THE UNIVERSITY
of EDINBURGH

Floral biology and taxonomic complexity in the genus *Alpinia* Roxb. (Zingiberaceae)

Surabhi Ranavat

Doctor of Philosophy

School of Biological Sciences
Institute of Evolutionary Biology
Royal Botanic Garden Edinburgh

The University of Edinburgh
January 2021

Declaration

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

Surabhi Ranavat

31st January 2021

Acknowledgments

I would like to thank my supervisor Alex Twyford, who has been very supportive and encouraging throughout the PhD. I'd like to thank Mark Newman, who familiarised me with the Zingiberaceae taxonomy and made me realise that taxonomy is not boring after all! And Vinita Gowda, who introduced me to the weird and wonderful world of plant ecology and evolution, and gingers in particular.

The past four years in Edinburgh have been wonderful, and I've got to learn so much from my friends. A big thank you to all my friends at the RBGE- Lucía, Flávia, Cynthia, Madhavi, Natalia, Kristine, Pakkapol, Hannah, Gustavo, Yun-Yu, Julieth, Andy, Wu. I miss our evenings cooking together, solving cryptic crosswords in the office and canteen, visits to the Orchard and ceilidh-ing together. To Lucia, Cynthia and Madhavi, for always being there for me.

To all my friends in Ashworth, especially from my PhD cohort- Jon, Christina, Saudamini, Tom, Megan, Eevi, Jim, Max, and Rory. I will miss our pub quizzes playing the large cards and Mafia at the Old Bell Inn, or just meeting in DDH for coffee and crosswords. My time in Ashworth has been wonderful thanks to you all.

Thanks to Max Brown for helping with the statistical analyses and Hannes Becher for helping with the genomic analyses. I would also like to thank Sujai Kumar and Mark Blaxter for allowing me to use the ClubAshworth cluster for genomic analyses.

To my flatmates (old and new) Yuchen, Emily, Ameil, Safiye, and Bea- this whole lockdown was bearable (dare I say pleasant) only because of you all. Thanks for all the wonderful evenings of food, music, and occasional dancing and for being wonderful office mates. And for always motivating me to work when I was feeling down.

I am grateful to Helen Yeats, Steve Willis, David Tricker, Paulina Maciejewska-Daruk, Louise Galloway, Sadie Barber, and Fiona Inches for helping me with the

crosses and harvest. Sophie Haupt for arranging space for my *Alpinia* babies in the glasshouse at Kings' buildings, Pat Watson and Bill Adams for helping me plant the seeds and looking after my plant babies from the hybrids at the Kings' Buildings glasshouses.

Thanks to members of the **Tr**opical **E**cology and **E**volution (TrEE Lab, IISER Bhopal)- Ajith, Prasanna, Saket and Preeti for helping me with fieldwork in the North-East India. The fieldwork would not have been possible without Siddhant Shetty and Susnata Salony- thanks for being great fieldwork companions!

I would like to thank John McNeill for his advice on nomenclature and E, K, BM, LINN, and LIV for the specimens.

I would also like to thank The Darwin Trust of Edinburgh for funding this PhD. I would like to thank the Davis Expedition Fund, Stanley Smith Horticultural Trust, and the Genetics Society for funding my fieldwork.

Special thanks to my examiners Mark Hughes and Chelsea Specht, for your valuable input to the thesis and interesting discussion during the viva. I had a wonderful time and learned so much more about Zingiberaceae and plant taxonomy, ecology and evolution in general.

Lastly, I'd like to thank my family for always being extremely supportive and caring.

Abstract

The tropics harbour the highest number of species in the world. Uncovering why and how these regions are so species-rich has been a central theme in evolutionary biology. When it comes to plants, about two-thirds of the plant species are present in the tropics. Research on tropical plants is impeded by an array of taxonomic challenges. To understand these challenges in more detail, I have focused on the genus *Alpinia* Roxb. from the Zingiberaceae. It is the largest genus in the ginger family (c.250 species) and one of the most taxonomically complex. Molecular phylogenetic studies have revealed that this genus is polyphyletic. Moreover, nomenclatural confusion is an issue in *Alpinia* as many species with validly published names were described before the development of the type concept which has led to taxonomic confusion regarding species identities. Several *Alpinia* species also possess intermediate morphological characters that cause difficulty in delimiting species. Understanding the extent of reproductive isolation and hybridisation is crucial in this case. These species also possess a unique stylar dimorphism termed flexistyly, a strategy used to promote outcrossing. Many studies have investigated the reproductive biology of flexistylous species in *Alpinia*, but not much is known about the genetic basis of this trait.

Accordingly, the aims of my thesis are:

- To understand the taxonomic challenges present in the tropics.
- To resolve taxonomic uncertainty in a group of Indian *Alpinia* species.
- To investigate the crossing barriers and the potential for hybridisation in the genus *Alpinia*.
- To elucidate the genetic basis of flexistyly.

I conducted a survey of plant biologists to investigate the taxonomic problems faced when working on tropical taxa, and reviewed the literature to find case studies where taxonomic issues impede research. I found that taxonomic challenges in the tropics are caused by practical issues associated with fieldwork and herbarium collections and biological factors such as rapid radiations, hybridisation, and phenotypic

plasticity. To resolve the nomenclatural confusion in a group of Indian *Alpinia* species, I examined a range of original material and have resolved the confusion associated with the name *A. bracteata* used by Roscoe and Roxburgh. I have also assigned a lectotype for the name *A. calcarata* (Haw.) Roscoe. To investigate crossing barriers, I performed artificial hybridisation between and within clades of *Alpinia* s.l. I found that *Alpinia* species show widespread interspecific cross compatibility, especially within clades, and in a few instances, between divergent clades as well as genera. I also found a negative correlation between the genetic distance and the seed set, but this correlation was not significant. To investigate the genetic basis of flexistyly, I assembled a draft genome of *Alpinia nigra* that served as a reference for downstream analysis. I used a Pool-Seq approach to investigate the allele frequency differences between bulks of the anaflexistylous and cataflexistylous morphs. Most of the genome showed low differentiation (average genome-wide F_{ST}=0.04), with no clear outlier regions. This is consistent with the lack of evidence for a large inversion. This might be due to the complex nature of the genomic region(s) that govern this trait or it could be a single gene that might be difficult to detect. The potential for hybridisation and the lack of reproductive isolation might be common in gingers, and overlooked in many tropical taxa. It may be a critical factor that caused taxonomic complexity within this tropical genus.

Lay summary

The tropics are home to the highest number of plant species in the world. Understanding why and how these regions harbour so many species is one of the main aims in evolutionary biology. Carrying out research on tropical plants is hindered by many challenges. The aim of this thesis is to understand the challenges faced by plant taxonomists when working in the tropics. For this, I conducted a survey of plant biologists and reviewed the literature to find the problems faced when working on tropical plant groups. My research found that these issues are caused by fieldwork and herbarium-related problems, or biological factors underlying the formation of new species.

To understand the challenges of tropical plant groups in more detail, I focussed on a genus from the ginger family called *Alpinia*. It is the largest genus in the family with more than 250 species spread throughout tropical and subtropical Asia. Studies using genetic tools have revealed that the relationships in this genus were more complex than previously thought. Some groups of species from *Alpinia* were more closely related to other genera than groups from the same genus. It is a difficult genus to work with, as many species have not been assigned the correct names. Therefore, one of my aims is to clarify the confusion associated with names in a group of *Alpinia* species from India. I examined the original material and description of species and found that the same names were applied to different species. Moreover, it is often difficult to distinguish species that are closely related due to similar traits. This could due to the formation of hybrid species. Therefore, I performed artificial pollination between closely-related and distant species to investigate the potential for cross-breeding. I found that closely related species are capable of setting seed more often although some crosses with distantly related species were successful too.

Alpinia species also exhibit a unique strategy called flexistyly, that prevents self-pollination. Many studies have investigated its ecological and adaptive importance, but little is known about its genetic basis. Therefore, I examined the genetic

differences between the floral types. I did not find any clear pattern of genetic differences as the region of the genome that controls this trait is difficult to detect.

As many tropical plants are now under threat, understanding basic taxonomy and the processes that have led to the formation of new species is crucial for conservation. Therefore, the taxonomic, ecological, and genomic studies presented in this thesis will be valuable information for conservation assessments.

Table of Contents

Chapter 1

General Introduction

1.1 Plant diversity in the tropics

The tropics harbour the greatest number of species in the world. The presence of a latitudinal diversity gradient, with a higher concentration of species near the equator and decreasing towards the poles, has been one of the most intensely studied of biodiversity patterns (Pianka, 1966; Hillebrand, 2004; Jablonski et al., 2006; Mittelbach et al., 2007; Brown, 2014; Mannion et al., 2014). Within the tropics, the Neotropics possess one-thirds of the total number of vascular plants in the world (118,308), followed by the Afrotropical region (56,451 species) and Southeast Asia (50,000) (Raven et al., 2020).

A plethora of historical and ecological hypotheses have been proposed to explain the latitudinal biodiversity gradient. There are two main contrasting hypotheses that have been widely discussed. One is the "out of the tropics" hypothesis, whereby lineages originate and undergo massive diversification in the tropics, and then disperse towards the poles (Jablonski et al., 2006; Rolland et al., 2014). The second is the "tropical niche conservatism" hypothesis states that most lineages originate in the tropics and have difficulties in dispersing and adapting to the temperate regions, thereby retaining more diversity in the tropics (Wiens and Donoghue, 2004; Buckley et al., 2010; Cooper et al., 2011; Romdal et al., 2013). Higher availability of energy leads to higher diversification rates in the tropics, higher availability of space causes more opportunities for isolation, and the stability of tropical climates leading to lower rates of extinction have also been hypothesised to play an important role in the origin of tropical diversity (Mittelbach et al., 2007 and references therein). In some studies, there is also an evidence for a balance between high extinction and speciation rates (Koenen et al., 2015).The tropics also harbour a wide range of habitat types from rainforests and dry forests to savannahs and deserts (Figure 1.1) (Antonelli and Sanmartín, 2011; Raven et al., 2020), with many biotic and abiotic evolutionary drivers at play in different regions (Antonelli and Sanmartín, 2011).

Figure 1.1: A few examples of different habitat types found in the tropics. A- Lowland evergreen forest (Nongkhyllem Wildlife Sanctuary, Meghalaya, India), B- Freshwater swamp (Keibul Lamjao National Park, Manipur, India), C- Grassland (Kas plateau, Maharashtra, India), D- Moist deciduous forest (near Wayanad, Kerala, India). Photographs by Surabhi Ranavat.

When it comes to large plant genera, some of the larger ones are predominantly present in the tropics such as *Begonia* L. (2000 species) (Hughes et al., 2015), *Bulbophyllum* Thouars (c. 2000 species), *Piper* L. (c. 2000 species) and *Psychotria* L. (c. 1600 species) (POWO, 2019). Due to their large size, studies at the genus-wide level are difficult to carry out (Frodin, 2004, but see Muñoz-Rodríguez et al., 2019), therefore, systematic analyses involving representative taxa from geographic or infrageneric groups have been done (Jaramillo et al., 2008; Hosseini et al., 2016; Moonlight et al., 2018), that have resolved many infrageneric relationships. Although many large genera are predominantly present in the tropics, some have a cosmopolitan distribution such as *Astragalus*, *Solanum*, *Euphorbia*, and *Senecio* (Frodin, 2004).

1.2 Evolutionary drivers of tropical diversity

The highly diverse tropical lineages tend to be separated across ecological gradients where they occupy specialised niches. This prevents the species from dispersing across barriers. This dispersal limitation causes to reproductive isolation between species, which in turn promotes allopatric diversification (Kozak and Wiens, 2010; Salisbury et al., 2012). Factors such as rapid shifts in climatic niches among species have been found to play an important role in driving diversification in the tropics (Kozak and Wiens, 2007; McCain, 2009; Cadena et al., 2012; Castro-Insua et al., 2018).

The formation of reproductive barriers is considered to be crucial for speciation to occur. Pre-pollination barriers include niche differentiation, phenological isolation and pollinator specialisation whereas post-pollination barriers include competition between conspecific and nonconspecific pollen, low hybrid seed formation, and low F1 viability and fertility (Baack et al., 2015). Barriers in both the stages act together to maintain species boundaries, but the strength of these barriers are highly variable (Scopece et al., 2010). This has been investigated in a few tropical groups by artificially pollinating taxa that has given an insight into the strength of reproductive barriers, and therefore, mechanisms that are involved in the evolution and maintenance of species (Palma-Silva et al., 2011; Chen, 2013; Pinheiro et al., 2013; Twyford et al., 2015).

Rapid radiation is one of the factors considered to be responsible for the large diversity of plant species present in the tropics. It is a process in which lineages diversify rapidly due to climatic changes, new ecological opportunities, or both (Richardson et al., 2001; Knope et al., 2012; Lagomarsino et al., 2016; Carlsen et al., 2018). The Pleistocene climatic changes in tropical regions led to forest disturbance and the resulting ecological change may have led to recent, rapid radiations (Harris et al., 2000; Richardson et al., 2001; Pennington et al., 2004). Similarly, the uplift of major mountains, such as the Andes or the Himalayas, has also played an important role in the rapid diversification of lineages (Hughes and Eastwood, 2006; Antonelli and Sanmartín, 2011; Wang et al., 2012). Speciation in the tropics may also have

accelerated due to the evolution of novel traits that promote diversification, or key innovations (Sanderson and Donoghue, 1994; Drummond et al., 2012; Silvestro et al., 2014).

Processes such as natural hybridisation are regarded important as drivers of biodiversity, as they are found to be more common than previously thought (Soltis and Soltis, 2009). Hybridisation is regarded as a creative force in plant evolution and is more common in some taxa than others (Ellstrand et al., 1996). While it is not currently considered to be a key driver of diversity in tropical groups, new studies are finding this to be more common (Košnar et al., 2010; Schley et al., 2020) and re-evaluating the role of hybridisation in species diversity.

1.3 Reproductive diversity in angiosperms

The remarkable diversity of angiosperms that the tropics harbour also displays a remarkable range of traits and floral forms. These traits are often adaptations to ensure pollination by means of a pollen vector (Waser and Ollerton, 2006; Barrett, 2010a). Pollinators are considered to be key drivers of angiosperm diversity in terms of adaptation as well as diversification (Stebbins, 1970; Van der Niet et al., 2014), where the structural variety of floral forms is presumed to have resulted from natural selection during interactions with floral vectors (Figure 2) (Harder and Johnson, 2009). When it comes to the tropics, the number of plant pollination systems present is considerably higher when compared with other latitudinal zones (Waser and Ollerton, 2006).

Most angiosperms are hermaphroditic in nature and thus fitness is determined by pollen export as well as pollen receipt (Lloyd, 1982). This sexual condition also comes with a reproductive cost where opportunities for self-fertilisation or sexual interference may occur. To circumvent this problem, angiosperms have evolved highly specialised structures or mechanisms that prevent pollen-pistil interference and promote outcrossing (Barrett, 2003). Herkogamy, the spatial separation of sexual organs within a flower, is one such strategy (e.g., *Primula* L. (Darwin, 1877), *Passiflora* L.(Dai and Galloway, 2011), *Melochia* L. (Faife-Cabrera et al., 2018)). In

some cases, reciprocal herkogamy might be present where the positions of the male and female sex organs are reciprocal. The separation of sexual function can also be temporal (dichogamy) where the timing of pollen release and the stigma receptivity are different (e.g., *Alpinia* Roxb. (Li et al., 2001), *Carya* Nutt. (Thompson and Romberg, 1985), *Campanula* L. (Inoue, 1990)).

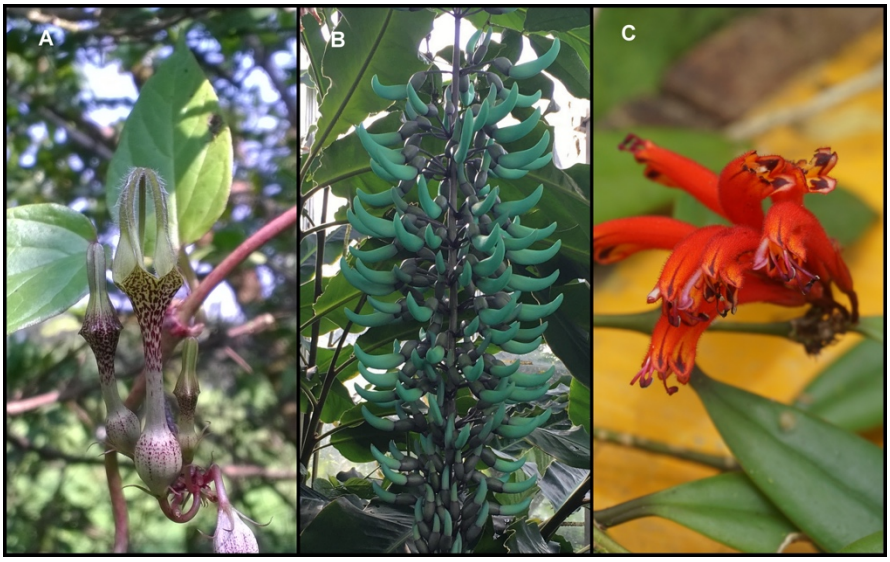

Figure 1.2: Examples of diverse tropical angiosperms adapted to different biotic pollinators. A- *Ceropegia vincifolia* adapted to fly pollination. B- *Strongylodon macrobotrys* adapted to bat pollination. C- *Aeschynanthus* sp. adapted to sunbird pollination. Photographs by Surabhi Ranavat.

Although hermaphroditism is the predominant sexual system in flowering plants, they can also be dioecious (separate sexes), although dioecy is relatively infrequent across angiosperms (~6% of species) (Ashman, 2007; Renner, 2014). Dioecious taxa are found in higher proportions in the tropical plants that have a woody growth form (Vamosi and Vamosi, 2004; Käfer et al., 2017). They also have an array of mating patterns, which can range from outcrossing to predominant selfing (except in dioecious taxa) (Barrett, 2002b). Selfing might be detrimental as it can lead to inbreeding depression (Charlesworth and Charlesworth, 1987) but can serve as a strategy for reproductive assurance under the absence of pollination (Eckert et al., 2007).

1.4 Zingiberaceae

The ginger family or the Zingiberaceae is the largest in the order Zingiberales with 58 genera and over 1800 species (Newman, pers. comm.). This family is mainly distributed in the tropics of South-East Asia, with a few genera extending as far as Africa and the Americas. Some genera such as *Roscoea*, *Cautleya*, and *Alpinia* are found in subtropical and even temperate regions (Kress and Specht, 2005). Species from this family are a common component of the understorey flora of wet tropical forests. They are characterised as herbaceous perennials with distichous leaves and inflorescence terminal on the leafy shoot or growing on a separate, leafless shoot. Most flowers are bisexual and zygomorphic that are open only for a day. In some cases, there are some species with separate hermaphrodite flowers and functionally male flowers on each inflorescence from *Alpinia* Clade III (Smith, 1990; clade system sensu Kress et al., 2005) and *Globba* section *Nudae* (Sangvirotjanapat et al., 2019, 2020). The flowers possess only one stamen, with the other stamens modified as petaloid staminodes (Larsen et al., 1998).

The ginger family is economically important as many species are used in medicine, horticulture, and cooking. Spices such as ginger (rhizome of *Zingiber officinale* Roscoe), turmeric (rhizomes of *Curcuma longa* L.) and cardamom (fruit of *Elettaria cardamomum* (L.) Maton) are widely used in cooking. Other spices such as galangal (rhizomes of *Alpinia galanga* (L.) Willd.) and Melegueta pepper (seeds of *Aframomum melegueta* (Roscoe) K.Schum.) are commonly used in their native regions (Rangsiruji, 1999; Leong-Škorničková and Newman, 2015). Species from the genera such as *Hedychium*, *Alpinia*, *Curcuma*, and *Etlingera* are commonly grown as ornamentals. *Curcuma* species are also used as a source of yellow dyes. These species are also important from an ethnobotanical perspective, as many local species are used as herbal medicines to treat several ailments (Daimei and Kumar, 2014; Phuong Hanh and Quoc Binh, 2014). Many species in this family are of pharmacological importance as they contain compounds that have anti-inflammatory, anti-bacterial and anti-fungal effects and are used to treat digestive problems (Jatoi et al., 2007; Chen et al., 2008; Kumar et al., 2011).

Figure 1.3: Floral diversity in the Zingiberaceae. A- *Zingiber rubens*, B- *Kaempferia rotunda* L., C- *Curcuma* sp., D- *Hedychium spicatum*, E- *Globba* sp., F- *Alpinia galanga* (L.) Willd., G- *Elettaria cardamomum*, H- *Etlingera linguiformis*. A-E are from subfamily Zingiberoideae and F-H from Alpinioideae. Photographs by Surabhi Ranavat.

Genera from the Zingiberaceae exhibit a variety of strategies to ensure reproduction. They reproduce sexually as well as asexually by means of rhizome or bulbils (Larsen et al., 1998). Vivipary, a phenomenon where the seed germinates within the fruit when it is still attached to the plant has also been observed in a few genera in the Zingiberaceae (see Ashokan and Gowda, 2018). In terms of sexual reproduction, species exhibit a wide variety of strategies to prevent self-pollination such as two-day flowers with a single, protandrous morph (Gao et al., 2004) or protandry and protogyny in two different floral morphs within a population (flexistyly) (Li et al., 2001). The extensive floral diversity in the Zingiberaceae (Figure 1.3) suggests that many interesting mating systems and floral traits are yet to be discovered in this family (Kress and Specht, 2005).

