



THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e. g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

- This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
- A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
- This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
- The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
- When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

The evolution of the legume flower: case studies in the early-branching papilionoid legumes (Papilionoideae, Leguminosae)

Gustavo Ramos de Oliveira

Doctor of Philosophy



School of Biological Sciences
Institute of Molecular Plant Sciences
Royal Botanic Garden Edinburgh

The University of Edinburgh

May 2024

The evolution of the legume flower: case studies in the early-branching papilionoid legumes (Papilionoideae, Leguminosae)

Gustavo Ramos de Oliveira

The University of Edinburgh
School of Biological Sciences
Institute of Molecular Plant Sciences
Royal Botanic Garden Edinburgh

Supervisors

Richard Toby Pennington ^{1,3}, Catherine Kidner ^{1,2}

¹Royal Botanic Garden Edinburgh

²The University of Edinburgh

³University of Exeter

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except when explicitly stated otherwise in the text, and that this work has not been submitted to any other degree or professional qualification except as specified.

A handwritten signature in blue ink, reading "Gustavo Ramos de Oliveira". The signature is written in a cursive style with a large initial 'G'.

Gustavo Ramos de Oliveira

30st May 2024

Abstract

Flowers are the reproductive structures of angiosperms and have played an important role in the diversification of land plants. In the legume family (Leguminosae or Fabaceae), the diversity of flower morphology is remarkable across its six subfamilies, but one particular type of architecture is considered the signature of the family: the papilionate flower. This typical pea flower with zygomorphic symmetry, five petals and ten stamens is exclusively found in the Papilionoideae subfamily, but genera with a range of non-papilionate floral morphologies are found across the early-branching lineages nested within papilionate lineages. Investigating the genetic changes in these divergent lineages can shed light on the underlying processes leading to the evolution of novel flower morphology.

CYCLOIDEA (CYC), DIVARICATA (DIV) RADIALIS (RAD) and DIV-and-RAD-Interacting-Factor (DRIF) have been identified as key genes regulating floral symmetry across flowering plants. In typical papilionate flowers, CYC like genes are expressed on the dorsal (upper) side of the flower and are specific to development of the standard petal, whereas DIV, RAD and DRIF activity is not well known. In this thesis, we use new genomic and transcriptomic data to explore the key symmetry-regulating genes in the papilionoid legume genera *Andira* (typical zygomorphic papilionate flower), *Cyathostegia* (single petal and multiple stamens), and *Cadia* (radially symmetrical flower). We produce draft genome assemblies and RNASeq to identify putative orthologs of CYC, DIV, RAD, and DRIF. We quantify the relative abundance of these genes in the different developmental stages and between species. Using these genomic and transcriptomic data in the context of previously published DNA sequence data for these genes from across legumes and other angiosperms we use a phylogenetic approach to understanding their evolution and possible function in Papilionoideae. We find that non-papilionate genera do not differ in copy number of CYC, DIV and RAD, suggesting, in contrast to prior studies, that higher copy number of these genes is not associated with the typical papilionate floral morphology. We identify genes differentially expressed between adaxial and abaxial domains of the *Cadia* flower, genes upregulated on the floral transition *Cadia* and *Cyathostegia*, and gene families differentially expressed between the species. We examine the expression pattern of candidate genes to establish if there is any evidence for neofunctionalisation of the retained paralogs. This work is placed into an ontogenetic developmental context with new data of floral ontogeny and published information from across basal legumes. We focus on petal heterochronical patterns

and androecium symmetry because these floral features have been shown to be variable in legumes and have only been studied in isolated fashion in classic ontogenetic studies. We show multiple independent evolutionary changes in both features, which underlines the general evolutionary lability of floral forms, for example overall symmetry, uncovered by prior phylogenetic studies.

This thesis also presents a taxonomic synopsis of *Aldina*, a genus of Papilionoideae with unusual non-papilionoid flower morphology. Flowers of *Aldina* are radially symmetrical, with numerous free stamens, unlike its sister genera *Andira* and *Hymenolobium*, whose zygomorphic flowers are typically papilionate and with 10 stamens. *Aldina* taxonomy is in need of a revision since the preliminary taxonomic studies on the genus that date back 70 years. *Aldina* is also ecologically important in the Amazon basin, but because it is poorly studied, and species boundaries are uncertain. Given the urgent need of filling in the gaps in knowledge in the plant diversity of unexplored regions, this thesis presents a taxonomic synopsis of the genus. Eighteen species are recognized in *Aldina*, with seven names newly placed in synonymy, new lectotypification of five names, and a new combination, *Aldina auyantepuiensis*, is proposed.

Lay summary

In the legume family (Leguminosae), the diversity of flower morphology is remarkable across its six subfamilies, but one particular type of architecture is considered the signature of the family; the papilionate flower. This typical pea flower with five petals, bilateral symmetry, and ten stamens is exclusively found in the Papilionoideae subfamily, but genera with a range of non-papilionate floral morphologies are found across the first lineages to diverge within the subfamily. Investigating the genetic changes in these divergent lineages can shed light on the underlying processes leading to the evolution of novel flower morphology.

In order to understand how such diversity in floral form emerged in Papilionoideae, I started to investigate the genetics underlying changes in floral morphology in different groups of flowering plants in order to find candidate genes that could explain flower diversity in papilionoid legumes. A set of candidate genes reported in the literature was designed including CYCLOIDEA (CYC), DIVARICATA (DIV) RADIALIS (RAD), and DIV-and-RAD-Interacting-Factor (DRIF) have been identified as key genes regulating floral symmetry across flowering plants. Four representatives of the subfamily Papilionoideae were selected taking into account their distinct morphologies and include two species of the typical bilateral pea-flowered *Andira*, the single-petal *Cyathostegia* and the radially symmetric *Cadia*. Each species had their genome sequenced in order to find our set of candidate genes. After employing several bioinformatic tools to mine genes in the genomes, I found copies of the candidate genes in all sequenced genomes. The confirmation that these genes were true copies came with the sequencing of the transcriptomes (RNA-seq) of our selected species. I generated phylogenetic trees with both genomic and transcriptomic sequences which supported each other in the tree among a wide sampling of other flowering plants, including sequences from other legume species.

The transcriptomic data was also important to investigate the expression of the candidate genes in the floral and vegetative tissues. I designed experiments where RNA was extracted from a series of developmental stages from our species, including vegetative buds, young flowers, and abaxial and adaxial domains of mature flowers. CYC, RAD, DIV and DRIF were found differentially expressed across species in the selected tissues and confirmed the role of these genes in controlling processes of flower development.

I have also analysed the flowers of our set of species from an ontogenetic perspective, which considers the developmental aspects of flower growth. A wide sampling of

developmental data from several papilionoid lineages was compiled and two important floral features were selected to be analysed, petal growth and the symmetry of stamens grouping. We concluded that these features have had an important role in the fixation of the typical pea flower in more derived lineages of legumes by stabilizing a slow petal growth and symmetrical development of the androecium. These features have a relative lability in the early-branching lineages of Papilionoideae, where some reversals are observed or undefined types of growth and symmetry are reported.

This thesis also presents a taxonomic synopsis of *Aldina*, a genus of Papilionoideae with unusual non-papilionoid flower morphology. Flowers of *Aldina* are radially symmetrical, with free numerous stamens, unlike its sister genera *Andira* and *Hymenolobium*, whose zygomorphic flowers are typically papilionate and with 10 stamens. *Aldina* taxonomy is in need of a revision since the preliminary taxonomic studies on the genus that date back 70 years. *Aldina* is also ecologically important in the Amazon basin, but because it is poorly studied, species boundaries are uncertain. Given the urgent need of filling in the gaps in knowledge in the plant diversity of unexplored regions, this thesis presents a taxonomic revision of the genus. Our taxonomic account of *Aldina*, therefore, contributes to understanding the species diversity in Amazonia, filling in the gaps of plant diversity in this under-explored region.

Acknowledgements

Writing this section is probably the hardest task to be done in this whole thesis. It is easier to start though when I think when this journey started and therefore, I am most grateful to the people who have made this four years' experience possible in first place: Domingos Cardoso, Tiina Särkinen, Edeline Gagnon, Zoe Goodwin, Peter Moonlight, and Toby Pennington. The opportunity to work with you in the Nordeste project in 2018 was, without a doubt, a great professional experience and it has marked a big change in my life for the next 5 years. I am so grateful that this work allowed me to visit the cold lands of Scotland for the first time.

I am very lucky to have met an amazing group of inspiring people that motivated me to apply for the PhD at the University of Edinburgh and work in this fabulous place which is the Royal Botanic Garden. I am of course talking about my fellow friends from the basement offices at the Botanic, I have learned much more than bioinformatics and genomics with you all, thank you for your support and friendship over the years: Lucia, Madhavi, Cynthia, Rebecca, Thibault, Surabhi, Natalia, Pakkapol, Junru, Charlie, Ozan, Flavia, and Hazel. There are many more people to mention, but at this stage my brain is completely melted to remember everyone. But be sure that you all made this journey one of the best experiences.

My PhD was funded by the Darwin Trust of Edinburgh, to whom I am very thankful for the opportunity to study at the University of Edinburgh. Thanks to Professor David Finnegan, who has been so interested in my project since the beginning of my studies.

I am very thankful to my team of supervisors Professor Toby Pennington and Dr. Catherine Kidner who have played a vital part in my academic progress. Thank you for all the shared knowledge and for being so patient and open to discuss work and the rantings of a PhD student.

Also thanks to the other members of my thesis committee including Andre Hudson for the feedbacks in my annual reviews.

I also want to thank to all the staff from RBGE whom in one way or another contributed to a pleasant life at the botanics. Laura Forrest, Michelle Hart, Kanae Nishii: thank you so much for all the technical support in the labs! Valeria, Agron, and Scott: always making my start at the botanics nicer with their kindness and wide smiles. Simon Milne, your interest in the PhD student's life has always been heart-warming, thank you for always checking how we were doing and for just being a nice person.

I want to thank to my family for being patient and dealing well with the distance during this period. We know it has not been easy years, but I am glad we still managed to offer each other support. Thank you being there and always receive me with so much love in my visits home.

Special thanks to the Johnston clan for making the last two years full of adventures in Scotland and becoming my Scottish family. Thank you Struan, especially in the last months of my PhD, you helped to make the final sprint calmer with your love and support.

Finally I want to thank to everyone who allowed me to borrow their brains during this journey. Your collaboration has made a big difference in the final look of this thesis! Thank you Lucía, Rowan, Flavia, Yumaly, Domingos, Catherine, and Toby!

Table of Contents

Chapter 1: Introduction and Literature review

1.1 The role of flower symmetry in the diversification of Legumes	1
1.2 Types of floral symmetry and evolutionary trends	3
1.3 Flower identity: a matter of expression	7
1.4 Conserved domains characterizes TCP gene family.....	13
1.5 MYB gene family	15
1.6 Thinking beyond the models: the need for studies of transcription factors in papilionoid legumes with unusual flowers	17
1.7 Expression of symmetry genes in legumes	17
1.8 Studies in flower ontogeny	18
1.9 Taxonomic studies of <i>Aldina</i> , a genus of Amazonian papilionoid legume trees with radial flowers	21
2. Objectives and approaches	24
References	26

Chapter 2: An investigation of symmetry genes in papilionoid legumes

2.1 Introduction	34
2.2 Material and methods	38
2.2.1 Plant material sampling	38
2.2.2 Data collection and sequence annotation	38
2.2.3 Whole genome sequencing	39
2.2.4 Genomic data QC and pre-processing	40
2.2.5 Genome assembly	42
2.2.6 Assembly process	43
2.2.7 K-mer analysis	44
2.2.8 Genome annotation	45
2.2.9 Candidate gene dataset	46
2.2.10 Phylogenetic analyses	46
2.2.11 Protein structure	47
2.3 Results	47

2.3.1 K-mer analysis: ploidy and levels of heterozygosity	47
2.3.2 Secondary protein structure of candidate genes.....	50
2.3.4 Phylogenetic analyses of TCP and MYB genes in papilionoid legumes.....	50
2.4 Discussion.....	58
References	62
Tables and Appendix.....	69

Chapter 3: Differential expression of symmetry genes in the early branching

Papilionoideae

Introduction	86
3.1 RNA sequencing and its applications	86
3.2 Comparative transcriptome analyses	88
3.3 Transcriptomics studies in legumes	88
3.4 Material and methods	90
3.4.1 RNA-seq analysis strategy	90
3.4.2 Plant material and isolation	90
3.4.3 RNA extraction, library preparation and RNA-seq	92
3.4.4 Pre assembly quality control	92
3.4.5 Alignment to draft genomes	93
3.4.6 <i>De novo</i> transcriptome assembly with trinity	94
3.4.7 Mikado transcriptome annotation	95
3.4.8 Transcripts quantification	95
3.4.9 Differential gene expression analysis	96
3.5 Results and discussion	
3.5.1 3D Analysis of contrast groups	100
3.5.2 Differential expression of candidate genes involved in flower symmetry	107
3.6 Conclusion	110
References	112
Tables and software codes.....	116

Chapter 4: An ontogenetic overview of the florally diverse Papilionoideae (Leguminosae): new insights into the evolution of androecium symmetry and petal heterochrony

Introduction	127
4.1 What else development tell us about floral differences in legumes.....	127
4.2 Premises and Hypotheses of this study.....	130
Material and Methods	
4.3 Target species	131
4.4 Scanning electron microscopy analysis	132
4.5 Ancestral character estimation	132
Results	
4.6 Ontogenetic development of androecium and petal growth in selected papilionoid taxa	133
4.7 Evolution of petal growth and androecium symmetry	139
Discussion	
4.8 From zygomorphic to radially symmetrical flowers	144
4.9 Ontogenetic stability of slow petal growth and symmetrical androecium underlying the evolutionary stability of the papilionate floral architecture in papilionoid legumes.....	148
4.10 Insights into differential rates of petal growth and androecium symmetry from transcriptomics	149
References	152
Tables	159

Chapter 5: A taxonomic synopsis of *Aldina*, a florally distinctive and poorly collected Amazonian genus of papilionoid legumes

Introduction	164
5.1 Taxonomic history	165
5.2 Phylogenetic affinities	167
5.3 Geographic Distribution, Habitat and Ecology	170
Material and Methods	173
5.4 Taxonomic treatments	174
References	200

Chapter 6: Conclusions and future work

6.1 General conclusions	207
6.2 Future work	209

Chapter 1

General introduction and literature review

Introduction

1.1 The role of flower symmetry in the diversification of legumes

Flowers are the reproductive structures of angiosperms (flowering plants), which represent ca. 90% of all living land plants (Sauquet et al., 2017). The elaboration of forms of flowers contributed to the extraordinary events of diversification along the evolutionary history of the angiosperms, and via coevolution also to the diversification of insect and other pollinators (Citerne et al., 2010; Sauquet et al., 2017). Even though flowers possess a well-conserved general ground plan, evolution has produced huge diversity within this basic whorl of organs with petals surrounding gynoecium and androecium in a wide range of variation in size, symmetry, colour, shape, degree of fusion, and number of parts. This diversity in general flower shape is related to variation in reproductive strategies and a wide range of pollination syndromes (Sauquet et al., 2017).

In the legume family (Leguminosae), the third largest family of angiosperms with about 800 genera and over 23,000 species (LPWG, 2017), the plasticity of floral characters is striking with the majority of species bearing zygomorphic (bilaterally symmetric) flowers. The great diversity in flower morphology has historically influenced the classification of legumes, from a long-standing traditional recognition of the three subfamilies Papilionoideae (Faboideae), Mimosoideae, and Caesalpinioideae (Bruneau et al., 2013) to a recent new phylogenetic classification based on six subfamilies: Duparquetioideae (1 genus/ 1 sp.), Cercidioideae (14 genera/ 400 spp.), Detarioideae (79 genera/ 760 spp.), Caesalpinioideae (163 genera/ 4680) and Papilionoideae (500 genera/ 14,000 spp.) (LPWG, 2017).

The newest legume classification is based upon monophyly and resolves, for example, the previously paraphyletic Caesalpinioideae by including within it, the mimosoid legumes (ex subfamily Mimosoideae) that are nested as a clade (LPWG, 2017). The types of floral symmetry found in these subfamilies differ (LPWG, 2017; Tucker, 2003). The Papilionoideae encompasses the vast majority of species with zygomorphic, papilionate flowers. This typical zygomorphic, papilionate flower has petals differentiated into an adaxial banner, two lateral wings and two partially fused petals forming the keel. Internal organs are also strongly bilateral; the androecium is generally tilted towards the adaxial side of the flower with nine stamens partially united into a proximal sheath, free halfway towards the tips and a single free stamen on the abaxial side of the flower (Fig. 1). Similar to the androecium, the carpel tends to be curved towards the banner petal (Tucker, 2003). In the Detarioideae and

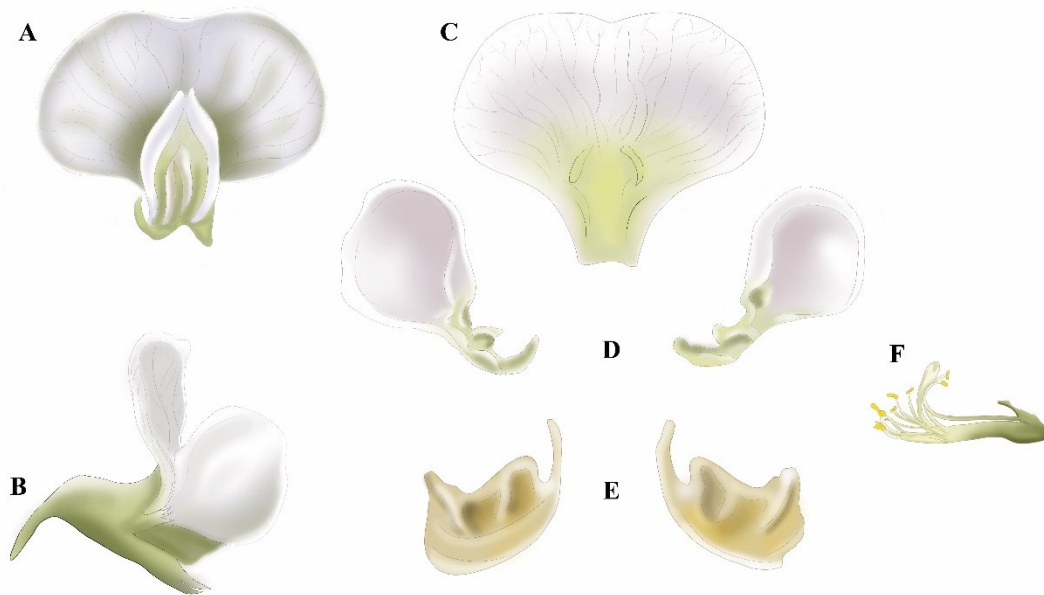


Figure 1. Depiction of the typical papilionate flower found in the Papilionoideae subfamily. A–B Frontal and lateral views show the zygomorphy plan of symmetry of the flower. C–F Show internal organ asymmetry in the adaxial banner petal (C), lateral wing petals (D), ventral keel petals (E), and androecium and pistil (F). The ornamentation of the wing and keel petals have two main functions: (i) the ridges, grooves, and occasional striations work as an interlocking mechanism that encloses and protects the stamens and pistil until pollinators trigger the exposition of these structures. The same structures (ii) work as a landing platform for insect pollinators, which use the grooves and pockets in the keel for stabilization. The body weight and buzzing of the pollinators trigger the releasing of the keel petals, exposing the stamens and pistil, which touches the insect’s abdomen making it a vector of pollen grains.

Duparquetioideae, flowers are bilaterally symmetric but do not have specialised petals. Flowers of Cercidioideae and Dialioideae are radially or slightly to strongly bilaterally symmetrical, and sometimes papilionate. Caesalpinioideae display a wide range of flower symmetry, ranging from near radial symmetry to bilaterally symmetric, but many genera have radial symmetry, especially in the mimosoid clade. (LPWG, 2017).

1.2 Types of floral symmetry and evolutionary trends

Flower diversity is thought to be the evolutionarily driven by pollinators, with floral symmetry playing an important role in shifts between flower types (Neal et al., 1998; Endress, 1999; Sargent 2004; Hileman 2014; Ojeda et al. 2017; Ruchisansakun et al., 2021). At least two main types of floral symmetry can be described: radial (actinomorphic or polysymmetric) and bilateral (zygomorphic or monosymmetric). Radial flowers are characterised by multiple planes of symmetry that mirror one another, such symmetry is illustrated by sunflowers (Asteraceae), although this is only true when considering the capitulescent inflorescence in its entirety or the disk flowers, since the ray flowers are zygomorphic because of the presence of a ligule petal. Zygomorphic flowers, as mentioned above, have only one plane of symmetry, resulting in a single mirror image. Pea flowers (*Pisum sativum* L., Leguminosae, Papilionoideae) make a good example of classic zygomorphic symmetry. This zygomorphic trait in particular is thought to have allowed the great success of legumes in pollinator specialisation and with subsequent speciation (Cubas, 2004; Sargent, 2004; Zhao et al., 2019).

The overall appearance of a papilionate flower is one of bilateral symmetry that can only be divided into a single plane. Although the typical zygomorphic symmetry is the signature of Papilionoideae, genera with unusual radially symmetric flowers are found across the phylogeny of the subfamily (Pennington et al., 2000, 2001; Cardoso et al., 2012) (Fig. 2). Interestingly, zygomorphy or actinomorphy are not pre-determined from the early stages of development. Ontogenetic studies show that papilionoid species with radially symmetric flowers undergo a zygomorphic phase in the early stages of organ development, such in seen in *Cadia purpurea* Forsk., which after a brief zygomorphic phase, have the floral organ primordia radially arranged (Tucker, 2002). Conversely, papilionoid legumes with final zygomorphic symmetry can undergo a radially symmetric phase in early development, as seen in Swartzoid legumes such as *Swartzia*, *Trischidium*, *Ateleia*, *Cyathostegia* and *Mildbraediodendron*, where a ring meristem is present in early development.

It has been hypothesised that plant-pollinator interactions have played an important role in the transitions from radial to bilateral symmetry and is among the drivers of diversification

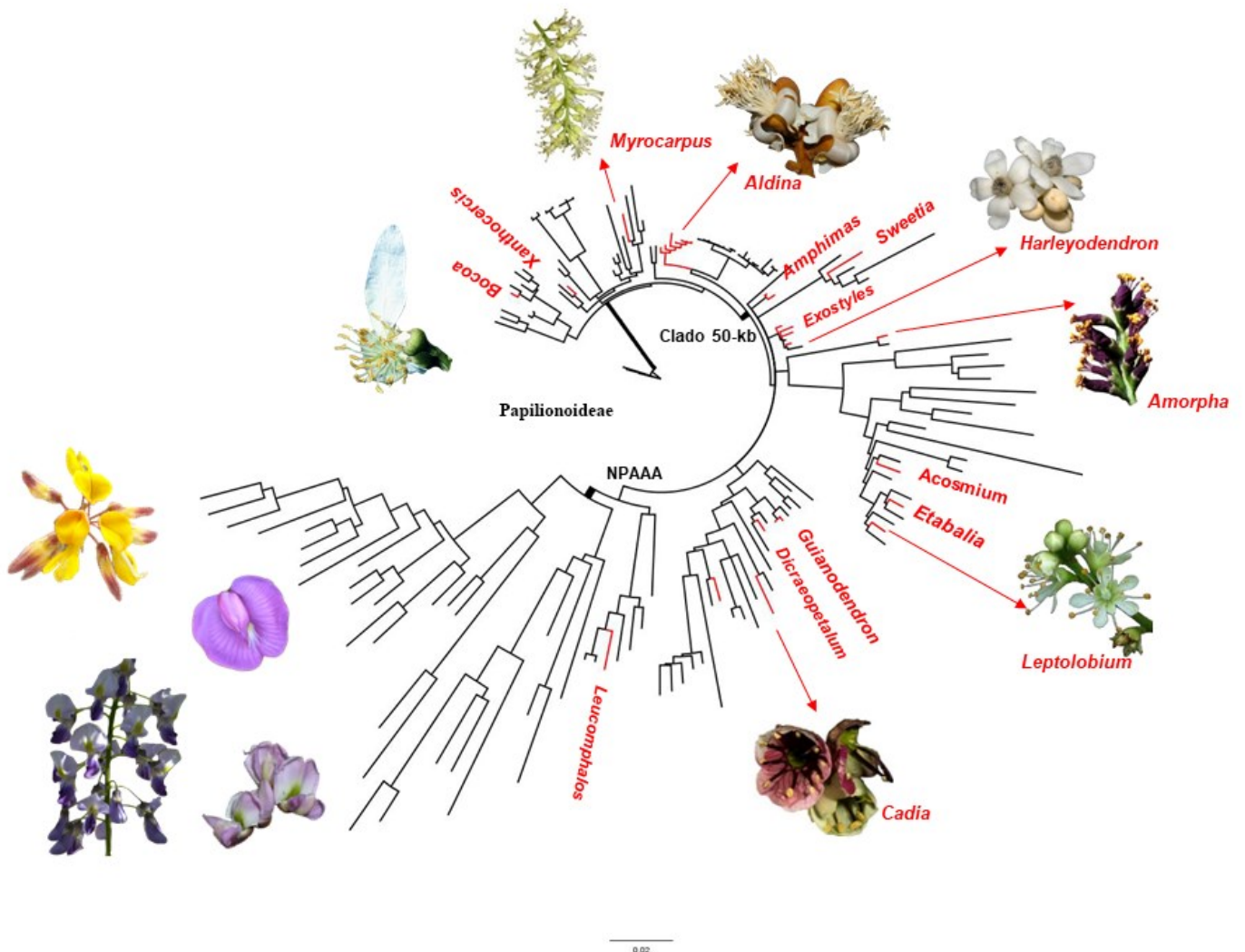


Figure 2. Phylogeny of Papilionoideae based on plastid markers. Genera with unusual non-papilionate flowers are represented by red branches or indicated by red arrows and are phylogenetically placed in clades where other lineages have typical zygomorphic (papilionate) flowers (not shown, but represented by branches in black). The papilionate flower fixed only in the NPAAA clade.

Chapter 1. General introduction and literature review

in species-rich plant clades (Sargent, 2004; Hileman, 2014; Armbruster, 2014; Jiang and Moubayidin, 2022). Radial symmetry is generally thought to be ancestral and bilateral symmetry to have evolved several times independently in the flowering plants, and so is generally assumed to be a derived character state (Hileman, 2014). However, recurrent cases of apparent reversal, from zygomorphic to actinomorphic, have been observed across angiosperms, where radially symmetric species evolved within zygomorphic clades (Pennington et al., 2001; MacMahon and Hufford, 2004; Sinjushin, 2018). Some clear examples of such reversals are seen across the lineages of papilionoid legumes (Leguminosae, Papilionoideae), where not only transitions from zygomorphy to actinomorphy are seen (Mansano et al., 2002; Tucker 2002; Cardoso et al., 2012b; Klitgaard et al., 1999, 2013; Bento et al., 2021), but also dramatic changes where the corolla is reduced to a single petal (Tucker, 1990, 1999, 2003; Leite et al., 2015; Paulino et al., 2013) or no evidence of a corolla is observed (Tucker et al., 2003; Torke & Mansano, 2009; Sinjushin, 2018).

The classification within the subfamily Papilionoideae was originally based on morphological observations. Most papilionoid genera have flowers with 5-petals, 10-stamens, with well-differentiated, strongly bilateral symmetry with a high degree of fused floral organs. In this typical papilionate flower, a dorsal petal is modified into a banner, two lateral petals into wings, and two ventral petals into a keel (often interlocked by claws or free). The androecium is generally diadelphous with nine united filaments and one partially free, and the gynoecium is typically born on a stipe. Such traits found in a typical pea flower were hypothesised to be derived and thought to have evolved from more open, radially symmetrical and other non-papilionate flowers (Polhill 1981, 1994; Tucker and Douglas 1994) that are characteristic, for example of Caesalpinioideae and Detarioideae. Therefore, genera nested in the large non-protein amino acid accumulating (NPAAA) clade of Papilionoideae, where the zygomorphic typical pea flower is predominant, are generally regarded as more morphologically derived relative to those of tribes such as Swartzieae and Sophoreae where genera with more open, radially symmetrical, non-papilionate flowers are more frequent (Klitgaard, 1999, 2013; Pennington et al., 2001; Wojciechowski et al., 2004; Cardoso et al., 2012; Bruneau et al., 2014).

In contrast to the canalised pea-flowered morphology found in major clades nested high in the papilionoid phylogeny such as the NPAAA clade, the early branching lineages of Papilionoideae, which includes tribes Sophoreae and Swartzieae, contain more floral diversity than the remaining papilionoid legumes. Whilst some papilionoids share the typical pea flower

Chapter 1. General introduction and literature review

morphology, others differ in floral symmetry and organ number. This is the case of genera such as *Amburana* Schwacke & Taub., *Bocoa* Aubl., *Cyathostegia* (Benth.) Schrerer, *Candolleodendron* R.S. Cowan and *Trischidium* Tul., which have flowers with a single banner petal (Cowan, 1981; Polhill, 1981; Ireland, 2007). Papilionoid genera that have flowers with divergent morphologies including single petals, numerous free stamens, and even radial symmetry were characterised as primitive and classified in the tribes Sophoreae and Swartzieae (Arroyo, 1981; Cowan, 1981; Polhill, 1981, 1994; Tucker and Douglas, 1994).

Pennington et al. (2001) and Wojciechowski et al. (2004) challenged this point of view using early DNA-based phylogenies of Papilionoideae, providing a starting point for studies involving lineages of papilionoid legumes marked by non-papilionate morphology including unusual floral symmetry. At least nine reversions from a papilionate flower to one with radial or undifferentiated architecture have been identified among the early-branching papilionoids when the polyphyly of Swartzieae, Sophoreae, and Dalbergieae (Pennington et al., 2001). The 23 genera of non-papilionate types listed by Pennington et al. (2000) can be put into two main morphological groups: one with a zygomorphic flower but atypical among papilionoids in having one petal and/or more than ten stamens (e.g. *Amburana*, *Ateleia*, *Bobgunnia*, *Bocoa*, *Cyathostegia*, *Exostyles* and *Swartzia*) and another group with radially symmetrical flowers (e.g. *Acosmium*, *Amphimas*, *Cadia*, *Dicraeopetalum*, *Harleyodendron*, *Holocalyx*, *Lecointea*, *Myrocarpus* and *Zollernia*).

Building on the studies of Pennington et al. (2001) and Wojciechowski et al. (2004), subsequent studies showed that groups that possessed these supposedly primitive traits, such as a radial corolla, unspecialized petals, and numerous free stamens did not form a single group, rather they were found phylogenetically distant from each other throughout the phylogeny of the Papilionoideae (Cardoso et al., 2012a, 2013) within clades essentially comprised by papilionate genera. Additional, more clade-focused studies involved the Amorphoid clade (McMahon and Hufford, 2004), the Lecointeoid clade (Mansano et al., 2004), the vataireoid genera *Luetzelburgia* Harms and *Sweetia* Spreng. (Cardoso et al., 2013a), the Dalbergioid genera *Acosmium* Schott, *Etaballia* Benth., *Inocarpus* J.R.Forst & G.Forst and *Riedeliella* Harms (Cardoso et al., 2012, Klitgaard et al., 2013), and the genistoid genera *Cadia* Forssk (Boatwright et al., 2008), *Guianodendron* Sch.Rodr. & A.M.G Azevedo and *Leptolobium* Vogel (Cardoso et al., 2012b).

Chapter 1. General introduction and literature review

Collectively, these phylogenetic studies have therefore over-turned the premise (Arroyo, 1981; Polhill, 1981; Tucker & Douglas, 1994) that non-papilionate flowers represent a signature of antiquity in the lineages of Papilionoideae (Pennington et al., 2001; Cardoso, 2012). Rather, a high integration of floral parts in the papilionate flower has been a more evolutionarily labile character state than previously considered (Ireland et al., 2007; Klitgaard et al., 2013; Bruneau et al., 2014; Cardoso et al., 2012a, 2013, 2015).

The heterogeneity in flower morphology in many clades of the early branching papilionoids suggests that changes in floral symmetry could involve key genes involved in the establishment of flower symmetry (Citerne et al., 2006; Feng et al., 2006). These genes may be transcription factors (TFs) involved in flower development because they are important elements mediating the transition between vegetative and floral meristems, organ identity, and flower symmetry (Heisler et al., 2022; Viola and Gonzalez, 2023).

1.3 Flower identity: a matter of expression

Transcription factors (TFs) are proteins that control the rate of transcription of genetic information from DNA to messenger RNA by binding to a specific DNA sequence. The main function of a TF is to regulate genes so that they are expressed in the target cells at the right time, place and in the right amount throughout the development of a certain tissue. In plants, TFs control many biological processes of development through binding to promoter regions and activating or suppressing transcription of their target genes (Boyden et al., 2012; Raimundo et al., 2013). This is especially true of the network of genes involved in the development of flowers which are encoded by large multigenic families, which by definition is a group of genes descended by duplication and variation from a common ancestral gene and that share similar functions (Corley et al., 2005; Raimundo et al., 2013).

One such multigenic family is the TCP family, which is named after its first four characterised members, namely TEOSINTE BRANCHED (TB1) from maize (*Zea mays*), CYCLOIDEA (CYC) from snapdragon (*Antirrhinum majus*), and the PROLIFERATING CELL NUCLEAR ANTIGEN FACTOR1 (PCF1) and PCF2 from rice (*Oryza sativa*). The TCP family constitutes a plant-specific group of regulators involved in the development of plants, with members being expressed early in seed germination throughout vegetative development until the formation of flowers and fruits.). TCP proteins can be classified into two types, class I (TCP-P) and class II (TCP-C), with the latter being divided into two subgroups, namely CINCINNATA (CIN) and angiosperm-specific CYC/TB1 (Fig. 3) (Li, 2015).

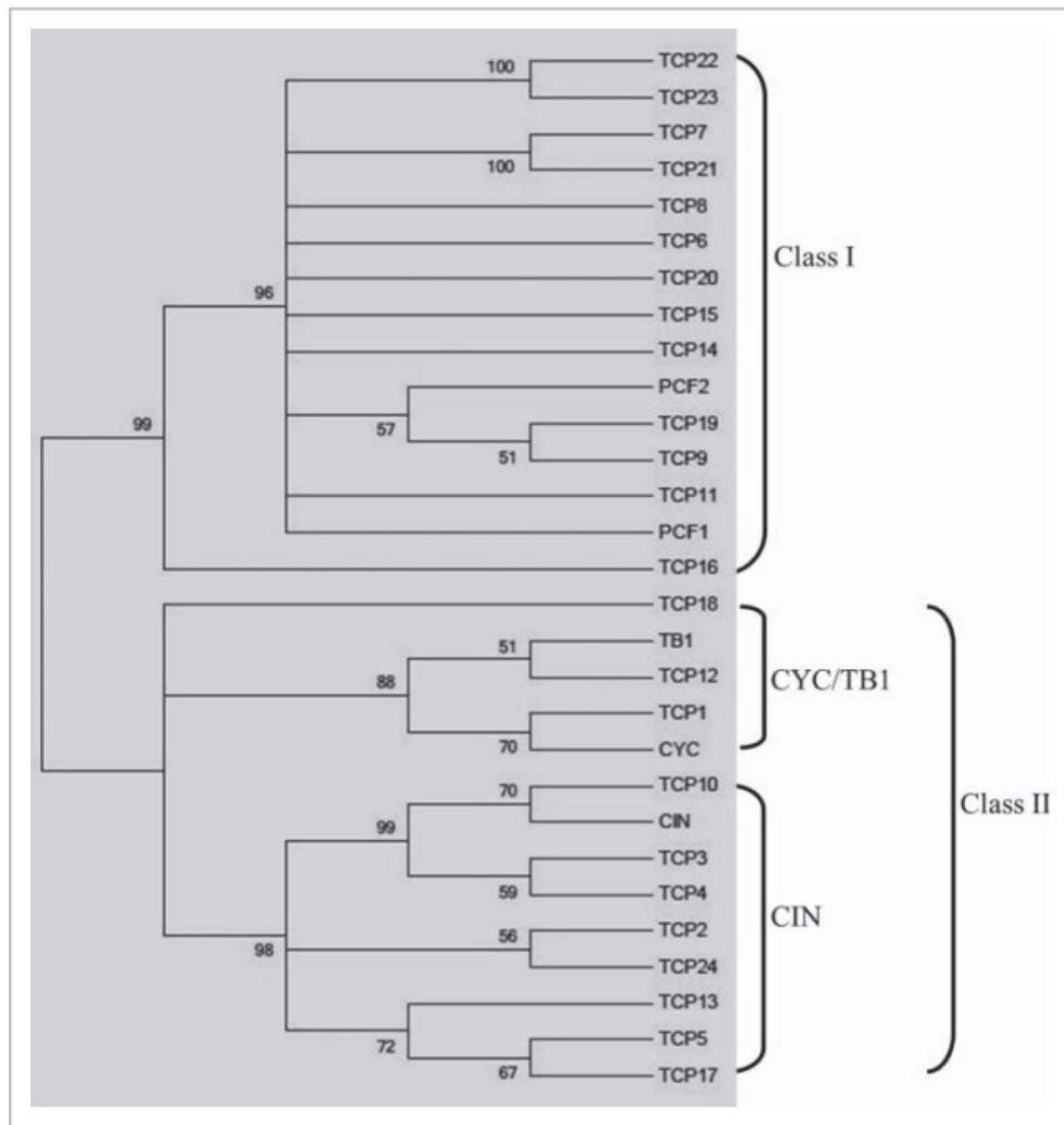


Figure 3. A phylogenetic tree of the *Arabidopsis* TCP protein family with different classes (I and II) and subgroups (CYC/TB1 and CIN). Adapted from Li 2015.

CYCLOIDEA TFs are remarkably ubiquitous in plant biological processes and are commonly involved in the expression or suppression of genes regulating flower development (Corley et al., 2005; Wang et al., 2008). The molecular structure of CYC involves a basic-Helix-Loop-Helix (bHLH) domain referred to as the TCP domain, which, along with an R domain, is conserved across the gene family (Martin-Trillo & Cubas, 2010). The TCP domain is necessary for specific binding to promoter elements whilst the R domain is generally involved in protein interaction (Cubas, 1999). Spatially, CYC genes are typically expressed in the adaxial domain of the developing meristem, enabling or antagonising the expression of other genes during organ development (Gaudin et al., 2000; Martin-Trillo & Cubas, 2010; Viola & Gonzalez, 2023).

CYC-like genes have been studied in different plant families, shedding light onto the evolutionary history of this class of genes. In core eudicots, for example, CYC-like genes have undergone two major duplication events and given rise to three clades, CYC1, CYC2, and CYC3 (Citerne et al., 2013; Howarth & Donoghue, 2006; Zhao et al., 2019). Each of these clades has undergone several gene duplications within many angiosperm clades (Carlson et al., 2011; Citerne et al., 2013; Martin-Trillo & Cubas, 2010; Spencer & Kim, 2017; Tahtiharju et al., 2012; Zhang et al., 2010). In legumes, CYC-like genes have experienced one duplication event before the evolution of Papilionoideae, giving rise to the copies *LEGCYC1* and *LEGCYC2* (Citerne et al., 2006). Within group 1, one another major duplication has occurred, giving rise to the copies *LEGCYC1A* and *LEGCYC1B* (Citerne et al., 2003) (Fig. 4). Relationships between sequences within these groups are complex mostly because of lack of resolution in phylogenies (Citerne et al., 2003).

Another important gene family with members involved in plant development is the myeloblastosis (MYB) TF family, found in all higher plant genomes, and comprising one of the largest TF families in flowering plants with number of genes varying from 157 in maize (Du et al., 2012a) to 252 in soybean (Du et al., 2012b). MYB genes regulate biological processes in a wide range of organisms, including viruses and animals. In plants, members of this gene family also play an important role in the establishment of flower symmetry in angiosperms and interact with members from the TCP family, either enabling or suppressing the formation of new proteins (Reardon et al., 2014; Hsin and Wang, 2018; Valoroso et al., 2019; Madrigal et al., 2019; Lucibelli et al., 2020).

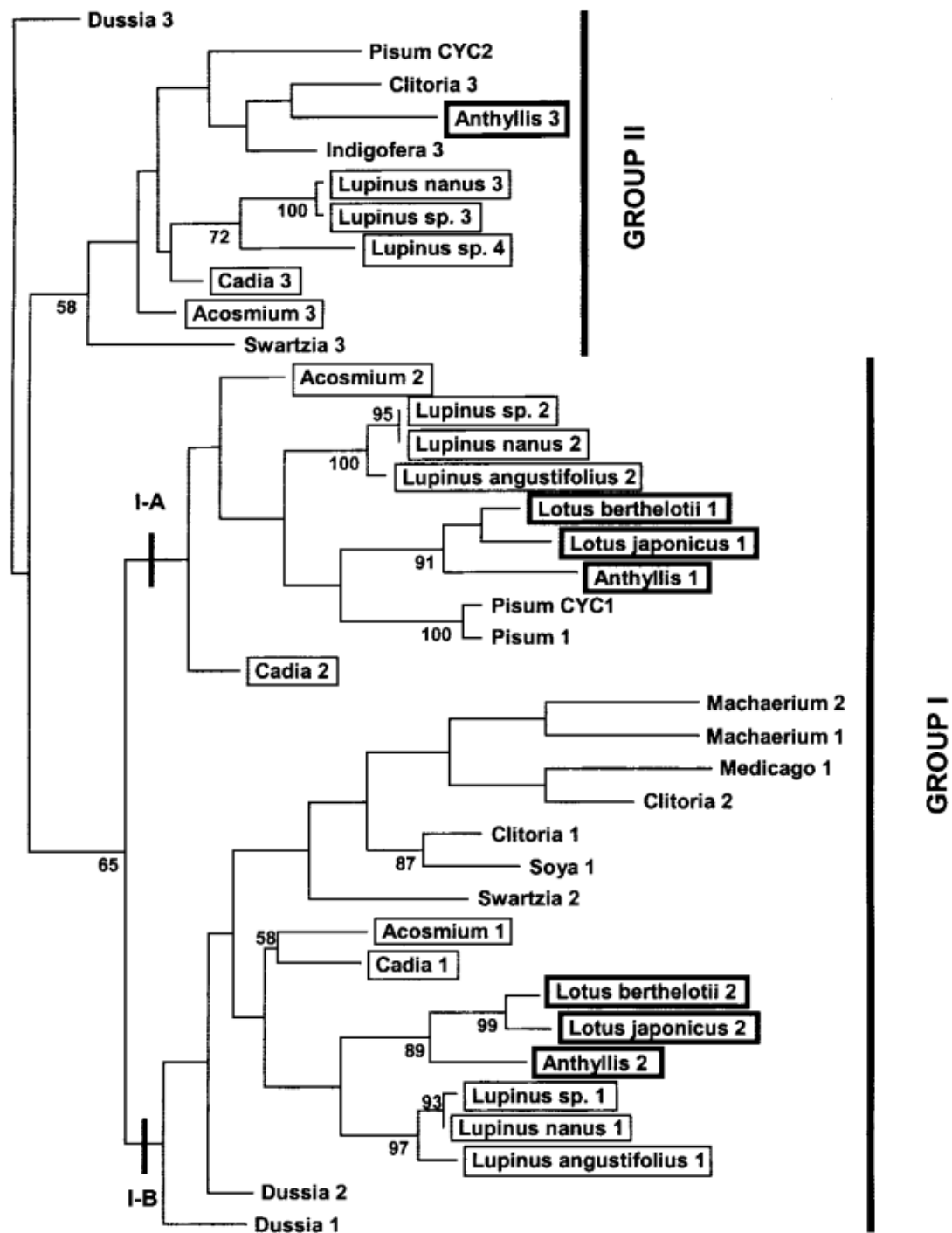


Figure 4. Unrooted phylogram of Maximum likelihood analysis of partial legume CYC sequences. The two main groups of *LEG*CYC (I and II) are highlighted, and one putative duplication event in group I is marked by A and B. Citerne et al. (2003).

Chapter 1. General introduction and literature review

The first comprehensive analysis of the molecular pathway underlying floral symmetry conducted with the snapdragon *Antirrhinum majus* L. identified four genes, *CYCLOIDEA* (*CYC*) and *DICHOTOMA* (*DICH*), and MYB, *RADIALIS* (*RAD*) and *DIVARICATA* (*DIV*), involved in the determination of flower symmetry (Corley et al., 2005). The floral asymmetry found in *A. majus* therefore involved an interplay between TCP and MYB transcription factors, where a pair of TCP (*CYC* and *DICH*) proteins are expressed in the dorsal region of the floral meristem, and activate *RAD* (MYB) in the dorsal domain. The *RAD* protein then antagonises the related *DIV* (MYB) protein, preventing *DIV* activity in dorsal regions (Fig. 5).

Subsequent studies confirmed a strong interplay between these genes and its correlation with the establishment of symmetry in different groups, such as Orchidaceae, Asteraceae, Plantaginaceae, Aristolochiaceae and Gesneriaceae (Galego and Almeida, 2002; Reardon et al., 2014; Gao and Zhang, 2017; Bello et al., 2017; Madrigal et al., 2019; Valoroso et al., 2019; Lucibelli et al., 2020). Studies involving individual genes in papilionoid legumes showed that *CYC* orthologues play a key role in determining dorsal identity in two species, *Lotus japonicus* and *Cadia purpurea* (Feng et al., 2006; Citerne et al., 2006).

In *Lotus* and *Cadia*, differential expression of *CYC*-like genes plays a role in the determination of organ identity. For example, in *C. purpurea* (actinomorphic) and *Lupinus nanus* (zygomorphic), two orthologues (*LegCYC1A* and *LegCYC1B*) of the floral symmetry gene *CYC* were found using degenerate Polymerase Chain Reaction (PCR) primers and cloning (Citerne et al., 2006). Even though *C. purpurea* shares in common the set of genes involved in petal differentiation with *L. nanus*, the radial symmetry of *C. purpurea* is the result of the overexpression of *LegCYC1B*, which is responsible for the development of a petal with banner morphology throughout the lateral and abaxial domains (Citerne et al., 2006). The atypical flowers of *C. purpurea* had frequently been considered as primitive, but phylogenetic studies suggested that they derive from relatives with zygomorphic or papilionate flowers (Pennington et al., 2001; MacMahon & Hufford, 2004).

Although differential expression of genes is a key factor determining the symmetry of morphologically distinct genera, mutations also play a role in the determination of flower types in legumes. Feng et al. (2006) conducted an experiment in *Lotus japonicus* where induced mutants had abnormal flower symmetry.

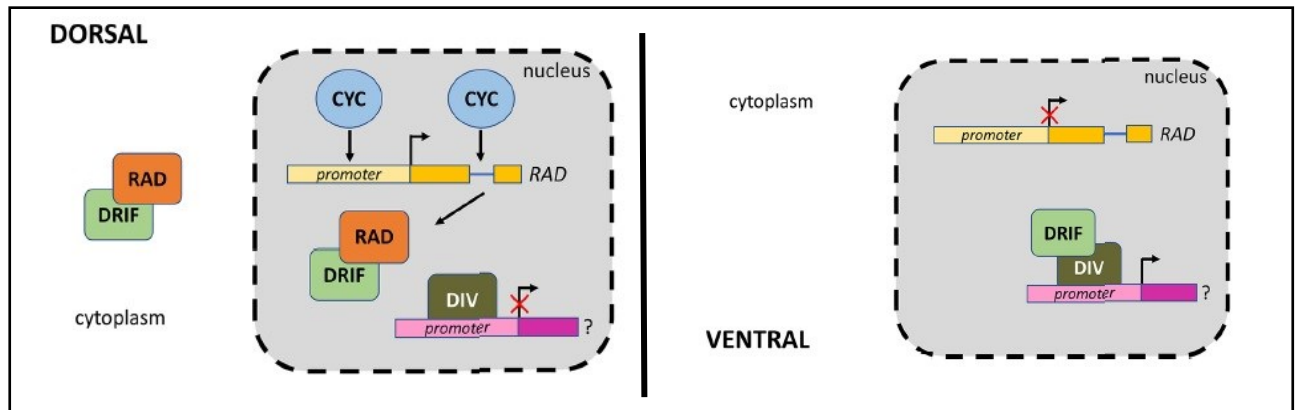


Figure 5. Model of interaction of CYC, DIV, RAD, and DRIF proteins in *Antirrhinum majus*. The DIV and DRIF proteins are expressed both in the dorsal and ventral parts of the flower. In the dorsal domain, the CYC protein interacts with the promoter and intron of the RAD gene, activating its expression. The RAD protein binds to DRIF and prevents the formation of the DIV/DRIF complex, thus inhibiting ventralization. In the ventral domain, the absence of RAD permits the formation of the DIV/DRIF complex that controls downstream genes involved in the ventralization. Valoroso et al., 2019 (modified from Raimundo et al., 2013).

A CYC-like mutant of *LjCYC2*, *squared standard 1 (sqt1)*, was identified and found to be responsible for an abnormal shape of the standard petal (squared rather than rounded). The mutant *sqt1* specifically affects the development of the standard petal without other major changes in other organs, apart from affecting the epidermal cell shape. The cells do not all have the typical conical morphology typical of the standard petal, but include jigsaw puzzle-shaped cells characteristic of the lateral petals in the wild type, indicating a correlation between change in the overall morphology of the dorsal petal and the reduction of dorsal identity. Modifications of petal micromorphology, such as changes in cell type, can have drastic implications to the biology of the species, such as shifts in the pollination system (Ojeda et al., 2017).

Moreover, *KEW1 (LjCYC3)*, a putative orthologue of the gene K in *Pisum sativum* (Pellew and Sverdrup, 1923), was also identified in *L. japonicus* with a specific role in the control of lateral petal identity (Feng et al., 2006). Mutations in *Kew1* led to keel-like lateral petals, a partial abaxialisation of the flower (Feng et al., 2006). *Kew1* was confirmed as a CYC gene after the study with *Pisum* that showed that the gene K resulted in modification of the lateral wings which resembled the keel petals (Wang et al., 2008). Double mutants of *LjCYC2* and *Kew1* had all ventralized petals and a much stronger phenotype than either individual mutation, indicating that the interaction of these two genes plays an important role in the

establishment of the dorsoventral axis, and ultimately in the highly specialised zygomorphy of the flower in both species, similar to the tug of war involved in the regulation of the dorsoventral axis in *Antirrhinum* via *RAD* and *DIV* interaction (Corley et al., 2005; Wang et al., 2008).

1.4 Conserved domains characterise TCP gene family

The TCP domain is the main molecular signature of the members in the TCP family with its non-canonical bHLH. TCP domains form two types of proteins, based on their differences within their TCP domains: one defined by *CYC* and *TB1* (also known as TCP-C class), and another defined by the PCFs (also known as TCP-P class) (Mártin-Trillo and Cubas, 2009). The molecular characterization of the TCP family involves conserved features such as two short stretches of residues in the basic region, hydrophobic residues along the apolar face of both α -helices, a tryptophan in helix II, and a helix-breaking glycine in the loop between the helices (Fig. 6). However, the residues in the loop and the hydrophilic residues of the helices are not as well conserved.

In addition to the bHLH domain, *CYC* and *TB1* have a second conserved region termed the R-domain (Fig. 6), which is rich in polar residues (arginine, lysine and glutamic acid). The R-domain is also predicted to form a hydrophobic α -helix. Both the bHLH and R domains are also predicted to form coiled coils, similar to those formed by leucine zippers (Cubas et al., 1999).

Another important TCP gene involved in the establishment of flower symmetry is *DICHOTOMA* (*DICH*). *DICH* was identified as a homologue of *CYC* in the snapdragon with a partially redundant function in the establishment of flower symmetry (Howarth and Donogue, 2006). Both genes are expressed in the dorsal region of the floral meristem throughout its development and maintaining the wild zygomorphic phenotype (Raimundo et al., 2013; Valoroso et al., 2019). Double mutants of these genes result in completely radially symmetric flowers with the corolla pieces resembling the ventral petal (a ventralized phenotype). *CYC* and *DICH* expression is also detectable prior to the development of floral organs at the junction of the inflorescence and floral meristem (Corley et al., 2005), indicating that these genes may also play a role in the early arrangement of the cells from the floral meristem. After all organs are initiated, their expression is limited to the two dorsal petals and staminode, with *DICH* having a more restricted expression in the dorsal half of the petals (Da Luo et al., 1999).

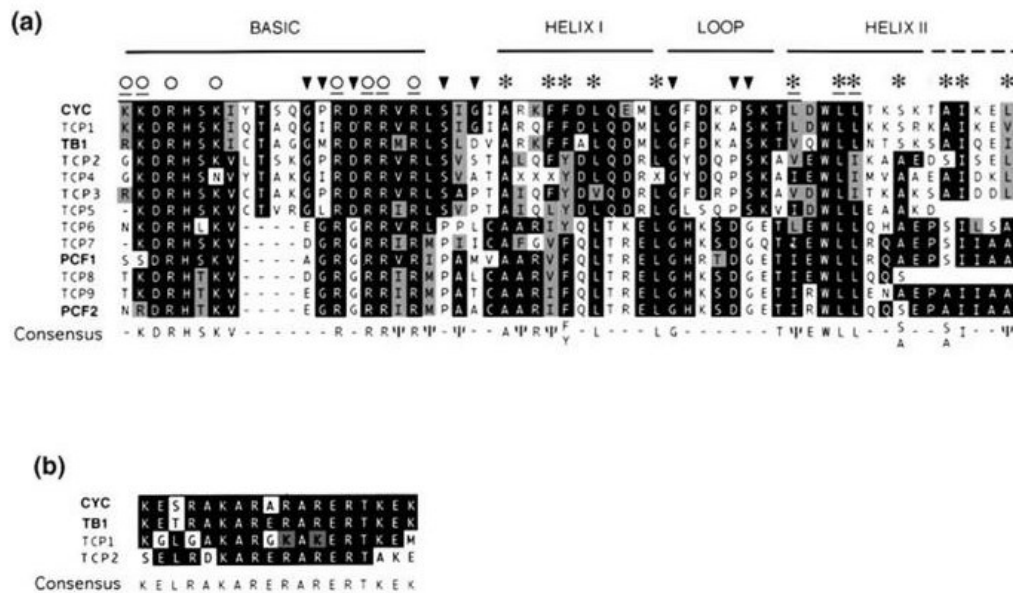


Figure 6. a- Representation of basic helix-loop-helix structure of the TCP domain with amino acids from members of the TCP family. Founder members (i.e. *Antirrhinum*, *Zea mays* and *Oryza sativa*, in order from top to bottom) of the TCP family are indicated in bold. Adapted from Cubas et al. 1999. **b-** Alignment of the R-domain in the CYC/TB1 subfamily. Adapted from Cubas et al. 1999

Although *CYC* and *DICH* partially overlap in function in the floral meristem of *A. majus*, their expression domains differ, and each also has a unique role, consistent with subfunctionalization following duplication, which is common in these cases (Gubtiz et al., 2003). Redundancy of *CYC* activity has also been reported in the zygomorphic papilionate-flowered genus *Lupinus*. The two *CYC* copies, *LegCYC1A* and *LegCYC1B*, were detected in the floral tissue of *Lupinus* in a similar pattern to *Antirrhinum*. Like in the snapdragon, both genes were also detected in floral meristems before organogenesis, on the adaxial side of the meristem (Citerne et al., 2006). Although *LegCYC1A* and *LegCYC1B* expression overlap in the adaxial part of the floral meristem, suggesting functional redundancy, *LegCYC1A* has a reduced expression domain when compared to *LegCYC1B*. In more advanced developmental stages, both copies are transcribed not only in the standard petal (dorsal region) but also in the dorsal anthers. These results were confirmed by RT-PCR and *in situ* hybridization experiments (Citerne et al., 2006).

Chapter 1. General introduction and literature review

In the radially symmetric *Cadia*, *LegCYC1A* has a moderate to weak expression limited to the dorsal petal in more advanced developmental stages, whereas *LegCYC1B* is expressed in all petals, showing an expansion of the expression domain to the lateral and ventral petals. Neither copy is expressed in the androecium of *Cadia*, unlike in *Lupinus*, which expression is noticed in more advanced stages in the standard petal and in the dorsal anthers (Citerne et al., 2006).

1.5 MYB gene family

MYB genes are part of a large gene family of transcription factors and they contain two conserved DNA-binding domain repeats R2-R3 DNA at their N-termini (Fig. 7). *MYB* genes play a key role in the development of plants, such as in secondary metabolism, hormone signal transduction, cell shape, organ development, and disease response (Du et al., 2009).

MYB genes are divided into five subfamilies, which include two important genes involved in the determination of flower symmetry: *RADIALIS* (*RAD*), in the I-box-like subfamily and *DIVARICATA* (*DIV*), in the R-R-type subfamily (Boyden et al., 2012). Studies in *Arabidopsis* showed that these lineages appear to be sister groups and were previously put together under the I-box-like subfamily based upon the similarity of *DIV*'s first *MYB* domain to the single *MYB* domain of *RAD* (Riechman and Ratcliffe, 2000). In a study conducted in Dipsacales, it was pointed out that the origin of I-box-like genes, such as *RAD*, were the result of the loss of the second *MYB* domain found in R-R-type genes such as *DIV* (Stevenson et al., 2006).

Gene duplication is frequent during the evolution of *RAD*- and *DIV*-like genes in Dipsacales. Three paralogs of the *RAD* lineage, *RAD1*, *RAD2*, and *RAD3*, as well as three paralogs of the *DIV* lineage, *DIV1*, *DIV2*, and *DIV3*, are recognized (Boyden et al., 2012). The gene duplications that gave rise to these paralogs were predicted to have occurred around the diversification of the Pentapetalae. Consistent results have shown that both the ECE clade of *CYC*-like TCP genes and the core R-R-type clade of *DIV*-like genes apparently duplicated twice before the diversification of the Pentapetalae (Howarth and Donoghue, 2006, 2009).

MYB genes have been well studied in other angiosperms, such as the model *Antirrhinum majus* (Luo et al., 1995; Almeida et al., 1997; Luo et al., 1999; Galego and Almeida, 2002; Corley et al., 2005; Stevenson et al., 2006). It was observed that the floral

symmetry of *A. majus* is produced by an orchestrated interplay of these genes, each of which are expressed in different domains in wild specimens, resulting in different patterns of symmetry in mutants (Corley et al., 2005). *RAD* promotes dorsal identity and *DIV* promotes ventral identity on the flowers of *Antirrhinum*. Antagonism involving these MYB-like proteins was found to be a mechanism regulating zygomorphy in the flowers of *Antirrhinum*. *RAD* proteins have an antagonistic effect on the formation of the DIV/DRIF (DIV and RAD interacting factor) complex depending on the presence or absence of *CYC* transcription factors. In the absence of *RAD*, the expression of *DIV* is expanded to the lateral and dorsal domains, resulting in a radially symmetric flower (peloric phenotype) (Corley et al., 2005; Raimundo, 2013; Lucibeli et al., 2020).

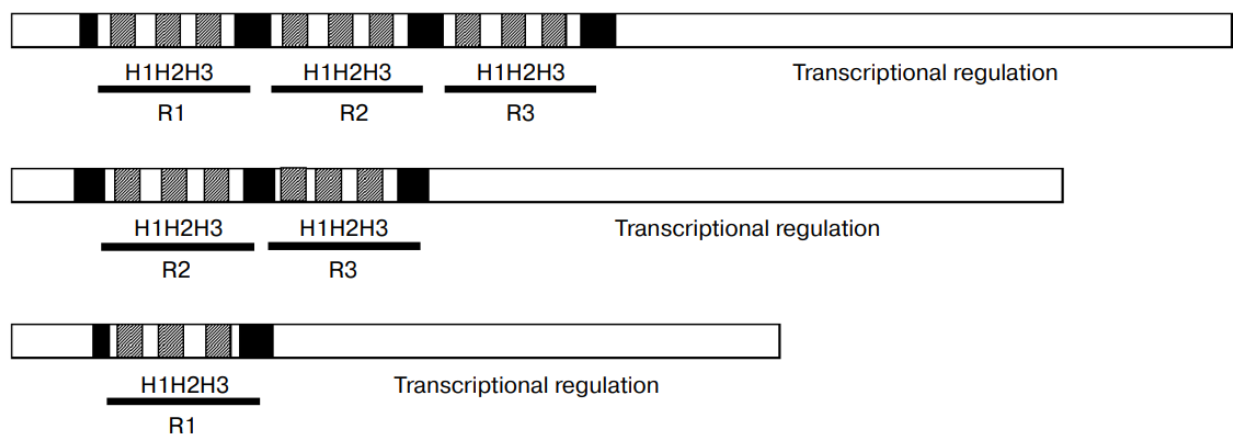


Figure 7. Representation of functional domains of the MYB protein. Shaded boxes represent the most conserved DNA binding domain in MYB proteins; R1, R2, and R3 are repeats of DNA binding domains; crosshatched/hatched bars indicate three α -helices in MYB R1, R2, and R3. Each repeat is about 50-53 amino acids long and encodes three α -helices, with the second and the third helices forming a helix-turn-helix (HTH) structure when bound to DNA. Adapted from Du et al. 2009.

As discussed in the section above, *CYC*-like genes independently play a role in the establishment of dorsoventral and internal organ asymmetries in legumes (Wang et al., 2008) but little is known about *MYB* genes. Therefore, knowing the key role of *DIV* and *RAD* in establishing symmetry in other groups, investigating *MYB* genes in typical and non-papilionate genera might reveal insights into the relevance of these genes in the establishment of different types of flower symmetry in the early-branching lineages of Papilionoideae.

1.6 Thinking beyond the models: the need for studies of transcription factors in papilionoid legumes with unusual flowers

Transcription factors, such as the ones involved in floral symmetry regulation, have been extensively investigated in many angiosperms. In legumes, these studies have been more focused on model species from the subfamily Papilionoideae and in economically important crops from which genomic data is widely available (Wang et al., 2008; Feng et al., 2008; Singh et al., 2013; Singh and Jain, 2014). The vast majority of these studies include legume species with limited morphological variation in the typical zygomorphic papilionoid flower and there is therefore limited knowledge of the genetic basis underlying the differences in the flower morphology found in papilionoid taxa with non-papilionate flowers (Citerne et al., 2003, 2006, 2010). This leaves a gap in the knowledge for all the other groups in the early branching lineages of Papilionoideae with morphological differences in their flowers. Revisiting the broad scope of floral diversity found in the early branching lineages of Papilionoideae through a genetic lens will help to understand how floral form evolved in this large family and in the context of the angiosperms in general.

In Chapter Two I investigate the genetic basis underlying the changes found in the flower morphology of four morphologically distinct non-model species from the early branching lineages of Papilionoideae; *Andira anthelmia* (typical zygomorphic papilionate), *Andira inermis* (typical zygomorphic papilionate), *Cadia purpurea* (radially symmetric, non-papilionate), and *Cyathostegia mathewsii* (zygomorphic, non papilionate with one petal). I use a candidate gene approach in order to identify putative *TCP* and *MYB* orthologues that may be recruited to the control of the development of flower symmetry. Newly generated genomic data from the target species is analysed and I investigate the correlation between different gene copy numbers and distinct morphology types, as well as the phylogenetic events that led to duplication/deletions of symmetry genes in these lineages.

1.7 Expression of symmetry genes in legumes

As discussed above, the investigations of *CYC* genes carried out in the radially symmetric *Cadia* by Citerne et al. (2003, 2006, 2010) clarified the importance of symmetry genes in legumes. *In situ* hybridization experiments have shown the activity of *CYC* genes throughout flower development in *Cadia*, providing evidence that, rather than a loss or gain of symmetry genes, the different levels of gene expression play a role in the determination of

Chapter 1. General introduction and literature review

flower morphology in papilionoid legumes with non-typical flower morphologies (Citerne et al., 2006).

With the advance of Next Generation Sequencing (NGS) technologies, including RNA-sequencing (RNA-seq), it has become easier to assess and quantify the abundance of genes expressed in specific tissues through the generation of whole transcriptomes. Targeted genes as well as novel genes can now be identified in specific tissues in different developmental stages without the technical complexity of in-situ hybridization experiments. Moreover, this is a convenient approach to study differentially expressed genes and to predict putative gene function (Wang et al., 2009; Ozsolak and Milos, 2011). One advantage of working with RNA-seq data is that a reference genome is not mandatory, since the generation of whole transcriptomes can be achieved by the *de novo* assembly of reads. That is often the case when working with non-model species, for which reference genomes are not normally available (Martin and Wang, 2011).

In Chapter Three I analyse the differential expression of the genes *CYC*, *DIV*, *RAD*, and related genes in vegetative and floral tissues. I investigate whether the expression of *CYC* and related genes correlates with variation in floral symmetry found in the radially symmetric *Cadia purpurea* (vegetative buds, young flower buds, mature flowers, and abaxial and adaxial regions), the zygomorphic papilionate *Andira inermis* (vegetative buds), and the zygomorphic single-petal *Cyathostegia mathewsii* (vegetative buds, young flower buds). Subsidiary questions are: what are the genes which show variation in expression across the dorsoventral axis in the flowers? Is the variation in floral morphology found in these species reflected in variation in gene expression patterns?

1.8 Studies in flower ontogeny

Investigating the genetic processes underlying the changes seen in distinct morphologies is key to understanding how different flower types emerged in the evolution of Papilionoideae. Coupling gene function to morphological ontogenetic data can provide stronger evidence of gene activity at specific points in space and time throughout flower development. Thus, in Chapter Four, I analyse the ontogeny of unusual floral morphologies across the early branching Papilionoideae.

Chapter 1. General introduction and literature review

The early branching lineages of Papilionoideae, for example, have several cases where the phylogenetic relationships of genera with unusual floral morphologies have been resolved (see discussion above) but no ontogenetic study has been carried out for those lineages. In addition to generating hypotheses about the genetic control of morphologies, detailed studies of ontogeny can help us to understand how these phylogenetic relationships can be supported from a developmental perspective. Floral development evidence has shed light on character homology in morphologically heterogeneous lineages such as those in early branching papilionoid legumes (Tucker, 2003; Prenner, 2004; Zimmerman et al., 2013; Bruneau et al., 2014; Prenner et al., 2015; Prenner & Cardoso, 2017).

The genus *Acosmium* Schott is an example that has previously encompassed five lineages, based on the superficial similarity of radially symmetric corollas, now recognised to be phylogenetically distant (*Dicraeopetalum* Harms, *Guianodendron* Sch.Rodr. & A.M.G. Azevedo, *Leptolobium* Vogel and *Sweetia* Spreng.). Cardoso et al. (2012) confirmed and demonstrated that radial flower symmetry evolved independently in all *Acosmium* segregates that are now recognised as separate genera. One of these segregates, the radially symmetric *Leptolobium*, has the nearly papilionate *Bowdichia* as a sister group, making an intriguing pair for which ontogenetic studies are lacking.

Klitgaard et al. (2013) investigated flower evolution in the morphologically heterogeneous Pterocarpus clade including four genera with radially symmetric flowers (*Acosmium*, *Etabalia*, *Inocarpus* and *Riedeliella*). This study concluded that actinomorphy evolved independently four times in the Pterocarpus clade. Ontogenetic studies have been carried out for some of these genera (Klitgaard, 1999; Klitgaard et al., 2013; Sinjushin, 2019; Bento et al., 2021). By comparing the floral ontogeny of the radially symmetric *Riedeliella* and its typical zygomorphic papilionate sister *Discolobium*, Bento et al. (2021) observed differences between the two genera from early stages in petal development, where *Riedeliella* developed unidirectional growth and *Discolobium* petal differentiation initiated by the petal standard, followed by the keel petals and finally by the wing petals.

Floral symmetry of both genera is expressed at different times throughout development, with the zygomorphy of *D. pulchellum* being expressed earlier compared to the actinomorphic symmetry of *R. graciliflora*, which occurs at a later developmental stage. Bento et al. also concluded, based on comparison with other unusual papilionoid genera with radially symmetric flowers, that the development of *R. graciliflora* involved several unique stages. No

Chapter 1. General introduction and literature review

ontogenetic data for the other members of the *Pterocarpus* clade has yet been investigated. Further ontogenetic comparison with the remaining radially symmetric *Inocarpus*, *Acosmium*, and *Etaballia* can shed light on the homology of these flowers and allow hypotheses on the genetic process underlying flower development in the clade.

Another example is found in the *Andira* clade. Ramos et al. (2016) found the phylogenetic placement of the radially symmetrical genus *Aldina* as the sister group to the typical zygomorphic papilionoid *Andira* and *Hymenolobium*. Although florally distinct, *Aldina* and *Andira* have similar drupoid legumes and the three genera share similar habitat preferences.

Collectively, these studies have contributed to our knowledge of the phylogenetic relationship of these lineages and elucidated discussions on the evolution of unusual flower types among the typical zygomorphic papilionoid lineages and there is a clear need for more ontogenetic study comparing distinct morphologies in these phylogenetic contexts.

Radially symmetric flowers have evolved recurrently across the phylogeny of Papilionoideae (Boatwright et al., 2008; Cardoso et al., 2012; Klitgaard et al., 2013; Ramos et al., 2016), however other unusual types of flower morphologies are also found across the early branching papilionoids. Leite et al. (2015) showed that in *Amburana cearensis* (Papilionoideae) the single petal corolla is the result of the abortion of four petal primordia in their early stages of development, demonstrating that, although apparently distinct in morphology from the papilionate flower, the genus still retains floral traits in its development consistent with the rest of the subfamily. From a genetic perspective, it seems possible that suppression of development of lateral and ventral domains in *Amburana*'s unusual single-petal flower may be the result of suppression of the expression of genes involved in the development of these domains, since the primordia of those petals are discernible in their early stages of development. In this scenario, studies using RNA-seq can potentially shed light on the activity of floral genes in *Amburana* flower. This approach would show the start and end of activity of putative floral genes in time and space throughout flower development.

Remarkably, in some legume lineages the whole whorl of petals can be lost, for example in the African-Madagascan genus *Cordyla*. In *Cordyla pinnata* (Sinjushin et al. 2018), no evidence of corolla initiation is visible at any stage, characterising the absence of petal primordia from their initiation. This species bears numerous stamens and a gynoeceum that is sporadically multicarpellate (Sinjushin, 2018). A polymerous androeceum and multicarpellate gynoeceum are associated with a polysymmetric perianth in some papilionoid genera (Mansano

Chapter 1. General introduction and literature review

et al., 2002; Sinjushin, 2018). These species tend to have an increased number of organs in different domains and are more prone to merism fluctuations (Sinjushin, 2018).

