BMC Plant Biology 17: 1-3.

Liu L, Li X, Jiang J, Nie Q, Zheng T, Wan W, Hsiang T, Sun Z and Zhou Y. 2020. First report of *Fusarium oxysporum* causing basal stem rot of *Clivia miniata* in China. Plant Disease 104(5): 1561.

Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L and Law M. 2012. Comparison of Next-Generation Sequencing Systems. Journal of Biomedicine and Biotechnology.

Lohmann J, Hong R, Hobe M, Busch M, Parcy F, Simon R and Weigel D. 2001. A molecular link between stem cell regulation and floral patterning in *Arabidopsis*. Cell 105: 793-803.

Love M, Huber W and Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15: 550.

Lu X, Kim H, Zhong S, Chen H, Hu Z and Zhou B. *De novo* transcriptome assembly for rudimentary leaves in *Litchi chinesis* Sonn. and identification of differentially expressed genes in response to reactive oxygen species. BMC Genomics 15(1): 1-4.

Luz A, Pretti I and Batitucci M. 2016. Comparison of RNA extraction methods for *Passiflora edulis* sims leaves. Revista Brasileira de Fruticultura 38: 226-232.

Ma J, Shen X, Liu Z, Zhang D, Liu W, Liang H, Wang Y, He Z and Chen F. 2018. Isolation and characterization of AGAMOUS-Like genes associated with double-flower morphogenesis in *Kerria japonica* (Rosaceae). Frontiers in Plant Science 9: 959.

Ma Q, Zhang W and Xiang Q. 2017. Evolution and developmental genetics of floral display-A review of progress. Journal of Systematics and Evolution 55(6): 487-491.

Malli R, Adal A, Sarker L, Liang P and Mahmoud S. 2019. De novo sequencing of the *Lavandula angustifolia* genome reveals highly duplicated and optimized features for essential oil production. Planta 249: 251-256.

Mandel M and Yanofsky M. 1995. Unravelling the mystery of double flowers. Plant Cell 7: 1763-1771.

Manni M, Berkeley M, Seppey M, Simão F and Zdobnov E. 2021. BUSCO Update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic and viral genomes. Molecular Biology and Evolution 38(10): 4647-4654.

Manning J. 2005. Pollination in *Clivia*. Clivia 7: 17-22.

Marks R, Hotaling S, Frandsen P and VanBuren R. 2021. Representation and participation across 20 years of plant genome sequencing. Nature Plants 7(12): 1571-1578.

Marshall C, Thompson V, Creux N and Harmer S. 2023. The circadian clock controls temporal and spatial patterns of floral development in sunflower. Elife 12: e80984.

Matimati I. 2017. Exploring the moisture sources for *Clivia mirabilis*. Veld and Flora 103(2): 55-57.

Mauracher S and Wagner J. 2021. Flower preformation in the nival plant *Ranunculus glacialis* L.: shoot architecture and impact of the growing season length on floral morphogenesis and developmental dynamics. Alpine Botany 131: 1-12.

McCarthy E, Mohamed A and Litt A. 2015. Functional divergence of APETALA1 and FRUITFULL is due to changes in both regulation and coding sequence. Frontiers in Plant Science 6: 1076.

Meyerowitz E, Smyth D and Bowman J. 1989. Abnormal flowers and pattern formation in floral development (Review). Development 106: 209-217.

Meyerowitz E. 2001. Prehistory and history of *Arabidopsis* research. Plant Physiology 125: 15-19.

Mi H, Muruganujan A, Casagrande J and Thomas P. 2013. Large-scale gene function analysis with the PANTHER classification system. Nature Protocols 8(8): 1551-1566.

Michaels S and Amasino R. 1999. FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. Plant Cell 11: 949-956.

Middleton L. 2015. South African consumers' selection criteria for ornamental plants: a market perspective. South African Journal of Plant and Soil 32(4): 253-255.

Mohrholz A, Sun H, Glöckner N, Hummel S, Kolukisaoglu Ü, Schneeberger K and Harter K. 2019. The striking flower-in-flower phenotype of *Arabidopsis thaliana* nossen (No-0) is caused by a novel leafy allele. Plants 8(12): 599.

Mondal T, Rawal H, Chowrasia S, Varshney D, Panda A, Mazumdar A, Kaur H, Gaikwad K, Sharma T and Singh N. 2018. Draft genome sequence of first monocot-halophytic species *Oryza coarctata* reveals stress-specific genes. Scientific Reports 8: 13698.

Montero-Vargas M, Escudero-Leyva E, Díaz-Valerio S and Chaverri P. 2020. Step-by-step pipeline for the ecological analysis of endophytic fungi using ITS nrDNA data. Current Protocols in Microbiology 56(1): 96.

Moreau F, Thevenon E, Blanvillain R, Lopez-Vidriero I, Franco-Zorrilla J, Dumas R, Parcy F, Morel P, Trehin C and Carles C. 2016. The Myb-domain protein ULTRAPATELA interacting factor 1 controls floral meristem activities in *Arabidopsis*. Development 143(7): 1108-1119.

Moyroud E and Glover J. 2017. The evolution of diverse floral morphologies. Current Biology 27: 941.

Murray B, Ran Y, De Lange P, Hammett K, Truter J and Swanevelder Z. 2004. A new species of *Clivia* (Amaryllidaceae) endemic to the Pondoland Centre of Endemism, South Africa. Botanical Journal of the Linnaean Society 146(3): 369-374.

Murray B, Wong C, Hammett K and Keith R. 2011. The karyotype of *Clivia mirabilis* analysed by differential banding and fluorescence in-situ hybridization. Plant Systematics and Evolution 293: 193-196.

Musara C, Aladejana E and Aladejana A. 2021. *Clivia miniata* (Lindl.) Bosse, (Amaryllidaceae): Botany, medicinal uses, phytochemistry and pharmacological properties. Journal of Applied Pharmaceutical Science 11(2): 2-18.

Naiker E. 2019. Prize-winning clivia could attract a hefty R25K price tag. Northern Natal News accessed in March 2023. (https://northernnatalnews.co.za/242181/prize-winning-clivia-attract-hefty-r25k-price-tag/)

Nashima K, Shirasawa K, Ghelfi A, Hirakawa H, Isobe S, Suyama T, Wada T, Kurokura T, Uemachi T, Azuma M, Akutsu M, Kodama M, Nakazawa Y and Namai K. 2021. Genome sequence of *Hydrangea macrophylla* and its application in analysis of the double flower phenotype, DNA Research 28: dsaa026.

Nelson A and Stewart C. 2020. Microbiota analysis using sequencing by synthesis: From library preparation to sequencing. Innate Lymphoid Cells 165-184.

Ng M and Yanofsky M. 2001. Function and evolution of the plant MADS-box gene family. Nature Reviews Genetics 3: 186-195.