Many genera in the family possess species that have overlapping morphological characters hence the species limits are difficult to discern, e.g., *Curcuma* and *Hedychium*, where many species are part of a complex (Záveská et al., 2012; Saryan et al., 2020). Moreover, many species in genera such as *Curcuma* (Záveská et al., 2012), *Globba* (Takano and Okada, 2002), *Kaempferia* (Nopporncharoenkul et al., 2017) are polyploid in nature therefore resolving species relationships is challenging. Some species also tend to hybridise in the wild whereby the progeny becomes naturalised (Leong-Škorničková et al., 2007).

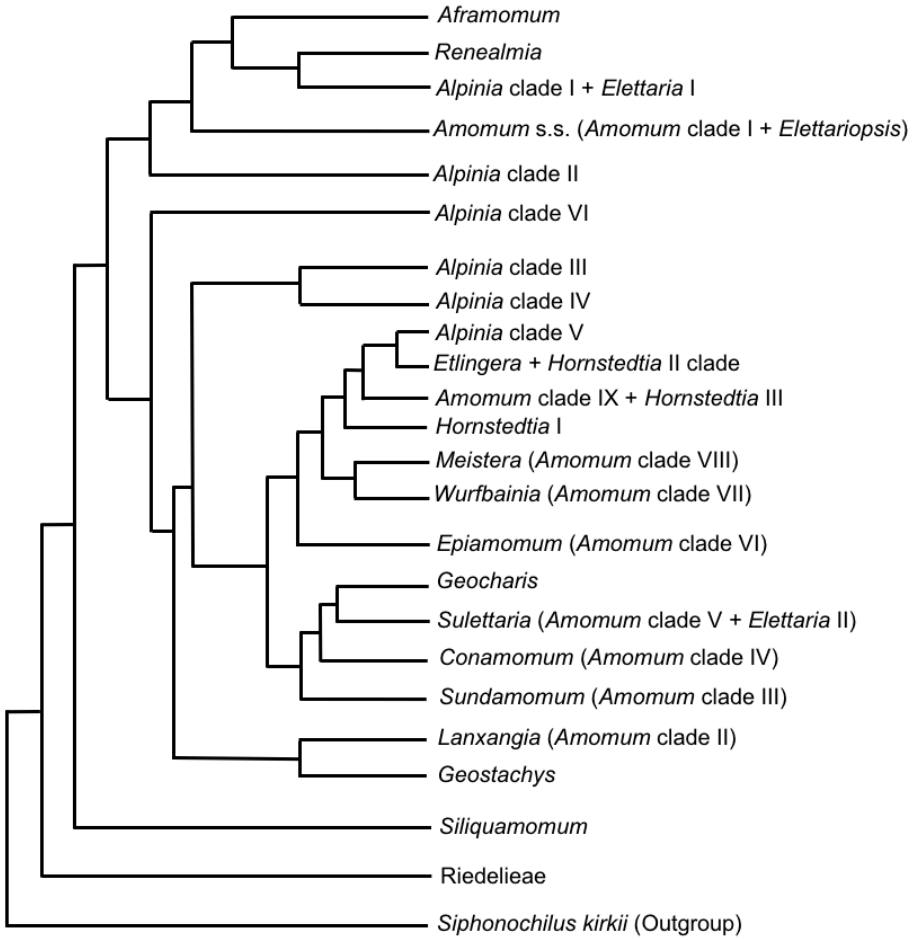

Figure 1.4: A condensed tree of the Alpinioideae from analysis of nrITS and matK sequence data (modified from de Boer et al., 2018). The clades are numbered according to Kress et al., 2005 (*Alpinia)* and de Boer et al., 2018 (*Amomum, Hornstedtia*).

Discerning boundaries even at the generic level has been problematic. Many genera in this family are taxonomically complex, with recent molecular work revealing the

polyphyletic nature of several of them (Xia et al., 2004; Kress et al., 2005, 2007). The same suite of morphological characters are found in multiple groups across the tree, which suggests that these characters are labile. Therefore, the generic boundaries need to be redefined. This work has already been done in the case of *Amomum* from the subfamily Alpinioideae, where the genus was previously found to be polyphyletic (Xia et al., 2004). *Amomum* formed nine clades across Alpinieae most of which have been elevated to the generic level (de Boer et al., 2018). Other genera such as *Alpinia*, *Elettaria* and *Hornstedtia* were also found to be polyphyletic, where some of their clades are more closely related to other genera than they are to each other (Figure 1.4).

Many species from the Zingiberaceae were described in the late 18th and early 19th century, which was before the development of the type concept (1867). The type is a specimen to which the name of an organism is permanently attached (Turland et al., 2018). In the absence of this concept, botanists assigned different names to the same species which caused taxonomic confusion (Leong-Škorničková et al., 2008). Additionally, the original descriptions of several species are vague which has led to misapplication of names (Leong-Škorničková et al., 2007). This problem is associated with many species that have been described from India, where many of the economically important ginger species are found. A large number of ginger species were described by botanists based in Europe, such as Roscoe, who never saw them in nature (Roscoe, 1807). Furthermore, Roxburgh, who did collect gingers in India, left relatively few herbarium specimens so we have to rely on his illustrations. Only a few of the 60 or so names published by Roxburgh have been typified. Early botanists also cited each other's work which has caused confusion in species concepts and nomenclature.

The most recent monographic account of the Zingiberaceae dates back to 1904 (Schumann, 1904), and there have been virtually no monographic studies since then (except for smaller genera such as *Aframomum* K. Schum. (Harris and Wortley, 2018)). In the latter half of the 20th century, it was difficult for botanists in neighbouring areas of China and India, where the same species may be expected, to

see each other's publications. These factors contribute to a lack of taxonomic work in the Zingiberaceae.

India is home to about 20 genera and over 200 species from the Zingiberaceae (Sabu, 2006) but more species are being described due to increased plant exploration in biodiversity hotspots such as North-East India (Thomas et al., 2013; Bhaumik et al., 2017; Joe et al., 2017; Thongam and Konsam, 2017; Hareesh and Sabu, 2018; Ashokan and Gowda, 2019; Odyuo et al., 2019; Sabu and Hareesh, 2020). Genera such as *Curcuma* L. and *Hedychium* J. Koenig have the highest diversity in India (Leong-Škorničková et al., 2007; Ashokan and Gowda, 2019). Given the complexity within the Zingiberaceae, more ecological and evolutionary studies are required to disentangle the confusion associated with the species described in the early days of ginger taxonomy.

1.5 Study system - *Alpinia* Roxb.

Alpinia[1] is the largest and one of the most taxonomically complex genus in the Zingiberaceae. It is made up of c. 250 species (Newman, pers. comm.) distributed from the Indian subcontinent to South-East Asia, China, Japan, the Pacific as far as Australia and Fiji (Figure 1.5) (Smith, 1990). *Alpinia* species are large herbs that are usually 2-3m tall except in some cases where they can be as tall as 10m in height (*Alpinia boia*, see (Burtt and Smith, 1972)). The inflorescence is usually terminal on the leafy shoot. The lateral staminodes are absent or represented by small, tooth-like structures (Figure 1.6). There are no characters that uniquely identify all species of *Alpinia* therefore plants that do not fit into any other genera in the Alpinieae tend to be added to this genus.

Within *Alpinia*, Clade I comprising three species is distributed across Sri Lanka and south India and is sister to genera such as *Renealmia*, *Aframomum*, and the Clade *Elettaria* I therefore represents a lineage that spreads from central America through

[1] As *Alpinia* is a polyphyletic genus, for the purpose of this thesis, *Alpinia* indicates *Alpinia sensu lato* unless stated otherwise.

Africa to south Asia and is found more basally branching. As this clade is closely related to genera present in Africa and tropical America, it is likely that its common ancestor travelled across the Indian Ocean with the breakup of Gondwana (Kress et al., 2005, 2007). *Alpinia* Clade II, comprising three species is distributed in continental Asia. It contains the type of the genus, *Alpinia galanga* (L.) Willd. (Rangsiruji, 1999; Kress et al., 2005). Species from *Alpinia* clade III are found in Fiji, the Caroline Islands, and Sulawesi and are generally east of Wallace's line (Kress et al., 2005). *Alpinia* Clade IV, the largest clade, is distributed throughout tropical Asia. The biogeographical patterns for this clade are not clear as is the case with morphological synapomorphies (Kress et al., 2005). *Alpinia* Clade V is made up of species found in Philippines, the Bismarck Archipelago, Australia, and the tropical Pacific (Kress et al., 2005, 2007). Alpinia Clade VI, comprising two species, is restricted to southern Thailand, peninsular Malaysia and Indonesia (Kress et al., 2005).

Although the origin of *Alpinia* is not known (Rangsiruji, 1999), the fossil genus *Zingiberopsis* (dated at 65 mya), assumed to be its nearest extant relative, was found in Western Northern America (Hickey and Peterson, 1978). Therefore, *Alpinia* and other Zingiberaceae taxa may have originated in Laurasia and migrated to Asia via Europe from the Middle to Late Eocene (Rangsiruji, 1999). Based on the phylogenetic tree in Figure 1.4, it is likely that *Alpinia* Clade I, *Elettaria* clade I, and African *Renealmia* evolved from a common ancestor (Rangsiruji, 1999; Kress et al., 2007). Species in *Alpinia* Clade II, that occur in India and China, are widely cultivated throughout South-East Asia. Therefore it is difficult to determine the native range of these species (Rangsiruji, 1999). Clade VI may have arisen post migration of *Alpinia* to South-East Asia in peninsular Malaysia and Indonesia (Rangsiruji, 1999). Clades III and V have presumably arisen due to the eastward and southeastward dispersal, and further northward direction towards the Philippines (Rangsiruji, 1999). Clade IV consists of several endemic island species but also species that are widespread across India, China, and Malesia (Rangsiruji, 1999). This implies that *Alpinia* s.l. may have undergone multiple dispersal events, and a study of the tribe Alpinieae is necessary in order to corroborate these assumptions.

Alpinia species are commonly found in low to mid-elevation forests, with some found even in montane forests. They are usually found in the forest understorey near light gaps, along water bodies, or at forest margins (Leong-Škorničková and Newman, 2015). They are predominantly pollinated by large bees but are pollinated by birds or even bats in some cases (Kress and Specht, 2005; Kress et al., 2005). Numerous species from this genus exhibit a unique floral dimorphism termed flexistyly, a strategy that combines temporal and spatial separation of sexual function via stylar movement. This strategy may have evolved to promote outcrossing and prevent pollen-pistil interference as *Alpinia* species are known to be self-compatible.

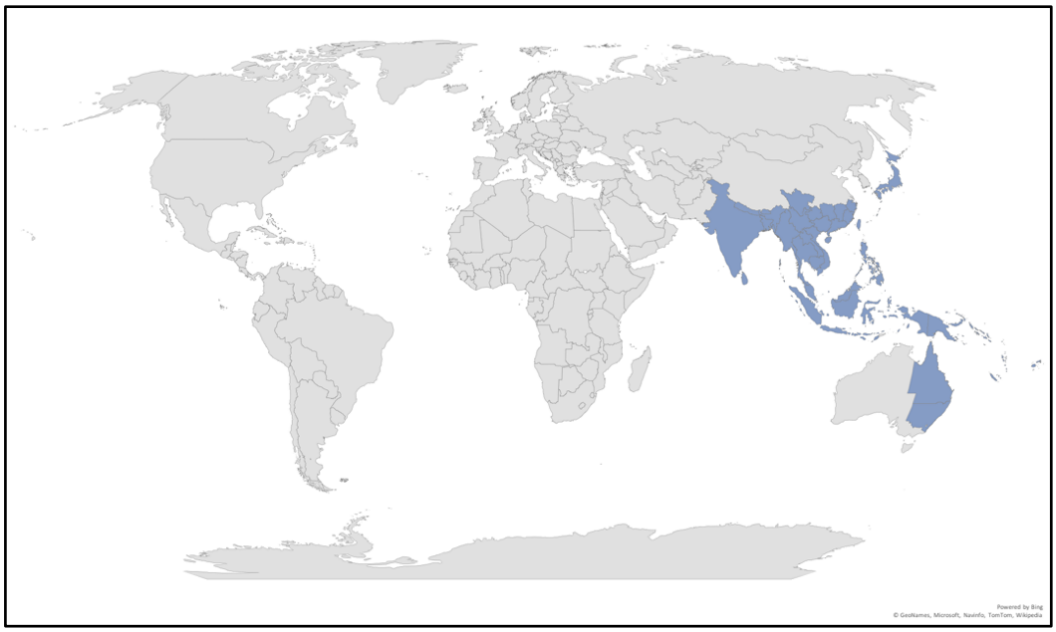

Figure 1.5: Geographical distribution of *Alpinia* species.

Alpinia serves as a good study system as it represents the challenges that are found in other tropical genera. Part of the reason why this genus is poorly studied is due to a lack of collections across its geographical range. Identification from herbarium specimens is challenging as the important distinguishing characters are poorly preserved (Ranavat, pers. obs.). Moreover, nomenclatural problems are common in this genus where the number of names applied (>500) are greater than the number of species present (The Plant List, 2013). When it comes to classification, *Alpinia* is regarded as a “dustbin” genus where all the species having a terminal inflorescence

that cannot be classified into any other closely related genera are included (Kress et al., 2005).

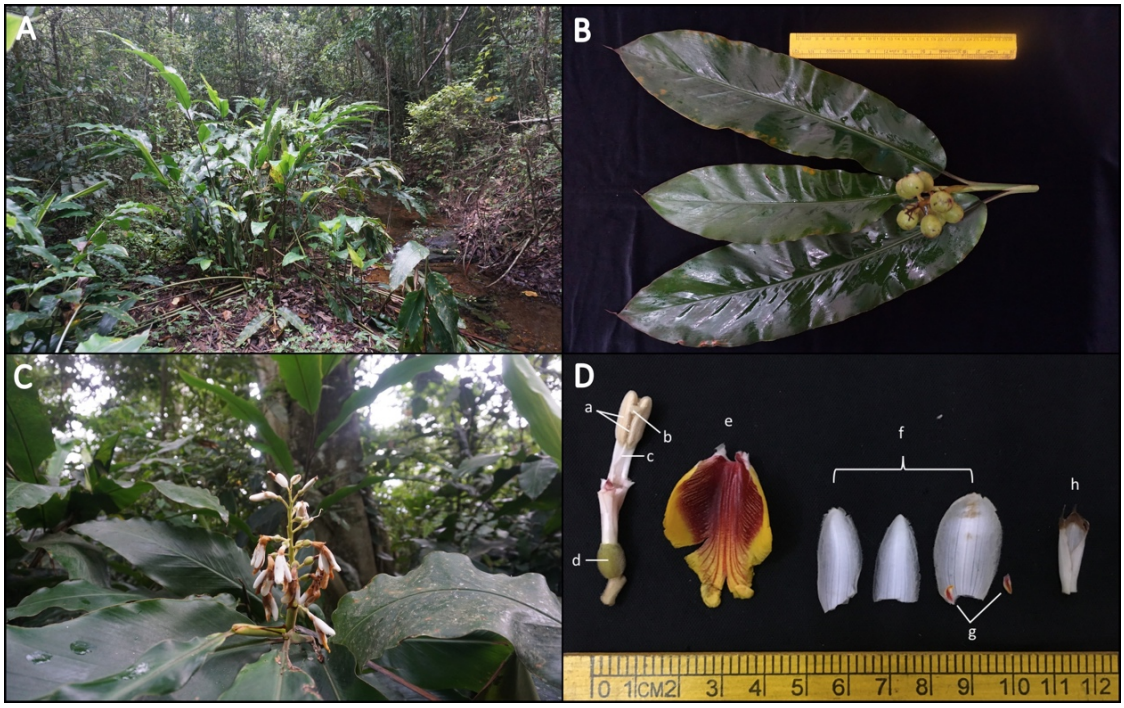

Figure 1.6: *Alpinia* cf. *malaccensis* found in Nelliyampathy, Kerala, India. A- Habitat (found along watercourses), B- Leaf and infructescence (scale=30cm), C- inflorescence terminal on the leafy shoot, D- dissected flower (a- anther, b- stigma, c- style, d- ovary e- labellum, f- corolla lobes, g- lateral staminodes, h- calyx). Photographs by Surabhi Ranavat.

Clarification of taxonomic uncertainty is essential for ecological and evolutionary studies, especially for genera such as *Alpinia*. An investigation of the evolutionary relationships within this genus revealed the polyphyletic nature of this genus (Rangsiruji et al., 2000; Kress et al., 2005). Consequently, *Alpinia* will ultimately be reduced to a small genus with only three species (represented by *Alpinia* Clade II, Figure 3). Therefore, its polyphyletic nature, combined with the difficulties in species delimitation and taxonomic problems and the presence of unique floral traits make *Alpinia* an interesting genus to study. In this thesis, I primarily study Clade IV, the largest clade in *Alpinia* (sensu Kress et al., 2005) which is still a cohesive group.

1.6 Objectives of the thesis

This thesis aims to document the taxonomic challenges researchers face when working on tropical plants and to tackle various aspects of the taxonomically complex tropical genus *Alpinia*. *Alpinia* is a polyphyletic genus that has a complicated taxonomic history, with many species being inaccurately circumscribed. Due to this, many names have been used incorrectly therefore untangling nomenclatural problems is key to understanding the diversity within this genus. Moreover, many species in *Alpinia* have similar morphological characters that make it difficult to tell species apart (S. Ranavat, pers. obs.). This may be due to the presence of hybridisation between species. Therefore carrying out cross-pollination across the different clades can reveal the potential for hybridisation across this genus. While natural hybridisation may or may not be common in *Alpinia*, it is known that species from this genus are self-compatible and therefore possess a stylar dimorphism termed flexistyly to prevent inbreeding. While it is known that cross-pollination between the two morph types often results in higher fruit set as compared to self-pollination or within-morph pollination, the genetic basis of this trait remains unknown.

With this background, my specific aims are:

- To understand the taxonomic challenges present in the tropics.
- To resolve taxonomic uncertainty in a group of Indian *Alpinia* species.
- To investigate the crossing barriers and the potential for hybridisation in *Alpinia*.
- To elucidate the genetic basis of flexistyly.

In chapter 2, I review the taxonomic challenges present in the tropics. I carried out a survey to investigate the taxonomic problems faced when working in the tropics, and review the literature to find case studies where taxonomic issues impede research.

To understand these challenges in more detail, I have focused on the genus *Alpinia*. In chapter 3, I resolve the nomenclatural confusion within a group of Indian *Alpinia*

species. Upon examination of original material and literature, I assign a lectotype for a species and clarify the confusion associated with the same names that were applied to different species.

Taxonomic complexity also arises when the species have intermediate morphological characters that cause difficulty in delimiting species, which is common in *Alpinia*. In chapter 4, I investigate the reproductive barriers within and between *Alpinia* clades. I artificially pollinated different species within and between clades of *Alpinia* to examine the potential for hybridisation across the genus.

Among the many strategies that exist to prevent self-pollination, a novel mechanism termed flexistyly is present in *Alpinia*. In chapter 5, I investigate the genetics of this floral dimorphism. I describe the sequencing and assembly of the draft genome of an *Alpinia* species and carry out NGS analysis to find the genomic region(s) that governs flexistyly.

In chapter 6, I synthesise the main conclusions from the four data chapters of the thesis, highlighting the need for more plant exploration and integrative research in the tropics, the challenges of genomic analyses in non-model plants, and the contribution of this study in a broader context.

Chapter 2

Taxonomic challenges in the tropics: understanding patterns and processes

This review was written in collaboration with Dr Alexandre Antonelli (Gothenburg Global Biodiversity Centre, Department of Biological & Environmental Sciences, University of Gothenburg and Royal Botanic Gardens, Kew, Richmond, Surrey, UK).

2.1 Introduction

The tropics harbour the greatest number of species in the world (Mittelbach et al., 2007; Brown, 2014). Understanding the causes and maintenance of such high diversity is one of the main goals of evolutionary biology. Of the estimated 374,000 plant species in the world (Christenhusz and Byng, 2016) about two-thirds are present in the tropics (Pimm and Joppa, 2015). Most of these species are concentrated in the Neotropics as compared to tropical Africa or Asia (Antonelli and Sanmartín, 2011). Such high diversity is often associated with taxonomic challenges for determining species identities. These challenges can have a profound impact on the study of biodiversity, systematics, conservation, ecology, evolution, among other fields.

Taxonomic challenges are not only limited to species/taxon complexes as previously discussed in Ennos et al. (2005) and Pinheiro et al. (2018), but extend to the many other, particularly species-rich, plant groups. While many taxonomic challenges are reported for plant groups present in temperate regions (Hollingsworth, 2003; Ennos et al., 2005, 2012; Robertson et al., 2010; Liu et al., 2011), this probably indicates a bias in the distribution of botanic institutions and researchers in temperate, relatively depauperate regions. It is clear, however, that the scale of the problem is just as large or likely to be larger (but less well-documented) in the tropics, where there are significant taxonomic challenges that act as impediments to research.

Solving taxonomic challenges in the tropics could have major implications for our understanding of ecological and evolutionary processes. For instance, in the world's largest tropical rain forest –Amazonia – it has been estimated that just 227 tree species account for over 50% of all individual trees in the Amazon (ter Steege et al., 2013). Are all of these 'hyperdominant' species true biological units or are some of them aggregates? And could some of the rarest species – such as the 6000 species estimated to have populations of less than 1000 individuals in ter Steege et al. (2013), or the c. 36.5% of all plants species identified as rare by Enquist et al. (2019) – in fact belong to more common, and therefore less threatened species?

In this review, we consider the patterns and the processes that contribute to the taxonomic challenges faced in tropical plants. Given this broad scope and limited documentation of many tropical groups, we conducted a survey that has helped us understand the issues that plant biologists face when working on tropical taxa (Box 1). Here, we first introduce the practical issues and biological factors underlying the complexity of taxa in the tropics. We then discuss the implications of these taxonomic challenges on aspects of biodiversity research and plant conservation. Finally, we consider prospects for resolving these challenges by using the latest tools and resources. Throughout this review, we use data from our survey and quotations from survey responses, in addition to our literature review.