Tucker (2003) presented a seminal revision on floral development in legumes, which is after 20 years still a reference to understand the ontogenetic patterns in the whole family. Because of the dramatic morphological diversity of flowers in the early branching lineages of Papilionoideae, it has become a challenge to interpret the morphological evolution of flowers in the subfamily, and new reviews are needed. Thus, in Chapter Four I compare the ontogenetic development of a series of papilionoid legumes with contrasting morphologies with particular focus on the evolution of androecium symmetry and petal growth. The aim of this study is to revisit the floral development across the Papilionoideae and find ontogenetic insights to shed light on the stabilisation of the typical papilionate flower in the most derived clades and hypothesise about the dramatic morphological changes observed in the early-branching lineages of the subfamily. Data from previously published ontogenetic studies such as from the radially symmetric *Cadia purpurea* (Tucker, 2002), the typical papilionate *Petaladenium urceoliferum* (Prenner et al., 2015) and the single-petal Swartzoid *Cyathostegia mathewsii* are used together with unpublished data generated by Dr. Gehard Prenner and Dr. Domingos Cardoso from several papilionoid species spanning from the early branching lineages to the most derived ones, including the megadiverse NPAAA clade.

1.9 Taxonomic studies of Aldina, a genus of Amazonian papilionoid legume trees with radial flowers

The final chapter of this thesis is a taxonomic account of the Amazonian papilionoid legume *Aldina*, which is a genus of Papilionoideae with unusual non-papilionate flower morphology. Flowers of *Aldina* are radially symmetrical, with free numerous stamens, unlike its sister genera *Andira* and *Hymenolobium*, whose zygomorphic flowers are typically papilionate and with 10 stamens. *Aldina* taxonomy is in need of a revision since the preliminary taxonomic studies on the genus that date back 70 years (Cowan, 1953, 1958, 1961, 1984; Gontscharov, 2006, 2009). Studying the taxonomy of *Aldina* is a way to understand the morphological boundaries of the genus and establish the delimitation of its species. This chapter therefore contributes to understanding the species diversity in Amazonia, filling in the gaps of knowledge of plant diversity in such unexplored regions.

Cataloguing and describing plants in species-rich regions such as the vast tropical Amazon lowland rainforest is fundamental to understanding patterns of diversity and are the

Chapter 1. General introduction and literature review

first step toward conservation prioritisation (Hopkins, 2007; Cardoso et al., 2017; Ramos et al., 2022). However, much taxonomy remains to be done in Amazonia, partly because vast remote regions are unexplored due accessibility constraints (Cardoso et al., 2015a), and when species exist, they can remain hidden for decades in herbaria without being studied by taxonomists (Bebber et al., 2010). The role of taxonomists in monographing poorly collected genera is fundamental to fill in the gaps of plant diversity. Species number estimates based on models can give a broad idea of species richness, but when the lists accessed are not taxonomically verified, they can erroneously lead to unrealistic numbers. For example, recent estimates have shown that the Amazon basin host up to 50,000 plant species, depending on which model is used and how the region is defined (Hubbell et al., 2008). Of these, between 6,000 and 16,000 species are predicted to be trees (Hubbell et al., 2008; ter Steege 2016) based upon models using data from inventory plots.

However, such ecological datasets are challenged by the high percentage of incorrectly named individuals and outdated species names (i.e, synonyms). Cardoso et al. (2017) showed that taxonomically verified lists that take into account taxonomic specialist knowledge and the most recent taxonomic revisions reduce those numbers to 14,003 species of which 6,727 are trees, representing ~11% of the 60,065 tree species estimated to occur worldwide (Cardoso et al.2017; Beech et al., 2017). These numbers may represent a better approximation to the reality of plant diversity in the Amazon and are more reliable due to taxonomically verified list of species published in taxonomic revisions, floras and checklists. In the face of accelerating habitat destruction, thorough monographic taxonomic revisions of plant groups in poorly-collected regions are therefore urgent priorities and are essential for further studies that lead to decision making in conservation.

Many legume genera with predominant distribution in the Amazon basin have been monographed, including *Inga* Mill. (Pennington, 1997), *Ormosia* Jacks. (Rudd, 1965), *Andira* Lam. (Pennington, 2003), *Swartzia* Schreb (Cowan, 1967), *Dimorphandra* Schott (Silva, 1986), *Parkia* R. Br. (Hopkins, 1986), and *Vatairea* Aubl. (Lima, 1982). These studies have contributed to fill in the gaps in taxonomic and geographical knowledge of these groups. However, many other Amazonian genera of Papilionoideae, remain understudied in the 21st century, despite legumes being the family that dominates Amazonia in terms of species diversity and abundance (ter Steege et al., 2013; Cardoso et al., 2017).

Chapter 1. General introduction and literature review

Plant groups of ecological importance should be prioritised in such taxonomic studies. *Aldina* Endl. is one example, which encompasses 18 species distributed in distinct habitats in the Amazon basin, from the seasonally flooded forests (*igapós*), white sand forests (*campinas* and *campinaranas*), *terra firme* forests, and higher altitude forests reaching up to 1000 m. Although *Aldina* has a widespread distribution (ter Steege, 2013, 2016), its representation in herbaria is relatively small when compared to other taxa with distribution in the same region. Part of the problem relies on the fact that the current knowledge we have of the genus dates back 70 years from Cowan's preliminary taxonomic studies (Cowan, 1953, 1958, 1961), since when much new herbarium material has accumulated. Although Cowan's published names remained stable, two varieties and one species recently needed nomenclatural changes because new collected material showed evidence to synonymize these taxa (Ramos et al., 2022). Thus, old species descriptions, often based in scant material, no longer reflect the current morphological limits of the species based on newly collected specimens. Moreover, the most recently described species are based on material cited in studies where putative new species were pointed out and for which vouchers are not cited (Gontscharov et al., 2006, 2009).

The current knowledge we have of the genus today is based on the collections of Richard Spruce, who collected extensively across the Amazon from 1850-1853 and from Bentham, who described the first species in the genus in *Flora Brasiliensis*. In the 20th century, Richard Cowan provided a preliminary revision of the genus in a series of publications in the project "The Botany of the Guayana Highland" (Cowan, 1953, 1958, 1961), based on the material collected by the American botanists H. R. Kunhardt (1948-1949), Bassett Maguire (1951-1953) and Julian Alfred Steyermark (1909-1988). Around ten new taxa were added to the genus, including seven new species and three varieties.

Later, new species were described from the Venezuelan Guayana (Cowan and Steyermark, 1984; Aymard 1998). Stergios and Cowan (1999) pointed out seven putative new *Aldina* species in the Papilionoideae section of the volume Flora of the Venezuelan Guyana, however no complete description or typification were provided (Steyermark et al. 1999). Seven new *Aldina* species have been recently described by Gontscharov et al. (2006, 2009), but the characters chosen to describe them showed continuity with previously described species (Cowan, 1953, 1958, 1961; Ramos et al., 2022).

Given the ecological importance of *Aldina* and its complex taxonomy, a taxonomic revision of the genus is needed. It is vital to include in a new revision more data on the

Chapter 1. General introduction and literature review

distribution of the species, material from new collections, complete morphological descriptions with identification keys, a taxonomic treatment, and illustrations. In Chapter Five, I thoroughly analyse global herbarium collections of *Aldina* in order to deliver a more thorough taxonomic revision of the genus.

2. Objectives and approaches

In this thesis, I ask, how did distinct flower morphologies evolve in the papilionoid legumes (Papilionoideae, Leguminosae)? To address this question I use an integrative approach involving genomic, transcriptomic, developmental, and morphological data, which are developed in the following chapters:

- Chapter Two: I use a candidate gene approach based on new genomic data generated from three legume species with distinct types of flower symmetry: *Andira anthelmia* (typical zygomorphic flower), *Andira inermis* (typical zygomorphic flower), *Cadia purpurea* (radially symmetrical), and *Cyathostegia matheewsii* (zygomorphic with one petal). I investigate the key transcription factors *CYCLOIDEA* (*CYC*), *DIVARICATA* (*DIV*), *RADIALIS* (*RAD* and *DIV*-and-*RAD*-Interacting-Factor (*DRIF*)). The objective of this chapter is to find orthologues of symmetry genes in the genomes of the target species using a phylogenetic approach to investigate their relatedness and the events of duplications that have led to different copy numbers and their implication in the development of non-papilionate flowers. The main question I aim to address from these analyses is: how does the copy number of symmetry genes correlate with floral diversity in papilionoid legumes? Genomes generated in this chapter are also used as references in Chapter Three for analyses of gene expression.
- Chapter Three: I investigate how the expression of symmetry genes might underlie flower development in morphologically distinct papilionoid legume species. RNA-seq data is used from a series of developmental stages from vegetative buds and flowers from the radially symmetric *Cadia purpurea* (vegetative buds, young flower buds, mature flowers, and abaxial and adaxial regions), the papilionate-flowered *Andira inermis* (vegetative buds), and the zygomorphic single-petalled *Cyathostegia matheewsii* (vegetative buds, young flower buds). This chapter analyses and compares the differential expression of symmetry genes in distinct developmental stages in order

Chapter 1. General introduction and literature review

to ask if expression levels of symmetry genes like CYC correlate with variation in floral symmetry in morphologically distinct species and if the variation in floral form is reflected in variation in gene expression patterns.

- Chapter Four: I use a comparative approach using new and previously published ontogenetic data from a series of developmental stages of a range of florally distinct species across the Papilionoideae subfamily. These included some of the taxa studied in Chapters Two and Three: such as the radially symmetric *Cadia purpurea* (ten stamens), *Cyathostegia mathewsii* (zygomorphic, single-petal, numerous stamens), and several unpublished ontogenetic data, for example from *Trischidium molle* (zygomorphic, single-petal, numerous stamens), *Ateleia guaraya* (zygomorphic, single-petal, ten stamens), and the papilionate-flowered species *Andira ormosioides* (ten stamens) and *Hymenolobium janeirensis* (ten stamens). The objective of this chapter is to compare the development of androecium symmetry and petal growth across a range of distinct flower types in order to gain insights on the stabilisation of these traits and what impact they had in the fixation of the typical papilionate flower in the derived lineages of Papilionoideae.
- Chapter Five: I present a taxonomic synopsis of the papilionoid genus *Aldina* Endl. The objective of this paper is to present a thorough revision of the 18 species of *Aldina* based on all specimens available in herbarium collections and new records from recent fieldwork. In this revision I provide an updated list of species with taxonomic comments, identification keys, distribution maps and diagnostic illustrations. This chapter is presented in the index as a publication titled “A Taxonomic Synopsis of *Aldina*, a Florally Distinctive and Poorly Collected Amazonian Genus of Papilionoid Legumes”.

In a final chapter, I summarise the main findings from the previous chapters, combining the genomic and transcriptomic results with flower ontogeny to discuss how morphologically unusual flowers evolved in a group where the majority of lineages possess the classic specialised papilionate flower.

References

- Arroyo, M. T. K. 1981. Breeding systems and pollination biology in Leguminosae. In: Polhill, R. M. & Raven, P. H. (eds) *Advances in Legume Systematics*, part 2. Royal Botanic Gardens, Kew, UK, pp 723–769.
- Armbruster, W. S. 2014. Floral specialization and angiosperm diversity: phenotypic divergence, fitness trade-offs and realized pollination accuracy. *AoB Plants* 6: plu003.
- Beech, E., M. Rivers, S. Oldfield, P. P. Smith. 2017. GlobalTreeSearch: The first complete global database of tree species and country distributions. *Journal of Sustainable Forestry* 36: 454–489.
- Bento, J. P. S., F. J. Kochanovski, A. L. B. Sartori. 2021. Floral ontogeny and secretory structures of *Discolobium pulchellum* and *Riedeliella graciliflora* (Leguminosae: Papilionoideae: Dalbergieae): two closely related genera with diverging floral morphology. *Plant Systematics and Evolution* 307: 6.
- Bello, M. A., P. Cubas, I. Alvarez, G. Sanjuanbenito, J. Fuertes-Aguilar. 2017. Evolution and expression patterns of CYC/TB1 genes in *Anacyclus*: phylogenetic insights for floral symmetry genes in Asteraceae. *Frontiers in Plant Science* 8: 589.
- Bruneau, A. et al. 2013. Legume phylogeny and classification in the 21st century: progress, prospects and lessons for other species-rich clades. *Taxon* 62(2): 217–248.
- Bruneau et al. 2014. Floral evolution in the Detarieae (Leguminosae): phylogenetics evidence for labile floral development in an early-diverging legume lineage. *International Journal of Plant Science* 175: 392–417.
- Boatwright, J. S., V. Slovin, B.-E. van Wyk, A. L. Schutte-Vlok, F. Forest, M. van der Bank. 2008. Systematic position of the anomalous genus *Cadia* and the phylogeny of the tribe Podalyrieae (Fabaceae). *Systematic Botany* 33(1): 133–147.
- Boyden, G. S., M. J. Donoghue, D. G. Howarth. 2012. Duplications and expression of RADIALIS-like genes in Dipsacales. *International Journal of Plant Sciences* 173(9): 971–983.
- Cardoso, D., L. P. de Queiroz, R. T. Pennington, H. C. de Lima, E. Fonty, M. F. Wojciechowski, M. Lavin. 2012a. Revisiting the phylogeny of papilionoid legumes: new insights from comprehensively sampled early-branching lineages. *American Journal of Botany* 99: 1991–2013.
- Cardoso, D., H. C. de Lima, R. S. Rodrigues, L. P. de Queiroz, R. T. Pennington, M. Lavin. 2012b. The realignment of *Acosmium* sensu stricto with the Dalbergioid clade

Chapter 1. General introduction and literature review

- (Leguminosae, Papilionoideae) reveals a proneness for independent evolution of radial floral symmetry among early branching papilionoid legumes. *Taxon* 61: 1057–1073.
- Cardoso, D., L. P. de Queiroz, H. C. de Lima, E. Suganuma, C. van den Berg, M. Lavin. 2013a. A molecular phylogeny of the vataireoid legumes underscore floral evolvability that is general to many early-branching papilionoid lineages. *American Journal of Botany* 100(2): 403–421.
- Cardoso, D. et al. 2013b. Reconstructing the deep-branching relationships of the papilionoid legumes. *South African Journal of Botany* 89: 58–75.
- Cardoso, D. et al. 2015. Filling in the gaps of the papilionoid legume phylogeny: the enigmatic Amazonian genus *Petaladenium* is a new branch of the early-diverging Amburaneae clade. *Molecular Phylogenetics and Evolution* 84: 112–124.
- Cardoso, D. et al. 2017. Amazon plant diversity revealed by a taxonomically verified species list. *The Proceedings of the National Academy of Sciences* 114(40): 10695–10700.
- Citerne, H., D. Luo, R. T. Pennington, E. Coen, Q. C. B. Cronk. 2003. A phylogenetic investigation of CYCLOIDEA-like TCP genes in the Leguminosae. *Plant Physiologist* 131(3): 1042–53.
- Citerne, H. L., R. T. Pennington, Q. C. B. Cronk. 2006. An apparent reversal in floral symmetry in the legume *Cadia* is a homeotic transformation. *Proceedings of the National Academy of Sciences* 103(32): 12017–12020.
- Citerne, H., Florian J., Sophie N., Catherine D. 2010. The evolution of floral symmetry. *Advances in Botanical Research* 54: 85–128.
- Citerne, H., M. L. Guilloux, J. Sannier, S. Nadot, C. Damerval. 2013. Combining phylogenetic and syntenic analyses for understanding the evolution of TCP ECE genes in eudictos. *PLOS ONE* (8), e74803.
- Corley, S. B., R. Carpenter, L. Copsey, E. Coen. 2005. Floral asymmetry involves an interplay between TCP and MYB transcription factors in *Antirrhinum*. *Proceedings of the International Academy of Sciences* 102(14): 5068–5073.
- Cowan, R. S. 1981. Swartzieae. Pp. 209–212 in *Advances in Legume Systematics*, eds. R. M. Polhill & P. H. Raven. Royal Botanic Gardens, Kew.
- Cubas, P. 2004. Floral zygomorphy, the recurring evolution of a successful trait. *BioEssays* 26(11): 1175–1184.
- Cubas, P., Nick L., John, D., Enrico C. 1999. The TCP domain: a motif found in proteins regulating plant growth and development. *The Plant Journal* 18(2): 215–222.

Chapter 1. General introduction and literature review

- Luo, D., R. Carpenter, L. Copsey, C. Vincent, J. Clark, E. Coen. 1999. Control of organ asymmetry in flowers of *Antirrhinum*. *Cell* 99: 367–376.
- Du, H., B. R. Feng, S. S. Yang, Y. B. Huang, Y. X. Tang. 2012a. The R2R3-MYB transcription factor gene family in maize. *PLoS One* 7(6): e37463.
- Du, H., S. S. Yang, Z. Liang, B. R. Fang, L. Liu, Y. B. Huang, Y. X. Tang. 2012b. Genome-wide analysis of the MYB transcription factor superfamily in soybean. *BMC Plant Biology*, 12: 106.
- Du, H., L. Zhang, L. Liu, X-F Tan., W-J Yang, Y-M Wu, Y-B Huang, Y-X Tang. 2009. Biochemical and molecular characterization of plan MYB transcription factor family. 2009. *Biochemistry* 74(1): 1–11.
- Endress, P. K. 1999. Symmetry in flowers: diversity and evolution. *International Journal of Plant Sciences* 160(S6): S3–S23.
- Feng X, Zhao Z, Tian Z, Xu S, Luo Y, Cai Z, Wang Y, Yang J, Wang Z, Weng L, Chen J, Zheng L, Guo X, Luo J, Sato S, Tabata S, Ma W, Cao X, Hu X, Sun C, Luo D. 2006. Control of petal shape and floral zygomorphy in *Lotus japonicus*. *Proceedings of the National Academy of Sciences* 103:4970–4975.
- Galego, L. & J. Almeida. 2002. Role of DIVARICATA in the control of dorsoventral asymmetry in *Antirrhinum* flowers. *Genes and Development* 16(7): 880–891.
- Gao, A. & W. Zhang. 2017. Evolution of RAD- and DIV-Like genes in plants. *International Journal of Molecular Sciences* 18: 1961.
- Gaudin, V. et al. 2000. The expression of D–Cyclin genes defines distinct developmental zones in snapdragon apical meristem and is locally regulated by the Cycloidea gene. *Plant physiology*, 122: 1137–1148.
- Gontscharov, M.Y. et al. 2006. On the new subtribe Aldiniinae of the Swartzieae and new species of the genus *Aldina*. *Botanicheskii Zhurnal* 91: 312–321.
- Gontscharov, M.Y., Yakovlev, G.P., Povydysh, M.N., 2009. Notes on the genus *Aldina* (Fabaceae). *Botanicheskii Zhurnal* 94: 267–275.
- Gubtiz, T., A. Caldwell, A. Hudson. 2003. Rapid molecular evolution of CYCLOIDEA-like genes in *Antirrhinum* and its relatives. *Molecular Biology and Evolution* 20(9): 1537–44.
- Hileman, L. C. 2013. Bilateral flower symmetry: how, when, and why? *Current Opinion in Plant Biology* 17: 146–152.

Chapter 1. General introduction and literature review

- Heisler, M. G., H. Jonsson, S. Wenkel, K. Kaufmann. 2022. Context-specific function of transcription factors controlling plant development: from leaves to flowers. *Current Opinion in Plant Biology* 69, 102262
- Hsin, K-T. and C-N Wang. 2018. Expression shifts of floral symmetry genes correlate to flower actinomorphy in East Asia endemic *Conandron ramondioides* (Gesneriaceae). *Botanical Studies* 59: 24.
- Hubbell, S. P. et al. 2008. How many tree species are there in the Amazon and how many of them will go extinct? *Proceedings of the National Academy of Sciences* 105: 11498–11504.
- Ireland, H. E., R. T. Pennington, J. Preston. 2000. Molecular systematics of the Swartzzieae. In P. S. Herendeen & A. Bruneau (eds). *Advances in Legume Systematics*. Royal Botanic Garden, Kew, 9: 217–231.
- Ireland, H. E. 2007. Taxonomic changes in the South American genus *Bocoa* (Leguminosae-Swartzieae): reinstatement of the name *Thrischidium* and a synopsis of both genera. *Kew Bulletin* 62: 333–349.
- Klitgaard, B. B. 1999. Floral ontogeny in tribe Dalbergieae (Leguminosae: Papilionoideae): *Dalbergia breasiliensis*, *Machaerium villosum* s.l. *Platymiscium floribundum*, and *Pterocarpus rotundifolius*. *Plant Systematics and Evolution* 219: 1–25.
- Klitgaard, B. B. et al. 2013. A detailed investigation of the Pterocarpus clade (Leguminosae: Dalbergieae): *Etabalia* with radially symmetrical flowers is nested within the papilionoid-flowered *Pterocarpus*. *South African Journal of Botany* 89: 128–142.
- Leite, V. G. et al. 2015. Floral development of the early-branching papilionoid legume *Amburana cearensis* (Leguminosae) reveals rare and novel characters. *International Journal of Plant Sciences* 176(1): 94–106.
- Li, S. 2015. The *Arabidopsis thaliana* TCP transcription factors: a broadening horizon beyond development. *Plant Signaling & Behavior* 10(7): e1044192.
- Luo, D., R. Carpenter, C. Vincent, L. Copsey, E. Coen. 1995. Origin of floral asymmetry in *Antirrhinum*. *Nature* (383): 794–799.
- Luo, D., R. Carpenter, L. Copsey, C. Vincent, J. Clark, E. Coen. 1999. Control of organ asymmetry in flowers of *Antirrhinum*. *Cell* (99): 367–76.
- Lucibelli, F., M. C. Valoroso, S. Aceto. 2020. Radial or Bilateral? The molecular basis of floral symmetry. *Genes* 11(4): 395.

Chapter 1. General introduction and literature review

- LPWG [Legume Phylogeny Working Group]. 2017. A new subfamily classification of the Leguminosae based on taxonomically comprehensive phylogeny. *Taxon*, 66:44–77.
- MacMahon, M. & Larry H. 2004. Phylogeny of Amorpheae (Fabaceae, Papilionoideae). *American Journal of Botany*, 91(8): 1219–1230.
- Madrigal, Y., J. F. Alzate, F. Gonzalez and N. Pabon-Mora. 2019. Evolution of RADIALIS and DIVARICATA gene lineages flowering plants with an expanded sampling in non-core eudicots. *American Journal of Botany* 106(3): 334–351.
- Mansano, V. F., V. Bittrich, A. M. G. A. Tozzi, A. P. Souza. 2004. Composition of the *Lecointea* clade (Leguminosae, Papilionoideae, Swartzieae), a re-evaluation based on combined evidence from morphology and molecular data. *Taxon* 53: 1007–1018.
- Martin-Trillo, M. & P. Cubas. 2010. TCP genes: a family snapshot ten years later. *Trends in Plant Science* 15(1): 31–9.
- Ojeda, D., R. Jaén-Molina, A. Santos-Guerra, J. Caujape-Castells, Q. Cronk. 2017. Temporal, but not spatial, changes in expression patterns of petal identity genes are associated with loss of papillate conical cells and the shift to bird pollination in Macaronesian *Lotus* (Leguminosae). *Plant Biology* 19(3):420-427.
- Ozsolak, F. & P. M. Milos. 2011. RNA sequencing: advances, challenges and opportunities. *Nature Reviews Genetics* 12(2): 87–98.
- Paulino, J. V., V. F. Mansano, S. P. Teixeira. 2013. Elucidating the unusual floral features of *Swartzia dipetala* (Fabaceae). *Botanical Journal of the Linnean Society* 173: 303–320.
- Pennington, R. T., B. B. Klitgaard, H. Ireland, M. Lavin. 2000. New insights into floral evolution of basal Papilionoideae from molecular phylogenies. In P. S. Herendeen & A. Bruneau [eds.], *Advances in Legume Systematics*, part 9, 233–248. Royal Botanic Gardens, Kew, UK.
- Pennington, R. T., M. Lavin, H. Ireland, B. Klitgaard, J. Preston, J. M. Hu. 2001. Phylogenetic relationships of basal papilionoid legumes based upon sequences of the chloroplast trnL intron. *Systematic Botany* 26: 537–556.
- Pellew, C. & A. Sverdrup. 1923. New observations on the genetics of peas. *Journal of Genetics* 13: 125–131.
- Prenner, G. 2004. Floral development in *Polygala myrtifolia* (Polygalaceae) and its similarities with Leguminosae. *Plant Systematics and Evolution* 249: 67–76.

Chapter 1. General introduction and literature review

- Prenner, G., D. Cardoso, C. E. Zartman, L. P. Queiroz. 2015. Flowers of the early-branching papilionoid legume *Petaladenium urceoliferum* display unique morphological and ontogenetic features. *American Journal of Botany* 102: 1780–93.
- Prenner, G. & Cardoso, D., 2017. Flower development of *Goniorrachis marginata* reveals new insights into the evolution of the early diverse detarioid legumes. *Annals of Botany* 119(3): 417–432.
- Polhill, R. M. 1981. Sophoreae. In: Polhill & R.M., Raven, P.H. (eds.), *Advances in Legume Systematics*. Royal Botanic Gardens, Kew, 1: 213–230.
- Polhill, R. M., 1994. Classification of the Leguminosae. In: Bisby, F.A., Buckingham, J., Harborne, J.B. (Eds.), *Phytochemical Dictionary of the Leguminosae. Plants and Their Constituents*. Chapman and Hall, London. 1: 25–27.
- Ramos, G., H. C. Lima, G. Prenner, L. P. Queiroz, C. E. Zartman, D. Cardoso. 2016. Molecular systematics of the Amazonian genus *Aldina*, a phylogenetic enigmatic ectomycorrhizal lineage of papilionoid legumes. *Molecular Phylogenetics and Evolution* 97: 11–18.
- Ramos, G. C. E. Zartman, H. C. Lima, R. T. Pennington, D. Cardoso. 2022. A taxonomic synopsis of *Aldina*, a florally distinctive and poorly collected Amazonian genus of Papilionoid legumes. *Systematic Botany* 47(4): 962–977.
- Raimundo, J., Sobral, R., Bailey, P., Azevedo, H., Galego, L., Almeida, J., E. Coen, M. M. R. Costa. 2013. A subcellular tug of war involving three MYB-like proteins underlies a molecular antagonism in *Antirrhinum* flower asymmetry. *The Plant Journal* 75: 527–538.
- Reardon, W., P. Gallagher, K. M. Nolan, H. Wright, M. C. Cardenosa-Rubio, C. Bragalini, C-S. Lee, D. A. Fitzpatrick, K. Corcoran, K. Wolff, J. M. Nugent. 2014. Different outcomes for the MYB floral symmetry genes *DIVARICATA* and *RADIALIS* during the evolution of derived actinomorphy in *Plantago*. *New Phytologist* 202: 716–725.
- Sauquet, H., et al. 2017. The ancestral flower of angiosperms and its early diversification. *Nature Communications* (8): 16047.
- Sargent, R. D. 2004. Floral symmetry affects speciation rates in angiosperms. *The Royal Society* 271: 603–608.
- Singh, V. K., R. Garg, M. Jain. 2013. A global view of transcriptome dynamics during flower development in chickpea by deep sequencing. *Plant Biotechnology Journal* (11): 691–701.
- Singh, V. K. & M. Jain. 2014. Transcriptome profiling for discovery of genes involved in shoot apical meristem and flower development. *Genomics Data* (2): 135–138.

Chapter 1. General introduction and literature review

- Sinjushin, A. A. 2018. Floral ontogeny in *Cordyla pinnata* (A.Rich) Milne-Redh. (Leguminosae, papilionoideae): away from stability. *Flora* 241: 8–15.
- Sinjushin, A. A. 2019. Notes on floral symmetry in the *Pterocarpus* clade (Leguminosae: Papilionoideae: Dalbergieae). *Wulfenia* 26: 175–188.
- Stevenson, C. E. M., N. Burton, M. M. R. Costa, U. Nath, R.A. Dixon, E. S. Coen, D.M. Lawson. 2006. Crystal structure of the MYB domain of the RAD transcription factor from *Antirrhinum majus*. *Proteins: Structure Function Bioinformatics* 65: 1041–1045.
- Torke, B. M. & V. F. Mansano. 2009. A phylogenetically based sectional classification of *Swartzia* (Leguminosae-Papilionoideae). *Taxon* 58: 913–924.
- Tucker, S. C. 1990. Loss of floral organs in *Ateleia* (Leguminosae, Papilionoideae). *American Journal of Botany* 77: 750–761.
- Tucker, S.C. & Douglas, A.W. 1994. Ontogenetic evidence and phylogenetic relationships among basal taxa of legumes. In: Ferguson, I.K., Tucker, S. (Eds.), *Advances in Legume Systematics*, Structural Botany. Royal Botanic Gardens, Kew. 6:11–32.
- Tucker, S. C. 1999. Evolutionary lability of symmetry in early floral development. *International Journal of Plant Sciences* 160(6): S25–S39.
- Tucker, S. C. 2002. Floral ontogeny in Sophoreae (Leguminosae: Papilionoideae). III. Radial symmetry and random petal aestivation in *Cadia purpurea*. *American Journal of Botany* 89(5): 748–757.
- Tucker, S. C. 2003. Floral ontogeny in *Swartzia* (Leguminosae: Papilionoideae: Swartzieae): distribution and role of the ring meristem. *American Journal of Botany* 90: 1271–1292.
- Valoroso, M. R., R. Sobral, G. Saccone, M. Salvemini, M. M. R. Costa, ad S. Aceto. 2019. Evolutionary conservation of the orchid MYB transcription factors DIV, RAD, and DRIF. *Frontiers in Plant Science* 10: 1359.
- Viola, I. L. & D. H. Gonzalez. 2023. TCP transcription factors in plant reproductive development: juggling multiple roles. *Biomolecules* 13(5): 750.
- Wang Z, Luo Y, Li X, Wang L, Xu S, Yang J, Weng L, Sato S, Tabata S, Ambrose M, Rameau C, Feng X, Hu X, Luo D. 2008. Genetic control of floral zygomorphy in pea (*Pisum sativum* L.). *Proceedings of the National Academy of Sciences of the U.S.A.* 105: 10414–10419.
- Wojciechowski, M. F., M. Lavin, M. J. Sanderson. 2004. A phylogeny of legumes (Leguminosae) based on analysis of the plastid *matK* gene resolves many well-supported subclades within the family. *American Journal of Botany* 91: 1846–1862.

Chapter 1. General introduction and literature review

Zhao, Z., J. Hu, S. Chen, Z. Luo, D. Luo, J. Wen, T. Tu, D. Zhang. 2019. Evolution of CYCLOIDEA-like genes in Fabales: insights into duplication patterns and the control of floral symmetry. *Molecular Phylogenetics and Evolution* 132: 81–89.

Chapter 2

An investigation of symmetry genes in papilionoid legume
flowers

Introduction

2.1 Looking for candidate genes: developmental genetics underlying morphologically unusual papilionoid flowers

Given the diversity of flower morphologies found in the legume family (Leguminosae), the early branching lineages of papilionoid legumes make a good case study to investigate flower evolution. Insightful developmental genetic studies have been conducted on the radially-symmetrical-flowered papilionoid legume genus *Cadia* Forsk. (Citerne et al., 2003, 2006, 2010), yet there is a lack of genetic studies of numerous other Papilionoideae genera with unusual floral morphologies. This chapter investigates florally distinct papilionoid species in order to gain insights into the genetic basis underlying the changes in flower symmetry found in legumes and angiosperms more generally. I originally aimed to sample genera belonging to the same clade that possessed morphologically distinct species, such as the *Andira* clade (*Andira* Lam., *Aldina* Endl. and *Hymenolobium* Benth.), but this plan was blocked by the poor availability of fresh material from these species, which is crucial to obtain high molecular weight DNA for high throughput sequencing. In light of that, I have selected species with distinct floral morphology and for which fresh material was available in the RBGE living collections. Thus, this chapter includes species from genera with morphologies that span the typical zygomorphic papilionate flower such as *Andira*, the radially symmetric *Cadia*, and the zygomorphic single-petalled with multiple stamens *Cyathostegia* (Benth.) Schery (Fig. 1).

Andira belongs to the *Andira* clade, along with the typically zygomorphic papilionate *Hymenolobium* and the radially symmetric *Aldina* (Ramos et al., 2016). The *Andira* clade is one example of several morphologically heterogeneous clades in the early branching papilionoids that encompass genera with papilionate flowers and others with more unusual non-papilionate morphologies. Two species of *Andira* were sampled, *A. inermis* and *A. anthelmia*, representing lineages with zygomorphic flowers, where there is a specialisation of the corolla into an adaxial petal (standard), two lateral wings, and two ventral petals (keel). Nine of the 10 stamens in this type of flower are half fused throughout their length with one stamen almost free from the androecium sheath, with the distal points curved upwards the standard petal. The gynoecium is enveloped by the androecium sheath and exposed at the tip. *Andira inermis* (W.Wright) DC. is distributed throughout the Caribbean, Mexico, Central and South America and Africa. In Africa, it occurs in wooded savanna and rain forest, whilst in Central America and the Caribbean coast of northern South America it is common in seasonally

dry tropical forests. Elsewhere in the Neotropics it grows in rain- and gallery forest, including inundated areas and mangrove (Pennington, 2003). *Andira anthelmia* (Vell.) Benth. is distributed along the coast of Brazil, in restinga forest and rainforest. *Andira anthelmia* is often found along riverbanks with seasonal flooding regimes (Pennington et al., 2003). Pollination in *Andira* involves various bee species, as has been observed in the flowers of *A. inermis* in Costa Rica (Frankie et al., 1976; Pennington, 2003). The size of the flowers seems to play a role in the specificity of pollinators for each species. *Andira inermis*, for example, has flowers of about 10–12 mm long, whereas those of *A. anthelmia* reach 19–24 mm long and are visited by large xylocopid bees, suggesting a different pollination biology (Pennington, 2003). The probable difference in pollination strategy is probably reflected in the phenology of *A. anthelmia*, where a high percentage of individuals were observed to be synchronously flowering in one area, in contrast to smaller-flowered species of *Andira*, where few individuals were observed flowering synchronously (Pennington, 2003).

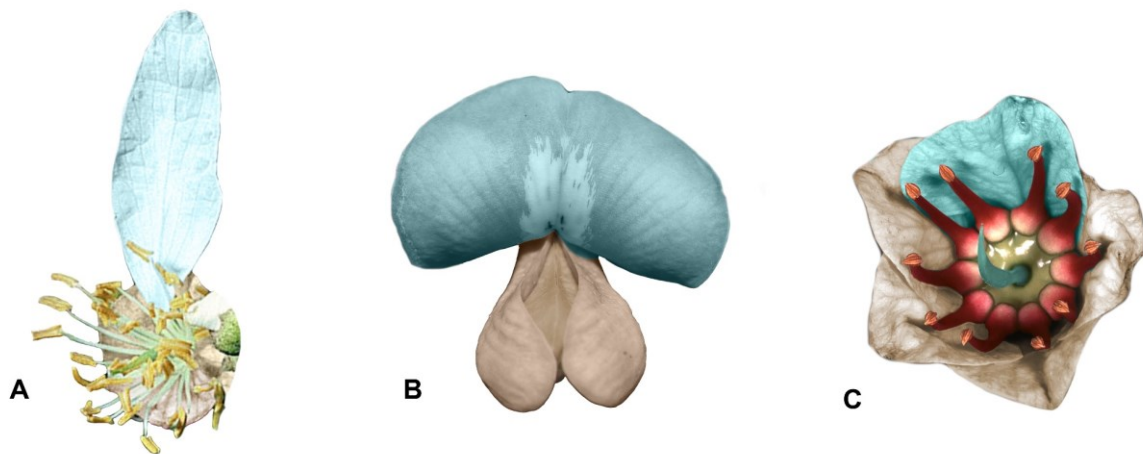


Figure 1. Abaxial and adaxial domains in three florally distinct papilionoid legume species. The adaxial domain is shown in blue and abaxial domain in sepia. A, *Cyathostegia mathewsii*; B, *Hymenolobium alagoanum*; C, *Cadia purpurea*.

Papilionoid legume species with the typical zygomorphic papilionate flowers tend to have bees as pollinators, which is an interaction that has been facilitated by the architecture of the flowers, especially the wing and keel petals, which work as a landing platform for bees seeking for nectar. Other papilionoid species with unusual flower types seem to have evolved different pollination syndromes, which might be related to the lack of the highly specialised

wing and keel petals. An example is the single-petalled, monotypic genus *Cyathostegia* that is also the focus of this chapter. *Cyathostegia mathewsii* is endemic to the dry inter-Andean valleys of Southern Ecuador and Northern Peru and has flowers with a single, erect, white petal on the adaxial side of the flower with free stamens varying in number from 16 to 26 and a cyathiform calyx that protects the internal whorls (Lewis et al., 2003). The pollination of *C. mathewsii* was observed by Lewis et al. (2003) to involve a single species of beetle, *Phyllophaga* sp. This finding was interesting considering that the closely related, single-petalled *Ateleia herbert-smithii* Pittier is wind pollinated (Janzen, 1989). Beetle pollination is not common in legumes and has only been reported in the mimosoid genus *Acacia* (Hawkeswood, 1983, 1989; Tybirk, 1993). Both *Cyathostegia* and *Ateleia*, together with *Trischidium* Tul., belong to the Swartzoid clade, which is largely comprised of genera displaying a tendency to lack petals combined with a high number of commonly free stamens (Cardoso et al., 2004). *Cyathostegia*, *Ateleia* and *Trischidium* are nested in the Ateleioid clade within the Swartzoid clade (Torke & Schaal, 2008) and are subject of ontogenetic study in Chapter Four.

The Genistoid *Cadia* is another well-known genus of florally distinctive, radially symmetrical papilionoid legumes. It comprises seven species, six endemic to Madagascar and one, *Cadia purpurea*, ranging from east and north-east Africa to southwestern Arabia (Du Puy et al., 2002). *Cadia purpurea*, which is also a subject of this chapter, has radially symmetric flowers, ten radially arranged stamens, and a central gynoeceum moderately curved towards the adaxial region of the flower. *Cadia* flowers produce abundant nectar but no scent, suggesting it might be pollinated by birds (Pennington et al., 2000; Boatwright et al., 2008). Little is known about the pollination biology of the early branching papilionoid legumes, but it is possible that deviations from the typical, zygomorphic flowers may be due to adaptation to different pollinators, as suggested in *Cadia* (Pennington et al., 2000).

From an ontogenetic perspective, the radial flowers of *Cadia* have been regarded by Tucker (2002) as neotenous because they retain early developmental features in late growth stages. Although *Cadia* also has unidirectional development in each floral whorl, like other papilionoid legumes, later organ internal differentiation remains uniform leading to its radially symmetric flowers (Tucker, 2002), whereas in other papilionoids the specialisation of internal organs rearranges the flower parts towards a zygomorphic pattern.

Citerne et al. (2003) identified four copies of the dorso-ventral symmetry gene CYCLOIDEA in *Cadia* using degenerate Polymerase Chain Reaction (PCR) with primers designed from available sequences from *Lotus japonicus*, compared with CYCLOIDEA's conserved TCP and R domain sequences from *Antirrhinum majus*. These sequences were analysed in an unrooted phylogenetic tree which showed the gene family was divided into two subgroups, the PCF group (one *Cadia* ortholog - LegCYC4) and a second group CYC/TB1 (three *Cadia* orthologs, LegCYC1A, LegCYC1B and LegCYC3). Although this study gave important insights in the events that led to CYC-like duplications in the early diversification of Papilionoideae, no further discussion was presented on the putative correlation between gene copy number and the morphological changes in the unusual flowers found across the early branching lineages of Papilionoideae.

The gene expression changes associated with *Cadia*'s distinct floral morphology were also studied by Citerne et al. (2006), who observed that CYC-like genes are expressed primarily in the dorsal domain of flowers in *C. purpurea* and the papilionate-flowered *Lupinus nanus*, which is also a member of the Genistoid clade. However, *LegCYC1B* has an expanded expression including the lateral and ventral domains in *Cadia*, which is likely to underlie the distinct *Cadia* radial flower morphology (Citerne et al., 2006).

This chapter investigates the genetic basis underlying changes in flower morphology of four morphologically distinct non-model species from the early branching lineages of Papilionoideae, including the typical zygomorphic *Andira anthelmia* and *A. inermis*, the radially symmetric *Cadia purpurea*, and the zygomorphic single-petaled *Cyathostegia mathewsii*. I use a candidate gene approach to identify putative orthologs for key *TCP* and *MYB* genes (described in Chapter One) that may have been recruited to play a role in the establishment of flower symmetry in these species. Genomic sequences were used to detect gene copy number variation between species and protein secondary structure was analysed in order to identify conserved or sub/neo-functionalization of these proteins. This chapter investigates whether duplication and/or reduction of *TCP/MYB* gene copy number correlates with the evolution of unusual floral morphologies in papilionoid legumes.

2.2 Material and Methods

2.2.1 Plant material sampling

Fresh plant material was sampled from the four species in this study from the living collections of the RBGE. Our sequencing strategy involved high coverage Whole Genome Sequencing (WGS) at 100× in order to generate draft genomes for our target species. Fresh material was used in order to minimise DNA degradation and increase the proportion of high molecular weight DNA (QIAGEN, 2015) for the WGS. This enables better quality libraries to be prepared, which results in improved DNA sequencing data (Fig. 2).

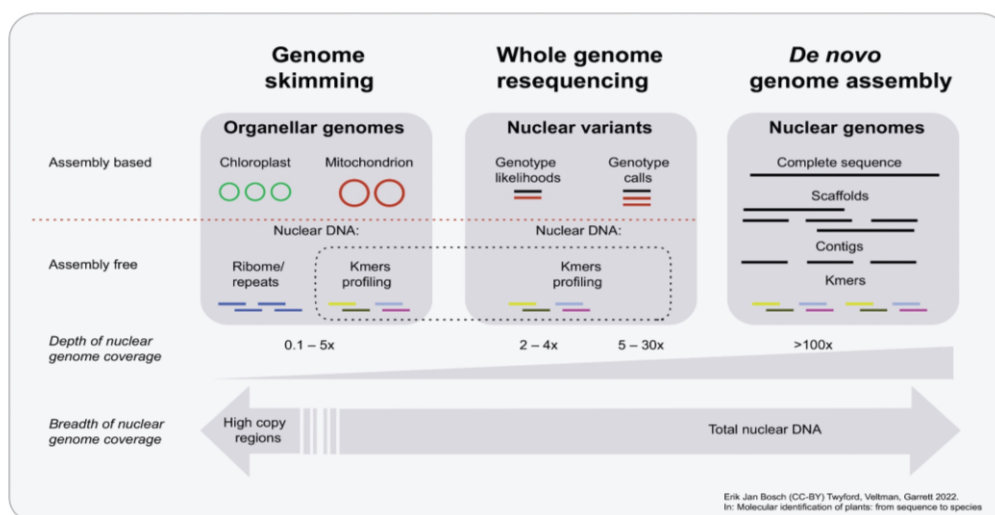


Figure 2. Flowchart showing different types of sequencing strategies (image originally published by Twyford et al. 2022).

Young healthy leaflets were obtained from each species. Leaflet tissue was collected using sterile tweezers and stored in silica gel for subsequent DNA extraction. Genomic tip 20G protocol was used for DNA extraction following the recommendations of Nishii et al. (2020) and associated protocols (Gunter, 2019; QIAGEN, 2015).

2.2.2 Data collection and sequence annotation

In this study, sequences of the TCP (CYC) and MYB gene families (DIV, RAD, DRIF) of 57 legume species generated in previous studies (Table 1, 2, 3, and 4) were selected from

NCBI's GenBank (Benson et al., 2018) to build protein sequence matrices. A new dataset bringing together RAD, DIV and DRIF sequences from legumes was assembled and analysed in a broad phylogenetic context for the first time. Sequences were obtained from different databases such as OrthoDB (<https://www.orthodb.org>), NCBI (<https://www.ncbi.nlm.nih.gov/>), and Phytozome (<https://phytozome-next.jgi.doe.gov/>).

Finding orthologs between different species can be challenging and one way to establish homology is to employ a phylogenetic approach (Emms & Kelly, 2019). The Basic Alignment Research Tool (BLAST; Altschul et al., 1990) was used to identify putative TCP and MYB orthologs in the genomes of our target species. Parameters such as percentage of identical positions, sequence overlap, and e-value were used as parameters to select putative orthologs and paralogs in the Blast searches. Firstly, previously published TCP sequences of *Arabidopsis thaliana* and MYB sequences from *Antirrhinum majus* were used. The candidate sequences were aligned to TCP and MYB sequences from *A. thaliana* and *A. majus* as these were the orthologs which had been best characterised functionally. Sequences clustered with CYC-like, DIV-like, and RAD-like genes were then selected for the downstream analyses. The same methodology was employed to obtain candidate sequences from the genomes assembled in this study.

2.2.3 Whole genome sequencing

Because no reference genome was available for the species studied here, sequencing was carried out at 100x coverage in order to cover as much as possible of the genome, including genes with multiple copies. Illumina NovaSeq 6000 PE150 technology was used to generate short-reads of 150 base pairs, resulting in 22 G of raw data per sample. A reference guided assembly was not possible for the lack of a reference genome and therefore a *de-novo* assembly approach was employed using k-mer to build a graph, then identify paths through the graphs to generate contigs. After the initial assembly, scaffolding of the contigs was performed to generate longer and contiguous sequences. The details of the assembly steps are discussed further.

Until the development of this thesis (2019–2024), ca. 920 genomes have been sequenced in Leguminosae based on Kew's C-values database. From this list, about 681 genomes are from papilionoid legumes, however only 12 taxa distributed in four genera represented early-branching lineages of Papilionoideae. In this chapter I have generated four new genomes spanning florally distinct species within the subfamily (Fig. 3).

2..2.4 Genomic data QC and pre-processing

Illumina raw reads were assessed for quality control using FastQC v0.11.9 (Andrews, 2010). In order to remove adapter sequences and low quality reads, Trimmomatic v0.39 was used with the default set of Illumina adapter sequences (Bolger et al., 2014). Cutadapt v3.1 (Martin, 2011) was also used to remove potential PCR duplicates, remaining primers, poly-A tails and other artefacts generated during amplification. Specific settings for the above mentioned softwares are given below.

Trimmomatic settings:

```
trimmomatic PE -phred33 Cy_mathewsii_1_fq.gz Cy_mathewsii_2_fq.gz  
ILLUMINACLIP:TruSeq3-PE.fas:2:30:10 LEADING:3 SLIDINGWINDOW:4:15 MINLEN:36
```

Cutadapt settings:

```
cutadapt -q 30 -m 50 -a AGATCGGAAGAGC -A AGATCGGAAGAGC -trim-n -o  
S1_EDSW200019175-1a_HNKLWDSXY_L1_1_trimmed.fq -p S1_EDSW200019175-  
1a_HNKLWDSXY_L1_2_trimmed.fq S1_EDSW200019175-1a_HNKLWDSXY_L1_1.fq  
S1_EDSW200019175-1a_HNKLWDSXY_L1_2.fq
```

After pre-processing, the trimmed reads were reassessed by an extra run of FASTQC to check if there was any improvement in the read quality. Pre-processed forward and reverse paired reads were then used as input for KmerGenie (Chikhi and Medvedev, 2014). This software breaks the datasets into sequences of length 'k' to allow quick analysis of the levels of variation in the sample, identifying repeats and estimating heterozygosity. These measures are key parameters for a full genome assembly. Different k-mer sizes were tested, from 21 to 121 bp in intervals of 10 by default. Output files were used as input for GenomeScope2 (Ranallo-Benavidez et al., 2020), which uses k-mer count data to adjust the analysed genomes to different models and provides together with the k-mer profiles, estimates of the genome size, % heterozygosity and % of genomic repeats estimation. Such data provides important information about the sequenced genomes and can be used to select the most appropriate genome assembler.

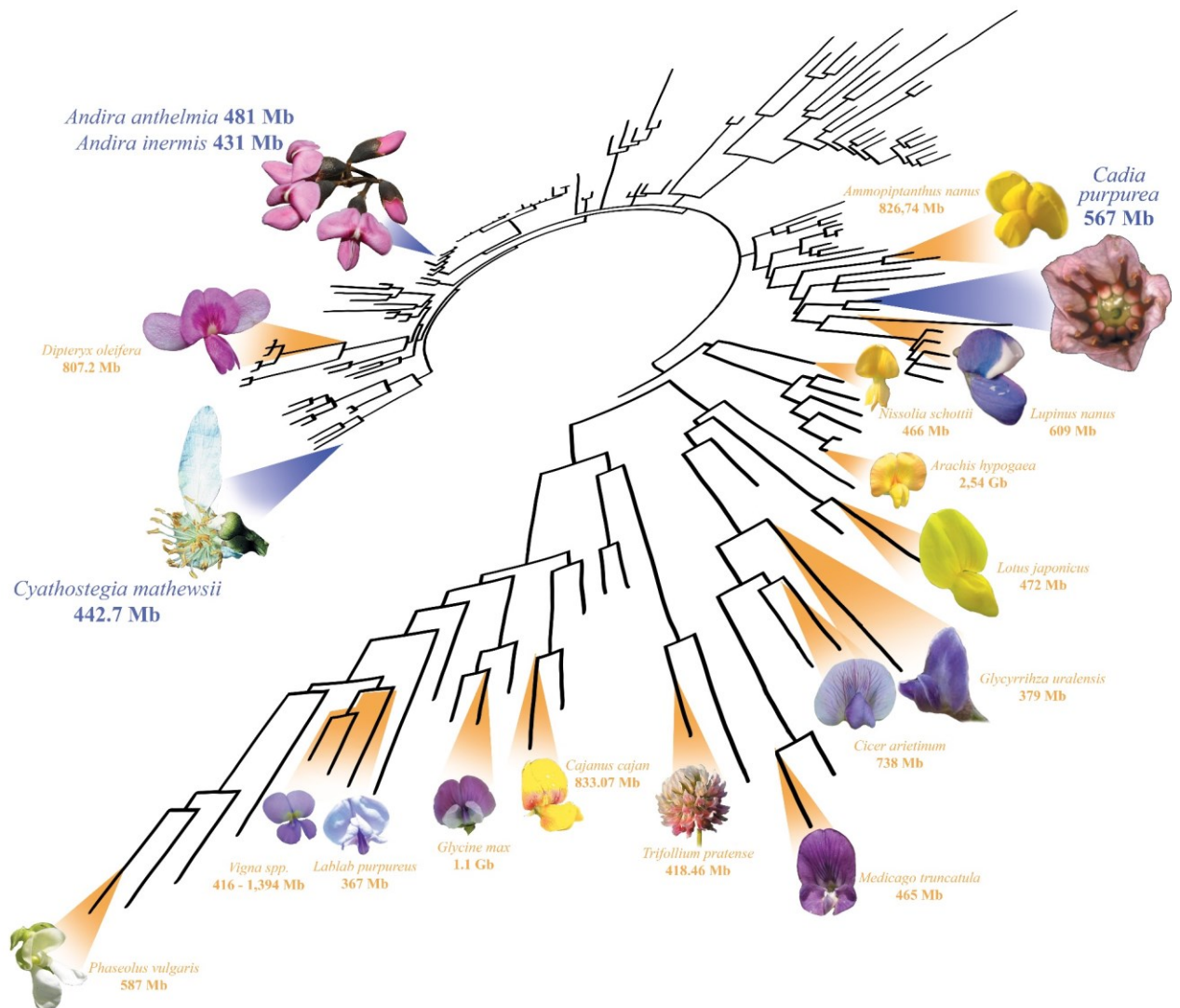


Figure 3. Phylogeny of the Papilionoideae showing species whose genomes have been sequenced up until 2019 (orange shading). Purple shadings indicate species whose genomes have been sequenced in this study.

2.2.5 Genome assembly

The choice of assembler employed in this chapter took into account that non-model species were used, for which a close reference genome is not available. In such cases *de novo* assemblers can be used to assemble genomes from pre-processed short reads. Two assemblers were used to assemble the genome of our target species: Platanus-allee 2.2.2 (Kajitani et al., 2019) and SPAdes (Prjibelski et al., 2020). Platanus-allee 2.2.2 is a *de novo* haplotype assembler and phasing tool, which assembles each haplotype from a diploid genome. Platanus-allee is especially useful to analyse highly divergent (heterozygous) regions in which haplotypes differ considerably (Kajitani et al., 2019).

Similar to Platanus-allee, SPAdes builds continuous and accurate sequences from short reads. The earlier version of SPAdes (Nurk et al., 2013) was designed to assemble bacterial genomes from short Illumina reads. The latest release has five protocols, including the classic bacterial assembly and hybrid assembly from short and long reads (Prjibelski et al., 2020). SPAdes 3.15.5 was used to *de novo* assemble the genomes of *A. inermis* and *A. anthelmia*, since the assembly algorithm from Platanus-allee did not perform well in the phasing step for those species. A segmentation fault seems to be the problem and is followed by the error message: “Error, SolveDBG exception!! Platanus_allee solve_DBG command failed.”. This error has been pointed out by users as a bug on earlier versions of the software and a solution has not been clearly addressed. The fact that Platanus-allee optimally performs with heterozygous genomes and failed with the two *Andira* species, which are more heterozygous than *Cadia* and *Cyathostegia*, is intriguing and a response from the authors of the software would shed light into the problem. However I tried to contact the authors, no response has been given.

The general steps of Platanus-allee *de novo* assembly involved:

1-Pre-processing: Trim the reads using a best fit program (e.g. Cutadapt, Trimmomatic). The paired-end and mate-pair reads should be trimmed using different programs as they have normally different orientation. **2-De Bruijn Graph construction:** The preprocessed sequencing data is used to build a de Bruijn graph. This graph represents the overlaps between short k-mers in the reads. **3-Graph simplification:** Platanus-allee performs graph

simplification to remove branches that are considered to be sequencing errors or repetitive regions. This simplification step reduces the complexity of the graph. **4-Contig assembly:** this step represents the initial draft of the genome assembly and uses the simplified de Bruijn graph to generate longer and contiguous sequences of DNA. **5-Scaffold building:** scaffolds are then built by considering mate-pair information from paired-end sequencing libraries. This step helps to connect contigs into more extensive genomic sequences, providing a more comprehensive representation of the genome. **6-Gap closing:** After scaffolding, Platanus-allee attempts to fill gaps in the assembly by using paired-end information and iterative processes, finishing the assembly process.

2.2.6 Assembly process

As shown above, the *de novo* assembly follows six steps. The tools used to assemble the genomes in this study are Platanus-allee 2.2.2 package plus additional software as described below. Command options for each step are described (see Appendix I). Platanus-allee was employed in the assembly of *Cadia* and *Cyathostegia*.

I – Trimming: For the assembly of our plant genomes I used *Cutadapt* (Martin, 2011) and *Trimmomatic* (Bolger et al., 2014) to remove unwanted sequences from the reads such as adapter sequences, primers, poly-A tails, and other artefacts from Illumina NGS data.

II – Contig generation: number of threads (-t) was set to 48, -m max memory in GB used for the assembly step was set to 140 GB, and the -tmp temp directory to use was set to scratch space.

III – Phasing: in this step I used the contigs generated in the previous step, along with trimmed reads (pair end) to scaffold the initial set of contigs. Again, most of the default options were used.

IV – Consensus: The final step is to close the gaps (reduce the N content in the genome introduced during scaffolding). The scaffolds generated in the previous step were needed.

In addition to Platanus-allee, I also employed the St. Petersburg genome assembler (SPAdes) to *Cadia*, *Cyathostegia* and *Andira anthelmia* and *Andira inermis*. SPAdes is known for its robustness in handling short-read sequencing data. Like Platanus-allee, SPAdes constructs de Bruijn graphs, enabling the resolution of complex genomic regions, such as repetitive elements, which can pose a challenge for assembly processes. The *de novo* genome assembly of spades involves several key steps:

I - *Error correction*: SPAdes performs error correction internally prior to assembly contigs, improving the accuracy of the reads.

II - *De Bruijn Graph construction*: the de Bruijn graph approach is used to represent overlapping sequences in the input reads. K-mers are extracted from the reads, and the de Bruijn graph is constructed based on the relationship between smaller fragments.

III - *Contig Assembly*: initial contigs are constructed from the de Bruijn Graphs, which identifies paths representing potential contigs.

IV - *Contig Extension*: SPAdes extends contigs by iteratively walking through the de Bruijn graph to incorporate additional reads and sequences. This helps in resolving ambiguities and improving contig lengths.

V - *Scaffold Construction*: paired-end information is used to order and orient contigs into scaffolds.

VI - *Repeat Resolution*: various algorithms are employed to identify and handle repetitive elements during the assembly process. This step is crucial for accurate assembly, especially in genomes with high repeat content.

VII - *Gap Filling*: in a final step, gaps within scaffolds are closed, which is important to improve the completeness of the assembly.

After performing *de novo* assembly of our genomes with both assemblers, I accessed the completeness of the assemblies using BUSCO (Table 5).

2.2.7 K-mer analysis

Trimmed output files containing both forward and reverse paired reads were used as input for KmerGenie (Chickhi and Medvedev, 2014). Such software performs k-mer counting using different k-mer sizes, taking from 21 to 121 bp in intervals of 10 by default. Output files were used as input for GenomeScope2 (Rannallo-Benavidez et al., 2020), which uses k-mer count data to adjust the analysed genomes to different models and provides together with the k-mer profiles, estimates of the genome size, % of heterozygosity and % of genomic repeats estimation.

2.2.8 Genome annotation

Genomes were annotated using BRAKER3 (bioRxiv 2023 Nov 27:2023.06.10.544449. doi: 10.1101/2023.06.10.544449. Preprint). BRAKER3 is designed for gene prediction and annotation in eukaryotic genomes, even with the absence of a reference genome. RNA-seq data and hints from protein alignments are used to improve the accuracy of protein prediction. The main step of the annotation path used by BRAKER3 is as follows:

I - *Data preparation*: when available, genomes sequences can be used in FASTA format. RNA-seq reads should be pre-processed including quality control, trimming, and filtering. A protein database can also be used in order to improve the gene prediction.

II - *Extraction of intron hints*: RNA-seq data is used to generate intron hints by aligning RNA-seq reads to the genome and identifying potential intron locations.

III - *Extraction of protein hints*: if a protein dataset is available, it should be align to the target genome to generate hints for protein-coding regions.

IV - *Training Augustus*: BRAKER3 uses the gene prediction tool Augustus. Augustus can be trained to use the intron and protein hints from the previous steps to improve its accuracy for gene prediction in the specific genome.

V - *Gene prediction*: Augustus is used with the trained parameters and additional evidence to predict genes in the genome. This step generates an initial set of gene predictions.

VI - *Creation of gene structures*: BRAKER3 combines the initial gene predictions with the extrinsic evidence to create a new set of gene structures. This step refines the gene predictions based on the additional information.

VII - *Final annotation*: the final annotation includes the refined gene structures along with information about introns and proteins, improving the overall accuracy of the annotated gene set.

Several output files are generated including the final annotated genome in GFF (General Feature Format) and GTF (Gene Transfer Format) format. These files can then be used for downstream analyses. In this study, I used the annotated files to search for the candidate genes described in detail in Chapter One.

2.2.9 Candidate gene dataset

Putative orthologs for our candidate genes were initially identified using the Basic Local Alignment Search Tool (BLAST, Altschul et al., 1990). Canonical sequences for each candidate gene were downloaded from NCBI (Table 1). The tblastn module from BLAST was used to find orthologous sequences using a set of translated query sequences against genomic sequences. Because low hits were retrieved for some genes, a transcriptome dataset was also used to confirm and/or retrieve the sequences captured from the genomes. Where putative orthologs were found in the transcriptomes, those were considered as putative copies.

2.2.10 Phylogenetic analyses

Translated sequences were used to generate matrices for all the CYC, RAD and DIV genes (Tables 1-2). The first set sequences were manually trimmed in order to only contain the conserved TCP and R domains, since regions located between these domains showed high variation and were difficult to align, as observed by Citerne et al. (2003). The second set involved complete sequences, including the extremes and the regions between domains. Sequences were manually aligned after a general automated alignment using MUSCLE in AliView 1.27 (Larsson, 2014) in order to avoid inconsistencies derived from the highly divergent regions between the conserved domains. Gaps were removed manually. The IQ-TREE software (Minh et al. 2020) was used to infer phylogenetic trees based on maximum likelihood with the following settings:

```
#Finding the best substitution model (-m MFP)
#Using ultrafast bootstrap approximation (-B 1000)
#Reducing impact of severe model violations (-bnni)
#With multiple cores -T AUTO
#For amino acid alignments -seqtype AA
#From http://www.iqtree.org/doc/Tutorial

for name in CYC_matrix_trimmed2.phy;
do iqtree -s "$name" -m MFP -B 1000 - bnni -T AUTO -seqtype AA -redo;
done
```

I contextualised the new legume sequences and rooted the trees with sequences from legumes, legume sister group families Polygalaceae, Surianaceae and Quillajaceae, and other angiosperm families for our genes of interest from prior studies (e.g., Citerne et al., 2003; Zhao et al., 2019; Madrigal et al., 2019). Phylogenetic trees are presented here as rooted, but in the face of high sequence divergence and difficulties in identifying a placement for the root in gene families where there has been high duplication (see Citerne et al., 2003), trees can be better

interpreted as unrooted networks. The CYC analysis was rooted using *Arabidopsis*. The RAD tree was rooted using *Oryza sativa* representing a monocot lineage of RAD. According to Madrigal et al. (2019), duplication events occurred before the monocot and eudicot split, therefore they have used sequences from the basal *Amborella*. I chose not to use the sequences from *Amborella* since Madrigal et al. (2019) reported difficulties in rooting these sequences, which could have implications in the topology of our tree. I did use two accessions from *Amborella* in the DIV phylogeny.

2.2.11 Protein structure

The structures of the candidate proteins were modelled using AlphaFold in a google collaborative notebook.

<https://colab.research.google.com/github/deepmind/alphafold/blob/main/notebooks/AlphaFold.ipynb>

Results

2.3.1 K-mer analysis: ploidy and levels of zygoty

With the outputs generated from Genomescope, k-mer distribution profiles were generated for the four species in this study with k-mer sizes varying from 21 to 121 in intervals of ten nucleotides. K-mer profiles provide useful information on the sequenced genomes such as heterozygosity and ploidy levels, repeat content, genome coverage (if previous information about the genome is unavailable) as well as potential contaminants that can be seen in the k-mer spectra. These features are relevant when choosing the assembling softwares, for example, some assemblers do not perform well with large or highly heterozygous genomes. Figure 4 shows the k-mer profiles for each species.

For *Cyathostegia mathewsii*, the k-mer spectra showed $k = 81$ as the best model with an estimated genome size of 442 Mbp. The single peak observed in the k-mer spectrum represents a continuous variation in the homozygous genome. For *Cadia purpurea*, $k = 81$ was also indicated as the best suited k-mer, indicating an estimated genome size of 567 Mbp. *Cadia*'s genome also showed a single peak in its kmer-spectrum, suggesting a homozygous genome. Taking into account the distribution of these species, both *Cyathostegia* and *Cadia* are found in relatively large populations (Lewis et al. 2003), which increase the chances of inbreeding

and therefore loss of allele variability. This explains the low heterozygosity levels of 0.31% in *Cyathostegia* and 0.28% in *Cadia* (Table 4).

Both *Cadia* and *Cyathostegia* showed relative differences when compared to the two *Andira* species. *Andira inermis* estimated genome size was 431 Mbp and heterozygosity level of 1.2% under a $k = 31$. *Andira anthelmia*'s best suit k -mer was $k = 21$ with an estimated genome size of 481 Mbp and heterozygosity level of 1.7%. The four species did not vary much in their genome size, but there is a clear difference between the heterozygosity levels of the two species of *Andira* and *Cadia* and *Cyathostegia* (Table 4).

The estimated repeat content found in the genome estimations, i. e. overrepresented sequences between genes, was uniform between species with the exception of *A. anthelmia*, which showed a length of 148 Mbp (Table 4). Genome repeat content is common and can make up a significant portion of genomes and represent tandem repeats, where the same pattern of nucleotides is repeated in DNA sequences consecutively, and transposable elements, which represent DNA sequences that can move within the genome. A few implications of genomes with a high repeat content include ambiguity in mapping sequencing reads to a reference genome. Since repetitive elements are often identical, this can lead to the wrong location of the correct read during the assembly process. Chimeric contigs, artificial sequences that combine parts from distinct genomic regions, can also be formed when reads from different copies of a repetitive element are incorrectly joined together, resulting in hybrid sequences that do not represent the true genomic structure.

Our data retrieved relatively different k -mer profiles for each species. *Andira inermis* and *Andira anthelmia* showed a double-peak curve whereas *Cadia purpurea* and *Cyathostegia mathewsii* showed a single-peak (Fig. 4). This reflects the more heterozygous genomes of the two species of *Andira*. The heterozygosity levels do not differ much between the two species, as their haploid (right) and diploid (left) peaks also show high variability in height (Fig. 4). The higher the peak, the more heterozygous the genome. No duplications are found in the genomes, as those would be represented by more peaks. It is noticeable in *Andira inermis* a pronounced error curve overlapping the reads coverage, which indicate a probable contamination.

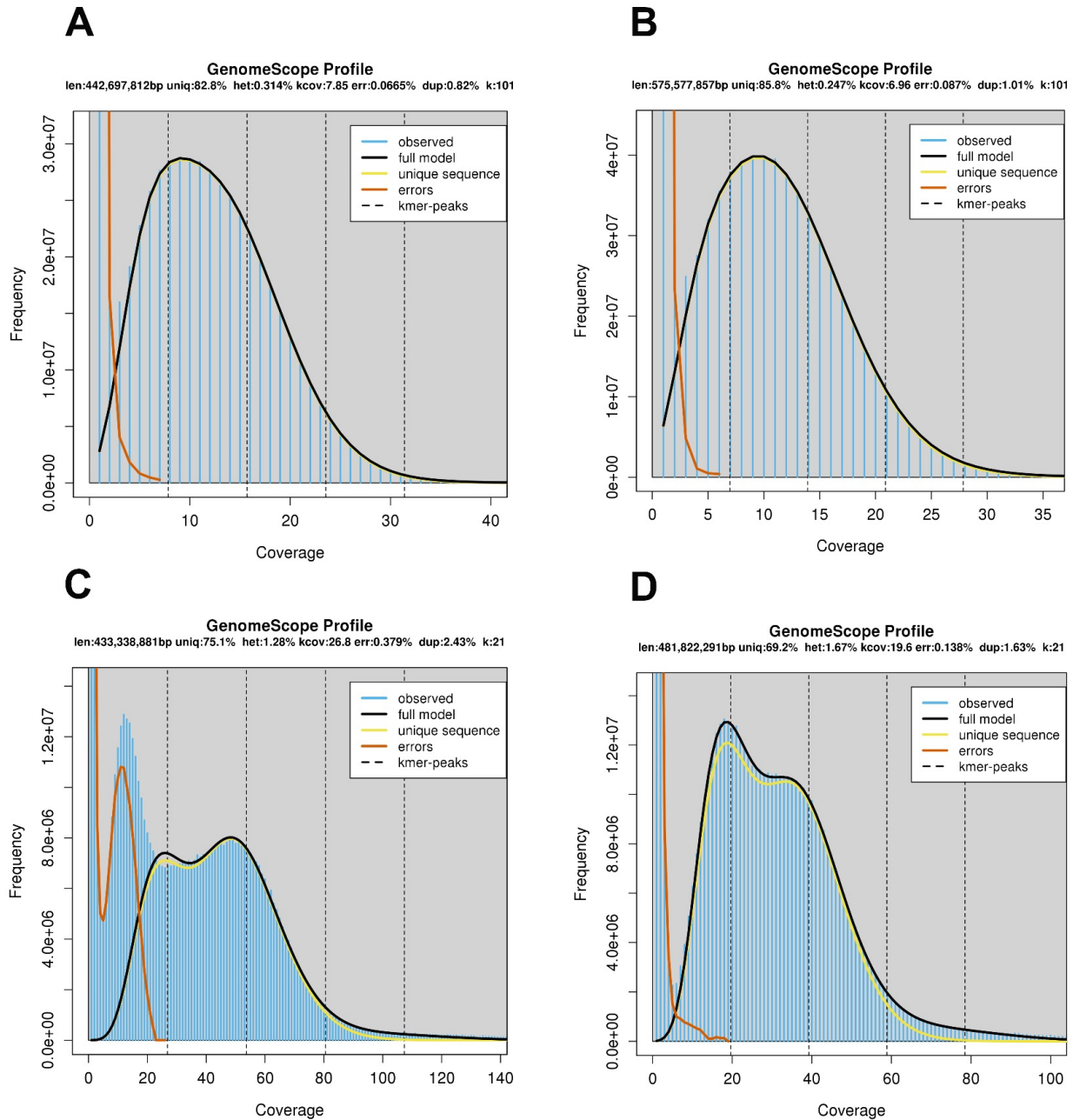


Figure 4. K-mer profiles of studied species generated by Genomescope. A. *Cyathostegia matheewsii*. B. *Cadia purpurea*. C. *Andira inermis* D. *Andira anthelmia*. The higher pics in each graph represents the homozygous portion of the genome whereas the smaller one indicates heterozygosity. The red line upon the read's coverage designates error and can indicate contamination, such as seen in *Andira inermis*.

2.3.2 Secondary protein structure of candidate genes

Protein secondary structure (PSS) prediction can be an important tool for studying protein structure and function. The protein secondary structure refers to the local conformation of the polypeptide backbone of proteins. There are basically two regular PSS states: alpha-helix (H) and beta-strand (E), and one irregular type: coil region (C). Overall, protein secondary structure can be regarded as a bridge that links the primary sequence and tertiary structure and thus, is used by many structure and functional analysis tools. The protein structures of the CYC, DIV and DRIF candidate proteins did not deviate from the structure of the orthologs in model species with confirmed function. However some changes were seen for RAD -like genes, which had longer N and C terminal tails (Figure 7). Further work would be needed to confirm if these are correct or due to errors in annotation.