Nitta K, Yasumoto A and Yahara T. 2010. Variation of flower opening and closing times in F1 and F2 hybrids of daylily (*Hemerocallis fulva*; Hemerocallidaceae) and nightlily (*H. citrina*). American Journal of Botany 97(2): 261-267.

Niwa Y, Ito S, Nakamichi N, Mizoguchi T, Niinuma K, Yamashino T and Mizuno T. 2007. Genetic linkages of the circadian clock-associated genes, *TOC1*, *CCA1* and *LHY*, in the photoperiodic control of flowering time in *Arabidopsis thaliana*. Plant and Cell Physiology 48(7): 925-937.

Omura H, Honda K and Hayashi N. 1999. Chemical and chromatic bases for preferential visiting by the cabbage butterfly, *Pieris rapae*, to rape flowers. Journal of Chemical Ecology 25: 1895-1906.

Onda Y and Mochida K. 2016. Exploring genetic diversity in plants using high-throughput sequencing techniques. Current Genomics 17(4): 358-367.

Pacific Bulb Society website accessed in November 2022 (https://www.pacificbulbsociety.org/pbswiki/index.php/Clivia).

Papenfuss H, Kumari A, Kulkarni M, Finnie J and Van Staden J. 2014. Smoke-water enhances *in vitro* pollen germination and tube elongation of three species of Amaryllidaceae. South African Journal of Botany 90: 87-92.

Patro R, Duggal G and Kingsford C. 2015. Salmon: accurate, versatile and ultrafast quantification from RNA-seq data using lightweight-alignment. BioRxiv 10: 021592.

Payá-Milans M, Poza-Viejo L, Martín-Uriz P, Lara-Astiaso D, Wilkinson M and Crevillén P. 2019. Genome-wide analysis of the H3K27me3 epigenome and transcriptome in *Brassica rapa*. GigaScience 8(12): 147.

Pelaz S, Ditta G, Baumann E, Wisman E and Yanofsky MF. 2000. B and C floral organ identity functions require SEPALLATA MADS-box genes. Nature 405(6783): 200-203.

Pelaz S, Tapia-López R, Alvarez-Buylla E and Yanofsky M. 2001. Conversion of leaves into petals in *Arabidopsis*. Current Biology 11(3): 182-184.

Pessoa D, Dos-Santos C, Vidal M, Baldani J, Tadra-Sfeir M, de Souza E and Simoes-Araujo J. 2021. *Herbaspirillum seropedicae* strain HRC54 expression profile in response to sugarcane apoplastic fluid. 3 Biotech 11(6): 1-3.

Ran Y, Hammett X and Murray B. 2001. Phylogenetic analysis and karyotype evolution in the genus *Clivia* (Amaryllidacea). Annals of Botany 87(6): 823-830.

Refahi Y, Traas J, Jonsson H and Zardilis A. 2021. Research data supporting "A multiscale analysis of early flower development in *Arabidopsis* provides an integrated view of molecular regulation and growth control". Developmental Call 56(4): 540-556.

Regel E. 1864. *Clivia miniata* Lindl. Amaryllideae. Gartenflora 14: 131-434.

Reid M. 2005. Flower development: From bud to bloom. Acta Horticulturae 105-110.

Roberts W and Roalson H. 2017. Comparative transcriptome analyses of flower development in four species of *Achimenes* (Gesneriaceae). BMC Genomics 18: 240.

Robinson M, McCarthy D and Smyth G. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26: 139-140.

Rothberg J, Hinz W, Rearick T, Schultz J, Mileski W, Davey M, Leamon J, Johnson K, Milgrew M, Edwards M and Hoon J. 2011. An integrated semiconductor device enabling non-optical genome sequencing. Nature 475(7356): 348-352.

Rouet S, Durand JL, Leclercq D, Bernicot MH, Combes D Escobar-Gutiérrez A and Barillot R. 2022. L-GrassF: a functional–structural and phenological model of *Lolium perenne* integrating plant morphogenesis and reproductive development *in silico*. Plants 4(2): diac012.

Rourke J. 2002. *Clivia mirabilis* (Amaryllidaceae: Haemantheae) a new species from Northern Cape, South Africa. Bothalia 32: 1-7.

Roy D, Tomo S, Modi A, Purohit P and Sharma P. 2020. Optimising total RNA quality and quantity by phenol-chloroform extraction method from human visceral adipose tissue: A standardisation study. MethodsX 7:101113.

Sadat-Hosseini M, Bakhtiarizadeh M, Boroomand N, Tohidfar M and Vahdati K. 2020. Combining independent *de novo* assemblies to optimize leaf transcriptome of Persian walnut. PloS One (4): e0232005.

Salimonti A, Forgione I, Sirangelo T, Puccio G, Mauceri A, Mercati F, Sunseri F and Carbone F. 2021. A complex gene network mediated by ethylene signal transduction TFs defines the flower induction and differentiation in *Olea europaea*. Genes 12(4): 545

SANBI website accessed in November 2022 (http://redlist.sanbi.org/genus.php?genus=2018).

Sanger F, Nicklen S and Coulson A. 1977. DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences 74(12): 5463-5467.

Sasaki K, Mitsuda N, Nashima K, Kishimoto K, Katayose Y, Kanamori H and Ohmiya A. 2017. Generation of expressed sequence tags for discovery of genes responsible for floral traits of *Chrysanthemum morifolium* by Next-Generation Sequencing technology. BMC Genomics 18: 683.

Sauquet H, Von Balthazar M, Magallon S, Doyle J, Endress P, Bailes E, de Morais E, Bull-Herenu K, Carrive L, Chartier M and Chomicki G. 2017. The ancestral flower of angiosperms and its early diversification. Nature Communications 8: 1-10.

Schilbert H, Rempel A and Pucker B. 2020. Comparison of read mapping and variant calling tools for the analysis of plant NGS data. Plants 9(4): 439.

Schneeberger K. 2014. Using next-generation sequencing to isolate mutant genes from forward genetic screens. Nature Reviews Genetics 15(10): 662-676.

Schwarz-Sommer Z, Huijser P, Nacken W, Saedler H and Sommer H. 1990. Genetic control of flower development by homeotic genes in *Antirrhinum majus*. Science 250(4983): 931-936.

Searle I, He Y, Turck F, Vincent C, Fornana F, Krober S, Amasino R and Coupland G. 2006. The transcription factor FLC confers a flowering response to vernalization by repressing meristem competence and systemic signalling in *Arabidopsis*. Genes and Development 20: 898-912.

Shawky E. 2016. Phytochemical and biological investigation of *Clivia nobilis* flowers cultivated in Egypt. Iranian Journal of Pharmaceutical Research 15(3): 531-535.

Shearman J and Tangphatsornruang S. 2021. Nanopore sequencing in agricultural and food applications. InHandbook of Nanotechnology Applications 443-459.