Box 1: Taxonomic complexity in the tropics survey

Our survey asked 19 questions that covered a range of topics, from the background of the researcher to the biological and technical considerations associated with the researcher's taxa of interest (full survey questions, responses and summary given in Appendix 2.1 and Appendix 2.2). Researchers were asked to give detailed responses for one taxonomic group of interest. The survey was advertised on social media channels and on email lists between April and September 2019. We received 174 responses, of which 145 were considered valid (Appendix 2.1). Analysis of the results was carried out in R version 3.6.1.

Our survey responses reflected a range of career stages (with 37% Principal Investigators and 32% students). Researchers who study species in a defined geographical area (22%) rather than taxonomic group (79%) are more likely to investigate species in the Neotropics (63 %) than the African or Asian tropics (37%), though over 60% of all researchers did not solely work on tropical taxa. In total, taxa of interest spanned 57 plant families.

Box 1: continued

Key results from the survey include:

- 88% of responses consider their group of interest taxonomically complex (as opposed to 5% that didn't, 6% that replied "maybe" or "don't know", and 1% that didn't respond).
- 92% of researchers need to do fieldwork for their research and 84% of these said one or more practical factors relating to fieldwork are extremely or very problematic.
- A diverse range of biological factors, including hybridisation, phenotypic plasticity, and cryptic speciation impact the ease of telling tropical species apart.
- Recent monographs, improved phylogenies and digitised herbarium specimens are some of the many resources that would help improve tropical plant research.
- In the free text comments in the survey, many researchers stated that the lack of basic knowledge of species limits and species distributions impedes their research.

We discuss the survey results in more detail in the main text and highlight important topics emerging from the survey responses.

2.2 The nature of taxonomic challenges

Tackling taxonomic challenges in the tropics is problematic due to both *practical issues* and *biological factors*. Practical issues include difficulties associated with fieldwork, lack of taxonomic expertise, herbarium and specimen-related problems, and identifying and delimiting species in large genera and families. Biological factors are inherent properties that underlie diversification of lineages that cause difficulties in species delimitation, which include rapid radiation, hybridisation, polyploidy, cryptic species, and these are discussed below. These biological and practical issues often occur in combination. For example, the challenge in delimiting species is a product of biological factors that blur species boundaries (such as hybridisation), and practical issues in applying a suitable species concept if there is

limited specimen availability. These issues are frequent in taxonomically challenging groups, as exemplified in Box 2.

Box 2: Taxonomic challenges in the genus *Alpinia* Roxb. (Zingiberaceae)

We illustrate the scale of taxonomic challenges in the tropics with an example from the genus *Alpinia* Roxb., the largest genus in the ginger family (Zingiberaceae). It comprises more than 230 species, and is widespread throughout tropical and subtropical Asia. This genus is defined only on the basis of plesiomorphic characters such as a terminal inflorescence (Kress et al., 2005). Molecular work shows this genus is polyphyletic within the tribe Alpinieae (Rangsiruji et al., 2000; Kress et al., 2005, 2007). Within any given clade, species can be very difficult to tell apart. Practical issues such as lack of type specimens or clear descriptions of species often cause problems in species identification. In many cases, the distinctive features of the plants are not preserved properly on herbarium specimens which causes further confusion in species identification. Biological factors such as the presence of overlapping morphological features adds further problems in delimiting species. For example, in the "Zerumbet" clade (sensu Kress et al., 2005), many species have been misidentified and many names have been applied incorrectly (Ranavat, pers. obs.). The name *Alpinia blepharocalyx* K. Schum. (now accepted as *A. roxburghii* Sweet) has been erroneously applied to what we suspect are different species. One defining character of this species is the presence of papery bracteoles. There are a few other species that also possess this character but because floral and/or vegetative characters of these species are poorly preserved on herbarium specimens (Figure 2.1), the determination of specimens is challenging. Furthermore, processes that obscure species boundaries such as hybridisation and rapid radiation may be present in this group but their occurrence has not been well-studied. This taxonomic disarray has limited the progress that can be made in evolutionary studies in this group.

Figure 2.1: Taxonomic confusion in the tropical ginger *Alpinia blepharocalyx* K. Schum (Zingiberaceae). All the specimens above were determined as *A. blepharocalyx* and possess papery bracteoles but the overall morphology is variable between specimens and suggests these are different species. It is difficult to determine herbarium specimens in this group due to the poor preservation of important morphological characters. A- a specimen of *A. blepharocalyx* from the Royal Botanic Garden Edinburgh (RBGE) (E00149213), B – a specimen of *A. blepharocalyx* from the Natural History Museum, London (BM013718324), C- isotype of *A. blepharocalyx* from the Royal Botanic Garden Edinburgh (RBGE) (E00149177).

2.2.1 Practical issues

Practical issues relate mostly to various aspects of fieldwork and specimens in the herbarium. These include the inaccessibility of certain geographic regions that leads to incomplete taxon sampling and the general undercollection of specimens. Undercollection of specimens in tropical regions is prominent in comparison with temperate regions, such as the United Kingdom, where the collection density of herbarium specimens per 100 km^2 is more than 500-fold higher than that of the Lao People's Democratic Republic (Newman et al., 2007). Species-rich areas in the tropics such as the Rio Marañón valley in the Andes harbour many narrow endemics that are undercollected, which leads to the underestimation of diversity in such regions (Särkinen et al., 2011). Most collections are made close to roads, cities and rivers, in regions of overall high income, reflecting accessibility to humans rather

than true biological patterns (Meyer et al., 2016; Zizka et al., 2020). This accessibility bias causes morphological, geographic and genetic gaps in the sampling of tropical plants. One of the respondents of the survey expressed: *'The lack of collections in areas that are difficult to access sometimes prevents us from assessing the extent of variation/plasticity of morphological characters. Several characters historically used for species delimitation in my genus turned out to be too variable to be reliable, once I had access to a greater number of specimens. This has caused many synonymies in my group.'* Thus, accessibility is a serious issue in tropical regions, specifically those which have a history of undercollection.

In general, fieldwork in the tropics is difficult due to many factors. According to the survey, most responses said that important issues included the difficulty in obtaining permits and poorly documented location details (Figure 2). In a recent localised survey conducted to understand why researchers did not broaden their collections across political boundaries, the primary cause of discouragement was an absence of information on how foreign researchers can initiate a collaboration with a local host (V. Gowda, unpublished data). Forestry officials across the globe have similar interests but differing research permit processes. Further, issues with research permits can range from – high research fees, unpredictable processing time, as well as research restrictions which may make the entire exercise futile.

Another practical hindrance in collection-based science is lack of *a priori* occurrence records for most tropical plants. The number of records of tropical plant species is limited on data portals such as the Global Biodiversity Information Facility (GBIF; http://www.gbif.org/) in comparison with temperate plants that have been extensively collected over the years. Most of the occurrence records in online data portals are also old and may not have relevance to current occurrence as the plants may no longer be present in those areas due to habitat destruction. Thus, current occurrence data is not available for most taxa in the tropics or is restricted to specific geographic regions (for example -forests in Borneo and Sumatra), or is inaccurate and of limited use. The information generated from occurrence records is often critical to accurately map species ranges, which in turn are invaluable for

conservation assessments (Feeley and Silman, 2011). Additionally, lack of funds to carry out surveys that may update information on plant occurrences and which aim to carry out regional botanical collections was acknowledged as a major impediment in tropical botany in our survey (Figure 2.2).

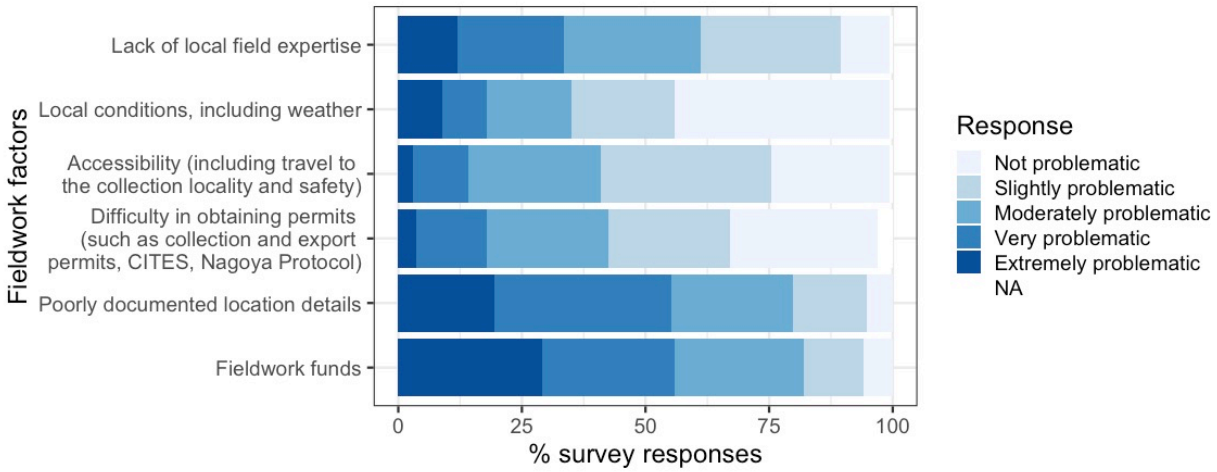

Figure 2.2: Survey results for the question 'How great a problem are the following factors in fieldwork?' associated with the tropics.

The lack of taxonomic expertise is a significant barrier when it comes to documenting the diversity in the tropics (Kim and Byrne, 2006). A study by Ahrends et al. (2011) found that better trained botanists tend to record more species, especially species that are critical from a conservation perspective. Identifying a new species requires a high level of taxonomic expertise as a less well-trained botanist might hesitate to record it as a new species or mistake a new species for one that has already been collected. Therefore, museums and herbaria should invest more towards the training of field botanists to improve recording efficiency and reliability of data (Ahrends et al., 2011).

Species identification and delimitation is often hindered by the lack of diagnostic morphological characters even when extensive collections are available. The few characters that may be informative are often poorly preserved on herbarium specimens. Preparing herbarium specimens in the wet tropics is difficult as specimens can become mouldy or brittle even after drying. Techniques such as air-

drying, treating the specimens with alcohol prior to drying, and artificial heat (Smith, 1971) are still being used but these techniques may cause damage to the specimens as well as the DNA (Särkinen et al., 2012). These difficulties make species identification from herbarium specimens particularly challenging. In our survey, 76 % of researchers said that their work involves the identification of unlabelled plant specimens. The likelihood that a researcher was able to identify a species depends on the type of specimens available. As shown in Figure 2.3, the probability of identifying a living or herbarium specimen with fruits and flowers is very high whereas specimens of vegetative material alone are much less likely to be identified. A small proportion of researchers felt that the identification of vegetative specimens was generally not feasible for their taxa of interest. This highlights the need for complete specimens showing reproductive structures in order to accurately identify species.

Large plant genera or families often present a greater array of challenges for taxonomists than small groups. Species identification and delimitation in these large tropical groups is by necessity more likely to focus on a geographic subset of species or a single clade of the larger group. Phylogenetic studies in these groups often include only a small number of representative taxa, with these helping to understand relationships between and within genera but not providing species-level resolution (Frodin, 2004). Tackling these large groups requires collaborative work involving researchers working in different geographical regions and focusing on different clades of such groups. Overall, diverse practical issues considerably limit tropical taxonomic research and our understanding of biological processes underlying diversification of tropical plant taxa.

2.2.2 Biological factors

Taxonomic challenges in tropical taxa are also a consequence of diverse biological factors but many of these are not well-studied (Duminil et al., 2012) and their relative importance has yet to be determined. As one of the participants of the survey expressed it: "*Many of the biological challenges referred to in this study probably*

apply to the group I focussed on, but given that they're long-lived and highly endemic tropical trees, the truth is that the issues probably haven't even been recognised yet." Despite the uncertainty, our survey suggests that researchers consider factors such as rapid and recent radiations, hybridisation, cryptic speciation, and phenotypic plasticity to cause problems in delimiting tropical plant species (Figure 2.4). Here, we discuss these factors and highlight their impact on species delimitation by focusing on illustrative examples.

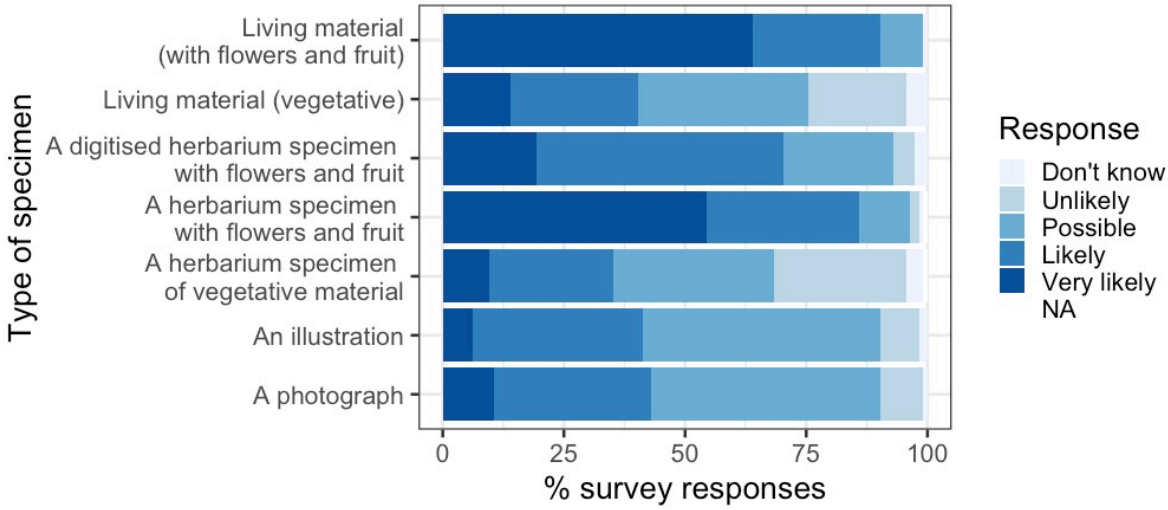

Figure 2.3: Survey results for 'How likely would you be able to identify the following material?' answered by researchers whose work involves identification of unlabelled tropical plant specimens.

2.2.2.1 Rapid radiation

Species delimitation is challenging when studying taxa that have undergone recent and rapid radiations (Shaffer and Thomson, 2007). This is a predominant issue in the Neotropics, where a significant proportion of the species are a result of these radiations (Hughes and Eastwood, 2006). These taxa are difficult to discriminate as they may have had insufficient time to evolve discrete morphological differences. Furthermore, molecular phylogenetic analysis may not help resolve species limits as insufficient sequence divergence may result in unresolved phylogenies (Richardson et al., 2001; Knope et al., 2012; Turner et al., 2013; Bell et al., 2015; Uribe-Convers and Tank, 2015). Such taxa may not have had sufficient time to accumulate reproductive barriers, which also provides an opportunity for hybridisation and

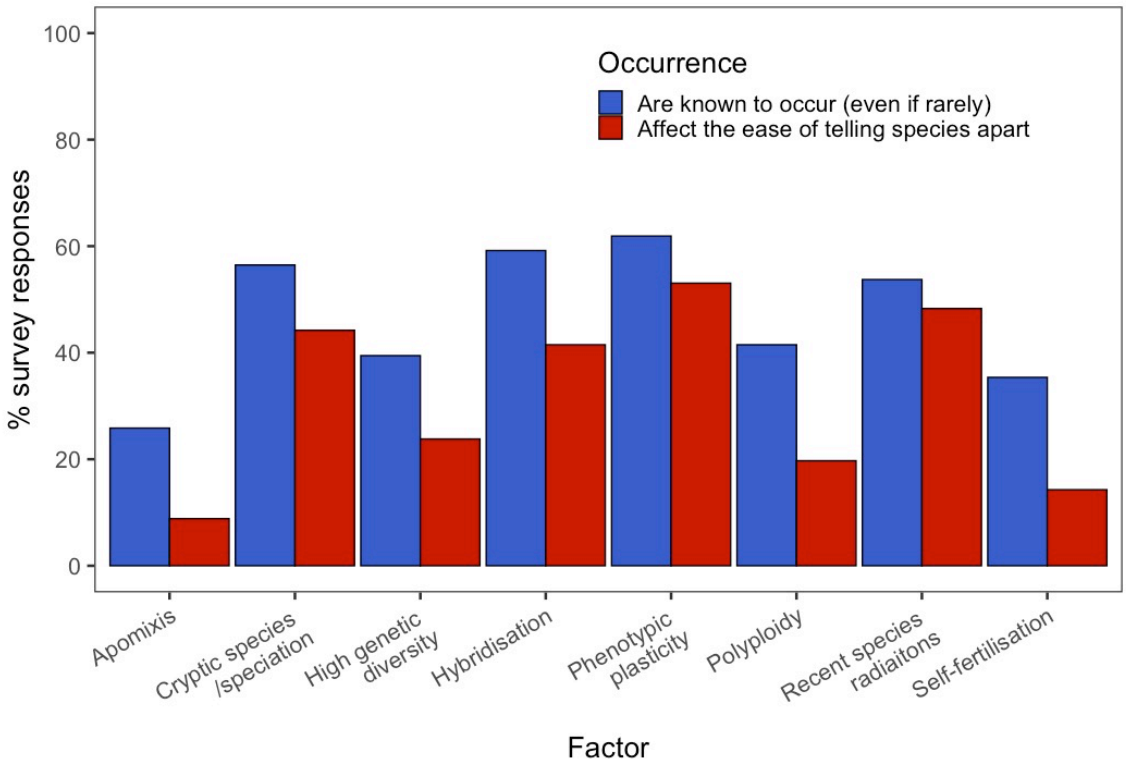

Figure 2.4: Survey results for 'From the following list of biological factors and processes, please tick those that are (a) known to occur (even if rarely), (b) Affect the ease of telling species apart.' The researchers were asked to answer this question based on one tropical plant group of interest.

introgression to occur (Seehausen, 2004; Genner and Turner, 2012; Wan et al., 2014; Valderrama et al., 2018).

A well-known example of an explosive radiation is the genus *Inga* (Fabaceae), which diversified as recently as 2-10 million years ago. It is a species rich tree genus (~300 spp.) found in Neotropical rainforests. A study based on plastid and nuclear markers found that the *Inga* phylogeny was unresolved as the genetic divergence in the sequences was low (Richardson et al., 2001; Nicholls et al., 2015). As the species within this genus are difficult to discriminate by their morphology (especially in the absence of flowers or fruits) or molecular markers, chemical fingerprinting or 'chemocoding' has been used to help delimit species (Endara et al., 2018). Targeted enrichment with hundreds of nuclear loci has also helped in resolving the relationships in this genus which was poorly resolved by Sanger sequenced loci (Nicholls et al., 2015).

Rapid plant radiations are often found in insular environments such as mountains or islands. For example, the genus *Pritchardia* (Arecaceae) is one of the most species-rich genera in Hawaii, comprising 27 species that are mainly single island endemics. Species delimitation in this genus has been problematic due to highly variable morphological characters that are affected by environmental conditions. A combination of plastid, nuclear and microsatellite markers along with morphological data was used to test if known species form distinct lineages. Most *Pritchardia* species showed little sequence differentiation which might be due to incomplete lineage sorting or ongoing gene flow between sympatric species (Bacon et al., 2012). Therefore, species limits in this genus are still ambiguous and illustrate the general issue of species delimitation in such a diverse tropical genus.

2.2.2.2 Hybridisation

Natural hybridisation in plants is a common phenomenon (Mallet, 2007; Goulet et al., 2017). It has a significant impact on the way species are delimited as it causes the sharing of genetic variation and obscures morphological differences between species (Soltis and Soltis, 2009). Hence, resolving species boundaries in hybridising taxa is challenging. Hybridisation is known to play a role in the evolutionary history of a few, well-studied tropical taxa (e.g., Bromeliaceae- Palma-Silva et al., 2011; Marques et al., 2014; Zanella et al., 2016; Goetze et al., 2017; Neri et al., 2018) but not much is known about its occurrence in most other tropical groups. As such, it could be more common than expected in these species-rich regions where many closely related, recently diverged species are sympatric (Schley et al., 2020).

Extensive hybridisation in recently radiated taxa can cause difficulties in telling species apart. Such is the case in the genus *Epidendrum*, which is the largest Neotropical genus in the Orchidaceae, comprising almost 1500 species. Circumscription of species in this genus is difficult due to the presence of extensive morphological diversity between species and even within species in a few cases (Pinheiro and Cozzolino, 2013). Many sympatric species in this genus have low pollinator specificity and an overlapping flowering period, along with weak reproductive barriers that have facilitated hybridisation (Pessoa et al., 2012; Vega et

al., 2013; Marques et al., 2014; Pinheiro et al., 2016). Moreover, some of these species hybridise even if there is a difference in ploidy level (Pinheiro et al., 2010). A study by Vega et al. (2013) found that the proportion of hybrids found in three sympatric *Epidendrum* species (*E. calanthum*, *E. cochlidium* and *E. schistochilum*) based on AFLP data was found to be about 75% or greater and included F_1, F_2 and backcrosses. Therefore, hybridisation is widespread in these sympatric *Epidendrum* species.

Hybridisation is often an underlying cause of difficulties in delimiting species within species complexes. This is the case in many tropical plant taxa such as the tree genus *Carapa* (Meliaceae). This genus has had a controversial taxonomic history as different numbers of species have been defined by different taxonomists over the years due to the lack of clear diagnostic morphological characters. A recent study based on morphological and nuclear markers defined 11 new species making it a total of 27 species in the genus (Kenfack, 2011). Species boundaries were tested in sympatric *Carapa* species complexes in Central Africa using a combination of morphological, nuclear and chloroplast markers and nuclear microsatellites. The species were difficult to identify on the basis of morphology as many of the diagnostic morphological characters were absent from the herbarium specimens that were analysed. The results indicate the presence of ancient hybridisation between *C. dinklagei* and *C. parviflora* as they share two chloroplast haplotypes in sympatric sites. Only *C. dinklagei* was found to be monophyletic based on ITS sequences. Four other morphospecies could not be distinguished on the basis of molecular markers and might represent a species complex with frequent gene flow (Duminil et al., 2012).