2.3.4 Phylogenetic analyses of TCP and MYB genes in papilionoid legumes

CYCLOIDEA

Only one CYC copy, an ortholog of CYC2 (*sensu* Zhao et al., 2019; LEG-CYC *sensu* Citerne et al. 2003) was found in *Cadia* genome (Fig. 5). Citerne et al. (2003), however, recovered four copies, which is supported by our transcriptomic sequences (see Chapter Three; Fig. 5; Table 6). Our transcriptomic sequences from *Cadia* (Chapter Three) represent a collection from different developmental stages, ranging from vegetative buds to mature flowers, which is the probable reason that all four CYC copies were recovered, three of CYC2 and one of CYC1 (*sensu* Zhao et al., 2019). The failure to recover the three other copies in the genome sequenced here is possibly due to an incomplete assembly of the genomes using Platanus-allee, whose completeness showed a low BUSCO of 87.8%. SPADES assembler was also used to verify any improvement in completeness, but only a small increase to 89.5% was achieved (Table 5).

Our genome sequence data for *Cyathostegia* did not allow retrieval of any CYC copies, which may reflect that the unannotated assemblies had low BUSCO, 88% for Platanus-allee

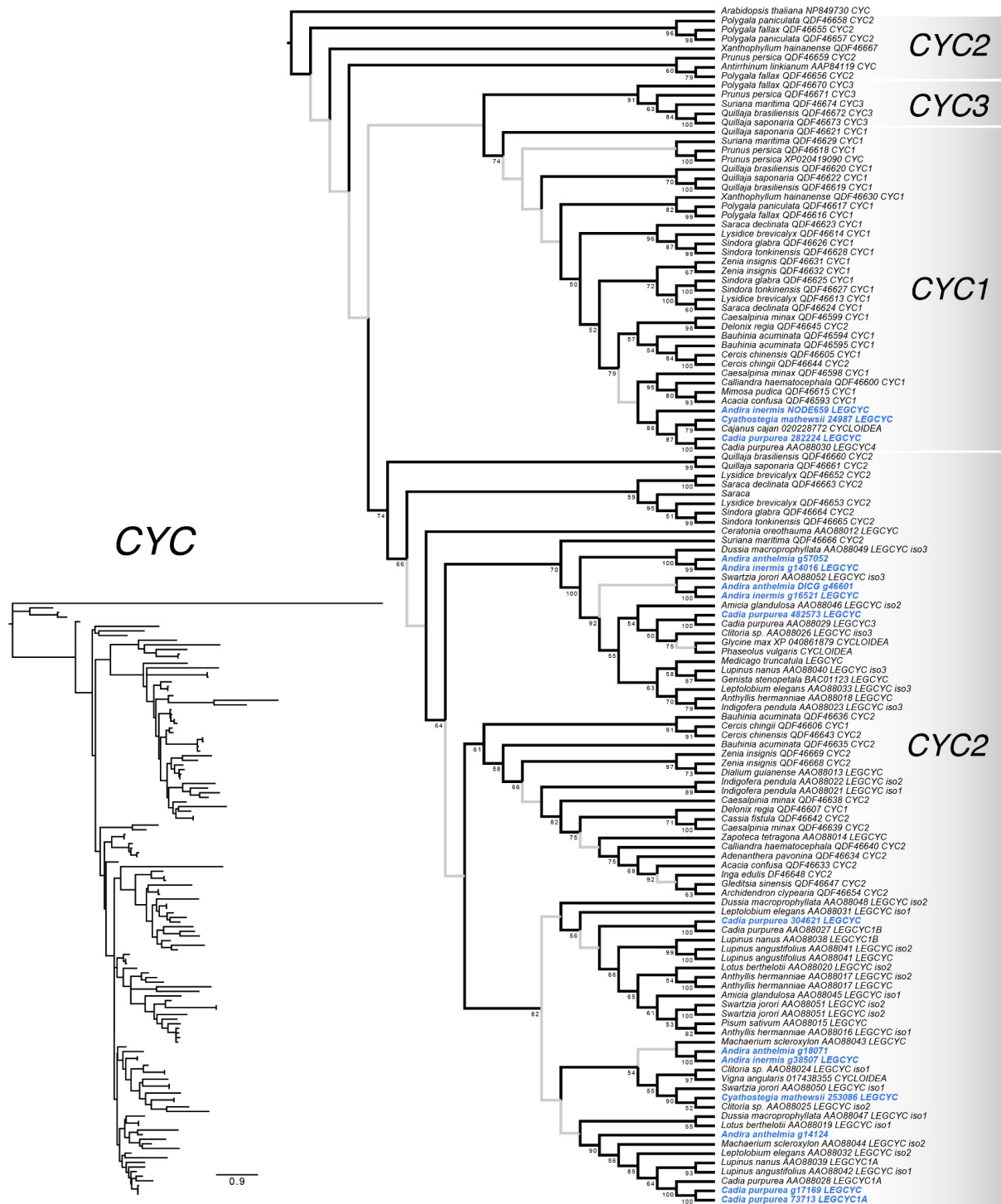


Figure 5. IQ-TREE maximum-likelihood phylogenetic tree and respective phylogram of *CYCLOIDEA* (CYC) protein sequences showing the placement of putatively new CYC orthologs (shaded in blue) of *Cadia purpurea*, *Cyathostegia mathewsii*, *Andira inermis* and *Andira anthelmia*. The CYC1, CYC2 and CYC3 lineages are highlighted. Outgroups are represented by the Fabales families Polygalaceae, Quilajaceae and Surianaceae, as well as *Arabidopsis*, *Prunus* and *Antirrhinum*. Legumes are clustered into two main groups represented by CYC1 and CYC2. The bootstrap values are given below branches, except for branches in grey that had <50% support.

and 93% for SPADES (Table 5). However, two CYC copies for *Cyathostegia mathewsii* were found in the transcriptomic sequences (see Chapter Three; Table 6), an ortholog of CYC1 and the other CYC2 (*sensu* Zhao et al., 2019). This confirms that the lack of matching genomic sequences does not reflect a loss of this gene family in *Cyathostegia*, but a possible artefact of the incomplete genome assembly. If sequencing spanned across more floral developmental stages, and maybe to a greater depth, more copies may have been revealed in the transcriptomes because it is possible that for *Cyathostegia*, other CYC orthologs are expressed in later stages of flower development. Nonetheless, at least one of the copies found in *Cyathostegia* falls phylogenetically close to another Swartzioidean genus, *Swartzia*, confirming its status as a CYC ortholog (Fig. 5).

Our genomic sequences from *Andira* recovered more CYC copies than those from *Cadia* and *Cyathostegia*. Four copies were found in both *Andira inermis* and *A. anthelmia*, representing copies of both CYC1 and CYC2 (*sensu* Zhao et al., 2019) (Table 6). Only one CYC copy was found in the *A. inermis* transcriptome while four were retrieved from the genome (Table 6). Transcriptomic sequences from *A. inermis* were assembled from vegetative buds only and this is perhaps the reason why other copies were not recovered- more copies are likely to be recovered from RNA from developing flower buds. No transcriptome sequences were available from *Andira anthelmia*.

RADIALIS

Our genomic data identified four copies of RAD in *Cadia* (Fig. 6). When complemented with transcriptomic sequences (see Chapter Three), three of these copies were confirmed in the transcriptome, and a fifth sequence was found only in the transcriptome. Taking both genomic and transcriptomic data together, five putative RAD copies were found (Fig. 6) (Table 7). Four copies were found nested in RAD3 clade (2 with genome and transcriptome support, one genome only, one transcriptome only) and one copy (genome and transcriptomic support) in RAD2 clade *sensu* Madrigal et al. (2019) (Fig. 6).

No putative RAD sequences were recovered from the *Cyathostegia* genome, but four putative orthologs from the transcriptome were recovered (see Chapter Three) (Table 7), all of which are found in the RAD3 clade (Fig. 6). Three sequences are clustered together and difficult to interpret since they have long branches suggesting high variation

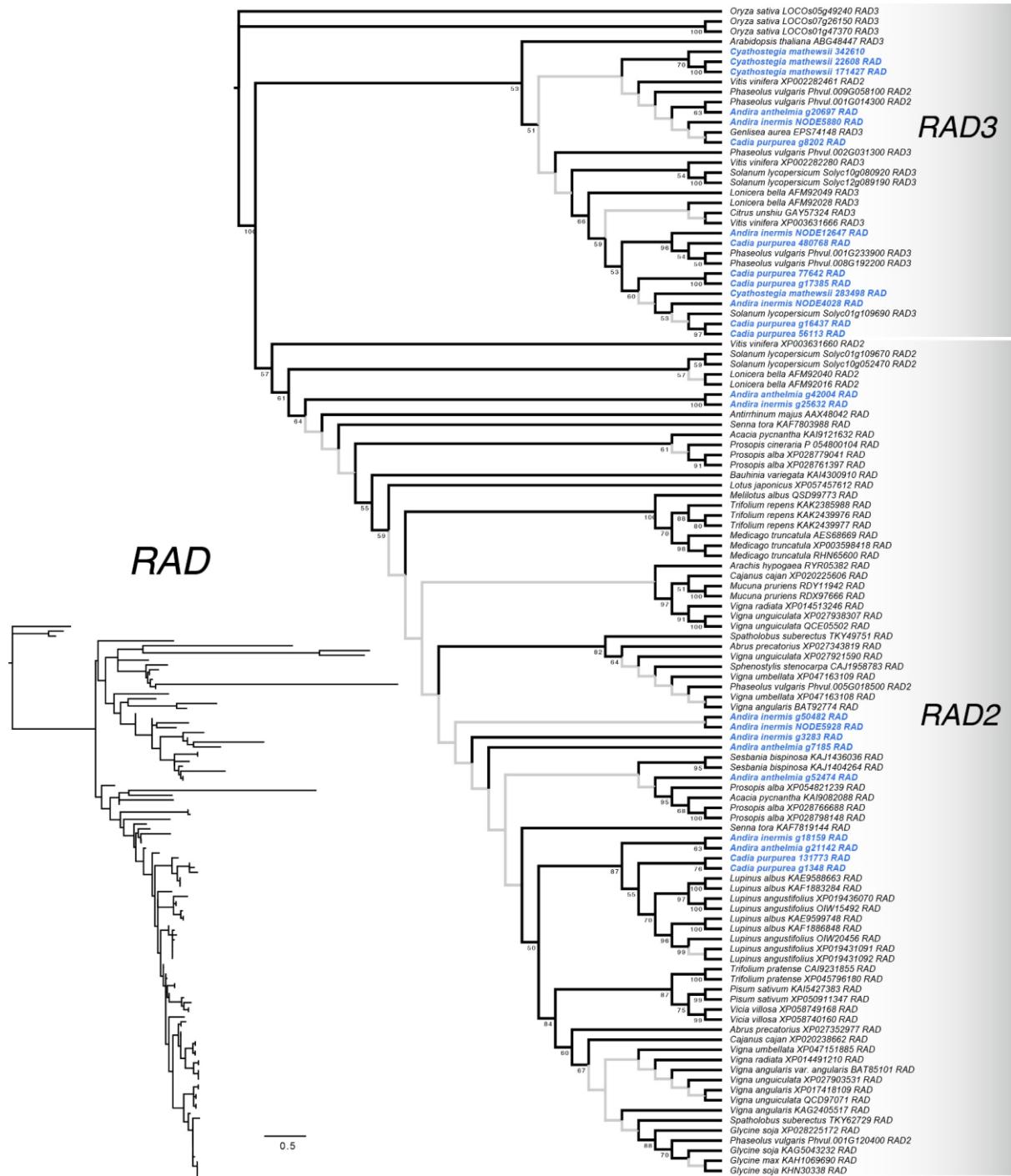


Figure 6. IQ-TREE maximum-likelihood phylogenetic tree and respective phylogram of *RADIALIS* (RAD) protein sequences showing the placement of putatively new RAD orthologs (shaded in blue) of *Cadia purpurea*, *Cyathostegia mathewsii*, *Andira inermis* and *Andira antheimia*. The RAD2 and RAD3 lineages *sensu* Madrigal et al. (2019) are highlighted. The great majority of legumes is clustered within RAD2 and RAD3. Outgroups are represented by accessions of the monocot *Oryza sativa*. The bootstrap values are given below branches, except for branches in grey that had <50% support.

relative to the rest of the clade. The predicted secondary structure of these isolated sequences (Cyathostegia_scaffold22608, Cyathostegia_scaffold342610, and Cyathostegia_scaffold171427) differ from the model in *Antirrhinum*, but it is possible to see the three characteristic helixes of the MYB family, however in a longer chain with another group of helixes. Only Cyathostegia_scaffold283498 secondary structure is more comparable to the model (Fig. 7).

Our genomic data suggested three RAD copies for *Andira inermis*, but when complemented with transcriptome data (Ch 3) seven RAD copies were found, with three copies nested in RAD3 clade and four in RAD2. From these, only one copy is supported by both transcriptomic and genomic sequences. The lack of transcriptomic pairs for *A. inermis* is probably due to the lack of expression in the vegetative buds for these copies, and floral buds might reveal more. Five copies were found in *A. anthelmia*'s genome with one copy in RAD3 and the remaining four in RAD2, confirming the high copy number of this locus in *Andira* species, even in the absence of transcriptome data (Table 7).

DIVARICATA

Our genomic data identified two copies of DIV in *Cadia* (Fig. 8). When complemented with transcriptomic sequences (Chapter Three), these two copies were confirmed, and two more sequences were found in the transcriptome. Taking both genomic and transcriptomic data together, four putative DIV copies were found. Except for the two isolated transcriptomic sequences, the two copies with a genomic pair were found close to sequences of the closely related genistoid legume *Lupinus*.

One putative copy of DIV was recovered from our *Cyathostegia* genome and the same copy from the transcriptome (Table 8). The single sequences from the genome and transcriptome suggest that that DIV may be only found in a single copy in *Cyathostegia*. However, a more deeply sequenced and fully assembled genome of *Cyathostegia* might reveal more RAD copies.

Our genomic data suggested three copies of DIV for *A. inermis*, and four copies in total when one extra transcriptomic copy is added. Three genomic sequences therefore have transcriptomic support and are scattered across the phylogeny. Similarly, four copies of DIV were recovered in *A. anthelmia*'s genome, which are placed alongside with *A. inermis* sequences (Fig. 8).

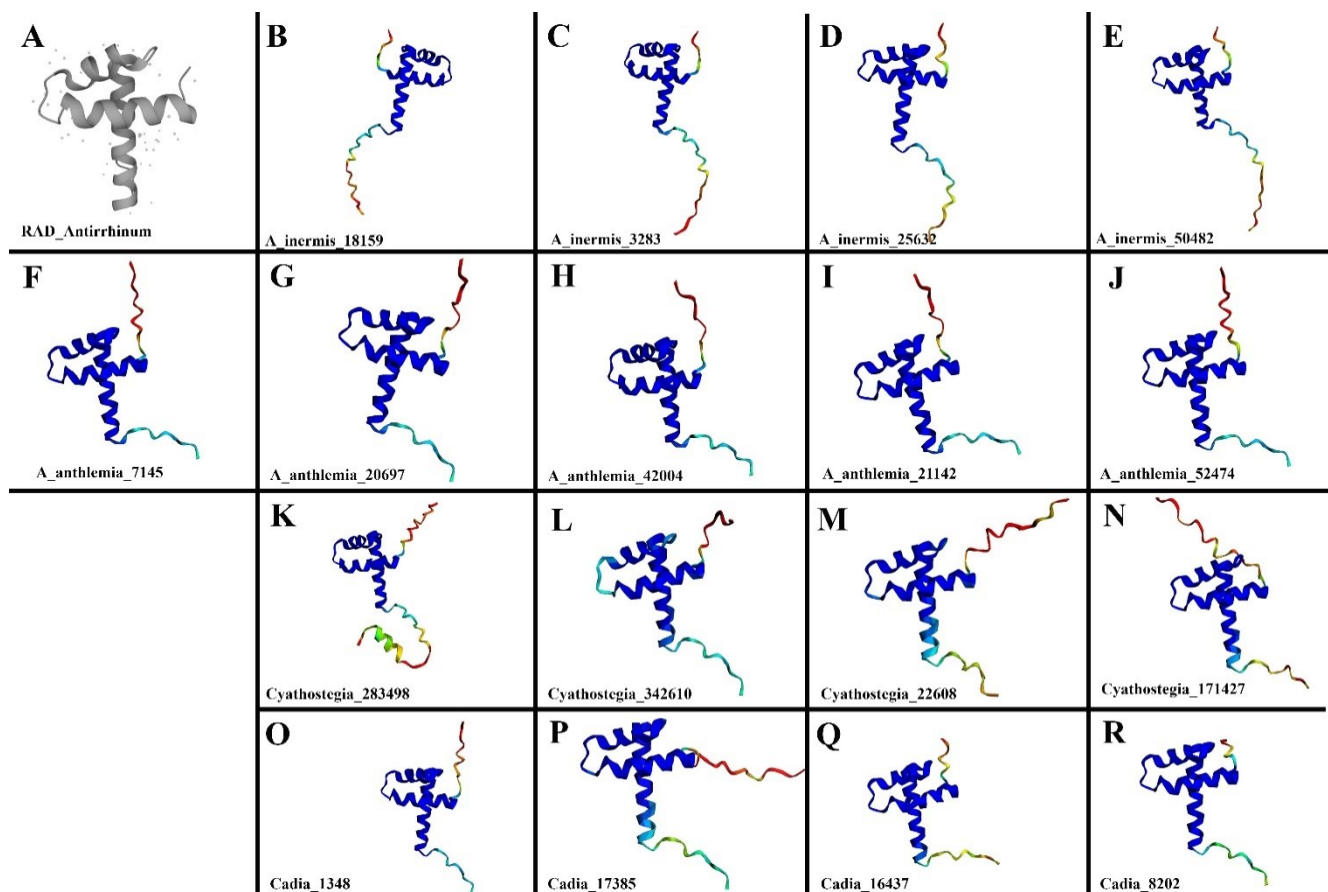


Figure 7. Secondary protein prediction generated by AlphaFold. A canonical RAD protein from *Antirrhinum majus* is depicted in A. B-E are paralogs of RAD of *Andira inermis* F-J are the five copies found in *Andira anthelmia*. K-N represents the four paralogs found in *Cyathostegia mathewsii* and O-R four paralogs of RAD for *Cadia purpurea*. Long N and C terminal tails are seen in some paralogs but the main structure of the proteins are well conserved across the different species.



DRIF

Our data from genome and transcriptome only recovered one copy of DRIF from both *Cadia* and *Cyathostegia* (Table 9). DRIF is not widely sampled in legumes, most of the available sequences are from model species of economic importance such *Glycine max*, *Cicer arietinum*, and *Pisum sativum*, from which a large range of genomic resources is available. Although not comprehensively sampled, our phylogeny is consistent with the placement of *Cadia* closely related to *Lupinus*, and *Cyathostegia* appears as sister group to all papilionoid legumes (Fig. 9).

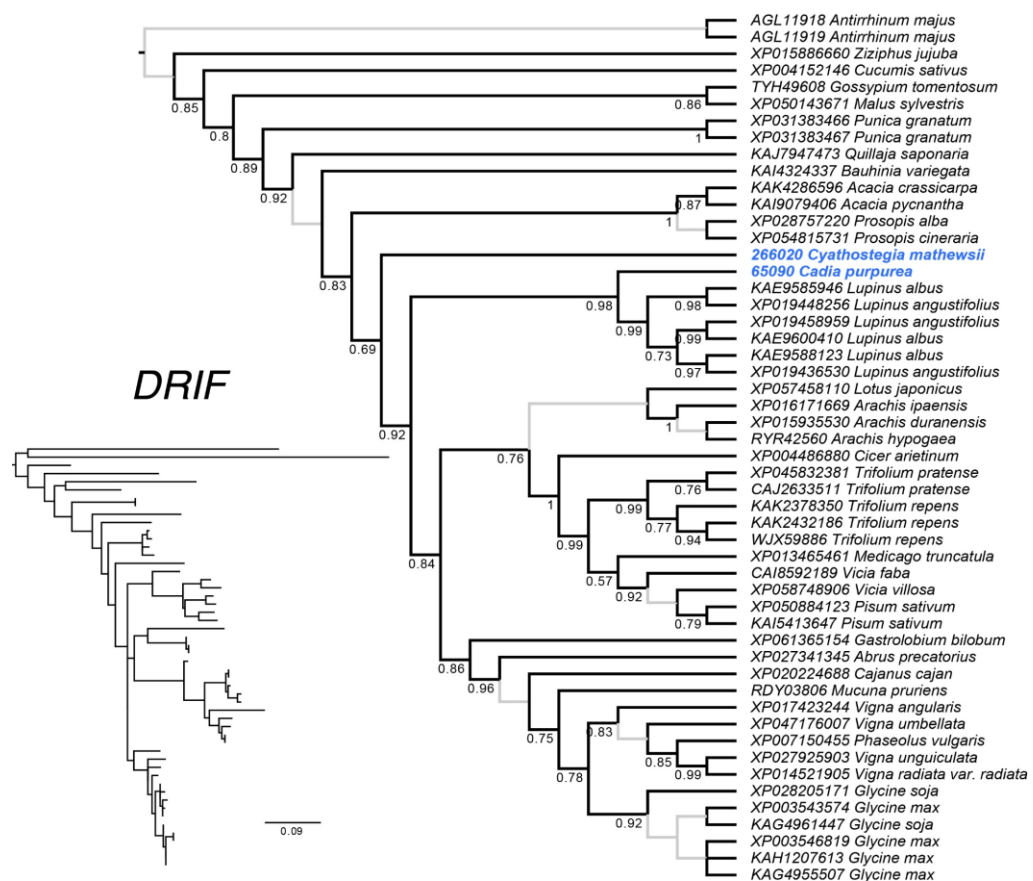


Figure 9. IQ-TREE maximum-likelihood phylogenetic tree and respective phylogram of *DIV-and-RAD-Interacting-Factor* (DRIF) protein sequences showing the placement of putatively new DRIF orthologs (shaded in blue) of *Cadia purpurea* and *Cyathostegia mathewsii*. No DRIF orthologs were recovered from either the genome or transcriptome of *Andira inermis* and *Andira anthelmia*. Given the lack of available DRIF sequences, our sampling was limited to only a few legume sequences and few eudicots. The bootstrap values are given below branches, except for branches in grey that had <50% support.

Discussion

Heterozygosity

Our data showed differences in heterozygosity amongst the genera and species examined. The *Andira* species (*A. anthelmia* 1.7%; *A. inermis* 1.2%) showed higher heterozygosity than *Cadia* (0.28%) and *Cyathostegia* (0.31%). These differences in heterozygosity may reflect differences in pollination biology. *Andira anthelmia* has been shown to be visited exclusively by large xylocopid bees in Bahia, Brazil, which are known to fly long distances (Pennington, 2003). In contrast the smaller flowered *A. inermis* is visited by numerous species of smaller bees (Frankie, 1976), which are likely to fly shorter distances and therefore be less effective in transporting pollen from distant trees that are likely to be genetically more different. In this pollination context, it is interesting to consider that *Cyathostegia* often grows as a pioneer species in dense stands (Pennington, pers. comm.), and its pollination by beetles (Lewis et al., 2003) may result in pollination between neighbouring plants as the beetles move through these stands which might lead to lower rates of heterozygosity. *Cadia* may be pollinated by birds (Pennington et al. 2000; Boatright et al., 2008), which fly long-distances, but this is not confirmed. The lower heterozygosity levels suggest though that pollination may be carried out by pollinators that fly short distances between stands, keeping gene flow only between local populations.

TCP and MYB genes in papilionoid legumes

CYCLOIDEA

Our genomes and transcriptomes show that a CYC1 ortholog is present in *Andira*, *Cyathostegia* and *Cadia*. Given the divergent morphologies of the flowers of these genera - *Andira* being standard papilionate, *Cyathostegia* with a single standard petal and multiple stamens, and *Cadia* radially symmetrical, the presence of a functional CYC1 copy in all of them may argue against a role in establishing zygomorphy, multiple stamens and in controlling petal number. Zhao et al. (2019) used a similar argument to dismiss a role for CYC1 in establishing zygomorphy in the legumes as a whole based on its uniform presence across subfamilies with different floral symmetries. CYC1 has been shown to be differentially expressed in *Cadia* and *Cyathostegia* floral and vegetative tissues (see Chapter Three), also

demonstrated by Bartlett and Specht in Zingiberales (2011), however Zhao et al. (2019) reported that no studies have associated CYC1 genes with floral zygomorphy.

For CYC2 (Citerne's LEG-CYC), Zhao et al. (2019) suggested that duplication and higher copy number is related to the establishment of zygomorphy in subfamily Papilionoideae. However, in the context of the new data presented here, that conclusion seems premature. Our new data for taxa with unusual floral form in Papilionoideae (*Cadia*, *Cyathostegia*), alongside existing data reported by Zhao et al. (2019) and Citerne et al. (2003) fails to reveal any clear correlation of CYC2 copy number with floral form in papilionoid legumes. There are multiple copies of CYC2 in the NPAAA clade where the typical papilionoid flower is uniform (see Chapter Four; e.g., three copies in *Anthyllis*, *Indigofera*; four in *Cajanus*, *Glycine*; see Fig. 5 and Table 1 in Zhao et al., 2019). However, there are also high copy numbers in taxa such as with radial flowers such as *Cadia* (4) and *Leptolobium elegans* [reported as *Acosmium subelegans* in Citerne et al. (2003) and Zhao et al. (2019)], and flowers with a single petal and multiple stamens (e.g., *Swartzia* with three copies). Furthermore, Zhao et al. (2019) do not highlight that there are non-papilionoids with zygomorphic corollas such as *Bauhinia* and *Caesalpinia* that also have multiple copies of CYC2 (see Zhao et al., 2019, Table 1). In failing to consider the full diversity of floral form across Papilionoideae and legumes more generally, the conclusion that multiple copies of CYC2 are central to legume zygomorphy seems simplistic. Citerne et al. (2006) showed that differential expression of the multiple CYC2 copies in *Cadia* resulted in a radially symmetrical corolla, highlighting the importance of gene expression based studies that are a focus for Chapter Three.

RADIALIS, DIV and DRIF

Madrigal et al. (2019) surveyed RAD and DIV in a phylogenetic context across all angiosperms, but in their paper legumes were represented only by *Phaseolus vulgaris* (see Maridgal et al. 2019, Supp. Table 1). A new dataset was assembled bringing together a unique dataset for RAD and DIV sequences that samples legumes more thoroughly. In general, these legume data confirm the high copy numbers of DIV and RAD found in other angiosperms by Madrigal et al. (2019).

Our data from genome and transcriptome recovered a varied number of RAD copies among our target species. In total, *Cyathostegia* has three copies, *Cadia* and *A. anthelmia* have five each, and *A. inermis* has seven. Madrigal et al. (2019) found that RAD underwent at least two duplications before the monocot and eudicot diversification, resulting in RAD1/RAD3-

like and RAD2-like orthologs. Except for *Cyathostegia*'s sequences, which are all clustered in RAD3, sequences from *Cadia*, *A. inermis* and *A. anthelmia* are found in both RAD3 and RAD2.

A varied number of putative DIV orthologs were found in our data. In a similar pattern as RAD, but with lower copy number, *Cyathostegia* has two copies, *Cadia* and *A. anthelmia* have four, and *A. inermis* has five. DIV sequences are found scattered across the phylogeny marking putative duplication events that gave rise to the number of copies that were identified. Single putative DRIF orthologs were found in *Cadia* and *Cyathotegia*. No DRIF was found in either *A. inermis* or *A. anthelmia*.

RAD, DIV, and DRIF have been shown to regulate the dorsoventral asymmetry flowers of *Antirrhinum* by the antagonising activity of DIV in the dorsal region via interaction with CYC (Corley et al., 2005). I expect that these genes also regulate development along the the dorsoventral axis in all our legume species. It is difficult to establish a correlation between copy number when these are so variable and the data must be based on draft genomes and transcriptomes. However, the typical papilionate *A. anthelmia* and *A. inermis* do have a higher number of RAD copies, which is known to determine dorsal identity in the floral meristem. The adaxialised radial flowers of *Cadia* with five copies of RAD could suggest that the expanded adaxial domain is correlated with increased copy number of RAD. However, copy number variation alone is not strong evidence, and the lack of such a correlation in *Andira* suggests the high copy number is not causa. Examination of the variation in expression patterns between gene family members is needed to test this hypothesis. Chapter Three uses RNASeq to gain further insights into the genetic variation underlying legume floral diversity.

In this chapter I have investigated the role of the candidate genes CYC, RAD, DIV, and DRIF in the flower development of radially symmetric *Cadia purpurea*, the single-petalled with numerous stamens *Cyathostegia mathewsii*, and the typical papilionated-flowered *Andira inermis* and *Andira anthelmia*. Genomes from *Cadia*, *Cyathostegia* and the papilionate *A.inermis* and *A. anthelmia* were sequenced in order to look for putative orthologs using a phylogenetic approach. Results showed a high copy number of CYC genes in the papilionate *A. inermis* (four copies) and *A. anthelmia* (four copies) and in the radially symmetric *Cadia* (four copies). Only two copies were of CYC were found in *Cyathostegia*. These results in the context of prior studies that sampled these TCP and MYB genes in other legumes (Citerne et al. 2003; Zhao et al. 2019) suggest that a similar number of multiple CYC copies found in papilionoid genera with unusual flowers morphologically argues against the premise that a

higher number of CYC copies is correlated with strong zygomorphic symmetry (Zhao et al. 2019). For example, Citerne et al. (2003) found high copy numbers in taxa with radial flowers such as *Cadia* and *Leptolobium* (Citerne et al. 2003). My study confirms the high copy number of CYC genes in *Cadia* found by Citerne et al. (2003) and additionally brings data from another morphologically unusual papilionoid genus *Cyathostegia*, which suggests that reduction or loss of petals might be related to a lower number of copies of CYC. A wider sampling including single-petal species and lineages that have completely lost their corolla will shed light into the nature of gene copy number underlying such morphologies.

This chapter brings additional data to understand the genetics of unusual flowers in papilionoid genera since the first insights brought to light by Pennington et al. (2001) and Wojciechowski et al. (2004), passing through Citerne's experiments with Real Time PCR and *In situ* hybridization. In this chapter I analysed data at the genomic level using additional transcriptomic data to support my findings.

References

- Arroyo, M. T. K. 1981. Breeding systems and pollination biology in Leguminosae. In: Polhill, R. M. & Raven, P. H. (eds) *Advances in Legume Systematics*, part 2. Royal Botanic Gardens, Kew, UK, pp 723–769.
- Armbruster, W. S. 2014. Floral specialization and angiosperm diversity: phenotypic divergence, fitness trade-offs and realized pollination accuracy. *AoB Plants* 6(0): plu003.
- Bartlett, M. E. & C. D. Specht. Changes in expression pattern of the *teosinte branched1-like* genes in the Zingiberales provide a mechanism for evolutionary shifts in symmetry across the order. *Evolution and Phylogeny* 98(2): 227–243.
- Bello, M. A., P. Cubas, I. Alvarez, G. Sanjuanbenito, J. Fuertes-Aguilar. 2017. Evolution and expression patterns of CYC/TB1 genes in *Anacyclus*: phylogenetic insights for floral symmetry genes in Asteraceae. *Frontiers in Plant Science* 8: 589.
- Benson, D. A., M. Cavanaugh, K. Clark, I. Karsch-Mizrachi, J. Ostell, K. D. Pruitt, E. W. Sayers. 2017. GenBank. *Nucleic Acids Research* 46(D1): D41–D47.
- Boatwright, S. J., V. Savolainen, B. E. Van Wyk, A. L. Schutte-Vlok, F. Forest, M. van der Bank. 2008. Systematic position of the anomalous genus *Cadia* and the phylogeny of the tribe Podalyrieae (Fabaceae). *Systematic Botany* 1: 133–147.
- Bolger, A. M., M. Lohse, B. Usadel. 2014. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* 30: 2114–2120.
- Bruneau, A. et al. 2014. Floral evolution in the Detarieae (Leguminosae): phylogenetics evidence for labile floral development in an early-diverging legume lineage. *International Journal of Plant Science* 175: 392–417.
- Cardoso, D. et al. 2015. Filling in the gaps of the papilionoid legume phylogeny: the enigmatic Amazonian genus *Petaladenium* is a new branch of the early-diverging Amburaneae clade. *Molecular Phylogenetics and Evolution* 84: 112–124.
- Cardoso, D. et al. 2013. Reconstructing the deep-branching relationships of the papilionoid legumes. *South African Journal of Botany* 89: 58–75.
- Cardoso, D., L. P. Queiroz, R. T. Pennington, H. C. Lima, E. Fonty, M. F. Wojciechowski, M. Lavin. 2012. Revisiting the phylogeny of papilionoid legumes: new insights from comprehensively sampled early-branching lineages. *American Journal of Botany* 99: 1991–2013.
- Chickhi, R., Medvedev P. 2014. Informed and automated k-mer size selection for genome assembly. *Bioinformatics* 30: 31–37.

- Citerne, H., Florian J., Sophie N., Catherine D. 2010. The evolution of floral symmetry. *Advances in Botanical Research*. 54: 85–128.
- Citerne, H. L. et al. 2006. An apparent reversal in floral symmetry in the legume *Cadia* is a homeotic transformation. *Proceedings of the International Academy of Sciences* 103(32): 12017–12020.
- Citerne, H., D. Luo, R. T. Pennington, E. Coen, Q. C. B. Cronk. 2003. A phylogenetic investigation of CYCLOIDEA-like TCP genes in the Leguminosae. *Plant Physiologist* 131(3): 1042–53.
- Corley, S. B., R. Carpenter, L. Copsey, E. Coen. 2005. Floral asymmetry involves an interplay between TCP and MYB transcription factors in *Antirrhinum*. *Proceedings of the International Academy of Sciences* 102(14): 5068–5073.
- Cowan, R. S. 1981. Swartzieae. Pp. 209–2012 in *Advances in Legume Systematics*, eds. R. M. Polhill & P. H. Raven. Royal Botanic Gardens, Kew.
- Cubas, P., Nick L., John, D., Enrico C. 1999. The TCP domain: a motif found in proteins regulating plant growth and development. *The Plant Journal* 18(2): 215–222.
- Luo, D., R. Carpenter, L. Copsey, C. Vincent, J. Clark, E. Coen. 1999. Control of organ asymmetry in flowers of *Antirrhinum*. *Cell* 99: 367–376.
- Du, H., B. R. Feng, S. S. Yang, Y. B. Huang, Y. X. Tang. 2012a. The R2R3-MYB transcription factor gene family in maize. *PLoS One* 7(6): e37463.
- Du, H., S. S. Yang, Z. Liang, B. R. Fang, L. Liu, Y. B. Huang, Y. X. Tang. 2012b. Genome-wide analysis of the MYB transcription factor superfamily in soybean. *BMC Plant Biol.* 12:106.
- Du, H., L. Zhang, L. Liu, X-F Tan., W-J Yang, Y-M Wu, Y-B Huang, Y-X Tang. 2009. Biochemical and molecular characterization of plan MYB transcription factor family. 2009. *Biochemistry* 74(1): 1–11.
- Du Puy, D. J., J.-N. Labat, R. Rabevohitra, J.-F. Villiers, J. Bosser, J. Moat. 2002. The Leguminosae of Madagascar. Royal Botanic Gardens, Kew.
- Feng, X, Zhao Z, Tian Z, Xu S, Luo Y, Cai Z, Wang Y, Yang J, Wang Z, Weng L, Chen J, Zheng L, Guo X, Luo J, Sato S, Tabata S, Ma W, Cao X, Hu X, Sun C, Luo D. 2006. Control of petal shape and floral zygomorphy in *Lotus japonicus*. *Proceedings of the National Academy of Sciences U.S.A.* 103: 4970–4975.

- Frankie, G. W., P. A. Opler, K. S. Bawa. 1976. Foraging behaviour of solitary bees: implications for out-crossing of Neotropical forest tree species. *Journal of Ecology* 64: 1049–1057.
- Galego, L. & J. Almeida. 2002. Role of DIVARICATA in the control of dorsoventral asymmetry in *Antirrhinum* flowers. *Genes and Development* 16(7): 880–891.
- Gaudin, V. et al. 2000. The expression of D–Cyclin genes defines distinct developmental zones in snapdragon apical meristem and is locally regulated by the Cycloidea gene. *Plant Physiology* 122: 1137–1148.
- Gubtiz, T., A. Caldwell, A. Hudson. 2003. Rapid molecular evolution of CYCLOIDEA-like genes in *Antirrhinum* and its relatives. *Molecular Biology and Evolution* 20(9): 1537– 44.
- Hawkeswood, T. J. 1983. Observations on *Pyrgoides dryops* (Blackburn) (Coleoptera: Chrysomelidae), a pollen-feeding beetle on *Acacia leiocalyx* (Domin) Pedley, at Brisbane, south-east Queensland. *The Victorian Naturalist* 100(4):156–158.
- Hawkeswood, T. J. 1989. Notes on *Diphucephala affinis* (Coleoptera: Scarabaeidae) associated with flowers of *Hibbertia* and *Acacia* in Western Australia. *Plant Systematic and Evolution* 168: 1–5.
- Hsin, K.-T. & C.-N. Wang. 2018. Expression shifts of floral symmetry genes correlate to flower actinomorphy in East Asia endemic *Conandron ramondioides* (Gesneriaceae). *Botanical Studies* 59: 24.
- Hilleman, L. C. 2014. Bilateral flower symmetry – how, when and why? *Current Opinion in Plant Biology* 17: 146–152.
- Hughes, C., R. Nyffeler, H. P. Linder. 2015. Evolutionary plant radiations: where, when, why and how? *New Phytologist* 207: 249–253.
- Janzen, D. H. 1989. Natural history of a wind-pollinated Central American dry forest legume tree(*Ateleia herbert-smithii* Pitier), In: C. H. Stirton & J. L. Zaruchi (eds.), *Advances in Legume Biology. Monographs in Systematic Botany from the Missouri Botanical Garden* 29: 293–376.
- Ireland, H. E. 2007. Taxonomic changes in the South American genus *Bocoa* (Leguminosae-Swartzieae): reinstatement of the name *Trischidium* and a synopsis of both genera. *Kew Bulletin* 62: 333–349.
- Ireland, H. E., Pennington, R. T., Preston, J. 2000. Molecular systematics of the Swartzieae. In: Herendeen, P.S., Bruneau, A. (Eds.), *Advances in Legume Systematics*. Royal Botanic Gardens, Kew, 9: 217–231.

- Jumper, J. et al. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* 596: 583–589.
- Kajitani, R., D. Yoshimura, M. Okuno, Y. Minakuchi, H. Kagoshima, A. Fujiyama, K. Kubokawa, Y. Kohara, A. Toyoda, T. Itoh. 2019. Platanus-alley is a de novo haplotype assembler enabling a comprehensive access to divergent heterozygous regions. *Nature Communications* 10: 1702.
- Klitgaard, B. B. et al. 2013. A detailed investigation of the Pterocarpus clade (Leguminosae: Dalbergieae): *Etabalia* with radially symmetrical flowers is nested within the papilionoid-flowered *Pterocarpus*. *South African Journal of Botany* 89: 128–142.
- Lavin, M. et al. 2005. Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the tertiary. *Systematic Biology* 54(4): 530–549.
- Larsson, A., 2014. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* 30(22): 3276–8.
- Leite, V. G. et al. 2015. Floral development of the early-branching papilionoid legume *Amburana cearensis* (Leguminosae) reveals rare and novel characters. *International Journal of Plant Sciences* 176(1): 94–106.
- Lewis, G. P., J. T. Knudsen, B. B. Klitgaard, R. T. Pennington. 2003. The floral scent of *Cyathostegia mathewsii* (Leguminosae, Papilionoideae) and preliminary observations on reproductive biology. *Biochemical Systematics and Ecology* 31(9): 951–962.
- Li, H., W. Wang, L. Lin, X. Zhu, J. Li, X. Zhu, Z. Chen. 2013. Diversification of the phaseoloid legumes: effects of climate change, range expansion and habit shift. *Frontiers in Plant Science* 4: 386.
- Li, Xin, W. Liu, L. Zhuang, Y. Zhu, F. Wang, T. Chen, J. Yang, M. Ambrose, Z. Hu, J. L. Weller, D. Luo. 2019. BIGGER ORGANS and ELEPHANT EAR-LIKE LEAF1 control organ size and floral organ internal asymmetry in pea. *Journal of Experimental Botany* 70(1): 179–191.
- Lucibelli, F., M. C. Valoroso, S. Aceto. 2020. Radial or bilateral? The molecular basis of floral symmetry. *Genes* 11(4): 395.
- MacMahon, M. & Larry H. 2004. Phylogeny of Amorphaeae (Fabaceae, Papilionoideae). *American Journal of Botany* 91(8): 1219–1230.
- Madrigal, Y., J. F. Alzate, F. Gonzalez, N. Pabon-Mora. 2019. Evolution of RADIALIS and DIVARICATA gene lineages flowering plants with an expanded sampling in non-core eudicots. *American Journal of Botany* 106(3): 334–351.

- Mansano, V. F., S. C. Tucker, A. M. G. A. Tozzi. 2002. Floral ontogeny of *Lecointea*, *Zollernia*, *Exostyles*, and *Harleyodendron* (Leguminosae, Papilionoideae: Swartzieae S.L.). *American Journal of Botany* 89 (10): 1553–1569.
- Martin, M. 2011. CUTADAPT removes adapter sequences from high-throughput sequencing reads. *EMBnet journal* 17(1): 10–12.
- Martin-Trillo, M. & P. Cubas. 2010. TCP genes: a family snapshot ten years later. *Trends in Plant Science* 15(1): 31–9.
- Miller, M. A., W. Pfeiffer, T. Schwartz, 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *Proceedings of the Gateway Computing Environments Workshop (GCE)*, New Orleans, LA. pp 1–8.
- Pennington, R. T., B. B. Klitgaard, H. Ireland, M. Lavin. 2000. New insights into floral evolution of basal Papilionoideae from molecular phylogenies. Pp. 233–248 in: *Advances in Legume Systematics* 9, eds. P.S. Herendeen and A. Bruneau. Kew: Royal Botanic Gardens.
- Pennington, R. T., M. Lavin, H. Ireland, B. Klitgaard, J. Preston, J. M. Hu. 2001. Phylogenetic relationships of basal papilionoid legumes based upon sequences of the chloroplast trnL intron. *Systematic Botany* 26: 537–556.
- Prenner, G. 2004. Floral development in *Polygala myrtifolia* (Polygalaceae) and its similarities with Leguminosae. *Plant Systematics and Evolution* 249: 67–76.
- Prenner, G. & D. Cardoso. 2017. Flower development of *Goniorrachis marginata* reveals new insights into the evolution of the early diverse detarioid legumes. *Annals of Botany* 119(3): 417–432.
- Prenner, G. et al. 2015. Flowers of the early-branching papilionoid legume *Petaladenium urceoliferum* display unique morphological and ontogenetic features. *American Journal of Botany* 102: 1780–93.
- Preston, J. C., M. A. Kost, L. C. Hilleman. 2009. Conservation and diversification of the symmetry developmental program among close relatives of snapdragon with divergent floral morphologies. *New Phytologist* 182(3): 751–762.
- Prjibelski, A., D. Antipov, D. Meleshko, A. Lapidus, A. Korobeynikov. 2020. Using SPAdes de novo assembler. *Current Protocols in Bioinformatics* 70(1): 1–29.
- Polhill, R. M. 1981. Sophoreae. In: Polhill R. M. & Raven, P. H. (eds.), *Advances in Legume Systematics*. Royal Botanic Gardens, Kew, 1: 213–230.

- Polhill, R. M. 1994. Classification of the Leguminosae. In: Bisby, F.A., Buckingham, J., Harborne, J. B. (Eds.), *Phytochemical Dictionary of the Leguminosae. Plants and Their Constituents*. Chapman and Hall, London. 1: 25–27.
- Ramos, G., H. C. Lima, G. Prenner, L. P. Queiroz, C. E. Zartman, D. Cardoso. 2016. Molecular systematics of the Amazonian genus *Aldina*, a phylogenetic enigmatic ectomycorrhizal lineage of papilionoid legumes. *Molecular Phylogenetics and Evolution* 97: 11–18.
- Rannalo-Benavidez, T. R., K. S. Jaron, M. C. Schatz. 2020. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nature Communications* 11: 1432.
- Raimundo, J., Sobral, R., Bailey, P., Azevedo, H., Galego, L., Almeida, J., E. Coen, M. M. R. Costa. 2013. A subcellular tug of war involving three MYB-like proteins underlies a molecular antagonism in *Antirrhinum* flower asymmetry. *The Plant Journal* 75: 527–538.
- Reardon, W., P. Gallagher, K. M. Nolan, H. Wright, M. C. Cardenosa-Rubio, C. Bragalini, C-S. Lee, D. A. Fitzpatrick, K. Corcoran, K. Wolff, J. M. Nugent. 2014. Different outcomes for the MYB floral symmetry genes *DIVARICATA* and *RADIALIS* during the evolution of derived actinomorphy in *Plantago*. *New Phytologist* 202: 716–725.
- Sauquet, H., et al. 2017. The ancestral flower of Angiosperms and its early diversification. *Nature Communications* 8: 16047.
- Sinjushin, A. A. 2018. Floral ontogeny in *Cordyla pinnata* (A. Rich) Milne-Redh. (Leguminosae, papilionoideae): away from stability. *Flora* 241: 8–15.
- Stamatakis, A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688–2690.
- Stamatakis, A., Hoover, P., Rougemont, J. 2008. A rapid bootstrap algorithm for the RAxML web-servers. *Systematic Biology* 75: 758–771.
- Tucker, S.C. & Douglas, A.W. 1994. Ontogenetic evidence and phylogenetic relationships among basal taxa of legumes. In: Ferguson, I.K., Tucker, S. (Eds.), *Advances in Legume Systematics*, Structural Botany. Royal Botanic Gardens, Kew. 6: 11–32.
- Torke, B. M. & B. A. Schaal. 2008. Molecular phylogenetics of the species-rich neotropical genus *Swartzia* (Leguminosae, Papilionoideae) and related genera of the swartzioid clade. *American Journal of Botany* 95: 215–228.
- Twyford, A. D., M. A. Veltman, P. Garret. 2022. Whole genome sequencing. In: H. de Boer, M. O. Rydmark, B. Verstraete, B. Gravendeel (Eds.), Chapter 16, *Molecular identification of plants: from sequence to species*. Advanced Books. 234–244.

- Tucker, S. C. 2003. Floral ontogeny in *Swartzia* (Leguminosae: Papilionoideae: Swartzieae): distribution and role of the ring meristem. *American Journal of Botany* 90: 1271–1292.
- Tybirk, K. 1993. Pollination, breeding system and seed abortion in some African acacias. *Botanical Journal of the Linnean Society* 112: 107–137.
- Varadi, M. et al. 2022. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research* 50(D1): D439–D444.
- Valoroso, M. R., R. Sobral, G. Saccone, M. Salvemini, M. M. R. Costa, S. Aceto. 2019. Evolutionary conservation of the orchid MYB transcription factors DIV, RAD, and DRIF. *Frontiers in Plant Science* 10: 1359.
- Wang, Z., Y. Luo, X. Li, L. Wang, S. Xu, J. Yang, L. Weng, S. Sato, S. Tabata, M. Ambrose, C. Rameau, X. Feng, X. Hu, D. Luo. 2008. Genetic control of floral zygomorphy in pea (*Pisum sativum* L.). *Proceedings of the National Academy of Sciences* 105: 10414–10419.
- Xu, S., Y. Luo, Z. Cai, X. Cao, X. Hu, J. Yang, D. Luo. 2013. Functional diversity of CYCLOIDEA-like TCP genes in the control of zygomorphic flower development in *Lotus japonicus*. *Journal of Integrative Plant Biology* 55(3): 221–231.
- Yuxiang, J. & L. Moubayidin. 2022. Floral symmetry: the geometry of plant reproduction. *Emerging Topics in Life Sciences* 6(3): 259–269.
- Wojciechowski, M. F., M. Lavin, M. J. Sanderson. 2004. A phylogeny of legumes (Leguminosae) based on analysis of the plastid *matK* gene resolves many well-supported subclades within the family. *American Journal of Botany* 91: 1846–1862.
- Zimmerman, E. et al. 2013. Floral ontogeny in Dialiinae (Caesalpinoideae: Cassieae), a study in organ instability. *South African Journal of Botany* 89: 188–209.
- Zhao, Z., J. Hu, S. Chen, Z. Luo, D. Luo, J. Wen, T. Tu, D. Zhang. 2019. Evolution of CYCLOIDEA-like genes in Fabales: insights into duplication patterns and the control of floral symmetry. *Molecular Phylogenetics and Evolution* 132: 81–89.
- Zhou, X-R., Y-Z Wang, J. F. Smith, R. Chen. 2008. Altered expression patterns of TCP and MYB genes relating to the floral developmental transition from initial zygomorphy to actinomorphy in *Bournea* (Gesneriaceae). *New Phytologist* 178: 532–543.
- Żmienko, A., A. Samelak, P. Kozłowski, M. Figlerowicz. 2014. Copy number polymorphism in plant genomes. *Theoretical and Applied Genetics* 127: 1–18.

Table 1. List of CYC sequences and corresponding GenBank accession number.

Sequence	GenBank Accession	Study	Sequence	GenBank Accession	Study
Papilionoideae					
Dialium guianensis 1	AY225811	Citerne et al. (2003)	Lupinus angustifolius 1	AY225839	Citerne et al. (2003)
Pisum 1	AY225813	Citerne et al. (2003)	Lupinus angustifolius 2	AY225840	Citerne et al. (2003)
Anthyllus 1	AY225814	Citerne et al. (2003)	Machaerium 1	AY225841	Citerne et al. (2003)
Anthyllus 1	AY225815	Citerne et al. (2003)	Machaerium 2	AY225842	Citerne et al. (2003)
Anthyllus 3	AY225816	Citerne et al. (2003)	Amicia 1	AY225843	Citerne et al. (2003)
Lotus berthelotii 1	AY225817	Citerne et al. (2003)	Amicia 2	AY225844	Citerne et al. (2003)
Lotus berthelotii 2	AY225818	Citerne et al. (2003)	Dussia 1	AY225845	Citerne et al. (2003)
Indigofera 1	AY225819	Citerne et al. (2003)	Dussia 2	AY225846	Citerne et al. (2003)
Indigofera 2	AY225820	Citerne et al. (2003)	Dussia 3	AY225847	Citerne et al. (2003)
Indigofera 3	AY225821	Citerne et al. (2003)	Swartzia 1	AY225848	Citerne et al. (2003)
Clitoria 1	AY225822	Citerne et al. (2003)	Swartzia 2	AY225849	Citerne et al. (2003)
Clitoria 2	AY225823	Citerne et al. (2003)	Swartzia 3	AY225850	Citerne et al. (2003)
Clitoria 3	AY225824	Citerne et al. (2003)	Cadia_1	scaffold_g17169	This study
Cadia 1	AY225825	Citerne et al. (2003)	Cadia_1.1	scaffold_73713	This study
Cadia 2	AY225826	Citerne et al. (2003)	Cadia_2	scaffold_304621	This study
Cadia 3	AY225827	Citerne et al. (2003)	Cadia_3	scaffold_482573	This study
Cadia 4	AY225828	Citerne et al. (2003)	Cadia_4	scaffold_282224	This study
Leptolobium 1	AY225829	Citerne et al. (2003)	Cyathostegia_1	scaffold_24987	This study
Leptolobium 2	AY225830	Citerne et al. (2003)	Cyathostegia_2	scaffold_253086	This study
Leptolobium 3	AY225831	Citerne et al. (2003)	Andira anthelmia_1	scaffold_g18071	This study
Lupinus sp. 1	AY225832	Citerne et al. (2003)	Andira anthelmia_2	scaffold_g14124	This study
Lupinus sp. 2	AY225834	Citerne et al. (2003)	Andira anthelmia_3	scaffold_g57052	This study
Lupinus sp. 3	AY225833	Citerne et al. (2003)	Andira anthelmia_4	scaffold_g46601	This study
Lupinus nanus 2	AY382156	Citerne et al. (2003)	Lupinus sp. 4	AY225835	Citerne et al. (2003)
Lupinus nanus 3	AY225838	Citerne et al. (2003)	Andira inermis 1	scaffold_659	This study

Chapter 2. An investigation of symmetry genes in legume flowers

Andira inermis 3	scaffold_g14016	This study	Andira inermis 4	scaffold_g38507	This study
Andira inermis 2	scaffold_g16521	This study	Cajanus cajan	XP_020228772	Varshney et al. (2012)
Glycine max	XP_040861879	Schmutz et al. (2010)	Phaseolus vulgaris	FE898085	Bhore et al. (unpublished)
Medicago truncatula	A0A072VQD1	Tang et al. (2014)	Genista stenopetala	BAC01123	Fukuda et al. 2003
Lotus berthelotii	AAO88020	Citerne et al. (2003)	Lotus berthelotii	AAO88019	Citerne et al. (2003)
Cercidioideae					
Bauhinia acuminata	QDF46594	Tu et al. (2019)	Cercis chinensis	QDF46605	Tu et al. (2019)
Bauhinia acuminata	QDF46595	Tu et al. (2019)	Cercis chinensis	QDF46644	Tu et al. (2019)
Bauhinia acuminata	QDF46636	Tu et al. (2019)	Cercis chinensis	QDF46606	Tu et al. (2019)
Bauhinia acuminata	QDF46635	Tu et al. (2019)	Cercis chinensis	QDF46643	Tu et al. (2019)
Caesalpinioideae					
Caesalpinia minax	QDF46599	Tu et al. (2019)	Delonix regia	QDF46607	Tu et al. (2019)
Caesalpinia minax	QDF46598	Tu et al. (2019)	Delonix regia	QDF46645	Tu et al. (2019)
Caesalpinia minax	QDF46638	Tu et al. (2019)	Cassia fistula	QDF46642	Tu et al. (2019)
Caesalpinia minax	QDF46639	Tu et al. (2019)	Calliandra haematocephala	QDF46640	Tu et al. (2019)
Acacia confusa	QDF46633	Tu et al. (2019)	Adenanthera pavonina	QDF46634	Tu et al. (2019)
Gleditsia sinensis	QDF46647	Tu et al. (2019)	Inga edulis	DF46648	Tu et al. (2019)
Zapoteca tetragona	AY225812	Citerne et al. (2003)	Archidendron clypearia	QDF46654	Tu et al. (2019)
Mimosa pudica	QDF46615	Tu et al. (2019)	Ceratonia oreothauma	AY225810	Citerne et al. (2003)
Detarioideae					
Saraca declinata	QDF46624	Tu et al. (2019)	Sindora tonkinensis	QDF46627	Tu et al. (2019)
Saraca declinata	QDF46623	Tu et al. (2019)	Sindora tonkinensis	QDF46628	Tu et al. (2019)
Saraca declinata	QDF46663	Tu et al. (2019)	Sindora tonkinensis	QDF46665	Tu et al. (2019)
Saraca declinata	QDF46662	Tu et al. (2019)	Lysidice brevicalyx	QDF46613	Tu et al. (2019)
Sindora glabra	QDF46625	Tu et al. (2019)	Lysidice brevicalyx	QDF46614	Tu et al. (2019)
Sindora glabra	QDF46626	Tu et al. (2019)	Lysidice brevicalyx	QDF46653	Tu et al. (2019)
Sindora glabra	QDF46664	Tu et al. (2019)	Lysidice brevicalyx	QDF46652	Tu et al. (2019)
Dialioideae					
Dialium guianensis	AY225811	Citerne et al. (2003)	Zenia insignis	QDF46632	Tu et al. (2019)
Zenia insignis	QDF46631	Tu et al. (2019)	Zenia insignis	QDF46669	Tu et al. (2019)

Zenia insignis	QDF46668	Tu et al. (2019)			
Brassicaceae					
Arabidopsis thaliana	NP_849730	Theologis et al. (2000)			
Surianaceae					
Suriana maritima	QDF46629	Tu et al. (2019)	Suriana maritima	QDF46666	Tu et al. (2019)
Suriana maritima	QDF46674	Tu et al. (2019)			
Polygalaceae					
Polygala paniculata	QDF46617	Tu et al. (2019)	Polygala fallax	QDF46655	Tu et al. (2019)
Polygala paniculata	QDF46658	Tu et al. (2019)	Polygala paniculata	QDF46657	Tu et al. (2019)
Polygala fallax	QDF46616	Tu et al. (2019)	Xanthophyllum hainanense	QDF46667	Tu et al. (2019)
Polygala fallax	QDF46670	Tu et al. (2019)	Xanthophyllum hainanense	QDF46630	Tu et al. (2019)
Polygala fallax	QDF46656	Tu et al. (2019)			
Quillajaceae					
Quillaja saponaria	QDF46621	Tu et al. (2019)	Quillaja brasiliensis	QDF46620	Tu et al. (2019)
Quillaja saponaria	QDF46622	Tu et al. (2019)	Quillaja brasiliensis	QDF46619	Tu et al. (2019)
Quillaja saponaria	QDF46673	Tu et al. (2019)	Quillaja brasiliensis	QDF46672	Tu et al. (2019)
Quillaja saponaria	QDF46661	Tu et al. (2019)	Quillaja brasiliensis	QDF46660	Tu et al. (2019)
Rosaceae					
Prunus persica	QDF46618	Tu et al. (2019)	Prunus persica	QDF46659	Tu et al. (2019)
Prunus persica	XP020419090		Prunus persica	QDF46671	Tu et al. (2019)

Table 2. List of DIV sequences and corresponding GenBank accession number.

Sequence	Genbank Accession	Study	Sequence	Genbank Accession	Study
Papilionoideae					
Lupinus angustifolius	XP019439408	Lowe & Eddy (2016)	Abrus precatorius	XP027335834	Hovde et al. (unpublished)
Lupinus angustifolius	OIW14207	Citerne et al. (2003)	Spatholobus suberectus	TKY47515	Qin, S. (unpublished)
Lupinus angustifolius	OIW16175	Hane et al. (2016)	Glycine soja	KAG4923320	Chu, J. (unpublished)
Lupinus angustifolius	XP019433336	Hane et al. (2016)	Cajanus cajan	XP020207747	Varshney et al. (2012)
Lupinus angustifolius	XP019433340	Hane et al. (2016)	Cajanus cajan	XP020207746	Varshney et al. (2012)

Lupinus angustifolius	XP019454444	Hane et al. (2016)	Cajanus cajan	XP020207744	Varshney et al. (2012)
Lupinus angustifolius	OIW04088	Hane et al. (2016)	Cajanus cajan	KYP33425	Varshney et al. (2012)
Lupinus albus	KAE9598753	Hufnagel et al. (2020)	Cajanus cajan	XP020207745	Varshney et al. (2012)
Lupinus albus	KAF1895030	Xu et al. (2020)	Cajanus cajan	XP020233753	Varshney et al. (2012)
Lupinus albus	KAF1863706	Xu et al. (2020)	Cajanus cajan	KYP48712	Varshney et al. (2012)
Lupinus albus	KAE9596177	Hufnagel et al. (2020)	Cajanus cajan	XP020233748	Varshney et al. (2012)
Phaseolus vulgaris	XP007163514	Schmutz et al. (unpublished)	Cajanus cajan	XP020233752	Varshney et al. (2012)
Phaseolus vulgaris	ESW35508	Schmutz et al. (unpublished)	Cajanus cajan	XP020233747	Varshney et al. (2012)
Phaseolus vulgaris	XP007163513	Schmutz et al. (unpublished)	Cajanus cajan	XP020233749	Varshney et al. (2012)
Phaseolus vulgaris	ESW35507	Schmutz et al. (unpublished)	Vigna unguiculata	XP027921959	Lonardi et al. (unpublished)
Zapoteca 1	AY225812	Citerne et al. (2003)	Vigna unguiculata	QCD98557	Xia et al. (unpublished)
Pisum 1	AY225813	Citerne et al. (2003)	Vigna radiata	XP014495343	Kang et al. (2014)
Anthyllus1	AY225814	Citerne et al. (2003)	Vigna umbellata	XP047176427	Kaul et al. (unpublished)
Anthyllus1	AY225815	Citerne et al. (2003)	Vigna umbellata	XP047176428	Kaul et al. (unpublished)
Anthyllus3	AY225816	Citerne et al. (2003)	Vigna angularis	XP017416957	Xiang et al. (unpublished)
Lotus berthelotii 1	AY225817	Citerne et al. (2003)	Vigna angularis	BAT85955	Sakai et al. (2015)
Lotus berthelotii 2	AY225818	Citerne et al. (2003)	Vigna angularis	KAG2406277	Xiang, H. (unpublished)
Indigofera 1	AY225819	Citerne et al. (2003)	Vigna angularis	KOM39134	Wan, P. (unpublished)
Indigofera 2	AY225820	Citerne et al. (2003)	Cicer arietinum	XP004490951	Varshney et al. (2013)
Indigofera 3	AY225821	Citerne et al. (2003)	Dussia 3	AY225847	Citerne et al. (2003)
Clitoria 1	AY225822	Citerne et al. (2003)	Swartzia 1	AY225848	Citerne et al. (2003)
Clitoria 2	AY225823	Citerne et al. (2003)	Swartzia 2	AY225849	Citerne et al. (2003)
Clitoria 3	AY225824	Citerne et al. (2003)	Swartzia 3	AY225850	Citerne et al. (2003)
Machaerium 1	AY225841	Citerne et al. (2003)	Cadia_1	scaffold_274555	This study
Lotus japonicus	XP057441833	Kamal et al. (unp.)	Cadia_2	scaffold_302396	This study
Lotus filicaulis	ATI97625	Boehm, M.M.A. (unp.)	Cadia_3	scaffold_459698	This study
Cadia 1	AY225825	Citerne et al. (2003)	Cadia_3.1	scaffold_14292	This study
Cadia 2	AY225826	Citerne et al. (2003)	Cadia_4	scaffold_6565	This study

Chapter 2. An investigation of symmetry genes in legume flowers

Cadia 3	AY225827	Citerne et al. (2003)	Cadia_4.1	scaffold_265731	This study
Cadia 4	AY225828	Citerne et al. (2003)	Cyathostegia_1	scaffold_22608	This study
Acosmium 1	AY225829	Citerne et al. (2003)	Cyathostegia_1.2	scaffold_22608	This study
Acosmium 2	AY225830	Citerne et al. (2003)	Andira_anthelmia_1	scaffold_56534	This study
Acosmium 3	AY225831	Citerne et al. (2003)	Andira_anthelmia_2	scaffold_14279	This study
Lupinus sp. 1	AY225832	Citerne et al. (2003)	Andira_anthelmia_3	scaffold_51007	This study
Lupinus sp. 2	AY225834	Citerne et al. (2003)	Andira_anthelmia_4	scaffold_43489	This study
Lupinus sp. 3	AY225833	Citerne et al. (2003)	Andira_inermis_1	scaffold_27947	This study
Lupinus sp. 4	AY225835	Citerne et al. (2003)	Andira_inermis_2	scaffold_27947	This study
Amicia 1	AY225843	Citerne et al. (2003)	Andira_inermis_3	scaffold_14	This study
Amicia 2	AY225844	Citerne et al. (2003)	Andira_inermis_4	scaffold_58629	This study
Dussia 1	AY225845	Citerne et al. (2003)	Andira_inermis_5	scaffold_839	This study
Dussia 2	AY225846	Citerne et al. (2003)	Andira_inermis_6	scaffold_41763	This study
Trifolium pratense	XP045786186	Bickhart et al. (DS)	Andira_inermis_7	scaffold_417	This study
Trifolium pratense	PNX95682	Istvanek et al. (2014)	Melilotus albus	QSD99686	Zhang et al.
Trifolium pratense	XP045829296	Bickhart et al. (DS)	Medicago truncatula	AFK48294	Krishnakumar et al. (DS 2012)
Trifolium pratense	CAJ2653318	Rodrigues & Vega (DS 2023)	Medicago truncatula	AES99526	Tang et al. (2014)
Trifolium pratense	WJX64117	Santangelo et al. (DS 2023)	Medicago truncatula	XP003616568	Pecrix et al. (2018)
Trifolium pratense	KAK2377129	Kuo & Olsen (DS 2023)	Medicago truncatula	RHN57078	Gamas & Carrere (DS 2018)
Trifolium repens	KAK2430921	Kuo & Olsen (DS 2023)	Vicia faba	CAI8586261	Kreplak (DS 2023)
Trifolium repens	WJX58837	Santangelo et al. (DS 2023)	Vicia villosa	XP058759979	Bickhart et al. (DS 2023)
Trifolium repens	WJX58836	Santangelo et al. (DS 2023)	Vicia villosa	XP058744855	Bickhart et al. (DS 2023)
Pisum sativum	KAI5434214	Gao, S. (2022)	Arachis ipaensis	XP016166941	Bertioli et al. (2016)
Pisum sativum	XP050904322	Gao, S. (2022)	Arachis ipaensis	XP016166940	Bertioli et al. (2016)
Pisum sativum	KAI5434213	Gao, S. (2022)	Arachis hypogaea	XP025666503	Bertioli et al. (2019)
Abrus precatorius	XP027331255	Hovde et al (DS 2018)	Arachis hypogaea	AHB59615	Chen et al (2014)
Abrus precatorius	XP027331254	Hovde et al (DS 2018)	Arachis hypogaea	RYR46245	

Chapter 2. An investigation of symmetry genes in legume flowers

<i>Abrus precatorius</i>	XP027331256	Hovde et al (DS 2018)	<i>Arachis hypogaea</i>	QHO27374	Zhuang et al. (unp.)
<i>Spatholobus suberectus</i>	TKY56831	Qin, S. (DS 2019)	<i>Arachis hypogaea</i>	XP025609237	Bertioli et al. (2019)
<i>Glycine max</i>	KAH1093298	Silva-Junior et al. (DS 2020)	<i>Vigna angularis</i>	KOM46939	Wan, P. (unp.)
<i>Glycine max</i>	KAH1211898	Kim et al. (2021)	<i>Vigna unguiculata</i>	QCE01304	Xia et al. (unp.)
<i>Glycine max</i>	NP001393425	Lucas et a. (2009)	<i>Vigna unguiculata</i>	XP027939354	Lonardi et al. (2017)
<i>Glycine max</i>	KRH14994	Schmutz et al. (2010)	<i>Vigna unguiculata</i>	XP027939351	Lonardi et al. (2017)
<i>Glycine max</i>	ABH02835	Liao et al., (unp.)	<i>Vigna unguiculata</i>	XP027939353	Lonardi et al. (2017)
<i>Glycine max</i>	APA28810	Kim & Woo (unp.)	<i>Vigna umbellata</i>	XP047174313	Kaul et al. (2019)
<i>Glycine max</i>	KAG4962252	Chu, J. (unp.)	<i>Vigna angularis</i>	XP017428068	Xiang, H. (unp.)
<i>Glycine max</i>	NP001235748	Shoemaker et al. (unp.)	<i>Vigna angularis</i>	BAT81144	Sakai et al. (2015)
<i>Glycine max</i>	KAG4964726	Chu, J. (unp.)	<i>Vigna angularis</i>	XP017428069	Xiang, H. (unp.)
<i>Arachis hypogaea</i>	QHO27375	Zhuang et al. (unp.)	<i>Vigna umbellata</i>	XP047174314	Xiang, H. (unp.)
<i>Sphenostylis stenocarpa</i>	CAJ1970981	Domelevo (DS 2023)	<i>Vigna angularis</i>	KAG2391237	Xiang, H. (unp.)
<i>Phaseolus vulgaris</i>	XP007141798	Schmutz et al (unp.)	<i>Phaseolus vulgaris</i>	Phvul.006G06 5700	Kavas et al. (2022)
<i>Phaseolus vulgaris</i>	ESW13792	Schmutz et al (unp.)	<i>Cicer arietinum</i>	XP004490951	Varshney et al. (2013)
Caesalpinioidae					
<i>Acacia pycnantha</i>	KAI9080697	Mclay et al. (unpublished)	<i>Prosopis alba</i>	XP028779755	Sudalaimuthua sari et al. (2022)
<i>Prosopis cineraria</i>	XP054798674	Sudalaimuthua sari et al. (2022)	<i>Prosopis alba</i>	XP028792630	Sudalaimuthua sari et al. (2022)
<i>Prosopis cineraria</i>	XP054795786	Sudalaimuthua sari et al. (2022)			
Detarioideae					
<i>Bauhinia variegata</i>	KAI4306793	Zhong et al (2022)	<i>Bauhinia variegata</i>	KAI4299451	Zhong et al (2022)
<i>Bauhinia variegata</i>	KAI4299452	Zhong et al (2022)			
Plantaginaceae					
<i>Antirrhinum majus</i>	AAL78742	Galego & Almeida (2002)	<i>Antirrhinum majus</i>	AAL78741	Galego & Almeida (2002)
Solanaceae					
<i>Solanum lycopersicum</i>	XP004241178	Hosmani et al (unp.)	<i>Solanum lycopersicum</i>	XP004249083	Hosmani et al (unp.)
Caprifoliaceae					
<i>Heptacodium miconioides</i>	ACR09740	Howarth and Donoghue (2009)	<i>Heptacodium miconioides</i>	ACR09741	Howarth and Donoghue (2009)

Heptacodium miconioides	ACR09746	Howarth and Donoghue (2009)			
Vitaceae					
Vitis vinifera	XP002280403	Shi et al. (2023)	Vitis vinifera	XP002273261	Shi et al. (2023)
Vitis vinifera	XP010654186	Shi et al. (2023)			
Ranunculaceae					
Aquileja coerulea	Aqcoe1G083200	Filiault et al. (2018)	Aquileja coerulea	Aqcoe2G103600	Filiault et al. (2018)
Orchidaceae					
Orchis italica	ARU86433	Valoroso et al. (2017)	Orchis italica	ARU86426	Valoroso et al. (2017)
Orchis italica	ARU86429	Valoroso et al. (2017)			
Poaceae					
Oryza sativa	XP015621817	Kawahara et al. (2013)	Oryza sativa	XP015621447	Kawahara et al. (2013)
Oryza sativa	XP015637475	Kawahara et al. (2013)			
Amborellaceae					
Amborella trichopoda	XP006837395	Victoria et al. (2013)	Amborella trichopoda	XP006838124	Victoria et al. (2013)

Table 3. List of RAD sequences and corresponding GenBank accession number.