Sheldon C, Burn J, Perez P, Metzger J, Edwards J, Peacock W and Dennis E. 1999. The FLC MADS box gene: A repressor of flowering in *Arabidopsis* regulated by vernalization and methylation. Plant Cell 11: 445-458.

Shendure J and Ji H. 2008. Next-generation DNA sequencing. Nature Biotechnology 26(10): 1135-1145.

Shi T, Rahmani R, Gugger P, Wang M, Li H, Zhang Y, Li Z, Wang Q, Van De Peer Y, Marchal K and Chen J. 2020. Distinct expression and methylation patterns for genes with different fates following a single whole-genome duplication in flowering plants. Molecular Biology and Evolution 37(8): 2394-2413.

Shore P and Sharrocks A. 1995. The MADS-box family of transcription factors. European Journal of Biochemistry 229: 1-3.

Singh V and Jain M. 2014. Transcriptome profiling for discovery of genes involved in shoot apical meristem and flower development. Genomics data 2: 135-138.

Smyth D, Bowman J and Meyerowitz E. 1990. Early flower development in *Arabidopsis*. The Plant Cell 2(8): 755-767.

Snyman M. 2010. Identification and expression analysis of flavonoid biosynthetic genes in the genus *Clivia*. Doctoral dissertation, University of the Free State.

Sornchai P, van Doorn W, Imsabai W, Burns P and Chanprame S. 2020. Dendrobium orchids carrying antisense ACC oxidase: small changes in flower morphology and a delay of bud abortion, flower senescence, and abscission of flowers. Transgenic Research 29(4): 429-442.

Spies P and Spies J. 2018. *Clivia* taxonomy revisited, using DNA barcode regions. Acta Horticulture 1201: 503-514.

Spies P, Grobler J and Spies J. 2011. A review of phylogenetic relationships in the genus *Clivia*. Philosophical Transactions in Genetics 1: 168-207.

Spies P, van der Westhuizen H, Stegman S, Watson M and Spies J. 2011. Barcoding *Clivia* for species identification. Clivia 13: 26-37.

Stegemann T, Kruse L, Brutt M and Ober D. 2019. Specific distribution of pyrrolizidine alkaloids in floral parts of comfrey (*Symphytum officinale*) and its implications for flower ecology. Journal of Chemical Ecology 45(2): 128-135.

Stegmann S. 2011. Genetic variation of *Clivia caulescens*. Doctoral dissertation, University of the Free State.

Strickler S, Bombarely A and Mueller L. 2012. Designing a transcriptome next-generation sequencing project for a non-model plant species. American Journal of Botany 99(2): 257-266.

Suárez-López P, Wheatley K, Robson F, Onouchi H, Valverde F and Coupland G. 2001. CONSTANS mediates between the circadian clock and the control of flowering in *Arabidopsis*. Nature 410(6832): 1116-1120.

Sun Y, Fan Z, Li X, Liu Z, Li J and Yin H. 2014. Distinct double flower varieties in *Camellia japonica* exhibit both expansion and contraction of C-class gene expression. BMC Plant Biology 14: 1-1.

Swanevelder D and Fisher R. 2009. *Clivia*: Nature and nurture. Briza.

Swanevelder Z. 2003. Diversity and population structure of *Clivia miniata* Lindl. (Amaryllidaceae): evidence from molecular genetics and ecology. MSc thesis. University of Pretoria, Pretoria.

Taheri S, Lee Abdullah T, Yusop M, Hanafi M, Sahebi M, Azizi P and Shamshiri R. 2018. Mining and development of novel SSR markers using next generation sequencing (NGS) data in plants. Molecules 23(2): 399.

Theissen G and Melzer R. 2007. Molecular mechanisms underlying origin and diversification of the angiosperm flower. Annals of Botany 100(3): 603-619.

Thomson B and Wellmer F. 2019. Molecular regulation of flower development. Current Topics in Developmental Biology. Academic Press 131: 185-210.

Tsaftaris A, Pasentsis K, Makris A, Darzentas N, Polidoros A, Kalivas A and Argirioua A. 2011. The study of the E-class SEPALLATA3-like MADS-box genes in wild-type and mutant flowers of

cultivated saffron crocus (*Crocus sativus* L.) and its putative progenitors. Journal of Plant Physiology 168(14): 1675-1684.

Tzeng T, Hsiao C, Chi P and Yang C. 2003. Two lily SEPALLATA-like genes cause different effects on floral formation and floral transition in *Arabidopsis*. Plant Physiology 3: 1091-1101.

Roy-Choudhury S, Roy S, Nag A, Singh S and Sengupta D. 2012. Characterization of an AGAMOUS-like MADS box protein, a probable constituent of flowering and fruit ripening regulatory system in banana.

Unamba C, Nag A and Sharma R. 2015. Next Generation Sequencing Technologies: The doorway to the unexplored genomics of non-model plants. Frontiers in Plant Science 6: 1074.

UniProt: the Universal Protein knowledgebase. 2023. Nucleic Acids Research 51(D1): D523-D531.

Upson T and Andrews S. 2004. The genus *Lavandula*. Royal Botanic Gardens Kew.

Varadaraju A, Ramadoss B, Gurunathan S, Balaram R, Ganesamurthy K and Ram S. 2021. TILLING by sequencing (TbyS) reveals mutations in flowering control genes that are associated with altered plant architecture in Mungbean (*Vigna radiata*). Genetic Resources and Crop Evolution 68(3): 849-864.

Vaseva I, Akiscan Y, Demirevska K, Anders I and Feller U. 2011. Drought stress tolerance of red and white clover–comparative analysis of some chaperonins and dehydrins. Scientia horticulturae 130(3): 653-659.

Vermeulen J, De Preter K, Lefever S, Nuytens J, De Vloed F, Derveaux S, Hellemans J, Speleman F and Vandesompele J. 2011. Measurable impact of RNA quality on gene expression results from quantitative PCR. Nucleic Acids Research 39(9): e63.

Villar L, Lienqueo I, Llanes A, Rojas P, Perez J, Correa F, Sagredo B, Masciarelli O, Luna V and Swanevelder Z. 2003. Diversity and population structure of *Clivia miniata* Lindl. (Amaryllidaceae): Evidence from molecular genetics and ecology.

Villar L, Lienqueo I, Llanes A, Rojas P, Perez J, Correa F, Sagredo B, Masciarelli O, Luna V and Almada R. 2020. Comparative transcriptomic analysis reveals novel roles of transcription factors and

hormones during the flowering induction and floral bud differentiation in sweet cherry trees (*Prunus avium* L. cv. Bing). PloS One 15(3): e0230110.

Vimont N, Fouché M, Campoy J, Tong M, Arkoun M, Yvin J, Wigge P, Dirlewanger E, Cortijo S and Wenden B. 2019. From bud formation to flowering: Transcriptomic state defines the cherry developmental phases of sweet cherry bud dormancy. BMC Genomics 20: 1.