Disturbance and degradation of habitats may bring closely related but previously geographically isolated species into secondary contact. This may facilitate gene flow between these species, especially when reproductive barriers are weak or absent, which can lead to an increase in the occurrence of hybridisation. Understanding the impact of hybridisation as constructive (e.g., facilitating range expansion) or destructive (e.g., causing species extinction) is essential for addressing conservation

concerns (Chunco, 2014; Todesco et al., 2016), especially in tropical regions that are undergoing rapid habitat degradation. With the advent of high-throughput sequencing technologies, detecting the occurrence of hybridisation and introgression is now increasingly possible in non-model plants (Twyford and Ennos, 2012).

2.2.2.3 Polyploidy

Polyploidy is a key driver of plant speciation (Wood et al., 2009) with most plant lineages now known to have a polyploid ancestry and many with multiple rounds of genome duplication (Soltis et al., 2014; Alix et al., 2017). Polyploidy causes taxonomic problems in many plant genera as it may lead to reticulate evolution due to hybridisation and introgression between species or may cause difficulty in defining species that have multiple ploidy levels (Soltis et al., 2007; Padilla-García et al., 2018). Although the frequency of polyploid plant species is known to decrease towards the equator (Rice et al., 2019), it is likely that many instances of polyploidy in the tropics remain to be identified, in particular in cooler, montane areas. Polyploidy may be predicted to confer an adaptive advantage on taxa that colonise harsh environments as found in other ecosystems (Soltis and Soltis, 2000; Brochmann et al., 2004).

Resolving relationships in genera comprising species with multiple ploidy levels is a difficult task. Phylogenetic inference is facilitated by phasing alleles in the nuclear genome that have been derived from different parental sources, but although some bioinformatic solutions have been developed (Andermann et al., 2018) this is not yet possible for taxa with high ploidy levels. The economically important genus *Curcuma* (Zingiberaceae) is made up of 120 species with a high variation in ploidy levels (2x-15x), and many species are of allopolyploid origin (Záveská et al., 2016). It has a complicated taxonomic history with ambiguous original descriptions of species, incorrect application of names and lack of type material. This, along with high intra- and interpopulation variation, ambiguity in diagnostic morphological characters and interspecific hybridisation coupled with polyploidy hinders how we discern species boundaries within this genus (Leong-Škorničková et al., 2007). Incongruency between nuclear and plastid markers within and between some

subgenera in this genus indicates incomplete lineage sorting and/or hybridisation which further complicates the evolutionary history of this genus (Záveská et al., 2012, 2016). The presence of multiple ITS paralogues in polyploid species impedes accurate phylogenetic reconstruction (Záveská et al., 2012). More detailed studies using genomic data and polyploid-aware analyses are required to disentangle the complex relationships within this genus.

Delimiting species in genera with polyploidy can be extremely difficult, especially if these taxa have formed as a result of hybridisation. One such example is the tropical tree genus *Afzelia* (Fabaceae) which is made up of 11 species, seven of which are found in Africa. The African species are either diploid or tetraploid. The African tetraploid species of this genus are difficult to delimit based on morphological and genetic markers, unlike the diploid species. These tetraploid species seem to have diverged more recently and have adapted to a different ecological niche than the diploid species. These species are sympatric and are found in rainforests whereas the diploids are allopatric and found in dry forests. While the tetraploids do form a monophyletic group, the resolution among them is poor and nuclear or plastid DNA haplotypes are shared by several species (Donkpegan et al., 2017). Given that several of these economically important species have been classified as vulnerable in the IUCN red list (Donkpegan et al., 2014), understanding the complex speciation histories of these species is crucial for planning appropriate conservation measures.

Polyploidy is commonly regarded as a mechanism of instant speciation as most polyploid species are immediately reproductively isolated from their parent species (Soltis et al., 2014; Alix et al., 2017). However, several studies have shown that polyploids can hybridise with the parent species in the face of weak reproductive barriers (Moraes et al., 2013; Kovalsky et al., 2018). The occurrence of polyploidy and its impact on the relationships within complex tropical taxa needs further investigation. A knowledge of the genetic features and reproductive biology of the polyploid species is essential to take the proper steps towards their conservation (Chen et al., 2019).

2.2.2.4 Cryptic species

Two or more distinct species that have been recognised as a single species due to morphological similarity are termed cryptic species. This is a significant impediment in estimating the number of species, understanding speciation patterns and in conservation assessments. The detection of cryptic species is often only possible with additional sources of data such as genetic markers (Hebert et al., 2004; Särkinen et al., 2011; Li, Tong, et al., 2016; Surveswaran et al., 2018; Ito et al., 2019). Recent methodological developments, such as Bayesian methods for species discovery and delimitation based on the multispecies coalescent model, are likely to contribute to this pursuit (Jones et al., 2015; Solís-Lemus et al., 2015). In some cases, this additional information can lead to the discovery not only of new species but of new genera as well (Gagnon et al., 2015).

As an example, the presence of species complexes in the genus *Nervilia* (Orchidaceae) has led to ambiguity in species boundaries. Species from the *N. adolphi/punctata* alliance cannot be distinguished on the basis of morphology alone. To resolve the species relationships and boundaries within this alliance, Gale et al. (2018) sampled 12 out of 27 species from this alliance found in Asia along with 20 unidentified samples from Thailand. Eleven species were corroborated using nuclear and plastid markers analysed with a Bayesian coalescence approach to delimit species. The 20 unidentified samples formed 3 distinct clades with one made up of two sub-clades that led to the circumscription of three new species.

Cryptic species are present in smaller tropical genera as well. Only two species in the African tree genus *Greenwayodendron* (Annonaceae) were previously described when an analysis of microsatellite markers and morphology led to the discovery of two new species and two subspecies. These subspecies are now recognised as distinct species, increasing the number of species in this genus from two to six. The taxonomic status of a few taxa remains unresolved partially due to the lack of fertile material (Lissambou et al., 2019). Therefore, integrative analyses can aid in discovering cryptic species and delimiting species that cannot be distinguished based on subtle morphological differences (Prata et al., 2018; Gomes et al., 2020).

As the tropics are the most species-rich regions in the world, they likely harbour many cryptic species that are yet to be discovered (Bickford et al., 2007; Funk et al., 2012). Cryptic diversity is widespread in other tropical taxa such as butterflies (Toussaint et al., 2015; Janzen et al., 2017), birds (Lohman et al., 2010; Younger et al., 2018), reptiles (Domingos et al., 2017; Portillo et al., 2018), and mammals (Hotaling et al., 2016; Rivera et al., 2018) and this pattern is likely to be similar in plants. Discovery of cryptic species is likely to inflate overall estimates of current species numbers and will better reflect the actual diversity present in the world. Recognition of cryptic species is crucial, especially from a conservation perspective as different species in a complex may require different strategies for conservation (Bickford et al., 2007).

2.2.2.5 Other biological factors and processes

Certain additional factors (including those depicted in Figure 2.4) might cause taxonomic issues in tropical taxa, but their occurrence is not well-documented, or they have not been studied in the context of species delimitation. One such factor is phenotypic plasticity, where organisms with the same genotype express a range of phenotypes under different environmental conditions (Sultan, 2000), and sometimes without any obvious ecological or environmental correlates. This may cause taxonomic uncertainty if the plastic morphological characters are used for species delimitation. For example, the four species of *Dracaena* (Asparagaceae) found in Madagascar, are known for their high morphological variability. This group has posed difficulties for taxonomic revisions and species identification in the field due to the presence of continuous morphological characters. Buerki et al. (2009) carried out a study to understand the relationship between morphology, biogeography and molecular markers within these species. They found that the most developed inflorescences (panicles) are found in humid environments where the availability of resources is high and the least developed ones (racemes) are found in arid regions. This could be due to the presence of phenotypic plasticity in these species or selective pressure acting on these traits. This could be tested by performing a common garden experiment or a reciprocal transplant to investigate the level of

plasticity, however this would require considerable effort given that species in *Dracaena* are long lived. This is the case for many tropical plant taxa (Chen and Schemske, 2015) hence studying the presence of phenotypic plasticity in tropical taxa is challenging. According to our survey, over 50% of the respondents feel that phenotypic plasticity affects the ease of telling species apart in their taxa of interest (Figure 2.4).

Another factor that is known to hinder species delimitation is apomixis, which is a form of asexual reproduction that results in seed formation from an unreduced and unfertilised egg cell or from the somatic cell of the ovule. Apomixis is widespread in angiosperms but more common in the larger families and genera such as Asteraceae, Poaceae, Rosaceae and *Ranunculus* (Ranunculaceae) (Hojsgaard et al., 2014). Defining species in apomictic taxa is problematic as these taxa have higher phenotypic diversity and represent reproductively isolated individual genotypes. Apomictic individuals can serve as pollen donors for their sexual progenitors if present in sympatry. This can lead to the formation of new apomictic and polyploid taxa that results in reticulate evolution. This can complicate the evolutionary history of taxa and obscure species delimitation (Ennos et al., 2005; Majeský et al., 2017; Hörandl, 2018). The occurrence of apomictic taxa is presumed to be higher in northern latitudes and such taxa have larger distributions with respect to their sexual relatives (Hörandl et al., 2008). But apomixis has been reported from many tropical families such as Dipterocarpaceae (Kaur et al., 1978), Bignoniaceae (Sampaio et al., 2013), and Melastomataceae (dos Santos et al., 2012). As there is a dearth of studies of apomixis in the tropics in the framework of species delimitation, not much is known about its role in causing taxonomic challenges.

Other forms of uniparental reproduction such as self-fertilisation can also have a significant impact on species delimitation. Selfing taxa often show higher morphological variation between populations than outcrossing taxa (Wright et al., 2013) which causes problems in delimiting species. Hence an understanding of population structure and mating systems is valuable knowledge for defining species limits and planning conservation measures.

2.3 Implications of taxonomic challenges

The problems caused by taxonomic challenges can have far reaching effects on biodiversity research and conservation. Issues such as carrying out conservation assessments, *ex situ* conservation and species reintroductions, estimating species diversity, and modelling species distributions can be compromised due to unclear taxonomy. All of these factors were found to be important in our survey (Figure 2.5). Identifying and prioritising taxa and regions for conservation requires detailed surveys, which might be a daunting task given the number of species-rich regions in the tropics. The information produced by the species number estimates and the models of species distributions can help us in informing conservation assessments and planning, and therefore is crucial information.

Habitat fragmentation and loss is on the rise in most tropical regions and *ex situ* conservation might be the only solution to conserving species from such regions. Botanical gardens are the major centres for *ex situ* conservation but they are mostly concentrated in the temperate regions of the northern hemisphere with some southern exceptions (e.g., many gardens in Australia and New Zealand). Carrying out conservation of tropical species in temperate regions is extremely difficult due to limited space and high energy costs for heating glasshouses (Mounce et al., 2017). In recent years, however, some tropical countries have been developing more botanic gardens. For example, the 6 gardens of the Botanic Gardens Organisation and the c. 18 gardens of the Department of National Parks, Wildlife and Plant Conservation in Thailand, and a network of botanic gardens in Indonesia, with a plan to establish one in each province in the country. While seed banking offers a space-efficient alternative to the conservation of adult plants, it is currently not feasible for many threatened taxa (Wyse et al., 2018).

Taxonomic issues in tropical plants also impact many studies in ecology and evolution. Uncertainty in species identification creates problems in almost every research area, from ecological surveys to genomic analysis. For example, there is a

global drive to sequence reference genomes for all eukaryotic genomes in the next 10 years (Lewin et al., 2018). However, such efforts are inhibited by our limited taxonomic knowledge of life on earth and in particular in the tropics. Only ~2.3 million of the projected 10-15 million Eukaryotic species have been described, and such global sequencing efforts rely on firm taxonomic knowledge. There is considerable uncertainty even in iconic and well-studied ecosystems, as illustrated by a complete species list for seed plant species from lowland Amazon rain forests only being published in 2017 (Cardoso et al., 2017). Detailed taxonomic treatments of tropical plants are essential for such genomic sequencing efforts to ensure complete taxonomic coverage and to ensure genome sequences are linked to a valid name.

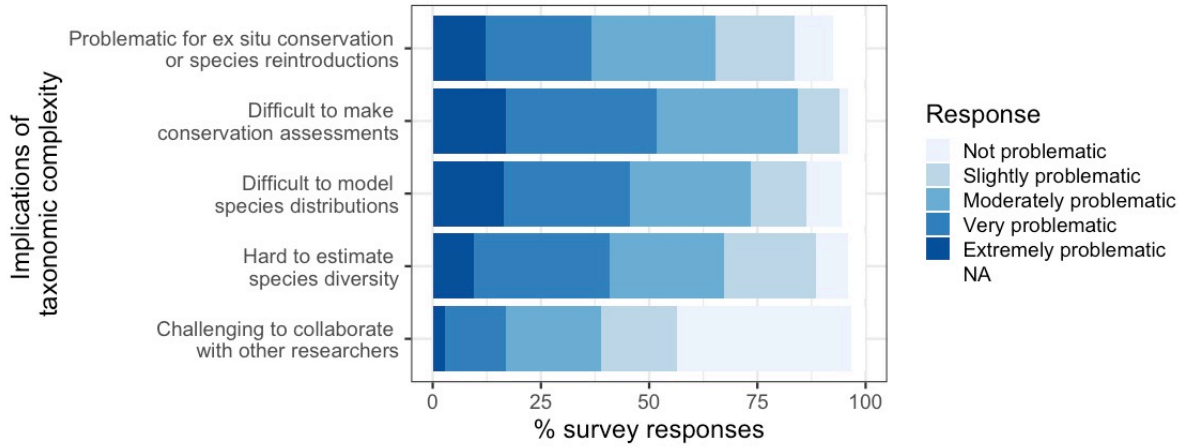

Figure 2.5: Survey results for the question 'To what extent does taxonomic complexity have practical implications in this group? Please tick those which apply'

2.4 Conclusions and prospects

Our survey and literature review reveal the extent and causes of taxonomic challenges in species-rich tropical regions, with many of these challenges poorly documented and understood. This is partly due to a lack of studies that have investigated the processes that hamper species delimitation in tropical taxa. Taxonomic bias is an important issue, as the studies that have investigated the factors causing difficulties in species delimitation in the tropics have focused on taxa that have been well-studied over a long period of time (such as Orchidaceae) or are economically important (such as Fabaceae and Poaceae; (Pinheiro et al., 2018)). This is a major impediment in understanding the patterns of taxonomic challenges

between temperate and tropical zones. There are numerous plant groups with no recent monographic work where a recent taxonomic treatment could form the foundation to improve conservation work.

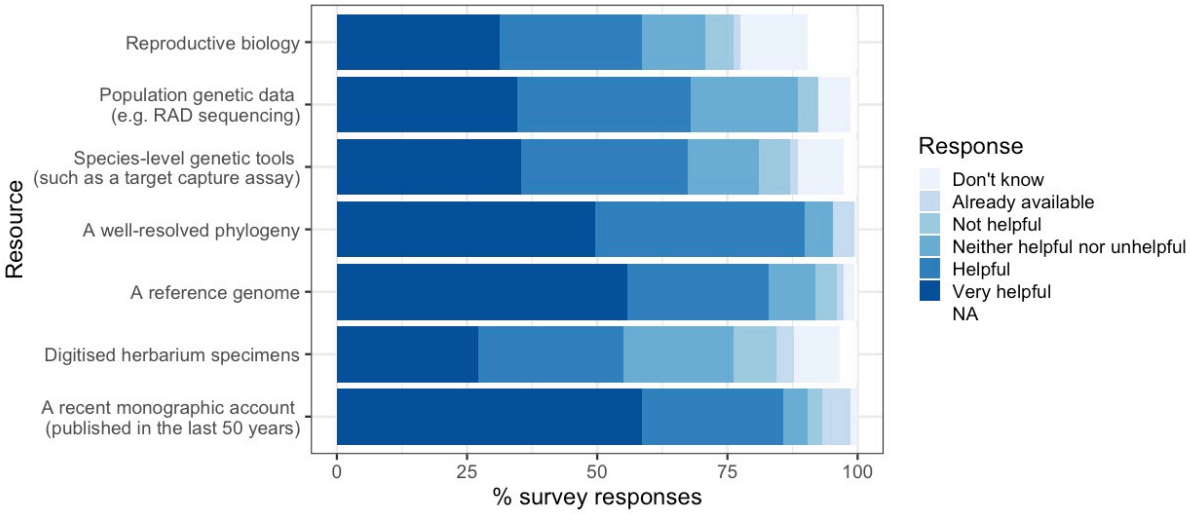

Figure 2.6: Survey results for 'How important would each of these developments be in helping improve species delimitation?'

Increase in taxonomic effort , i.e. the collection and description of species, is vital for groups that are not well-documented as many species still await discovery or description, including many narrowly distributed taxa. The survey results indicate that resources such as a recent monographic account or a reference genome would be most helpful (Figure 2.6). Both of these require a large effort, especially for large and widespread genera and families. Initiatives such as The Legume Phylogeny Working Group (The Legume Phylogeny Working Group et al., 2013) have helped in resolving relationships within large families, hence collaboration of researchers working on these large taxa is important. Producing monographs that integrate morphological and genetic data is key for tropical taxa and one such example is the recent monograph of the genus *Ipomoea* L. (Convolvulaceae) that has led to description of 63 new species in this large genus (Muñoz-Rodríguez et al., 2019). As the number of herbaria in the world has increased over the years, it may not be possible for researchers to study all the collections from around the world (Goodwin et al., 2015). Hence digitisation can enable taxonomists to access specimens from

across the world with ease. Resources such as JSTOR Global Plants (https://plants.jstor.org/), iDigBio (https://www.idigbio.org/), and DiSSCo (https://www.dissco.eu/) are actively recording and digitising herbarium specimens or collaborating with different museums, botanic gardens, and institutes from around the world.

More studies using an integrative systematic approach should be implemented in order to unravel the complex evolutionary processes underlying the diversification of tropical taxa. Clarification of the taxonomic uncertainty that persists in many tropical genera is essential baseline information for conservation work, and can be achieved through a combination of population genetic and phylogenetic approaches, together with morphological and geographic examinations. The availability of tools and resources in this era of high-throughput sequencing has made it easier to disentangle species relationships and delimit challenging taxa using vast amounts of genomic data and broad specimen sampling. Genetic tools such as targeted enrichment (Nicholls et al., 2015; Heyduk et al., 2016), RADseq (Tripp et al., 2017; Clugston et al., 2019) and GBS (Pérez-Escobar et al., 2020) can aid in producing well-resolved phylogenies. As these genomic resources are becoming more affordable, it will be easier to obtain large datasets and carry out robust analyses to delimit species in these problematic groups. These studies may also unearth new species. In some cases, these large genetic datasets may not help in delimiting species but can help in identifying the evolutionary processes that have resulted in diversification.

Although we have specifically focussed on examples and issues associated with tropical plant taxa, many of these issues apply to taxonomic research in general. Biological factors discussed above such as polyploidy, hybridisation, apomixis, and phenotypic plasticity are found in plant taxa that occur the tropics and other regions alike. Even in cases where the flora is well-documented and the extent of hybridisation is known) (Stace et al., 2015), telling species apart is challenging in complex taxa, especially in larger genera (Hollingsworth, 2003; Ennos et al., 2005; Mráz et al., 2011; Padilla-García et al., 2018; Wang et al., 2018; Walter et al., 2020). Utilisation of NGS techniques mentioned above can similarly help in understanding

species relationships to an extent across taxonomically complex plant groups (Twyford and Ennos, 2012).

Given that fragmentation and loss of natural habitats is rapid and widespread, documentation of all life on earth is the need of the hour. Development of tools such as Machine Learning (ML) based species identification through image recognition (Wäldchen and Mäder, 2018) and ML-based cluster analyses to find species boundaries and cluster characteristics (Saryan et al., 2020) can help speed up species identification and delimitation. Digital repositories should be encouraged and a standard format should be recommended for digital images such that uniformity can be achieved across institutions. The current pandemic has allowed us to appreciate plants blooming from our enclosed spaces. This further emphasizes the need for digital repositories for plants with a potential to store detailed morphologies which is similar to the "morphobank" model which has been successful among palaeontologists and zoologists, in general (https://morphobank.org/). Digital access to global, well-curated data has a potential to improve our understanding of taxa and their variations throughout their geographic range. We conclude that efforts to improve processes that can aid taxonomy or in understanding lineage diversifications will ultimately allow us in planning conservation measures and improving our understanding of the biodiversity around us.

Chapter 3

Resolving taxonomic uncertainty in a group of Indian *Alpinia* species

3.1 Introduction

Alpinia Roxb. is the largest genus in the Zingiberaceae comprising more than 250 species (Kress et al., 2005). It is widespread in tropical and subtropical Asia and extends as far as the Western Pacific islands and northern Australia. The centres of diversity of this genus are in Malesia and Indochina (Larsen, 2005). It belongs to the tribe Alpinieae where the genera possess characters such as the pseudostems rising up to 8m, alternate leaves with the plane of distichy of leaves perpendicular to the rhizome, and reduced or absent lateral staminodes. It is an economically important genus as many species are used as ornamentals, in medicine and in cooking (Rangsiruji et al., 2000; Leong-Škorničková and Newman, 2015).