Sequence	Genbank Accession	Study	Sequence	Genbank Accession	Study
Papilionoideae			Vigna umbellata	XP047163108	
Lotus japonicus	XP057457612	Kamal et al. (2020)	Vigna angularis	BAT92774	
Melilotus albus	QSD99773	Chen et al. (2021)	Vigna angularis	BAT85101	
Trifolium repens	KAK2385988	Kuo & Olsen (2023)	Vigna angularis	XP017418109	
Trifolium repens	KAK2439976	Kuo & Olsen (2023)	Vigna angularis	KAG2405517	
Trifolium repens	KAK2439977	Kuo & Olsen (2023)	Spatholobus suberectus	TKY62729	
Trifolium pratense	XP045796180	Bickhart et al. (unp.)	Spatholobus suberectus	TKY49751	
Trifolium pratense	CAI9231855	Cubillos et al. (DS)	Abrus precatorius	XP027352977	
Medicago truncatula	AES68669	Young et al. (2011)	Abrus precatorius	XP027343819	
Medicago truncatula	XP003598418	Pecrix et al. (2018)	Sphenostylis stenocarpa	CAJ1958783	
Medicago truncatula	RHN65600	Pecrix et al. (2018)	Phaseolus vulgaris	Phvul.009G058100	
Arachis hypogaea	RYR05382	Chen et al. (2019)	Phaseolus vulgaris	Phvul.002G031300	
Cajanus cajan	XP020238662	Varshney et al. (2011)	Phaseolus vulgaris	Phvul.001G233900	
Cajanus cajan	XP020225606		Phaseolus vulgaris	Phvul.008G192200	

Chapter 2. An investigation of symmetry genes in legume flowers

<i>Mucuna pruriens</i>	RDY11942		<i>Phaseolus vulgaris</i>	Phvul.005G018500	
<i>Mucuna pruriens</i>	RDX97666		<i>Phaseolus vulgaris</i>	Phvul.001G120400	
<i>Vigna radiata</i>	XP014513246		<i>Phaseolus vulgaris</i>	Phvul.002G031300	
<i>Vigna radiata</i>	XP014491210		<i>Andira inermis</i>	NODE5880	
<i>Vigna unguiculata</i>	XP027921590		<i>Andira inermis</i>	NODE12647	
<i>Vigna unguiculata</i>	XP027938307		<i>Andira inermis</i>	NODE4028	
<i>Vigna unguiculata</i>	QCE05502		<i>Andira inermis</i>	g25632	
<i>Vigna unguiculata</i>	XP027903531		<i>Andira inermis</i>	g50482	
<i>Vigna unguiculata</i>	QCD97071		<i>Andira inermis</i>	NODE5928	This study
<i>Vigna umbellata</i>	XP047151885		<i>Andira inermis</i>	g3283	This study
<i>Sesbania bispinosa</i>	KAJ1436036		<i>Andira inermis</i>	g18159	This study
<i>Sesbania bispinosa</i>	KAJ1404264		<i>Andira anthelmia</i>	g52474	This study
<i>Cadia purpurea</i>	g1348	This study	<i>Andira anthelmia</i>	g20697	This study
<i>Cadia purpurea</i>	131773	This study	<i>Andira anthelmia</i>	g21142	This study
<i>Cadia purpurea</i>	56113	This study	<i>Andira anthelmia</i>	g42004	This study
<i>Cadia purpurea</i>	g16437	This study	<i>Andira anthelmia</i>	g7185	This study
<i>Cadia purpurea</i>	g17385	This study	<i>Lupinus albus</i>	KAE9588663	Hufnagel et al. (2020)
<i>Cadia purpurea</i>	77642	This study	<i>Lupinus albus</i>	KAF1883284	Xu et al. (2020)
<i>Cadia purpurea</i>	480768	This study	<i>Lupinus albus</i>	KAE9599748	Hufnagel et al. (2020)
<i>Cadia purpurea</i>	g8202	This study	<i>Lupinus albus</i>	KAF1886848	Xu et al. (2020)
<i>Pisum sativum</i>	KAI5427383	Yang et al. (2022)	<i>Lupinus angustifolius</i>	XP019436070	Hane et al. (2016)
<i>Pisum sativum</i>	XP050911347	Yang et al. (2022)	<i>Lupinus angustifolius</i>	OIW15492	Hane et al. (2016)
<i>Vicia villosa</i>	XP058740160	Fuller et al. (2023)	<i>Lupinus angustifolius</i>	OIW20456	Hane et al. (2016)
<i>Vicia villosa</i>	XP058749168	Fuller et al. (2023)	<i>Lupinus angustifolius</i>	XP019431091	Hane et al. (2016)
<i>Glycine soja</i>	KAG5043232	Chu et al. (2021)	<i>Lupinus angustifolius</i>	XP019431092	Hane et al. (2016)
<i>Glycine soja</i>	KHN30338	Guo et al. (2021)	<i>Glycine soja</i>	XP028225172	Xi et al. (2019)
			<i>Glycine max</i>	KAH1069690	Silva-Junior et al. (unp.)
Caesalpinioideae					
<i>Senna tora</i>	KAF7803988		<i>Prosopis alba</i>	XP028761397	Kong & Paterson (DS)
<i>Senna tora</i>	KAF7819144		<i>Prosopis alba</i>	XP028779041	Kong & Paterson (DS)
<i>Bauhinia variegata</i>	KAI4300910		<i>Acacia cineraria</i>	P054800104	
<i>Prosopis alba</i>	XP028798148	Kong & Paterson (DS)	<i>Acacia pycnantha</i>	KAI9082088	Mclay et al. (2022)
<i>Prosopis alba</i>	XP028766688	Kong & Paterson (DS)	<i>Acacia pycnantha</i>	KAI9121632	Mclay et al. (2022)

Chapter 2. An investigation of symmetry genes in legume flowers

Prosopis alba	XP054821239	Sudalaimuthua sari et al. (2022)			
Plantaginaceae					
Antirrhinum majus	AAX48042	Corley et al. 2005			
Solanaceae					
Solanum lycopersicum	Solyc10g080920		Solanum lycopersicum	Solyc12g089190	
Solanum lycopersicum	Solyc01g109690		Solanum lycopersicum	Solyc10g052470	
Solanum lycopersicum	Solyc01g109670				
Vitaceae					
Vitis vinifera	XP002282461	Shi et al. (2023)	Vitis vinifera	XP003631666	Shi et al. (2023)
Vitis vinifera	XP002282280	Shi et al. (2023)	Vitis vinifera	XP003631660	Shi et al. (2023)
Poaceae					
Oryza sativa	LOC0s05g49240		Oryza sativa	LOC0s01g47370	
Oryza sativa	LOC0s07g26150				
Caprifoliaceae					
Lonicera bella	AFM92049	Boyden et al (2012)	Lonicera bella	AFM92040	Boyden et al (2012)
Lonicera bella	AFM92028	Boyden et al (2012)	Lonicera bella	AFM92016	Boyden et al (2012)
Rutaceae					
Citrus unshiu	GAY57324	Shimizu et al. (2017)			
Lentibulariaceae					
Genlisea aurea	EPS74148	Leushkin et al. (2013)			

Table 4. Estimation of genome size, heterozygosity, genome repeat content, genome unique length from k-mer count data based on the Illumina reads for the species *Cyathostegia mathewsii*, *Cadia purpurea*, *Andira inermis* and *Andira anthelmia*. Estimates calculated from Genoscope.

Illumina-seq	<i>C. mathewsii</i> <i>k</i> = 81	<i>Ca. purpurea</i> <i>k</i> = 81	<i>A. inermis</i> <i>k</i> = 31	<i>A. anthelmia</i> <i>k</i> = 21
Genome size (Mbp)	442	567	431	481
Heterozygosity (%)	0.31	0.28	1.2	1.7
Repeat length (Mbp)	76	77	83	148
Unique length (Mbp)	366	490	347	333
Model fit (%)	99	98	95	96
Read error rate (%)	0.1	0.1	0.4	0.1

Chapter 2. An investigation of symmetry genes in legume flowers

Table 5. Genome assembly metrics of *Cyathostegia mathewsii*, *Cadia purpurea*, *Andira inermis* and *Andira anthelmia*. Buscos were assessed using the Metaeuk gene predictor from Galaxy.

Assembler	Platanus allee									
Species	Total no. scaffolds	Largest scaffold (bp)	N50 (bp)	N base count (bp)	BUSCO completeness (%)	No. complete BUSCOs	No. complete and single-copy BUSCOs	No. complete and duplicated BUSCOs	No. fragmented BUSCOs (%)	No. missing BUSCOs (%)
<i>Cyathostegia mathewsii</i>	430207391	514875	30058	48130	88.9	4770	4426	344	111	485
<i>Cadia purpurea</i>	585136375	305852	17114	121517	87.8	4711	4479	232	165	490
<i>Andira inermis</i>										
<i>Andira anthelmia</i>										

Assembler	Spades assembler									
Species	Total no. scaffolds	Largest scaffold (bp)	N50 (bp)	N base count (bp)	BUSCO completeness (%)	No. complete BUSCOs	No. complete and single-copy BUSCOs	No. complete and duplicated BUSCOs	No. fragmented BUSCOs (%)	No. missing BUSCOs (%)
<i>Cyathostegia mathewsii</i>	424041376	620768	55740	1219451	93	5002	4570	432	61	303
<i>Cadia purpurea</i>	573609435	562682	42732	795802	89.5	4803	4495	308	180	383
<i>Andira inermis</i>	714226929	99033	1651	3219866	77.3	4150	3677	473	406	810
<i>Andira anthelmia</i>	664812375	402916	4452	2579448	81.3	4362	3793	569	351	653

Table 6. Copy numbers and sequence length of CYC in *Cyathostegia mathewsii*, *Cadia purpurea*, *Andira inermis* and *Andira anthelmia*.

Putative ortholog sequences	Scaffold ID	Seq length (aa)	Type of sequence
<i>Cyathostegia mathewsii</i> 1	scaffold_24987	391	transcriptome
<i>Cyathostegia mathewsii</i> 2	scaffold_253086	397	transcriptome
<i>Cadia purpurea</i> 1	scaffold_g17169	350	genome
<i>Cadia purpurea</i> 1.1	scaffold_73713	350	transcriptome
<i>Cadia purpurea</i> 2	scaffold_304621	403	transcriptome
<i>Cadia purpurea</i> 3	scaffold_482573	367	transcriptome

<i>Cadia purpurea</i> 4	scaffold_282224	369	transcriptome
<i>Andira inermis</i> 1	scaffold_659	299	transcriptome
<i>Andira inermis</i> 2	scaffold_g16521	385	genome
<i>Andira inermis</i> 3	scaffold_g14016	369	genome
<i>Andira inermis</i> 4	scaffold_g38507	434	genome
<i>Andira anthelmia</i> 1	scaffold_g18071	428	genome
<i>Andira anthelmia</i> 2	scaffold_g14124	420	genome
<i>Andira anthelmia</i> 3	scaffold_g57052	332	genome
<i>Andira anthelmia</i> 4	scaffold_g46601	386	genome

Table 7. Copy numbers and sequence length of RAD in *Cyathostegia mathewsii*, *Cadia purpurea*, *Andira inermis* and *Andira anthelmia*.

Putative ortholog sequences	Scaffold ID	Seq length (aa)	Type of sequence
<i>Cyathostegia mathewsii</i> 1	scaffold_22608	83	transcriptome
<i>Cyathostegia mathewsii</i> 2	scaffold_171427	86	transcriptome
<i>Cyathostegia mathewsii</i> 3	scaffold_342610	77	transcriptome
<i>Cyathostegia mathewsii</i> 4	scaffold_283498	81	transcriptome
<i>Cadia purpurea</i> 1	scaffold_g8202	72	genome
<i>Cadia purpurea</i> 2	scaffold_480768	81	transcriptome
<i>Cadia purpurea</i> 3	scaffold_16437	81	transcriptome
<i>Cadia purpurea</i> 3.1	scaffold_56113	81	transcriptome
<i>Cadia purpurea</i> 4	scaffold_g17385	93	genome
<i>Cadia purpurea</i> 4.1	scaffold_77642	96	transcriptome
<i>Cadia purpurea</i> 5	scaffold_g1348	85	genome
<i>Cadia purpurea</i> 5.1	scaffold_g31773	89	genome
<i>Andira inermis</i> 1	scaffold_5928	87	transcriptome
<i>Andira inermis</i> 1.1	scaffold_g50482	80	genome
<i>Andira inermis</i> 2	scaffold_5880	97	transcriptome
<i>Andira inermis</i> 3	scaffold_4028	100	transcriptome
<i>Andira inermis</i> 4	scaffold_12647	98	transcriptome
<i>Andira inermis</i> 5	scaffold_g25632	78	genome
<i>Andira inermis</i> 6	scaffold_g3283	80	genome
<i>Andira inermis</i> 7	scaffold_g18159	80	genome
<i>Andira anthelmia</i> 1	scaffold_g20697	90	genome
<i>Andira anthelmia</i> 2	scaffold_g21142	85	genome
<i>Andira anthelmia</i> 3	scaffold_g52474	85	genome
<i>Andira anthelmia</i> 4	scaffold_g7185	85	genome
<i>Andira anthelmia</i> 5	scaffold_g42004	78	genome

Table 8. Copy numbers and sequence length of DIV in *Cyathostegia mathewsii*, *Cadia purpurea*, *Andira inermis* and *Andira anthelmia*.

Putative ortholog sequences	Scaffold ID	Seq length (aa)	Type of sequence
<i>Cadia purpurea</i> 1	scaffold_274555	330	transcriptome
<i>Cadia purpurea</i> 2	scaffold_302396	305	transcriptome
<i>Cadia purpurea</i> 3	scaffold_459698	305	transcriptome
<i>Cadia purpurea</i> 3.1	scaffold_14292	303	genome
<i>Cadia purpurea</i> 4	scaffold_6565	319	genome
<i>Cadia purpurea</i> 4.1	scaffold_265731	322	transcriptome
<i>Cyathostegia mathewsii</i> 1	scaffold_22608	290	transcriptome
<i>Cyathostegia mathewsii</i> 1.2	scaffold_22608	292	genome
<i>Andira anthelmia</i> 1	scaffold_56534	144	genome
<i>Andira anthelmia</i> 2	scaffold_14279	307	genome
<i>Andira anthelmia</i> 3	scaffold_51007	302	genome
<i>Andira anthelmia</i> 4	scaffold_43489	307	genome
<i>Andira inermis</i> 1	scaffold_27947	146	genome
<i>Andira inermis</i> 2	scaffold_11557	143	genome
<i>Andira inermis</i> 3	scaffold_14	307	transcriptome
<i>Andira inermis</i> 4	scaffold_58629	307	genome
<i>Andira inermis</i> 5	scaffold_839	310	transcriptome
<i>Andira inermis</i> 6	scaffold_41763	305	genome
<i>Andira inermis</i> 7	scaffold_417	305	transcriptome

Table 9. Copy numbers and sequence length of DRIF in *Cyathostegia mathewsii* and *Cadia purpurea*. No DRIF orthologs were found for *Andira inermis* and *Andira anthelmia*.

Putative ortholog sequences	Scaffold ID	Seq length (aa)	Type of sequence
<i>Cadia purpurea</i> 1	scaffold_65090	236	transcriptome
<i>Cyathostegia mathewsii</i> 1	scaffold_266020	237	transcriptome

Appendix I.

De novo assembly using Platanus-allee 2.2.2.

Contig assembly usage:

```
platanus_allee assemble -f PE_1.fq PE_2.fq 2>assemble.log
```

Input format:

Uncompressed and compressed (gzip or bzip2) files are accepted for -f option.

Options available:

-o STR	: prefix of output files (default out, length <= 200)
-f FILE1 [FILE2 ...]	: reads file (fasta or fastq, number <= 100)
-t INT	: number of threads (<= 100, default 1)
-m INT	: memory limit for making kmer distribution (GB, >=1, default 16)
-tmp DIR	: directory for temporary files (default .)
-k INT	: initial k-mer size (default 32)
-K FLOAT	: maximum-k-mer factor (maximum-k = FLOAT*read-length, default 0.5)
-s INT	: step size of k-mer extension (>= 1, default 20)
-n INT	: initial k-mer coverage cutoff (default 0, 0 means auto)
-c INT	: minimum k-mer coverage (default 2)
-a FLOAT	: k-mer extension safety level (default 10.0)
-u FLOAT	: maximum difference for bubble crush (identity, default 0)
-d FLOAT	: maximum difference for branch cutting (coverage ratio, default 0.5)
-e FLOAT	: k-mer coverage depth (k = initial k-mer size specified by -k) of homozygous region (default auto)

Outputs:

PREFIX_contig.fa

PREFIX.kmerFrq.tsv

PREFIX is specified by -o.

III – Phasing: in this step we used the contigs generated in the previous step, along with trimmed reads (pair end) to scaffold the initial set of contigs. Again, most of the default options were used.

Haplotype phasing usage:

```
platanus_allee phase \  
-c out_contig.fa \  
-IP1 PE_1.fq PE_2.fq \  
-OP2 MP_1.fq MP_2.fq \  
-p PacBio.fq \  
-x 10X_barcoded.fq \  
2>phase.log
```

Input format:

Uncompressed and compressed (gzip or bzip2) files are accepted for `-c`, `-ip`, `-IP`, `-op`, `-OP`, `-p`, `-x` and `-X` option.

Options available:

-o STR	: prefix of output file and directory (do not use "/", default out, length <= 200)
-c FILE1 [FILE2 ...]	: contig (or scaffold) file (fasta format)
-ip{INT} PAIR1 [PAIR2 ...]	: lib_id inward_pair_file (interleaved file, fasta or fastq)
-IP{INT} FWD1 REV1 [FWD2 REV2 ...]	: lib_id inward_pair_files (separate forward and reverse files, fasta or fastq)
-op{INT} PAIR1 [PAIR2 ...]	: lib_id outward_pair_file (interleaved, fasta or fastq)
-OP{INT} FWD1 REV1 [FWD2 REV2 ...]	: lib_id outward_pair_files (separate forward and reverse files, fasta or fastq)
-p PAIR1 [PAIR2 ...]	: long-read file (PacBio, Nanopore) (fasta or fastq)
-x PAIR1 [PAIR2 ...]	: linked-reads files (paired-ends, 10x Genomics) (interleaved, fasta or fastq)
-X FWD1 REV1 [FWD2 REV2 ...]	: linked-reads files (paired-ends, 10x Genomics) (separate forward and reverse files, fasta or fastq)
-t INT	: number of threads (default 1)
-tmp DIR	: directory for temporary files (default .)
-i INT	: number of iterations (default 2)
-l INT	: minimum number of links to scaffold (default 3)
-k INT	: minimum number of links to phase variants (default 1)
-s INT1 [INT2 ...]	: mapping seed length for short reads (default 32 64 96)
-mapper FILE	: path of mapper executable file (default minimap2, only effective with -p option)
-minimap2_sensitive	: sensitive mode for minimap2 (default, off; only effective with -p option)

Outputs:

PREFIX_allPhasedScaffold.fa (including sequences below)
PREFIX_primaryBubble.fa
PREFIX_secondaryBubble.fa
PREFIX_nonBubbleHetero.fa
PREFIX_nonBubbleOther.fa
PREFIX_consensusInput.fa (for "consensus" command (-c))

IV – Consensus: The final step is to close the gaps (reduce the N content in the genome introduced during scaffolding). Here we needed the scaffolds generated in the previous step.

Consensus scaffold construction usage:

```
-c out_consensusInput.fa \  
-IP1 PE_1.fq PE_2.fq \  
-OP2 MP_1.fq MP_2.fq \  
-p PacBio.fq -x 10X_barcode.fq \  
-x 10X_barcode.fq \  
2>consensus.log
```

Input format:

Uncompressed and compressed (gzip or bzip2) files are accepted for -c, -ip, -IP, -op, -OP, -p, -x and -X option.

Options available:

```
-o STR : prefix of output file (default out, length <= 200)  
-c FILE1 [FILE2 ...] : input_scaffolds (fasta format)  
-ip{INT} PAIR1 [PAIR2 ...] : lib_id inward_pair_file (interleaved file, fasta or fastq)  
-IP{INT} FWD1 REV1 [FWD2 REV2 ...] : lib_id inward_pair_files (separate forward and reverse files, fasta or  
fastq)  
-op{INT} PAIR1 [PAIR2 ...] : lib_id outward_pair_file (interleaved, fasta or fastq)  
-OP{INT} FWD1 REV1 [FWD2 REV2 ...] : lib_id outward_pair_files (separate forward and reverse files, fasta or fastq)  
-p PAIR1 [PAIR2 ...] : long-read file (PacBio, Nanopore) (fasta or fastq)  
-x PAIR1 [PAIR2 ...] : linked-reads files (paired-ends, 10x Genomics) (interleaved, fasta or fastq)  
-X FWD1 REV1 [FWD2 REV2 ...] : linked-reads files (paired-ends, 10x Genomics) (separate forward and reverse  
files, fasta, or fastq)  
-t INT : number of threads (default 1)
```

-tmp DIR	: directory for temporary files (default .)
-e FLOAT	: coverage depth of homozygous region (default auto)
-L INT	: maximum fragment length of tag (10x Genomics) (default 200000)
-s INT1 [INT2 ...]	: mapping seed length for short reads (default 32 64 96)
-I INT	: minimum number of links to scaffold (default 3)
-mapper FILE	: path of mapper executable file (default, minimap, only effective with -p options)
minimap2_sensitive	: sensitive mode for minimap2 (default, off; only effective with -p option)
-no_partial	: not close gaps partially, i.e. only close ones completely (default, off)

Outputs:

PREFIX_consensusScaffold.fa (final)

Chapter 3

Differential expression of symmetry genes in the early
branching Papilionoideae

Introduction

3.1 RNA sequencing and its applications

RNA sequencing (RNA-seq) analysis represents a significant advancement in our ability to study gene expression at a molecular level. Prior to the development of RNA-seq, *in situ* hybridization was a popular method to visualize the location of specific gene activity within cells and tissues. The technique involves the use of labeled RNA or DNA probes that could hybridize with complementary sequences in the target RNA as a means of physically localizing the expression of target genes via fluorescence microscopy. It is still used alongside RNA-seq. RNA-seq can be done on specific tissue types (e.g., developing flower primordia) and can therefore give organ-specific information on gene transcription. The degree of spatial information on gene expression which can be obtained is dependent on the ability to isolate specific tissues and stages, with the most extreme level being single cell RNA-Seq, requiring laser dissection of fixed tissues. RNASeq has an advantage over in-situ analysis of providing global transcriptional information. This allows us to discover novel genes with interesting expression patterns and provides us with the sequences of all the genes expressed in the sampled tissues. Depending on expression level and coverage, this can give more reliable sequence data than a genome skim. Whole Genome Sequencing (Chapter Two) can provide the set of genes, but partial genomes may miss many sequences, and generation of full, chromosomally contiguous genomes is still expensive and informationally challenging.

As discussed in Chapter One, in-situ hybridisation has been employed in the study of flower development in legumes, including in *Lotus*, *Pisum*, and *Lupinus* (Feng et al., 2006; Citerne et al., 2006; Wang et al., 2008). For example, the two CYC genes *LegCYC1A* and *LegCYC1B* were detected in *Lupinus nanus*, with expression visible on the adaxial side of the floral meristem before organogenesis (Citerne et al., 2006). These genes, although overlapping in function, were both exclusively detected during the development of the standard petal. These findings suggested a correlation between expression of *LegCYC1A* and *LegCYC1B* and the development of zygomorphy in the flower of *L. nanus* through their involvement in adaxial identity in legumes. In *Antirrhinum*, CYC activates the MYB gene *RADIALIS* (*RAD*) to promote floral dorsal identity. CYC and RAD activity antagonize the expression of a third gene responsible for the determination of ventral identity, *DIVARICATA* (*DIV*), limiting DIV activity to the ventral domain. The antagonistic interplay of the genes results in the zygomorphic wild morphology of *Antirrhinum* (Corley et al., 2005). These interactions were

discovered by employing *in situ* hybridization, mutant analysis, and transgenics which showed precisely the regions where each gene was expressed and how their expression was affected by the activity of each gene.

Although *in situ* hybridization provided valuable insights into the spatial-temporal distribution of gene expression, the methodology was challenging to accurately detect expression levels and was labour-intensive, requiring the use of autoradiography or fluorescence microscopy, which depend upon technically difficult thin-sectioning and subsequent tissue labelling. For these reasons, it has not been employed in the study of non-model legumes such as papilionoids with unusual floral morphologies. Now, quantifying the expression of genes has become simpler with the employment of RNA-seq and the generation of full transcriptomes. This sequencing technique can offer considerable insights into many biological processes in plants by providing detailed information of the level of transcription of specific genes in specific organs. RNA-seq uses Next Generation Sequencing (NGS) technologies to sequence and quantify transcripts (Wang et al., 2009). This type of sequencing is especially suitable for non-model organisms, since a reference genome is not required for downstream analyses, although it is useful for several analyses such as genome assembly and annotation, mapping and identification of syntenic genes (Wang et al., 2009; Korpelainen et al., 2014).

There is a great need to understand the evolution and adaptations of the legume species beyond a few model species. The number of good and draft genomes for legumes is continually increasing, but to link the data therein to cellular, anatomical and morphological changes it is needed to couple genome sequence with transcriptome analysis to understand the relationship between gene function and regulation of gene expression (Jackson et al., 2011; Garg & Jain, 2013). Assembling RNA-seq reads into a transcriptome providing transcribed mRNA sequence information can be used for identifying differentially expressed genes or investigating isoforms and alternative splicing events (Korpelainen et al., 2014). Despite having identical genetic content, different cells in a multicellular organism show very different patterns of gene expression that underlie a wide range of physical, biochemical, and developmental differences among them. Thus, through comparative transcriptome analyses amongst different tissues and/or cell types and developmental stages, the way in which transcriptional change contributes to various cellular processes may be understood. For example, RNA-seq derived transcriptomes from a developmental series have been used for comparative transcriptomics to

identify candidate regulators of phyllotaxy in *Antirrhinum* (Wang et al., 2017a) and shoot apical meristem and floral development in legumes (Singh & Jain, 2014).

Flower development is one of the major developmental processes in angiosperms. However, whilst there have been some transcriptomic studies in model legumes and crops such as chick-pea (e.g., Singh et al., 2013), little is known about the molecular mechanisms underlying flower development in non-model legumes. Employing RNA-seq in various stages of flower development and vegetative tissues can shed light on the differential expression of genes involved in these structures, revealing biological processes and molecular functions during flower development (Singh & Jain, 2014).

3.2 Comparative Transcriptome Analyses

In the context of this study of legume flower morphology, the goal of comparative transcriptome analysis is to identify conserved co-expressed genes in two or more species (Mohavedi et al., 2011; Mutwil et al., 2011). Sequence homology and syntenic relationships are traditionally used to define gene orthology between species (Emms et al., 2015; Lyons et al., 2008). Gene expression patterns are also important to validate homology and comparative transcriptome analyses combine a comparison of expression patterns between homologous genes in different species (Lee et al., 2019). Homologous genes have been reported to be expressed at different developmental stages, in different tissue types, or under different stress conditions in different species (Mohavedi et al., 2011; Yao et al., 2007). This documented divergence of expression patterns provides crucial evidence for the existence of functional divergence of homologous genes across species (Roulin et al., 2013). Therefore, comparative transcriptomic analysis is an important tool for distinguishing those genes that have retained functional conservation from those that have undergone functional divergence.

3.3 Transcriptomics studies in legumes

The molecular basis of flower development has been extensively studied in model plants such as *Arabidopsis thaliana* and rice (*Oryza sativa*). Similarity search has revealed that most of the known genes involved in flower development are conserved in soybean (*Glycine max*), chickpea (*Cicer arietinum*), *Medicago truncatula*, and *Lotus japonicus* (Dong et al., 2005; Hecht et al., 2005; Jung et al., 2012; Singh et al., 2013). However, the morphological analyses of legume flowers have shown that flower development is somewhat dissimilar in

legumes to in other angiosperms though genetic analysis shows it is controlled by the same ABC model (Tucker, 1987, 2003; Hidalgo et al., 2022).

Studies involving gene expression through transcriptome analyses have been carried out in crop and model legumes such as *Glycine max* (Libaut et al., 2010), *Medicago truncatula* (Gao et al., 2022), *Lotus corniculatus* (Feng et al., 2006), and *Cicer arietinum* (Kudapa et al., 2018) to uncover the overall and specific transcriptional activity of genes across various tissues and/or organs and developmental stages. Several candidate genes putatively involved in important traits, such as nodule, flower, and seed development, have been identified.

Singh et al. (2013) presented a global view of transcriptome dynamics in chickpea based on the RNA-seq transcriptome of three vegetative tissues and eight stages of flower development. This study identified several genes expressed preferentially and specifically during different stages of flower development. As discussed before, there is evidence to show the role of differential expression of lineage-specific TCP and MYB genes in flower development, which might be responsible, at least partly, for distinct floral symmetry in legumes (Feng et al., 2006; Citerne et al., 2006; Wang et al., 2008; Garg & Jain, 2011; Singh et al., 2013). This approach is of special interest when comparing closely related species that bear distinct flower types.

In this chapter I present a comparative transcriptome approach where I investigate putative symmetry genes and related proteins (described in Chapter Two), and a general RNASeq analysis to study the expression of genes involved in the development of flowers across morphologically distinct species. This study involves a series of developmental stages from flowers and vegetative tissues of the morphologically distinct legume species *Cadia purpurea* (radially symmetric), *Cyathostegia mathewsii* (zygomorphic with one petal), and *Andira inermis* (typical zygomorphic papilionate). The main goal of this chapter is to analyse differentially expressed genes across a series of contrast groups, including vegetative and floral tissues, throughout the development of flowers with distinct morphologies.

Full transcriptomes were compared from: A - *Cadia* (adaxial vs abaxial region of the flower); B - *Cadia* (vegetative buds vs floral buds); C - *Cyathostegia* (vegetative bud vs floral bud); D - *Cadia*, *Cyathostegia*, and *Andira* (vegetative buds) in order to identify genes involved in the determination of floral organs resulting in unusual flower types in the early-branching lineages of papilionoid legumes. The main question raised in this chapter are: what are the genes which show variation in expression across the dorsoventral axis in the flowers? Is the

variation in floral morphology found in these species reflected in variation in gene expression patterns?

3.4 Material and Methods

3.4.1 RNA-seq analysis strategy

Analysis of RNA-seq data involved four steps: read processing (i.e. trimming), assembling, quantification of transcripts and differential expression analysis. Read preprocessing involved removing low quality reads and Illumina adaptor sequence, which if not filtered can cause misassembly (Andrews, 2010). The transcriptome assembly process follows a similar approach to that employed in Chapter Two for genome assembly. A *de novo* assembly was carried out, consisting of breaking down RNA-seq reads into smaller k-mers, allowing the reconstruction of transcripts using the de Bruijn graph approach (see Chapter Two). The Trinity assembler (Granheer et al., 2013) was used for *de novo* assembly.

Because Trinity assembler generates a large number of assembled sequences reflecting splice variants as well as potentially allelic variants and chimeric assemblies of these, I filtered the transcriptomes by using the Mikado v2.3.2 pipeline (Venturini et al., 2018) in order to facilitate the generation of transcript quantification data with Salmon (Patro et al., 2017) and to carry out downstream differential gene expression (DEG) analysis via 3D-RNASeq (Guo et al., 2020). The complete pipeline to run Mikado is found in Annex 1.

3.4.2 Plant material and isolation

In this chapter I analyse the species introduced in Chapter Two with the exception of *Andira anthelmia*, whose single specimen had a single branch from which I tried to extract RNA but yield was insufficient. Different procedures were used to harvest floral and vegetative tissues from our target species growing in the living collections at the Royal Botanic Garden Edinburgh (Fig. 1). For *Cadia purpurea*, the accession 19942001 was used; for *Cyathostegia mathewsii* 19981783, and for *Andira inermis* 20030344 accession numbers were used. Table 1 gives a summary of the material collected from each species.

Floral phenology posed a constraint to the sampling of *Andira inermis*. There was consistent monitoring of the species in the RBGE living collections but the specimens did not flower over the period 2019 to 2023. Only two species of *Andira* have been known to flower



Figure 1. Sampling of flower developmental series from A–C. *Cadia purpurea* (accession 19942001), mature flowers and flower buds are shown. D–F. *Cyathostegia mathewsii* (accession 19981783), flower buds are shown. Samples were collected and immediately embedded into liquid nitrogen canisters to be transported to laboratories

for immediate RNA extraction. The species are found in the arid glasshouse at the Royal Botanic Garden Edinburgh.

under glass in the UK (Pennington et al., 2000, 2018) and that flowering is occasional - *A. anthelmia* has flowered only once since its planting under glass in 1991 (Pennington et al., 2018). Because of the lack of floral buds, samples of vegetative buds were used for the RNA-seq experiments, since floral symmetry genes have been reported to be expressed in these tissues and to be involved in vegetative symmetry (Martin-Trillo & Cubas, 2010; Bello et al., 2017; Zhao et al., 2019).

3.4.3 RNA extraction, library preparation and RNA-seq

Harvested tissues were immediately placed in liquid nitrogen for subsequent RNA extraction. Total RNA was extracted from each tissue using the QIAGEN Plant Mini Kit according to manufacturer's instructions (Leebens-Mack and Carrigan, 2019), grinding tissues in pre-chilled (in liquid nitrogen) mortar and pestle. Total RNA concentration and purity were determined using a Nanodrop 1000 Spectrophotometer (Thermo Fisher Scientific) and Bioanalyzer 2100 (Agilent). Library preparation and sequencing were conducted by Novogene UK, Cambridge. NovaSeq6000 platform and Illumina Sequencing PE150 technology were used to generate data with 20M reads, 6Gb data.

3.4.4 Pre assembly quality control

Read quality was initially assessed prior to processing using FASTQC. Trimmomatic-0.35 (Bolger et al., 2014) was then used for adapter and quality trimming with the following settings:

```
trimmomatic PE -phred33 -threads 2 ILLUMINACLIP:TruSeq3-PE2.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36.
```

The pre-processed reads were stored in fastq.gz format and quality checked again with FASTQC (Appendix 2).

3.4.5 Alignment to draft genomes

Transcriptomes can be assembled as either reference-guided (mapping-first) or *de novo* (assembly-first) (Grabherr et al., 2011; Martin et al., 2011; Raghavan et al., 2021). The reference-guided approach is appropriate for model organisms for which good quality, annotated genomes are available, or where there is genomic data of a closely related species. Reads can then be mapped to a reference genome to reconstruct the corresponding transcripts (Martin et al., 2011). Mapping to the genome is also a good quality control step to confirm lack of contamination and assess gene space coverage.

BAM alignment files - In order to generate a BAM file, a genome index was generated prior to the mapping of the RNA reads as follows:

```
STAR --runThreadN 6 --runMode genomeGenerate \  
--sjdbFileChrStartEnd ./SJ.out.tab.txt \  
--genomeDir path/to/new/directory/STAR_index_Cy_1index \  
--genomeFastaFiles path/to/genome/file/genome.fasta \  
--genomeChrBinNbits 15 \  
--limitGenomeGenerateRAM 31114570677 \  
--genomeSAindexNbases 13
```

In the next step, STAR maps the forward and reverse RNA-seq reads to the indexed genome:

```
STAR --runThreadN 3 \  
--genomeDir /path/to/STAR_index_Cy_1 \  
--readFilesCommand zcat \  
--readFilesIn /path/to/forward/and/reverse/RNA-seq/reads \  
--outSAMtype BAM SortedByCoordinate
```

The previous files are then used to generate the GFF and GTF files. In a first step, GMAP builds a genome database with:

```
gmap_build -d genome <path/to/genome/fasta/file>
```

After that, the genome is mapped against its transcriptome to generate a GFF3 file with the command:

```
gmap -d genome -f 2 /path/to/transcriptome.Trinity.fasta > gff3_gene.gff
```

The resulting GFF3 file was then used as an input file to Stringtie, which generated a GTF file that was also fed into Mikado (complete pipeline in Annex 1). Stringtie also takes as input spliced alignments in coordinate-sorted BAM format to produce a GTF output which consists of assembled transcript structures and their estimated expression levels (FPKM/TPM). The following command can be used to generate the GTF file:

```
stringtie /path/to/sorted_BAM_files.bam -G /path/to/GFF_file.gff -o finaloutput.gtf
```

In order to assess the quality and distribution of the mapped reads and assembled transcripts generated with STAR, two types of alignment were carried out for each tissue. Firstly, a BAM file was generated containing the alignment of the reads to its respective genome. The quality metrics were assessed using the command *flagstat* provided by the Samtools software suite (Danecek et al., 2021). Among the relevant alignment metrics, the output of *flagstat* included information such as: the total number of reads in the alignment; the number of reads that mapped uniquely to the reference genome; the number of reads that mapped to multiple locations in the reference genome; the percentage of reads that were properly paired; the percentage of reads that were mapped concordantly (both pairs mapped correctly); the number of reads that failed to align to the reference genome (table 2). These metrics give an overview of the quality and characteristic of the alignment, allowing assessment of whether the alignment is successful and reliable for further analysis.

3.4.6 De novo transcriptome assembly with Trinity

De novo transcriptome assembly is reference-free and reconstructs transcripts directly from the reads. This approach is the only one possible if there is no close reference genome but

it is open to producing artefactual sequences through conflation of allelic and paralog variation and the generation of chimeric sequences. This is a particular problem in gene families with highly conserved domains, such as MYB transcription factors.

Although a reference-guided approach promises maximum sensitivity, it depends on correct read-to-reference alignment, a task that is complicated by splicing, sequencing errors, and the lack or incompleteness of the reference genome (Grabherr et al., 2011). Conversely, *de novo* assembly approaches do not require any read-reference alignments, which is convenient when genomic data is not available, is gapped, or is fragmented, as in the case of draft genomes (Grabherr et al., 2011). During gene mining in Chapter Two I observed that the assembled genomes for our target species are not complete, therefore in this chapter I use the filtered mikado set of transcriptomes as a reference for the downstream analyses in the 3D-RNASeq app.

3.4.7 Mikado transcriptome annotation

After generating transcriptomes for each developmental stage in *Andira*, *Cadia*, and *Cyathostegia*, annotation was carried out to refine and improve the assembled transcripts generated by Trinity. Trinity tends to assemble many more transcripts than there are genomic loci as it produces transcripts for each splice variant and can generate chimeric transcripts. To simplify the assembled transcriptome, I used Mikado v2.3.2 (Venturini et al., 2018). Mikado analyses RNA-seq assemblies by looking for overlapping transcripts based on their genomic position. Overlapping transcripts are then analyzed to locate the gene loci that will form the final annotation.

Pre-settings and pre-files for Mikado

Mikado requires a set of bam alignments, gff and gtf files and the swissport database. The main files used in our Mikado runs were a GFF and a GTF generated by the Genomic Mapping and Alignment Program (Wu and Watanabe, 2005) - GMAP - and Stringtie (Pertea et al., 2016), respectively.

3.4.8 Transcripts Quantification

Quantifying transcripts after *de novo* assembly allows the measurement of the abundance of individual genes or transcripts across different samples. This information is vital

for understanding the expression patterns of genes, identifying highly expressed genes, and comparing gene expression levels between different conditions or experimental groups (Raghavan et al., 2022). In this study, quantification was carried out using Salmon (Patro et al., 2017) using the BAM file from each sample (table 2). This method involved mapping the RNA-Seq reads to the assembled transcripts and then counting the number of reads that align to each transcript. The read counts were used for downstream differential expression analyses.

These files result from preceding data, such as transcripts relative abundance, given in transcription per million (TPM), calculated using Salmon (Patro et al., 2017) with transcripts as input files. STAR 2.7.11 was also used (Dobin et al., 2013) to generate Binary Alignment Map (BAM) files using RNA-seq reads and the mikado selected transcripts as input files. These preliminary files were generated following the steps below:

TPM quantification - The following commands were used to build a transcriptome index, which is then used to map forward and reverse reads into a quantification file:

```
salmon index -t path/to/transcriptome.Trinity_mikado.fasta -i salmon index
```

```
then;
```

```
salmon quant -i salmon_index -l A\
```

```
-1 Cy_1_forward_L3_1.fq.gz Cy_1_forward_L4_1.fq.gz \
```

```
-2 Cy_1_reverse_L3_2.fq.gz Cy_1_reverse_L4_2.fq.gz \
```

```
-p 8 --validateMappings -o path/to/output/file
```

3.4.9 Differential gene expression analysis

In order to analyze differentially expressed genes in the tissues of our target species, I employed the 3D RNA-seq pipeline (Fig. 2) developed by the University of Dundee at the James Hutton Institute (Guo et al., 2020). The 3D RNA-seq app is used to quantify RNA-seq reads to perform differential gene expression in samples from different experimental conditions. In this chapter I have used contrast conditions that represent different developmental stages found throughout flower growth such as abaxial vs adaxial flower regions, and vegetative vs young flower buds. For species that did not have floral tissues available I also analyzed vegetative buds across species.

Data pre-processing

Prior to evaluating the correlation between different tissues and establishing contrast groups, a stringent sample number cut-off was set in the 3D-RNA-seq app to filter low

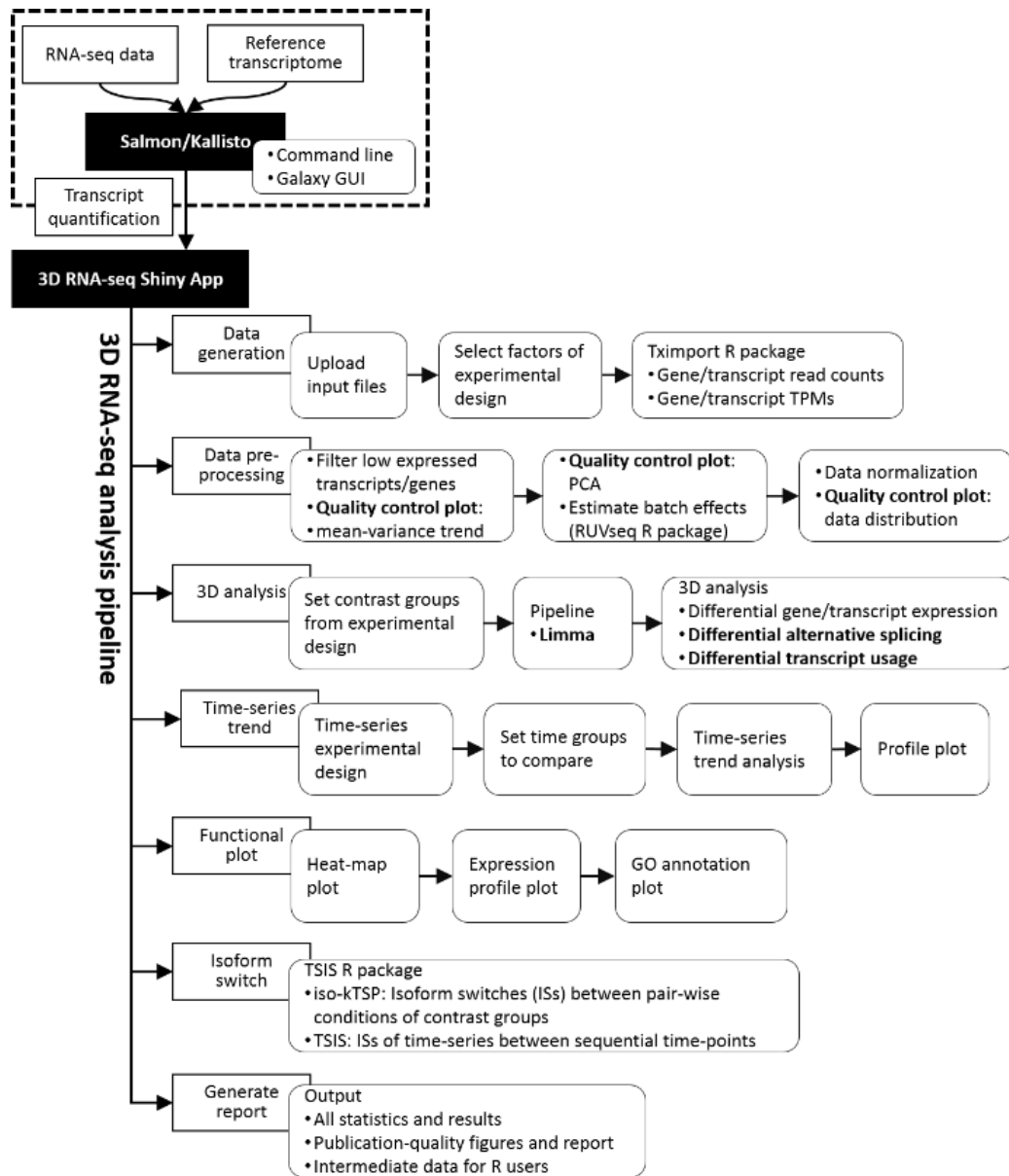


Figure 2. 3D RNA-seq General Pipeline. Prior to analyzing Salmon quantification files, data pre-processing is an important step because it creates files based on the available samples to further generate contrast groups that will be compared in the 3D analyses. Filtering of lowly expressed transcripts can be selected by adjusting sample cut-offs in order to filter only highly expressed genes. PCAs are then used to verify the distribution of transcripts by treatment, which informs the level of correlation between replicates, allowing to identify any inconsistency with sampling. The main downstream analyses are then performed in the 3D analysis step.

expressed transcripts for all species (Fig. 3), except *Andira inermis* where only vegetative buds were available and which was therefore analyzed separately. A principal component analysis (PCA) was carried out using the triple replicates from each species in order to find correlation between the different tissues. Replicates involved vegetative buds, young flower buds, and abaxial and adaxial regions of mature flowers (Fig. 4).

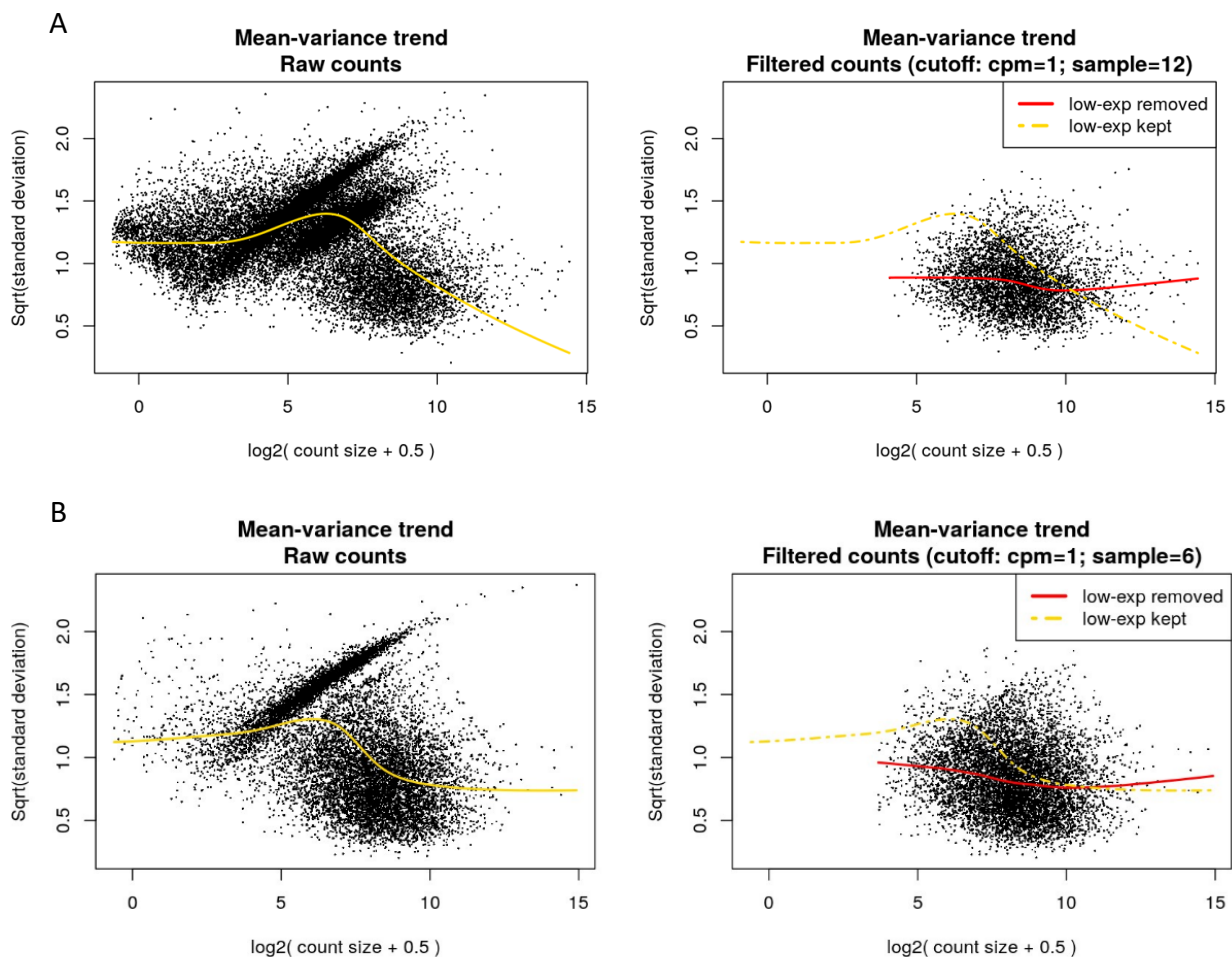


Figure 3. Gene mean variance trend before and after filtering low expressed transcripts. A – *Cadia purpurea* at a 12 sample cut-off. B – *Cyathostegia mathewsii*.

The first contrast group involved the abaxial and adaxial flower tissues of *Cadia*. Because RNA-seq data may have variation in sample quality due to, for example, degradation or contamination of specific samples, outliers can appear in the PCA plots. In such cases different pipelines can be used for balance. Here the limma-voom pipeline was used, which is

capable of managing more complex experimental designs and has more robust predictions without losing stringency (Ritchie et al. 2013).

3.5 Results and Discussion

I estimated and removed batch effects caused by biological replicates from the same tissues found in separate clusters in the PCA plot (Fig. 4). Adaxial flower samples were found near young flower bud samples, which suggests that both tissues are more similar in gene expression profiles to each other than to the abaxial flower and vegetative tissues. Abaxial replicates form a distinct cluster, indicating that this tissue has a novel suite of gene expression profiles compared to other tissues and stages. Identifying the distinctive genes will provide clues about the development of abaxial floral fate.

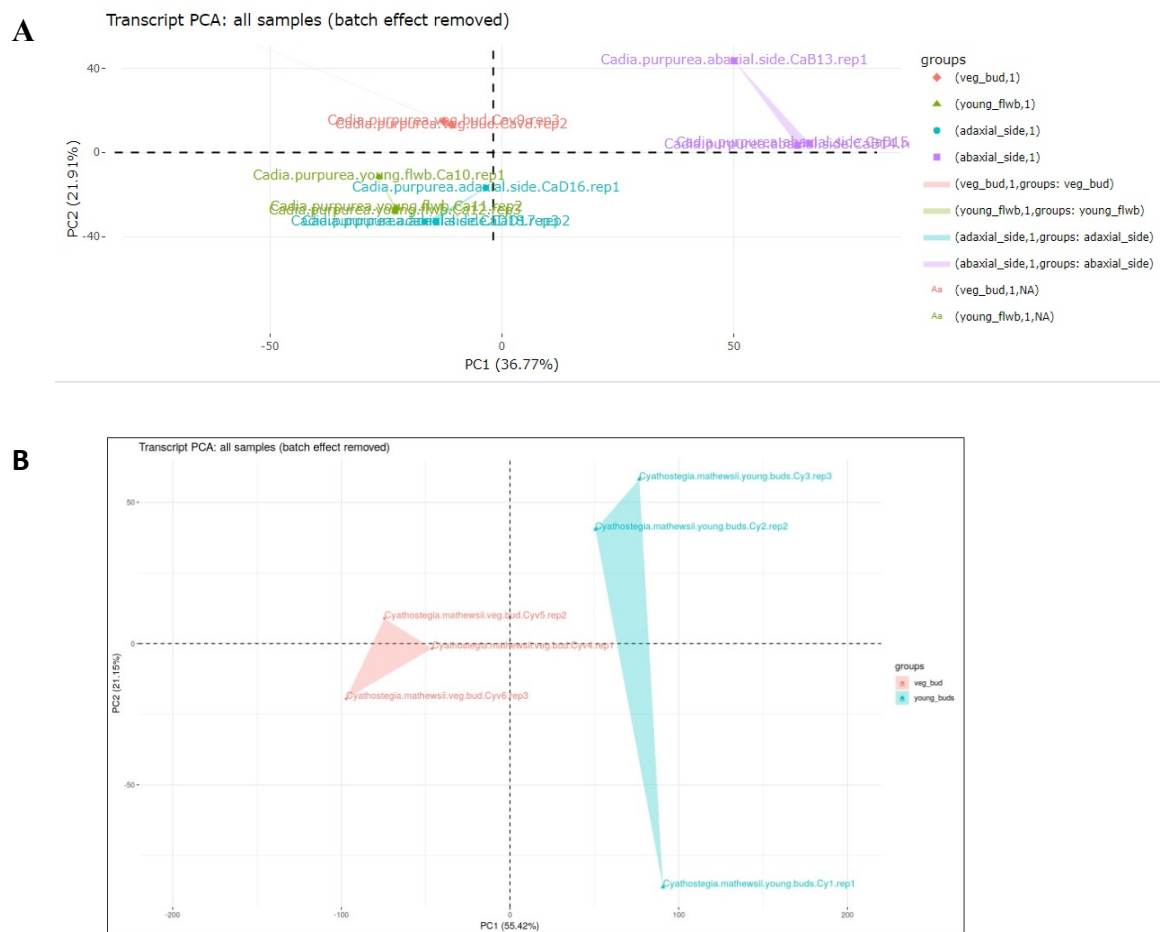


Figure 4. Principal component analysis (PCA) of floral and vegetative tissues of A-*Cadia*: young flower buds, abaxial and adaxial flower domains and vegetative buds. B- *Cyathostegia*: young flower buds and vegetative buds. The correlation between points shows the common origin of replicates from different tissues.

3.5.1 3D Analysis of contrast groups

The complete set of differentially expressed genes was analyzed to give an overview of the variation and subsequently filtered to examine the differentially expressed transcription factors. 1050 up-regulated and 1016 down-regulated genes were found for *Cadia* in the abaxial vs adaxial contrast group. Most genes in these tissues were housekeeping genes involved in photosynthesis, nutrient uptake, nucleotide metabolism, lipid and secondary metabolism (Fig. 5). This indicates that the two sides of the flower are distinct not only in morphology, but also in physiology. Genes involved in processes relevant to the development of flowers were also on the list, such as plant organogenesis, cell division, and phytohormone action.

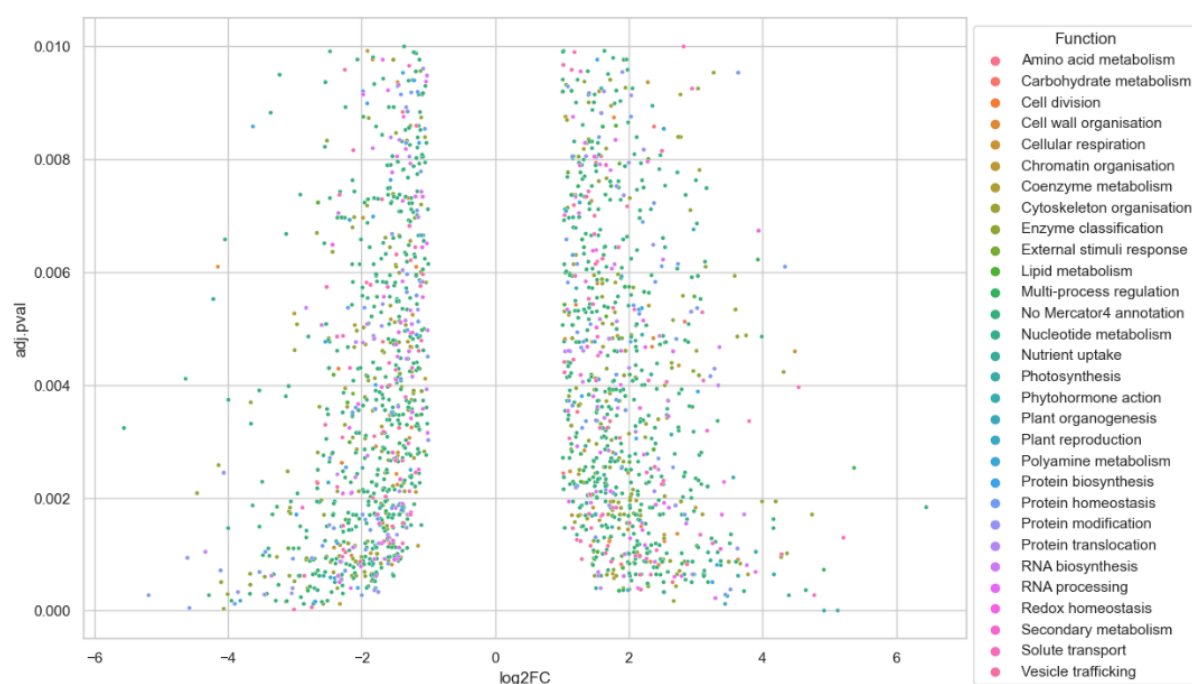


Figure 5. Volcano plot representing all genes found across abaxial vs adaxial contrast group in *Cadia*. Y axis (adj.pval) represents the adjusted *p* value and x axis (log2FC) the relative gene expression. Genes up-regulated are found at the right hand side of the plot and down-regulated genes at the left side of the plot. Genes are coloured according to function.

The top 30 up-regulated genes were filtered (Appendix, Table 3) and found several genes involved in the perception and signaling of the plant hormone ethylene - three orthologs of ethylene receptor ETR/ERS, two EIN2 ethylene signal transducer orthologs, an ortholog of ethylene signal transducer (EIN3/EIL), and an ortholog of EIN3-binding F-box protein (Appendix, Table 3:47).

Our results comparing vegetative buds, young flower buds, abaxial and adaxial floral tissues of *Cadia* corroborates the role of ethylene across flower development, with a higher expression of several genes involved in the production, signaling and response to this phytohormone found in the adaxial domain, suggesting that ethylene may be key to the distinctness of the abaxial domain (Fig. 6). Ethylene is known to be responsible for flower senescence in many species (Vaishak et al., 2019). One possible explanation for the pattern observed in the data is that the adaxial region of *Cadia*'s flower might senesce earlier than the abaxial side.

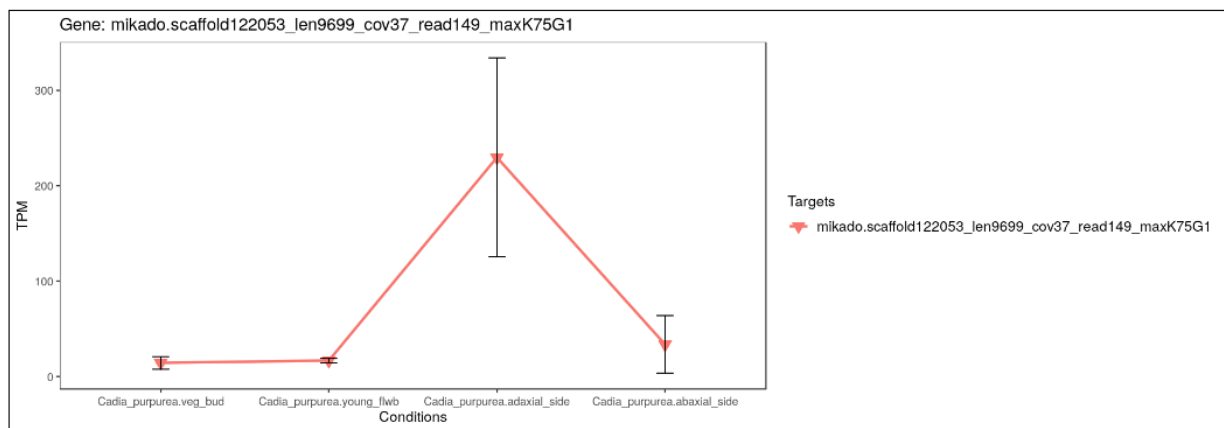


Figure 6. Relative abundance of *EIN3* transcription factor in different tissues of *Cadia purpurea*. Expression of *EIN3* has an overall increase across flower development with a pic in expression in the abaxial side of the flower.

Ethylene has an important role in the development of plants, acting together with other hormones for the development of leaves, flowers, and fruits. Particularly in flower development, the expression of ethylene biosynthesis genes seems to be linked to development of floral organs (Lqbal et al., 2017). In tobacco, ACC oxidase (*ACO*), an enzyme which catabolic products include ethylene, this phytohormone was expressed in early developing stigma, style, and ovary (De Martinis & Mariani, 1999). In tomato, *LeACO1,2,3* and *4* and *LeACS1A* transcripts were detected in pistils (Llop-Tous et al., 2000).

Ethylene has also been shown to be involved in the development of reproductive organs in the China rose *Hibiscus rosa-sinensis*, with *HrsETR* and *HrsERS* transcript levels being shown differentially in the pistil and stamens, petals, and ovary with different temporal patterns, suggesting a possible tissue-specific role (Trivelline et al., 2011). Studies in pineapple and *Arabidopsis* have shown the activity of ethylene in the very early developmental stages of flowers, with higher expression in bract primordia, inflorescence, floral meristems, developing petals and ovules (Sakai et al., 1998; Li et al., 2016).

In the common pea, *Pisum sativum*, ethylene has been reported to play a role in flower senescence. Pollination did not seem to trigger a senescence syndrome in the corolla, instead a differential increase in ethylene content was observed in different floral organs across the development of pollinated and unpollinated flowers. These results suggested that corolla senescence in *P. sativum* is endogamously regulated, rather than dependent on pollination (Orzáez et al. 1999).

To focus on developmentally important genes the top 30 differentially expressed transcription factors were filtered in Adaxial versus Abaxial tissues in the *Cadia* flower (Table 3 and 4). This set included a gene encoding a B-box protein-like transcript (*BBX*). The B-box proteins have a conserved domain in many proteins involved in cell growth control, differentiation and transcriptional regulation among animals and plants (Crocco & Botto, 2013). For example, in *Arabidopsis thaliana*, *CO/AtBBX1* plays a central role in the development of flowers by photoperiod control (Onouchi et al., 2000; Samach et al., 2000).

BBX expression is also related to ethylene, mentioned above. The major roles of joint activity of *BBX* and ethylene are not well explored, but recent studies in *Arabidopsis* have shown the action of *BBX24* proteins to mediate ethylene positive regulation of flowering (Vaishak et al., 2019). Our data confirms both ethylene and *BBX* genes being up-regulated during flower development in *Cadia* (Figs. 6 and 7A), indicating that these genes are important regulators of flower development.

B-Box transcription factors play an important role in mediating the interaction between light and hormones and, among other functions, is involved in the flowering timing with the regulation of the CONSTANS (CO) domain (Vaishak et al., 2019). In *Arabidopsis*, the interaction between CO and another protein, *FLOWERING LOCUS T* (FT), which is a mobile signaling protein that acts as a long-distance messenger that moves from leaves to the shoot

apical meristem, where it induces the transition from vegetative to reproductive growth (Samach et al., 2000).

BBX proteins were listed among the top 30 up-regulated TFs in other *Cadia* contrast groups, showing a higher expression in the vegetative buds when compared with the other tissues (Fig. 7A). This outcome agrees with the function of these proteins since the vegetative buds represent a transition to a reproductive phase. Moreover, a clear decrease in the expression of *BBX* can be observed throughout *Cadia* development (Fig. 7A). Although not listed among the most up or down-regulated genes, both *FT* and *FLOWERING LOCUS K* (FK) were expressed in all tissues, however following a slightly different profile (Fig. 7B, C). Although related to the *BBX* genes, variation in *FT-like* and *FK-like* expression are not significant.

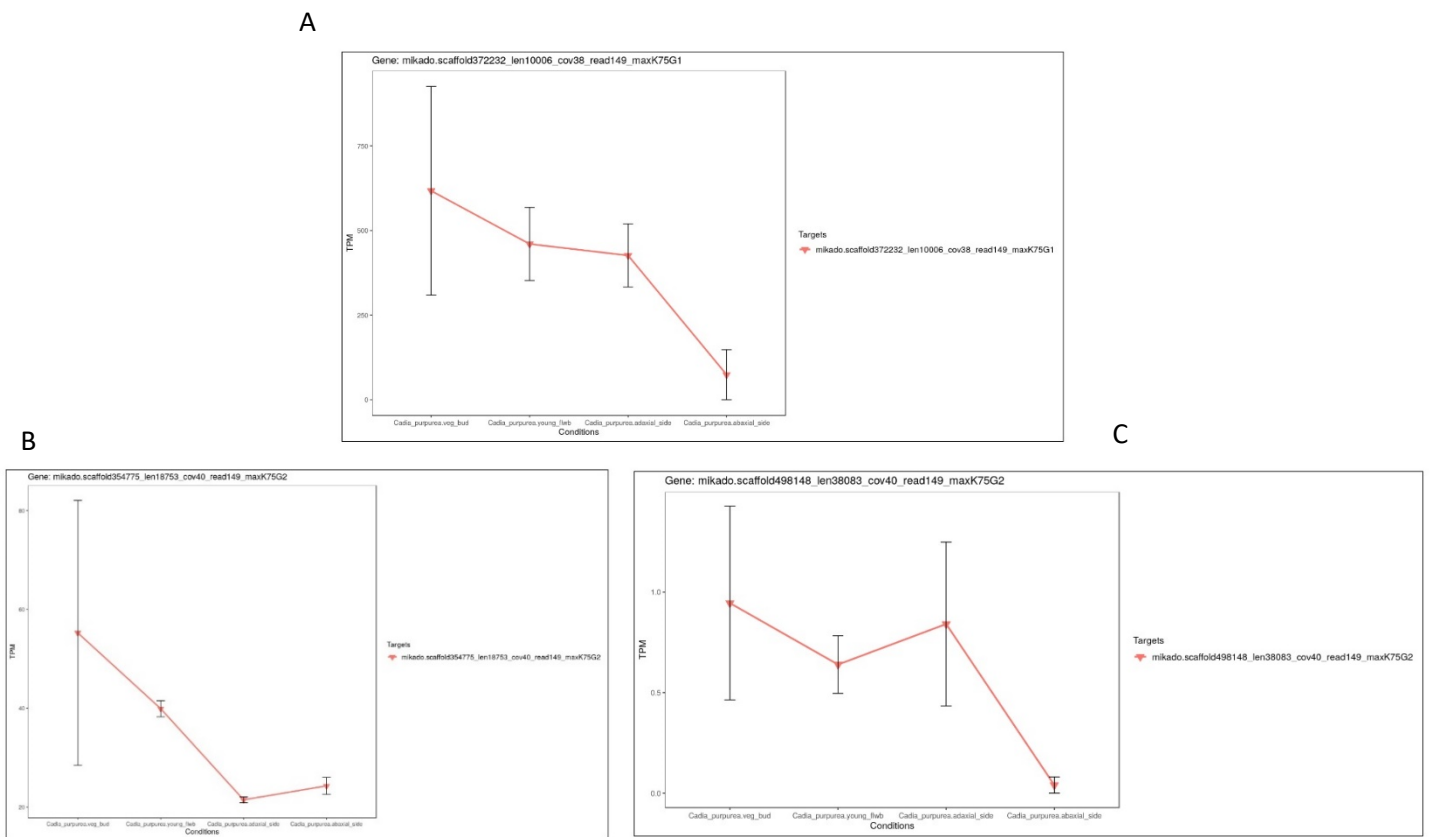


Figure 7. Relative expression of *BBX* (A), *Flowering Locus K* (B) and *Flowering Locus T* (C) across flower development in *Cadia purpurea*. *BBX* is highly expressed in *Cadia*'s vegetative buds, but decreased in the following developmental stages, with a pronounced drop in the abaxial side of the flowers. The expression tendency similar for *FK* and unclear for *FT*.

Another contrast group involved *Cadia*'s vegetative buds compared to the young flower. 93 down-regulated genes and 58 up-regulated were identified, from which five were transcription factors. The distribution of these genes is presented in Fig. 8 by function. The top 30 upregulated genes in flower buds relative to vegetative buds included plant *Cryptochromes* and *PLATZ* genes. *PLATZ* transcription factors are zinc dependent DNA binding proteins, binding to AT-rich sequences and functions in transcriptional repression. In pea, *PLATZ1* was found to be preferentially expressed in the active cell division regions, including the root tip and terminal buds. Although *PLATZ1* proteins act as a transcriptional repressor by nonspecifically binding to AT-rich sequences in pea (Nagano et al., 2001), its orthologue in *Arabidopsis*, *ORE15*, acts as a transcriptional activator, mostly participating in the regulation of leaf development, but also with transcripts highly accumulated in seedlings, roots, and floral organs (Kim et al., 2018). *PLATZ* genes are generally involved in processes where cell proliferation is active, such as meristematic tissues where cell division is constantly active. This is the case of developing floral organs and vegetative buds, where meristematic activity is high throughout organ development.

Among the *PLATZ* transcription factors, the floral gene *PERIANTHIA PLATZ* was found. Mutations of *PERIANTHIA* (*PAN*) gene in *Arabidopsis* have been shown to transform flowers from tetramerous to largely pentamerous (Chuang et al., 1999). Given that this gene is up-regulated in *Cadia*'s floral tissues, it is likely that *PERIANTHIA PLATZ* plays a role in the establishment of *Cadia*' pentamerous corolla.

Plant *Cryptochromes* (*CRYs*) are UV-A/blue light photoreceptors that play an important role in regulating blue light-mediated physiological responses (Yang et al., 2017). *CRY* genes have been analyzed in pea and soybean with activity in the regulation of stem elongation, leaf expansion and photoperiodic flowering in these species (Zhang et al., 2008). These genes might work in connection with the previously discussed *BBX* and *ethylene* by signaling floral tissue development through light mediation.