Vlk D and Repkova J. 2017. Application of next generation sequencing in plant breeding. Czech Journal of Genetics and Plant Breeding 53: 89-96.

Vorster P. 1994. *Clivia nobilis*. Flowering Plants of Africa 53: 70-74.

Waki T, Kodama M, Akutsu M, Namai K, Ligo M, Kurokura T, Yamamoto T, Nashima K, Nakayama M and Yagi M. 2018. Development of DNA markers linked to double-flower and hortensia traits in *Hydrangea macrophylla*. The Horticulture Journal 87(2): 264-273.

Wang C, Hou X, Qi N, Li C, Luo Y, Hu D, Li Y and Liao W. 2022. An optimized method to obtain high-quality RNA from different tissues in *Lilium davidii* var. unicolor. Scientific Reports 12: 2825.

Wang H, Lu Y, Zhang T, Liu Z, Cao L, Chang Q, Liu Y, Lu X, Yu S, Li H, Jiang J, Liu G, Sederoff H, Sederoff R, Zhang Q and Zheng Z. 2023 The double flower variant of yellow horn is due to a LINE1 transposon-mediated insertion. Plant Physiology 191(2): 1122-1137.

Wang H, Yang Y, Zhang Y, Zhao T, Jiang J, Li J, Xu X and Yang H. 2021. Transcriptome analysis of flower development and mining of genes related to flowering time in tomato (*Solanum lycopersicum*). Internal Journal of Molecular Science 22(15): 8128.

Wang Q, Wang Y, Sun L, Gao F, Sun W, He J, Gao X and Wang L. 2012. Direct and indirect organogenesis of *Clivia miniata* and assessment of DNA methylation changes in various regenerated plantlets. Plant Cell Reports 31(7): 1283-1296.

Wang Y, Li Y, Yan X, Ding L, Shen L and Yu H. 2020. Characterization of C, A and D-Class MADS-Box genes in orchids. Plant Physiology 184(3): 1469-1481.

Wang Z, Gerstein M and Snyder M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics 10: 57-63.

Wellmer F and Riechmann J. 2010. Gene networks controlling the initiation of flower development. Trends in Genetics 26(12): 519-527.

Wessinger C and Hileman L. 2016. Accessibility, constraints, and repetition in adaptive floral evolution. Developmental Biology 419: 175-183.

Williams V, Victor J and Crouch N. 2013. Red-listed medicinal plants of South Africa: Status, trends, and assessment challenges. South African Journal of Botany 86 :23-35.

Williamson M. 2008. Differential responses of tillers to floral induction in perennial ryegrass (*Lolium perenne* L.): implications for perennially. Doctoral dissertation, Massey University.

Winter J. 2000. The natural distribution and ecology of *Clivia*. Clivia Club, Kirstenbosch National Botanical Garden, Cape Town. Clivia Yearbook 2: 5-9.

Wollmann H, Mica E, Todesco M, Long J and Weigel D. 2010. Reconciling the interactions between APETALA2, miR172 and AGAMOUS with the ABC model of flower development. Development 137(21): 3633-3642.

Wozniak N and Sicard A. 2018. Evolvability of flower geometry: Convergence in pollinator driven morphological evolution of flowers. Seminars in Cell and Developmental Biology 79: 3-15.

Xu D, Hou Z, Deng N, Li Y, Yang L, Li S and Wang X. 2022. Transcriptome analysis of *Morus alba* L. flower reveals important genes of floral sex differentiation. Horticulturae 8(7): 625.

Xu X, Crow M, Rice B, Li F, Harris B, Liu L, Demesa-Arevalo E, Lu Z, Wang L, Fox N and Wang X. 2021. Single-cell RNA sequencing of developing maize ears facilitates functional analysis and trait candidate gene discovery. Developmental Cell 56(4): 557-568.

Yamada K, Saraike T, Shitsukawa N, Hirabayashi C, Takumi S and Murai K. 2009. Class D and B-sister MADS-box genes are associated with ectopic ovule formation in the pistil-like stamens of alloplasmic wheat (*Triticum aestivum* L.). Plant Molecular Biology 71: 1-14.

Yang H, Zhang Y, Zhen X, Guo D, Guo C and Shu Y. 2020. Transcriptome sequencing and expression profiling of genes involved in daylily (*Hemerocallis citrina* Borani) flower development. Biotechnology and Biotechnological Equipment 34: 542-548.

Yue J, Zhu C, Zhou Y, Niu X, Miao M, Tang X, Chen F, Zhao W and Liu Y. 2018. Transcriptome analysis of differentially expressed unigenes involved in flavonoid biosynthesis during flower development of *Chrysanthemum morifolium* 'Chuju'. Scientific Reports 8: 13414.

Zeng F, Biligetu B, Coulman B, Schellenberg M and Fu Y. 2017. RNA-Seq analysis of plant maturity in crested wheatgrass (*Agropyron cristatum* L.). Genes 8(11): 291.

Zhang C, Cui L and Fang J. 2022. Genome-wide association study of the candidate genes for grape berry shape-related traits. BMC Plant Biology 22(1): 42.

Zhang H, Wei Y, Shen J, Lai B, Huang X, Ding F, Su Z and Chen H. 2014. Transcriptomic analysis of floral initiation in litchi (*Litchi chinensis* Sonn.) based on *de novo* RNA sequencing. Plant Cell Reports 33: 1723-1735.

Zhao L, Chen P, Liu P, Song Y and Zhang D. 2021. Genetic effects and expression patterns of the nitrate transporter (NRT) gene family in *Populus tomentosa*. Frontiers in Plant Science 12: 661635.

Zhao Y, Broholm S, Wang F, Rijpkema A, Lan T, Albert V, Teeri T and Elomaa P. 2020. TCP and MADS-box transcription factor networks regulate heteromorphic flower type identity in *Gerbera hybrida*. Plant Physiology 184(3): 1455-1468.

Zhou X, Fei Z, Thannhauser T and Li L. 2011. Transcriptome analysis of ectopic chloroplast development in green curd cauliflower (*Brassica oleracea* L. var. botrytis). BMC Plant Biology 11: 169.