The name *Alpinia* was first used by Linnaeus in 1753 for the tropical American species *Alpinia racemosa* named in honour of Italian botanist Prospero Alpino. Later, *Alpinia* was mostly used for Asiatic species by Roxburgh (Roxburgh, 1810) and others, while *Renealmia* L.f. was used for American and West African species. In 1904, Schumann (Schumann, 1904) produced a comprehensive revision of *Alpinia*. Although he referred to *Alpinia* as that of Linnaeus rather than Roxburgh, he applied this name only to Asiatic species. In 1922, Merrill revived the name *Alpinia* L. for American species against *Renealmia* and accepted the name *Languas* J.Koenig (1783) for the Asiatic species (Merrill, 1922). To avoid the confusion caused by this transfer, the name *Renealmia* L.f. was conserved for American species in 1930 so as to apply the name *Alpinia* Roxb. to the Asiatic species. As *Alpinia* Roxb. was a later homonym of *Alpinia* L., the name *Alpinia* would have to be replaced by *Languas* J.Koenig. To avoid the confusion caused by the transfer of names, Rehder proposed to conserve the name *Alpinia* Roxb. against *Alpinia* L. and *Languas* J.Koenig (Rehder et al., 1935) for this large and economically important genus.

The circumscription of this genus on the basis of morphology has always been problematic as the species are distinguished only on the basis of pleisiomorphic characters in the tribe Alpinieae, such as a terminal inflorescence (Kress et al., 2005), although some *Alpinia* species possess radical inflorescences (e.g. *Alpinia melichroa* K.Schum.). The first comprehensive classification of *Alpinia* was done by Schumann

in 1904, where it was classified into five subgenera and 27 sections. Later, Holttum (1950) divided *Alpinia* into four genera (*Cenolophon*, *Alpinia*, *Catimbium*, and *Languas*), but this was only for the species occurring in the Malay Peninsula. As some of the generic names were illegitimate, *Alpinia* sensu Schumann was revived and an infrageneric classification was produced by Smith (1990), where the genus was divided into two subgenera, nine sections and 12 subsections. Therefore, different botanists classified this genus in dissimilar ways (Rangsiruji et al., 2000).

Clarification of relationships within this genus was essential due to these disagreements. (Rangsiruji et al., 2000) carried out a molecular phylogenetic study to test Smith's infrageneric classification. The results of the study were discordant with the previous classification and indicated that *Alpinia* may be paraphyletic with respect to *Renealmia*. This was confirmed in a study by Kress et al. (2005), where they found *Alpinia* to be polyphyletic across six clades within the tribe Alpinieae. Molecular phylogenetic studies have found other genera in the Alpinieae to be polyphyletic, such as *Amomum* Roxb., which has recently been split into several genera (de Boer et al., 2018). Similarly, *Alpinia* needs to be divided into several genera. The type species, *Alpinia galanga* (L.) Willd., belongs to Clade II sensu Kress et al. (2005) which also contains the species *A. nigra* and *A. conchigera*. Hence, the genus *Alpinia* will be reduced to just three species.

Taxonomic problems are widespread in *Alpinia*. Certain species in this genus have validly published names but, because they were described before the development of the type concept, their type specimens may be missing or difficult to interpret, which leads to inaccurate circumscription of species. The protologues are often vague and lacking in detail which causes further confusion. As a result, many names have been applied incorrectly which has led to taxonomic confusion within this genus. The other problem with *Alpinia* and other taxa from Zingiberaceae is the difficulty in preservation of morphological characters on herbarium specimens (Leong-Škorničková and Newman, 2015). *Alpinia* leaves are often quite large and the inflorescence is bulky and difficult to press. If the specimen is not pressed immediately in the field, the leaves tend to roll and the fruits and delicate flowers fall

off the inflorescence. This results in herbarium specimens being difficult to determine as many important morphological characters are poorly preserved or lost. The taxonomic uncertainty caused by these issues hinders research as assigning the correct names is essential for systematic, ecological and evolutionary studies.

The aim of this chapter is to resolve the nomenclatural confusion in a taxonomically challenging group of Indian *Alpinia* species from subgenus *Alpinia* subsection *Catimbium* (sensu Smith, 1990) or the Zerumbet clade (sensu Kress et al., 2005) (Figure 1). These species were first described in the early 19th century where a few names were applied incorrectly. In many cases, the type material itself is missing or does not correspond to the species as they are presently known. Resolving this confusion is essential to prevent further incorrect application of names.

This study was limited partly because of the difficulty in accessing material in the locations from where the species were described originally, partly for a lack of material in accessible herbaria, and finally because of restrictions in movement during the COVID-19 pandemic.

3.1.1 *Alpinia* species in India

Alpinia species are found in the wet zones of India, i.e., Western Ghats and North-East India. Sabu and Mangaly revised the South Indian Alpinias where they described eight species (Mangaly and Sabu, 1992) whereas Tripathi revised the North-East Indian Alpinias and reported seven species (Tripathi, 2002). The *Alpinia* species found in India belong to Clades I, II and IV, and V (sensu Kress et al., 2005) and subgenus *Alpinia* based on Smith, (1990) (Table 3.1). Some of these species are widely cultivated so it is difficult to ascertain their native range, such as *A. galanga* and *A. calcarata*. Species such as *A. zerumbet* and *A. purpurata* are non-native but are widely cultivated as ornamental plants.

The status of some of the species reported from India is controversial. For example, the name *A. malaccensis* (Burm.f.) Roscoe has been applied to Indian species for what could possibly be a different species as it was first described from Ambon. The

Indian taxa are morphologically similar to *A. malaccensis* but a thorough examination is required to resolve the uncertainty associated with this name.

Table 3.1: A list of *Alpinia* species found in India.

Species name	Infrageneric placement (sensu Smith, 1990)	Clade (sensu Kress et al., 2005)	Distribution in India	Source
Alpinia fax B.L Burtt & R.M.Sm.	Sect. *Fax*	I	Kerala	(Sasidharan, 2004; Nayar et al., 2006)
Alpinia abundiflora B.L.Burtt & R.M.Sm.	Sect. *Fax*	I	Tamil Nadu, Kerala	(Henry et al., 1989; Mangaly and Sabu, 1992)
Alpinia galanga (L.) Willd.	Sect. *Alpinia* subsect. *Alpinia*	II	Western Ghats and North-East India; widely cultivated	(Hooker, 1894)
Alpinia nigra (Gaertn.) B.L.Burtt	Sect. *Alpinia* subsect. *Allughas*	II	North-East India	(Hooker, 1894)
Alpinia conchigera Griff.	Sect. *Alpinia* subsect. *Strobidia*	II	North-East India	(Hooker, 1894)
Alpinia calcarata (Haw.) Roscoe	Sect. *Alpinia* subsect. *Catimbium*	IV	Western Ghats and North-East India	(Mangaly and Sabu, 1992)
Alpinia malaccensis (Burm.f.) Roscoe	Sect. *Alpinia* subsect. *Catimbium*	IV	Western Ghats and North-East India	(Hooker, 1894; Mangaly and Sabu, 1992; Chaudhari, 1993)
Alpinia mutica Roxb.	Sect. *Alpinia* subsect. *Catimbium*	IV	Karnataka, Kerala and North-East India	(Mangaly and Sabu, 1992)
Alpinia ovoidocarpa H.Dong & G.J.Xu	Sect. *Alpinia* subsect. *Catimbium*	IV	Arunachal Pradesh	(Sabu et al., 2008)
Alpinia smithiae M.Sabu & Mangaly	Sect. *Alpinia* subsect. *Catimbium*	IV	Kerala	(Sabu and Mangaly, 1991)

Table 1: continued

Alpinia roxburghii Sweet	Sect. *Alpinia* subsect. *Catimbium*	IV	Arunachal Pradesh, Assam, Manipur, Meghalaya, Nagaland Mizoram and Tripura	(Hooker, 1894)
Alpinia zerumbet (Pers.) B.L.Burtt & R.M.Sm.	Sect. *Alpinia* subsect. *Catimbium*	IV	Cultivated; non-native	(Mangaly and Sabu, 1992)
Alpinia purpurata (Vieill.) K.Schum.	Sect. Guillainia	V	Cultivated; non-native	S. Ranavat, pers. obs.
Alpinia manii Baker	Sect. *Allughas* subsect. ? *Strobidia*	?	Andaman and Nicobar Islands	(Hooker, 1894)

3.1.2 Confusion associated with the name *Alpinia bracteata*

The name *Alpinia bracteata* was first used by Roscoe in 1815 (Roscoe, 1815). The description of this species was based on Chinese drawings belonging to Lord Stanley. The illustration of this species in Roscoe's Monandrian plants of the order Scitaminae (Roscoe, 1828) was based on the Chinese drawings. Roscoe's description in Monandrian plants stated that it was unlike any other known species of *Alpinia* and that its habit was more like a *Costus* than that of an *Alpinia*. This illustration is the only original material for *Alpinia bracteata* Roscoe but it is badly drawn. The name *Alpinia bracteata* Roscoe is currently placed in synonymy under *Alpinia calcarata* (Haw.) Roscoe.

The name *Alpinia bracteata* was also applied to a species by Roxburgh in Flora Indica (Roxburgh, 1820). This species was described as one of the smallest species of *Alpinia* that has green bracts and is native to the eastern parts of Bengal. This name was accepted by Schumann (1904) but was wrongly ascribed to Hort. Bengal.

2: 1811, where only the name appeared. As the name *Alpinia bracteata* Roxb. is a later homonym of *Alpinia bracteata* Roscoe, it is an illegitimate name (Figure 3.1).

3.1.3 Application of the name *Alpinia calcarata*

Alpinia calcarata was first published as *Renealmia calcarata* by Haworth in 1805 (Haworth, 1805), which is its basionym. Its specific character was described as 'Renealmia with lance-sword-shaped leaves, and an erect terminal racemus of flowers.' A. B. Lambert communicated the plant to Haworth and stated that its native country is Coromandel. It was then moved to *Alpinia* by Roscoe in 1807 (Roscoe, 1807). Roscoe described it as 'nectario ovato-oblongo apice semibifido, foliis ensiformibus, capsula hirta', which is a rather vague description. The name *Alpinia calcarata* was wrongly applied in Edwards' Botanical Register 1816 (Edwards, 1816). The illustration does not correspond with what was originally described as *Alpinia calcarata*, hence this was an apparent misidentification. Schumann later synonymised *Alpinia bracteata* Roscoe under *A. calcarata* (Haw.) Roscoe (Schumann, 1904).

Alpinia calcarata (Haw.) Roscoe, Trans. Linn. Soc. London 8: 347. 1807; Wu & Larsen, Fl. China 24: 336. 2000; T.S. Nayar et al., Fl. Plants of Kerala 840. 2006; Rama Rao M., Flowering plants of Travancore 404. 1914; Baker, Fl. Brit. India 6: 254. 1892; Mangaly, JK and Sabu M, Rheedea, 2(1): 38-51.1992. – *Renealmia calcarata* Haw., Bot. repos. 6, pl. 421. 1805. – *Languas calcarata* (Haw.) Merr, Lingnan Sci. J. 5: 51. 1927. (Type cult. Liverpool, no specimen)
– *Alpinia spicata* Roxb., Asiat. Res. 11: 356. 1810. (Type: India, Ind. Orient., *Roxburgh* BM [barcode: BM000603478])
– *Alpinia cernua* Sims, Bot. Mag. 44: t. 1900. 1817. (Type not located)
– *Alpinia alata* A.Dietr., Sp. pl., ed. 6, 45. 1831. (Type not located)
– *Alpinia simsii* Gasp., Ann. Civili Regno Due Sicilie 4: 4. 1833. (Type not located)
– *Alpinia erecta* Lodd. ex Steud., Nomencl. Bot., ed. 2, 1: 62. 1840.– *Catimbium erectum* (Redouté) Juss. ex T.Lestib., Ann. Sci. Nat. Bot., II 15: 342. 1841. – *Renealmia erecta* (Redouté) Boos, Schönbrunn's Fl. 2. 1816. – *Globba erecta* Redouté, Liliac. t.174. 1807. (Type not located).

Specimens examined: Illustration in Monandr. Pl. Scitam. 9/10, Pl. 70. 1826, India: *Wallich* 6577 (K) [K001124240- K001124246], *Wight* 3047 (E) [E00149216], *Wight* WC1700 (E), [E00149218], Ind. Orient. *Roxburgh* [BM000603478] (BM), West Bengal: Calcutta Colitur in Horto Botanico prop. Calcuttam & China *Buchanan-Hamilton* 9 [E00149221](E), Tamil Nadu: Palamcotton *Wight* 1855 (E) [E00149215], Northern Division *Elliot* [E00149217] (E) Madras *Anon* LINN-HS6-11 (LINN), Sri Lanka: North Western Provine, Kurunegala District, Dummalasuriya *Jayasuriya* 1191 [E00149223] (E).

3.1.4 Usage of the name *Alpinia roxburghii* Sweet

In 1826, Sweet made a new name for *Alpinia bracteata* Roxb., which was *Alpinia roxburghii* (Sweet, 1826). The description stated '*bracteata* F.I. non L.T' which means *bracteata* of Flora Indica (1:61, 1820) not of Trans. Linn. Soc. London (11: 281, 1815).

In 1904, Schumann described a species called *A. blepharocalyx* based on the specimen collected by Henry (no. 11692) in Yunnan (Schumann, 1904). It was described as *A. blepharocalyx* and is currently placed in synonymy under *A. roxburghii* Sweet as it was described later.

A comparison of the original descriptions of *A. blepharocalyx* K. Schum. and *A. bracteata* Roxb. indicates that *A. bracteata* Roxb. (now *A. roxburghii* Sweet) has shorter pseudostems (0.9m) than *A. blepharocalyx* (2.5m). Variation of the length of pseudostems within an *Alpinia* species is common but this requires an examination of multiple specimens across its geographical range. As mentioned previously, *A. bracteata* Roxb. was described as the smallest of East Indian Alpinias with green bracts (bracteoles), which are its distinguishing characters from other Alpinias of the region. Whether *A. roxburghii* is a different species than *A. blepharocalyx* needs further investigation.

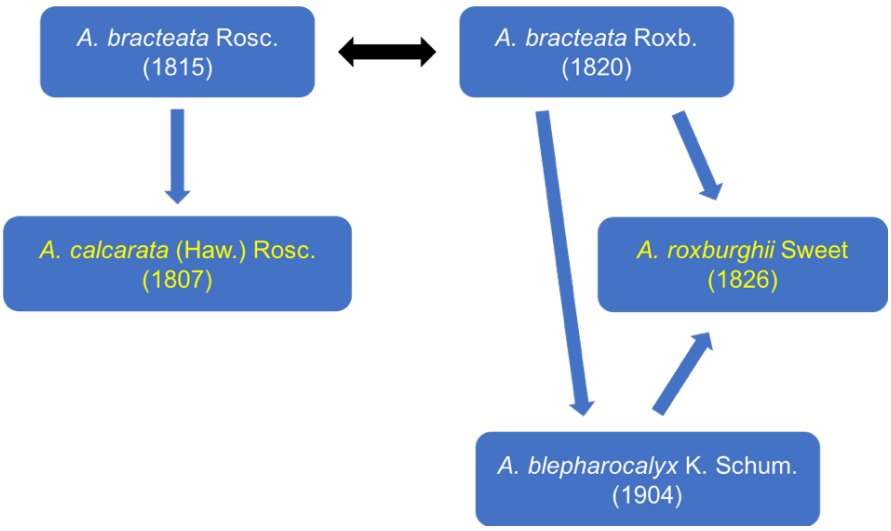

Figure 3.1: Taxonomic confusion associated with the name *A. bracteata.* The black arrow indicates the use of the same epithet for two different species. The accepted names are highlighted in yellow and the blue arrows indicate the synonyms.

Alpinia roxburghii Sweet, Hort. Brit. 390. 1826. – *Alpinia bracteata*, Roxb., Fl. Ind. 1: 61. 1820, *non* Roscoe, 1815; DB Deb, The flora of Tripura State, 2: 368. 1983; Mitra JN, Flowering Plants of Eastern India, 1:255. 1958. (Type- India, Ind. Orient., *Roxburgh* 154 BM [barcode: BM000958163]).

– *Alpinia blepharocalyx* K.Schum. Pflanzenr. IV, 46 (Heft 20): 334. (1904); Wu & Larsen, Fl. China 24: 339. 2000; Ahmed ZU, Encyclopedia of Flora and Fauna of Bangladesh 12: 456. 2008. – *Languas blepharocalyx* (K.Schum.) Hand.-Mazz. Symb. Sin. 1322. 1936.

(Type: China, Yunnan, Szemao woods, 1400 m. alt., *A. Henry* 11962, holo K [barcode: K000815864]; iso E [barcode: E00149177], iso W [barcode:W 1922-0008582]).

Specimens examined: Bangladesh: Silhet *de Silva* Wall. Cat. 6547e [BM013718322] (BM), China: Yunnan province, Szemao woods *A. Henry* 11962 (E, K, W), Guizhou Province, Kwichow *Cavalerie* 2980 (E); Yunnan Province, between Keng Hung and Muang Hing, *Rock* 2612 (E), Yunnan Province, Baoshan Pref., Hills N.W. of Tengyueh, W. Yunnan *Forrest* 29618 (E), Yunnan Province, Baoshan Pref., Hills to the south of Tengyueh, W. Yunnan *Forrest* 8591 (E), Yunnan Province, Baoshan

Pref., Hill to the west of Tengyeuh, W. Yunnan *Forrest* 7612 (E), Yunnan Province: Baoshan Pref., Hills to the south of Tengyueh, W. Yunnan *Forrest* 8245 (E), Yunnan Province, Yunnan-Sen, District Lofou *Cavalerie* 3489 (E), Yunnan Province *Wang & Liu* 83034 (E), *Henry* 11494 (E), Yunnan Province, Baoshan Pref., Shweli valley *Forrest* 12043 (E), India: India orientalis, *Roxburgh* 154 [BM000958163] (BM), Lao People's Democratic Republic: Forests surrounding Nakai NBCA area office, Khammouan *Newman* et al. LAO 345 (E), Myanmar: Chappedong orae Tenasserim, *Wallich* 6578 [K001124247] (K). Kachin State, Bhamo District, Momauk *Lace* 5728 (E), Shan State, Keng Tung *MacGregor* 224 (E). Nepal: Mechi, 5 miles E. of Soktim tea estate *Stainton* [BM013718323] (BM), Thailand: Chiang Mai, Eastern slopes of Doi Suthep; c. 1500 m Hosseus 508 [BM013718307] (BM), Chiang Mai *Newman* 1010 (E), Chiang Mai *Kerr* 559 (E), Vietnam: Langbian Peaks, South Annam *Kloss* [BM013718311] (BM), Lao Cai, Sa Pa District, lower slopes of Fansipan *van der Werff* 14435 (E), Quang Ninh:Taai Wong Mo Shan, near Chuk-phai, Ha-coi, Tonkin *Ts'ang* 29108 (E), Lam Dong, Lac Duong District, slopes of Mount Lang Bian *Newman* 134 (E), Vinh Phuc Province, Tam Dao 2 *Škornicková* 854 (E), Vinh Phuc, Tam Dao N. P. *Bình & Dang* 1440 (E), Lao Cai Province, Sapa, Ban Khoang, wet places in Ban Khoang village *Rushforth* 4450 (E),

3.2 Methods

Specimens and illustrations labelled as *A. calcarata* (Haw.) Roscoe, *A. bracteata* Roscoe, *A. bracteata* Roxb., *A. blepharocalyx* K. Schum., and *A. roxburghii* Sweet and other Indian *Alpinia* species were examined at E, K, BM, LINN, and LIV. Digitised specimens on JSTOR Global Plants (https://plants.jstor.org/) were also examined. The original descriptions and correspondence associated with them were examined in the herbaria mentioned above and on Biodiversity Heritage Library (https://www.biodiversitylibrary.org/).

3.3 Results

The original specimen of *A. bracteata* Roscoe in Monandr. Pl. Scitam. (Roscoe, 1828) is badly drawn, especially the vegetative parts which do not resemble *Alpinia* or any other species from the Zingiberaceae (Figure 3.2). As the habit was described

Figure 3.2: An illustration of *Alpinia bracteata* Roscoe in Monandr. Pl. Scitam. 9/10, pl. 70 (1828). This illustration is based on a Chinese drawing that belonged to Lord Stanley and is badly drawn. Image courtesy- Biodiversity Heritage Library (www.biodiversitylibrary.org). Contributed by Missouri Botanical Garden.

as being more like a *Costus* than *Alpinia*, it is possible that the Chinese drawing used to describe the species was not an accurate illustration of the species. Hence, an epitype for *A. bracteata* Roscoe must be chosen. On examination of herbarium specimens, we found that the relative usage of *Alpinia bracteata* Roscoe against

A. bracteata Roxb. is low, but this name also has not been applied often. Hence, the name *A. roxburghii* Sweet should be applied for specimens labelled as *A. bracteata* Roxb.

Figure 3.3: Specimen chosen as the lectotype of A. calcarata (Haw.) Roscoe. This is the original illustration of *Renealmia calcarata* (basionym of *A. calcarata*) in The Botanists' repository (Pl. 421, 1805). Image courtesy- Biodiversity Heritage Library (www.biodiversitylibrary.org). Contributed by Missouri Botanical Garden.

As *A. calcarata* (Haw.) Roscoe was described before the development of the type concept, we propose to assign the original illustration published as *Renealmia calcarata* Haw. in The Botanists' Repository (Haworth, 1805) as the lectotype (Figure 3.3). The protologue itself is vague and the only distinguishing character mentioned is lance-sword-shaped leaves and an erect terminal racemus of flowers. An epitype will also be useful and can be a specimen collected from the region in south-east India (preferably from the Coromandel region). The epitype can also be a specimen from a herbarium in India that has been collected from South India.