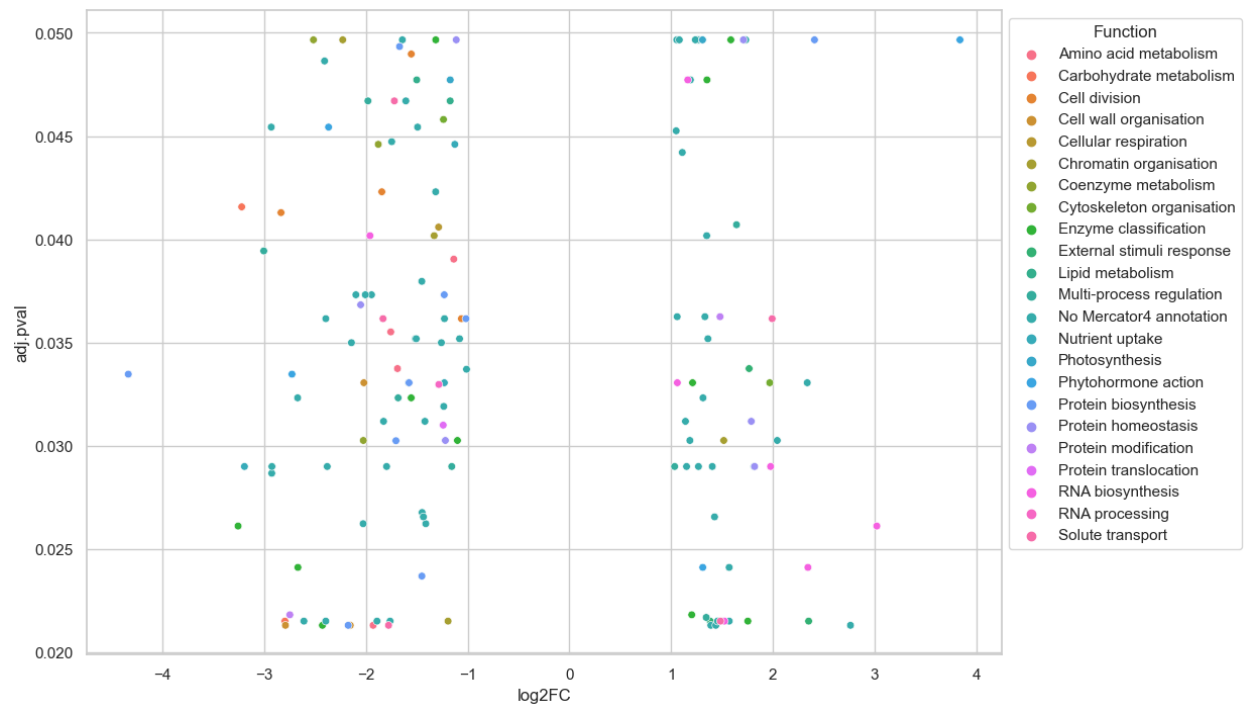


Figure 8. Distribution of up and down regulates transcripts in the contrast group comparing vegetative buds and young flower buds in *Cacia*. The majority of transcripts observed are related to maintenance of cellular function. We shortlisted our sampling to the top 30 upregulated genes and present the ones which have relevance for flower growth.

Our analysis of *Cyathostegia* vegetative buds and young flower buds showed only 18 genes being differentially expressed in the vegetative buds compared to young flower buds. Sixteen of these genes were shown up-regulated and only two down-regulated. This may be the result of over-stringent filtering in the analysis, or a poor sequencing from one of the samples, leading to rejection from the analysis due to incomplete data. A more lax approach to the analysis might perhaps recover relevant genes involved in flower development, since low expressed genes could have been excluded because of filter stringency. This is the probable cause for the very few genes detected, since the expression profiles of the candidate genes discussed in Chapter Two were found in low levels in the tissues analyzed (see discussion in the next section).

To allow comparison across species I summed the TPM by gene annotation and looked for types of genes which differed greatly in expression between species either in the flowers or in the vegetative tissues (Fig. 9).

Transcripts from the: RADIALIS-like (15.5.3.2.1) and TCP (15.5.7.7.1) gene families are much higher in *Cadia* flower buds than in *Cyathostegia* buds are. This variation in expression of key symmetry genes could well be related to the variation in floral morphology between the species (Fig. 9A). The variation between species is due to the variation in the number of transcripts annotated in each species. Mapman annotation found 7 RAD-like genes in *Cadia* and only one in *Cyathostegia*. It found 24 TCP-like genes in *Cadia* and none in *Cyathostegia*.

Transcripts in *Cyathostegia* flower buds are higher compared to *Cadia* from genes related to the homeobox meristem identity gene BEL (15.5.3.1.4.3) and to the boundary gene, NAC (15.5.7.1), as well as genes related to the flowering time gene family, CONSTANS (15.5.4.5.1.1) (Fig. 9A). These may be good candidates for the variation in floral form and development.

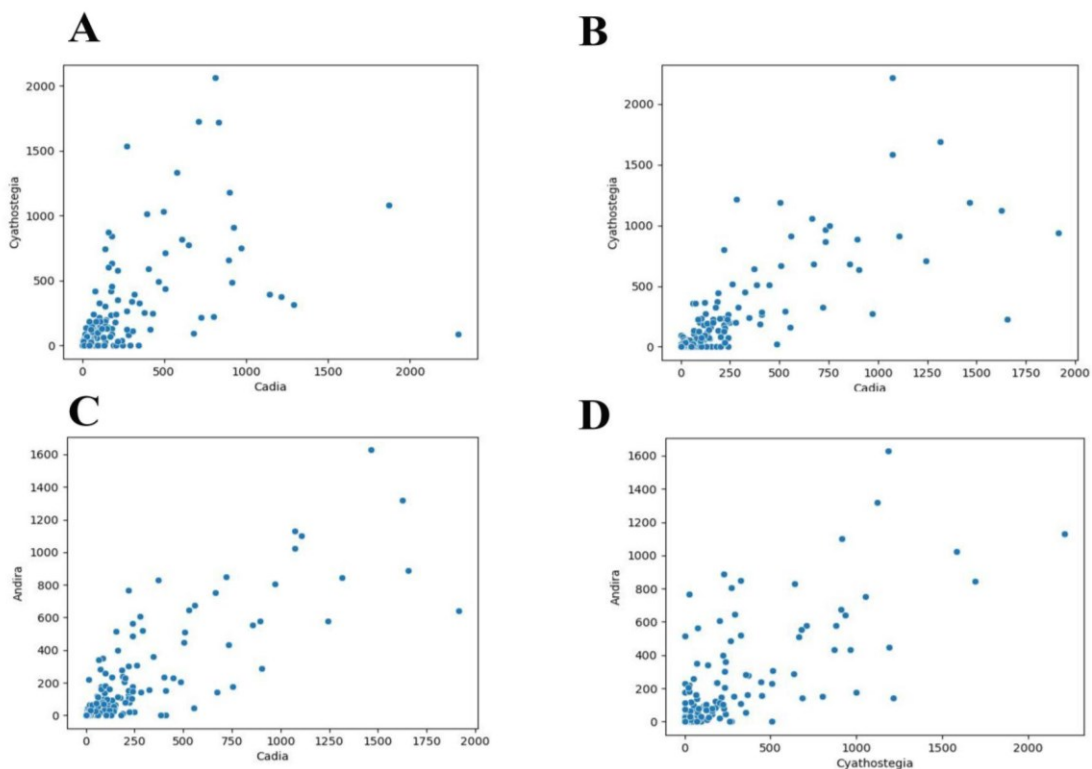


Figure 9. Summary graphs comparing expression of transcripts in the contrasts: A. floral tissues of *Cadia* and *Cyathostegia*; B. vegetative buds of *Cadia* and *Cyathostegia*; C. vegetative buds of *Andira* and *Cadia*; D. vegetative buds of *Andira* and *Cyathostegia*. Expression is given in TPM

Comparing gene expression patterns across vegetative buds (Fig. 9) brings up an Ethylene response gene family which is only expressed in *Cadia* and *Cyathostegia*, and transcripts belonging to the gene family of FAR1 - a nuclear localized protein involved in far red light response signaling, which are also missing in *Andira*. In contrast *Andira* has high expression of transcripts related to LIGHT-DEPENDENT SHORT HYPOCOTYLS 3 and a LOB gene which might be involved in boundary specification. CONSTANS-like genes and BEL-like genes are also expressed at higher levels in *Cyathostegia* vegetative buds compared to *Cadia* vegetative buds. GROWTH REGULATING FACTOR 5 (GR5) also has higher expression in *Cyathostegia* vegetative buds when compared to *Cadia*. GR5 stimulates cell division and is also correlated with photosynthesis and leaf longevity in *Arabidopsis* (Vercruyssen et al., 2015). The presence of high expression of GR5 in vegetative buds likely indicates high photosynthetic activity in these tissues.

3.5.2 Differential expression of candidate genes involved in flower symmetry

In addition to analyzing the overall profile of genes differentially expressed in the contrast groups mentioned above, I looked into the expression of the transcription factors discussed in Chapter One and Two - CYC, DIV, RAD, and DRIF - to gain insights on how gene expression may be related to the distinct types of floral morphologies found in our group of species. Even though a thorough representation of floral tissues for all the species is presented, the expression of these genes in the vegetative bud tissues is also discussed.

All the CYC gene family members have a low level of expression across our study species. In *Cadia* (Fig. 10) m282224 the CYC1 ortholog is variable in expression between replicates, as is m304621 (LegCYC1B). This suggests a dynamic expression pattern which changes between exact developmental stages, or between conditions. m73713 (LegCYC1B) and m482573 (LegCYC3) are more evenly expressed, but altogether the levels are so low that this variation is not significant. All four are expressed in all three replicates in the adaxial domain, but only in one replicate in the abaxial domain. This supports an adaxial function. All but m482573 (legCYC3) are also expressed in vegetative buds, confirming a broader role than floral symmetry, which may constrain their evolution. There is no strong evidence from expression patterns for neo-functionalisation in the duplicated genes.

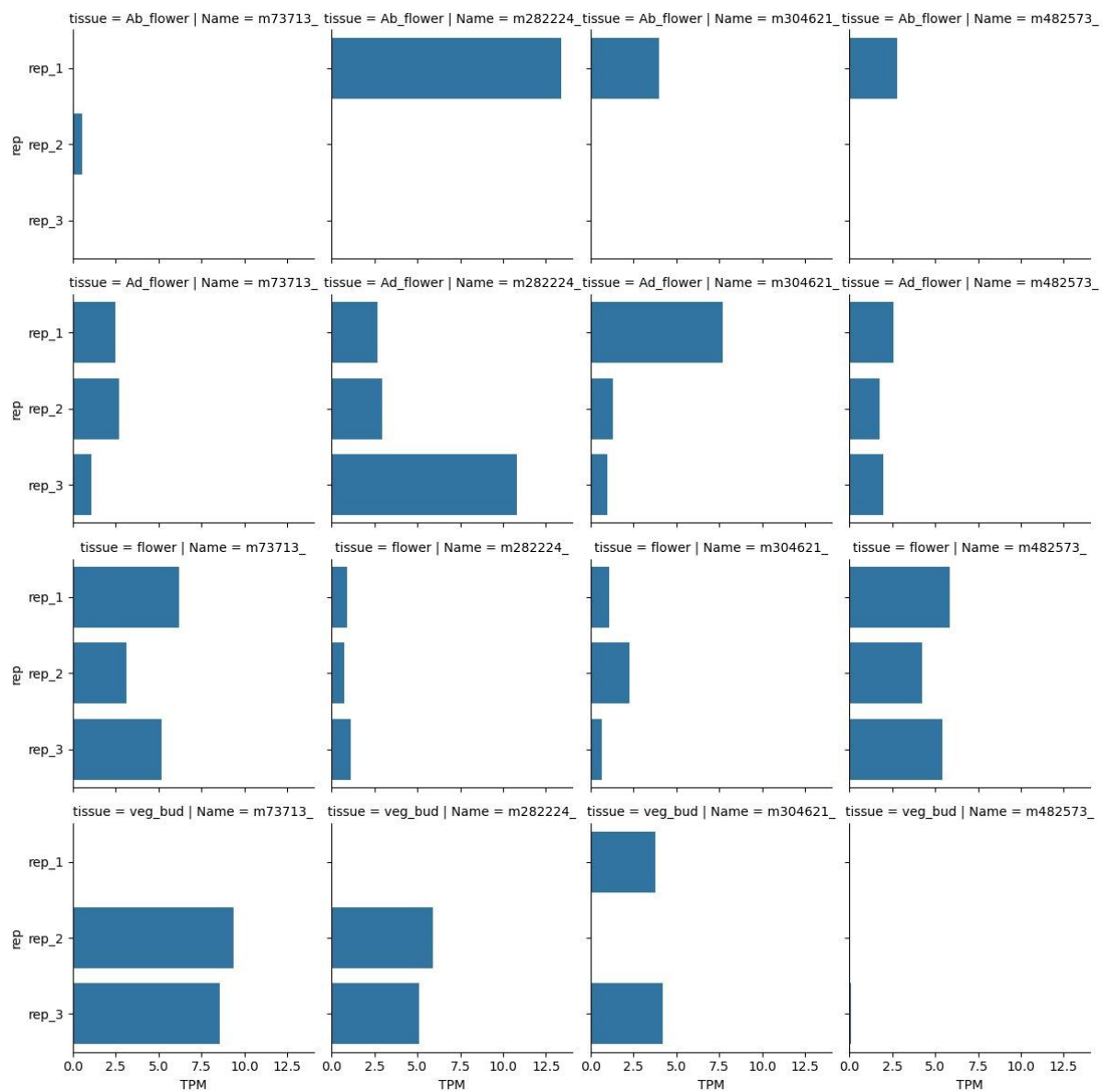


Figure 10. Relative expression of CYC transcripts in TPM for all tissues in *Cadia*. Columns represent different CYC paralogs and lines are labeled for each tissue. Replicates are ordered from 1-3. Ab_flower = abaxial region of flower, Ad_flower = adaxial region of flower; flower = young flower buds; veg_bud = vegetative flower buds.

RAD shows more variation between paralogs (Fig. 11). There is very little expression in the adaxial and abaxial flowers. What there is, is more in the adaxial domain based on expression in all replicates in the adaxial flower but only one of the abaxial replicates. In the younger flower buds there is a strong expression of m131773 and some expression of m56113. The expression of m131773 is highly upregulated on the floral transition.



Figure 11. Summary of relative expression of CYC, DIV, RAD, and DRIF for all tissues in *Cadia*. Genes are represented in columns with their respective paralogs. Tissues are ordered in the lines with their respective replicates. Discussion is provided in the main text.

DIV gene family members show similar patterns to CYC, though at a generally higher expression level. Higher expression was detected in the adaxial domain than in abaxial except for m360942, in the flower and in the vegetative bud. This is not in keeping with the expected patterns, which should show adaxial-abaxial variation based on the role of these genes in other taxa (see Ch 1, 2). Strong adaxial-abaxial variation may not be seen as expected due to sampling older tissue, or maybe only m360942 is truly DIV.

DRIF is similar to CYC, showing more reliable expression in the adaxial samples, expression in the flower and the vegetative bud, and little variation between the two paralogs. Generally there is only very low expression, so the variation between tissues is likely not significant. It may be active only in a subset of cells.

In *Cyathostegia*, no difference between paralogs of CYC, DIV or RAD is seen, and have data only for one copy of DRIF (Fig. 12). Expression for all the genes is seen in only one of the replicates for floral buds. These could be stochastic, or may be linked to technical issues

with some replicates. DIV shows strong variation between replicates in the vegetative buds. This could be linked to changes in expression during the floral transition.

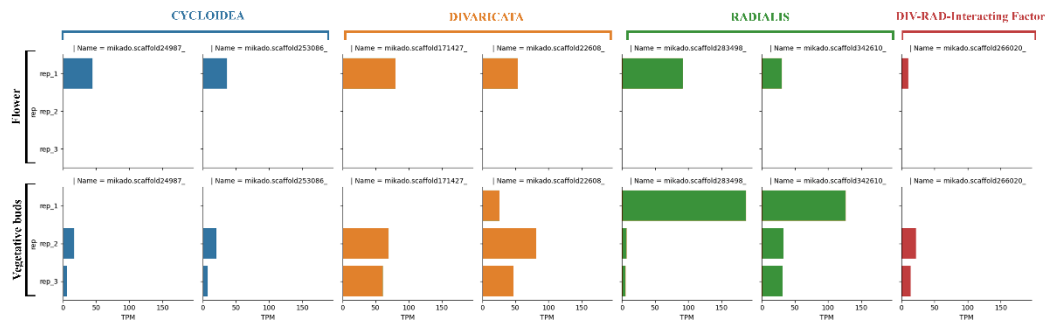


Figure 12. Summary of relative expression of CYC, DIV, RAD, and DRIF for all tissues in *Cyathostegia*. Genes are represented in columns with their respective paralogs. Tissues are ordered in the lines with their respective replicates

In *Andira* no data for CYC family members has been found, but there is clear variation in expression between paralogs of RAD, with m_12647 consistently high across replicates and m4028 low. For DIV m14 is higher than the other two paralogs. This supports evolution of functional variation between the paralog lineages.

3.6 Conclusion

Through use of RNA-Seq I have been able to establish that floral development in our study species does not just vary in morphology but also in many other aspects of physiology and metabolism. I have identified ethylene signaling and response as an important axis of variation between *Cadia* and *Cyathostegia* along with pathways involved in regulation of meristem size, boundary formation and response to light. Although I do not find absolute differences in expression patterns of our candidate genes there is support for their importance in dorsoventrality and as dynamic elements of the gene regulatory networks during the floral transition and early floral development.

As discussed in Chapter Two, copy number was variable among different morphologies, suggesting that copy number does not play a role in the establishment of either papilionate flowers or unusual morphologies such as the radially symmetric flower of *Cadia* or the single petal in *Cyathostegia*. Here I do not find strong evidence for gene expression changes associated with specific morphologies, but I do find evidence of some variation in expression patterns between paralogs which could explain why multiple paralogs have been retained.

This chapter also contributes to the knowledge of canonical genes involved in flower development in legumes as well as novel genes with putative role in the transition between vegetative and reproductive organs. Previous studies have used Real Time PCR and *in situ* hybridization to detect and observe the local expression of CYC in *Cadia* (Citerne et al., 2006). However successful and providing interesting insights, the lack of a broader range of morphology types represented a constraint to our extrapolations across morphologically distinct species. Thus, in this chapter I provide a perspective from the quantification of expression of these genes and elaborate on the differential expression observed across different developmental stages and between distinct species. My findings provide additional knowledge to the dynamic role and multiple classes of genes being responsible for the flower development of distinct legume species.

References

- Andrews, S. 2010. FastQC: a quality control tool for high throughput sequence data. Available at: <http://bioinformatics.babraham.ac.uk/projects/fastqc/>
- Bello, M. A., P. Cubas, I. Alvarez, G. Sanjuanbenito, J. Furtés-Aguilar. 2017. Evolution and expression patterns of CYC/TB1 gene in *Anacyclus*: phylogenetic insights for floral symmetry genes in Asteraceae. *Frontiers in Plant Science* 8: 589.
- Chuang, C.-F., M. P. Running, R. W. Williams, E. M. Meyerowitz. 1999. The *PERIANTHIA* gene encodes a bZIP protein involved in the determination of floral organ number in *Arabidopsis thaliana*. *Genes and Development* 13(3): 334–344.
- Crocco, C. D. & J. F. Botto. 2013. BBX proteins in green plants: insights into their evolution, structure, feature and functional diversification. *Gene* 531(1): 44–52.
- Emms, D.M. & S. Kelly. 2015. OrthoFinder: solvin fundamental biases in whole genome comparisons dramatically improves orthoroup inference accuracy. *Genome Biol.* 16: 157.
- Feng X, Zhao Z, Tian Z, Xu S, Luo Y, Cai Z, Wang Y, Yang J, Wang Z, Weng L, Chen J, Zheng L, Guo X, Luo J, Sato S, Tabata S, Ma W, Cao X, Hu X, Sun C, Luo D. 2006. Control of petal shape and floral zygomorphy in *Lotus japonicus*. *Proceedings of the National Academy of Sciences* 103: 4970–4975.
- Danecek, P., J. K. Bonfield, J. Liddle, J. Marshal, V. Ohan, M. O. Pollard, A. Whitwham, T. Keane, S. A. McCarthy, R. M. Davies, H. Li. 2021. Twelve years of SAMtools and BCFtools. *GigaScience* 10(2): giab008.
- De Martinis, D. & C. Mariani. 1999. Silencing gene expression of the ethylene-forming enzyme results in a reversible inhibition of ovule development in transgenic tobacco plants. *Plant Cell* 11: 1061–1072.
- Dong, Z., O. Danilevskaya, T. Abadie, C. Messina, N. Coles, M. Cooper. 2012. A gene regulatory network model for floral transition of the shoot apex in maize and its dynamic modeling. *PloS One* 7: e43450.
- Dong, Z. C., Z. Zhao, C. W. Liu, J. H. Luo, J. Yang, W. H. Huang, X. H. Hu, T. L. Wang, D. Luo. 2005. Floral patterning in *Lotus japonicus*. *Plant Physiology* 137: 1272–1282.
- Garg, R. & Jain M. 2013. Transcriptome analyses in legumes: a resource for functional genomics. *The Plant Genome* 6(3): 1–9.
- Gao, Y., B. Selee, E. L. Schnabel, W. L. Poehlman, S. A. Chavan, J. A. Frugoli, F. A. Feltus. 2022. Time series transcriptome analyses in *Medicago truncatula* shoot and root tissue during early nodulation. *Frontiers in Plant Science* 13: 861639.

- Grabherr, M. G. et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* 29(7): 644–52.
- Hecht, V. et al. 2005. Conservation of *Arabidopsis* flowering genes in model legumes. *Plant Physiology* 137: 1420–1434.
- Hidalgo, M., C. Ramos, J. Vázquez-Regalado, G. Zolla. 2022. Closing the gap in the “ABC” model in legumes: a review. *Legume Research* 45: 1465–1475.
- Jackson, B. C. 2011. Recombination-suppression: how many mechanisms for chromosomal speciation? *Genetica* 139: 393–402.
- Jung, C. H., C. E. Wong, M. B. Singh, P. L. Bhalla. 2012. Comparative genomic analysis of soybean flowering genes. *PLoS One* 7: e38250.
- Kim, J. H., J. Kim, S. E. Jun, S. Park, R. Timilsina, D. S. Kwon, Y. Kim, S.-J. Park, J. Y. Hwang, H. G. Nam, G.-T Kim, H. R. Woo. 2018. ORESARA15, a PLATZ transcription factor, mediates leaf growth and senescence in *Arabidopsis*. *New Phytologist* 220(2): 609–623.
- Korpelainen, E., J. Tuimala, P. Somervuo, M. Huss, G. Wong. 2014. RNA-seq Data Analysis. CRC Press. Florida. USA.
- Kudapa, H., V. Garg, A. Chitkineni, R. K. Varshney. 2018. The RNA-Seq-based high resolution gene expression atlas of chickpea (*Cicer arietinum* L.) reveals dynamic spatio-temporal changes associated with growth and development. *Plant Cell Environment* 41(9): 2209–2225.
- Lee, J., L. S. Heath, R. Grene, S. Li. 2019. Comparing time series transcriptome data between plants using a network module finding algorithm. *Plant Methods* 15: 61.
- Leebens-Mack, J. & C. Carrigan. 2019. RNA isolation from plant tissue protocol 1: Qiagen RNeasy Plant Mini Kit. GigaScience Press.
- Li, Y., Q.S. Wu, X. Huang, S. H. Liu, H. N. Zhang, Z. Zhang, G. M. Sun. 2016. Molecular cloning and characterization of four genes encoding ethylene receptors associated with pineapple (*Ananas comosus* L.) flowering. *Frontiers in Plant Science* 7: 710.
- Liao, Y., G.K. Smyth, W. Shi. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30(7): 923–930.
- Libaut, M., A. Farmer, T. Joshi, K. Takahashi, R. J. Langley, L. D. Franklin, J. He, D. Xu, G. May, G. Stacey. 2010. An integrated transcriptome atlas of the crop model *Glycine max*, and its use in comparative analyses in plants. *The Plant Journal* 53: 86–99.

- Llop-Tous, I., C. S. Barry, D. Grierson. 2000. Regulation of ethylene biosynthesis in response to pollination in tomato flower. *Plant Physiology* 123: 971–978.
- Lqbal, N., N. A. Khan, A. Ferrante, A. Trivellini, A. Francini, M. I. R. Khan. 2017. Ethylene plant growth, development and senescence: interaction with other phytohormones. *Frontiers in Plant Science* 8: 475.
- Lyons, E., B. Pedersen, J. Kane, M. Alam, R. Ming, H. Tang, et al. 2008. Finding and comparing syntenic regions among *Arabidopsis* and the outgroups papaya, poplar, and grape: CoGe with rosids. *Plant Physiology* 148: 1772–1781.
- Martin-Trillo, M. & P. Cubas. 2010. TCP genes: a family snapshot ten years later. *Trends in Plant Science* 15(1): 31–9.
- Mohadevi, S., Y. van de P., K. Vandepoele. 2011. Comparative network analysis reveals that tissue specificity and gene function are important factors influencing the mode of expression evolution in *Arabidopsis* and rice. *Plant Physiology* 156: 1316–1330.
- Mutwil, M., S. Klie, T. Tohge, F. M., Giorgi, O. Wilkins, M.M. Campbell, et al. 2011. PlaNet: combined sequence and expression comparisons across plant networks derived from seven species. *Plant Cell* 23: 895–910.
- Nagano, Y., H. Furuhashi, T. Inaba, Y. Sasaki. 2001. A novel class of plant-specific zinc-dependent DNA-binding protein that binds to A/T-rich DNA sequences. *Nucleic Acids Research* 29: 4097–4105.
- Onouchi, H., M. J. Igeno, C. Perilleux, K. Graves, G. Coupland. 2000. Mutagenesis of plants overexpressing CONSTANS demonstrates novel interactions among *Arabidopsis* flowering-time genes. *Plant Cell* 12: 885–900.
- Orzaéz, D., R. Blay, A. Granell. 1999. Programme of senescence in petals and carpels of *Pisum sativum* L. flowers and its control by ethylene. *Planta* 208: 220–226.
- Pennington, R. T., G. P. Lewis, M. Marsh. 2000. *Andira inermis* subsp. *inermis* (Leguminosae-Papilionoideae). *Curtis Botanical Magazine* 17: 188–194.
- Pennington, R. T., H. C. Lima, F. Inches, N. Watherston. 2018. *Andira anthelmia* (Leguminosae-Papilionoideae). *Curtis Botanical Magazine* 35(2): 125–133.
- Raghavan, V., L. Kraft, F. Mesmy, L. Rigerte. 2022. A simple guide do *de novo* transcriptome assembly and annotation. *Briefings in Bioinformatics* 23(2): 1–30.
- Ritchie, M. E., B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, G. K. Smyth. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* 43(7): e47.

- Roulin, A., P. L. Auer, M. Libaut, J. Schlueter, A. Farmer, G. May, G. Stacey, R. W. Doerge, S. A. Jackson. The fate of duplicated genes in a polyploid plant genome. *The Plant Journal* 73(1): 143–153.
- Sakai, H., J. Hua, Q. G. Chen, C. Chang, L. J. Medrano, A. B. Bleecker, E. Y. Meyerowitz. 1998. ETR2 is an ETR1-like gene involved in ethylene signaling in *Arabidopsis*. *Proceedings of the National Academy of Sciences U.S.A.* 95: 5812–5817.
- Samach, A., H. Onouchi, S. E. Gold., G. S. Ditta, Z. Schwarz-Sommer, M. Yanofsky, G. Coupland. 2000. Distinct roles of CONSTANS target genes in reproductive development in *Arabidopsis*. *Science* 288(5471): 1613–1616.
- Singh, V.K., R. Garg, M. Jain. 2013. A global view of transcriptome dynamics during lower development in chickpea by deep sequencing. *Plant Biotechnology Journal* 11(6): 691–701.
- Singh, V. K. & M. Jain. 2014. Transcriptome profiling for discovery of genes involved in shoot and apical meristem and flower development. *Genome Data* 2: 135–138.
- Trivelline, A., A. Ferrante, P. Vernieri, G. Serra. 2011. Effects of abscisic acid on ethylene biosynthesis and perception in *Hibiscus rosa-sinensis* L. *Journal of Plant Growth Regulation* 30: 175–184.
- Venturini, L., C. Shabbonam, G. G. Kaithakottil, D. L. Mapleson, D. Swabreck. 2018. Leveraging multiple transcriptome assembly methods for improved gene structure annotation. *GigaScience* 7(8): giy093.
- Wang, Z., M. Gerstein, M. Snyder. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10: 57–63.
- Wang, D., G. Cao, P. Fang, L. Xia, B. Cheng. 2017. Comparative transcription analysis of different *Antirrhinum* phyllotaxy nodes identifies major signal networks involved in vegetative-reproductive transition. *PloS One* 12(6): e0178424.
- Yao, X., H. M., L. Wang, D. Zhang. 2007. Genome-wide comparative analysis and expression pattern of TCP gene families in *Arabidopsis thaliana* and *Oryza sativa*. *Journal of Integrative Plant Biology* 49: 885–897.
- Zhang, Q., H. Li, R. Li, R. Hu, C. Fan, F. Chen, Z. Wang, X. Liu, Y. Fu, C. Lin. 2008. Association of the circadian rhythmic expression of *GmCRY1a* with a latitudinal cline in photoperiodic flowering of soybean. *Proceedings of the Academy of Natural Sciences* 105: 21028–21033.

Chapter 3. Differential expression of symmetry genes in the early branching Papilionoideae

Table 1. List of species and their respective tissues used in this study.

Species/tissue	Young flower buds	Mature flowers	Vegetative buds	Abaxial/adaxial
<i>Cadia purpurea</i>	3 replicates	3 replicates	3 replicates	3 replicates
<i>Cyathostegia mathewsii</i>	3 replicates	3 replicates	3 replicates	
<i>Andira inermis</i>			3 replicates	

Table 2. Basic alignment metrics of the reads against their respective genomes using *flagstat* from Samtools.

	Ca_10	CaB_13 (abaxial)	CaD_16 (adaxial)	Cav_7 (veg. bud)	Cy_1 (young buds)	Cyv_4 (veg. buds)	Andi_19 (veg. buds)
QC-passed reads + QC failed reads	83550704 + 0	75012114 + 0	50152220 + 0	1393874 + 0	40018140 + 0	1450426 + 0	8809246 + 0
Secondary	830798 + 0	744810 + 0	452422 + 0	191748 + 0	344926 + 0	175876 + 0	70980 + 0
Supplementary	0 + 0	0 + 0	0 + 0	0 + 0	0 + 0	0 + 0	0 + 0
Duplicates	0 + 0	0 + 0	0 + 0	0 + 0	0 + 0	0 + 0	0 + 0
Mapped 100.00% ; N/A	83550704 + 0	75012114 + 0	50152220 + 0	1393874 + 0	40018140 + 0	1450426 + 0	8809246 + 0
Paired in sequencing	82719906 + 0	74267304 + 0	49699798 + 0	1202126 + 0	39673214 + 0	1274550 + 0	8738266 + 0
read 1	41359953 + 0	37133652 + 0	24849899 + 0	601063 + 0	19836607 + 0	637275 + 0	4369133 + 0
read 2	41359953 + 0	37133652 + 0	24849899 + 0	601063 + 0	19836607 + 0	637275 + 0	4369133 + 0
Properly paired 100% ; N/A	82719906 + 0	74267304 + 0	49699798 + 0	1202126 + 0	39673214 + 0	1274550 + 0	8738266 + 0
With itself and mate mapped	82719906 + 0	74267304 + 0	49699798 + 0	1202126 + 0	39673214 + 0	1274550 + 0	8738266 + 0
Singletons (0.00% ; N/A	0 + 0	0 + 0	0 + 0	0 + 0	0 + 0	0 + 0	0 + 0
With mate mapped to a different chr	0 + 0	0 + 0	0 + 0	0 + 0	0 + 0	0 + 0	0 + 0
With mate mapped to a different chr (mapQ>=5)	0 + 0	0 + 0	0 + 0	0 + 0	0 + 0	0 + 0	0 + 0

Table 3. Top 30 up-regulated genes in the contrast adaxial vs abaxial in *Cadia*.

	target	contrast	adj.pval	log2FC	up.down	BINCODE	NAME	DESCRIPTION
729	mikado.scaffold236868_le n6622_cov32_read149_m a...	Cadia_purpurea.adaxial_ side- Cadia_purpurea.abaxial_	0.006219	3.929978	up-regulated	27.6.2.3	Multi-process regulation.phosphoinositide sign...	mercator4v6.0: phosphatidylinositol 4-kinase *
762	mikado.scaffold298723_le n45656_cov42_read149_ m...	Cadia_purpurea.adaxial_ side- Cadia_purpurea.abaxial_	0.00673	3.938841	up-regulated	10.1.2	Redox homeostasis.reactive oxygen scavenging.c...	mercator4v6.0: catalase & prot-scriber: catala...
763	mikado.scaffold298723_le n45656_cov42_read149_ m...	Cadia_purpurea.adaxial_ side- Cadia_purpurea.abaxial_	0.00673	3.938841	up-regulated	50.1.11	Enzyme classification.EC_1 oxidoreductases.EC_...	mercator4v6.0: EC_1.11 oxidoreductase acting o...
1216	mikado.scaffold111401_le n39713_cov42_read149_ m...	(Cadia_purpurea.adaxial_ side+Cadia_purpurea.a b...	0.004856	3.987385	up-regulated	35.1	No Mercator4 annotation.other annotation avail...	mercator4v6.0: not classified & prot-scriber: ...
1061	mikado.scaffold100632_le n114711_cov41_read149_ ...	(Cadia_purpurea.adaxial_ side+Cadia_purpurea.a b...	0.001934	3.992752	up-regulated	50.3.4	Enzyme classification.EC_3 hydrolases.EC_3-4 h...	mercator4v6.0: EC_3.4 hydrolase acting on pept...
187	mikado.scaffold432839_le n39996_cov48_read149_ m...	Cadia_purpurea.adaxial_ side- Cadia_purpurea.abaxial_	0.00146	4.162151	up-regulated	35.1	No Mercator4 annotation.other annotation avail...	mercator4v6.0: not classified & prot-scriber: ...
224	mikado.scaffold448978_le n20198_cov41_read149_ m...	Cadia_purpurea.adaxial_ side- Cadia_purpurea.abaxial_	0.00162	4.163387	up-regulated	35.1	No Mercator4 annotation.other annotation avail...	mercator4v6.0: not classified & prot-scriber: ...
47	mikado.scaffold366585_le n22799_cov41_read149_ m...	Cadia_purpurea.adaxial_ side- Cadia_purpurea.abaxial_	0.000642	4.169431	up-regulated	11.5.2.4. 2	Phytohormone action.ethylene.perception and si...	mercator4v6.0: substrate(EIN3) adaptor of SCF ...
1063	mikado.scaffold261033_le n7830_cov32_read149_m a...	(Cadia_purpurea.adaxial_ side+Cadia_purpurea.a b...	0.001934	4.196632	up-regulated	50.3.2	Enzyme classification.EC_3 hydrolases.EC_3-2 g...	mercator4v6.0: EC_3.2 glycosylase & prot-scrib...
989	mikado.scaffold444070_le n17906_cov45_read149_ m...	(Cadia_purpurea.adaxial_ side+Cadia_purpurea.a b...	0.000951	4.277694	up-regulated	50.2.5	Enzyme classification.EC_2 transferases.EC_2-5...	mercator4v6.0: EC_2.5 transferase transferring...
988	mikado.scaffold444070_le n17906_cov45_read149_ m...	(Cadia_purpurea.adaxial_ side+Cadia_purpurea.a b...	0.000951	4.277694	up-regulated	7.3.1	Coenzyme metabolism.S-adenosyl methionine (SAM...	mercator4v6.0: S-adenosyl methionine synthetas...
1531	mikado.scaffold99278_le n24895_cov33_read149_m a...	(Cadia_purpurea.adaxial_ side+Cadia_purpurea.a b...	0.000999	4.284694	up-regulated	24.1.2.3. 1	Solute transport.primary active transport.P- ty...	mercator4v6.0: P3A-type proton-translocating A...
1183	mikado.scaffold148491_le n4347_cov39_read149_m a...	(Cadia_purpurea.adaxial_ side+Cadia_purpurea.a b...	0.004227	4.31435	up-regulated	50.2.3	Enzyme classification.EC_2 transferases.EC_2-3...	mercator4v6.0: EC_2.3 acyltransferase & prot-s...
719	mikado.scaffold122053_le n9699_cov37_read149_m a...	Cadia_purpurea.adaxial_ side- Cadia_purpurea.abaxial_	0.006092	4.334615	up-regulated	27.2.13.1	Multi-process regulation.Programmed Cell Death...	mercator4v6.0: programmed cell death cysteine ...
718	mikado.scaffold122053_le n9699_cov37_read149_m a...	Cadia_purpurea.adaxial_ side- Cadia_purpurea.abaxial_	0.006092	4.334615	up-regulated	19.4.1.3	Protein homeostasis.proteolysis.cysteine pepti...	mercator4v6.0: Legumain-type asparaginyl endop...
1532	mikado.scaffold261033_le n7830_cov32_read149_m a...	(Cadia_purpurea.adaxial_ side+Cadia_purpurea.a b...	0.001014	4.365622	up-regulated	50.3.2	Enzyme classification.EC_3 hydrolases.EC_3-2 g...	mercator4v6.0: EC_3.2 glycosylase & prot-scrib...

Chapter 3. Differential expression of symmetry genes in the early branching Papilionoideae

1427	mikado.scaffold41971_len29870_cov42_read149_m a...	(Cadia_purpurea.adaxial_side+Cadia_purpurea.a b...	0.000271	4.392667	up-regulated	35.1	No Mercator4 annotation.other annotation avail...	mercator4v6.0: not classified & prot-scriber: ...
1832	mikado.scaffold276935_le n14440_cov31_read149_m...	(Cadia_purpurea.adaxial_side+Cadia_purpurea.a b...	0.004593	4.483983	up-regulated	2.4.2.2	Cellular respiration.oxidative phosphorylation...	mercator4v6.0: alternative oxidase *(AOx) & pr...
1164	mikado.scaffold120190_le n9356_cov41_read149_m a...	(Cadia_purpurea.adaxial_side+Cadia_purpurea.a b...	0.003958	4.537838	up-regulated	24.1.2.3.1	Solute transport.primary active transport.P-ty...	mercator4v6.0: P3A-type proton-translocating A...
929	mikado.scaffold41971_len29870_cov42_read149_m a...	(Cadia_purpurea.adaxial_side+Cadia_purpurea.a b...	0.000363	4.647499	up-regulated	35.1	No Mercator4 annotation.other annotation avail...	mercator4v6.0: not classified & prot-scriber: ...
248	mikado.scaffold46581_len10742_cov42_read149_m a...	Cadia_purpurea.adaxial_side-Cadia_purpurea.aba...	0.001701	4.738947	up-regulated	50.2.4	Enzyme classification.EC_2 transferases.EC_2-4...	mercator4v6.0: EC_2.4 glycosyltransferase & pr...
9	mikado.scaffold495629_le n84595_cov43_read149_m...	Cadia_purpurea.adaxial_side-Cadia_purpurea.aba...	0.000272	4.772244	up-regulated	24.2.1.1.10	Solute transport.carrier-mediated transport.DM...	mercator4v6.0: nucleotide sugar transporter *(...
968	mikado.scaffold375803_le n32981_cov41_read149_m...	(Cadia_purpurea.adaxial_side+Cadia_purpurea.a b...	0.000725	4.915715	up-regulated	27.6.2.3	Multi-process regulation.phosphoinositide sign...	mercator4v6.0: phosphatidylinositol 4-kinase *...
1415	mikado.scaffold107478_le n23129_cov41_read149_m...	(Cadia_purpurea.adaxial_side+Cadia_purpurea.a b...	0.000001	4.919766	up-regulated	50.1.3	Enzyme classification.EC_1 oxidoreductases.EC_...	mercator4v6.0: EC_1.3 oxidoreductase acting on...
1414	mikado.scaffold107478_le n23129_cov41_read149_m...	(Cadia_purpurea.adaxial_side+Cadia_purpurea.a b...	0.000001	4.919766	up-regulated	11.7.1.7	Phytohormone action.jasmonic acid.biosynthesis...	mercator4v6.0: 4,5-didehydro-jasmonic acid red...
909	mikado.scaffold107478_le n23129_cov41_read149_m...	(Cadia_purpurea.adaxial_side+Cadia_purpurea.a b...	0.000003	5.123756	up-regulated	50.1.3	Enzyme classification.EC_1 oxidoreductases.EC_...	mercator4v6.0: EC_1.3 oxidoreductase acting on...
908	mikado.scaffold107478_le n23129_cov41_read149_m...	(Cadia_purpurea.adaxial_side+Cadia_purpurea.a b...	0.000003	5.123756	up-regulated	11.7.1.7	Phytohormone action.jasmonic acid.biosynthesis...	mercator4v6.0: 4,5-didehydro-jasmonic acid red...
1020	mikado.scaffold99278_len24895_cov33_read149_m a...	(Cadia_purpurea.adaxial_side+Cadia_purpurea.a b...	0.001293	5.208281	up-regulated	24.1.2.3.1	Solute transport.primary active transport.P-ty...	mercator4v6.0: P3A-type proton-translocating A...
1099	mikado.scaffold236868_le n6622_cov32_read149_m a...	(Cadia_purpurea.adaxial_side+Cadia_purpurea.a b...	0.002528	5.368804	up-regulated	27.6.2.3	Multi-process regulation.phosphoinositide sign...	mercator4v6.0: phosphatidylinositol 4-kinase *...
283	mikado.scaffold224_len19313_cov40_read149_max K...	Cadia_purpurea.adaxial_side-Cadia_purpurea.aba...	0.001832	6.448107	up-regulated	35.1	No Mercator4 annotation.other annotation avail...	mercator4v6.0: not classified & prot-scriber: ...

Table 4. Top 30 down-regulated genes in the contrast adaxial vs abaxial in *Cadia*.

	target	contrast	adj.pval	log2FC	up.down	BINCODE	NAME	DESCRIPTION
448	mikado.scaffold369312_len16002_cov42_read149_m..	Cadia_purpurea.adaxial_side- Cadia_purpurea.abaxial...	0.00323 4	- 5.55452 4	down- regulated	35.1	No Mercator4 annotation.other annotation avail...	mercator4v6.0: not classified & prot- scriber: ...
142 8	mikado.scaffold98549_len11965_cov43_read149_ma..	(Cadia_purpurea.adaxial_side+Cadia_purpurea.ab..	0.00027 1	- 5.18585 6	down- regulated	19.4.2.1.1	Protein homeostasis.proteolysis.serine peptida...	mercator4v6.0: protease *(SBT1) & prot- scriber...
552	mikado.scaffold473497_len43046_cov37_read149_m..	Cadia_purpurea.adaxial_side- Cadia_purpurea.abaxial...	0.00410 8	- 4.63111 8	down- regulated	35.1	No Mercator4 annotation.other annotation avail...	mercator4v6.0: not classified & prot- scriber: ...
151 1	mikado.scaffold494480_len27249_cov45_read149_m..	(Cadia_purpurea.adaxial_side+Cadia_purpurea.ab..	0.00093 6	- 4.60484 2	down- regulated	11.11.1.8.2	Phytohormone action.signalling peptides.NCRP (...)	mercator4v6.0: TDIF-peptide receptor *(PXY) & ...
151 3	mikado.scaffold494480_len27249_cov45_read149_m..	(Cadia_purpurea.adaxial_side+Cadia_purpurea.ab..	0.00093 6	- 4.60484 2	down- regulated	50.2.7	Enzyme classification.EC_2 transferases.EC_2-7...	mercator4v6.0: EC_2.7 transferase transferring...
151 2	mikado.scaffold494480_len27249_cov45_read149_m..	(Cadia_purpurea.adaxial_side+Cadia_purpurea.ab..	0.00093 6	- 4.60484 2	down- regulated	18.4.1.11	Protein modification.phosphorylation.TKL prote...	mercator4v6.0: LRR-XI protein kinase & prot-sc...
912	mikado.scaffold76191_len15047_cov42_read149_ma..	(Cadia_purpurea.adaxial_side+Cadia_purpurea.ab..	0.00004 5	- 4.57444 3	down- regulated	19.2.2.12.4.2. 8	Protein homeostasis.ubiquitin- proteasome syste...	mercator4v6.0: substrate adaptor *(NPY) of CUL...
107 1	mikado.scaffold99695_len56850_cov40_read149_ma..	(Cadia_purpurea.adaxial_side+Cadia_purpurea.ab..	0.00208	- 4.46079 9	down- regulated	50.2.7	Enzyme classification.EC_2 transferases.EC_2-7...	mercator4v6.0: EC_2.7 transferase transferring...
100 4	mikado.scaffold75958_len15824_cov44_read149_ma..	(Cadia_purpurea.adaxial_side+Cadia_purpurea.ab..	0.00104 2	- 4.33637	down- regulated	15.5.2.6	RNA biosynthesis.DNA-binding transcriptional r...	mercator4v6.0: GRF transcription factor & prot...
8	mikado.scaffold41971_len29870_cov42_read149_ma..	Cadia_purpurea.adaxial_side- Cadia_purpurea.abaxial...	0.00027 2	- 4.28263 8	down- regulated	35.1	No Mercator4 annotation.other annotation avail...	mercator4v6.0: not classified & prot- scriber: ...
189 6	mikado.scaffold179652_len10169_cov40_read149_m..	(Cadia_purpurea.adaxial_side+Cadia_purpurea.ab..	0.00552	- 4.21761 4	down- regulated	35.1	No Mercator4 annotation.other annotation avail...	mercator4v6.0: not classified & prot- scriber: ...
721	mikado.scaffold387996_len73276_cov42_read149_m..	Cadia_purpurea.adaxial_side- Cadia_purpurea.abaxial...	0.00609 2	- 4.14862 6	down- regulated	21.1.2.2	Cell wall organisation.cellulose.cellulose- hem...	mercator4v6.0: regulatory protein *(COB) of ce...
168 6	mikado.scaffold101309_len20616_cov40_read149_m..	(Cadia_purpurea.adaxial_side+Cadia_purpurea.ab..	0.00257 7	- 4.13857 4	down- regulated	50.2.7	Enzyme classification.EC_2 transferases.EC_2-7...	mercator4v6.0: EC_2.7 transferase transferring...
966	mikado.scaffold98549_len11965_cov43_read149_ma..	(Cadia_purpurea.adaxial_side+Cadia_purpurea.ab..	0.00071 3	- 4.10848 7	down- regulated	19.4.2.1.1	Protein homeostasis.proteolysis.serine peptida...	mercator4v6.0: protease *(SBT1) & prot- scriber...
20	mikado.scaffold267316_len72178_cov41_read149_m..	Cadia_purpurea.adaxial_side- Cadia_purpurea.abaxial...	0.00050 5	- 4.10018 4	down- regulated	50.1.1	Enzyme classification.EC_1 oxidoreductases.EC_...	mercator4v6.0: EC_1.1 oxidoreductase acting on...
911	mikado.scaffold399914_len23871_cov29_read149_m..	(Cadia_purpurea.adaxial_side+Cadia_purpurea.ab..	0.00003 4	- 4.06087 5	down- regulated	50.2.7	Enzyme classification.EC_2 transferases.EC_2-7...	mercator4v6.0: EC_2.7 transferase transferring...
108 8	mikado.scaffold494480_len27249_cov45_read149_m..	(Cadia_purpurea.adaxial_side+Cadia_purpurea.ab..	0.00244 1	- 4.06008 7	down- regulated	11.11.1.8.2	Phytohormone action.signalling peptides.NCRP (...)	mercator4v6.0: TDIF-peptide receptor *(PXY) & ...
108 9	mikado.scaffold494480_len27249_cov45_read149_m..	(Cadia_purpurea.adaxial_side+Cadia_purpurea.ab..	0.00244 1	- 4.06008 7	down- regulated	18.4.1.11	Protein modification.phosphorylation.TKL prote...	mercator4v6.0: LRR-XI protein kinase & prot-sc...

Chapter 3. Differential expression of symmetry genes in the early branching Papilionoideae

1090	mikado.scaffold494480_len27249_cov45_read149_m..	(Cadia_purpurea.adaxial_side+Cadia_purpurea.ab..	0.002441	- 4.060087	down-regulated	50.2.7	Enzyme classification.EC_2 transferases.EC_2-7...	mercator4v6.0: EC_2.7 transferase transferring...
1297	mikado.scaffold382136_len2035_cov53_read149_ma..	(Cadia_purpurea.adaxial_side+Cadia_purpurea.ab..	0.006576	- 4.042053	down-regulated	35.1	No Mercator4 annotation.other annotation avail...	mercator4v6.0: not classified & prot-scriber: ...
11	mikado.scaffold338924_len67926_cov39_read149_m..	Cadia_purpurea.adaxial_side-Cadia_purpurea.ab...	0.00029	- 4.002068	down-regulated	50.4.1	Enzyme classification.EC_4 lyases.EC_4-1 carbo...	mercator4v6.0: EC_4.1 carbon-carbon lyase & pr...
1030	mikado.scaffold399988_len12331_cov44_read149_m..	(Cadia_purpurea.adaxial_side+Cadia_purpurea.ab..	0.00146	- 3.995104	down-regulated	35.1	No Mercator4 annotation.other annotation avail...	mercator4v6.0: not classified & prot-scriber: ...
505	mikado.scaffold376093_len30337_cov33_read149_m..	Cadia_purpurea.adaxial_side-Cadia_purpurea.ab...	0.003735	- 3.991015	down-regulated	35.1	No Mercator4 annotation.other annotation avail...	mercator4v6.0: not classified & prot-scriber: ...
5	mikado.scaffold344922_len7595_cov31_read149_ma..	Cadia_purpurea.adaxial_side-Cadia_purpurea.ab...	0.000176	- 3.944744	down-regulated	35.1	No Mercator4 annotation.other annotation avail...	mercator4v6.0: not classified & prot-scriber: ...
1423	mikado.scaffold76191_len15047_cov42_read149_ma..	(Cadia_purpurea.adaxial_side+Cadia_purpurea.ab..	0.000119	- 3.896469	down-regulated	19.2.2.12.4.2.8	Protein homeostasis.ubiquitin-proteasome syste...	mercator4v6.0: substrate adaptor *(NPY) of CUL...
917	mikado.scaffold210837_len41785_cov38_read149_m..	(Cadia_purpurea.adaxial_side+Cadia_purpurea.ab..	0.000329	- 3.848992	down-regulated	50.2.7	Enzyme classification.EC_2 transferases.EC_2-7...	mercator4v6.0: EC_2.7 transferase transferring...
916	mikado.scaffold210837_len41785_cov38_read149_m..	(Cadia_purpurea.adaxial_side+Cadia_purpurea.ab..	0.000329	- 3.848992	down-regulated	18.4.1.6.1	Protein modification.phosphorylation.TKL prote...	mercator4v6.0: LRR-VI-1 protein kinase & prot-...
3	mikado.scaffold174228_len21504_cov36_read149_m..	Cadia_purpurea.adaxial_side-Cadia_purpurea.ab...	0.000176	- 3.813177	down-regulated	17.1.3.1.23	Protein biosynthesis.ribosome biogenesis.small...	mercator4v6.0: component *(uS19) of small ribo...
12	mikado.scaffold119071_len24762_cov43_read149_m..	Cadia_purpurea.adaxial_side-Cadia_purpurea.ab...	0.000298	- 3.668862	down-regulated	35.1	No Mercator4 annotation.other annotation avail...	mercator4v6.0: not classified & prot-scriber: ...
958	mikado.scaffold73045_len37447_cov43_read149_ma..	(Cadia_purpurea.adaxial_side+Cadia_purpurea.ab..	0.000651	- 3.666042	down-regulated	50.2.7	Enzyme classification.EC_2 transferases.EC_2-7...	mercator4v6.0: EC_2.7 transferase transferring...

Annex 1

Pipeline to prepare files for Mikado:

In order to run Mikado a few steps need to be carried out to prepare the files. Those steps will involve the softwares Gmap and Stringtie.

Gmap – Genomic Mapping and Alignment Program

A standalone program for mapping and aligning cDNA sequences to a genome. The resulting cDNA-genomic alignments not only reveal the intron-exon structure of the genes, but also facilitate the study of splicing mechanics and such transcript-based phenomena as alternative splicing, single nucleotide polymorphisms, and cDNA insertions and deletions.

I - Set up a conda environment for Gmap (called it “gammapping”)

conda install -c “bioconda/label/cf201901” gmap (this one has the right perl version to make gmap run)

II - Create a conda environment for Stringtie (called it “stringtie”)

```
conda install -c bioconda stringtie
```

In the gmapping environment

III - Build a genome database

gmap build -d genome (genome fasta file)

```
gmap build -d genome “/home/gramos/scratch/wgs_novogene/genomes/cyathostegia_genome.fasta”
```

note: genome fasta file means reference genome

IV - Run gmap to generate a gff3 file

```
gmap -d genome -f 2 “/home/gramos/scratch/wgs_novogene/transcriptomes/cyathostegia_trinity.fasta” > gff3_gene.gff
```

note: although the module is called genome, the input here is the transcriptome assembly file from trinity output. This line will produce gff3 output files.

A bash script might look like:

```
#!/bin/sh
#SBATCH --partition=medium
#SBATCH --cpus-per-task=8
#SBATCH --mem=10G
#sbatch --wrap
gmap_build" -d genome "/home/gramos/scratch/wgs_novogene/genomes/cyathostegia_genome.fasta"
gmap -d genome -f 2 "/home/gramos/scratch/wgs_novogene/transcriptomes/cyathostegia_trinity.fasta" >
gff3_gene.gff
```

V – Stringtie

Stringtie takes as input spliced alignments in coordinate-sorted SAM/BAM/CRAM format and produces a GTF output which consists of assembled transcript structures and their estimated expression levels (FPKM/TPM) and base coverage values).

To prepare the GTF files from the merged and sorted bam outputs:

```
Stringtie "/home/gramos/scratch/RNA-seq_novogene/Processing/Aligned_sortedByCoord.out.bam" -G
"/home/gramos/scratch/wgs_novogene/transcriptomes/gff3_gene.gff" -o finaloutput.gtf
```

Once the **.gff3** and the **.gtf** files are generated, we can run **Mikado**.

I - Input file preparation

Create a text file = list.txt

Enter the input files (should be started at the top line of the .txt file)

gff3_gene.gff	tri	True	0	False	True	False	False
cy1_stringtie.gtf	str	True	0	False	True	False	False

II - Mikado configuration

Run the following command:

```
#!/bin/sh
#SBATCH --partition=medium
#SBATCH --cpus-per-task=8
#SBATCH --mem=10G
#sbatch --wrap

mikado configure --list mikado.txt --reference cyathostegia_genome.fasta --mode permissive --scoring plant.yaml
--copy-scoring plant.yaml --junctions junctions.bed -bt uniprot_sprot_plants.fasta configuration.yaml
```

mikado.txt: pre prepared input list file

cyathostegia_genome.fasta: reference genome file

plant.yaml: configuration file from the Mikado installed directory (generally in the apps folder where conda install softwares)

uniprot_sprot_plants.fasta: protein database

This script will create 2 outputs:

1. **plant.yaml**
2. **configuration.yaml**

III - Prepare Mikado

```
mikado prepare --json-conf configuration.yaml
```

This will create the following outputs:

1. **cyathostegia.fasta.fai**
2. **mikado_prepared.fasta**
3. **mikado_prepared.gtf**
4. **prepare.log**

IV - Blast of candidate transcripts

Install blast separately before running this part. Three commands will be run:

- 1- Uncompress the SwissProt database

```
gzip -dc uniprot_sprot_plants.fasta.gz > uniprot_sprot_plants.fasta
```

output: uniprot_sprot_plants.fasta

2- Prepare the database for BLAST

```
makeblastdb -in uniprot_sprot_plants.fasta -dbtype prot -parse_seqids > blast_prepare.log
```

output = blast_prepare.log

uniprot_sprot_plants.fasta.pdb

uniprot_sprot_plants.fasta.phr

uniprot_sprot_plants.fasta.pin

uniprot_sprot_plants.fasta.pog

uniprot_sprot_plants.fasta.pos

uniprot_sprot_plants.fasta.pot

uniprot_sprot_plants.fasta.ptf

uniprot_sprot_plants.fasta.pto

3- Execute BLAST:

```
blastx -max_target_seqs 5 -outfmt "6 qseqid sseqid  
pident length mismatch gapopen qstart qend sstart  
send eval evalue bitscore ppos btop" -num_threads 2  
-query mikado_prepared.fasta  
-db uniprot_sprot_plants.fasta  
-out mikado_prepared.blast.tsv
```

Output: mikado_prepared.blast.tsv

V- ORF calculation for the transcripts

The line below is for bacterial genomes

```
prodigal -i mikado_prepared.fasta -g 1 -o mikado.orfs.gff3 -f gff
```

Output = mikado.orfs.gff3

The line below is for eukaryotic genomes

Step 1.

```
TransDecoder.LongOrfs -t mikado_prepared.fasta
```

Output: created in 2 folders:

mikado_prepared.fasta.transdecoder_dir

mikado_prepared.fasta.transdecoder_dir._checkpoints_longorfs/

separate output file [piperliner.4714.cmds](#)

Step 2.

```
Transdecoder.Predict -t mikado_prepared.fasta
```

Output: created in 2 folders

Additional files were created in mikado_prepared.fasta.transdecoder_dir and in mikado_prepared.fasta.transdecoder_dir._checkpoints

Final outputs: mikado_prepared.fasta.transdecoder

- mikado_prepared.fasta.transdecoder.bed
- mikado_prepared.fasta.transdecoder.cds
- mikado_prepared.fasta.transdecoder.gff3
- mikado_prepared.fasta.transdecode.pep

pipeliner.146007.cmds

rename: mikado_prepared.fasta.transdecoder.bed = junctions.bed

move: longest_orfs.cds.best_candidate.gff3.revised_starts.gff3 to ped
and rename=mikado.orfs.gff3

VI- Mikado serialise

```
mikado serialise --json-conf configuration.yaml  
--xml mikado_prepared.blast.tsv --orfs mikado.orfs.gff3  
--blast_targets uniprot_sprot_plants.fasta  
--junctions junctions.bed
```

Outputs: mikado.db (take it from installed directory)

VII- Mikado pick

```
mikado pick --configuration configuration.yaml --subloci-out  
mikado.sunloci.gff3
```

Copy the mikado.db from installed directory to working directory

```
cp "/mnt/shared/scratch/gramos/apps/conda/lib/python3.9/site-  
packages/Mikado/tests/mikado.db" /mnt/shared/scratch/gramos/mikado2/
```

Outputs: mikado.loci.gff3
 mikado.loci.metrics.tsv
 mikado.loci.scores.tsv
 pick.log
 mikado.subloci.scores.tsv
 mikado.subloci.metrics.tsv
 mikado.subloci.gff3

Chapter 4

An ontogenetic overview of the florally diverse
Papilionoideae (Leguminosae): new insights into the
evolution of androecium symmetry and petal heterochrony

Introduction

4.1 What development tells us about floral differences in legumes

Comparative developmental or ontogenetic studies have been fundamental to our understanding of the evolution of the remarkable diversity of floral architecture in tropical angiosperms (Endress 1994; Marazzi & Endress 2008), as well as for revealing the homology among florally distinctive lineages that have often appeared in phylogenetically surprising positions (Tucker 1994, 2002, 2004; Bello et al. 2007; Rudall et al. 2007; Prenner & Klitgaard 2008; Bruneau et al. 2014). Conducting comprehensive floral ontogenetic investigations for undersampled or inadequately studied taxa, along with reexamining published data, is both timely and especially relevant. This effort is particularly crucial for the economically, ecologically, and evolutionarily successful Papilionoideae genera of the Leguminosae family.

Enhanced sampling across clades has significantly improved the reconstruction of the phylogeny within the species-rich subfamily Papilionoideae. This progress is evident through the integration of data from a limited number of plastid loci (e.g., Pennington et al. 2001; Wojciechowski et al. 2004; Cardoso et al. 2012, 2013, 2015) and the comprehensive exploration of both nuclear and plastid phylogenomics (Zhao et al. 2021; Choi et al. 2022). These new phylogenetic data have facilitated an evolutionary comparative examination of morphological characters, yielding noteworthy insights into the evolution of ecologically important features such as extrafloral nectaries (Marazzi et al. 2012) and the evolutionary persistence of legume–rhizobia symbiosis (Werner et al. 2015).

Currently encompassing approximately 14,000 species and 500 genera across its pantropical distribution, and notably in the tropics (Lewis et al. 2005; Legume Data Portal 2023), the Papilionoideae legume subfamily showcases a wide spectrum of floral morphologies. They stand out as a prime example of morphological diversity among genera, characterised by a wide variation in floral architecture (Pennington et al. 2000; Cardoso et al. 2013). Although there is wide variation in floral morphology, the typical zygomorphic papilionate flower is the signature of the majority of the genera of the Papilionoideae (LPWG, 2017). The specialisation of the corolla with an adaxial petal (banner), two lateral petals (wings) and two ventral petals (keel) distinguishes the Papilionoideae from the other five subfamilies. Internal organs are also strongly bilateral; the androecium is diadelphous, generally tilted towards the adaxial side of the flower, with nine stamens partially united into

a proximal sheath, free halfway towards the tips and a single free stamen on the adaxial side of the flower. Similar to the androecium, a generally single carpel, borne on a stipe, also tends to be curved towards the banner petal (Endress, 1996; Tucker, 2003). Although terms largely differ when referring to a particular side of the flower in Papilionoideae, the literature more often uses the orientation of the floral organs, where the vexillum or standard petal is on the adaxial side, as are the vexillary stamen and carpellary cleft. In the papilionate flowers this is readily distinguishable, but in papilionoid genera with unusual morphologies such as radially-symmetrical-flowered genera (see Chapter One), with radially arranged internal organs, the position of the carpellary cleft commonly points toward the adaxial side (standard) of the flower (Fig. 1). The highly specialised papilionate architecture is thought to have allowed the success of legumes via pollinator specialisation and subsequent speciation (Cubas 2004; Sargent 2004; Zhao et al. 2019). Variation in the basic plan of a papilionate flower is observed to different degrees, either in micro structures of the specialised petals, such as auricles, transverse ridges, folds, pegs, and corresponding pits that interlock petals together (Tucker 2003). Other cases involve more dramatic changes, resulting in unusual floral morphologies, such as radially symmetric flowers (Mansano et al. 2002; Tucker 2002; Cardoso et al. 2012b; Klitgaard et al. 1999, 2013; Bento et al. 2021), reduced corolla to a single petal (Tucker 1990, 1999, 2003; Leite et al. 2015; Paulino et al. 2013) and absence of petals altogether (Tucker et al. 2003; Torke & Mansano 2009; Sinjushin 2018).

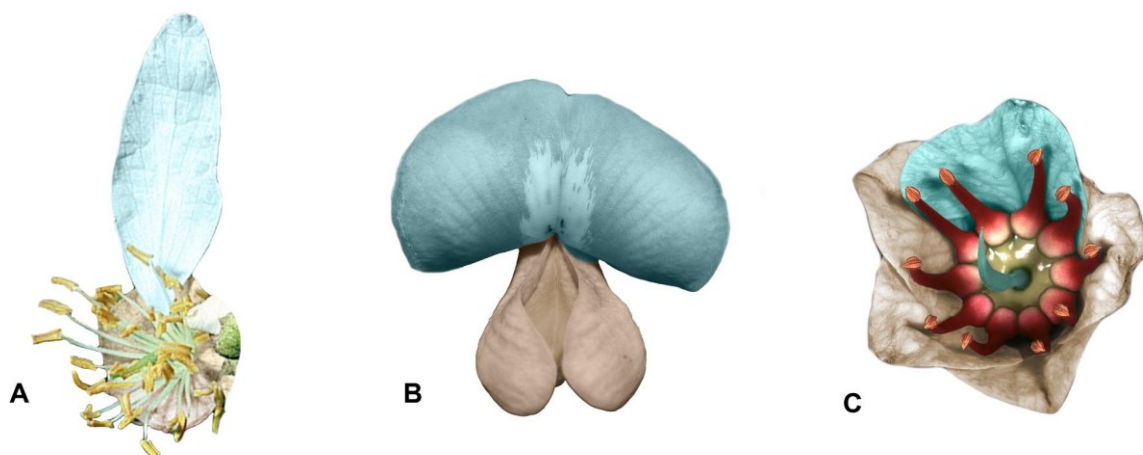


Figure 1. Abaxial and adaxial domains in three florally distinct papilionoid legume species. The adaxial domain is shown in blue and abaxial domain in sepia. A, *Cyathostegia mathewsii*; B, *Hymenolobium alagoanum*; C, *Cadia purpurea*.

The wealth of floral ontogenetic data available for papilionoid legumes (e.g., Tucker 1987, 1988, 2003; Mansano et al. 2002; Prenner 2004a, b), including recent results for phylogenetically enigmatic lineages that branch off at deep nodes (Leite et al. 2014; Prenner et al. 2015; Leite et al. 2015; Sinjushin 2018), are likewise extremely useful for large-scale comparison and as sources for potential synapomorphic characters of still obscure relationships. For example, Paulino et al. (2014) performed a comparative analysis of the unusual polycarpellate gynoecium in legumes and found that it evolved seven times, including in the early branching papilionoid genera *Aldina* and *Swarztia*. Prenner & Cardoso (2017) presented a comprehensive developmental series of the Rosaceae-like flowers found in the monotypic Brazilian Detarioideae legume genus *Goniorrhachis*, together with an ancestral character estimation of the gynoecium evolution. They found distinctive evolutionary patterns of style bending, but more remarkable was that the young style in all Detarioid legumes bends in the abaxial direction, and therefore it was a newly described synapomorphy for the subfamily.

Two ontogenetically interesting aspects of the evolution and diversification of the morphologically variable flowers of the Papilionoideae that still remain underexplored are the androecium symmetry and petal growth with respect to other organs (or petal heterochrony). Prenner (2004a) meticulously examined androecial symmetry in 91 species across 27 tribes of the Papilionoideae, from published literature on floral development and newly generated data by scanning electron microscopy (SEM) and studying dried herbarium specimens. He found that the initiation of the inner stamen whorl displayed asymmetry in 39 species spanning 11 tribes, contrasting with 52 species from 19 tribes where the androecium is initiated symmetrically, with the adaxial stamen lying precisely in the median plane. Notably, both symmetric and asymmetric androecia were observed in early branching papilionoid lineages. In contrast, taxa belonging to the predominantly derived Millettieae-Phaseoleae and Hologalegina exhibited exclusively symmetric androecia. Prenner (2004a) also highlighted that the occurrence of both androecium types was independent of staminal fusion. He put forward two hypotheses regarding the evolution of asymmetric androecia: (i) symmetric androecia are derived within Papilionoideae, and/or (ii) asymmetry arose independently several times.

The variation observed in developmental rate of floral organs and tissues has been widely described as heterochrony (Raff & Wray 1989; Lord 1991; Endress 1994). Comparative

floral ontogeny in distinct yet closely related taxa have revealed that variation in flower architecture can be elucidated by heterochronies or the changes in the rates of organ initiation and development (Endress 1994; Tucker 1999, 2003; Prenner 2004b; Vasconcelos et al. 2018). Perhaps the most well known examples of dramatic changes in floral architecture since early ontogeny are found in flowers exhibiting bilateral or zygomorphic characteristics, but that commence as "regular" and shift to irregular at a subsequent stage. Numerous families with taxa showcasing such a radial organ initiation but maturing into zygomorphic flowers because of heterochronic organ development include Bignoniaceae, Labiatae, Polygalaceae, Lecythidaceae, and Ranunculaceae, as well as Leguminosae (Endress 1996; Tucker 1999). Because such alterations in developmental rhythms promote differential investments in organs involved in adaptive features for plant reproduction (Li & Johnston 2000), heterochronic processes are suggested to promote flexibility in floral strategies (Vasconcelos et al. 2017) and may lead to recurrent evolutionary switches in pollination biology, allowing a lineage to successfully explore new ecological opportunities (Lovo et al. 2021).

While substantial strides have been taken in elucidating the origin and evolution of the remarkable floral diversity within the Papilionoideae, this study revisits their floral ontogeny, with particular focus on describing evolutionary switches and conservatism of petal growth and androecium symmetry. It briefly describes the development of the petal and androecium whorls for several previously unstudied, morphologically and evolutionarily key genera of papilionoid legumes. Additionally, special attention is given to reconstructing the ancestral evolution of petal heterochrony and androecium symmetry.

4.2 Premises and Hypotheses of this study

In this chapter, I review the floral ontogeny of papilionoid legume species with contrasting morphologies, with focus mostly on describing the evolution of heterochronical patterns and androecium symmetry. The aim of this study is to revisit taxa of unusual flower morphologies across the subfamily and shed light onto the ontogenetic pathway of these lineages in order to gain insights on the evolutionary significance of morphological changes. Ontogenetic data from previously published studies (e.g., on the radially-symmetrical-flowered *Cadia purpurea* and the single-petaled *Cyathostegia mathewsii*, which are subject of Chapters Two and Three) are used together with newly obtained, unpublished data generated by Dr. Gerhard Prenner and Dr. Domingos Cardoso for 17 papilionoid genera, including the single-petaled *Trischidium molle* and *Ateleia guaraya* that are closely related to *Cyathostegia*

mathewsii (Cardoso et al. 2013), and the papilionate-flowered *Andira ormosioides* and *Hymenolobium janeirense* that are more closely related to the radially-symmetrical-flowered *Aldina* (Ramos et al. 2016). *Cadia* and *Cyathostegia* are also of special interest in this study since genomic and transcriptomic data has been generated for these species in Chapters Two and Three, where I identify and quantify the expression of candidate genes regulating flower development. The specific objectives of this study are threefold: (i) to provide an in-depth ontogenetic analysis of petal and androecium development; (ii) to compare ontogenetic stages with related Papilionoideae species and, more broadly, with radially symmetrical taxa from other legume lineages; (iii) to identify potential ontogenetic synapomorphies for the morphologically intricate early branching clades of the Papilionoideae. To achieve these goals, a maximum likelihood approach was employed over a comprehensively-sampled plastome phylogeny of the Papilionoideae to estimate ancestral character states of two selected floral developmental traits. I specifically address the hypotheses that symmetric androecia are derived within Papilionoideae and asymmetric androecia evolved independently several times, and that ontogenetic stability of slow petal growth and symmetrical androecium led to the evolutionary specialisation and ecological success of the papilionate floral architecture in the largest clade of Papilionoideae.