Appendix A

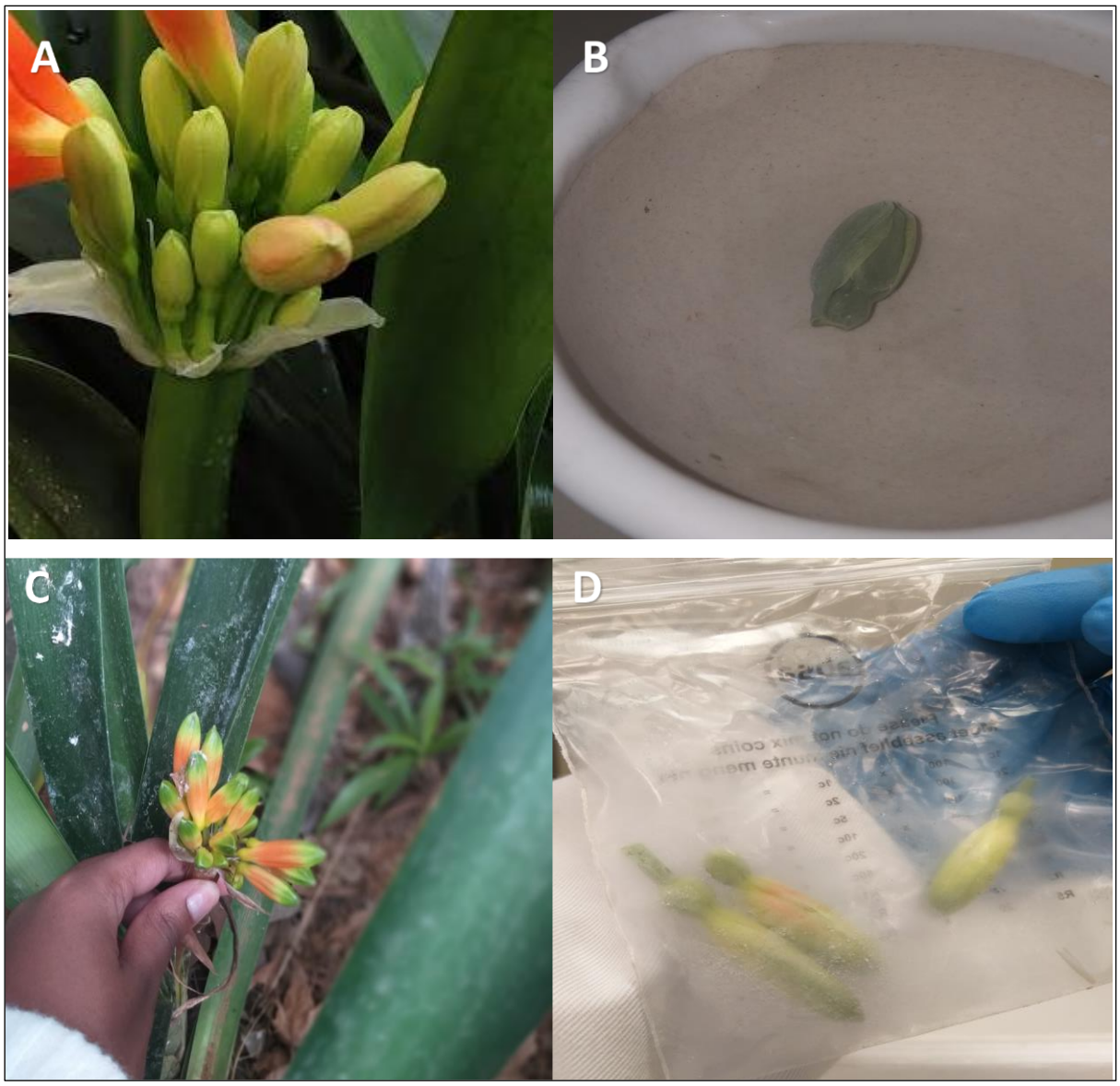

Figure S1: *Clivia* samples used in the current study. A and B show the non-pendulous flower buds whereas C and D show the pendulous flower buds.

Table S1: Nanodrop results depicted as average concentration, average 260/280 ratios and average 260/230 ratios.

Sample ID	Average Concentration (ng/µl)	Average 260/280	Average 260/230
Intermediate bud 1	2265.29	2.02	2.02
Intermediate bud 2	1965.33	2.03	2.34
Intermediate bud 3	2472.79	2.01	2.28
Miniata bud 1	1229.01	2.04	2.19
Miniata bud 2	1171.53	2.05	2.24
Miniata bud 3	987.55	2.03	2.32
Tiny bud 1	1046.17	2.03	1.79
Tiny bud 2	1231.80	2.05	2.17
Tiny bud 3	776.36	2.05	2.16

Table S2: Scaffold length distribution in the *de novo* assembled bud samples.

A: Intermediate buds		
Minimum scaffold length	No. of scaffolds	Total scaffold length
100	398,674	239,309,249
250	317,109	221,087,386
500	136,316	158,918,187
1 KB	58,395	105,083,937
2.5 KB	8,642	29,608,240
5 KB	614	3,941,692
10 KB	25	306,549
Sub total	919,775	758,255,240
B: Miniata buds		
Minimum scaffold length	No. of scaffolds	Total scaffold length
100	309,449	223,003,611
250	261,009	212,138,288
500	131,490	167,132,219
1 KB	63,603	119,631,208
2.5 KB	11,135	38,716,113
5 KB	906	5,688,302
10 KB	28	355,738
Sub total	468,171	766,665,479
C: Tiny buds		
Minimum scaffold length	No. of scaffolds	Total scaffold length
100	309,956	203,823,820
250	253,682	191,266,291
500	119,777	144,933,967
1 KB	54,957	99,848,207
2.5 KB	18,253	28,427,710
5 KB	641	4,024,540
10 KB	18	227,392
Sub total	757,284	672,551,927
Total	**2,145,230**	**2,197,472,646**

Table S3: ExN50 statistics of the main assembly showing the ExN50 value corresponding to the highest Ex value as well as the number of genes at a specific Ex value.

Ex	ExN50	Num_genes
1	3138	1
2	7488	4
3	5625	7
4	3138	12
---	---	---
88	2712	33412
89	2731	36326
90	2746	39766
91	**2754**	**43906**
92	2742	49040
93	2701	55598
94	2615	64315
95	2478	76684
96	2273	96112

```
Inter_Buds_trimmed_2.fastq,Miniata_Buds_trimmed_2.fastq,Tiny_Buds_trimmed_2.fastq -S SPAdes_transcri
Time loading reference: 00:00:01
Time loading forward index: 00:00:00
Time loading mirror index: 00:00:01
Multiseed full-index search: 06:35:42
147102407 reads; of these:
  147102407 (100.00%) were paired; of these:
    49302802 (33.52%) aligned concordantly 0 times
    24305207 (16.52%) aligned concordantly exactly 1 time
    73494398 (49.96%) aligned concordantly >1 times
    ----
    49302802 pairs aligned concordantly 0 times; of these:
      8464990 (17.17%) aligned discordantly 1 time
    ----
    40837812 pairs aligned 0 times concordantly or discordantly; of these:
      81675624 mates make up the pairs; of these:
        8139116 (9.97%) aligned 0 times
        2558975 (3.13%) aligned exactly 1 time
        70977533 (86.90%) aligned >1 times
97.23% overall alignment rate
Time searching: 06:35:46
Overall time: 06:35:47
```