3.4 Conclusions

The genus *Alpinia* has always had a complicated taxonomic history and continues to present challenges to taxonomists. The lack of extensive sampling of species across their geographical range has caused difficulties in telling species apart within certain sections in this genus. Moreover, many species in *Alpinia* remain undescribed based on the examination of herbarium specimens and live specimens. Taking detailed notes in the field, and taking high quality images are essential when collecting *Alpinia* specimens.

Resolving confusion among the names applied in *Alpinia* is the first step towards untangling the complicated taxonomic history of this genus. Along with this, thorough morphological and molecular analyses to understand species relationships and determining species is crucial in *Alpinia*. The extent of hybridisation in this genus also needs to be investigated as many species within the same clade have overlapping morphological characters.

Alpinia and many other genera from the Zingiberaceae have not been recently revised in India. There may be several *Alpinia* species in India that are yet to be described but this may be difficult due to some of the taxonomic challenges mentioned in Chapter 2. This work will lead to clarification of names which will aid in the revision of this genus. Like the *A. bracteata*, *A. calcarata* and *A. roxburghii* confusion, many other species names in *Alpinia* need to be resolved, such as the name *A. malaccensis* (Burm.f.) Roscoe, which has been applied to all species that have similar morphological characters right from the Western Ghats in India to China and Indonesia.

Note: Following Art. 30.9 in the Shenzen Code, this chapter is not intended as an effective taxonomic publication.

Chapter 4

Crossing barriers and the potential for hybridisation in the genus *Alpinia* Roxb.

4.1 Introduction

Reproductive isolation is central to the speciation process (Coyne and Orr, 2004). Understanding the types of barriers that contribute to reproductive isolation and how they evolve is crucial for understanding diversification (Lowry et al., 2008). Reproductive barriers occur at different stages and can be categorised as either prezygotic or postzygotic. Prezygotic barriers include niche differentiation, non-overlapping phenology and pollinator differences, and the interactions between pollen and pistil that impede interspecific mating (competition with conspecific pollen) or fertilisation (failure of nonconspecific pollen to fertilise eggs) (Rieseberg and Willis, 2007). Postzygotic barriers include zygote abortion, hybrid inviability and infertility (Baack et al., 2015). All these barriers combined contribute towards reproductive isolation. Postzygotic barriers are particularly important as they are the final barrier to interspecies mating.

A key finding from studies of many plant groups is that the strength of reproductive isolation increases with genetic distance (e.g., *Silene*- Moyle et al., 2004; *Coreopsis*- Archibald et al., 2005; food-deceptive orchids- Scopece et al., 2007; *Helianthus* and Madiinae- Owens and Rieseberg, 2014). This is due to the low divergence between closely related species where the species may not have accumulated reproductive barriers (Moyle et al., 2004) as these barriers accumulate over time. This correlation particularly applies to postzygotic barriers, where genetic incompatibilities can accumulate rapidly as species diverge. The number of genic incompatibilities, also known as Dobzhansky-Muller incompatibilities (DMI), between taxa increases much faster than linearly with time (Orr, 1995; Matute et al., 2010; Moyle and Nakazato, 2010; Larcombe et al., 2015). These incompatibilities are considered to be among the main causes of postzygotic reproductive isolation as they lead to hybrid inviability or sterility (Orr and Turelli, 2001). This is caused as a result of interactions that cause low viability of hybrid seeds or reduced pollen and/or ovule fertility.

Natural hybridisation is common in plants, with about 25% of plant species involved in hybridisation with at least one other species (Mallet, 2005, 2007). This natural hybridisation can have a wide range of outcomes. It can be beneficial as it can lead to

adaptive introgression where it increases the fitness of the recipient species (Abbott et al., 2013) or it can have detrimental effects where extensive introgression can drive a species to extinction (Rhymer and Simberloff, 1996). Hybridisation also has a significant impact on the way species are delimited as it causes the sharing of genetic variation and obscures morphological differences between species (Soltis and Soltis, 2009). Therefore, the presence of hybridisation can cause difficulties in phylogenetic analyses, and in taxonomy. Studying the occurrence and impact of hybridisation is critical, particularly for rare and endangered species where hybridisation may be a potential cause of population decline and extinction.

Alpinia Roxb. is the largest and most taxonomically challenging genus in the ginger family (Zingiberaceae). It is made up of more than 230 species, and the genus is widespread throughout tropical and subtropical Asia. These species are distinguished only on the basis of plesiomorphic characters in the tribe Alpinieae such as a terminal inflorescence on the leafy shoot. *Alpinia* species are usually found in the forest understorey along water bodies, and are pollinated by carpenter bees, birds and bats. This genus has had a complex taxonomic history with molecular analyses showing that it is polyphyletic, forming six clades in the tribe Alpinieae (Figure 2, Kress et al., 2005, Chapter 1). Many species within clades have similar morphological characters and are difficult to tell apart. Discerning generic boundaries within the tribe Alpinieae has always been challenging, and molecular studies over the years have shed a light on the complexity of these relationships (Kress et al., 2005, 2007). The polyphyletic nature of several genera within this tribe has challenged generic concepts and there is a need for recircumscription of old genera and the description of several new genera. Some of this work has already been carried out in genera such as *Amomum* Roxb. (de Boer et al., 2018) and a few groups of *Alpinia* as well (Docot et al., 2019).

To date, very few studies have examined the occurrence of hybridisation in *Alpinia* (but see Liu and Wang, 2009; Liu et al., 2009) and the potential role it may play in blurring species boundaries and causing taxonomic complexity. Most genera in the Alpinieae are diploids with the same chromosome number (2n = 48) (Ramachandran,

1969; Rice et al., 2015), suggesting that ploidy differences are unlikely to be a barrier to hybridisation. The presence of intermediate morphology, low pollen viability, overlapping phenology and sympatry of parent species indicated hybridisation of *Alpinia* species in Taiwan. Using nuclear and plastid markers, Liu et al. (2009) confirmed that naturally occurring morphologically intermediate individuals were hybrids.

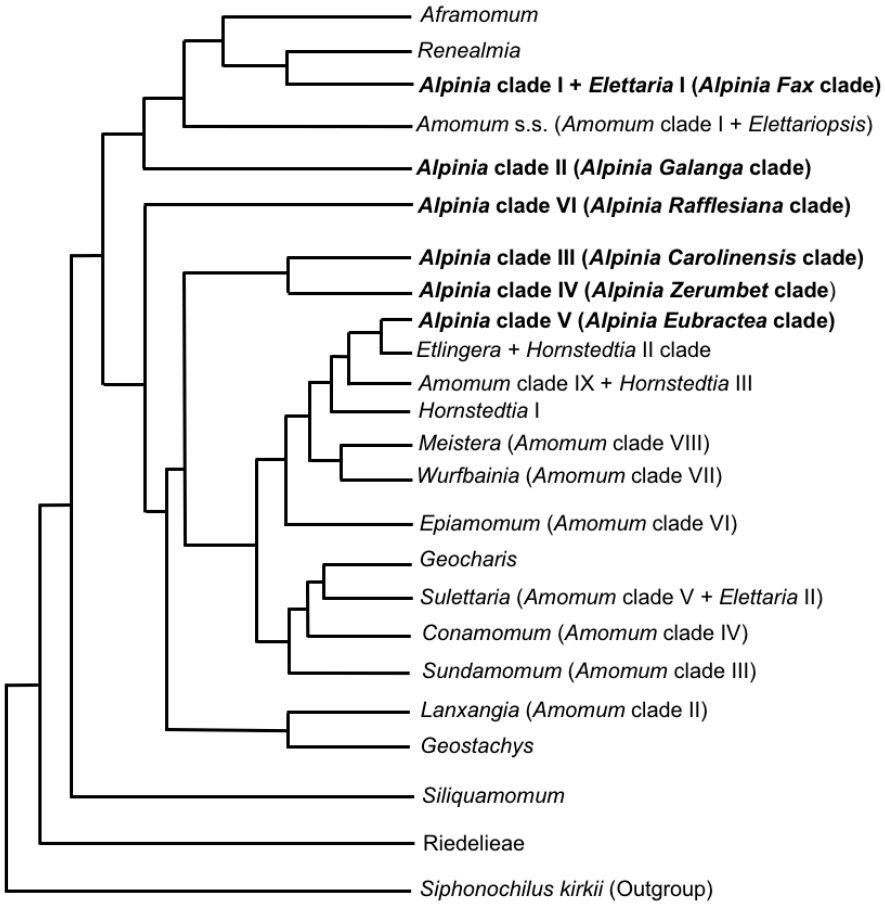

Figure 1: A condensed tree of the subfamily Alpinioideae. The six polyphyletic clades of *Alpinia* are highlighted in bold. Adapted from the molecular phylogeny of *Alpinia* based on ITS and matK markers by Boer et al (2018).

In some cases, more than two species were involved in hybridisation. The occurrence of hybridisation in Taiwanese *Alpinia* might be a result of secondary contact due to habitat disturbance. Interspecific crosses can lead to the generation of hybrids that have improved floral and vegetative characters than the parental species. As *Alpinia* is an economically important genus and many species are grown as ornamental

plants, the knowledge that *Alpinia* species are capable of hybridisation can be utilised in horticulture.

This chapter uses artificial hybridisation experiments to investigate the potential for hybridisation within and between *Alpinia* clades. Investigating the cross-compatibility of species within and between clades is essential information for understanding whether there is strong postzygotic reproductive isolation between these species. Given the complexity of relationships within and between genera in Alpinieae, understanding the extent of reproductive isolation may also help us improve our understanding of the mechanisms that prevent gene flow between taxa. To investigate this, I addressed the following research questions:

1. Are species within and between clades of *Alpinia* cross-compatible?
2. Is there a correlation between the genetic distance and seed set?

If morphologically distinct taxa are cross-compatible this suggests the potential for natural hybridisation, which may have been overlooked in this species-rich group. However, if the strength of reproductive isolation increases with genetic distance then we may expect hybridisation to be limited only to closely related species pairs.

4.2 Methods

4.2.1 Manual pollination experiments

The tropical glasshouses and research collections at the Royal Botanic Garden Edinburgh include more than a 100 accessions of *Alpinia* along with many other species of Zingiberaceae. Most of the accessions flower in the spring and therefore crosses were performed from March-May. To investigate reproductive barriers and the potential for hybridisation within *Alpinia*, the following types of crosses were performed in 2017 and 2018:

- Self (S)
- Cross (C)- between different individuals of the same species.
- Close Outcross (CO)- between species within the same clade

- Far Outcross (FO)- between species in different clades or between closely related genera.
- Emasculated (E)- removal of anthers before their dehiscence.

Figure 2: A few of the *Alpinia* accessions used for manual hybridisation at the Royal Botanic Garden Edinburgh. (A) *Alpinia* aff. *hainanensis* from China, (B) *Alpinia pricei* from Taiwan, (C) *Alpinia* aff. *melichroa* from Indonesia, (D) *Alpinia polyantha* from Vietnam, (E) *Alpinia purpurata* from Solomon Islands, (F) *Alpinia rafflesiana* from Malaysia. Accessions A, B, C and D are from clade IV, E is from clade V and F is from clade VI sensu Kress et al., 2005.

For all treatments, the flowers were emasculated and the pollen was collected and deposited on the stigma. In total, 44 maternal and 47 paternal accessions were used for crossing (Appendix 4.1). Each treatment was replicated depending on the number of flowers open per day on the parent species, with the aim of achieving a minimum of five crosses per parental combination where possible. The crosses were carried out at different times of the day, from early morning until late afternoon. Each accession was observed at different times of day to check for the occurrence of flexistyly and the type of morph. Each cross was assigned a unique ID and the flowers were tagged with this ID and the date. The cross type, maternal and paternal accessions, time of crossing, date, and position of the maternal flower on the inflorescence were recorded. Flexistyly, a unique floral dimorphism found in several genera in the

Alpinieae (discussed in Chapter 5) was also recorded as it may affect pollen viability and stigma receptivity. After crossing, the accessions were monitored to check for fruits and fallen tags and aborted fruits. The fruits ripened four to six months after crossing. Once collected, the seeds per fruit were counted. About 1000 crosses were performed in 2017 and 2018. Species from genera such as *Adelmeria*, *Meistera*, and *Etlingera* were also included.

To check for seed viability, 10 seeds per fruit were randomly selected from the successful crosses made in 2017, and germinated at the glasshouses at Kings' Buildings. The seeds came from 25 maternal parent accessions, and were sown in 262 pots. The seeds were first soaked in water for a few days to remove the dry aril coating the seeds. Once cleaned, five seeds per pot were sown in the potting mix of potting bark, propagation bark, perlite, charcoal, osmocote and dolomitic limestone. The number of germinated seedlings per pot was counted between May and September 2018.

4.2.2 DNA extraction, PCR and sequencing

My aim was to use obtain genetic data, to (1) estimate genetic distances for parents used in crosses, (2) confirm experimental crosses were of hybrid origin, (3) assign some unnamed individuals to their clades. For these purposes, I selected the nuclear ribosomal Internal Transcribed Spacer (ITS) as a marker. This marker has been commonly used to resolve relationships within *Alpinia* (Rangsiruji et al., 2000; Kress et al., 2005) and to discern generic boundaries within Alpinieae (Kress et al., 2007; de Boer et al., 2018), often in combination with other plastid markers.

The leaf tissue of the accessions mentioned in Appendix 4.1, along with the hybrids germinated in the Kings' Buildings glasshouses from the successful crosses listed in Appendix 4.3 were collected. The leaves were divided into pieces of about $2cm^2$ and dried in desiccating silica gel. About 18-20mg of dried leaf material was used for genomic DNA extraction using the Qiagen DNeasy Plant Mini Kit following the manufacturer's protocol. The primers used to amplify ITS were ITS5 (forward primer): 5'-GGAAGTAAAAGTCGTAACAAGG-3' and ITS4 (RP): 5'-

TCCTCCGCTTATTGATATGC-3' (White et al., 1990). Each 25µl PCR amplification used 1µl of template DNA, 0.5µl each of forward and reverse primers (10µM), 12.5µl of Taq 2X Master Mix (New England Biolabs), and 10.5µl of dH_2O. The following steps were used for amplification- initial denaturation at 95°C for 5 min, followed by 35 cycles of denaturation at 95°C for 30sec, annealing at 48°C for 40 sec and extension at 72°C for 1 min, and final extension at 72°C for 5min. The PCR products were visualised on a 2% agarose gel stained with SYBRSafe. The successfully amplified PCR products were then cleaned with a SAP-Exo Kit (Jena Bioscience) following manufacturer's protocol and submitted to Edinburgh Genomics (University of Edinburgh, UK) for Sanger sequencing on the ABI3730 using the ITS5 and ITS4 primers.

19/35 parental accessions and 0/32 hybrids sequenced had high quality sequences that were used to calculate genetic distances between accessions. For the sequences that were poor quality, the PCR product was diluted before the clean-up and re-submitted to Edinburgh Genomics and to Eurofins (Wolverhampton, UK).

4.2.3 Sequence alignment and genetic distance

The forward and reverse sequences were assembled using the *De Novo* Assembly option and then quality checked and edited manually in Geneious v. 8.1.9 (https://www.geneious.com). Sequences were aligned using Automatic alignment (Global alignment using free end gaps) followed by manual edits. A Neighbour-Joining tree was generated based on the alignment and the patristic distance matrix was used to obtain pairwise genetic distance.

4.2.4 Statistical analyses

Generalised linear mixed-effect models (GLMMs) were used to evaluate the effect of crossing treatment, flexistyly, time of crossing, inflorescence position, and genetic distance on the crossing success and the seed set. The response variables were analysed as either Poisson or Binomial. The analyses were done in R version 3.6.1 using the package lme4 (Bates et al., 2015). The first model was to estimate the

success of crosses where the response variable was success (1 for success, 0 for failure to set fruit), the fixed effects were the cross type, inflorescence position, and flexistyly morph type (anaflexistyly or cataflexistyly). The second model compared the seed-set for the different cross types with cross type, inflorescence position and morph type, time of the cross as fixed effect and ordered accession numbers and names of maternal and paternal species as the random effects. The correlation between the genetic distance and seed set was analysed by a GLMM with genetic distance as the fixed effect and ordered accession numbers and names of maternal and paternal species as the random effects. The relationship between genetic distance and seed set was also analysed. All the graphs were plotted using ggplot2 (Wickham, 2016).

4.3 Results

4.3.1 Crossing summary

198/1000 crosses were successful (78/284- self, 3/11- cross, 105/376- close outcross, 12/312- far outcross) (all crosses listed in Appendix 4.2 and successful crosses in Appendix 4.3). Accessions such as *Alpinia* aff. *haenkei* (19991898A) and *Alpinia* cf. *malaccensis* (20081105A), set seeds for all types of crosses for most of the replicates except emasculation. Some accessions had a higher proportion of successful selfed crosses such as *Alpinia* aff. *melichroa*. Autonomous self-pollination was observed in a few cataflexistylous accessions such as 20001434A, *Alpinia warburgii* (20091014A), *Alpinia* aff. *melichroa* (20080412A), and *Alpinia shimadai* (19934175C). It was also observed in accessions that lacked flexistyly such as *Alpinia* aff. *haenkei* (19991898A) where the stigma was in close proximity with the anthers. Some accessions such as *Adelmeria* cf. *pinetorum* (19972512A) and *Alpinia purpurata* (20070002B) did not set seed for any cross type. The germination success per successful cross was 19/55 for selfs, 1/3 for cross, 29/61 for close outcross, and 3/8 for far outcross (Appendix 4.4).

Figure 4.3: Fruits growing on some of the *Alpinia* accessions in the glasshouses at the Royal Botanic Garden Edinburgh.

4.3.2 Comparison of the success of different cross types

The success of self-pollinated crosses compared to the other cross types indicated that close outcrosses were more successful in terms of the number of crosses but this was not significant (Table 1). The only cross type that was significantly different from the selfs was far outcross. In terms of the average seed set per cross type, self-pollinated crosses had a higher seed set than the other cross types (Figure 4.4). The average seed set for self was 34, cross was 29, and 18 for both close and far outcross. Even though the average seed set in close outcross is low, there is a high variance in the seed set.

Table 1: Result of the generalised linear mixed model for the effect of cross type on the success of experimental *Alpinia* crosses. The model compares the success of self-pollinated (represented by the intercept) with other cross types. The random effects are represented by ordered species accession numbers or species names of female (pollen recepient) and male (pollen donor) into a single factor. The model assumes binomial residuals. *** p < 0.001, ** p < 0.01, * p < 0.05.

Fixed effects	**Estimate**	**Std. Error**	**z value**	**Pr(>|z|)**
(Intercept)	-1.6772	0.3973	-4.221	0.0000243***
cross_type - close_outcross	0.1634	0.4423	0.369	0.7118
cross_type - cross	-0.6304	1.1182	-0.564	0.5729
cross_type - emasculated	-2.5601	1.4924	-1.715	0.0863
cross_type - far_outcross	-2.4387	0.5578	-4.372	0.0000123***
Random effects				
Groups	Name	Variance	Std.Dev.	
species_accession_female_male	(Intercept)	1.692	1.301	
species_name_female_male	(Intercept)	1.313	1.146	

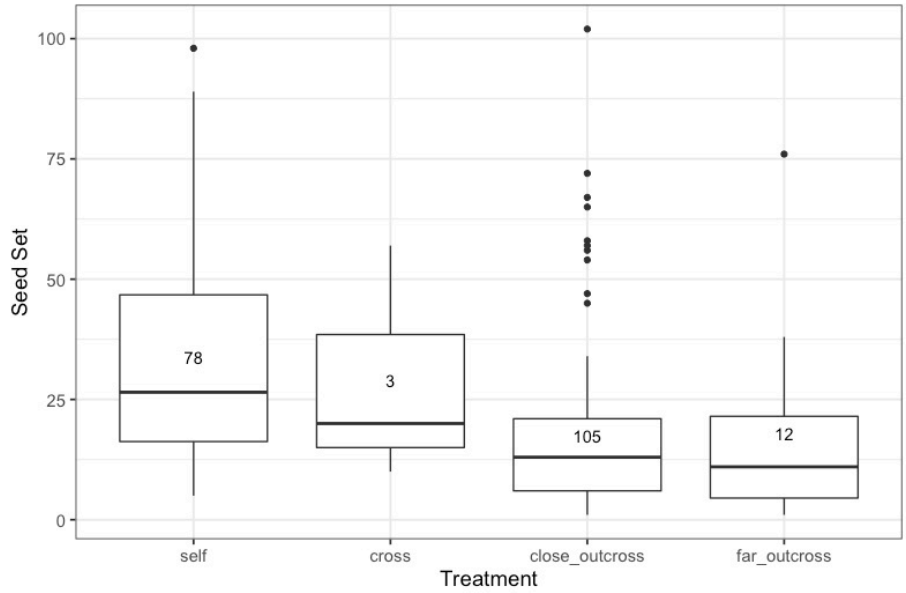

Figure 4.4: Mean seed set for the various crossing treatments performed on the different *Alpinieae* accessions at the Royal Botanic Garden Edinburgh. The numbers on the box indicate the number of successful crosses for each cross type.

4.3.3 Comparison of the seed set of different cross types

The number of seeds produced for close and far outcross was significantly different from the self seed set (Table 2). Factors such as position of the flower on the

inflorescence, time of the cross, and flexistyly did not have a significant impact on the seed set. Although the inflorescence position is not significant, its value indicates that there might be a negative correlation between number of seed set and inflorescence position (Table 2). This indicates that flowers on higher inflorescence positions produce fewer seeds. Although the success of close outcrosses is higher than self, the average number of seeds set in the close outcrosses is lower.