Material and Methods

4.3 Target species

To develop fuller understanding of the evolution of ontogenetic development of androecium symmetry and petal growth across the papilionoid legumes, several morphologically and evolutionarily key genera for which ontogenetic data were lacking or still incomplete were newly investigated: *Aldina latifolia*, *Andira ormosioides*, *Ateleia guaraya*, *Camoensia scandens*, *Cleobulia multiflora*, *Dipteryx rosea*, *Eysenhardtia orthocarpa*, *Harpalyce brasiliensis*, *Hymenolobium janeirense*, *Leptolobium bijugum*, *Luetzelburgia andradelima*, *Monopteryx uauacu*, *Olneya tesota*, *Poecilanthe itapuana*, *Pterodon abruptus*, *Trischium molle* and *Uleanthus erythrinoides*. Flowers and flower buds of these taxa were collected and immediately fixed and stored in 70% ethanol. These samples came from extensive field expeditions across most Brazilian phytogeographic domains, including Amazonia, Atlantic Forest, Caatinga, and Cerrado. All specimens were vouchered (Table 1)

and are accessible at ALCB and/or HUEFS herbaria (acronyms according to Thiers et al. 2023). The flower samples preserved in 70% ethanol, which were left after dissection for ontogenetic analysis, were also deposited in the spirit collection of the Royal Botanic Gardens, Kew.

The abovementioned genera were targeted in this study so as to fill in gaps of ontogenetic data across as much as possible of the florally diverse early branching papilionoid legumes (Cardoso et al. 2013). The ample variation in their floral architecture spans from truly papilionate flowers with zygomorphic symmetry (e.g., *Andira*) to non-papilionate flowers with both zygomorphic symmetry, involving petal loss (e.g., *Trischidium*) or clear differentiation of the standard petal (e.g., *Camoensia*), and radial symmetry, which is mostly marked by undifferentiation of all petals (e.g., *Aldina*, *Leptolobium*).

4.4 Scanning electron microscopy analysis

In preparation for scanning electron microscopy (SEM), the same methods as described in Prenner et al. (2015) and Prenner & Cardoso (2017) were used. As such, flowers and floral buds underwent dissection in 70% ethanol, followed by dehydration to 100% ethanol and critical-point drying using an Autosamdri-815B critical-point dryer. Subsequently, the dried specimens were affixed to Aluminum stubs using transparent nail polish and coated with platinum using a Quorum Q150T sputter coater. High-resolution SEM images were captured using a Hitachi S-4700-II cold-field emission SEM, and the acquired images were post-processed using Adobe Photoshop 2023 v. 25.0.

4.5 Ancestral character estimation

To investigate the evolution of the androecium symmetry and petal heterochrony in the context of the major papilionoid legume lineages, I comprehensively reviewed published ontogenetic data. In addition I studied some key taxa, leading to new information for 17 genera, most of which from early branching clades of the Papilionoideae (Table 1). Based on these observations, I have defined in the papilionoid flowers three different patterns related to androecium symmetry: symmetrical (sy), asymmetrical (as) and undefined (un), which correspond to flowers with more than ten stamens, sometimes even a profusion of stamens such as in *Aldina* and *Swarztia* (Table 1). Petal heterochrony was defined as having slow growth (sg), fast (fg) and a third state (ab) for flowers with a reduced or absent corolla (Table 1).

Because our ancestral-state estimations were made at genus level (i.e. terminals in the phylogeny correspond to a single representative of each genus), I scored the states for each genus only when this was known to occur in at least one of the species studied ontogenetically. However, I believe this is not a problem since the traits investigated here are largely conserved at genus level (Pennington et al. 2000).

A maximum likelihood approach was employed, implemented in the R package *corHMM* version 2.3 (Beaulieu et al. 2013; Beaulieu et al. 2020), to estimate the ancestral states of each ontogenetic trait. The hidden rates model (HRM) was selected using the 'corHMM' function, as it accommodates variable transition rates between states along different branches of the tree (Beaulieu et al. 2013). This choice is particularly crucial in current macroevolutionary reconstructions, given the significant variation in diversification patterns observed across clades of the Papilionoideae (e.g., Lavin et al. 2005; Koenen et al. 2013). Within the HRM, the distinct rate classes are considered as 'hidden' states during the estimation of ancestral character states.

Ancestral estimations were made on a maximum likelihood phylogenomic tree of plastome data (Choi et al. 2022). This plastome phylogeny was selected because it combines a large sample of Papilionoideae genera as well as excellent resolution of the major clades, the latter having also been largely confirmed by phylogenomic analyses of nuclear genes (Zhao et al. 2021). For the ancestral-state estimations, terminal genus taxa of the original tree of Choi et al. (2022) lacking information on floral ontogeny were pruned from the tree using a custom R function *phylotraitmatch* (Cardoso, unpubl. data), which also incorporates the *drop.tip* function from the *treeio* package (Wang et al. 2020). All figures displaying the ancestral state estimations were plotted using functions from the R packages *ape* (Paradis & Schliep 2019) and *phytools* (Revell 2012).

Results

4.6 Ontogenetic development of androecium and petal growth in selected papilionoid taxa

Our review of published ontogenetic studies, together with the new data described here for the first time, reports the development of androecium and petal growth in 132 species from across all main clades of papilionoid legumes (Table 1).

Fast petal growth was found to occur in 15 genera: *Trischidium*, *Ateleia*, *Cyathostegia*, *Swartzia*, *Bobgunnia*, *Angylocalyx*, *Castanospermum*, *Monopteryx*, *Dussia*, *Petaladenium*, *Myroxylon*, *Myrocarpus*, *Amburana*, *Harleyodendron*, *Machaerium*, *Riedeliella* and *Pterocarpus*. Corolla development in species with this type of growth is characterised by an earlier development of petal primordia relative to the androecium inception. In pentamerous species, like *Monopteryx uauacu* (Fig. 2P-T), five primordia are observed where the adaxial, lateral, and ventral petals will develop from. Stamens are only noticeable after a short enlargement phase of petals, where the two ventral petals surpass the lateral and dorsal ones in size. At this point, the first abaxial stamens primordia are initiated and the dorsal ones only initiate later when petals have considerably enlarged and start to differentiate. In *M. uauacu*, stamens never outgrow petals and this is observed throughout the complete developmental series.

Although phylogenetically closely related to *Monopteryx*, *Dipteryx*, *Pterodon* and *Taralea* have slow petal growth (Table 1). This pattern is evidenced in the new data brought here where we can see simultaneous growth of the five petal primordia with the first whorl of stamens in *Dipteryx rosea* (Fig. 2U-Y). The petals in this species undergo a phase of slow development where only the banner petal enlarges and stamens readily outgrow the remaining petals. The fast petal growth of *Pterodon abruptus* is, however, somewhat slower than in *M. uauacu* (Fig. 2Z-AD). This can be interpreted as an intermediate growth pattern, where petals are slightly bigger in size than stamens during early organ development.

In the single-petaled *Trischidium molle* and *Ateleia guaraya* (Fig. 2F-J, K-O), a single petal primordium is observed from the very early developmental stages in the floral meristem. In these species, the androecium inception only starts after the initial development of a ring meristem, from which stamens primordia develop. Thus, petal growth is fast and androecium never surpasses the corolla in size. Although phylogenetically close and morphologically similar to *Trischidium* and *Ateleia*, petal growth in *Cyathostegia mathewsii* is slow (Fig. 2 A-E). Initially, the single petal primordium develops at a similar pace as the androecium in the early developmental stages, but stamens readily outgrow the single petal, which only resumes

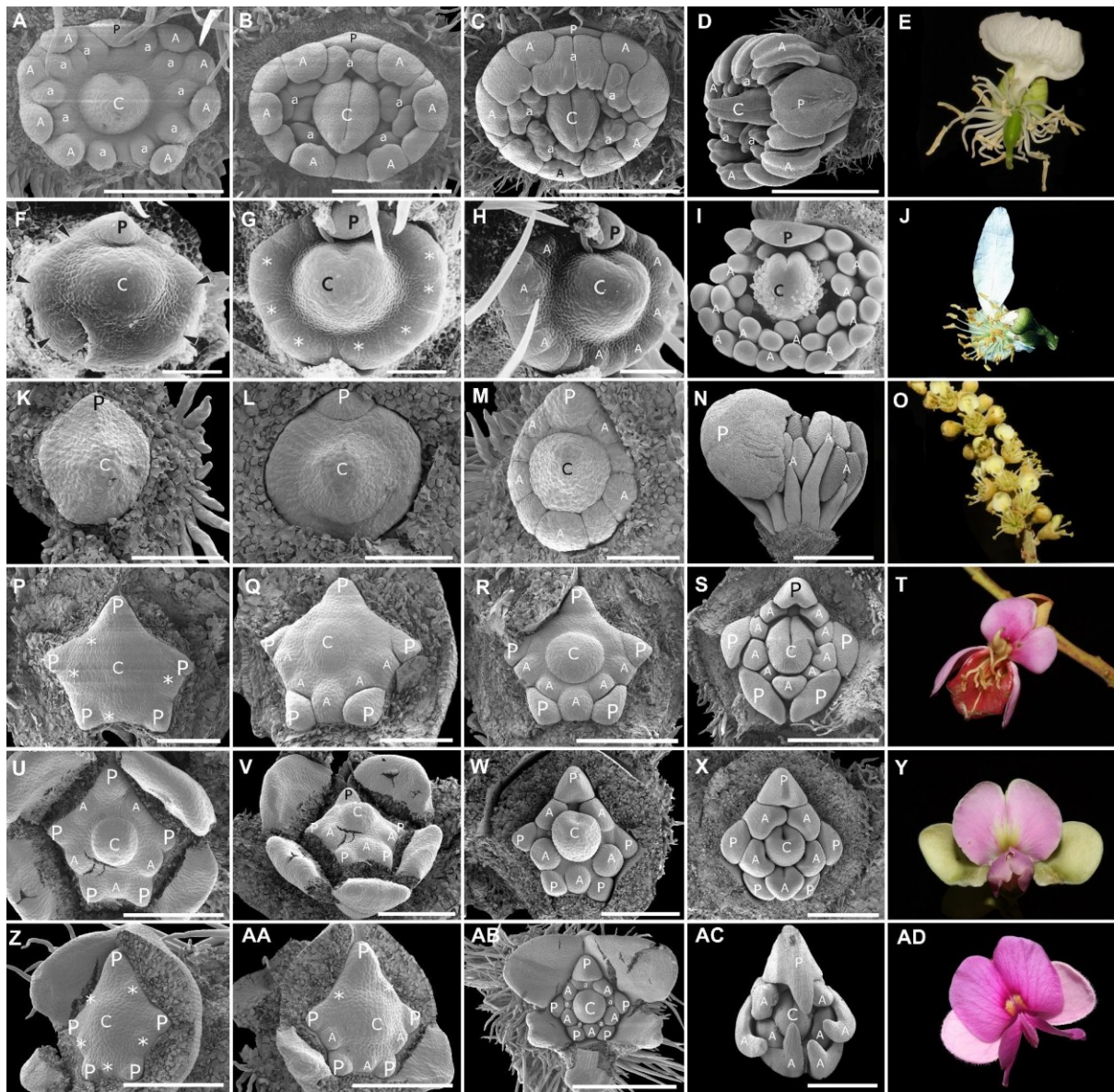


Figure 2. Androecium and petal development, and flower architecture in *Trischidium molle*, *Cyathostegia mathewsii*, *Ateleia guaraya*, *Monopteryx uauco*, *Dipteryx rosea* and *Pterodon abruptus*. A–E, *Trischidium molle*. A, Young flower bud with two staminal whorls growing at different paces and young carpellary cleft points to the single petal. It is not possible to establish a clear symmetry plan in androecium growth. B–D, Stamens outgrow petal development until petal growth resumes in later stages. F–J, *Cyathostegia mathewsii*. F, Fast petal growth relative to stamens whorl in the developing floral meristem with early evidence of a ring meristem (arrows). G, Androecium starts differentiation in the ring meristem with individual growth of stamens primordia (asterisk). H–I, Petal maintains faster development until the final developmental stages. K–O, *Ateleia guaraya*. K, A single petal primordium is seen in early floral meristem along with a developed carpel primordium. L, Petal primordium shows steady growth. M, Androecium primordium starts a late growth relative to the petal and an ordered symmetrical arrangement of the stamens is seen. N, Stamens never outgrow petal development. P–T, *Monopteryx uauco*. P, Five clear petal primordia can be seen and individual stamens primordia are noticeable (asterisks) since early developmental stages. Q, Abaxial flower region grows faster than adaxial, evidenced by enlargement of abaxial petal primordia and stamens. R, Corolla and androecium

Chapter 4. An ontogenetic overview of the florally diverse Papilionoideae (Leguminosae): new insights into the evolution of androecium symmetry and petal heterochrony

maintain initial growth difference throughout final stages with petal growing at a faster pace. S, Although displaying asymmetry in the early stages, the androecium is fully symmetrical from mid-development. U–Y, *Dipteryx rosea*. U, Five petals primordia can be seen along with slightly smaller stamens. V, Androecium catches up with petal growth and presents a symmetrical arrangement from this stage. W, Androecium surpass petal growth, excepting the adaxial petal, and stamens arrangement is fully symmetrical at this stage. X, Androecium and corolla undergo an enlargement phase, but stamens maintain a faster growth. Z–AD, *Pterodon abruptus*. Z, Five clear petal primordia are seen in early development with indication of an androecium meristematic activity (asterisks). AA, Petal primordia growth is faster relative to androecium, which is now more evident in the abaxial side of the floral meristem. AB, Androecium reaches the same size of petal, except for the banner. Androecium displays a symmetrical arrangement. AC, Petal growth overcome androecium development in the later stages of development. A, antesealous stamens; a, antepetalous stamens; C, carpel; P, petal; Scale bar = 500 μm in C. 400 μm in AB. 300 μm in S, X, AC. 200 μm in A–B, R, U–W, Z, AA. 100 μm in K–M, I, P, Q. 50 μm in F–H. 1 mm in D, I. All photos displaying floral architectures by Domingos Cardoso, except for *Cyathostegia* (J), by Gwilym Lewis.

growth after a long interval where stamens have reached a longer length and anthers are differentiated.

The radially-symmetrical-flowered *Leptolobium bijugum*, *Cadia purpurea* and *Eysenhardtia texana* show varied types of petal growth (Fig. 4). Among these species, only *L. bijugum* has fast petal growth, which can be observed from early developmental stages where petal primordia develop at a faster pace than the stamen primordia (Fig. 4A–E). Although *C. purpurea* presents well developed petal primordia in early stages, androecium development takes place quickly with the initiation of both antesealous and antepetalous stamens. At this point petals and stamens grow more or less at the same pace, but stamens differentiate quickly and anthers are readily distinguishable (Fig. 4K–O). Given the undifferentiated nature of *Cadia* petals, it is hard to accurately pinpoint how delayed the corolla develops relative to the androecium other than by using organ size comparison. *Eysenhardtia texana* shows evidence of slow petal growth from the very early developmental stages, where stamens primordia readily outgrow petals, which only take over growth in the final developmental stages. The second whorl of stamens do not grow as fast as the first whorl, being slightly smaller than petals; at this stage though, anthers start to differentiate, evidencing an advanced stage of development.

Slow petal growth is predominant in all genera from within the Non-Protein Amino Acid-Accumulating (NPAAA) clade (Table 1). For example, the new data shown here for *Cleobulia multiflora* and *Olneya tesota* clearly have developed petal primordia at early

developmental stages, but a first whorl of stamens readily takes over growth and surpasses petals in size (Fig. 4U-Y, Z-AD). Cases of reduction or loss of corolla were also reported but rare, as exemplified by the genera *Mildbraediendron*, *Cordyla* and *Parryella* (Table 1).

I also found that both symmetric and asymmetric androecia occur across all papilionoid legume genera irrespective of their floral architecture being papilionate or non-papilionate (Table 1). A symmetrically-growing androecium is the more common pattern, particularly across the genera with typically zygomorphic papilionate flowers from within the NPAAA clade, such as *Cleobulia* and *Olneya* (Fig. 4, U-Y, Z-AD).

Although phylogenetically closely related, the genera *Monopteryx*, *Dipteryx*, *Pterodon*, and *Taralea* presented divergent types of androecium symmetry. The early developmental stages of the stamens primordia in *Monopteryx* show an asymmetrical pattern given that the androecium initiates abaxially in the floral meristem (Fig. 2P-R). Towards more advanced stages, a direction to a symmetrical arrangement is observed and at later stages stamens are all arranged in a symmetrical position relative to the gynoecium and corolla (Fig. 2S). Androecium growth initiates asymmetrically in *Dipteryx* and the arrangement of stamens continues to grow in this fashion until the final stages (Fig. 2U-Y). The asymmetry is observed in the lack of an adaxial antesealous stamen (Fig. 2X). *Pterodon* androecium asymmetry is similar to *Dipteryx* with the adaxial antepetalous stamen primordium growing on the left side of the remaining adaxial meristematic tissue (Fig. 2AB).

Other morphologically heterogeneous groups also differ in androecium growth. That is the case of the *Andira* clade, comprised by the typically-zygomorphic-flowered *Andira* and *Hymenolobium*, and the radially-symmetrical-flowered *Aldina*. Undefined androecium symmetry is found in *Aldina* given the numerous stamens found in all species in the genus. *Andira* presents a symmetrical androecium which is not fully visible at the early developmental stages for the lack of the uppermost stamen in the adaxial region. In later stages all stamens primordia are developed and arranged in a clear symmetrical pattern, evidenced by the position of the carpellary cleft (Fig. 3, F-J). *Hymenolobium*, on the other hand, presents an asymmetrical androecium symmetry which is only visible with a developed carpellary cleft, whose groove is not aligned with the abaxial and adaxial stamens (Fig. 3K-O).

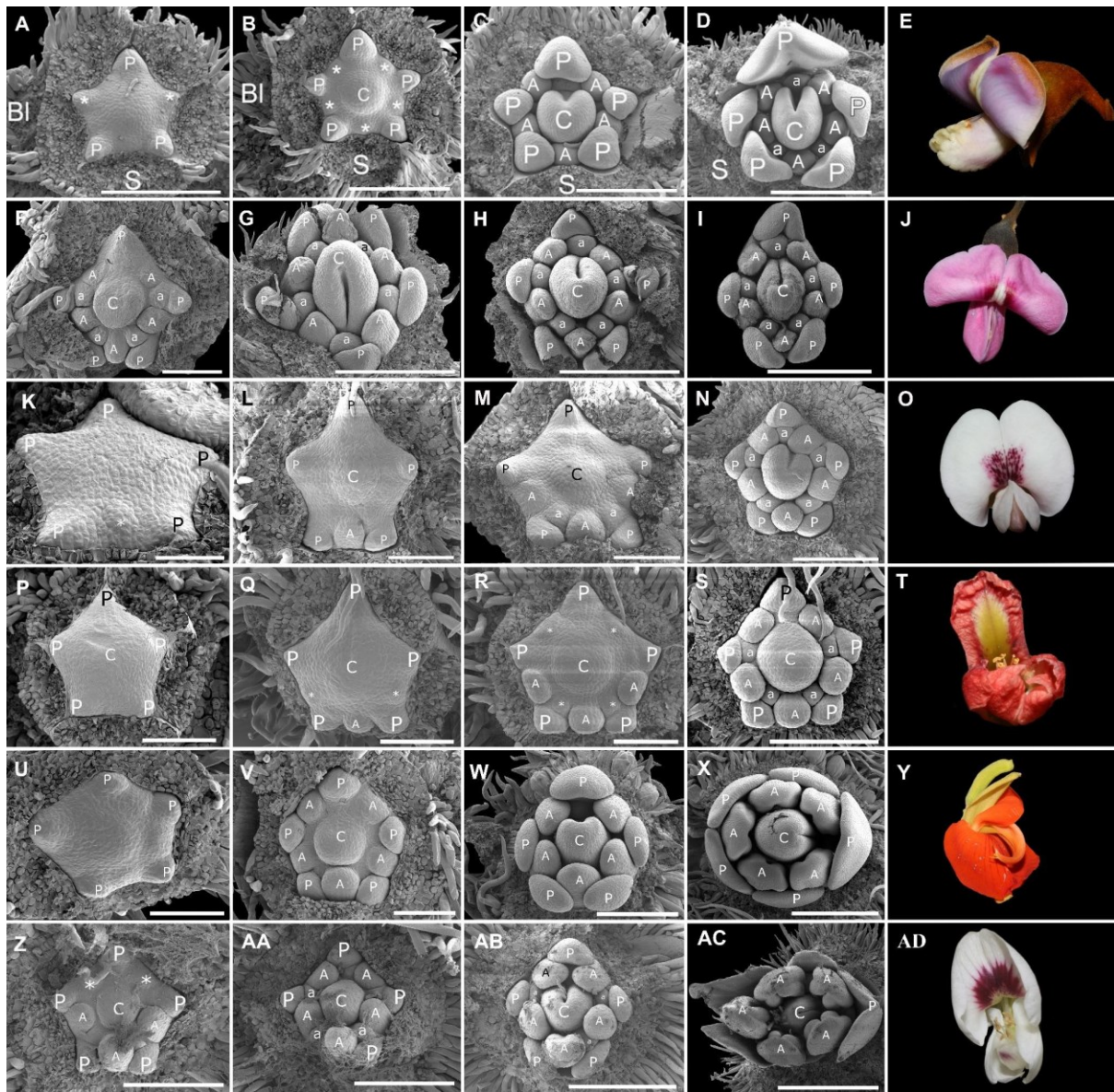


Figure 3. Androecium and petal development, and flower architecture in *Petaladenium urceoliferum*, *Andira humilis*, *Hymenolobium janeirense*, *Luetzelburgia andrade-limae*, *Harpalyce brasiliiana* and *Poecilanthe itapuana*. Bracteoles and sepals are removed. A–E, *Petaladenium urceoliferum*. A, Early development of corolla with the two abaxial petals and standard more developed than lateral petals (asterisks). B, Petals develop steadily and stamen primordia are initiated with a symmetrical arrangement. C, Petals undergo an enlargement phase whilst stamens grow at a slower pace. D, Petal growth is not surpassed by androecium development, which symmetry is maintained throughout the final stages of growth. F–J, *Andira humilis*. F, Petal and stamen primordia develop at a similar pace at this stage with abaxial organs being more developed. G, Petals enlarge whilst stamens grow in length. The androecium assumes a symmetrical arrangement at this stage. H–I, Petals and carpel undergo an enlargement phase while stamens grow in length surpassing the petal at the final stages. The

Chapter 4. An ontogenetic overview of the florally diverse Papilionoideae (Leguminosae): new insights into the evolution of androecium symmetry and petal heterochrony

symmetrical arrangement of stamens is maintained. K–O, *Hymenolobium janeirense*. K–L, early development of petals and stamens primordia. M, More developed stamens are found abaxially growing faster than petal development. N, Petals growth is slower relative to stamens from early stages and at this point only grow by enlargement. The symmetrical androecium is clear at this stage. P–T, *Luetzelburgia andrade-limae*. P–Q, Petals and stamens primordia grow at a somewhat similar pace at early stages following the pattern of more developed organs on the abaxial side of the floral meristem. R, Abaxial stamens quickly outgrow the petal primordia. S, The symmetrical androecium is more evident and petals keep a slow growth at this stage. U–Y, *Harpalyce brasiliensis*. U, Petal primordia are visible in early development and no evidence of stamen growth is seen. V–W, Once stamens are initiated, they rapidly outgrow the petals. The symmetry of the androecium is visible from early developmental stages. X, Petal growth is visibly slower relative to stamens at this stage except for the standard which has enlarged and grown in length. Z–AD, *Poecilanthe itapuana*. Z, Slow petal growth with stamens developing at similar pace at early stages. The symmetrical arrangement of stamens is already visible from this stage. AA–AB, Stamens primordia grow in length rapidly outgrowing the petals, which will grow in length when anthers have already started to differentiate in further stages (AC). A, antesealous stamens; a, antepetalous stamens; Bl, bracteoles; C, carpel; S, sepals; P, petal. Scale bar = 500 µm in J, AC. 400 µm in. 300 µm in A–C, G–I, N, X. 200 µm in F, L, S, W, AA–AB. 100 µm in M, P–R, U–V, and Z. 50 µm in K. All photos displaying floral architectures by Domingos Cardoso.

Undefined androecium symmetry, where the androecium is composed of a profusion of more than 10, is reported within *Aldina*, *Bobgunnia*, *Bocoa*, *Camoensia* (Fig. 4F–J), *Candoleodendron*, *Cyathostegia* (Fig. 2F–J), *Fairchildia*, *Swartzia*, and *Trischidium* (Fig. 2AE) (Table 1).

4.7 Evolution of petal growth and androecium symmetry

Our ancestral state estimation for the heterochronical patterns of petal growth (Fig. 5) revealed that fast growth is the ancestral state in Papilionoideae and largely marked the Swartzioideae genera, as well as Angylocalyceae and Ambureneae lineages of the ADA clade. Slow petal growth evolved relatively early in the subfamily. It firstly appeared in the Swartzioideae *Trischidium* and independently in the node defining the small clade comprising the Dipterygeae genera *Taralea*, *Pterodon* and *Dipteryx*; and then in the ancestral node defining the large clade of Cladrastis + 50kb inversion clade. Multiple independent evolutionary reversals to fast petal

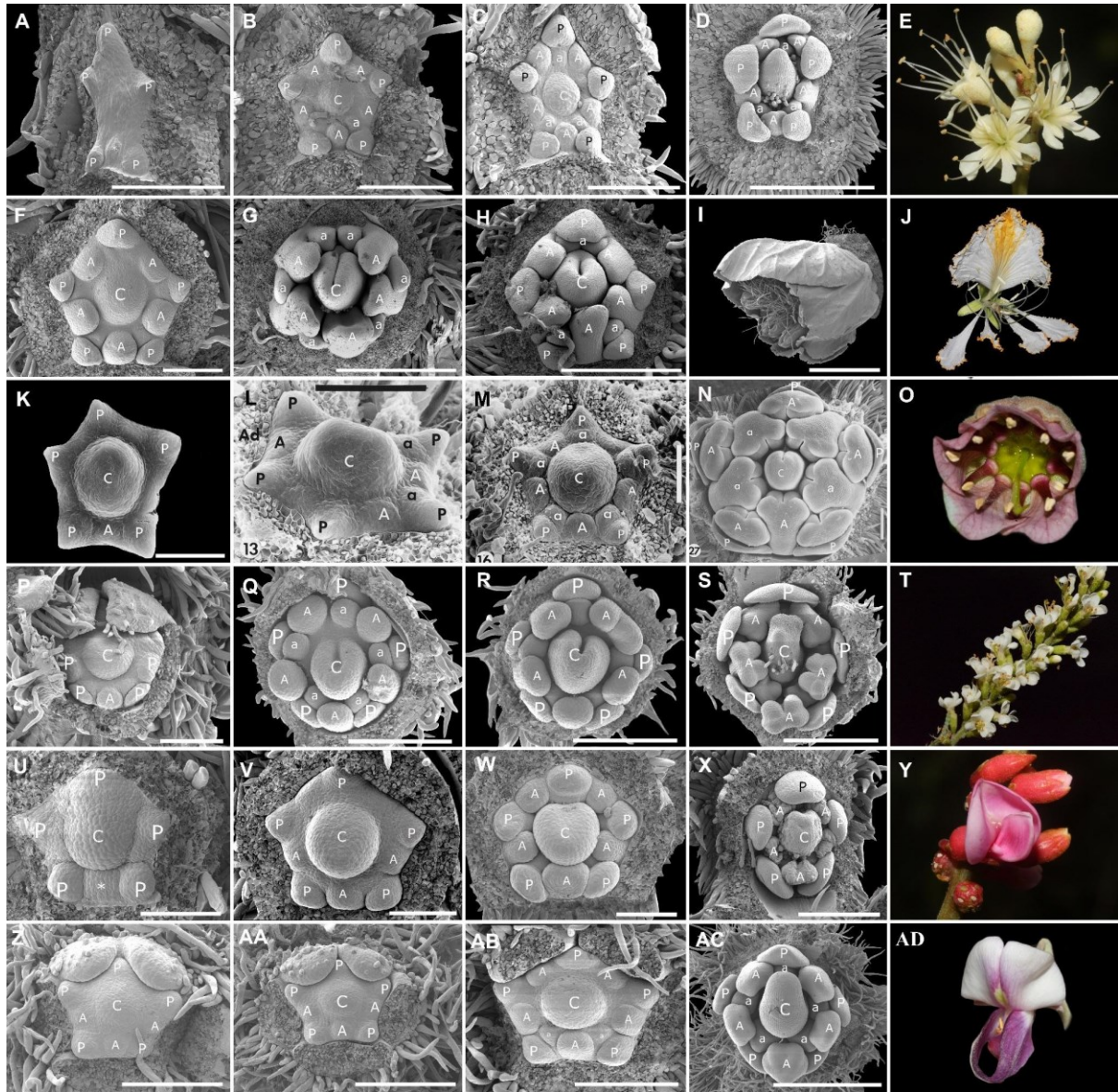


Figure 4. Androecium and petal development, and flower architecture in *Leptolobium bijugum*, *Camoensia scandens*, *Cadia purpurea*, *Eysenhardtia texana*, *Cleobulia multiflora* and *Olneya tesota*. A–E, *Leptolobium bijugum*. Early development of five petal primordia and stamens are seen in A–B. Petals have a fast growth relative to the Androecium, which is symmetrically arranged throughout development until late stages. F–J, *Camoensia scandens*. F, Early petals and stamens primordia grow at a somewhat similar pace, although only the first whorl of stamens is developed at this stage. G, First whorl of stamens grows in length while the second whorl is initiated. Petal removed. H–I, At this stage stamens outgrow petals, which are only slightly bigger than the second whorl of stamens. The androecium symmetry is not defined. K–O, *Cadia purpurea*. K–L, Petals and all stamens primordia are seen from early developmental stages. M, Stamens develop faster at this stage outgrowing petals, which

Chapter 4. An ontogenetic overview of the florally diverse Papilionoideae (Leguminosae): new insights into the evolution of androecium symmetry and petal heterochrony

only surpass the androecium development at later stages. *Cadia*'s androecium is asymmetrical in early development, assuming a radial symmetry on final stages. P–T, *Eysenhardtia texana*. P–Q, Initial development is marked by the development of all internal organs with stamens primordia growing faster than petals. R, The androecium asymmetry is seen early by the direction of the carpel cleft relative to adaxial stamens and banner petal. Petal development is slow when compared to androecium, which outgrow the corolla from early growth stages. S, The androecium asymmetry is more evident in later developmental stages when internal organs are fully differentiated. U–Y, *Cleobulia multiflora*. U–V, early development is marked by differential growth in petals and stamens. W, Although petals are initiated firstly, a succession of stamens primordia initiates from the abaxial side of the flower taking over growth over the petals. The symmetrical androecium is only evident after all internal organs are initiated. X, The slow growth of petals is seen from early stages until late stages where stamens show full differentiation whereas petals are still enlarging. Z–AD, *Olneya tesota*. Z, The early petal initiation suggests fast petal growth, but as soon as stamens are more evident (AA), they readily outgrow petal growth. AB, The symmetrical androecium is observed when most stamens have initiated and the carpellary cleft is more evident. AC, The slow petal growth and androecium symmetry are clearly visible as these organs differentiate in later stages. A, antesepalous stamens; a, antepetalous stamens; C, carpel; P, petal. Ad, adaxial side of flower meristem. Scale bar = 500 µm in D, G–H, N, and AC. 300 µm in F and S. 200 µm in A–C, Q–R, X, and Z–AB. 100 µm in K–M, P, and U–W. 2 mm in I. All photos displaying floral architectures by Domingos Cardoso.

growth occurred in the Exostyleae genus *Harleyondendron*, as well as the Dalbergioid genera *Machaerium*, *Riedeliella* and *Pterocarpus*. The species-rich NPAAA clade, where all genera are known to the consistent presence of papilionate flowers, with no single example of radial floral symmetry or any type of non-papilionate floral architecture, is also largely marked by the evolutionary conservatism of a slow petal growth.

The ancestral estimation of androecium symmetry (Fig. 6) revealed symmetrical androecium as the ancestral state for the entire subfamily. This ancestral trait shifted multiple times to asymmetrical androecia across the Papilionoideae phylogeny, more notably within early branching clades outside the large NPAAA clade. Despite such evolutionary lability of the asymmetrical androecium, this trait still marks large groups such as the dalbergioids including the Amorpheae, and most genistoids, although in this clade the Genisteeae genera reversed back to a symmetrical androecium (Fig. 6). Three independent evolutionary reversals

Chapter 4. An ontogenetic overview of the florally diverse Papilionoideae (Leguminosae): new insights into the evolution of androecium symmetry and petal heterochrony

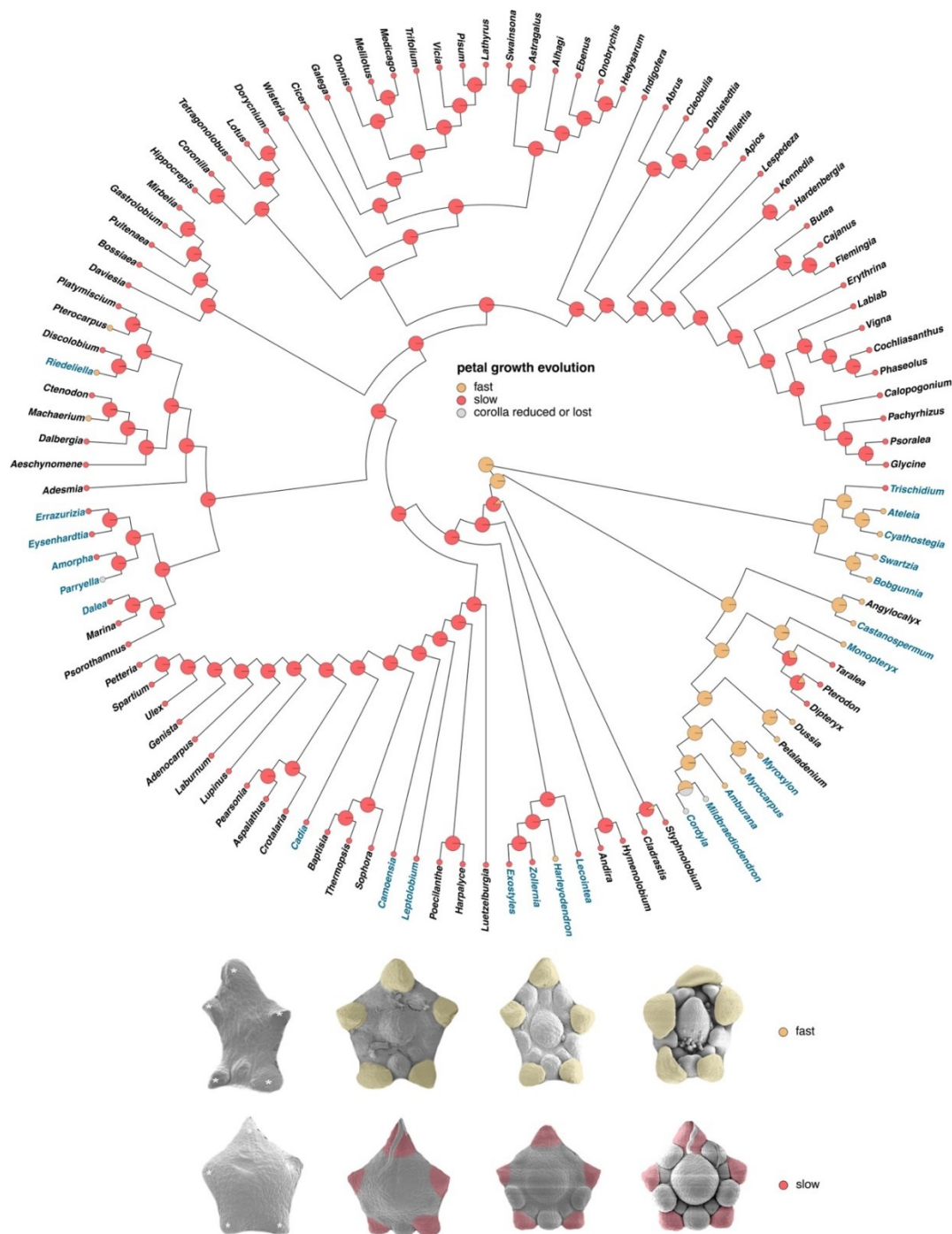


Figure 5. Evolution of petal growth across a plastome phylogeny of the Papilionoideae (sensu Choi et al. 2022), based on a hidden rates model (HRM) of ancestral estimation. Pie diagrams at each internal node indicate the proportional likelihoods of each state. Nodes represented in yellow indicate where ontogenetically-fast growing petals unambiguously evolved; nodes in red indicate slow petal growth; nodes in grey highlight nodes with reduced or lost corolla. Terminal branches in blue represent genera with non-papilionate floral architecture (e.g. radial symmetry, one petal, petals undifferentiated or completely absent). Representative SEM images of sequential ontogenetic series of slow and fast petal growths are shown below.

Chapter 4. An ontogenetic overview of the florally diverse Papilionoideae (Leguminosae): new insights into the evolution of androecium symmetry and petal heterochrony

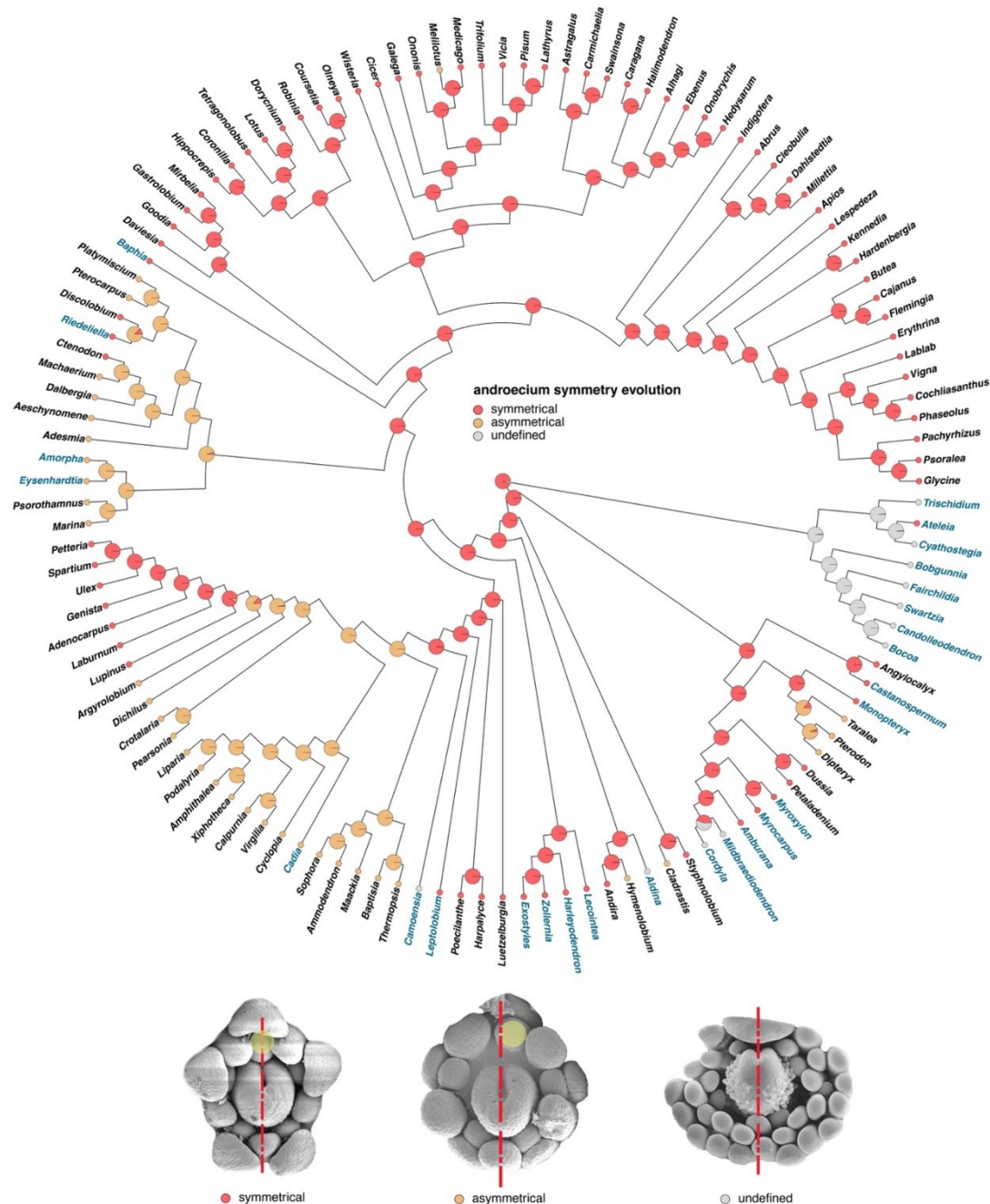


Figure 6. Evolution of androecium symmetry across a plastome phylogeny of the Papilionoideae (sensu Choi et al. 2022), based on a hidden rates model (HRM) of ancestral estimation. Pie diagrams at each internal node indicate the proportional likelihoods of each state. Nodes represented in red indicate where symmetrical androecium since early ontogeny evolved; nodes in yellow indicate evolution of asymmetrical androecia; nodes in grey highlight nodes with undefined symmetry, i.e. flower marked by a profusion of stamens where no fixed number can be readily defined. Terminal branches in blue represent genera with non-papilionate floral architecture (e.g. radial symmetry, one petal, petals undifferentiated or completely absent). Representative SEM images of symmetrical and asymmetrical androecia and undefined androecium symmetry are shown below.

into symmetrical androecium were also observed in the Dalbergioid clade: *Ctenodon*, *Riedeliella* and *Discolobium*.

A general look into floral architecture and the evolution of androecium symmetry reveals no clear evolutionary pattern, rather both types of symmetry seem to be found randomly in both typically papilionate species and radially symmetrical ones. For example *Cadia* exhibits radially symmetrical flower and asymmetrical androecium; *Hymenolobium* has a zygomorphic papilionate flower with an asymmetrical androecium; the papilionate-flowered *Andira* has a symmetrical androecium; and *Leptolobium* has a radially symmetrical flower with symmetrical androecium. However, just like in the apparent evolutionary conservatism of petal growth in the NPAAA clade, I did not find any single evolutionary shift into asymmetrical androecia across this large clade marked by a highly specialised papilionate floral architecture with symmetrical androecium since early ontogeny.

Undefined symmetry because of multiple stamens largely marked the Swartztioid clade, with an unusual reversal into symmetrical androecium in *Ateleia*, and multiple instances of independent evolution found scattered in *Cordyla*, *Mildbraediodendron*, *Aldina* and *Camoensia*. All of these genera with undefined androecium symmetry are also marked by varying non-papilionate floral architectures.

Discussion

4.8 From zygomorphic to radially symmetrical flowers

The typical zygomorphic papilionate flower is pentamerous, with most taxa having a full set of organs: alternating whorls of sepals, petals, two stamen whorls and a single carpel (Tucker 2003). Where organs appear to be fewer or lacking, ontogeny shows that all are either initiated and subsequently suppressed during development (Leite et al. 2015) or are lost altogether (Torke & Mansano 2009; Sinjushin 2018). In most Papilionoideae, all floral organs are initiated in a unidirectional successive order in each whorl starting on the abaxial side (next to the subtending bract) (Tucker 2003). Development of different organs may overlap in some papilionoids, where stamens start to initiate before the last petals have been initiated (Tucker, 2003). The pattern of organogenesis differs in the tribe Swartzieae, where a ring meristem

gives rise to numerous stamen primordia. Swartzoid genera such as *Swartzia* have a single petal initiated or no evidence of floral meristem is observed (Torke and Mansano, 2009). The carpel is generally single, but a few cases of polycarpelly are observed in distinct lineages.

Lineages with unusual floral architecture are found scattered across the early branching lineages of the Papilionoideae phylogeny (see Ch 1; Pennington et al. 2001; Wojciechowski et al. 2004; Cardoso et al. 2012, 2013; LPWG, 2017) and include up to 37 genera: *Acosmium*, *Aldina*, *Alexa*, *Amburana*, *Amorpha*, *Amphimas*, *Angylocalyx*, *Apoplanesia*, *Ateleia*, *Bobgunnia*, *Bocoa*, *Cadia*, *Camoensia*, *Candoleodendron*, *Castanospermum*, *Cordyla*, *Cyathostegia*, *Dicraeopetalum*, *Etaballia*, *Exostyles*, *Eysenhardtia*, *Guianodendron*, *Harleyodendron*, *Holocalyx*, *Inocarpus*, *Lecointea*, *Leptolobium*, *Mildbraediodendron*, *Myrocarpus*, *Myroxylon*, *Parryella*, *Pictetia*, *Pterocarpus*, *Riedeliella*, *Swartzia*, *Xanthocercis* and *Zollernia* (Beyra-Matos & Lavin, 1999; Pennington et al. 2000; Boatwright et al. 2008; Cardoso et al. 2012; Klitgaard et al. 2013). Phylogenetic studies in legumes have provided new insights into the evolution of such florally divergent genera (Ireland et al. 2000, 2007; Pennington et al. 2000, 2001; Mansano et al. 2004; MacMahon & Hufford 2004; Boatwright et al. 2008; Cardoso et al. 2012, 2013, 2015; Klitgaard et al. 2013; Ramos et al. 2016). These studies have revealed the relationship of genera that are phylogenetically close, often sister genera, with drastic differences in floral morphologies. This new knowledge has drawn attention to morphological puzzles that the flowers of these lineages pose, motivating more research on the floral ontogeny field in legumes (Tucker 1990, 1994, 1999, 2002, 2003; Klitgaard 1999; Mansano et al. 2002; Torke & Mansano 2009; Paulino et al. 2014; Leite et al. 2015; Sinjushin 2018; Bento et al. 2021).

Genera with radially symmetrical flowers have been thought to be primitive among Papilionoideae being traditionally classified in the tribes Swartzieae and Sophoreae based on a combination of characters such as radial symmetry, fused sepals, numerous and free stamens and unspecialized petals (Cowan 1981; Polhill 1981, 1994). Phylogenetic studies have changed this view by showing that flower morphology is evolutionarily labile among papilionoid lineages (Pennington et al. 2000; Klitgaard et al. 2013; Bruneau et al. 2014; Cardoso et al. 2012a, 2013, 2015). Pennington et al. (2000) listed 23 genera of non-papilionate genera in Papilionoideae, from which at least nine reversions have occurred from a papilionate flower to one with radial or undifferentiated morphology: *Acosmium*, *Amphimas*, *Cadia*,

Dicraepetalum, *Harleyodendron*, *Holocalyx*, *Lecointea*, *Myrocarpus* and *Zollernia* (Pennington et al. 2001).

Molecular phylogenetic studies on the florally heterogeneous dalbergioids that densely sampled all genera with radially symmetric flowers (*Acosmium*, *Etaballia*, *Inocarpus* and *Riedeliella*) demonstrated that actinomorphy evolved independently at least four times in the clade (Lavin et al. 2001; Cardoso et al. 2012a, 2013; Klitgaard et al. 2013). By comparing the floral ontogeny of the radially symmetric *Riedeliella* and its typically zygomorphic papilionate sister genus *Discolobium*, Bento et al. (2021) observed differences between the two genera from early stages in petal development, where *Riedeliella* developed unidirectional growth and *Discolobium* petal differentiation initiated by the petal standard, followed by the keel petals and finally by the wing petals. Floral symmetry of both genera is expressed at different times throughout development, with the zygomorphy of *D. pulchellum* being expressed earlier compared to the actinomorphic symmetry of *R. graciliflora*, which occurs at a later developmental stage. Bento et al. (2021) also concluded, based on comparison with other unusual papilionoid genera with radially symmetric flowers, that the development of *R. graciliflora* involved several unique stages. Further ontogenetic comparison with the remaining radially symmetric *Inocarpus*, *Acosmium*, and *Etaballia* can shed light on the homology of these flowers and allow hypotheses on the genetic process underlying flower development in the clade.

A similar comparison was carried out by Citerne et al. (2006) under a genetic approach with the phylogenetically close *Lupinus nanus* (typical papilionate) and *Cadia purpurea* (radially symmetrical). CYCLOIDEA genes were found to play a role in the establishment of dorsal petal identity in *Lupinus* and *Cadia*. Even though both species are drastically distinct in flower form, *C. purpurea* presented the same set of genes involved in standard petal differentiation as *L. nanus*. RT-PCR experiments showed that the radial symmetry of *C. purpurea* is the result of the overexpression of the CYC copy *LegCYC1B*, which is responsible for the development of a petal with banner morphology throughout the lateral and abaxial domains (Citerne et al. 2006). This result suggested that the dorsalized corolla of *Cadia*, thought to be an evolutionary reversal, is a homeotic transformation, where all petals have acquired dorsal identity.

In some papilionoids, flowers show a reduced number of floral organs. For example, Leite et al. (2015) showed that in *Amburana cearensis* (Papilionoideae) the single petal corolla

is the result of the abortion of four petal primordia in their early stages of development, demonstrating that, although apparently distinct in morphology from the typical papilionate flower, the genus still retains floral traits in its development consistent with the rest of the subfamily. From a genetic perspective, it seems possible that suppression of development of lateral and ventral domains in *Amburana*'s unusual single-petalled flower may be the result of suppression of genes involved in the development of these domains, since the primordia of those petals are discernible in their early stages of development.

The pattern of petal aestivation, i.e., the way petals fold in the flower bud, is imbricate in Papilionoideae in a pattern called descending cochleate, in which the standard petal margins overlap the adjacent wing margins, which overlaps the adjacent keel margins. This pattern contrasts with that of Caesalpinioideae (ascending cochleate) and mimosoid genera (valvate, no overlap). Some papilionoids with unusual flower morphology, have a random pattern of petal aestivation. This is the case of the radially symmetric *Cadia purpurea*, whose petals overlap by chance throughout development (Tucker 2002, 2003). This type of aestivation is probably related to the fact that all petals margin in *C. purpurea* grow straight outward like the standard petal. Throughout development, chance determines which petal overlaps another (Tucker 2002, 2003). Petal development in *Cadia* has been investigated by Citerne et al. (2006), who gave the first genetics insights about the development of *Cadia* flower. It was observed that the expansion of expression of the transcription factor *LegCYC1B*, responsible for the establishment of the dorsoventral axis in legumes, is responsible for the lateral and ventral petals acquiring dorsal identity (Citerne et al. 2006). This study showed the important role of gene expression in the determination of overall floral morphology. In Chapter Three, I used transcriptomic data to investigate the expression of CYC and related genes in the typical papilionate *Andira* and the morphologically distinct *Cadia* and the single-petalled *Cyathostegia*. Our results showed that CYC was differentially expressed among these species in different developmental stages through flower growth, suggesting that a higher or lower expression level is associated with distinct morphology types.

4.9 Ontogenetic stability of slow petal growth and symmetrical androecium underlying the evolutionary stability of the papilionate floral architecture in papilionoid legumes

The reconstructed evolution of petal heterochrony and androecium symmetry revealed highest evolvability within the early branching Papilionoideae lineages (Figs. 5-6). As previously hypothesised by Prenner (2004a), the ancestral estimations confirmed that asymmetrical androecium arose independently several times, and that both androecium types evolved independently of staminal fusion or floral architecture. This finding largely coincides with the high floral evolutionary lability in these early diverging clades, where many genera have evolved their own non-papilionate floral architecture that greatly deviates from the highly specialised papilionate flower (see discussion above; Pennington et al. 2000; Cardoso et al. 2013). Such dramatic floral changes involve varying levels of corolla symmetry, petal loss, petal transformation or organ suppression, as well as profusion of stamens and different degrees of staminal fusion.

While Prenner (2004a) previously proposed that symmetrical androecia would be a derived trait within Papilionoideae, my findings contradict this hypothesis. The evidence presented here demonstrates that symmetrical androecia have been present since the origin of the Papilionoideae subfamily (Fig. 6). Interestingly, despite contradicting Prenner's (2004) initial suggestion, this ontogenetic trait appears to be highly conserved in more derived clades. Specifically, both symmetrical androecium and slow petal growth are evolutionarily fixed characteristics within the species-rich NPAAA clade. Remarkably, in this clade, the typical zygomorphic papilionate flower has also remained unchanged, never switching to any other unusual non-papilionate floral architecture. The highly diversified NPAAA clade of papilionoid legumes encompasses almost 70% (9464 spp.) of the species diversity in the subfamily (Cardoso et al. 2012a, 2013; LPWG 2017). Indeed, most megadiverse papilionoid genera are found in this clade, such as *Astragalus* (3068 species), *Indigofera* (735), *Oxytropis* (614), *Trifolium* (291) and *Vicia* (254). Remarkably, the diversification of the NPAAA clade is not just underscored by an ontogenetic stability of slow petal development and symmetrical androecium (Figs. 5-6) underlying the conservatism of a specialised papilionate floral architecture. This lineage is known to have acquired highly stable symbioses with nitrogen-fixing bacteria, although the evolution of stable fixers is likely to have multiple origins across the legume family (Werner et al. 2015).

Floral heterochrony has been hypothesised to promote stability of reproductive strategies in large groups with morphologically homogeneous flowers. For example, Vasconcellos et al. (2017) demonstrated that reproductive strategies in *Eugenia* species are associated with superficially similar flowers. Subtle changes in developmental rates have altered the proportions between floral organs throughout the evolutionary history of the group. The heterochronies observed in the floral organs of *Eugenia* species might have played a role in subtle breeding system changes (affecting differential production of male and female parts) and phenology (floral development rate changes) (Vasconcellos et al. 2017). Consequently, the observed uneven diversity within the tribe Myrteae, traditionally attributed primarily to the key innovation of fleshy fruit, is also proposed to have been influenced by heterochronic processes that result in alterations in flower structure (Vasconcellos et al. 2017). The role of heterochrony in pollination biology has also been reported in Trimezieae (Iridaceae). Lovo et al. (2021) explored the ontogenetic floral lability of Trimezieae and concluded that heterochronic processes may have led to repeated evolutionary shifts in the pollination biology of the tribe, which has allowed the lineage to explore new ecological opportunities.

The question of whether the evolutionary stability of ontogenetic traits has played a role in the extraordinary ecological and evolutionary success of the papilionoid NPAAA lineages, akin to the ancient symbiosis between legumes and nitrogen-fixing bacteria, remains an open inquiry. This deserves further exploration, particularly in the context of a more extensive ontogenetic dataset encompassing numerous papilionoid genera that have not been scrutinised ontogenetically.

4.10 Insights into differential rates of petal growth and androecium symmetry from transcriptomics

Changes in floral morphology are mainly induced by shifts in space, caused by the time of initiation (heterochrony), pressure of organs, and alterations of the size of the floral meristem, operating independently or in parallel with the expression of important genes involved in floral development (Ronse De Craene 2018; Buendia-Monreal & Gillmor 2018). For example, it was shown in Brassicaceae that the evolution of a monosymmetrical corolla from the polysymmetrical ancestral flower involved a heterochronic shift in the expression of

CYCLOIDEA (*CYC2*) genes from early adaxial expression in ancestral floral meristem to a later adaxial expression in petal development (Busch et al. 2012).

Changes in the timing of activation and repression of gene expression is named transcriptional heterochrony and is often caused by changes in *cis*-regulatory gene regions (Pham et al. 2017). Throughout the transition from vegetative to reproductive tissues, meristem stages are defined by distinct molecular states and point to modified organ maturation schedules underlying morphological variation. In the nightshade family (Solanaceae), the comparison of transcriptomes of meristem maturation of five domesticated and wild species revealed a peak of expression divergence during this reproductive transition. The subtle differences of transcriptional programs observed were driven by heterochronic shifts of dynamic genes during a transitional window of meristem maturation, which led to morphological variation in inflorescence architecture (Lemmon et al. 2016).

In the papilionoid genus *Lotus*, changes in timing of expression of *CYC*-like genes are associated with modifications in petal growth and lateralization of the dorsal and ventral petals. These temporal shifts in expression have contributed to changes in petal size and micromorphology leading to a transition from bee pollination to bird pollination in four species of the genus (Ojeda et al. 2017).

Ronse de Craene (2018) emphasised the importance of heterochrony for the development of common stamen-petal primordia, linking changes in growth rate with delayed petal initiation. Common stamen-petal primordia is the name given to the floral meristem from which organs initiate. In the case of stamens and petals, instead of a creation of a new primordium in a sequence of organ inception, the dividing primordium does not lead to the creation of a new organ, but to its absorption and loss, a phenomenon described by Ronse de Craene et al. (1993) as a case of negative dédoublement.

Further studies are essential to illuminate the potential heterochrony associated with floral organs, including the androecium and petals, and to unravel the genetic factors responsible for regulating these organs. In Chapter Three I showed several genes, including those involved in the regulation of floral development, differentially expressed at different stages throughout flower development. Our results showed that the expression of some genes is shoscheduled for specific floral tissues (e.g. vegetative buds, young flower buds, abaxial and adaxial regions), with time and place for initiation and end of expression. More detailed studies

Chapter 4. An ontogenetic overview of the florally diverse Papilionoideae (Leguminosae): new insights into the evolution of androecium symmetry and petal heterochrony

with ontogenetic data from floral organs and their independent expression profile are vital to establish a putative heterochrony between floral organs, such as the ones discussed in this chapter, and the expression of genes regulating these organs. It is key that new studies, comparatively, investigate papilionoid species with distinct morphologies in order to gain insights on the correlation of morphological and genetic heterochrony.

References

- Beaulieu, J. M., B. C. O'Meara, M. J. Donoghue. 2013. Identifying hidden rate changes in the evolution of a binary morphological character: the evolution of plant habit in campanulid angiosperms. *Systematic Biology* 62: 725–737.
- Beaulieu, J.M., O'Meara, B.C., Oliver, J. & Boyko, J. 2020. *corHMM*: Hidden Markov Models of character evolution. R package version 2.3. <https://CRAN.Rproject.org/package=corHMM>. (Accessed 03.08.2020).
- Bento, J. P. S., F. J. Kochanovski, A. L. B. Sartori. 2021. Floral ontogeny and secretory structures of *Discolobium pulchellum* and *Riedeliella graciliflora* (Leguminosae: Papilionoideae: Dalbergieae): two closely related genera with diverging floral morphology. *Plant Systematics and Evolution* 307: 6.
- Beyra-M., A., M. Lavin. 1999. Monograph of *Pictetia* (Leguminosae-Papilionoideae) and review of the *Aeschynomeneae*. *Systematic Botany Monographs* 56: 1–93.
- Boatwright, J. S., V. Savolainen, B.-E. van Wyk, A. L. Schutte-Vlok, F. Forest, M. van der Bank. 2008. Systematic position of the anomalous genus *Cadia* and the phylogeny of the tribe Podalyrieae (Fabaceae). *Systematic Botany* 33: 133–147.
- Bruneau, A. et al. 2014. Floral evolution in the Detarieae (Leguminosae): phylogenetics evidence for labile floral development in an early-diverging legume lineage. *International Journal of Plant Science* 175: 392–417.
- Buendia-Moreau, G. & C. S. Gillmor. 2018. The times they A-changin': heterochrony in plant development and evolution. *Frontiers in Plant Science* 9: 1349.
- Busch, A., Horn S., Mühlhausen A., Mummenhoff K., Zachgo S. 2012. Corolla monosymmetry: evolution of a morphological novelty in the Brassicaceae family. *Molecular Biology and Evolution* 29:1241–1254.
- Cardoso, D., L. P. Queiroz, R. T. Pennington, H. C. de Lima, E. Fonty, M. F. Wojciechowski, M. Lavin. 2012a. Revisiting the phylogeny of papilionoid legumes: new insights from comprehensively sampled early-branching lineages. *American Journal of Botany* 99: 1991–2013.
- Cardoso, D., H. C. Lima, R. S. Rodrigues, L. P. Queiroz, R. T. Pennington, M. Lavin. 2012b. The realignment of *Acosmium* sensu stricto with the dalbergioid clade (Leguminosae, Papilionoideae) reveals a proneness for independent evolution of radial floral symmetry among early branching papilionoid legumes. *Taxon* 61: 1057–10.

- Cardoso, D. et al. 2013. Reconstructing the deep-branching relationships of the papilionoid legumes. *South African Journal of Botany* 89: 58–75.
- Cardoso, D. et al. 2015. Filling in the gaps of the papilionoid legume phylogeny: the enigmatic Amazonian genus *Petaladenium* is a new branch of the early-diverging Amburaneae clade. *Molecular Phylogenetics and Evolution* 84: 112–124.
- Citerne, H. L., R. T. Pennington, Q. C. B. Cronk. 2006. An apparent reversal in floral symmetry in the legume *Cadia* is a homeotic transformation. *Proceedings of the National Academy of Sciences* 103: 12017–12020.
- Cubas, P. 2004. Floral zygomorphy, the recurring evolution of a successful trait. *BioEssays* 26: 1175–1184.
- Choi, I.-S., D. Cardoso, L. P. Queiroz, H. C. Lima, C. Lee, T. A. Ruhlman, R. K. Jansen, M. F. Wojciechowski. 2022. Highly resolved papilionoid legume phylogeny based on plastid phylogenomics. *Frontiers in Plant Science* 13: 823190.
- Ronse de Craene, L. P. & Smets E. 1993. Dédoublément revisited: towards a renewed interpretation of the androecium of the Magnoliophytina. *Botanical Journal of the Linnean Society* 113: 285–350.
- Ronse de Craene, L. P. 2018. Understanding the role of floral development in the evolution of angiosperm flowers: clarifications from a historical and physico-dynamic perspective. *Journal of Plant Research* 131: 367–393.
- Endress, P. K. 1994. Diversity and evolutionary biology of tropical flowers. Cambridge University Press, Cambridge, UK.
- Endress, P. K. 1999. Symmetry in flowers: diversity and evolution. *International Journal of Plant Sciences* 160(S6): S3–S23.
- Ireland, H. E., R. T. Pennington, J. Preston. 2000. Molecular systematics of the Swartzzieae. In P. S. Herendeen and A. Bruneau (eds). *Advances in Legume Systematics*. Royal Botanic Garden, Kew, 9: 217–231.
- Ireland, H. E. 2007. Taxonomic changes in the South American genus *Bocoa* (Leguminosae-Swartzieae): reinstatement of the name *Thrischidium* and a synopsis of both genera. *Kew Bulletin* 62: 333–349.
- Klitgaard, B. B. 1999. Floral ontogeny in tribe Dalbergieae (Leguminosae: Papilionoideae): *Dalbergia brasiliensis*, *Machaerium villosum* s.l. *Platymiscium floribundum*, and *Pterocarpus rotundifolius*. *Plant Systematics and Evolution* 219: 1–25.

- Klitgaard, B. B., et al. 2013. A detailed investigation of the Pterocarpus clade (Leguminosae: Dalbergieae): *Etabalia* with radially symmetrical flowers is nested within the papilionoid-flowered *Pterocarpus*. *South African Journal of Botany* 89: 128–142.
- Koenen, E. J. M., J. M. de Vos, G. W. Atchison, M. F. Simon, B. D. Schrire, E. R. Souza, L. P. Queiroz, C. E. Hugues. 2013. Exploring the tempo of species diversification in legumes. *South African Journal of Botany* 89: 19–30.
- Lavin, M., P. S. Herendeen, M. F. Wojciechowski. 2005. Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the Tertiary. *Systematic Biology* 54: 575–594.
- Lemmon, Z. H., S. J. Park, K. Jiang, J. Van Eck, M. C. Schatz, Z. B. Lippman. 2016. The evolution of inflorescence diversity in the nightshades and heterochrony during meristem maturation. *Genome Research* 26: 1676–1686.
- Leite V. G., V. F. Mansano, S. P. Teixeira. 2014. Floral ontogeny in Dipterygeae (Fabaceae) reveals new insights into one of the earliest branching tribes in papilionoid legumes. *Botanical Journal of the Linnean Society* 174: 529–550.
- Leite, V. G. et al. 2015. Floral development of the early-branching papilionoid legume *Amburana cearensis* (Leguminosae) reveals rare and novel characters. *International Journal of Plant Sciences* 176: 94–106.
- Li, P. & Johnston, M. O. 2000. Heterochrony in plant evolutionary studies through the twentieth century. *Botanical Review* 66: 57–88.
- LPWG [Legume Phylogeny Working Group]. 2017. A new subfamily classification of the Leguminosae based on taxonomically comprehensive phylogeny. *Taxon* 66: 44–77.
- Lovo, J., S. Alcantara, T. N. Vasconcelos, M. G. Sajo, P. J. Rudall, G. Prenner, A. J. C. Aguiar, R. Mello-Silva. 2021. Evolutionary lability in floral ontogeny affects pollination biology in Trimezieae. *American Journal of Botany* 108: 828–843.
- Lord, E. M. 1991. The concepts of heterochrony and homeosis in the study of floral morphogenesis. *Flowering Newsletter* 11: 4–13.
- MacMahon, M., Larry H. 2004. Phylogeny of Amorpheae (Fabaceae, Papilionoideae). *American Journal of Botany* 91: 1219–1230.
- Mansano, V. F., S. C. Tucker, A. M. G. A. Tozzi. 2002. Floral ontogeny of *Lecointea*, *Zollernia*, *Exostyles*, and *Harleyodendron* (Leguminosae: Papilionoideae: Swartzieae s.l.). *American Journal of Botany* 89: 1553–1569.

- Mansano, V. F., V. Bittrich, A. M. G. A. Tozzi, A. P. Souza. 2004. Composition of the *Lecointea* clade (Leguminosae, Papilionoideae, Swartzieae), a re-evaluation based on combined evidence from morphology and molecular data. *Taxon* 53: 1007–1018.
- Marazzi, B. & P. K. Endress. 2008. Patterns and development of floral asymmetry in *Senna* (Leguminosae, Cassiinae). *American Journal of Botany* 95: 22–40.
- Marazzi, B., D. Ané, M. Simon, A. Delgado-Salinas, M. Luckow, M. J. Sanderson. 2012. Locating evolutionary precursors on a phylogenetic tree. *Evolution* 66: 3918–3930.
- Ojeda, D. I., Jaén-Molina R., Santos-Guerra A., Caujape-Castells J., Cronk Q. 2017. Temporal, but not spatial, changes in expression patterns of petal identity genes are associated with loss of papillate conical cells and the shift to bird pollination in Macaronesian *Lotus* (Leguminosae). *Plant Biology* 19: 420–427.
- Paradis, E., Schliep K. 2019. *ape* 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35: 526–528.
- Paulino, J. V., V. F. Mansano, S. P. Teixeira. 2013. Elucidating the unusual floral features of *Swartzia dipetala* (Fabaceae). *Botanical Journal of the Linnean Society* 173: 303–320.
- Paulino, J. V., Prenner G., Mansano V. F., Teixeira S. P. 2014. Comparative development of rare cases of a polycarpellate gynoecium in an otherwise monocarpellate family, Leguminosae. *American Journal of Botany* 101: 572–586.
- Pennington, R. T., B. B. Klitgaard, H. Ireland, M. Lavin. 2000. New insights into floral evolution of basal Papilionoideae from molecular phylogenies. In P. S. Herendeen & A. Bruneau [eds.], *Advances in Legume Systematics*, part 9, 233–248. Royal Botanic Gardens, Kew, UK.
- Pennington, R. T., M. Lavin, H. Ireland, B. Klitgaard, J. Preston, J. M. Hu. 2001. Phylogenetic relationships of basal papilionoid legumes based upon sequences of the chloroplast *trnL* intron. *Systematic Botany* 26: 537–556.
- Pham, T., S. M. Day, W. J. Glassford, T. M. Williams, M. Rebeiz. 2017. The evolutionary origination of a novel expression pattern through an extreme heterochronic shift. *Evolution and Development* 19: 43–55.
- Prenner, G. 2004a. New aspects in floral development of Papilionoideae: initiated but suppressed bracteoles and variable initiation of sepals. *Annals of Botany* 93: 537–545.
- Prenner, G. 2004b. The asymmetric androecium in Papilionoideae (Leguminosae): definition, occurrence, and possible systematic value. *International Journal of Plant Sciences* 165: 499–510.

- Prenner, G. & Klitgaard B. B. 2008. Towards unlocking the deep nodes of Leguminosae: floral development and morphology of the enigmatic *Duparquetia orchidacea* (Leguminosae, Caesalpinoideae). *American Journal of Botany* 95: 1349–1365.
- Prenner, G., D. Cardoso, C. E. Zartman, L. P. Queiroz. 2015. Flowers of the early-branching papilionoid legume *Petaladenium urceoliferum* display unique morphological and ontogenetic features. *American Journal of Botany* 102: 1780–93.
- Prenner, G. & D. Cardoso. 2017. Flower development of *Goniorrachis marginata* reveals new insights into the evolution of the early diverse detarioid legumes. *Annals of Botany* 119: 417–432.
- Raff, R. A., Wray, G. A. 1989. Heterochrony: developmental mechanisms and evolutionary results. *Journal of Evolutionary Biology* 2: 409–434.
- Ramos G, Lima H. C., Prenner G., Queiroz L. P., Zartman C. E., Cardoso D. 2016. Molecular systematics of the Amazonian genus *Aldina*, a phylogenetically enigmatic ectomycorrhizal lineage of papilionoid legumes. *Molecular Phylogenetics and Evolution* 97: 11–18.
- Revell, L. M. 2012. *phytools*: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* 3: 217–223.
- Saarela, J. M., Rai, H. S., Doyle, J. A., et al. 2007. Hydatellaceae identified as a new branch near the base of the angiosperm phylogenetic tree. *Nature* 446: 312–315.
- Sinjushin, A. A. 2018. Floral ontogeny in *Cordyla pinnata* (A.Rich.) Milne-Redh. (Leguminosae, papilionoideae): away from stability. *Flora* 241: 8–15.
- Sargent, R. D. 2004. Floral symmetry affects speciation rates in angiosperms. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 271: 603–608.
- Thiers, B. (continuously updated) Index Herbariorum: a global directory of public herbaria and associated staff. New York Botanical Garden's Virtual Herbarium. Available at <<http://sweetgum.nybg.org/ih/>>. Access on 30 Dec 2023.
- Torke, B. M. & V. F. Mansano. 2009. A phylogenetically based sectional classification of *Swartzia* (Leguminosae-Papilionoideae). *Taxon* 58: 913–924.
- Tucker, S. C. 1987. Floral initiation and development in legumes. In: Stirton C. H. ed. *Advances in legume systematics*, Part 3. Kew: Royal Botanic Gardens, 183–239.
- Tucker, S. C. 1988. Loss versus suppression of floral organs. In: Leins P, Tucker S. C., Endress PK. eds. *Aspects of floral development*. Berlin: J. Cramer, 69–82.
- Tucker, S. C. 1990. Loss of floral organs in *Ateleia* (Leguminosae, Papilionoideae). *American Journal of Botany* 77: 750–761.