Figure S2: Read representation statistics of *de novo* assembled transcriptome

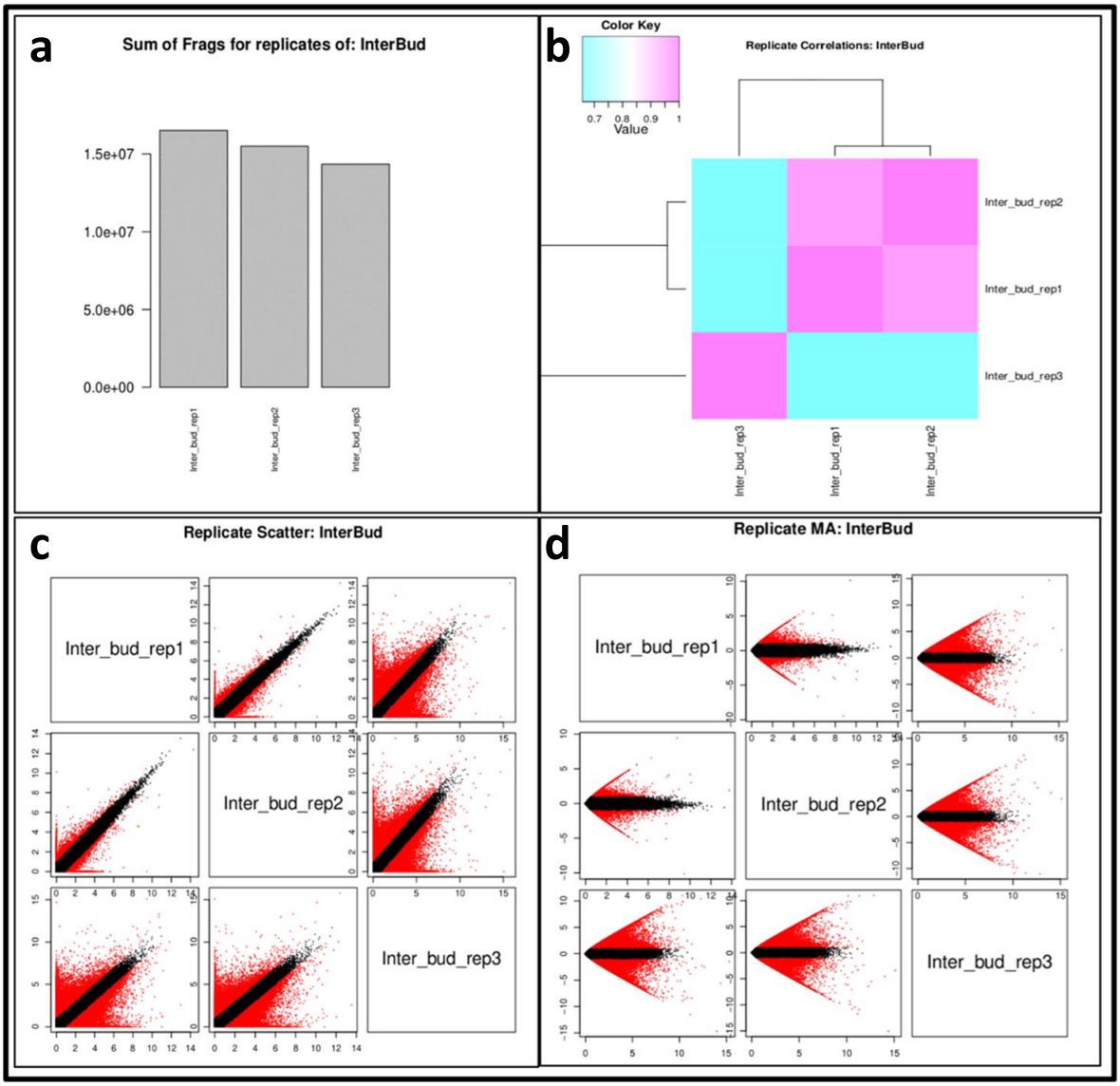

Figure S3: The comparison of biological replicates. Plots (**a**) shows the sum of mapped fragments, (**b**) replicate Pearson correlation heatmap, (**c**) pairwise comparisons of replicate log(CPM) values where data points highlighted in red indicate a >2-fold change and (**d**) pairwise MA plot of log(CPM) vs log(fold_change).