Table 2: Result of the generalised linear mixed model for effect of cross type on the seed set of experimental *Alpinia* crosses. The model compares the number of seeds produced for non-flexistylous accessions that were self-pollinated with other cross types and anaflexistylous or cataflexistylous accessions, position on the inflorescence of the flower crosses, time of the cross. The random effects are represented by ordered species accession numbers or species names of female (pollen recepient) and male (pollen donor) into a single factor. The model assumes Poisson residuals. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Fixed effects	**Estimate**	**Std. Error**	**z value**	**Pr(>|z|)**
(Intercept)	3.256506	0.320312	10.167	<2e-16 ***
cross_type - close_outcross	-0.388261	0.197069	-1.97	0.0488 *
cross_type - cross	-0.149785	0.426733	-0.351	0.7256
cross_type - far_outcross	-0.698824	0.313536	-2.229	0.0258 *
inflo_position	-0.00163	0.002423	-0.673	0.5011
time_hr	0.016418	0.0156	1.052	0.2926
flexistyly_female – Anaflexistyly	0.556861	0.363455	1.532	0.1255
flexistyly_female - Cataflexistyly	-0.276885	0.362068	-0.765	0.4444
flexistyly_female - Anaflexistyly:time_hr	-0.025668	0.020871	-1.23	0.2188
flexistyly_female - Cataflexistyly:time_hr	-0.032853	0.020808	-1.579	0.1144
Random effects				
Groups	**Name**	**Variance**	**Std. Dev.**	
species_accession_female_male	(Intercept)	0.3091	0.556	
species_name_female_male	(Intercept)	0.1479	0.3846	

4.3.4 Effect of genetic distance on seed set

The pairwise genetic distances ranged from 0-0.121 for 19 accessions that had high quality nrITS sequences (Appendix 4.5). The species pair *Alpinia* cf. *malaccensis* (20081105A, Clade IV) and *Alpinia mutica* (19901470, Clade IV) had a genetic distance of zero whereas the species pair *Alpinia roxburghii* (20120190B, Clade IV) and *Alpinia purpurata* (20070002B, Clade V) had the highest genetic distance of

0.121. The relative seed set decreases as the genetic distance increases (Figure 5), but this correlation is not significant (Table 3).

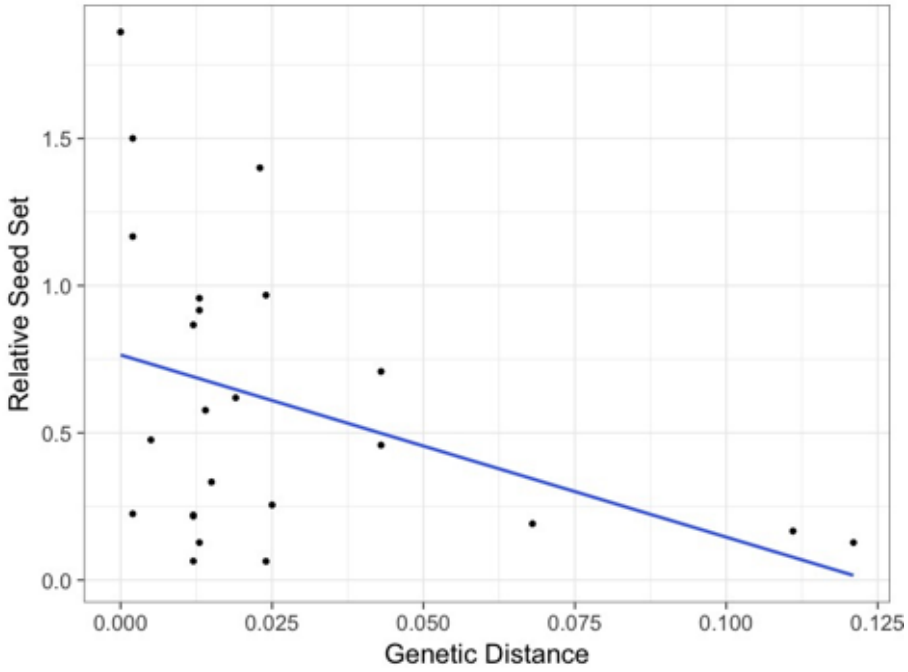

Figure 5: Relative seed set versus genetic distance for crosses between *Alpinia* species. Relative seed set is the ratio of the close or far outcross seed set by the self seed set for that particular maternal accession. The genetic distance was measured using nrITS.

Table 3: Result of the generalised linear mixed model for the effect of genetic distance on the seed set of experimental *Alpinia* crosses. The model compares the number of seeds with genetic distance (measured using nrITS). The random effects are represented by unique ID for each cross and ordered species accession numbers or species names of female (pollen recepient) and male (pollen donor) into a single factor. The model assumes Poisson residuals. *** p < 0.001, ** p < 0.01, * p < 0.05.

Fixed effects	**Estimate**	**Std. Error**	**z value**	**Pr(>|z|)**
(Intercept)	2.6344	0.1939	13.588	<2e-16***
genetic_distance	-1.277	5.7024	-0.224	0.823
Random effects				
Groups	Name	Variance	Std.Dev.	
cross_no	(Intercept)	0.23	0.4796	
species_accession_female_male	(Intercept)	0.2202	0.4693	
species_name_female_male	(Intercept)	0.3805	0.6168	

4.4 Discussion

4.4.1 *Alpinia* species show cross-compatibility under experimental conditions

My experimental crosses show that closely related *Alpinia* species can set seed when crossed artificially. Although most far outcrosses failed, a few of them were successful, which indicates that distantly related species can potentially cross. The proportion of successful crosses was comparable for self (27.4%) and close outcrosses (27.9%) but self-pollinated accessions set the highest average seed. A negative correlation between the genetic distance and seed set was observed and the far outcross treatment had the lowest proportion of successful crosses (3.8%). Seed set was observed to decrease with increasing genetic distance which is consistent with several studies that have investigated this correlation, but this correlation was not significant (Table 3). My results show that closely related species are cross-compatible and successfully set seeds, with some evidence that more distantly related species can set seed in a few instances.

Self-pollinated accessions set the highest average seed, indicating that accessions from several *Alpinia* and other closely related genera are self-compatible (Figure 4). This is consistent with studies that have carried out self-pollination in the Alpinieae (Li et al., 2002; Zhang et al., 2003; Wang et al., 2005; Cui and Li, 2015), but there were many accessions where no self-pollinated flowers set fruit (Appendix x). This indicates that self-incompatibility in *Alpinia* is polymorphic or that there might be some technical issues that have affecting pollination success. These technical issues may include factors such as pollen viability or stigma receptivity. In the self-pollinated treatments for many of the artificial pollination carried out in *Alpinia* and closely related genera, seed set for self-pollinated flowers was lower than the outcrossed seed set indicating inbreeding depression (Li et al., 2002; Ren et al., 2007; Sun et al., 2007; Jia et al., 2015).

Delayed selfing in cataflexistylous morphs is possible if there is some pollen remaining on the anthers. When the style curves downwards, it might come into

contact with the stigma leading to self-pollination. This has been reported from *Alpinia galanga*, where the cataflexistylous morph is self-compatible and this is a strategy for reproductive assurance in the absence of pollinators (Cui and Li, 2015).

Most of the unsuccessful treatments were the ones that failed to set fruit at all, but in some cases, postpollination postzygotic barriers such as zygote abortion were observed. This abortion of fruit is generally associated with shortage of resources or limited resources allocated to female function (Stephenson, 1981). Alternatively, the fruits might have fallen off when there were other crosses done on that accession. The inflorescence position was measured from the bottom, and there was a negative (but non-significant) correlation between inflorescence position and number of seeds. *Alpinia* and other closely related genera have an acropetal inflorescence development, where it is often presumed that the probability of fruit set or number of seeds per flower is higher for early opening flowers (Medrano et al., 2000). This is true for species with one flower per cincinnus but might not be true for species such as *A. galanga*, which have many flowers per cincinnus and flower in acropetal waves. Some genera in the Zingiberaceae such as *Globba* L. have evolved a separation of the sexes, hermaphrodite flowers at the base of the inflorescence and functionally male flowers above (Sangvirotjanapat et al., 2019). This also occurs in *Alpinia* sect. *Myriocrater* and *Alpinia* sect. *Monopleura* (Burtt and Smith, 1972). Therefore, the position of the flower on the inflorescence might have an impact on the seed set.

It was not possible to verify the crosses for the accessions that set seed across all treatments by using genetic markers as most of the hybrids had poor quality sequences. It is possible that while crossing, some self-pollen was deposited on the stigma (if the anthers had already dehisced depending on the time of the cross), which would have resulted in a mixed pollen load, with fruit for several crosses for those particular accessions. This might be unlikely, as I ensured that I provided plenty of supplementary pollen. Autonomous self-pollination is especially not possible for anaflexistylous accessions as the stigma is curved above the anther once the anther releases pollen. Based on the overall results, several *Alpinia* accessions

can cross therefore have not accumulated DMIs. But the divergent taxa may have accumulated these incompatibilities as the far outcrosses were not as successful as the other type of crosses.

4.4.2 Reproductive barriers within the Alpineae

Reproductive barriers within the Zingiberaceae have not been well studied. The most widely studied genus from an evolutionary perspective in the Zingiberales is *Costus* (Costaceae), where postpollination barriers such as lower conspecific pollen adhesion, reduced germination rates, slower pollen tube growth and pollen tube attrition are responsible for the lack of hybrids between sympatric species even though these species share pollinators (Yost and Kay, 2009).

Other than postpollination prezygotic and postzygotic barriers such as failure to set fruit, zygote abortion and low seed set, *Alpinia* and other closely related genera might possess are range of other, prezygotic, reproductive barriers such as non-overlapping distributions and non-overlapping phenology. Many of the accessions crossed in this study have a wide variation in floral size and form which might indicate non-overlapping pollinators. Clade IV is the largest and most well-represented in this study. The species from other clades such as Clade I were not present in the collection at RBGE and the species from Clade II did not have an overlapping flowering period with the rest of the clades. The non-overlapping flowering period here might serve as a reproductive barrier in the wild. Other postpollination reproductive barriers such as F_1 viability and fertility could not be tested as the F_1s take at least two years to produce flowers and fruits.

Factors such as non-overlapping floral phenology and niche differentiation might be the most effective reproductive barriers in *Alpinia* (Liu et al., 2009). Flexistyly, a unique floral dimorphism within the Alpinieae, is an effective mechanism to prevent self-pollination but it might not prevent interspecific gene flow. The presence of flexistyly might affect stigma receptivity and pollen viability as the stigma is considered to be receptive when the style is curved below the anther and the pollen might be viable only after it has dehisced from the anthers.

Floral size differentiation has also been reported from *Alpinia nieuwenhuizii* in Borneo where the floral size varied between habitats. The smaller floral type was found in forest floors and the large one along roadsides and riversides and they also had different legitimate pollinators. This indicates that floral size dimorphism might also play an important role in reproductive isolation in *Alpinia* species (Takano et al., 2013).

4.4.3 Evolutionary significance of reproductive isolation in *Alpinia*

Only a few studies have investigated the presence of hybridisation within *Alpinia* (Liu and Wang, 2009; Liu et al., 2009). Given the difficulty in delimiting species based on morphology, investigating the occurrence of hybridisation is necessary as this may blur species boundaries in the wild. Although most accessions crossed in this study are allopatric (Appendix 4.1), this study indicates that there is the possibility that where species do co-occur they may be cross compatible and this may allow the formation of hybrids. Hybridisation may also be possible as most species within *Alpinia*, or even the Alpinieae have the same chromosome number (2n=48). As *Alpinia* is a horticulturally important genus, generating new hybrids can be useful for ornamental plants. This has been previously done to generate a range of hybrids, including intergeneric hybrids such as *Alpinia purpurata* x *Etlingera elatior* (Luc-Cayol and Fereol, 1997).

In conclusion, *Alpinia* species show widespread interspecific cross compatibility, especially within clades, and in a few instances, between divergent clades as well as genera. Whether this cross-compatibility in the glasshouses translates to hybridisation between *Alpinia* species in the wild still needs to be tested. As habitat destruction due to anthropogenic effects is on the rise, it may be possible for many species to encounter nonconspecific pollen due to secondary contact which might result in hybridisation.

Chapter 5

Elucidating the genetic basis of flexistyly, a novel reproductive trait in *Alpinia*

5.1 Introduction

Angiosperms possess a remarkable range of floral traits and forms that enable them to reproduce successfully (Barrett, 2010b). As flowering plants are mainly hermaphroditic in nature, opportunities for self-fertilisation can occur. This can lead to reduction in viability and fertility of offspring, or inbreeding depression (Charlesworth and Charlesworth, 1987). Therefore, strategies to avoid self-pollination are widespread in angiosperms and many of these reproductive strategies have evolved to promote outcrossing by means of a pollen vector (Barrett, 2010a). The evolution of these reproductive strategies might also be driven by a range of other factors to prevent conflicts that interfere with pollen deposition on the stigma and pollen removal from the anthers, which may not be necessarily associated with self-pollination (Barrett, 2002a).

Among many such floral traits are stylar polymorphisms that promote insect–mediated cross-pollination and prevent pollen wastage due to interference between male and female function within a bisexual flower (Barrett, 2010a). The most widely studied stylar polymorphism is heterostyly, where each population comprises two (distyly) or three (tristyly) floral morphs that differ reciprocally in style and stamen lengths. This polymorphism was first recognised as early as the 16th century in the genus *Primula* L. (Ornduff, 1992). Heterostyly was also one of Darwin's research interests. His work lead to the publication of *The different forms of flowers on plants of the same species* (Darwin, 1877) which has laid a strong foundation for the research on heterostyly over the years.

Heterostyly has multiple independent origins within flowering plants and is found in at least 199 genera from 28 families (Barrett, 2002b; Naiki, 2012). Distyly was thought to be a diallelic trait, with the short styled morph dominant (S/s) and the long styled morph recessive (s/s). Three genes at the *S* locus were thought to control the trait- G (Griffel (style) length), P (Pollen size and number) and A (Anther position) (Lewis and Jones, 1992) held together by tight linkage due to suppressed recombination (Charlesworth, 2016). Recent work on the genetics of heterostyly in *Primula* has revealed that the short-styled morphs are hemizygous and not

heterozygous for the *S* locus. The *s* haplotype lacks a 278 kb sequence that contain genes such as *CYP734A50* (Cytochrome P450) that controls style length and *GLO2* (short-morph specific *GLOBOSA* gene) that controls anther height (Huu et al., 2016, 2020; Li, Cocker, et al., 2016). The hemizygous nature of distyly was also found in other systems such as *Fagopyrum* (Yasui et al., 2012), *Linum* (Ushijima et al., 2012), and *Turnera* (Shore et al., 2019). Therefore, we are now learning more about the genetic basis of this floral trait in a range of plant groups. Many other forms of stylar polymorphism exist such as enantiostyly, inversostyly, stigma-height dimorphism (where the stamen levels are the same), and flexistyly (summarised in Barrett (2010a)).

Among these stylar polymorphisms, the only one that involves stylar movement is flexistyly. This phenomenon was first reported by Li et al., (2001) where they observed this new mechanism in nine species of *Alpinia.* It is known to reduce pollen-pistil interference and promote outcrossing . Species displaying flexistyly possess two morphs that exhibit spatial as well as temporal separation of sexual function. The morphs can either be cataflexistylous (protandrous, cata=downwards), in which the anthers disperse pollen in the morning, while the style is curved above the anther, or they can be anaflexistylous (protogynous, ana=upwards), in which the style is curved below the anther while the anther does not disperse pollen (Figure 5.1). Around midday, the style of the cataflexistylous flower curves below the anther and receives cross pollen whereas the anther halts pollen release and the style of the anaflexistylous flower curves above the anther and the anthers begin to release pollen (Li et al., 2001, 2002; Zhang et al., 2003).

Stylar movement in *Alpinia* was observed as early as 1820 in Roxburgh's Flora Indica where he noted that *Alpinia galanga* possessed a style that curved above the anther and recurved under the anther afterwards (Roxburgh, 1820). But this phenomenon was not observed in detail until 1996 when Cui et al. discovered reciprocal stylar movement in *Amomum tsao-ko* (now accepted as *Lanxangia tsaoko*) (Cui et al., 1996). They observed the two floral morphs and found that only one morph type can be borne on a particular plant. The response of flowers was found to

be dependent on photoperiod and the flowers were found to be self-compatible when artificially pollinated. The fruit set in the presence of only one morph type was found to be low hence it is pollinator dependent for reproductive success via this mechanism.

Figure 5.1: Flexistyly in *Alpinia* accessions at the Royal Botanic Garden Edinburgh. A- Cataflexistylous morph (style curved above the anther along with anther dehiscence (*Alpinia* cf. *malaccensis,* Accession no. 19751793A). Picture taken at at 10:22am. B- Anaflexistylous morph (style curved below the anther and no anther dehiscence) (Alpinia roxburghii, accession no. 20120190B). Picture taken at 10:30am.

Flexstyly has been reported from more than 24 species in Zingiberaceae subfamily Alpinioideae, mostly from *Alpinia* s.l. (Kress et al., 2005). It is also found in several genera such as *Amomum* s.l. , *Etlingera*, *Paramomum*, *Plagiostachys*, and *Siliquamomum* in the tribe Alpinieae (Cui et al., 1996; Kress et al., 2005; Leong Škorničková et al., 2014; Jia et al., 2015; Poulsen and Phonsena, 2017). The two morphs are always present in a 1:1 ratio in natural populations and hence it is presumed that two alleles at a single locus control flexistyly (Renner, 2001). A similar pattern is seen in walnuts (Gleeson, 1982) and pecans (Thompson and Romberg, 1985), where protandry is the recessive homozygote and protogyny is

either the heterozygote or dominant homozygote, but the heterozygote does not occur due to the lack of selfing.

Alpinia is a non-model plant genus that is economically important and the species from this genus are rich in compounds that have many pharmacological uses. *Alpinia* species are ancient polyploids that have undergone diploidisation with 2n = 48. Previous estimates of genome sizes in *Alpinia* include a 1C value of 2.29 pg for *Alpinia nigra* (Basak and Rangan, 2018). NGS analyses have been carried out for other closely related genera such as *Amomum*, where a target capture approach was used to construct the phylogeny of the genus, and genome size evolution was investigated. They found genome sizes ranging from 3.54-15.66 pg in *Amomum* s. s. (Hlavatá et al., in press). So far however, there is no complete genome assembly for *Alpinia* or the wider Zingiberaceae. The nearest available genomes are of *Curcuma longa* (Chakraborty et al., 2020), and *Musa* spp. from the banana family (Musaceae, Zingiberales) (D'Hont et al., 2012; Davey et al., 2013; Wu et al., 2016; Belser et al., 2018).

Whole-genome sequencing of pools of individuals or Pool-seq is a powerful and cost-effective approach to determine allele frequency differences across the genome (Futschik and Schlötterer, 2010). It allows for the analysis of genome-wide polymorphism data at a low cost and has a wide range of applications from bulked segregant analysis to genome evolution and can be used for model as well as non-model species (Schlötterer et al., 2014). When it comes to flexistylous species, a comprehensive study of their pollination and reproductive ecology has been done where the studies have investigated the cross compatibility of the two morphs and the significance of stylar movement (Li et al., 2001, 2002; Zhang et al., 2003; Takano et al., 2005; Ren et al., 2007; Sun et al., 2011; Aswani and Sabu, 2014; Jia et al., 2015; Yang et al., 2016) but little is known about the genetic basis of this trait. Therefore, my aims are:

1.To generate a reference genome for *Alpinia*.

2.To find the region(s) with allele frequency differences between the ana and cata-morph using a Pool-seq approach.

Here, I have used Pool-seq data to detect genomic regions where there are allele frequency differences between the two morph types. As a complex phenotypic trait, I expect flexistyly to be controlled by multiple but tightly linked loci. Based on this, my expectation is that there will be few regions of the genome where there are many F_{ST} outliers. Moreover, if this trait is hemizygous, I expect to see a coverage difference between the pools in regions of the genome underlying flexistyly.

5.2 Methods

5.2.1 Study system

Alpinia Roxb. is the largest and most taxonomically challenging genus in the Zingiberaceae. It comprises more than 230 species throughout the tropics and subtropics (Smith, 1990). Many species in this genus are flexistylous, one being *Alpinia nigra* (Gaertn.) B.L.Burtt, which is distributed from the Indian subcontinent to China (South Yunnan) (Figure 5.2). This species was chosen to study the genetics of flexistyly as large populations are easily found in its native range.

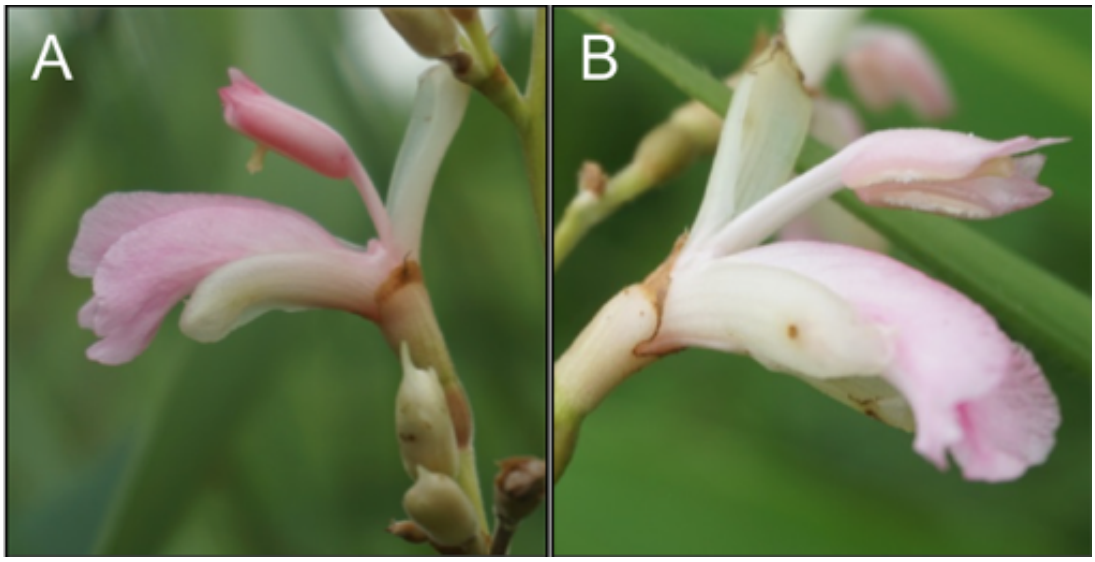

Figure 5.2: Flexistylous morphs of *Alpinia nigra* at Pakke Tiger Reserve, Arunachal Pradesh, India. A- Anaflexistylous (protogynous) morph (picture taken at 9:40am), B- Cataflexistylous (protandrous) morph (picture taken at 8:45am).