- Tucker, S. C. & Douglas, A. W. 1994. Ontogenetic evidence and phylogenetic relationships among basal taxa of legumes. In: Ferguson, I. K., Tucker, S. (Eds.), *Advances in Legume Systematics*, Structural Botany, Vol. 6. Royal Botanic Gardens, Kew, pp. 11–32.
- Tucker, S. C. 1999. Evolutionary lability of symmetry in early floral development. *International Journal of Plant Sciences* 160: S25–S39.
- Tucker, S. C. 2002. Floral ontogeny in Sophoreae (Leguminosae: Papilionoideae). III. Radial symmetry and random petal aestivation in *Cadia purpurea*. *American Journal of Botany* 89: 748–757.
- Tucker, S. C. 2002. Floral ontogeny of *Cercis* (Leguminosae: Caesalpinioideae: Cercideae): Does it show convergence with papilionoides? *International Journal of Plant Sciences* 163: 75–87.
- Tucker, S. C. 2003. Floral development in legumes. *Plant Physiology* 131: 911–926.
- Tucker, S.C. 2003. Floral ontogeny in *Swartzia* (Leguminosae: Papilionoideae: Swartzieae): distribution and role of the ring meristem. *American Journal of Botany* 90: 1271–1292.
- Vasconcelos, T. N. C., E. J. Lucas, J. E. Q. Faria, G. Prenner. 2018. Floral heterochrony promotes flexibility of reproductive strategies in the morphologically homogeneous genus *Eugenia* (Myrtaceae). *Annals of Botany* 121: 161–174.
- Wang, L., T. T.-Y. Lam, S. Xu, Z. Dai, L. Zhou, T. Feng, P. Guo, C.W. Dunn, B. R. Jones, T. Bradley, H. Zhu, Y. Guan, Y. Jiang, G. Yu. 2020. *treeio*: an R package for phylogenetic tree input and output with richly annotated and associated data. *Molecular Biology and Evolution* 37: 599–603.
- Werner, G. D. A., W. K. Cornwell, J. H. C. Cornelissen, E. T. Kiers. 2015. Evolutionary signals of symbiotic persistence in the legume–rhizobia mutualism. *Proceedings of the National Academy of Sciences of the USA* 112: 10262–10269.
- Wojciechowski, M. F., M. Lavin, M. J. Sanderson. 2004. A phylogeny of legumes (Leguminosae) based on analysis of the plastid *matK* gene resolves many well-supported subclades within the family. *American Journal of Botany* 91: 1846–1862.
- Zhao, Z., J. Hu, S. Chen, Z. Luo, D. Luo, J. Wen, T. Tu, D. Zhang. 2019. Evolution of CYCLOIDEA-like genes in Fabales: insights into duplication patterns and the control of floral symmetry. *Molecular Phylogenetics and Evolution* 132: 81–89.
- Zhao, Y., R. Zhang, K.-W. Jiang, J. Qi, Y. Hu, J. Guo, R. Zhu, T. Zhang, A. N. Egan, T.-S. Yi, C.-H. Huang, H. Ma. 2021. Nuclear phylotranscriptomics and phylogenomics support

Chapter 4. An ontogenetic overview of the florally diverse Papilionoideae (Leguminosae): new insights into the evolution of androecium symmetry and petal heterochrony

numerous polyploidization events and hypotheses for the evolution of rhizobial nitrogen-fixing symbiosis in Fabaceae. *Molecular Plant* 14: 748–773.

Chapter 4. An ontogenetic overview of the florally diverse Papilionoideae (Leguminosae): new insights into the evolution of androecium symmetry and petal heterochrony

Table 1. Patterns of androecium symmetry and petal heterochrony in Papilionoideae. sy = symmetrical; as = asymmetrical; un = undefined; sg = slow growth; fg = fast growth; ab = reduced or absent corolla.

Species	Androecium symmetry	Petal growth	Flower architecture	Reference
<i>Abrus precatorius</i>	sy	sg	papilionate	Prenner (2013)
<i>Adenocarpus decorticans</i>	sy	sg	papilionate	Prenner & Cardoso (unpubl. data)
<i>Adesmia ciliata</i>	as	sg	papilionate	Moco & Mariath (2009)
<i>Aeschynomene sensitiva</i>	as	sg	papilionate	Sampaio et al. (2013)
<i>Aldina latifolia</i>	un	unknown	non-papilionate	Prenner & Cardoso (unpubl. data)
<i>Alhagi persarum</i>	sy	sg, vex sw. fg	papilionate	Khodaverdi et al. (2014)
<i>Amburana cearensis</i>	sy	fg, 1 adax only	non-papilionate	Leite et al. (2015)
<i>Ammodendron</i>	as	unknown	papilionate	
<i>Amorpha fruticosa</i>	as	sg, 1 adax only	non-papilionate	Tucker 1987; Prenner (2002)
<i>Amphithalea intermedia</i>	as	unknown	papilionate	
<i>Andira ormosioides</i>	sy	sg	papilionate	Prenner & Cardoso (unpubl. data)
<i>Angylocalyx braunii</i>	sy	fg	papilionate	Prenner & Cardoso (unpubl. data)
<i>Apios americana</i>	sy	sg, vex fg	papilionate	Prenner (2002)
<i>Argyrolobium speciosum</i>	as	unknown	papilionate	
<i>Aspalathus nivea</i>	unknown	sg	papilionate	Prenner & Cardoso (unpubl. data)
<i>Astragalus compactus</i>	sy	sg, vex fg	papilionate	Naghiloo et al. (2012)
<i>Ateleia herbertsmithii</i>	sy	fg, 1 adax only	non-papilionate	Tucker (1990)
<i>Ateleia guaraya</i>	sy	fg, 1 adax only	non-papilionate	Prenner & Cardoso (unpubl. data)
<i>Baphia polyantha</i>	sy	unknown	non-papilionate	Prenner (2004)
<i>Baptisia australis</i>	as	sg	papilionate	Prenner (2002)
<i>Bobgunnia madagascariensis</i>	un	fg, 1 adax only	non-papilionate	Tucker (2003b)
<i>Bocoa viridiflora</i>	un	unknown	non-papilionate	
<i>Bossiaea cordigera</i>	unknown	sg	papilionate	Tucker (2006)
<i>Butea monosperma</i>	sy	sg	papilionate	Tucker (1987)
<i>Cadia purpurea</i>	as	sg	non-papilionate	Tucker (1987)
<i>Cajanus cajan</i>	sy	sg	papilionate	Sinjushin (2023)
<i>Calopogonium sp.</i>	unknown	sg	papilionate	Prenner & Cardoso (unpubl. data)
<i>Calpurnia intrusa</i>	as	unknown	papilionate	Prenner (2004)
<i>Camoensia scandens</i>	un	sg	non-papilionate	Prenner & Cardoso (unpubl. data)
<i>Candolleodendron brachystachyum</i>	un	unknown	non-papilionate	

Chapter 4. An ontogenetic overview of the florally diverse Papilionoideae (Leguminosae): new insights into the evolution of androecium symmetry and petal heterochrony

Species	Androecium symmetry	Petal growth	Flower architecture	Reference
<i>Carmichaelia flagelliformis</i>	sy	unknown	papilionate	Prenner (2004)
<i>Castanospermum australe</i>	sy	fg	non-papilionate	Tucker (1993)
<i>Cicer arietinum</i>	sy	sg	papilionate	Prenner (2002)
<i>Cladrastis sinensis</i>	as	sg	papilionate	Prenner (2002)
<i>Cleobulia coccinea</i>	sy	sg	papilionate	Prenner & Cardoso (unpubl. data)
<i>Cochliasanthus caracalla</i>	sy	sg, vex fg	papilionate	Prenner & Cardoso (unpubl. data)
<i>Cordyla africana</i>	un	ab	non-papilionate	Sinjushin (2018)
<i>Coursetia glandulosa</i>	sy	unknown	papilionate	Prenner (2004)
<i>Coronilla varia</i>	sy	sg	papilionate	Sokoloff et al. (2007)
<i>Crotalaria pallida</i>	as	sg	papilionate	Prenner (2002)
<i>Ctenodon falcatus</i>	sy	sg	papilionate	Sampaio et al. (2013)
<i>Cyathostegia</i>	un	fg, 1 adax only	non-papilionate	Prenner (2004)
<i>Cyclopia genistoides</i>	as	unknown	papilionate	
<i>Dahlstedtia pinnata</i>	sy	sg	papilionate	Teixeira et al. (2009)
<i>Dalbergia brasiliensis</i>	as	sg	papilionate	Klitgaard (1999)
<i>Dalea candida</i>	unknown	sg	non-papilionate	McMahon & Hufford (2002)
<i>Daviesia cordata</i>	sy	sg	papilionate	Prenner (2004)
<i>Dichilus lebeckioides</i>	as	unknown	papilionate	Leite et al. (2014)
<i>Dipteryx alata</i>	as	sg	papilionate	Bento et al. (2021)
<i>Discolobium pulchellum</i>	sy	sg	papilionate	Prenner (2002)
<i>Dorycnium germanicum</i>	sy	sg	papilionate	Prenner & Cardoso (unpubl. data)
<i>Dussia macrophyllata</i>	sy	fg	papilionate	Prenner (2002)
<i>Ebenus cretica</i>	sy	sg	papilionate	McMahon & Hufford (2005)
<i>Errazurizia megacarpa</i>	unknown	sg	non-papilionate	Tucker (2003c)
<i>Erythrina herbacea</i>	sy	sg	papilionate	Mansano et al. (2002)
<i>Exostyles venusta</i>	sy	sg	non-papilionate	Prenner & Cardoso (unpubl. data)
<i>Eysenhardtia orthocarpa</i>	as	sg	non-papilionate	Prenner & Cardoso (unpubl. data)
<i>Fairchildia panamensis</i>	un	unknown	non-papilionate	Prenner & Cardoso (unpubl. data)
<i>Flemingia strobilifera</i>	sy	sg	papilionate	Prenner (2002)
<i>Galega officinalis</i>	sy	sg	papilionate	Tucker (2006)
<i>Gastrolobium truncatum</i>	sy	sg	papilionate	Prenner (2002)
<i>Genista sagittalis</i>	sy	sg	papilionate	Crozier & Thomas (1993)
<i>Glycine max</i>	sy	sg	papilionate	
<i>Goodia lotifolia</i>	sy	unknown	papilionate	
<i>Halimodendron halodendron</i>	sy	unknown	papilionate	Tucker (2006)
<i>Hardenbergia violacea</i>	sy	sg	papilionate	Mansano et al. (2002)

Chapter 4. An ontogenetic overview of the florally diverse Papilionoideae (Leguminosae): new insights into the evolution of androecium symmetry and petal heterochrony

Species	Androecium symmetry	Petal growth	Flower architecture	Reference
<i>Harpalyce brasiliana</i>	sy	sg	papilionate	Tucker (1987)
<i>Hedysarum flavescens</i>	sy	sg	papilionate	Sokoloff et al. (2007)
<i>Hippocrepis comosa</i>	sy	sg	papilionate	Prenner & Cardoso (unpubl. data)
<i>Hymenolobium janeirense</i>	as	sg	papilionate	Prenner & Cardoso (unpubl. data)
<i>Indigofera suffruticosa</i>	sy	sg, vex fg	papilionate	Paulino et al. (2011)
<i>Kennedia nigricans</i>	sy	sg	papilionate	Prenner (2002)
<i>Lablab purpureus</i>	sy	sg	papilionate	Prenner & Cardoso (unpubl. data)
<i>Laburnum alpinum</i>	sy	sg	papilionate	Prenner (2002)
<i>Lathyrus latifolius</i>	sy	sg	papilionate	Prenner (2003b)
<i>Lecointea hatschbachii</i>	sy	sg	non-papilionate	Mansano et al. (2002)
<i>Leptolobium bijugum</i>	sy	sg	non-papilionate	Prenner & Cardoso (unpubl. data)
<i>Lespedeza thunbergii</i>	sy	sg	papilionate	Prenner (2002)
<i>Liparia splendens</i>	as	unknown	papilionate	Prenner (2004)
<i>Lotus corniculatus</i>	sy	sg	papilionate	Prenner (2003a)
<i>Luetzelburgia andradelimae</i>	sy	sg	papilionate	Prenner & Cardoso (unpubl. data)
<i>Lupinus affinis</i>	sy	sg	papilionate	Tucker (1984, 1987)
<i>Maackia amurensis</i>	as	unknown	papilionate	Prenner (2004)
<i>Machaerium villosum</i>	as	fg	papilionate	Klitgaard (1999)
<i>Marina maritima</i>	as	sg	papilionate	McMahon & Hufford (2002)
<i>Medicago truncatula</i>	sy	sg	papilionate	Benlloch et al. (2003)
<i>Melilotus alba</i>	as	sg	papilionate	Hirsch et al. (2002)
<i>Milletia pinnata</i>	sy	sg	papilionate	Tucker (1987)
<i>Mirbelia oxylobioides</i>	sy	sg	papilionate	Tucker (2006)
<i>Mildbraediodendron excelsum</i>	un	ab	non-papilionate	Prenner & Cardoso (unpubl. data)
<i>Monopteryx uauco</i>	sy	fg	non-papilionate	Prenner & Cardoso (unpubl. data)
<i>Myrocarpus frondosus</i>	sy	fg	non-papilionate	Prenner & Cardoso (unpubl. data)
<i>Myroxylon balsamum</i>	sy	fg	non-papilionate	Tucker (1993)
<i>Olneya tesota</i>	sy	sg	papilionate	Prenner & Cardoso (unpubl. data)
<i>Onobrychis melanotricha</i>	sy	sg, vex sw. fg	papilionate	Khodaverdi et al. (2014)
<i>Ononis spinosa</i>	sy	sg	papilionate	Tucker (1987)
<i>Pachyrhizus tuberosus</i>	sy	sg, adax fg	papilionate	Prenner & Cardoso (unpubl. data)
<i>Parryella filifolia</i>	unknown	ab	non-papilionate	McMahon & Hufford (2005)
<i>Pearsonia cajanifolia</i>	as	sg	papilionate	Prenner & Cardoso (unpubl. data)
<i>Petaladenium urceoliferum</i>	sy	fg	papilionate	Prenner et al. (2015)
<i>Petteria ramentacea</i>	sy	sg	papilionate	Prenner (2002)

Chapter 4. An ontogenetic overview of the florally diverse Papilionoideae (Leguminosae): new insights into the evolution of androecium symmetry and petal heterochrony

Species	Androecium symmetry	Petal growth	Flower architecture	Reference
<i>Pisum sativum</i>	sy	sg	papilionate	Tucker (1987, 1989)
<i>Platymiscium floribundum</i>	as	sg	papilionate	Klitgaard (1999)
<i>Podalyria argentea</i>	as	unknown	papilionate	Prenner (2004)
<i>Poecilanthe itapuana</i>	sy	sg	papilionate	Prenner & Cardoso (unpubl. data)
<i>Psoralea pinnata</i>	sy	sg	papilionate	Tucker (1987); Tucker & Stirton (1991)
<i>Psorothamnus kingii</i>	as	sg	papilionate	McMahon & Hufford (2002)
<i>Pterocarpus rotundifolius</i>	as	fg	papilionate	Klitgaard (1999)
<i>Pterodon pubescens</i>	as	sg	papilionate	Leite et al. (2014)
<i>Pultenaea daphnoides</i>	unknown	sg	papilionate	Tucker (2006)
<i>Riedeliella graciliflora</i>	sy	fg	non-papilionate	Bento et al. (2021)
<i>Robinia pseudacacia</i>	sy	unknown	papilionate	Prenner (2004)
<i>Sophora tomentosa</i>	as	sg	papilionate	Tucker (1994)
<i>Spartium junceum</i>	sy	sg	papilionate	Prenner (2002)
<i>Styphnolobium japonicum</i>	sy	sg	papilionate	Tucker (1987); Prenner (2002)
<i>Swainsona formosa</i>	sy	sg	papilionate	Prenner & Cardoso (unpubl. data)
<i>Swartzia aureosericea</i>	un	fg, 1 adax only	non-papilionate	Tucker (2003)
<i>Taralea oppositifolia</i>	as	sg	papilionate	Leite et al. (2014)
<i>Tetragonolobus purpureus</i>	sy	sg	papilionate	Tucker (1987)
<i>Thermopsis lanceolata</i>	as	sg	papilionate	Prenner (2002)
<i>Trifolium repens</i>	sy	sg	papilionate	Tucker (1987)
<i>Trischidium molle</i>	un	sg, 1 adax only	non-papilionate	Prenner & Cardoso (unpubl. data)
<i>Ulex europaeus</i>	sy	sg	papilionate	Prenner (2002)
<i>Vicia faba</i>	sy	sg	papilionate	Prenner & Cardoso (unpubl. data)
<i>Vigna radiata</i>	sy	sg	papilionate	Tucker (1987)
<i>Virgilia divaricata</i>	as	unknown	papilionate	Prenner (2004)
<i>Wisteria sinensis</i>	sy	sg	papilionate	Naghiloo & Dadpour (2010)
<i>Xiphotheca tecta</i>	as	unknown	papilionate	Prenner (2004)
<i>Zollernia ilicifolia</i>	sy	sg	non-papilionate	Mansano et al. (2002)

Chapter 5

A taxonomic synopsis of *Aldina*, a florally distinctive and poorly collected Amazonian genus of papilionoid legumes

Introduction

This chapter is a taxonomic synopsis of the Amazonian papilionoid legume *Aldina*, which is a genus of Papilionoideae with unusual non-papilionoid flower morphology. Flowers of *Aldina* are radially symmetrical, with free numerous stamens, unlike its sister genera *Andira* and *Hymenolobium*, whose zygomorphic flowers are typically papilionate and with 10 stamens. *Aldina* taxonomy is in need of a revision since the preliminary taxonomic studies on the genus that date back 70 years (Cowan, 1953, 1958, 1961, 1984; Gontscharov, 2006, 2009). This chapter therefore contributes to understanding the species diversity in Amazonia, filling in the gaps of plant diversity in this under-explored region. Eighteen species are recognized in *Aldina*, with seven names newly placed in synonymy, new lectotypification of five names, and a new combination, *Aldina auyantepuiensis*, is proposed. This chapter has been published in Systematic Botany: Ramos, G.A., Zartman, C.E., Lima, H.C., Pennington, R.T., Cardoso, D.B.O.S. (2022). Taxonomic Synopsis of *Aldina*, a Florally Distinctive and Poorly Collected Amazonian Genus of Papilionoid Legumes. Systematic Botany 47: 1-16. DOI:10.1600/036364422X16674053033859.

Cataloging and describing the plants in species-rich regions such as the vast tropical Amazon lowland rainforest are fundamental to understanding patterns of diversity and are the first step toward conservation prioritization (Hopkins 2007; Cardoso et al. 2017). However, much taxonomy remains to be done in Amazonia, partly because huge remote regions are unexplored (Cardoso et al. 2015a), and when specimens exist, they can remain hidden for decades in herbaria without being studied by taxonomists (Bebber et al. 2010). In the face of accelerating habitat destruction, thorough monographic revisions of plant groups in poorly-collected regions are therefore urgent priorities.

The taxonomy of many Amazonian genera of the early-branching papilionoid legumes remains understudied in the 21st century, despite legumes being the family that dominates Amazonia in terms of species diversity and abundance (ter Steege et al. 2013; Cardoso et al. 2017). Examples of genera that need taxonomic revision include *Dussia* Krug & Urb. ex. Taub, *Dipteryx* Schred., and the little-known neotropical genus *Aldina* Endl. that is the subject of this synopsis.

Preliminary taxonomic studies on *Aldina* (Cowan 1953, 1958, 1961) were based on specimens gathered in Colombia, Venezuela, and Guyana, and less is known about

species distributed in Brazil. Since then, more species have been described, increasing our knowledge of the genus (Gontscharov et al. 2006, 2009; Stergios & Aymard 2008; Aymard 1998; Cowan and Steyermark 1984; Steyermark 1967, 1999). However, most of the new names published have been based on single specimens, which often fell under the morphological continuity of previously described species. With the possibility of species number overestimation, a taxonomic revision of the names proved essential.

Here, we have comprehensively revised the taxonomy and nomenclature of all *Aldina* species by analyzing the available collections, including historical specimens and types, as well as newly collected specimens from across the Brazilian Amazon. In this synopsis, we have updated the species delimitation and descriptions, especially for those previously understudied and taxonomically confusing because they were represented only by a few specimens or the type. The taxonomic treatment presented here includes an identification key and full synonymy for all *Aldina* species, plus taxonomic comments that include information on how to distinguish morphologically similar species. Our study is a starting point toward a comprehensive taxonomic revision of the entire genus in the Neotropics.

Aldina is a neotropical genus of 18 species distributed in Amazonia, encompassing regions of Colombia, Venezuela, Guyana, and Brazil (Fig. 1). Collections from the state of Acre, close to Peru and Bolivia, suggest that the genus might have its distribution expanded to those countries when some remote peripheral areas of the Amazon Basin have been explored. *Aldina* species are trees up to 30 m tall, often buttressed, with odd-pinnate leaves, radially symmetric flowers with undifferentiated petals, calyx entire in flower bud and splitting in irregular lobes as the bud opens, free and numerous stamens, and tardily-dehiscent nucoid fruits (Barroso, 2004) with an undifferentiated endocarp, the mesocarp occasionally fibrous or spongy (Fig. 2). In a family largely marked by the evolution of root-symbiotic nitrogen fixation (Werner et al. 2014; Sprent et al. 2017), the genus *Aldina* is among the few ecologically successful lineages that nodulate and also associate with ectomycorrhizae.

5.1 Taxonomic History—*Aldina insignis* is the type species and was described in 1840 by Stephan F. L. Endlicher, who named it after the doctor and botanist Tobias Aldinus. In the same year, Bentham (1840) had described the genus *Allania* based on the same plant, naming it after the botanist Allan Cunningham. The name “*Alania*”, a close orthographic variant of *Allania*, however, had been previously assigned by Endlicher

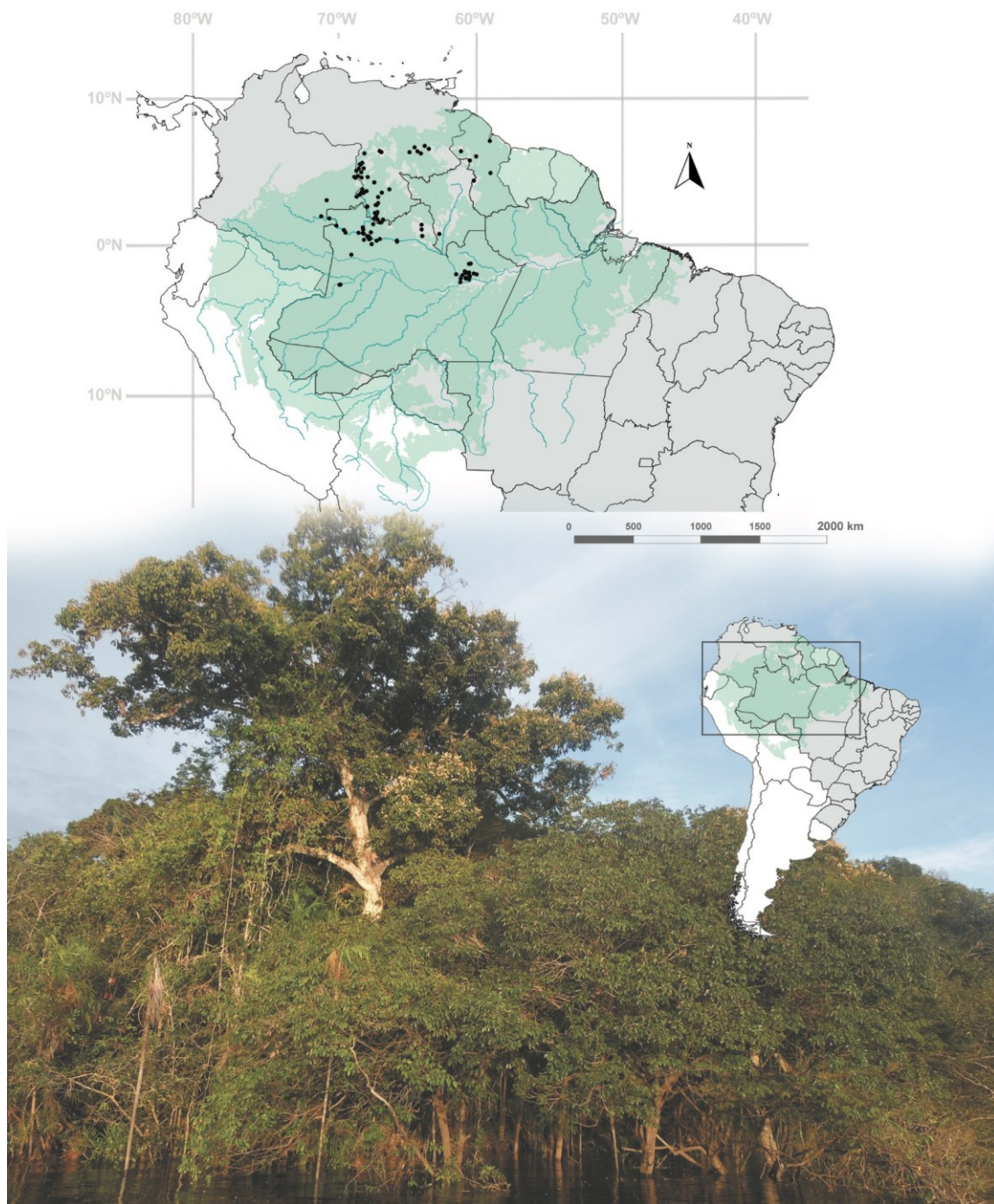


FIG. 1. Distribution of *Aldina* in the Neotropics (Colombia, Brazil, Guyana, and Venezuela). Collections are represented by black dots. The green shaded area represents the limits of the Amazonian lowland rain forest biome as delimited by Cardoso et al. (2017). A flowering individual of *Aldina latifolia* during the rainy season in the Rio Negro is shown in the background.

(1836) to a species in the family Liliaceae. This apparent ‘slip of the pen’ (Rehder et al. 1935) was corrected by Meisner (1842).

The name *Aldina* has been used by other authors over time. In the Leguminosae, a homonym described by Adanson (1763) was based on material of the papilionoid genus *Brya* P.Browne (Browne 1756). Since Adanson’s (1763) name has not been adopted, it has remained without an associated specific name and was rejected based on the Art. 14 of the International Code of Nomenclature for algae, fungi, and plants (Rehder et al. 1935; Turland et al. 2018).

Rehder et al. (1935) also suggested in the same proposal the rejection of the name *Aldinia* Scop., a synonym of *Justicia* L. (Acanthaceae), which has not been adopted and also remains without a specific name.

In view of the fact that *Aldina* Endl. has been universally accepted and has no synonym to take its place, *Allania* Benth. (1840) being a later homonym of *Alania* Endl. (1836), the former name has been conserved.

Aldina has been classified historically in the tribe Swartzieae which, along with the tribe Sophoreae, comprise genera of Papilionoideae that retain a set of floral traits traditionally considered plesiomorphic in the family, such as buds with an entire calyx splitting into irregular lobes as the flower opens, radially symmetric or non-papilionate flowers with unspecialized or undifferentiated petals, and free and numerous stamens (Polhill 1981; Ireland 2005).

5.2 Phylogenetic Affinities—The phylogenetic placement of *Aldina* came into question in molecular phylogenetic studies on the early-branching Papilionoideae (Doyle et al. 1997; Pennington et al. 2001; Cardoso et al. 2012, 2015b), which have consistently diverged from traditional morphology-based classifications. *Aldina* was first sampled in a molecular phylogeny of the Leguminosae based on plastid *rbcL* sequences, which placed the genus in a large clade comprised of genera with a 50-kb inversion in the chloroplast genome (Doyle et al. 1997). Further phylogenetic analysis of plastid *trnL* intron sequences, involving comprehensive sampling of the tribes Sophoreae and Swartzieae, placed *Aldina* among the early-branching Papilionoideae, all of them known

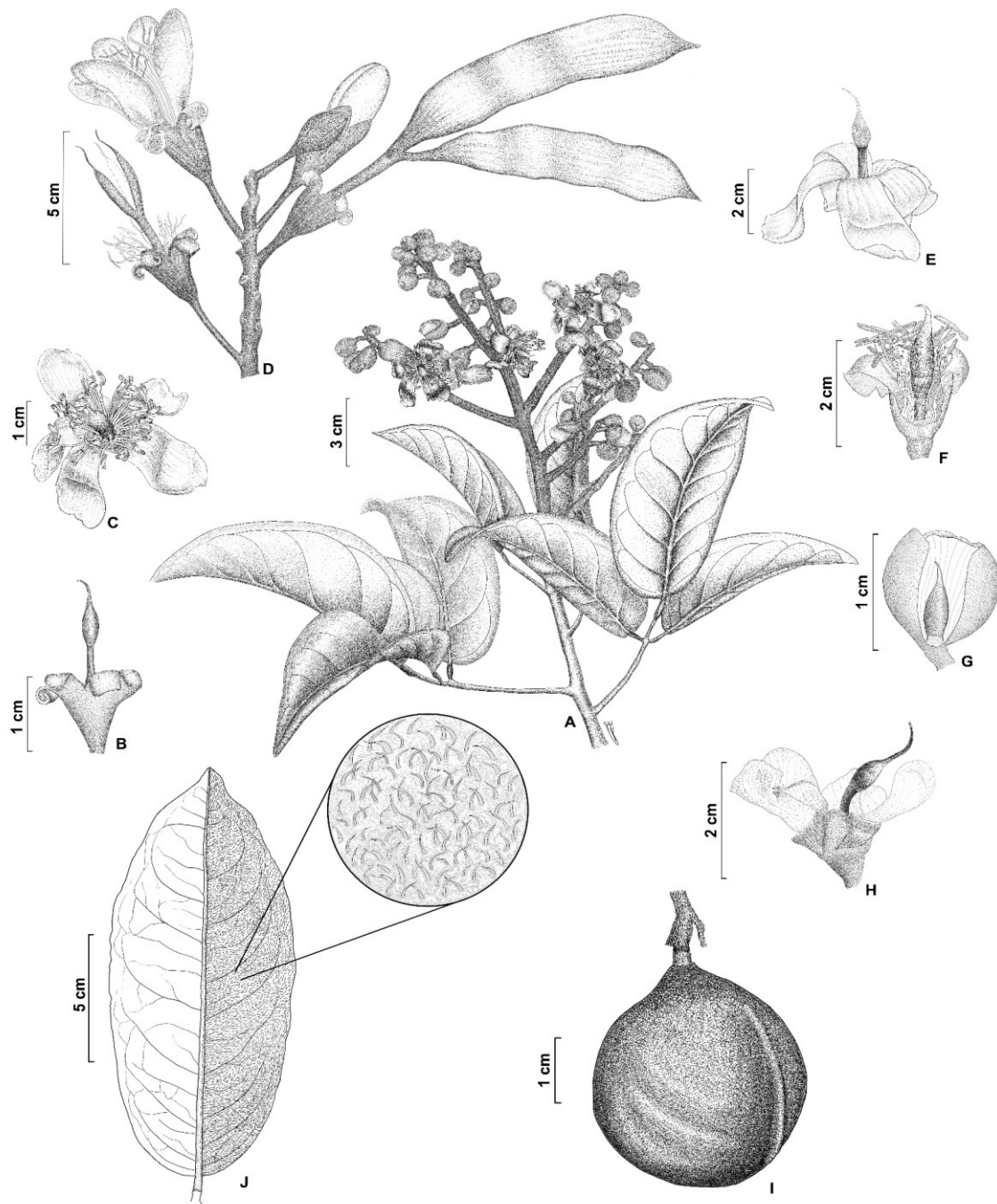


FIG. 2. Diagnostic characters of *Aldina* species. A. *A. heterophylla*. Fertile branch with the typical reflexed flower buds in the inflorescence axes and navicular (boat-shaped) leaves. B–H. Flowers (most of them with petals, stamens, or part of the calyx removed) showing the variation in ovary morphology across the genus. B. *A. berryi* is one of the two species

in the genus with glabrous ovary, occasionally with few hairs at its base. C. *A. heterophylla* has short appressed ferruginous hairs covering the stipe, joint, and ovary, whereas the style is glabrous; the species also bears the lowest number of stamens in the genus (ca. 36 vs. ca. 60). D. *A. diplogyne* has the largest flowers in the genus (ca. 6 cm long) and is the only species with multicarpellate gynoeceium, bearing up to (2)3–4 appressed-ferruginous-pubescent functional pistils per flower. E. *A. latifolia* has a gynoeceium in which only the style is glabrous; the stipe, joint, and ovary are covered with appressed-pale golden hairs. F. *A. kunhardtiana* has a calyx with arachnoid hairs on the inner side, the outer side is readily distinguishable by the velutinous dark-brown-ferruginous indumentum; also its gynoeceium is covered with appressed-ferruginous hairs, except for the style which is glabrous and is often recurved at the tip. G. The sessile ovary of *A. occidentalis* with ca. 1 mm long glabrous stipe readily distinguishes the species from the others in the genus. H. *A. macrophylla* has a unique gynoeceium with a stipe and joint covered with appressed-golden hairs and a glabrous dark brown ovary. I. The glabrous nucoid fruit of *A. heterophylla* often displays a subglobose to globose shape throughout its development. Other species of the genus tend to have oblong fruits in the early developmental stages such as shown here in *A. diplogyne*. J. The leaflet of *A. kunhardtiana* showing the conspicuous venation on the abaxial face and a close-up of the decumbent to erect trichomes intermixed over a flaking coat of white wax.

for lacking the 50-kb plastid inversion (Ireland et al. 2000; Pennington et al. 2001). Therefore, an Aldinoid clade comprised of *Aldina*, *Amburana* Schwacke & Taub. (2 spp.), and *Mildbraediodendron* Harms (1 sp.) was defined within the Papilionoideae (Ireland et al. 2000).

The incongruence between the phylogenies derived from *rbcL* (Doyle et al. 1997) and *trnL* intron (Ireland et al. 2000; Pennington et al. 2001) sequences was partially resolved by phylogenetic reassessments of *Aldina* from newly acquired *matK* and *trnL* intron sequences, which led to the conclusion of a sampling error in the previous analysis of *trnL*. These analyses confirmed *Aldina* in the 50-kb inversion clade (Cardoso et al. 2012, 2015b), but the resulting phylogenies remained poorly resolved. *Aldina* has since been placed in a large polytomy with other small clades (*Amphimas*, *Andira*, *Dermatophyllum*, Lecointeoid, Vataireoid), the species-rich lineages Dalbergioid,

Genistoid, and the NPAAA (non-protein amino acid accumulating) clade (Cardoso et al. 2012, 2013, 2015b).

In a recent, more comprehensive phylogenetic study of the early-branching papilionoid lineages, more light was shed on the phylogenetic placement of *Aldina* by analyzing combined data of the plastid loci *rbcL*, *trnL* intron, and *matK* (Ramos et al. 2016). Unexpectedly, in this study, *Aldina* was resolved as sister to the papilionate-flowered genera *Andira* and *Hymenolobium*, which compose the Andira clade. Such a relationship was also confirmed by a plastome-scale phylogenetic analysis of the Papilionoideae (Choi et al. 2022). Although *Aldina*'s placement is now resolved in the Papilionoideae phylogeny, no study has considered species relationships within the genus.

5.3 Geographic Distribution, Habitat and Ecology—*Aldina* is confined to the Amazon basin, which covers a vast area of relatively poorly-botanized lowland tropical rainforest in Brazil (Hopkins 2007; Cardoso et al. 2017), a country recognized for its high floristic diversity and endemism (Forzza et al. 2012). Although *Aldina* is a poorly collected genus, some species are broadly distributed in the Amazon basin. An example is *Aldina discolor*, found in the patchily distributed white-sand-forests (*campinas* and *campinaranas*) of Brazil, Colombia, and Venezuela. A recent ecological estimate listed *A. discolor* as one of the 227 most dominant species that make up half the individual trees in the Amazon (ter Steege et al. 2013). Moreover, in the Guiana Shield, *Aldina* is listed among the 10 most abundant genera (ter Steege et al. 2006).

Furthermore, other species of *Aldina* can be found in the black water floodplains (*igapós*), upland (*terra-firme*) (Fig. 3), and montane forests of the Amazonian tepuis along the Guiana Shield Region. Although at first sight the Amazon may look like an endless homogeneous type of vegetation, it harbors a mosaic of ecologically distinct vegetation physiognomies that are shaped by many factors such as altitude, river drainage, soil composition, and precipitation (Oliveira-Filho et al. 2021; Pires and Prance 1985). As shown for other Amazonian plant taxa (Fine et al. 2010; Fine and Baraloto 2016; Draper et al. 2016), the distribution of *Aldina* species seems to be constrained by their preferences for particular habitat types such as the periodically-flooded Amazonian black water forests (*igapós*), which demand physiological and morphological adaptations. For example, *A. latifolia* Spruce ex Benth., a widely distributed species along the Negro and

Vaupés Rivers, has evolved a highly specialized, dramatically enlarged nucoid fruit with a spongy mesocarp as an adaptation for flotation along rivers and streams. The combination of such features optimizes dispersal capacities as the seeds germinate only once the fruits have settled on a riverbank during the dry seasons. However, not all species of *Aldina* are adapted to flooded habitats. In fact, most of them are found in non-flooded habitats, from low altitudes to the higher elevations of montane forests. For example, *A. microphylla* M.Yu.Gontsch. & Yakovlev occurs over the Serra do Aracá tepui at elevations reaching 800 m. Other species such as *A. kunhardtiana* R.S.Cowan and *A. polyphylla* Ducke are found in the upland (*terra-firme*) forests: the most predominant vegetation of the Amazon basin, comprising ca. 80% of its total area (Oliveira-Filho et al. 2021).

Aldina heterophylla is also found in white-sand forests, but it is only known from the vicinities of Manaus and Presidente Figueiredo, which makes it difficult to evaluate its actual predominance in these forests. The white-sand forests are patchily distributed through the matrix of the *terra-firme* forests, comprising 5–10% of the entire Amazon's coverage (Stropp et al. 2011). This unusual habitat, regionally called *campina* (open vegetation) or *campinarana* (canopy formed) in Brazil, has the most nutrient-poor soils known (Fine and Baraloto 2016; Janzen 1974) as the result of leaching. Such challenging environmental conditions, along with their island-like configuration, are often invoked to explain the lower species diversity of white-sand forests. Nevertheless, a high proportion of plant species (42%) are restricted to that vegetation type (Oliveira-Filho et al. 2021; Stropp et al. 2011; Vicentini 2016; Fine and Bruna 2016; Fine and Baraloto 2016). Therefore, white-sand floras have a unique functional composition (Fortunel et al. 2014), suggesting important differences in ecosystem processes related to carbon and nutrient cycles.

Aldina discolor and *A. heterophylla* have achieved ecological success over the nutrient-poor soils of the white-sand forests because of their association with ectomycorrhizal fungi, particularly *Cantharellus* Adans. ex Fr. (Henkel et al. 2002; ter Steege et al. 2006; Smith et al. 2011; Roy et al. 2016). Ectomycorrhizal symbioses are more prevalent on white sands than elsewhere in the lowland Amazon, they confer to the host the ability to acquire more nutrients and water, and hence increase the chances of plant species to thrive (Roy et al. 2016).

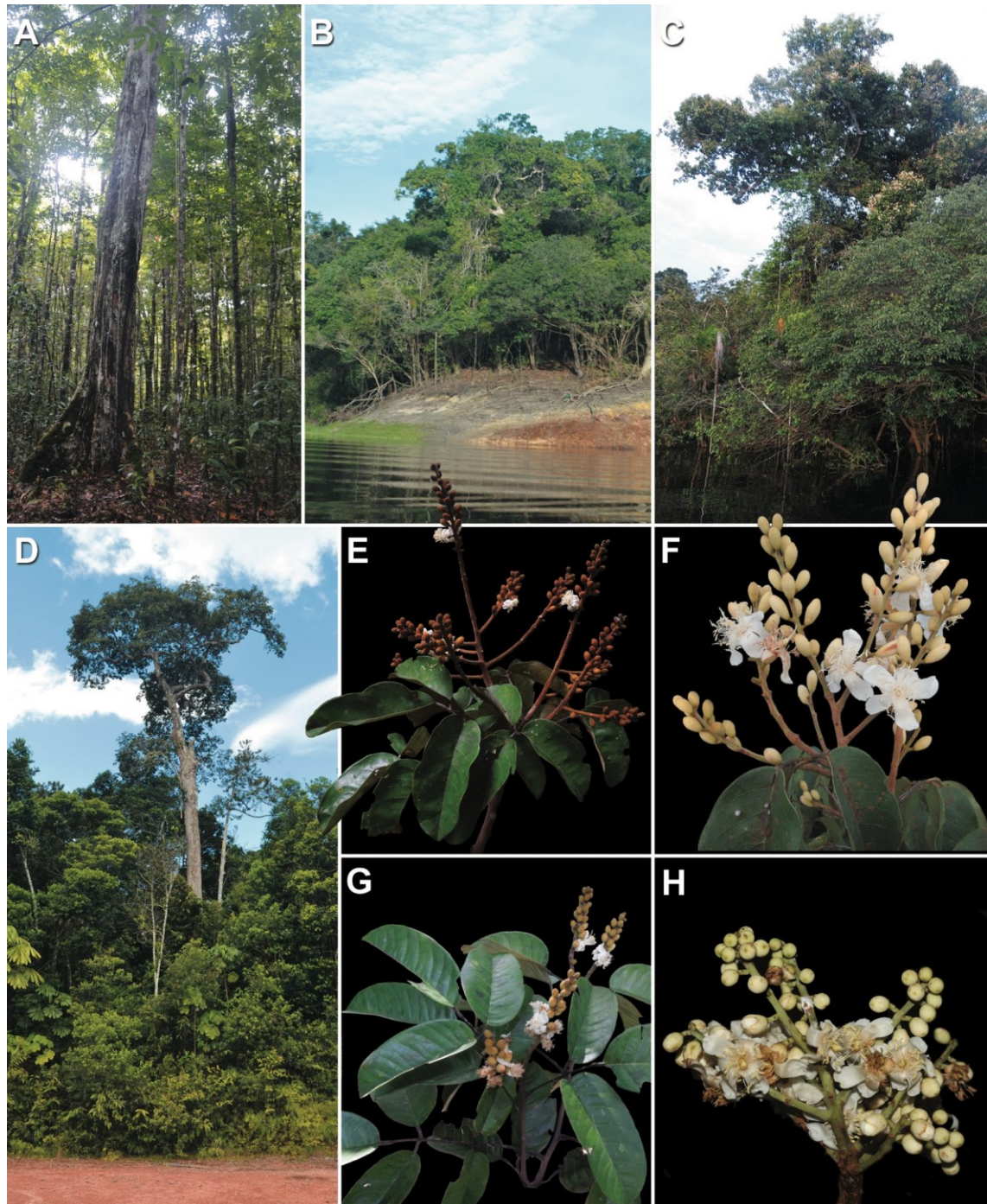


FIG. 3. Habitat and morphological characterization of *Aldina* species. A. White-sand forest (*campinarana*) with an individual of *A. discolor* at São Gabriel da Cachoeira, Brazil. B–C. Seasonally flooded forest (*igapó*) in the same locality in the dry and rainy season, respectively, from the black waters of Rio Negro in Manaus, Brazil. A flowering individual of *A. latifolia* with pale branches is shown by a red arrow. D. Terra-firme forest at São Gabriel da Cachoeira, Brazil showing its typical clay soil and an outstanding individual of *A. kunhardtiana*. E. Inflorescence of *A. kunhardtiana* showing its typical velvety rusty-brown axes. F–G. The pale yellowish flower

buds of *A. latifolia* and *A. discolor*, respectively, are readily distinguishable from the green background of the forest. H. Inflorescence of *A. heterophylla* showing its typical deflexed flower buds.

Understanding the complex taxonomy of *Aldina* is particularly important because of its ecological importance in the species rich, yet relatively poorly-collected Amazon rain forest: a biome globally recognized by its exceptionally high floristic diversity, endemism, and ecological importance (Cardoso et al. 2017). The taxonomic synopsis of *Aldina* presented here addresses the status of dubious taxa, geographic distribution, and identification of all species in the genus in order to shed light on diversity patterns and as the first step for conservation prioritization.

MATERIALS AND METHODS

A thorough survey of physical herbarium specimens plus virtual specimen images was conducted to gather information for the taxonomic descriptions, flowering and fruiting phenologies, habitat, and geographical distributions. We revised specimens in the following herbaria known to hold significant collections of *Aldina* (acronyms according to Thiers et al. 2015 [continuously updated]): ALCB, COL, E, F, HUEFS, IAN, INPA, K, MO, NY, OXF, P, R, RB, and US. Other collections were revised from virtual databases such as specieslink.net and plants.jstor.org.

The general terminology for the morphological descriptions (e.g., phyllotaxy, leaf shape, indumentum, venation, and fruit types) is based primarily on Harris and Harris (1994) and Barroso (2004). Specific terminology (e.g., leaf division, flower morphology, and habit) follows LPWG (2017) and other recent taxonomic revisions of papilionoid legume genera (e.g., Klitgaard 2005; Pirie et al. 2009; Cardoso et al. 2014).

We carried out extensive field research in remote and underexplored areas in the upper Rio Negro and Vaupés of western Amazonia, where new records of *Aldina* species were made along with photographs, including of habitats. Preparation of plant specimens followed Mori et al. (2011). All collected material was deposited at ALCB herbarium, Universidade Federal da Bahia (UFBA). Duplicate specimens were sent to HUEFS,

INPA, and RB. Flowers and leaves obtained during our fieldwork were fixed in 70% ethanol to enable accurate examination of the specimens for description and illustrations.

For the nomenclatural revision, the type designations and all the nomenclatural decisions followed the International Code of Nomenclature for algae, fungi, and plants (ICN, Turland et al. 2018).

5.4 TAXONOMIC TREATMENT

ALDINA Endl., Gen. Pl. 1322. 1840, nom. cons., Bull. Misc. Inform. Kew. 350. 1935. *Allania* Benth., J. Bot. (Hooker) 2: 91. 1840.—TYPE: *Aldina insignis* Endl., Repert. Bot. Syst. 1(5): 843. 1843; non *Alania* Endl. (1836) (Liliaceae).

Aldina E. Mey., nom. illeg., Comm. Pl. Afr. 171. 1835. = *Vachellia* Wight & Arn.

Aldina Adans., nom. rejec., Civ. Nat. Hist. Jamaica. 299. 1756. = *Brya* P.Browne

Trees, frequently with buttresses. **Roots** nodulating and ectomycorrhizal-associated. **Stems** frequently releasing red sap when cut. **Leaves** odd-pinnate, often 1-foliolate when nearer the inflorescences, commonly 3–9(–11)-foliolate, exstipulate; petiole and rachis terete, commonly green, sometimes darkening; leaflets opposite to subopposite, or sometimes clearly alternate, mostly coriaceous, sometimes chartaceous, base rounded to cordate, a subtle sometimes dense white wax coating beneath; venation brochidodromous. **Inflorescences** terminal or axillary, racemose or paniculate-racemose, densely covered by minute appressed hairs including the outer face of calyx; bracts and bracteoles minute; pedicels short. **Flowers** bisexual, actinomorphic (radially symmetrical), fragrant; calyx glabrous on the inner surface except for crisped villous hairs on tip of the lobes, entire in bud, opening about halfway in 2–4 deflexed, irregular segments, tube turbinate; petals (3–)4–6, imbricate in the bud, white to yellowish, glabrous, oblong, obovate and oblanceolate, usually concave to cucullate with folding impressions from stamens at the tips; stamens numerous, mostly ca. 60, rarely around 36, glabrous, white, essentially free, although commonly with a sheath at the base of ca. 3 mm long, laterally adnate to the hypanthium cup, the anthers large, uniform, narrowly oblong to linear, opening lengthwise by slits; gynoecium 1–3 carpellate, arising from the

bottom of hypanthium cup, sessile or more often borne on a distinct stipe, covered with appressed golden hairs or glabrous, swollen-articulate at junction with the somewhat oblong, 2–5-ovulate ovary, covered by golden appressed hairs or entirely glabrous; style glabrous, subulate, straight or usually hooked; stigma punctiform. **Fruit** a nuroid legume, 5–9 cm long, 4–6 cm diameter, tardily dehiscent, thick, spongy when adapted to water dispersal, globose to subglobose, glabrous to tomentose; 1–4-seeded.

KEY TO *ALDINA* SPECIES

1. Leaflets completely glabrous on both surfaces; terminal (occasionally axillary) inflorescence with single racemes growing from the leaf axis or multiple racemes growing from the tip of the branch 2
1. Leaflets pubescent on the lower surface, sometimes obviously but often minutely so; inflorescence in elongated, open, multiple racemes or a terminal panicle 8
2. Flowers showy, calyx tube 2.5–3 cm long, petals 3–4 × 2 cm; gynoecium 2–4-carpellate, i.e., multiple ovaries born on a single stipe; fruits elongated, sausage-shaped to thickly fusiform and noticeably constricted between the 2–4 seeds; leaflets coriaceous, cordate at the base *A. diplogyne*
2. Flowers half as large or smaller than above; gynoecium unicarpellate, i.e. a single ovary born on the stipe (see *A. insignis*); fruits ovoid to oblong, 1–2-seeded; leaflets chartaceous to coriaceous, but not cordate at the base 3
3. Leaflets shiny on the lower surface, broadly ovate to oblong-ovate; calyx and unopened buds sparsely pale yellowish puberulent with appressed trichomes; calyx tube truncate, claviform at the base; fruit semi-globose to ellipsoid *A. latifolia*
3. Leaflets pallid, rarely shiny on the lower surface, oblong to ovate-elliptic, obtuse-mucronate; calyx and unopened buds uniformly densely pubescent; calyx tube truncate, barrel-shaped at the base; fruit ovoid or not as above 4
4. Mature buds 7–8 × 5 mm, apiculate and commonly deflexed with respect to the main inflorescence axis; leaflets shiny on both surfaces; inflorescences in panicles *A. heterophylla*

4. Mature buds 10–20 × 7–10 mm, not apiculate and not deflexed (commonly patent); leaflets pallid on lower surface; racemes 5
5. Ovary mostly glabrous and black; leaflets chartaceous, gray-greenish on lower surface *A. berryi*
5. Ovary pubescent; leaflets subcoriaceous to coriaceous, silvery gray-greenish or opaque-glaucous on lower surface 6
6. Leaves mostly unifoliolate; fruits retaining pubescence in later stages of development *A. auyantepuiensis*
6. Leaves 3–5-foliolate; fruit glabrescent towards late stages of development 7
7. Leaflets subcoriaceous, 8.8–12 × 3.5–4.5 cm, oblong-ovate, base obtuse-cuneate, apex acuminate; inflorescence and flower buds glabrous, stipe 5–12 mm long
A. barnebyana
7. Leaflets coriaceous, 4.5–7 × 2.2–3.8 cm, elliptic, oblong or ovate, base obtuse to rounded, apex acute to obtuse, acuminate, the acumen 1–1.5 cm long; inflorescence and flower buds tan-brown, appressed-pubescent, stipe 3.5 mm long *A. microphylla*
8. Ovary glabrous or glabrescent 9
8. Ovary uniformly densely pubescent 10
9. Gynoecium as long as or longer than the stamen filaments at anthesis; leaflets appressed-puberulent on lower surface; ovary glabrescent; petals 5, 15–20 × 6–8 mm *A. latifolia*
9. Gynoecium noticeably shorter than the stamen filaments at anthesis; leaflets appressed-puberulent, somewhat uniformly glaucous on lower surface; ovary glabrous; petals 4, 20–25 × 10 mm *A. macrophylla*
10. Leaflets brown-villous with erect trichomes on lower surface; inflorescence notably dark, golden-brown-velutinous; calyx splitting in 2 lobes *A. kunhardtiana*

- 10. Leaflets appressed-puberulent, either pale or with a pruinose, waxy coat on lower surface; inflorescence and flower buds tan-brown, appressed-puberulent; calyx splitting in 3–5 lobes 11
- 11. Stipe subsessile, not longer than 1 mm long 12
- 11. Stipe longer than 1 mm long 13
- 12. Leaflets 10–12 × 6–7.5 cm, obtuse-mucronulate to rounded, pale on the lower surface with a pruinulose, waxy coat; calyx tube cupular, 7–8 × 9 mm at anthesis; petals ca. 15 mm long and strongly cucullate; stipe 1 mm long *A. occidentalis*
- 12. Leaflets 16–17 × 7–8 cm, obtuse-acuminate, concolorous and slightly shiny, waxy coat not prominent on lower surface; calyx tube funnel-shaped, 10 × 8–9 mm at anthesis; petals 20–25 mm long and somewhat flattened; stipe 7–15 mm *A. latifolia*
- 13. Petals 3 or 4 (6 in *A. elliptica*) 14
- 13. Petals 5 or 6 15
- 14. Leaflet venation strongly raised, conspicuous on both surfaces; petals 2–2.5 × 1–1.8 cm; leaflets 7.5–18 × 4.5–9 cm, oblong to oblong-ovate, scantily appressed-puberulent but shiny and the same color on the lower surface *A. reticulata*
- 14. Leaflet venation inconspicuous, usually obscure on upper surface; petals 1–2 cm long; leaflets 8–9 × 5–6 cm, elliptic, obtuse to rounded at the apex, glaucous-beneath *A. elliptica*
- 15. Leaflets acute to acuminate apically, markedly discolor 16
- 15. Leaflets obtuse to rounded-obtuse, concolorous to only slightly discolor 17
- 16. Leaflets tessellate and glaucous appressed-puberulent on lower surface; petals 5, 2–2.3 cm long *A. discolor*
- 16. Leaflets not tessellate, golden-pubescent on lower surface; petals 6, ca. 1.8 cm long *A. aurea*

17. Branchlets, petioles, rachis, and pulvinules densely golden pubescent, glabrous when mature; the pulvinules 10–25 mm long; leaflets broadly ovate to orbicular; petals 5 *A. paulberryi*
17. Branchlets, petioles, rachis, and pulvinules glabrous; the pulvinules 8–20 mm long; leaflets oblong; petals 5 18
18. Leaflets oblong or oblong-ovate, 11–22 × 5.5–10.5 cm; inflorescences in racemes, sometimes with multiple axes growing from the same point, flowers 3–5 × 2.5–5.5 cm *A. insignis*
18. Leaflets elliptical or lanceolate-oblong, 8.5–10 × 3.5–4.5 cm; inflorescences paniculate, flowers 1.5–3 × 1–2.5 cm, elongate 19
19. Leaflets lanceolate-oblong, 8–16 × 3–4 cm; inflorescences loosely paniculate, ca. 30 cm long, pulvinules 10 mm long *A. polyphylla*
19. Leaflets elliptical, 8.5–10 × 3.5–4.5 cm; inflorescences paniculate, 15–23 cm long, pulvinules 15–20 mm long *A. petiolulata*

ALDINA AUREA R.S.Cowan, Mem. New York Bot. Gard. 10 (1): 145. 1958. TYPE: VENEZUELA. Terr. Amazonas, Along left fork of Caño Yutaje, Cerro Yutaje, 1250 m, 12 Feb 1953, *B. Maguire & C.K Maguire 35186* (holotype: NY barcode NY00231174!; isotypes: F photo barcode F0058821F!, K barcode K000500850!, S photo barcode S-R-9321!, US barcode US00002864!, VEN photo barcode VEN43757!).

Distribution and Habitat—This species is only known from Venezuela, where the type collection was made at the tepui Cerro Yutaje at 1250 m elevation.

Taxonomic Comments—*Aldina aurea* can be readily distinguished from all other species in the genus by the conspicuously glaucous lower surface of the leaflets, a character that is otherwise only found, though more discreetly, in *A. discolor* and *A. elliptica*. The latter species was described from the same tepui as *A. aurea*. The typical tessellate pattern on the abaxial surface of the broadly oblong leaflets of *A. discolor* are

not found in *A. aurea*. *Aldina elliptica* has markedly elliptical leaflets, in comparison to the oblong-lanceolate to somewhat narrowly oblong leaflets of *A. aurea*.

Aldina auyantepuiensis (Pittier ex H.S.Irwin) G.Ramos & D.B.O.S.Cardoso, **comb. nov.**, Third Conf. Interamer. Agric. Caracas, 365. 1945. TYPE: VENEZUELA. Bolivar: Cumbre del cerro Guaiquinima, a lo largo del afluente del río Carapo, 1 km río arriba del Salto Szczerbanari, 5°44'4"N, 63°4'8"W, parte sur-oriental del cerro, 730–750m, 23-24 May 1978 (fl, fr), *J.A. Steyermark, P. Berry, G.C.K. & Dunsterville 117245* (lectotype [here designated]: US barcode US00459404!; isolectotype: MO barcodes MO2781106 & MO2781107!).

Aldina latifolia Benth. var. *auyantepuiensis* Pittier ex H.S.Irwin, Acta Bot. Venez. 2(5–8): 225. 1967. TYPE: VENEZUELA. Bolivar: Auyan-tepui, vecindades de Guayaraca, en la primera (inferior) meseta (hombrilo) arriba del valle de Kamarata, seccion pedregosa, en bosque enano, 1,000 m, 18 May 1964, *J.A. Steyermark 94191* (holotype: NY!; isotypes: F barcode F0058831F photo!, K barcode K000500828!, US barcode US00002859!), **syn. nov.**

Aldina auyantepuiensis Pittier, Third Conf. Interamer. Agric. Caracas, 365. 1945. TYPE: VENEZUELA. Guayana: altiplanice del Auyantepui, 1000 m., sobre arenizcas, 5 May 1937, *F. Cardona 41* (holotype: NY!), nom. nudum.

Aldina auyantepuiensis is a tree up to 20 m tall and is readily distinguished from all other *Aldina* species by the unique combination of unifoliolate, ovate-oblong leaves measuring 8–15 × 3–7.5 cm, these often clustered at the tips of the branchlets, commonly close to the inflorescences; the pubescence of the fruit from development to maturity is also a noteworthy trait that distinguishes the species from its tepui counterparts.

Distribution and Habitat—This species is known from three tepuis in Venezuela, growing on the summits of Auyan tepui and its two neighboring tepuis Guaiquinima and Acopán at elevations ranging from 730–1600 m elevation.

Representative Specimens Examined—**Venezuela**—BOLIVAR: Cumbre del cerro Guaiquinima, a lo largo del afluente del río Carapo (1 km río arriba del Salto

Szczerbanari) y los penascos de arenisa al sur del río, 5°44'4"N, 63°41'8"W, parte sur-oriental del cerro, 730–750 m, 23-24 May 1978 (fl, fr), *J.A. Steyermark, P. Berry, G.C.K. & E. Dunsterville 117301* (US); Mount Auyantepui, among sandstone rocks, 1000 m, 5 May 1937 (fl), *F. Cardona 41* (US); Icabarú, carretera Santa Elena de Uairen, km 90, 14-16 Dec 1969 (fl), *L.M. Berti & J.B.B. 2508* (US); Auyantepui, crece en pequeño bosque en la terasa de Guayaraca, 1100 m, Apr 1956 (fl), *Vareschi & Foldats 4687* (NY); Caroni, Guayana, Cerro Acopán, 1600 m, Oct 1947 (st), *F. Cardona 2258* (US).

Taxonomic Comments—*Aldina auyantepuiensis* was first noted as a distinct species by Pittier et al. (1945). Pittier's binomial is a *nomen nudum* according to Art. 38 of the International Code of Nomenclature for algae, fungi, and plants (Turland et al. 2018), because the publication does not provide a description or diagnosis of the taxon in the protologue nor by reference to a previous valid publication. Therefore, the name *Aldina auyantepuiensis* Pittier is invalid.

Later, Irwin (1967) recognized *A. auyantepuiensis* as the variety *Aldina latifolia* var. *auyantepuiensis* Pittier ex H.S.Irwin and used the combination of unifoliolate leaves and the pubescence of the developing fruit to differentiate the new taxon from the typical variety, *A. latifolia* var. *latifolia*. The glabrescent fruits of *A. latifolia* are only found in the very early expansion of the ovary towards fruit maturity, whereas the developed fruits of the newly recognized variety have persistent hairs. The clustering of unifoliolate leaves at the branch apices is only seen elsewhere in *A. heterophylla* and *A. berryi*, but these two species are clearly distinct.

We recognize here *Aldina auyantepuiensis* at species level based on the unique combination of unifoliolate leaves concentrated at the tip of the branchlets, fruits with an appressed indumentum, and often axillary racemes. *Aldina auyantepuiensis* is morphologically similar to *A. berryi*, also from Guaiquinima tepui, but it can be distinguished by its developing fruits with appressed golden hairs (vs. fruits black and glabrous in *A. berryi*), and its exserted ovary, with golden appressed hairs, and borne on a 12–15 mm long stipe and with appressed golden hairs (vs. ovary non-exserted, glabrous, and borne on ca. 5 mm long, glabrous stipe). Both species have glabrous unifoliolate leaves, but these are more evidently clustered at the branch apices in *A. auyantepuiensis*, which has mostly ovate leaflets (vs. ovate or oblong-ovate in *A. berryi*).

ALDINA BARNEBYANA M.Yu.Gontsch. & Yakovlev, Bot. Zhurn. (Moscow & Leningrad) 94(2): 270 (–271). 2009. TYPE: VENEZUELA. Bolivar, Dtto Heres. Base central de Guaiquinima-Tepui, bosques húmedos primarios, ao largo de la quabrada, El Trueno 90 km al S de la Paragua. A 500 msnm, aprox. 6°4' ; 63°22'. 6°4', 12 Mai 1987, *G. Aymard 5771* (holotype: NY barcode NY00231163!; isotypes: FLAS photo barcode FLAS199545!, MO barcode 3590503!).

Distribution and Habitat—This species is only known from Venezuela, where it grows in primary forest at 500 m elevation on the Guaiquinima tepui.

Taxonomic Comments—*Aldina barnebyana* is only known from its type collection and can be readily distinguished from other tepui species by having oblong-ovate leaflets, a glabrous peduncle and long pedicels and stipe.

ALDINA BERRYI R.S.Cowan & Steyer., Ann. Missouri Bot. Gard. 71(1): 312–313. 1984. TYPE: VENEZUELA. Bolivar, cumbre, Cerro Guaiquinima, sector suroeste central, bosque de galeria y bosque alto a lo largo del afluyente suroccidental del Rio Carapo, common in the tall forest of trees 20-25 m tall, 950 m, 5°45'N, 63°35'W, 26 May 1978, *J.A. Steyermark, P. Berry, G.C.K. Dunserville & E. Dunserville 117468* (holotype: NY barcode NY00215052!; isotypes: NY barcode NY00005016!, MO barcode 3242296!, U photo barcode U0003459!, US barcode US00090965!).

Distribution and Habitat—This species is endemic to the state of Bolivar in Venezuela. The type collection was recorded from Cerro Guaiquinima at 950 m elevation. Two other collections, Cerro Kampe (*O. Huber 12071*) and Cerro Camaron (*G. Aymard & A. Fernandez 7104*), suggest that the species prefers to grow at high altitudes around 850-950 m elevation.

Representative Specimens Examined—Venezuela.—BOLIVAR: Distrito Heres, arbustales densos sobre dorso de cuesta inclinado hacia el NW, a aprox. 5 km al WNW

del caserio de Chiguao, 06°07'N, 63°05'W, 800 m, 10 May 1987 (fl), *O. Huber 12221* (NY!, US!); Bolivar, Distrito Heres, cumbre del Cerro Kampe, dorso de cuesta ubicado a 25 km al SE de la punta SE del Cerro Guaiquinima, aprox. 70 km al SW del caserio Chiguao, 05°36'N, 63°21'W, 950 m, 04 May 1987 (fl), *O. Huber 12071* (NY!); Arbustales e bosques sobre altiplanice de arenisca de grano grueso en la cumbre del Cerro Camaron (al S-W del complejo Guaiquinima tepui), 29 Oct 1988 (fl), 850 m, *G. Aymard & A. Fernández 7104*(MO!).

Taxonomic Comments—*Aldina berryi* is distinct from the other species in the genus by its unique combination of glabrous vegetative parts and black, glabrous ovary. These features are otherwise only found separately in *A. microphylla* (glabrous leaflets), *A. heterophylla* (glabrous leaflets), *A. latifolia* (glabrous leaflets), and *A. macrophylla* (glabrous and black ovary). The leaflets of *A. berryi* are plain whereas in *A. heterophylla* they are boat-shaped, furthermore the inflorescence of the latter is very distinctive. When compared to *A. microphylla*, the leaflets of *A. berryi* are bigger 7–9.5 3 3–5 (vs. 4.5–7 3 2.2–3.8 cm) and different in shape from ovate to oblong-ovate (vs. elliptic, oblong or ovate). The fruits of *A. berryi* are rounded whereas in *A. latifolia* they are found rounded and with a spongy mesocarp adapted to floatation. *A. berryi* is distinguishable from *A. macrophylla* by having a glabrous stipe (vs. appressed pubescent golden stipe) (Fig. 2).

ALDINA DIPLOGYNE Stergios & Aymard, Harvard Papers in Botany, 13(1): 29–33. 2008. TYPE: VENEZUELA. Bolivar: municipio Heres, cumbre del cerro Kampe; 25 km al SE de la punta y SE del Cerro Guaiquinima, 70 km al SO del caseiro Chiguao, 05°87'N, 63°21'W, 900 m, 14 May 1987, *O. Huber 12072* (holotype: NY barcode NY03754824!; isotypes: MO barcode MO3891114!, NY barcodes NY00231161! & NY00231162!).

Distribution and Habitat—This species is endemic to the state of Bolivar in Venezuela. *Aldina diplogyne* is known only from the Guaiquinima tepui, where the species is found in shrubby savannas over sandstone outcrops at 300–900 m elevation.

Representative Specimens Examined—Venezuela.—BOLIVAR: distrito Heres, savanas arboradas sobre areniscas conglomeraticas, 2 km de la margen izquierda del medio Rio Paragua, 102 km al S-M de la Paragua, 06°01'N, 63°47'W, 300 m, 14 May

1987 (fl, fr), *G. Aymard* 5959 (NY, MO); Distrito Piar, arbustal sobre dorso de cuesta en la Serrania Senkopiren, aprox. 5 km al E del Rio Caroni y la 35 km al S de Canaima, 05°55'N, 62°48'W, 900m, Jul 1989 (fl, fr), *O. Huber* 12965 (NY); P.N. Canaima, río Cucurital, oeste de la meseta Auyántepeui, 05°54'N, 62°00'W, 430 m, 2 Apr 2000 (st), *R. Duno & R. Riina* 1392 (VEN).

Taxonomic Comments—*Aldina diplogyne* is clearly distinct by its large showy flowers with a remarkable multicarpellate gynoeceum bearing 2–4 functional pistils per flower (Fig. 2). Although this species resembles *Aldina latifolia* Spruce ex Benth. and *Aldina insignis* Endl. by their large mature flower buds, its multicarpellate gynoeceum is enough to separate it from both species. Moreover, the flower pedicels of *A. diplogyne* are longer and strongly vertically striate or wrinkled. There is one collection of *A. insignis* (*N.Y. Sandwith* 1574) with a 2-pistillate gynoeceum, reinforcing the similarity with *A. diplogyne*, but new collections are needed to determine if that is a consistent character. The woody fruits of *A. diplogyne* are also unique, having a botuliform shape, sometimes constricted between seeds, densely covered with a short ferrugineous-puberulent indumentum, although glabrescent when mature.

ALDINA DISCOLOR Spruce ex. Benth., Fl. Bras. 15(2): 12. 1870. TYPE: BRAZIL. Amazonas, “*prope Panuré ad Rio Uaupés*”, Oct 1852–Jan 1853, *R. Spruce* 2802 (lectotype [here designated]: K barcode K000500840!; isoelectotypes: BM barcode BM000931991 photo!, BR barcodes BR5189557 photo! & BR5189885 photo!, E barcode E00663841!, F barcode F0058822F photo!, GH barcode GH00057521 photo!, LD barcode LD1757043 photo!, K barcodes K000500841! & K000500842!, NY barcode NY00231172!, P barcodes P02771882! & P02771883!, TCD barcode TCD0004615 photo!).

Aldina stergiosii M.Yu.Gontsch. & Yakovlev, Bot. Zhurn. (Moscow & Leningrad) 91: 314. 2006. TYPE: BRAZIL. Amazonas: Upper Rio Negro basin. Morro de Seis Lagos., s.d., *A.C. Weber & A. Knob* 1717 (holotype: NY barcode NY01163747!), **syn. nov.**

Distribution and Habitat—This species is widely distributed in the lowland white-sand forests of the northwestern part of the Amazon basin of Brazil, Venezuela, and Colombia. Ter Steege et al. (2013) listed *A. discolor* as a hyperdominant species of the Amazon basin. This was confirmed by our fieldwork along the upper Rio Negro, where the species is frequent in the region of São Gabriel da Cachoeira and the Vaupés river system.