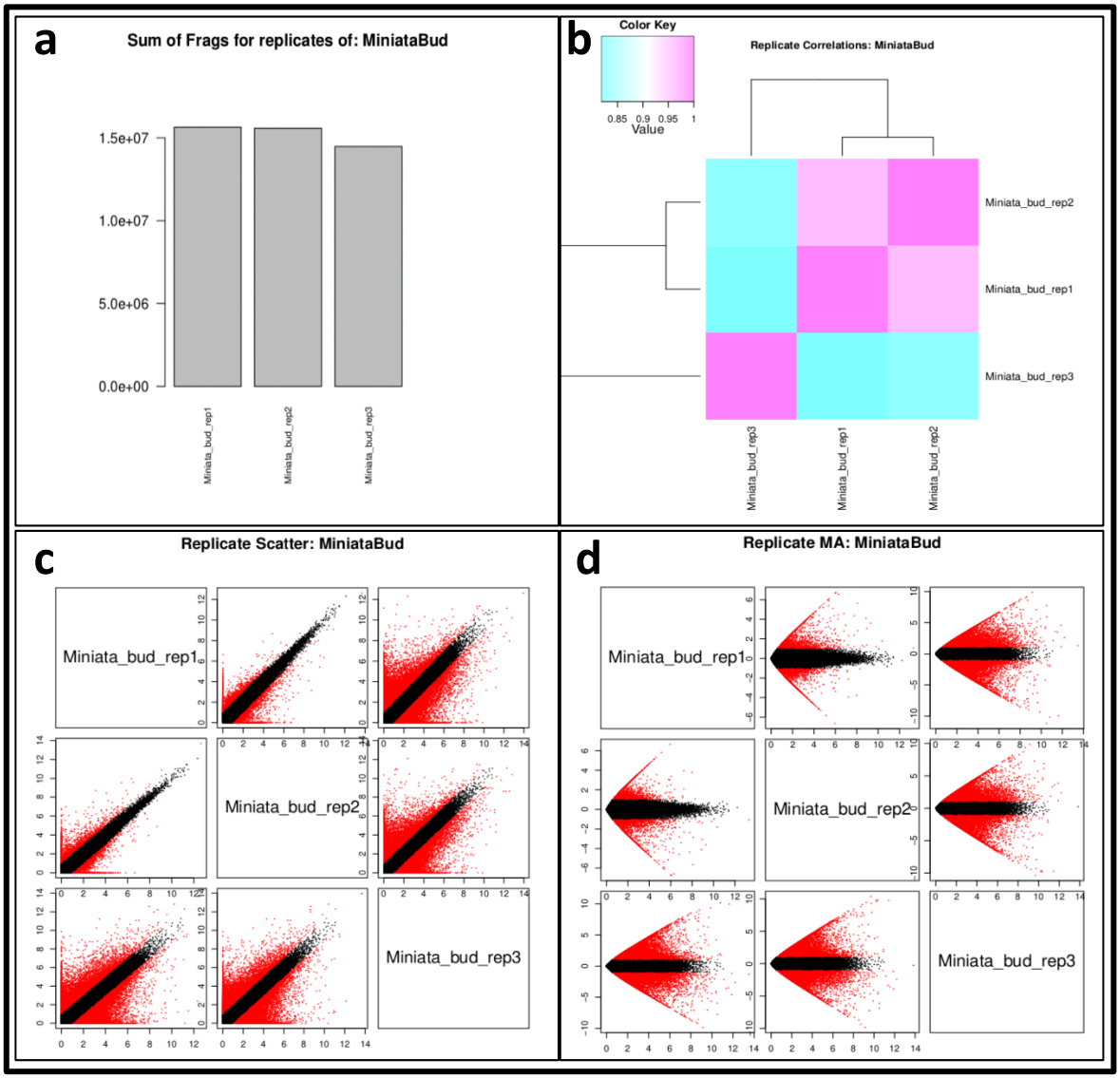

Figure S3.2: The comparison of miniata bud's biological replicates. Plots (**a**) shows the sum of mapped fragments, (**b**) replicate Pearson correlation heatmap, (**c**) pairwise comparisons of replicate log(CPM) values where data points highlighted in red indicate a >2-fold change and (**d**) pairwise MA plot of log(CPM) vs log(fold_change).

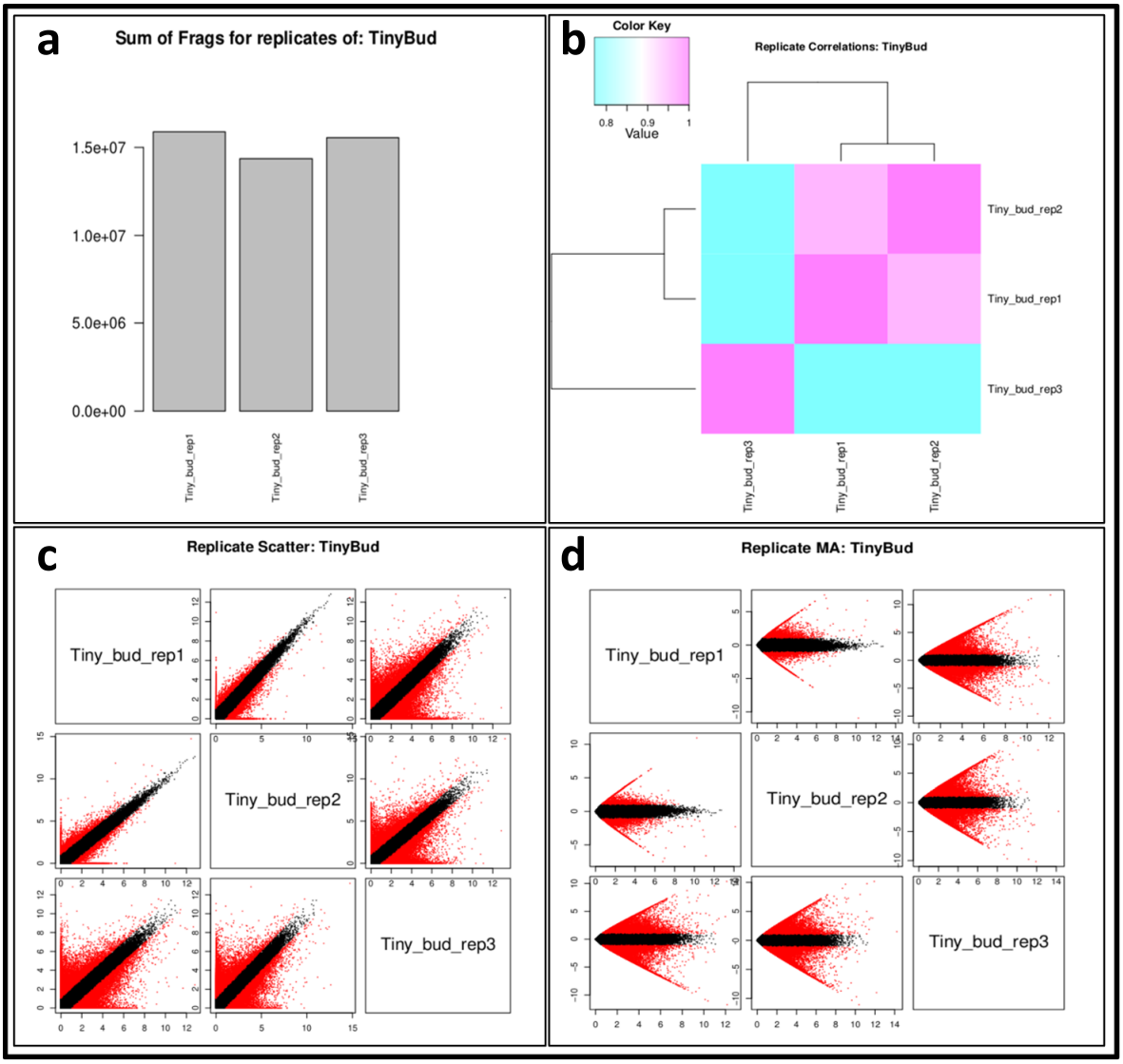

Figure S3.3: The comparison of tiny bud's biological replicates. Plots (**a**) shows the sum of mapped fragments, (**b**) replicate Pearson correlation heatmap, (**c**) pairwise comparisons of replicate log(CPM) values where data points highlighted in red indicate a >2-fold change and (**d**) pairwise MA plot of log(CPM) vs log(fold_change).