Representative Specimens Examined—**Brazil.**—AMAZONAS: São Gabriel da Cachoeira, Alto Rio Negro, Itacoatiara-mirim, 00°09'56"S, 66°59'54"W, 5 Apr 2013 (fl), *D. Cardoso et al.* 3391 (ALCB, HUEFS, INPA); São Gabriel da Cachoeira, Alto Rio Negro, Serra da Bela Adormecida, 00°16'27"S, 66°50'15"W, 1 Apr 2013 (fl), *D. Cardoso et al.* 3358 (ALCB, HUEFS, INPA); São Gabriel da Cachoeira, Rio Uaupés, Ipanuré, caminho para Cova do Diabo, 17 Nov 1987 (fl, fr), *H.C. de Lima et al.* 3245 (NY, RB); São Gabriel da Cachoeira, Lago Amaro, embocadura do igarapé Tuari, 6 Nov 1987 (fr), *C. Farney* 1912 (INPA, MO, NY); São Gabriel da Cachoeira, upper Rio Negro Basin, 00°18'N, 66°42'W, s.d. (fl), *A.C. Weber & A. Knob* 1717 (NY!). **Venezuela.**—AMAZONAS: Rio Casiquiare, sabanas arbustivas com arena blanca del Caño Momoni, 10–22 Feb 1989 (fl), *B. Stergios et al.* 13238 (US); ANTURES: Santa Rosa de Ucata, 04°24'N, 67°48'W, Apr 1989 (fr), *G.A. Romero et al.* 1869 (NY); San Francisco del Sipapo, arbustal de areia branca, 04°33'N, 67°38'W, s.d. (fr), *P. Berry et al.* 5455 (MO). **Colombia.**—GUAINIA: Rio Guaviare, 31 May 1976 (st), *R.T. Alvaro* 437 (INPA); Guaviare, Rio Inirida, 31 Mai 1976 (st), *R.T. Alvaro* 396 (INPA).

Taxonomic Comments—This species has been frequently misidentified as *A. heterophylla* (e.g., *Stropp* 2014) when sterile because of their shared distribution in white-sand forests and the superficial similarity of their leaves. However, *A. discolor* is found in the white-sand forests of the upper Rio Negro, extending to western Amazonia whereas *A. heterophylla* is only found in the campinaranas of Manaus, Presidente Figueiredo, and surrounding areas of the lower Rio Negro. Moreover, the leaflets of *A. discolor* are covered by white to golden appressed hairs on the lower surface and are broadly oblong or ovate with a rounded to short-acuminate apex, whilst those of *A. heterophylla* are narrowly ovate to oblong-lanceolate with a short-acuminate apex and are completely glabrous. A subtle to dense white-wax coat can also be found on the lower leaflet surface of *A. discolor*. On the leaflets of *A. heterophylla* the wax is arranged in minute, obscure

white dots. Floral morphology also distinguishes the species; *A. heterophylla* bears the smallest flower buds in the genus 0.5–0.8(–1) cm long (vs. 1–2.5 cm long in the other species) that are attached to the inflorescence axis in a nodding position (vs. patently attached in *A. discolor*). In addition, the number of stamens per flower in *A. discolor* is higher than in *A. heterophylla* (ca. 60 vs. ca. 36) ().

Aldina stergiosii is newly synonymized here as it shares with *A. discolor* similar racemose inflorescences, generally with multiple axes growing from a single point, ovate leaflets with rounded bases and short acuminate to rounded apices and somewhat tessellate abaxial surfaces.

ALDINA ELLIPTICA R.S.Cowan, Mem. New York Bot. Gard. 10(1): 146. 1958.

TYPE: VENEZUELA. Terr. Amazonas, near Intermediate Camp, Cerro Yutaje, 1,000 m, 4 Feb 1953, *B. Maguire & C.K. Maguire* 35055 (holotype: NY barcode NY00231170!; isotypes: F barcode F0058825F photo!, G barcode G00364668 photo!, K barcodes K000500849! & K000500848, RB barcode RB00539361!, S barcode S-R-9322 photo!, US barcode US00002863!, VEN barcode VEN43758 photo!).

Distribution and Habitat—This species is distributed in Venezuela and was described from the same mountain as *A. aurea*, Cerro Yutaje. In the original publication of *A. elliptica*, the species is described as a dominant tree, growing at 1000 m elevation.

Representative Specimen Examined—Venezuela.—AMAZONAS: Antures, 12 km WSW of Cerro Autana, 150 m, 16 Jul 1980 (fl), *O. Huber & S.S. Tillett* 5370 (US).

Taxonomic Comments—This distinctive species is similar to *A. discolor* but distinguishable by its elliptical leaflets with rounded to obtuse apices, and smaller petals. Additionally, *A. elliptica* leaflets are evenly glaucous beneath, and not tessellate as in *A. discolor*. *Aldina aurea*, a species from the same tepui complex, also shares the glaucous abaxial leaflet surfaces, but the leaflets of *A. elliptica* are markedly elliptic whereas those of *A. aurea* are oblong-lanceolate.

ALDINA HETEROPHYLLA Spruce ex Benth. Fl. Bras. 15(2):13. 1870. TYPE: BRAZIL.

Amazonas, “*in vicibus Barra, Prov. Rio Negro*”, Feb 1851, *R. Spruce 1369* (lectotype [here designated]: K barcode K000500846!; isolectotypes: BM barcode BM000931996 photo!, E barcode E00663842!, F barcodes F0058826F photo! & F0BN001879F photo!, FI barcode FI004899 photo!, G barcode G00364480 photo!, GH barcode GH00057522 photo!, GOET barcode GOET004631 photo!, K barcode K000500847!, LD barcode LD1775203 photo!, LE barcode LE00002441 photo!, NY barcode NY00231173!, P barcodes P02771877! & P02771878!, RB barcode RB00539362, TCD barcode TCD0004616 photo!).

Distribution and Habitat—This species is only known from Brazil in the lowland white-sand forests of Manaus, Presidente Figueiredo and their surrounding vicinities in the state of Amazonas. *Aldina heterophylla* is usually found in large populations and in association with ectomycorrhizal fungi of the genus *Cantharellus* sp.

Representative Specimens Examined—Brazil.—AMAZONAS: Manaus, Reserva de Desenvolvimento Sustentável do Tupé, margem do Rio Negro, 2°58'56"S, 60°14'27"W, 24 Mar 2013 (fl), *D. Cardoso et al. 3288* (ALCB, HUEFS, INPA); Manaus, Reserva Floresta Adolpho Ducke, Marco 264, 02°55'46"S, 59°58'29"W, 8 Nov 2008 (st), *L.P. de Queiroz et al. 13900* (HUEFS); Manaus, Itacoatiara, Reserva Florestal Adolpho Ducke, km 26, 02°53'S, 59°58'W, 12 Feb 1995 (fl), *Hopkins, M.J.G. et al. 1542* (RB, INPA, IAN, K, MBM, MO, NY, SP, U); Presidente Figueiredo, Rodovia para Balbina, estrada vicinal do km 13, 02°00'58"S, 59°54'26"W, 18 Feb 2016 (fl), *H.C. de Lima & P. Assunção 8199* (RB); Manaus, Campus do INPA, Reserva Biológica de Campina, BR 174, km 60, 02°38'00"S, 60°01'00"W, 8 Mar 1996 (fl), *C.A.C. Ferreira 11216* (INPA); Acajatuba, Margem S do Rio Negro, 03°03'S, 60°36'W, 20 Apr 1986 (fl), *G.T. Prance et al. 29957* (K, NY, US); Presidente Figueiredo, 4 km da margem esquerda do Rio Uatumã, 01°02'S, 59°60'W, 23 Mar 1986 (fl), *C.A.C. Ferreira et al. 6966* (INPA, K, NY, US); Presidente Figueiredo, Rio Uatumã, Igarapé Capitu, 01°02'S, 59°60'W, 29 Mar 1986 (fl), *C.A.C. Ferreira et al. 7016* (HUEFS, INPA, K, NY, US).

Taxonomic Comments—When flowering, *A. heterophylla* is distinguished from the other species of the genus by its dense branched panicles, and the attachment of its subglobose small flower buds (ca. 8 mm long) that are deflexed on the inflorescence axes

(Fig. 3). Furthermore, the leaflets are generally boat-shaped (navicular) (Fig. 4) with the abaxial face subtly dotted with white wax. *Aldina heterophylla* is morphologically similar to *A. occidentalis* Ducke in inflorescence structure and small flower buds, but the leaflets of *A. heterophylla* are completely glabrous, whereas those of *A. occidentalis* are pubescent abaxially. The differences in flower bud size and vegetative traits are sufficient to separate the two species.

ALDINA INSIGNIS (Benth.) Endl., *Allania insignis* Benth., Repert. Bot. Syst. 1(5): 843. 1843. TYPE: GUYANA. On the upper Essequibo and Rupununi, 1838, *R.H. Schomburgk 524* (holotype: K barcode K000500844!; isotypes: BM barcode BM000931995 photo!, E barcode E00663843!, F barcode V0058827F & V0058828F photos!, G barcode G00364792 photo!, GH barcode GH00057523 photo!, K barcodes K000500843 & K000500845, NL barcode 601280 photo!, NY barcode NY00231168!, P barcode P02771875!, TCD barcode TCD0004619 photo!, US barcodes US00002862 & US00459387!).

Aldina insignis var. *retusa* R.S.Cowan, Mem. New York Bot. Gard. Bot. 8: 105. 1953. TYPE: GUYANA. Makreba falls, Kurupung river, upper Mazaruni region, 25 Feb 1939, *A.S. Pinkus 266* (holotype: NY barcode NY00231166!; isotypes: MO barcode MO1174963!, US barcodes US00002861 & US00459389!), **syn. nov.**

Distribution and Habitat—This species is only known from Guyana, where specimens have been collected along the riverbanks and rapids of the Essequibo, Ireng, Kurupung, Kuribrong, and Rupununi rivers in the Mazaruni and Potaro-Siparuni region. *Aldina insignis* is commonly found along the riverbanks of the Makouria river, a tributary of the Essequibo, and frequently along the Ireng river, between Waipa and Sand Hill rapids. The collection *Pires et al. s.n. US no. 13943* from Roraima, Brazil was originally identified as *A. insignis*, but after analyzing the material we correct its identification to *A. latifolia*.

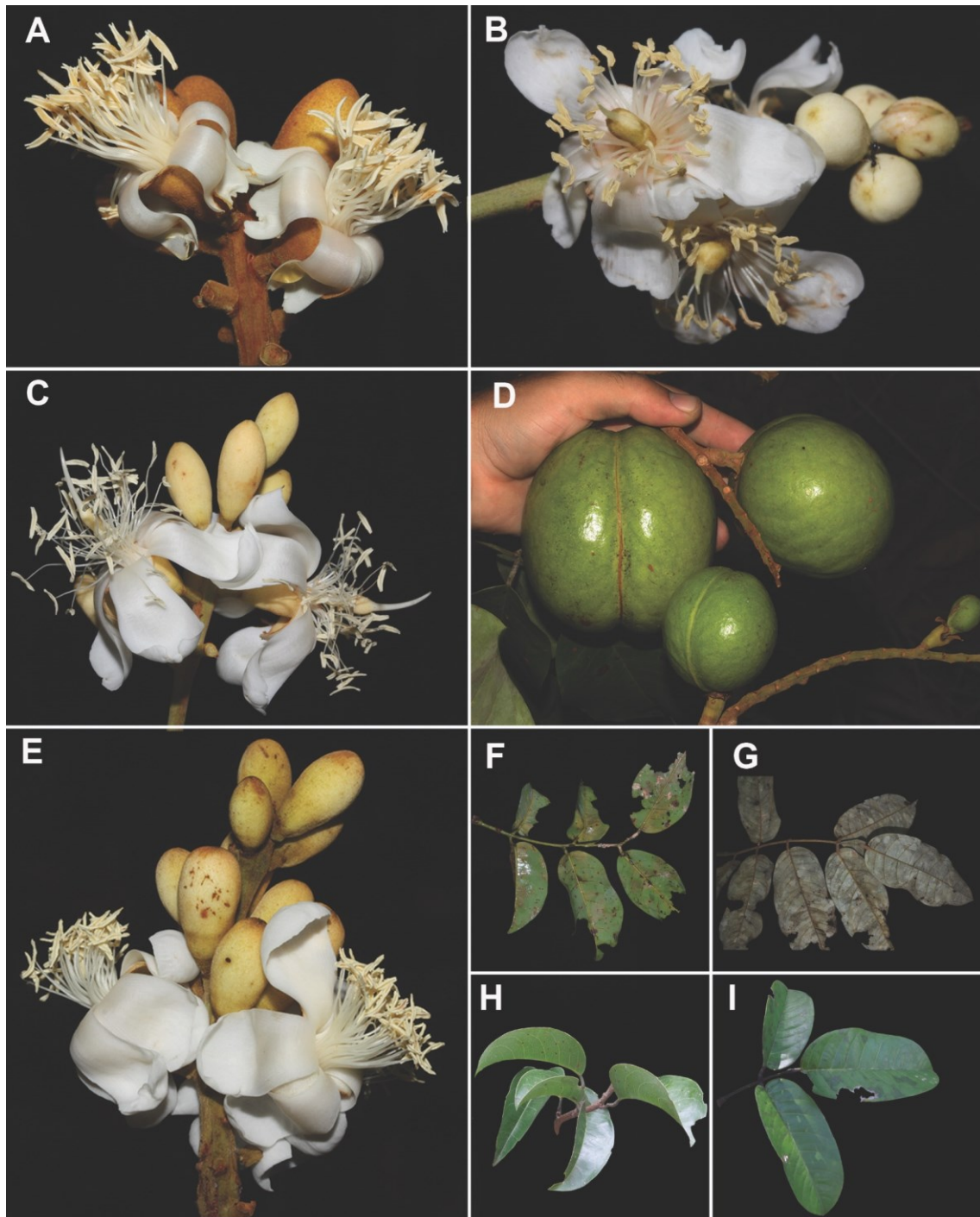


FIG. 4. Morphological diversity of *Aldina*. A, B, C and E. Showy actinomorphic flowers with numerous free stamens of A: *A. kunhardtiana*, B: *A. heterophylla*, C: *A. latifolia*, and E: *A. discolor*. D. Nucoid fruits of *A. latifolia*, the spongy mesocarp (not shown) allows dispersal by flotation along rivers and streams. F–I. Multi-foliolate leaves of F: *A. latifolia*, G: *A. kunhardtiana*, and 3-foliolate leaves of H: *A. heterophylla* and I: *A. discolor*.

Representative Specimens Examined—Guyana: Ireng river, between Waipa and Sand Hill rapids, southern Pakaraima mountains, 19 Sep 1961 (fl), *B. Maguire* 46236 (US, NY); Sukabi river, affluent of Ireng river, Ando savanna, southern Pakaraima mountains 16 Sep 1961 (fl), *B. Maguire* 46212 (US); Essequibo river, Makouria river, frequent on riverbanks, Sep 1937 (fl), *N.Y. Sandwith* 1574 (NY); Upper Mazaruni river region, between Kako and Akawaio indian village on the Kako river near its junction with the Mazaruni river and Karowtipu, 05°45'N, 60°35'W, 500 m, s.d. (fl, fr), *B. Boom* 7317 (NY); Potaro-Siparuni, Kuribrong river, V-4 8, 05°23'48''N, 59°32'00''W, 01 Apr 2011 (fl), *C.E. Zartman et al.* 9420 (INPA).

Taxonomic Comments—The showy, large flowers of *A. insignis* resemble *A. diplogyne*. Indeed, these are the two species with the largest flowers (3–6 cm long) in the genus. However, the flower buds of *A. insignis* are broader, globose and borne on slightly shorter pedicels. Furthermore, *A. insignis* does not have a multicarpellate gynoeceum as in *A. diplogyne*.

Aldina insignis var. *retusa* was described based on its retuse leaflet tips, but some specimens have a mixture of acuminate and retuse apices. Also, a second character used to separate the variety, ‘thin walled trichomes collapsing after drying’ (Cowan, 1953), was not consistent, but the result of the drying process of the specimen. In consequence, the variety is placed into synonymy here.

ALDINA KUNHARDTIANA R.S.Cowan, Mem. New York Bot. Gard. 8(2): 106. 1953. TYPE: VENEZUELA. Territorio Amazonas: Cerro Sipapo (Paráque), small tree in mixed forest, inflorescence and calyx brown, petals and stamens white, Base Camp and Intermediate Camp, 21 Jan 1949, *B. Maguire & L. Politi* 28485 (holotype: NY barcode NY00231178!; isotypes: BM barcode BM000931994 photo!, COL barcode COL00001691 photo!, F barcode F0058829F photo!, K barcode K000500830!, MO barcode MO022350!, NY barcode NY00005017!, RB barcodes RB00548722 & RB00549497!, US barcode US00002860!).

Distribution and Habitat—*Aldina kunhardtiana* is distributed in the lowlands of Brazil, Colombia, and Venezuela, in terra-firme forests. It is frequently found in mixed non-inundated forests on red clay soils and occasionally along non-inundated river-banks.

Representative Specimens Examined—**Brazil.**—AMAZONAS: Alto Rio Negro, São Gabriel da Cachoeira, comunidade Itacoatiara-Mirim, 00°09'43"S, 67°00'30"W, 28 Mar 2013 (fl), *D. Cardoso et al.* 3300 (ALCB, HUEFS, INPA); Rio Negro, Rio Arary, Cachoeira do Uapuhý, terra alta, argila, 2 Oct 1952 (fl), *R.L. Froés* 21317 (IAN, US); São Gabriel da Cachoeira, estrada para Cueui, BR 307, destacamento do Matarucá, 28 Nov 1987 (fl), *H.C. de Lima* 3336 (INPA, MO, NY, RB); São Gabriel da Cachoeira, foz do Rio Içana, acima da comunidade Boa Vista, 00°29'18"N, 67°21'20"W, 28 Nov 1987 (fl), *H.C. de Lima* 3160 (INPA, RB). **Venezuela.**—AMAZONAS: Yatua-Casiquiare, Rio Pacimoni, 12 Jan 1958 (fl), *B. Maguire* 42622 (IAN, NY, RB, US); Yavita-Maroa road, ca. 5 km west of Yavita, 2°53'39"N, 67°27'57"W, 23 May 1996 (st), *G. Aymard et al.* 11033 (MO). **Colombia.**—GUAINIA: Guaiania River, along River Maroa and Victorino, 4°57'00"N, 67°24'00"W, 14 Apr 1953 (fr), *B. Maguire & J.J. Wurdack* 35714 (NY, US); Vaupés, Rio Negro, San Felipe, 01°54'47"N, 67°03'39"W, 12 Dec 1947 (fl), *R.E. Schultes & F. López* 9345 (US).

Taxonomic Comments—*Aldina kunhardtiana* is distinct by the velvety, brown-colored indumentum of its inflorescences. Furthermore, the species has multi-jugate leaves (up to 11-foliolate) (Figs. 3, 4) and the leaflets can be easily recognized by the indumentum on the abaxial surface with an even layer of white wax and erect contorted golden-brownish curly hairs (Fig. 2). There are only two other species, *A. insignis* and *A. polyphylla*, that have multi-jugate leaves as in *A. kunhardtiana*, but the indumentum of *A. kunhardtiana* is unique.

ALDINA LATIFOLIA Spruce ex Benth., Fl. Bras. 15(2): 12. 1870. TYPE: BRAZIL.

Amazonas, Manaus, “in vicinibus Barra, Prov. Rio Negro, high rocky shores”, 1851, *R. Spruce* 1355 [lectotype (designated by Ramos and Cardoso 2016): K barcodes K000500834 & K000500835! [two sheets]; isolectotypes: BM barcode BM000931993 photo!, F barcode F0058830F photo!, FI barcode FI004910 photo!, GH barcode GH00057524 photo!,

LD barcode LD1756979 photo!, LE barcode LE00002442 photo!, NY barcode NY00231179!, OXF!, P barcodes P02771864 & P2771872!, TCD barcode TCD0004618 photo!.

Aldina latifolia var. *pubescens* R.S.Cowan, Mem. New York Bot. Gard. 8(2): 107. 1953. TYPE: BRAZIL. Amazonas: “Prope San Gabriel da Cachoeira, ad Rio Negro, Brasiliae Borealis, Jan-Aug 1852”, R. Spruce 2077 (holotype: NY barcode NY00231167!; isotypes: K barcode K000500840!, BM barcode BM000931992 photo!, P barcodes P02771865, P02771866 & P02771867!, TCD barcode TCD0004617 photo!), **syn. nov.**

Aldina amazonica M.Yu.Gontsch. & Yakovlev, Bot. Zhurn. (Moscow & Leningrad) 91: 318. 2006. TYPE: BRAZIL. Amazonas: Camanaus, igarapé Caçabu, mata baixa, solo arenoso, 27 Oct 1978, O.C. Nascimento 813 (holotype: US barcode US00459450!), **syn. nov.**

Aldina rio-negrae M.Yu.Gontsch. & Yakovlev, Bot. Zhurn. (Moscow & Leningrad) 94(2): 271. 2009. TYPE: VENEZUELA. Amazonas: Selvas pluviales ribereñas del Rio-Negro, entre S. Carlos de Rio-Negro y la boca de Rio-Cassiquiare. 10-22 Feb 1989, B. Stergios, G. Aymard, K. Kubitzki & E. Melguera 13096 (holotype: NY barcode NY00231108!), **syn. nov.**

Distribution and Habitat—*Aldina latifolia* is distributed in the lowlands of Brazil, Venezuela, and Colombia. This is the most frequently collected species in the genus, perhaps because of its large and showy inflorescences that stand out against the dark green background of the igapó forests along the Rio Negro, Vaupés, Inírida, Guáinia, Atacavi, and their tributaries. The enlarged hydrochoric fruits of *A. latifolia* have a spongy mesocarp (Fig. 4) to optimize long distance dispersal along rivers and streams. The seeds only germinate once the fruits have settled on a riverbank during the dry season, therefore the species is commonly found along river margins.

Representative Specimens Examined—**Brazil.**—AMAZONAS: São Gabriel da Cachoeira, depois da Ilha das Flores, 00°03'20"N, 67°17'46"W, 8 Apr 2003 (fl), D. Cardoso et al. 3438 (ALCB, HUEFS, INPA); Manaus, Reserva de Desenvolvimento do Tupé, 03°02'37.1"S, 60°15'12"W, 6 Apr 2018 (fl, fr), G. Ramos et al. 154 (ALCB, HUEFS, INPA, RB); Manaus, Rio Tarumanzinho, 03°00'52"S, 60°06'01"W, 7 Jul 1976

(fr), *G.T. Prance* 23736 (INPA, NY, US). **Venezuela.**—AMAZONAS: San Carlos de Rio Negro, igapó forest along margin of Rio Negro, 0.5 km of Caño Cuweje, 01°56'N, 67°03'W, 2 May 1980 (fr), *J.J. Clark* 7556 (MO, NY); Caño Mesaque, a tributary on the south side of Rio Atacavi, 02°59'50"N, 67°06'05"W, May 1996 (fr), *P. Berry* 6228 (MO); Guainia River, in forest along Pimichin-Yavita, road near Pimichin, Apr 1953 (fl), *B. Maguire & J.J. Wurdack* 35595 (NY, US). **Colombia.**—VAUPES: Vaupés River, Trubon waterfall, region S-E of Mitu, 01°15'07"N, 70°14'01"W, Nov 1945 (fl), *P.H. Allen* 3404 (US); Mariapiri, Inirida River, Raudal Alto, 02°22'06"N, 69°45'00"W, Feb 1953 (fr), *A. Fernandez* 2080 (US).

Taxonomic Comments—The name *Aldina latifolia* was first used on herbarium labels by Spruce [*Spruce* 1355 (K) and *Spruce* 2077 (OXF)]. Later on, Bentham added another collection by Martius [*Martius s.n.* (M)] to Spruce's collections for the formal description of the species in *Flora Brasiliensis* (1870) with a total of three syntypes. Given the lack of a single type, the name has been recently assigned a lectotype (Ramos and Cardoso 2016).

Cowan (1953) observed that only two species in the genus have glabrous leaflets, *A. latifolia* and *A. heterophylla*, yet he pointed out that these species are so distinct that there is no possibility of confusion. *Aldina latifolia* can be readily distinguished by the combination of its oblong pale yellow flower buds, glabrous leaflets (occasionally pubescent) and hydrochoric fruits with a spongy mesocarp. Cowan described the variety *Aldina latifolia* var. *pubescens* based on a note by Richard Spruce stating that the specimen *Spruce* 2077 might represent a variety based on the longer leaflets and the puberulent indumentum covering the vegetative parts (Cowan 1953). After revision of new collections and comparison with the type material of *A. latifolia* var. *pubescens*, we observed that leaflet size is continuous between the two taxa and the indumentum is ephemeral in the varietal specimens as well as in the typological species, with hairs being naturally caducous or shed after drying. Therefore, these characters are considered infra-specific variation, and we thus place *A. latifolia* var. *pubescens* in synonymy.

Aldina amazonica M.Yu.Gontsch. & Yakovlev and *Aldina rio-negrae* M.Yu.Gontsch. & Yakovlev are also synonymized here. *Aldina amazonica* description is based on both collections of *A. latifolia* (*Nascimento* 813) together with *Andira* sp. (*Prance* 9049 [fruits], 15011[leaves]; the two collections of Prance were later on

identified as *Andira micrantha* Ducke (Pennington, 2003), therefore the mix of material invalidates the description of *A. amazonica*. *Aldina rio-negrae* is also newly synonymized here given that the characters distinguishing it from *A. latifolia* var. *pubescens* (e.g., leaflet apex attenuate vs. acuminate, flower bud 12 mm long vs. 12–15 mm long, and gynophore 4–5 vs. 2–3 times longer) now fall into a broader concept of *A. latifolia*.

ALDINA MACROPHYLLA Spruce ex Benth., Fl. Bras. 15(2): 13. 1870. TYPE: VENEZUELA. “*Habitat ad ripas fluminis Pacimoni*”, 1853–1854, *R. Spruce* 3349 (lectotype [here designated]: K barcode K000500826!; isolectotypes: BM barcode BM000931990, E barcode E00663845!, F barcodes V0058832F & V28288F, MO barcode MO2840759!, NY barcode NY00231171!, GH barcode GH00057526, LD barcode LD1759076, K barcodes K000500824, K000500825 & K000500827!, P barcodes P02771861, P02771862 & P02771863, RB barcode RB00539364, W barcode W18890004637 photo!).

Aldina aquae-nigrae M.Yu.Gontsch. & Yakovlev, Bot. Zhurn. (Moscow & Leningrad) 94(2): 273. 2009. TYPE: VENEZUELA. Amazonas, selva pluviales de rebalse del Bajo Emoni, que desemboca en el Rio Siapa, 18 Feb - 4 Mar 1986, *B. Stergios & G. Aymard* 9239 (holotype: NY!; isotypes: NY barcodes NY00231160 & NY00231164!, MO barcode MO5992824!, FLAS barcode FLAS167009 photo!), **syn. nov.**

Aldina yapacanensis R.S.Cowan, Mem. New York Bot. Gard. 8(2): 108. 1953. *Aldina macrophylla* var. *yapacanensis* (R.S.Cowan) Stergios, Fl. Venez. Guayana 5: 248. 1999. TYPE: VENEZUELA. Amazonas, Caño to Cerro Yapacana, along the margin, 6 Jan 1951, *B. Maguire, R.S. Cowan & J. Wurdack* 30758 (holotype: NY barcode NY00231176!; isotypes: K barcode K000500829!, US barcode US00002853!, VEN barcode VEN31332 photo!), **syn. nov.**

Distribution and Habitat—*Aldina macrophylla* is only known from Venezuela, occurring in white-sand forests near riverbanks of the rivers Pacimoni, Yatua, Casiquiare, Siapa, and Yapacana and is commonly found in gallery forests on white sand along

canyons. The species was previously recorded in Brazil based on the single collection *Schultes 9936*, but after detailed examination this collection is reidentified as *A. latifolia*.

Representative Specimens Examined—Venezuela.—AMAZONAS: Rio Pacimoni – Yatua, Casiquiare, occasional on near Rio Yatua, 01°28'32"N, 66°29'00.0"W, 100–140 m, 6 Feb 1954 (fl), *B. Maguire 37553* (RB, NY); Rio Siapa, edge of savanna on left bank of Cano Hechimoni 8 km above mouth, 01°56'24"N, 66°4'20"W, Feb 1954 (fl), *B. Maguire 37620*(NY); Rio Siapa, edge of savanna on left bank of Cano Hechimoni 8 km above mouth, 100–130 m, 01°56'24"N, 66°04'20"W, *B. Maguire 37618* (NY); Along river Yapacana cano from laguna to junction with river, alto rio Orinoco, 150 m, 20 Mar 1953 (fl), *B. Maguire 34586* (US, NY); Rio Siapa, Selvas ribeirinhas do Rio Emoni, afluyente do Rio bajo Siapa, 02°10'N, 66°18'W, 22 Jan 1992 (fl), *B. Stergios et al. 15499* (MO, NY); Atabapo, southeastern bank of the middle part of caño yagua at cucurital de yagua, 03°36'N, 66°34'W, 120 m, 8 May 1979 (fl), *G. Davidse 17344* (US); Rio Cassiquiare, sabana arbustiva de arena blanca, 8 vueltas arriba de la boca en el Caño Chimoni (“Sabana de Basil”), 02°01'60"N, 66°24'00.0"W, 10 to 22 Feb 1989 (fl), *B. Stergios 13285* (US); Rio Pacimoni, Gallery forest on and around small rock islands, white sand and rocks, 01°44'19"N, 66°30'26"W, 86 m, 26 Jan 2005 (fl), *K.M. Redden 3342* (US); Rio Dimiti, rio Negro basin, 00°42'37"N, 66°38'28"W, 12 May 1948 (fr), *R.E. Schultes 9936* (US).

Taxonomic Comments—The glabrous and dark brown ovary of *A. macrophylla* and appressed golden hairs covering the stipe are the strongest distinguishing characters of the species (Fig. 2). A glabrous ovary has also been reported in *A. yapacanensis* (Cowan et al. 1953), although the name was reduced to synonymy of *A. macrophylla* (Cowan et al. 1958) due to the continuous variation in flower parts and leaflets. Later, Stergios (in Steryermark et al. 1999) suggested *A. yapacanensis* should be treated at varietal level within *A. macrophylla*, as var. *yapacanensis* based on the attachment of the bracteoles on the peduncle, the pubescence of the gynophore, and whether it is exserted. However, these characters are not consistent, since they vary according to the developmental stage of the flower, thus we synonymize Stergios’ variety. *Aldina aquae-negrae* M.Yu.Gontsch. & Yakovlev was recognized based on leaflet size and style length. However, both characters vary continuously between it and *A. macrophylla* and are of no taxonomic value. In addition, the glabrescent ovary supports the synonymy of this name

under *A. macrophylla*, since the character is only found elsewhere in *A. berryi*, which is not found anywhere else in the genus.

ALDINA MICROPHYLLA M.Yu.Gontsch. & Yakovlev, Bot. Zhurn. (Moscow & Leningrad) 91: 319. 2006. TYPE: BRAZIL. Amazonas: Serro Araca, W slope of southern massif, montane humid forest below wall, top of western serra, 800 m, 0°40'N, 63°18'W, 20 Mar 1984, J. Pipoly & G. Samuels 6846 (holotype: US barcode US00459448!; isotype: INPA barcodes INPA196430 & INPA 134109!).

Distribution and Habitat—This species is only found in Brazil and has been recorded from the Serra do Aracá tepui. The few records of *A. microphylla* indicate that the species occurs at elevations of 800 m elevation in montane humid forest on the slopes and summit of the tepui.

Representative Specimen Examined—Brazil.—AMAZONAS: Serra do Aracá, platô (ou encosta?), 200 km ao Norte de Barcelos, Mar 1984 (fl), M.G. da Silva 7151 (INPA, K, NY).

Taxonomic Comments—This species is vegetatively similar to *A. heterophylla*, especially in its small glabrous leaflets. The inflorescence of *A. microphylla* is racemose whereas *A. heterophylla* has dense panicles. Another difference between the two species is that *A. microphylla* bears larger, patent flower buds (13–20 mm long) versus the small, deflexed buds (5–10 mm long) of *A. heterophylla*. The two species also differ in their ecology; *A. microphylla* is often associated with montane forests while *A. heterophylla* grows in lowland sandy forests.

ALDINA OCCIDENTALIS Ducke, Arq. Inst. Biol. Veg. 4(1): 16–17. 1938. TYPE: BRAZIL. Amazonas: São Paulo de Olivença, Rio Solimões, “*silva terris altis loco humido prope rivulum*”, 27 Feb 1932, A. Ducke s.n. RB no. 24051 (lectotype [here designated]: RB barcode RB00539365!; isoelectotypes: K barcode K000500833!, P barcode P02771859!, RB

barcode RB00545298!, S barcode S-R-9324 photo!, U barcode U0003460 photo!, US barcode US00002858!).

Distribution and Habitat—This species is only known from São Paulo de Olivença in the western Brazilian Amazon. The two known collections of *Aldina occidentalis* describe the species as occurring in non-flooded terra-firme forests on siliceous soil.

Representative Specimen Examined—**Brazil.**—AMAZONAS: São Paulo de Olivença, mata de terra firme silicosa, 12 Apr 1944 (fl), *A. Ducke 1499* (IAN photo!, NY!, R!, RB!, US!).

Taxonomic Comments—*Aldina occidentalis* can be readily distinguished from the other species of the genus by the combination of small subglobose flower buds and a subsessile ovary born on a ca. 1 mm long stipe. Ducke (1939) pointed out the resemblance of this species with *A. heterophylla* because of the shared branched panicles and globose flower buds. However, *A. occidentalis* has an erect indumentum on the leaflet lower surface, whereas *A. heterophylla* has completely glabrous leaflets. Moreover, the difference in the size of the flower buds (5–8 mm long in *A. heterophylla* vs. 10–13 mm long in *A. occidentalis*) and their position on the inflorescence rachis (deflexed vs. patent), as well as the subsessile ovary of *A. occidentalis* (1 mm long vs. 6 mm long), adequately support the recognition of two distinct species.

ALDINA PAULBERRYI Aymard, Novon 8(4): 330. 1998. TYPE: VENEZUELA. Amazonas: Rio Orinoco, cerca de Macuruco, bosques húmedos, 200 m, 3 Mar 1976, *P. Berry 2125* (holotype: VEN photo!; isotypes: MO barcodes MO2669869 & MO2737198!, NY!).

Distribution and Habitat—This species is only known from Venezuela from a single collection in lowland humid forest at 200 m elevation.

Taxonomic Comments—*Aldina paulberryi* is most similar morphologically to *A. reticulata* from Venezuelan Amazon montane forest (Cowan, 1953, 1961), but it differs mainly in having broadly ovate to orbicular leaflets with inconspicuous tertiary venation (vs. oblong to oblong-ovate leaflets and prominent, strongly reticulate tertiary venation).

ALDINA PETIOLULATA R.S.Cowan, Mem. New York Bot. Gard. 10 (4): 70. 1961.

TYPE: VENEZUELA. Amazonas: frequent in slope forest near Camp 3, alt. 700 – 800 m, Cerro de la Neblina, Río Yatua, 20 Dec 1957, *B. Maguire, J.J. Wurdack & C.K. Maguire* 42555 (holotype: US barcode US00002857!; isotypes: COL barcode COL000001692 photo!, F barcode F0058833F, GH barcode GH00257061 photo!, IAN barcode IAN114760 photo!, K barcode K000500823!, LE barcodes LE00002444 photo! & LE00002443 photo!, MICH barcode MICH1107083 photo!, MO barcode MO022351!, NY barcodes NY00231175! & NY00005018!, P barcode P01817989!, RB barcodes RB00545297 & RB00539366!, S barcode S-R-9325 photo!, SP barcode SP000885 photo!, U barcode U0003461 photo!, UC barcode UC1221528 photo!, US barcode US00002857!, VEN barcode VEN52399 photo!).

Distribution and Habitat—This species is known from Venezuela and Colombia based on two collections. The type locality is in Venezuela in slope forests close to the Cerro de la Neblina at 700–800 m elevation, while the second collection is from terra-firme forest in the National Park of Cahuinari, Colombia.

Representative Specimen Examined—Colombia.—AMAZONAS: Parque Nacional Cahuinari, Río Cahuinari, rebalse alto, bosque de tierra firme, 01°28'S, 70°45'W, 26 Feb 1991 (fl), *Fundacion Biológica Puerto Rastrojo* 4833 (K).

Taxonomic Comments—The relatively large, 15–20 mm long pulvinules of *A. petiolulata* make it readily distinguishable from the other species of the genus. Cowan (1961) suggested a close similarity of *A. occidentalis* with *A. petiolulata*, both of which share globose flower buds. Nevertheless, *Aldina petiolulata* can be readily distinguished from *A. occidentalis* by its unique longer, slender pulvinules, and in having an ovary stipe ca. 10 mm long (vs. ca. 1 mm long).

ALDINA POLYPHYLLA Ducke, Arqv. Inst. Biol. Veg. 4(1): 17. 1938. TYPE: BRAZIL.

Amazonas, “*loco Uarurá, super Uanuacá, silva non inundabili*”, 14 Nov

1936, *A. Ducke s.n. RB no. 35083* (lectotype [here designated]: K barcode 000500832!; isoelectotypes: K barcodes K000500831! & K000858684!, NY barcode NY00231177!, P barcode P02771858!, RB barcodes RB00539368! & RB00545296!, S barcode S-R-9326 photo!, U barcode U0003462 photo!, US barcodes US00002855!, US00002856! & US00432644!).

Distribution and Habitat—This species is only known from Brazil and has been collected in the vicinities of São Gabriel da Cachoeira and Santa Isabel do Rio Negro. *Aldina polyphylla* is another species from lowland terra-firme forest, rarely occurring in flooded areas.

Representative Specimens Examined—**Brazil.**—AMAZONAS: São Gabriel da Cachoeira, Rio Curicuriari, 00°22'06"S, 67°17'05"W, 29 Jul 1991 (fl), *G. Martinelli & J. Ramos 14601* (ALCB, HUEFS, RB); São Gabriel da Cachoeira, mata de terra firme, 00°07'13"S, 67°02'53"W, 5 Aug 1989 (fl), *L. Coêlho & O.P. Monteiro 12* (INPA); Santa Isabel do Rio Negro, “*silva riparia rarius inunbilis*”, 10 Jun 1936 (fl), *A. Ducke s.n. RB no. 35082* (RB, US).

Taxonomic Comments—The large leaves, up to 36 cm long, with up to 11 leaflets that are narrowly oblong and measure 5–16 × 1.5–4.5 cm readily identify this species.

ALDINA RETICULATA R.S.Cowan, Mem. New York Bot. Gard. 8(2): 107. 1953.

TYPE: VENEZUELA. Cerro Sipapo, montane mixed forest above Caño Grande, 1 km NW of Savanna camp, 28 Dec 1948, *B. Maguire & L. Politi 27973* (holotype: NY barcode NY00231165!; isotypes: BM barcode BM000931989 photo!, F barcode F0058834F photo!, US barcode US00002854!, VEN barcode VEN31739 photo!).

Aldina speciosa M.Yu.Gontsch. & Yakovlev, Bot. Zhurn. (Moscow & Leningrad) 91: 316. 2006. TYPE: VENEZUELA. Amazonas: Sierra Cuao-Sipapo, 16 Feb 1993, *A. Groeger 754* (holotype: NY barcode NY01546440!; isotype: VEN barcode VEN283603 photo!), **syn. nov.**

Distribution and Habitat—This species is currently only known from Venezuela. *Aldina reticulata* grows in montane forest of Cerro Sipapo (Paráque), savannas, and near the river Guayapo.

Representative Specimens Examined—**Venezuela**—AMAZONAS: Cerro Sipapo, savanna vicinity Base Camp, 30 Dec 1948 (fl), *Maguire & L. Politi* 28046 (NY); savanna near Base Camp, 8 Feb 1949 (fl, fr), *B. Maguire & L. Politi* 28820 (NY); River Guayapo, 4°18'N, 67°28'W, 120 m, s.d. (st), *E. Foldats & J. Velazco* 9321 (NY).

Taxonomic Comments—This species is morphologically similar to *A. latifolia*, but it differs in having strongly reticulate-veined leaflets, a lax racemose inflorescence, the ovary stipe considerably shorter (2–5 mm long in *A. reticulata* vs. 7–15 mm long in *A. latifolia*), and young fruits with pale golden hairs, glabrescent (vs. fruit glabrous during the first stages of development in *A. latifolia*). *Aldina speciosa* M.Yu.Gontsch. & Yakovlev, another species described from the Cuao-Sipapo mountain range, is newly synonymized here based on its similar oblong leaflets with an acute apex and rounded base, prominent tertiary venation, and relatively short ovary stipe. The racemose inflorescence found in the type collection of *A. speciosa* is more branched than the racemes found in *A. reticulata*, but there is character overlap between the two.

ACKNOWLEDGMENTS

We thank the curators of the cited herbaria for making their collections available; Luciano Paganucci de Queiroz, Juliana Rando and an anonymous reviewer for their critical comments and suggestions on the manuscript. We also thank Honalto (Bire), Moisés, Abraão, and Adriel for invaluable support during fieldwork along the upper Negro and Valpes rivers and for sharing their knowledge of the local flora. The first author thanks Jefferson Valsko da Silva and Briggittthe Melchor for support and cheerful company during fieldwork in Manaus, and Rosy, Jorgete, Luke, and Eli for being such amazing hosts. The first author also thanks the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes-Brazil) for the M.Sc. scholarship, The Smithsonian National Museum of Natural History, Washington D.C. for the Cuatrecasas Fellowship Award, which allowed the examination of their *Aldina* collections, and the New York Botanical Garden and Missouri Botanical Garden for access to specimens. DBOSC's

research in legume systematics is supported by grants from CNPq (Research Productivity Fellowship no. 308244/2018-4; Universal no. 422325/2018-0) and FAPESB (Universal no. APP0037/2016).

AUTHOR CONTRIBUTIONS

GR and DBOSC designed the project. DBOSC supervised the study. GR, DBOSC, and CEZ collected specimens in the field. GR and DBOSC revised herbarium specimens. GR prepared morphological descriptions, figures, and wrote the first draft of the manuscript. All authors revised and approved the manuscript.

LITERATURE CITED

- Adanson, M. 1763. Familles des Plantes. *Academie des Sciences, de la Société Roiale de Londres, Censeur Roial* (2): 328.
- Aymard, G. 1998. A new species of *Aldina* (Fabaceae: Caesalpinioideae) from Venezuelan Amazonia. *Novon* 8: 330–331.
- Barroso, G. M., M. P. Morin, A. L. Peixoto, and C. L. F. Ichaso. 2004. *Frutos e sementes: morfologia aplicada à sistemática de dicotiledôneas*. Viçosa: Editora UFV.
- Bebber, D. P., M. A. Carine, J. R. I. Wood, A. H. Wortley, D. J. Harris, G. T. Prance, G. Davidse, J. Paige, T. D. Pennington, N. K. B. Robson, and R. W. Scotland. 2010. Herbaria are a major frontier for species discovery. *Proceedings of the National Academy of Sciences* 107: 22169–22171.
- Bentham, G. 1840. Contributions towards a flora of South America: enumeration of plants collected by Mr. Schomburgk in British Guiana. *The Journal of Botany* 2(10): 91.
- Bentham, G. 1870. Leguminosae, Swartzieae et Caesalpinieae. Pp. 12–13 in *Flora Brasiliensis*, eds. C. F. P Martius, C. F. P Eichler, I. Urban.
- Browne, P. 1756. The civil and natural history of Jamaica in three parts. *The civil and natural history of Jamaica*. 299.
- Cardoso, D., L. P. de Queiroz, R. T. Pennington, H. C. de Lima, E. Fonty, M. F. Wojciechowski, and M. Lavin. 2012. Revisiting the phylogeny of papilionoid

- legumes: new insights from comprehensively sampled early-branching lineages. *American Journal of Botany* 99: 1991–2013.
- Cardoso, D., R. T. Pennington, L. P. de Queiroz, J. S. Boatwright, B. E. van Wyk, M. F. Wojciechowski, and M. Lavin. 2013. Reconstructing the deep-branching relationships of the papilionoid legumes. *South African Journal of Botany* 89: 58–75.
- Cardoso, D.B.O.S., L. P. Queiroz, and H. C. Lima. 2014. A taxonomic revision of the South American papilionoid genus *Luetzelburgia* (Fabaceae). *Botanical Journal of the Linnean Society* 175: 328–375.
- Cardoso, D., J. G. Carvalho-Sobrinho, C. E. Zartman, D. L. Komura, and L. P. Queiroz. 2015a. Unexplored Amazonian diversity: rare and phylogenetically enigmatic tree species are newly collected. *Neodiversity* 8: 55–73.
- Cardoso, D., W. M. São-Mateus, D. T. Cruz, C. E. Zartman, D. L. Komura, G. Kite, G. Prenner, J. J. Wieringa, A. Clark, G. Lewis, R. T. Pennington, and L. P. de Queiroz. 2015b. Filling in the gaps of the papilionoid legume phylogeny: the enigmatic Amazonian genus *Petaladenium* is a new branch of the early-diverging Amburaneae clade. *Molecular Phylogenetics and Evolution* 84: 112–124.
- Cardoso, D., T. Särkinen, S. Alexander, A. M. Amorim, V. Bittrich, M. Celis, D. C. Daly, P. Fiaschi, V. A. Funk, L. L. Giacomini, R. Goldenberg, G. Heiden, J. Iganci, C. L. Kelloff, S. Knapp, H. C. de Lima, A. F. P. Machado, R. M. dos Santos, R. Mello-Silva, F. A. Michelangeli, J. Mitchell, P. Moonlight, P. L. R. de Moraes, S. A. Mori, T. S. Nunes, T. D. Pennington, J. R. Pirani, G. T. Prance, L. P. de Queiroz, A. Rapini, R. Riina, C. A. V. Rincon, N. Roque, G. Shimizu, M. Sobral, J. R. Stehmann, W. D. Stevens, C. M. Taylor, M. Trovó, C. van den Berg, H. van der Werff, P. L. Viana, C. E. Zartman, and R. C. Forzza. 2017. Amazon plant diversity revealed by a taxonomically verified species list. *Proceedings of the National Academy of Sciences* 114: 10695–10700.
- Choi I-S, D. Cardoso, L. P. Queiroz, H. C. Lima, C. Lee, T.A. Ruhlman, R. K. Jansen and M. F. Wojciechowski. 2022. Highly resolved papilionoid legume phylogeny based on plastid phylogenomics. *Frontiers in Plant Science* 13: 823190.
- Cowan, R. S. 1953. *Aldina* (Leguminosae, Caesalpinioideae). Pp. 104–108 in *The botany of the Guayana highland*, eds. B. Maguire, R. S. Cowan, and J. J. Wurdack. Memoirs of The New York Botanical Garden vol 8.

- Cowan, R. S. 1958. *Aldina* (Leguminosae, Caesalpinioideae). Pp. 145–147 in *The Botany of the Guayana highland*, eds. B. Maguire, R. S. Cowan, and J. J. Wurdack. Memoirs of the New York Botanical Garden, vol 10. Bronx: New York Botanical Garden.
- Cowan, R. S. 1961. *Aldina* (Leguminosae, Caesalpinioideae). Pp. 70 –71 in *The botany of the Guayana Highland*, eds. B. Maguire, R. S. Cowan, and J. J. Wurdack. Memoirs of the New York Botanical Garden, vol 10. Bronx: New York Botanical Garden.
- Cowan, R. S. and Steyermark. 1984. *Aldina* (Leguminosae, Caesalpinioideae). *Annals of the Missouri Botanical Garden* 71: 312–313.
- Doyle, J. J., J. L. Doyle, J. A. Ballenger, and J. D. Palmer. 1997. The distribution and phylogenetic significance of a 50-kb chloroplast DNA inversion in the flowering plant family Leguminosae. *Molecular Phylogenetics and Evolution* 5: 429–438.
- Draper, F. C., F. R. C. Costa, G. Arellano, O. L. Phillips, A. Duque, M. J. Macía, H. ter Steege, G. P. Asner, E. Berenguer, J. Schietti, J. B. Socolar, K. G. Dexter, P. M. Jørgensen, J. S. Tello, W. E. Magnusson, T. R. Baker, C. V. Castilho, A. Monteagudo-Mendoza, P. V. A. Fine, K. Ruokolainen, E. N. H. Coronado, G. Aymard, N. Dávila, M. S. Sáenz, M. A. R. Paredes, J. Engel, C. Fortunel, C. E. T. Paine, J.-Y. Goret, A. Dourdain, P. Petronelli, E. Allie, J. E. G. Andino, R. J. W. Brien, L. C. Pérez, A. G. Manzatto, N. Y. P. Zambrana, J.-F. Mollino, D. Sebatier, J. Chave, S. Fauset, R. G. Villacorta and C. Baraloto. 2021. Amazon tree dominance across forest strata. *Nature Ecology and Evolution* 5: 757–767.
- Endlicher, S. F. L. 1836. *Genera Plantarum Secundum Ordines Naturales Disposita* 151.
- Fine, P. V. A., R. Garcia-Villacorta, N. C. A. Pitman, I. Mesones and S. W. Kembel. 2010. A floristic study of the white-sand forests of Peru. *Annals of Missouri Botanical Garden* 97: 283–305.
- Fine, P. V. A. and C. Baraloto. 2016. Habitat endemism in white-sand forests: insights into the mechanisms of lineage diversification and community assembly of the Neotropical flora. *Biotropica* 48: 24–33.
- Fine, P. V. A. and E. M. Bruna. 2016. Neotropical white-sand forests: origins, ecology and conservation of a unique rain forest environment. *Biotropica* 48: 5–6.
- Fortunel, C., C. E. T. Paine, P. V. A. Fine, N. J. B. Kraft, and C. Baraloto. 2014. Environmental factors predict community functional composition in forests across Amazonia. *Journal of Ecology* 102: 145–155.

- Forzza, R., J. F. A. Baumgratz, C. E. M. Bicudo, D. A. L. Canhos, A. A. Carvalho Jr., M. A. N. Coelho, A. F. Costa, D. P. Costa, M. G. Hopkins, P. M. Leitman, L. G. Lohmann, E. N. Lughadha, L. C. Maia, G. Martinelli, M. Menezes, M. P. Morim, A. L. Peixoto, J. R. Pirani, J. Prado, L. P. de Queiroz, S. Souza, V. C. Souza, J. R. Stehman, L. S. Sylvestre, B. M. T. Walter, and D. C. Zappi. 2012. New Brazilian floristic list highlights conservation challenges. *Bioscience* 62: 39–45.
- Gontscharov, M. Yu., G. P. Yakovlev, and M. N. Povydysh. 2006. On the new subtribe Aldiniinae of the Swartzieae and new species of the genus *Aldina*. *Botanicheskii Zhurnal* 91: 312–321.
- Gontscharov, M. Yu., G. P. Yakovlev, and M. N. Povydysh. 2009. Notes on the genus *Aldina* (Fabaceae). *Botanicheskii Zhurnal* 94: 267–275.
- Harris, J. G. and M. W. Harris. 1994. *Plant identification terminology, an illustrated glossary*. Spring Lake Publishing, Utah.
- Henkel, T. W., J. Terborgh, and R. J. Vilgalys. 2002. Ectomycorrhizal fungi and their leguminous hosts in the Pakaraima Mountains of Guyana. *Mycological Research* 106: 515–531.
- Hopkins, M. J. G. 2007. Modelling the known and unknown plant biodiversity of the Amazon basin. *Journal of Biogeography* 34: 1400–1411.
- Ireland, H. E., R. T. Pennington, and J. Preston. 2000. Molecular systematics of the Swartzieae. Pp. 217–231 in *Advances in Legume Systematics*, eds. P. S. Herendeen and A. Bruneau. Royal Botanic Gardens vol 9. Kew.
- Ireland, H. E. 2005. Swartzieae. Pp. 214–225 in *Legumes of the World*, eds. G. Lewis, B. Schreire, B. Mackinder, and M. Lock. Royal Botanic Gardens. Kew.
- Janzen, D. H. 1974. Tropical blackwater rivers, animals, and mast fruiting by the Dipterocarpaceae. *Biotropica* 6: 69–103.
- Klitgaard, B. B. 2005. *Platymiscium* (Leguminosae: Dalbergieae): biogeography, systematics, morphology, taxonomy and uses. *Kew Bulletin* 60: 321–400.
- LPWG [Legume Phylogeny Working Group]. 2017. A new subfamily classification of the Leguminosae based on a taxonomically comprehensive phylogeny. *Taxon* 66: 44–77.
- Meisner, C. F. 1842. *Plantarium vascularium genera: secundum ordines naturales digestaeorumque differentiae et affinitates tabulis disgnosticis expositae*. 304.

- Mori, S. A., A. Berkov, and C. A. Grace. 2011. *Tropical Plant Collecting: From the Field to the Internet*. NYBG Press.
- Oliveira-Filho, A. T., K. G. Dexter, R. T. Pennington, M. F. Simon, M. L. Bueno, and D. M. Neves. 2021. On the floristic identity of Amazonian vegetation types. *Biotropica* 53: 767–777.
- Pennington, R. T., M. Lavin, H. Ireland, B. Klitgaard, J. Preston, and J. M. Hu. 2001. Phylogenetic relationships of basal papilionoid legumes based upon sequences of the chloroplast *trnL* intron. *Systematic Botany* 26: 537–556.
- Pennington, R. T. 2003. Monograph of *Andira* (Leguminosae-Papilionoideae). *Systematic Botany Monographs* 64: 1–143.
- Pires, J. M. and Prance, G. T. 1985. The vegetation types of the Brazilian Amazon. Pp. 109–144 in *Amazonia: key environment series*, eds. G. T. Prance and T. E. Lovejoy. Pergamon Press.
- Pirie, M. D., B. B. Klitgaard, and R. T. Pennington. 2009. Revision and biogeography of *Centrolobium* (Leguminosae – Papilionoideae). *Systematic Botany* 34: 345–359.
- Polhill, R. M. 1981. Sophoreae. Pp. 213–230 in *Advances in Legume Systematics*, eds. R. M. Polhill and P. H. Raven. Royal Botanic Gardens, Kew.
- Ramos, G., H. C. de Lima, G. Prenner, L. P. de Queiroz, C. E. Zartman, and D. Cardoso. 2016. Molecular systematics of the Amazonian genus *Aldina*, a phylogenetic enigmatic ectomycorrhizal lineage of papilionoid legumes. *Molecular Phylogenetics and Evolution* 97: 11–18.
- Ramos, G. and D. Cardoso. 2016. Towards a taxonomic revision of the papilionoid legume genus *Aldina*: lectotypification of a name published in Flora Brasiliensis. *Neodiversity* 9: 1–3.
- Rehder, A., C. A. Weatherby, R. Mansfeld, and M. L. Green. 1935. Conservation of later generic homonyms. *Bulletin of Miscellaneous Information* (6-9): 350.
- Roy, M., H. Schimann, R. Braga-Neto, R. A. E. da Silva, J. Duque, D. Frame, F. Wartchow, and M. A. Neves. 2016. Diversity and distribution of ectomycorrhizal fungi from Amazonian lowland white-sand forests in Brazil and French Guiana. *Biotropica* 48: 90–100.
- Smith, M. E., T. W. Henkel, M. C. M. Aime, A. K. Fremier, K. Alex, and R. Vilgalys. 2011. Ectomycorrhizal fungal diversity and community structure on three co-

- occurring leguminous canopy tree species in a Neotropical rainforest. *New Phytologist* 192: 699–712.
- Sprent, J., J. Ardley, and E. K. James. 2017. Biogeography of nodulated legumes and their nitrogen fixing symbionts. *New Phytologist* 215: 40–56.
- Stergios, B. D. and G. A. Aymard. 2008. A striking new species of *Aldina* (Fabaceae-Swartzieae-Aldiniinae) from the Venezuelan Guayana Highlands. *Harvard Papers in Botany* 13: 29–33.
- Steyermark, J. A., P. E. Berry, K. Yatskievych, and B. K. Holst. 1967. Flora del Auyan-Tepui. 5: 248.
- Steyermark, J. A. 1999. Fabaceae. Pp. 213–230 in *Flora of the Venezuelan Guayana*, eds. G. A. Aymard et al. Missouri Botanical Garden, Timber Press.
- Stropp, J., P. van der Sleen, P. A. Assunção, A. L. Silva, and H. ter Steege. 2011. Tree communities of white-sand and terra-firme forests of the upper Rio Negro. *Acta Amazonica* 41: 521–544.
- ter Steege, H., N. C. A. Pitman, O. L. Philips, J. Chave, D. Sabatier, A. Duque, J. F. Molino, M. F. Prévost, R. Spichiger, H. Castellanos, P. von Hildebrand, and R. Vásquez. 2006. Continental-scale patterns of canopy tree composition and function across Amazonia. *Nature* 443: 444–447.
- ter Steege, H., N. C. A. Pitman, D. Sabatier, C. Baraloto, R. P. Salomão, J. E. Guevara, O. L. Phillips, C. V. Castilho, W. E. Magnusson, J. F. Molino, A. Monteagudo, P. N. Vargas, J. C. Montero, T. R. Feldpausch, E. N. Honorio-Coronado, T. J. Killeen, B. Mostacedo, R. Vasquez, R. L. Assis, J. Terborgh, F. Wittmann, A. Andrade, W. F. Laurance, S. G. W. Laurance, B. S. Marimon, B. H. Marimon Jr., I. C. G. Vieira, I. L. Amaral, R. Brienner, H. Castellanos, D. C. López, J. F. Duivenvoorden, H. F. Mogollón, F. D. A. Matos, N. Dávila, R. García-Villacorta, P. R. S. Diaz, F. Costa, T. Emilio, C. Levis, J. Schietti, P. Souza, A. Alonso, F. Dallmeier, A. J. D. Montoya, M. T. F. Piedade, A. Araujo-Murakami, L. Arroyo, R. Gribel, P. V. A. Fine, C. A. Peres, M. Toledo, G. A. Aymard C., T. R. Baker, C. Cerón, J. Enge, T. W. Henkel, P. Maas, P. Petronell, J. Stropp, C. E. Zartman, D. Daly, D. Neill, Ma. Silveira, M. R. Paredes, J. Chave, D. A. L. Filho, P. M. Jørgensen, A. Fuentes, J. Schöngart, F. C. Valverde, A. Di Fiore, E. M. Jimenez, M. C. P. Mora, J. F. Phillips, G. Rivas, T. R. van Andell, P. von Hildebrand, B. Hoffman, E. L. Zent, Y. Malhi, A. Prieto, A. Rudas, A. R. Ruschell, N. Silva, V. Vos, S. Zent, A. A. Oliveira, A. C. Schutz, T.

- Gonzales, M. T. Nascimento, H. Ramirez-Angulo, R. Sierra, M. Tirado, M. N. U. Medina, G. van der Heijden, C. I. A. Vela, E. V. Torre, C. Vriesendorp, O. Wang, K. R. Young, C. Baider, H. Balslev, C. Ferreira, I. Mesones, A. Torres-Lezama, L. E. U. Giraldo, R. Zagt, M. N. Alexiades, L. Hernandez, I. Huamantupa-Chuquimaco, W. Milliken, W. P. Cuenca, D. Pauletto, E. V. Sandoval, L. V. Gamarra, K. G. Dexter, K. Feeley, G. Lopez- Gonzalez, and M. R. Silman. 2013. Hyperdominance in the Amazonian tree flora. *Science* 342: 1243092.
- Thiers, B. (continuously updated) *Index Herbariorum: a global directory of public herbaria and associated staff*. New York Botanical Garden's Virtual Herbarium. Available at <<http://sweetgum.nybg.org/ih/>>. Accessed on 31 Aug 2018.
- Turland, N. J., J. H. Wiersema, F. R. Barrie, W. Greuter, D. L. Hawksworth, P. S. Herendeen, S. Knapp, W. -H. Kusber, D. -Z. Li, K. Marhold, T. W. May, J. McNeill, A. M. Monro, J. Prado, M. J. Prince, and G. F. Smith (eds.). 2018. International Code of Nomenclature for algae, fungi, and plants (Shenzhen Code) adopted by the Nineteenth International Botanical Congress Shenzhen, China, July 2017. *Regnum Vegetabile* 159.
- Werner, G.D.A., W. K. Cornwell, J. I. Sprent, J. Kattge, and E. T. Kiers. 2014. A single evolutionary innovation drives the deep evolution of symbiotic N₂-fixation in angiosperms. *Nature Communications* 5: 4087.

Chapter 6

General conclusions and future work

6.1 General conclusions

In this thesis, I have addressed how the distinct flower morphologies found in the papilionoid legumes (Papilionoideae, Leguminosae) have evolved. Through a series of studies presented as chapters, here I have used an integrative approach involving genomic, transcriptomic, developmental, and morphological data (Chapters Two, Three and Four). There is a considerable focus on the TCP family of transcription factors and their immediate network, which early studies (e.g., Citerne et al., 2003) showed to be involved in floral form in papilionoid legumes. There have been recent reviews of such genes across legumes (e.g., Zhao et al., 2019; CYCLOIDEA (CYC)) and angiosperms more generally (e.g., Madrigal et al., 2019; DIVARICATA (DIV) and RADIALIS (RAD)). However, these either did not sample many legumes (Madrigal et al., 2019; just *Phaseolus vulgaris*), or failed to consider that papilionoid legumes actually encompass a diversity of floral morphologies because not all papilionoid genera have typical papilionate corollas. This thesis is distinct by focusing on these genera, which can have radial symmetry, multiple stamens, or lack petals (see Chapter One). Additionally it provides a taxonomic account of one such genus with radial flowers, *Aldina* (Chapter Five).

In Chapter Two, I investigated the CYC, RAD, DIV and DRIF in a phylogenetic context. We sequenced the genomes of *Cadia purpurea* (radially symmetric), *Cyathostegia mathewsii* (single petal and multiple stamens) and the papilionate *Andira inermis* and *A. anthelmia* in order to look for putative orthologs of these genes using a phylogenetic approach. Results showed high copy number of CYC genes in the papilionate *A. inermis* (four copies) and *A. anthelmia* (four copies) and in the radially symmetric *Cadia* (four copies). Only two copies of CYC were found in *Cyathostegia*. These results in the context of prior studies that sampled these TCP and MYB genes in other legumes (Citerne et al., 2003; Zhao et al., 2019) suggest that a similar number of multiple CYC copies found in papilionoid genera with unusual flowers morphologically argues against the premise that a higher number of CYC copies is correlated with strong zygomorphic symmetry (Zhao et al. 2019). For example, there are also high copy numbers in taxa with radial flowers such as *Cadia* (four copies – see Chapter Two) and *Leptolobium elegans* (Citerne et al. 2003). This is a similar number of copies of CYC2 as found in the Non-Protein Amino Acid Accumulating (NPAAA) clade where the typical papilionoid flower is uniform across all species. In failing to consider the full diversity of floral form across Papilionoideae and legumes more generally, the conclusion of Zhao et al. (2019) that multiple copies of CYC2 are central to legume zygomorphy seems simplistic.

In Chapter Three, I analysed the differential expression putative genes involved in floral development and the candidate genes *CYC*, *DIV*, *RAD*, and *DRIF* in a series of vegetative and floral tissues from *Cadia purpurea*, *Cyathotegia mathewsii* and *Andira inermis*. Our results showed that several genes are differentially expressed in these species. For example, the expression of the phytohormone ethylene, the B-box protein-like (BBX) and CONSTANS (CO) were found to be up-regulated in the floral tissues of *Cadia*. These genes have been reported to influence flower development via mediation with light. We also found *CYC* orthologues to have low expression and with variability between replicates in *Cadia*, suggesting a dynamic expression pattern between developmental stages. All four *CYC* orthologues are expressed in the adaxial domains for the three replicates of each tissue, but only in one replicate in the abaxial domain, supporting adaxial function of the gene. All but one copy are also expressed in vegetative buds, confirming a broader role than just floral symmetry, which may constrain their evolution. Because the expression pattern between paralogues is largely homogeneous, there is no strong evidence for neo-functionalization in the duplicated genes. *RAD* showed more variation between paralogues with overall low expression between adaxial and abaxial regions of mature flowers. However, more replicates in the adaxial region showed higher expression relative to abaxial, confirming *RAD*'s activity in dorsal floral domains. *DIV* had a similar pattern as *CYC*, though at a generally higher expression level with more activity in the adaxial than abaxial regions for all but one paralog. This result does not agree with *DIV*'s pattern of expression, which is commonly associated with abaxial identity. It is possible that *DIV*'s role in dorsiventral symmetry is retained by only a single paralog. *DRIF* was expressed in adaxial samples, mature flowers and the vegetative buds with little variation between paralogues. Expression overall was quite low, meaning that variation between tissues is likely not significant.

In Chapter Four, I reviewed the floral ontogeny of papilionoid legume species with contrasting morphologies to revisit taxa of unusual flower morphologies across the subfamily and shed light onto the ontogenetic pathway of these lineages in order to gain insights on the evolutionary significance of morphological changes. The focus is mostly on describing the evolution of petal heterochronical patterns and androecium symmetry because these floral features have been shown to be variable in legumes and have only been studied in isolated fashion in classic ontogenetic studies. This is the first study that gathers ontogenetic data into a comparative phylogenetic framework to explain the lability of floral forms in Papilionoideae through the ontogenetic analysis of petal growth and androecium symmetry. Our results

revealed that petal heterochrony and androecium symmetry have high evolvability within early-branching Papilionoideae. The ancestral character state estimations showed that the asymmetrical androecium arose independently several times, and that both androecium types evolved independently of staminal fusion or floral architecture. For the heterochronical patterns of petal growth, our estimations revealed fast growth in the ancestral state in Papilionoideae, with slow petal growth having evolved relatively early in the subfamily. Multiple independent evolutionary reversals to fast growth occurred in some clades. The species-rich NPAAA clade, where all genera have the consistent presence of papilionate flowers, is also largely marked by the evolutionary conservatism of a slow petal growth and symmetric androecium. These findings largely coincide with high floral evolutionary lability in these early-diverging clades, where many genera have evolved their own non-papilionate floral architecture that greatly deviates from the highly specialised papilionate flower found in most derived lineages such as the NPAAA.

In Chapter Five, I presented a taxonomic synopsis of the Amazonian papilionoid legume *Aldina*, which is a genus with unusual non-papilionate flower morphology. Flowers of *Aldina* are radially symmetrical, with numerous free stamens, unlike its sister genera *Andira* and *Hymenolobium*, whose zygomorphic flowers are typically papilionate and with 10 stamens. *Aldina* taxonomy is in need of a revision since the preliminary taxonomic studies on the genus date back 70 years (Cowan 1953, 1958, 1961, 1984; Gontscharov 2006, 2009). This chapter therefore contributes to understanding the species diversity in Amazonia, filling in the gaps of plant diversity in this under-explored region. Eighteen species are recognized in *Aldina*, with seven names newly placed in synonymy, new lectotypification of five names, and a new combination, *Aldina auyantepuiensis*, is proposed.

6.2 Future work

This thesis used different approaches to shed light on the morphological diversity in flower form found across the Papilionoideae subfamily of the legumes. Future studies will also need to take such an interdisciplinary approach combining genome sequencing and transcriptomics, if we are to understand the genetic control and evolution of floral form in the species rich and economically important papilionoid legumes.

Production of fully annotated, chromosomally continuous genomes was beyond the scope of this PhD project. Whilst our genomes of *Cadia*, *Cyathostegia* and *Andira* were

sufficiently complete to allow identification of multiple loci for some of the candidate TCP and MYB genes, it is likely that in some cases more gene copies are present, and understanding their chromosomal location would be of great use in understanding copy number. Whilst in Chapter Two I was able to produce phylogenies of CYC, RAD, DIV and DRIF that sampled more legume species than prior studies, to fully understand variation in copy number across the legumes, more taxa need to be surveyed. Such additional surveys should include more papilionoid legume genera with unusual, non-papilionoid corollas, but interestingly, there are few studies of non-model and crop genera in the NPAAA clade, which has a constrained papilionoid morphology. Also, in legumes more widely, the study of Zhao et al. (2019) included very few genera from subfamily Detarioideae, which is florally heterogeneous, and it is interesting to note that Zhao et al. (2019) do not highlight that there are non-papilionoids with zygomorphic corollas such as *Bauhinia* and *Caesalpinia* that also have multiple copies of CYC2 (see Zhao et al., 2019, Table 1).

The RNASeq work suggests a number of possible future approaches. A full developmental series would be a great value in identifying the key players in the establishment of floral form in legumes. This is dependent on obtaining flowering material, either from greenhouse collections with ideal cultivation conditions or, better, from collection in the wild with local RNA extraction to limit the dangers of sample degradation in transit. In-situ hybridisation would also be useful in establishing details of expression patterns, particularly in genes with very low levels of expression - here the limitation is again flowering material. A key point we have revealed here is the variation in copy number of these key dorso-ventral patterning genes. Although we do not find strong evidence for neofunctionalisation, with more detailed work on expression, or with functional characterisation we may be able to discover why legumes have retained the high copy number of these genes, and if they are truly redundant.

The wealth of available floral developmental studies in legumes since Dr. Shirley Tucker's pioneering works have enabled large-scale comparative analyses to better understand the homologies among the highly diverse legume flowers. Here we have contributed with new ontogenetic data from morphologically key papilionoid genera exhibiting a range of divergent floral architecture to explore evolutionary shifts and conservatism of floral ontogeny across the Papilionoideae phylogeny. While we have gained insight into the evolution of the petal heterochrony and androecium symmetry in this florally heterogeneous subfamily, there are several other ontogenetically significant traits yet to be described and analysed across numerous

genera. For example, a future evolutionary estimation of traits like directionality of organ development (e.g., Tucker 1987), presence or suppression of petal primordia in cases of petal losses (e.g., Tucker 1988; Leite et al. 2015), the very rare open carpel (e.g., Prenner et al. 2015), as well as the multicarpellate gynoeceium (e.g., Paulino et al. 2014) could be valuable in gaining deeper insights into origins and diversification of the Papilionoideae's unique floral morphologies.

Although I have presented in Chapter Five the thus far most complete taxonomic assessment of the previously taxonomically and nomenclaturally understudied genus *Aldina*, this requires more work to develop into a full taxonomic monograph of the genus. The taxonomic synopsis of *Aldina* is a start point towards its comprehensive revision that will include detailed morphological descriptions, new photographic records and illustrations for all species. A large amount of morphological data has been indeed compiled during my visits to important herbarium collections, which will certainly be fundamental to my ongoing systematic studies of *Aldina*.