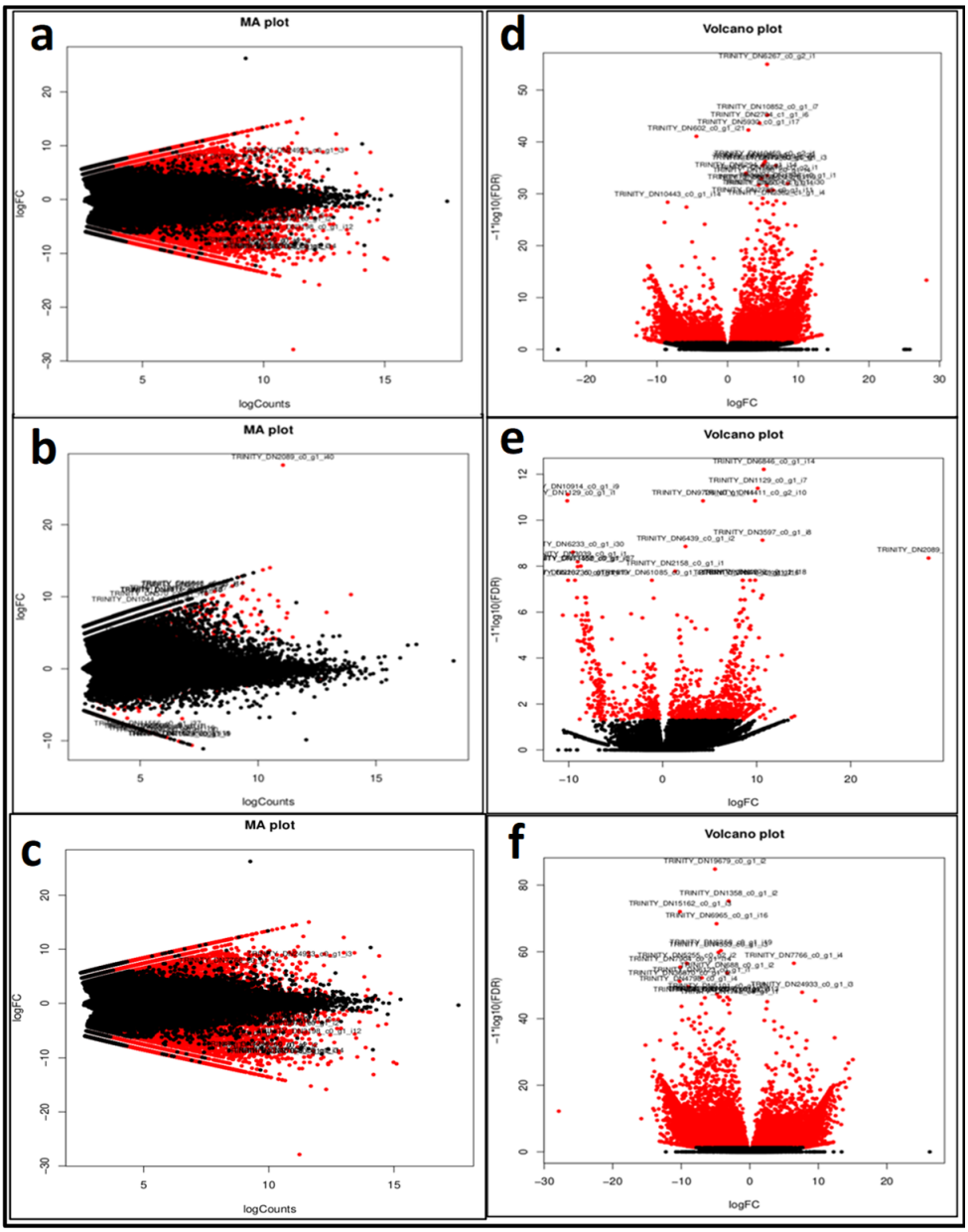

Figure S4: Pairwise comparisons of transcript expression in different tissues. The MA plots depict logFC (average log fold change) against logCounts (average log of counts) between (**a**) Miniata buds vs Inter buds (**b**) Inter buds vs Tiny buds and (**c**) Tiny buds vs Miniata buds. Volcano plots depict differentially expressed transcripts relative to FDR (False discovery rate) for (**d**) Miniata buds vs Inter buds (**e**) Inter buds vs Tiny buds and (**f**) Tiny buds vs Miniata buds.

Command Lines

Trimming

$ bbduk.sh in1=read1 in2=read2 out1=outr1 out2=outr2 ktrim=r k=23 mink=11 hdist=1 tpe tbo ref=adapter_file

$ bbduk.sh in1=outr1 in2=outr2 out1=outr1_trimmed out2=outr2_trimmed qtrim=rl trimq=20 maq=20 ftl=21 ftr2=4 entropy=0.1 outs=trimmed_singles

Blast

$blastx -query fasta -db swissprot -out proteins.txt -evalue 1e-10 -word_size 3 -gapopen 11 -gapextend 1 outfmt '6 std qlwn stittle' -max_target_seq 1 -html -num_threads 4

Assembly

$ bowtie2-build Trinity.fasta Trinity.fasta

$ bowtie2 -p 10 -q --no-unal -k 20 -x Trinity.fasta -1 reads_1.fq -2 reads_2.fq 2>align_stats.txt | samtools view -@10 -Sb -o bowtie2.bam

$TRINITY_HOME/util/TrinityStats.pl Trinity.fasta

$TRINITY_HOME/util/misc/contig_ExN50_statistic.pl transcripts.TMM.EXPR.matrix Trinity.fasta transcript | tee ExN50.transcript.stats

Estimating transcript abundance

$TRINITY_HOME /util/align_and_estimate_abundance.pl --transcripts Trinity.fasta --seqType fq --left reads_1.fq --right reads_2.fq --est_method salmon --aln_method bowtie --trinity_mode --prep_reference --output_dir_salmon

Building matrices

$TRINITY_HOME /util/abundance_estimates_to_matrix.pl --est_method salmon --quant_files file.listing_target_files.txt

$TRINITY_HOME/util/abundance_estimates_to_matrix.pl --est_method salmon --gene_trans_map

trinity_out_dir.Trinity.fasta.gene_trans_map --quant_files salmon.quant_files.txt --name_sample_by_basedir

$TRINITY_HOME/util/misc/count_matrix_features_given_MIN_TPM_threshold.pl trans_matrix.TPM.not_cross_norm | tee trans_matrix.TPM.not_cross_norm.counts_by_min_TPM

Comparing replicates

$TRINITY_HOME/Analysis/DifferentialExpression/PtR --matrix counts.matrix –samples samples.txt --log2 --CPM --min_rowSums 10 --compare_replicates

Principal component analysis

$TRINITY_HOME/Analysis/DifferentialExpression/PtR --matrix Trinity_trans.counts.matrix -s samples.txt --min_rowSums 10 --log2 --CPM --center_rows --prin_comp 3

Differential gene expression

$TRINITY_HOME/Analysis/DifferentialExpression/run_DE_analysis.pl –matrix salmon.isoform.counts.matrix -method salmon -samples_file main_samples.txt

$TRINITY_HOME/Analysis/DifferentialExpression/analyze_diff_expr.pl -matrix salmon.isofrom.TMM.EXPR.matrix -P 1e-3 -C 2 -max_gene_clust 30 000

Trinotate

$ Trinotate.init

$ autoTrinotate.pl --Trinotate_sqlite exp_1.sqlite --transcripts expr_1_trinity_assembly.fasta --gene_to_trans_map expr1_trinity_assembly.gene_to_trans_map --conf ufs_trinotate_conf.txt –CPU 6

*__All command lines used can be found on Github (https://github.com/trinityrnaseq/trinityrnaseq/wiki) and/or UFS docs (https://docs.hpc.ufs.ac.za/).__