5.2.2 Sample collection

Samples were collected in July 2018 from Pakke Tiger Reserve near the Seijosa region (26°57' N 92°59' E) located in East Kameng district, Arunachal Pradesh (Collection permit number- CWL/G/13(95)/2011-12/Pt.V/342/-26) in India. The populations of *A. nigra* were found at an elevation of 100-200m in the periphery of the reserve (Figure 5.3). They are found in marshy regions and are widespread in the reserve.

Figure 5.3: Population of *Alpinia nigra* at Pakke Tiger Reserve, Arunachal Pradesh, India.

Leaf tissues of 51 anaflexistylous and 63 cataflexistylous individuals of *Alpinia nigra* were collected along with one individual for the reference genome (Appendix 5.1). As *Alpinia* and other Zingiberaceous plants reproduce vegetatively by underground rhizomes, the tissues from the plants collected were at least 1 m apart to minimise collecting from the same genet. Samples were collected across a distance of 2 km (Figure 5.4). The tissues were cut into 1 cm^2 pieces and dried in silica gel.

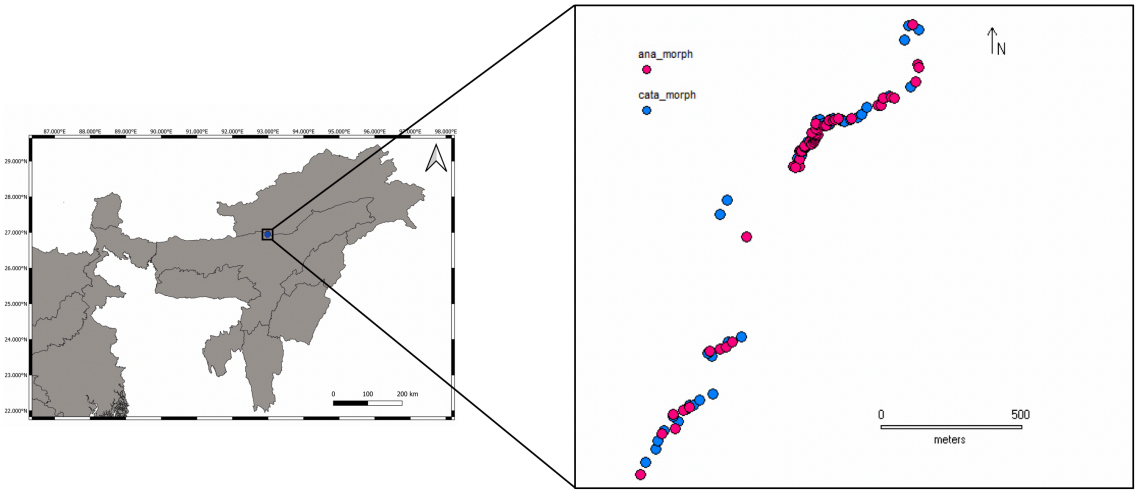

Figure 5.4: Sample collection of the ana- and cata-morphs of *Alpinia nigra* from Pakke Tiger Reserve, Arunachal Pradesh, North-East India (Pink- ana-morph, blue- cata-morph, green- unknown morph). The unknown morphs were not used for Pool-seq.

5.2.3 Reference genome assembly

5.2.3.1 DNA extraction (genome)

DNA was extracted from a silica dried leaf tissue from a cataflexistylous individual collected from Pakke Tiger Reserve using a modified CTAB protocol (Doyle and Doyle, 1987). Eight such extractions were done for leaf tissue from the same individual followed by gel extraction using QIAEX II (Qiagen, Valencia, CA, USA) using the manufacturer's protocol. The DNA from all the samples was pooled and purified using Genomic Tips (Qiagen, Valencia, CA, USA). The sample was then concentrated using the Savant SpeedVac (Thermo Fisher Scientific, USA). An Illumina Tru-Seq Nano gel-free library with a 550bp insert size was prepared from the DNA sample. The genome was sequenced at Edinburgh Genomics (UK) using NovaSeq 6000 (Illumina) with 250bp paired-end run at a 100x coverage.

5.2.3.2 Genome assembly

The quality of the raw reads was checked using FastQC v0.11.7 (Andrews, 2010). The raw reads did not show Illumina adapter contamination. K-mer Analysis Toolkit (KAT) (Mapleson et al., 2016) was used to check the k-mer occurrences in the raw

data using the hist tool. The genome was subsequently assembled using DISCOVAR *de novo* (Weisenfeld et al., 2014) with default parameters. The quality of the assembly was evaluated using QUAST v5.0.2 (Gurevich et al., 2013).

To check for the presence of DNA contamination from bacteria and fungi, I used BlobTools v1.1.1 (Laetsch and Blaxter, 2017). BlobTools helps infer the taxonomy of each contig and compares the GC proportion to the read coverage to check for contigs that have a skewed GC content and coverage along with contaminant DNA. It requires an assembly file, a coverage file, and a hits file. To obtain the coverage file, the genome was indexed and the raw reads were mapped back to the assembled genome using BWA (Burrows-Wheeler Aligner) (Li and Durbin, 2009) to generate a SAM (Sequence Alignment/Map) file. Samtools (Li et al., 2009) was used to covert the SAM file to a BAM (Binary Alignment/Map) file using the 'view' tool followed by sorting and indexing the files. For the hits file, a nucleotide similarity search using blastn (v2.9.0+) was performed against the NCBI nt database and a Diamond blastx (Buchfink et al., 2015) translated nucleotide sequence similarity search against the Uniprot database. The BlobTools module taxify was used to convert the output of the protein similarity searches into a BlobTools compatible input. The assembly file, coverage (BAM) file, and the protein and nucleotide hits file were used to create a blobDB and the contigs were annotated using the parameter '-x bestsumorder' to generate a blobplot. No significant contaminants were found so no sequences were filtered out. The completeness of the assembly was evaluated using BUSCO v3.0.2 (Seppey et al., 2019) with the embryophyta lineage dataset. Smudgeplot (Ranallo-Benavidez et al., 2020) was used to estimate the ploidy and copy number of genomic regions of *Alpinia nigra*.

5.2.4 Pool-seq

5.2.4.1 DNA extraction

DNA was extracted from 51 anaflexistylous and 63 cataflexistylous individuals using the Qiagen DNeasy Plant Mini kit (Qiagen, Valencia, CA, USA) following the manufacturer's protocol. The DNA was quantified using the Qubit dsDNA BR Assay

kit on the Qubit 2.0 Fluorometer. Two equimolar pools of the respective ana- and cata-morph were prepared. Due to DNA degradation, the pools were gel extracted using Zymoclean Large Fragment DNA Recovery kit (Zymo Research, Irvine, CA, USA) and concentrated using the Savant SpeedVac (Thermo Fisher Scientific, USA). The samples were sequenced by Novogene (UK) with 150bp paired-end reads on an Illumina NovaSeq 6000 instrument, aimed at generating 50x sequencing coverage per pool.

5.2.4.2 Genome mapping and analysis

The raw reads were quality checked using FastQC v0.11.8. The reads were aligned to the reference genome using BWA to generate a SAM file. Samtools v1.9 was used to convert the files from SAM to BAM using the option 'view' followed by sorting and indexing the files. PoPoolation2 (Kofler et al., 2011) was used to compare the allele frequency differences between the two pools by following the tutorial (https://sourceforge.net/p/popoolation2/wiki/Tutorial/). Ambiguously mapped reads were removed using Samtools v1.9 using the option view, -q 20 to include reads with mapping quality > 20 and then sorted. Samtools was used to create an mpileup with the BAM files from the two different pools which was then synchronised. F_{ST} values were calculated using a sliding window approach with window sizes 1 kbp and 10 kbp with a step size of 1 kbp for both windows with min-coverage 10, max-coverage 500 and min-count 10. Larger window sizes smooths out the noise but reduces the signal and opposite is the case for smaller window sizes hence different window sizes were used. Fisher's exact test was used to test the significance of differences in the allele frequencies. The F_{ST} and Fisher's exact test files were converted into an IGV (Integrated Genomics Viewer) compatible format. To plot the F_{ST}, the IGV files were converted to be compatible with R using the perl script Genome_R_script.pl (https://github.com/Gammerdinger/Manhattan_plots) and Manhattan plots were generated using R v3.6.3. As chromosomal level information was not available for the genome assembly, the plots were made using only the position data of the genome (i.e. along each contig in turn). To calculate the average coverage per scaffold, I used Samtools v1.9 with the option 'depth' and used the python script avg_deps.py (Becher, pers. comm.).

5.3 Results

5.3.1 *Alpinia nigra* genome assembly

A reference genome of *Alpinia nigra* was generated from a cataflexistylous individual from a wild population using a short-read Illumina platform. Each file (forward and reverse) contained 528,651,368 reads. The pre-assembly quality check indicated that no sequences were of poor quality (Phred score > 30) and there was no adapter contamination. The 27-mer spectrum (Figure 5.5) indicated that the genome is heterozygous as there are two main peaks.

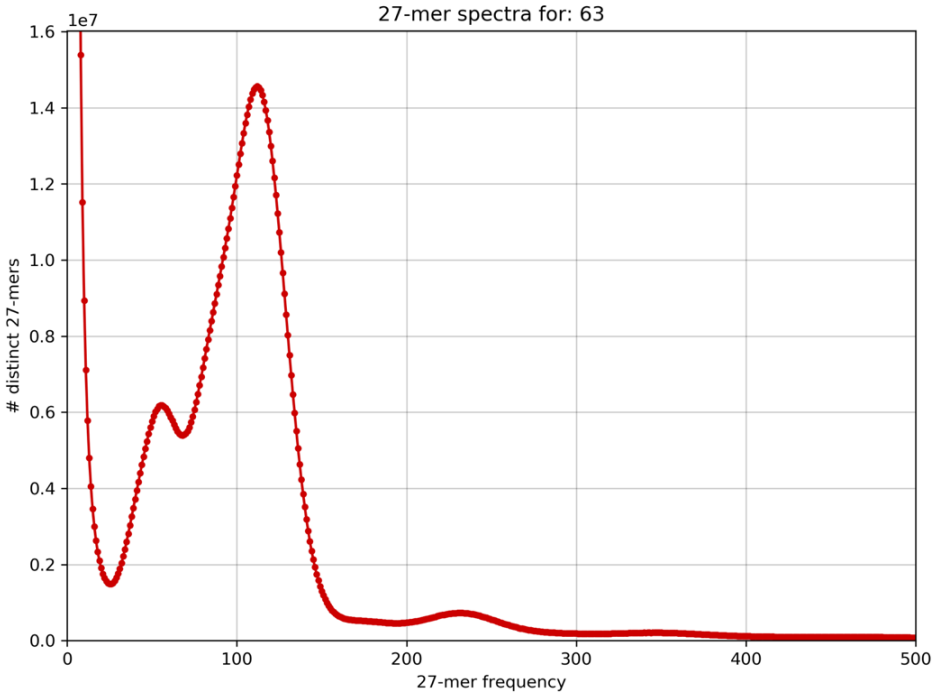

Figure 5.5: A 27-mer distribution plot for *Alpinia nigra*. The presence of two peaks indicates that this genome is heterozygous. The heterozygous peak is at x=56 and the homozygous peak is at x=109.

The peak at x=56 represents the k-mers that are unique in the reference and the peak at x=109 represents the k-mers that are present twice in the genome. The first peak with coverage near 1 represents the unique k-mers of low frequency present due to sequencing errors (Li and Harkess, 2018). The smaller peak at x= 222 indicates that

there are duplicates present. As expected for an outcrossing species, the k-mer spectra shows moderate heterozygosity, estimated as 1.06%.

Discovar *de novo* produces an assembly where polymorphisms are represented by bubbles. These bubbles usually contain heterozygous sites or somatic mutations, which are then flattened. The final assembly file contains these flattened lines, that produced a total of 2.07Gbp in 1 kb+ contigs with a read coverage of 127.5 based on the DISCOVAR *de novo* assembly statistics.

Table 5.1: Assembly metrics for the *Alpinia nigra* genome assembled using DISCOVAR *de novo*. The assembly metrics were estimated using QUAST v5.0.2 (Gurevich et al., 2013) and the completeness of the assembly was estimated with BUSCO v3.0.2 using the embryophyta_odb10 BUSCO set. *C= complete [S= single copy, D=duplicated], F=fragmented, M=missing, n=number of orthologues in comparison.

Total length (>= 0 bp) (Mb)	2389
Total length (>= 1000 bp) (Mb)	2119
No. of contigs(>= 0 bp)	1072070
No. of contigs (>= 1000 bp)	158389
N50 Length (Kb)	48.9
Longest contig (Kb)	591.2
BUSCO score*	C:91.1%[S:84.1%,D:7.0%],F:4.9%,M:4.0%,n:1375

Assembly metrics were also produced using QUAST that are listed in Table 5.2. Many contigs smaller than 1000bp were present. The complete (single and duplicated) BUSCOs were over 91% which implies that the assembly completeness is high. N50 is an important metric for assembly contiguity and is dependent on the type of data used. As the *Alpinia* genome assembly is based on short reads that cannot resolve long repeat regions, the scaffold N50 is 48.9 kbp. Sequence filtering was not required as the blobplot showed very low contamination (Figure 5.6).

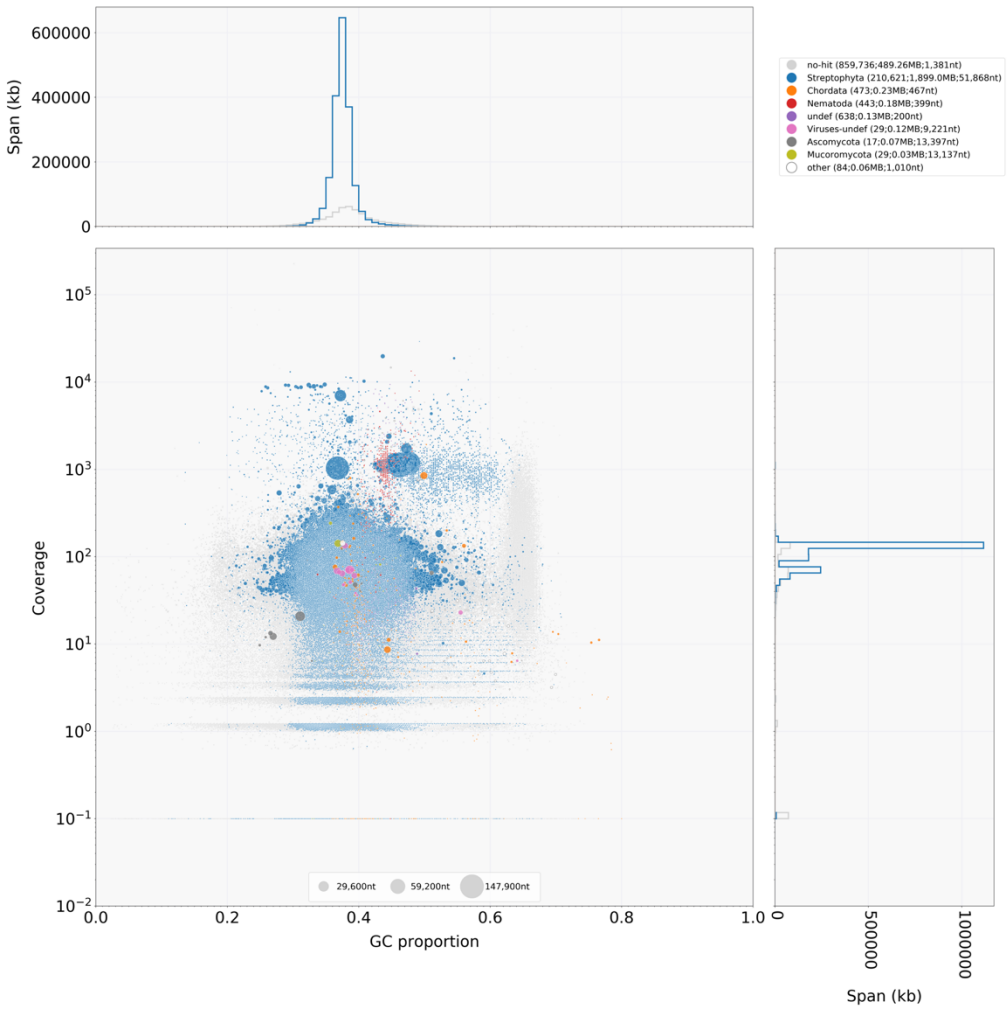

Figure 5.6: Blobplot for the *Alpinia nigra* reads using Blobtools v1.1.1. This plot was produced using the hits obtained from blastn and Diamond hits of NCBI nucleotide or UniProt proteome databases. These hits were used to assign taxonomy (using bestsumorder). 97.69% of the reads mapped back to the assembly out of which 84.79% belonged to Streptophyta and 12.73% were no-hits. The values in the brackets of the legend represent count, sum length and n50.

I used a graphical representation of heterozygous k-mer pairs, implemented in Smudgeplots, to show the haplotype structure and estimate ploidy (Figure 5.7). As 84% of the k-mer pairs are consistent with a diploid haplotype structure (i.e. with a 1:1 ratio of the two haplotypes), this species is confirmed as diploid. Triploid and tetraploid haplotype structures are also present albeit in extremely low proportions. These smudges match the peaks in the k-mer spectrum (Figure 5.5).

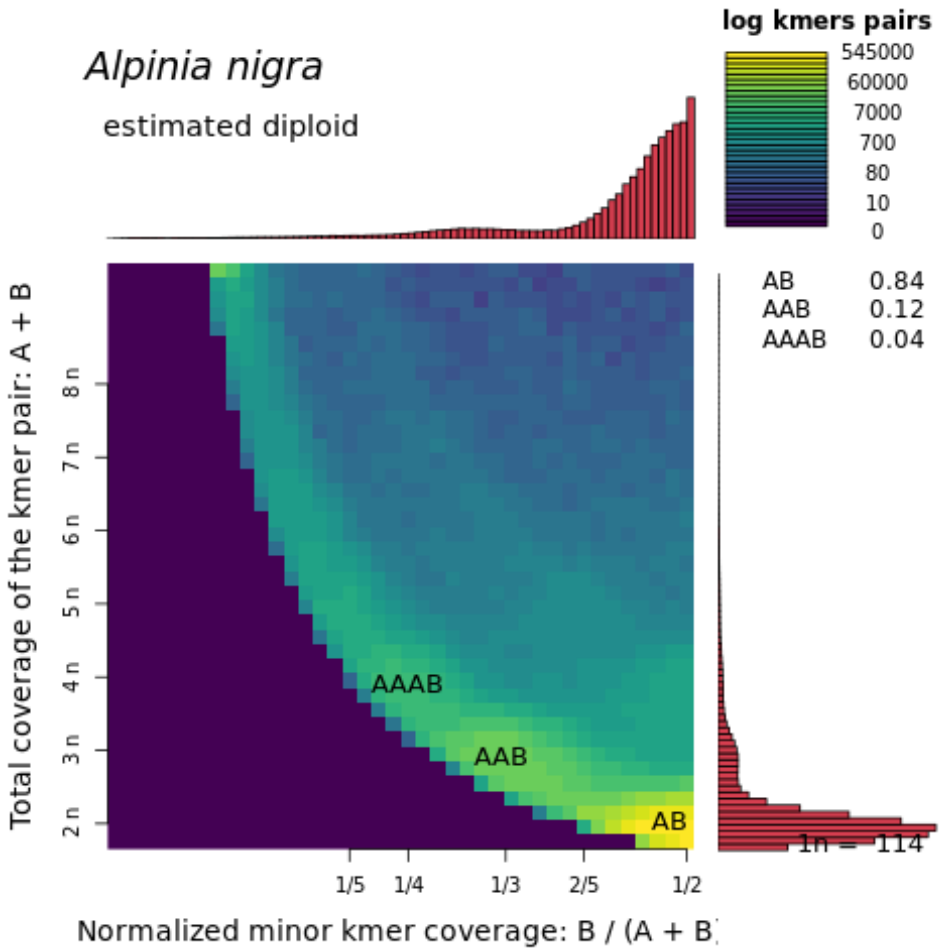

Figure 5.7: Smudgeplot for *Alpinia nigra* for k=21. This tool estimates the ploidy and monoploid coverage. The heat of the smudge indicates how frequently the haplotype structure is represented in the genome. The bar plots represent the sequencing coverage.

5.3.2 Pool-seq

Two pools of anaflexistylous (51 individuals) and cataflexistylous (63 individuals) morphs from a single wild population (same as the reference genome) were sequenced using Illumina short reads. There were no obvious coverage differences between scaffolds of anaflexistylous and cataflexistylous pools for the first 200,000 scaffolds (Appendix 5.2). About 92% of reads from anaflexistylous pool and 90% of reads from cataflexistylous pool mapped back to the reference genome. As recommended in the Popoolation2 tutorial, the ambiguously mapped reads were removed, leaving only reads mapping to a single location for downstream analysis.

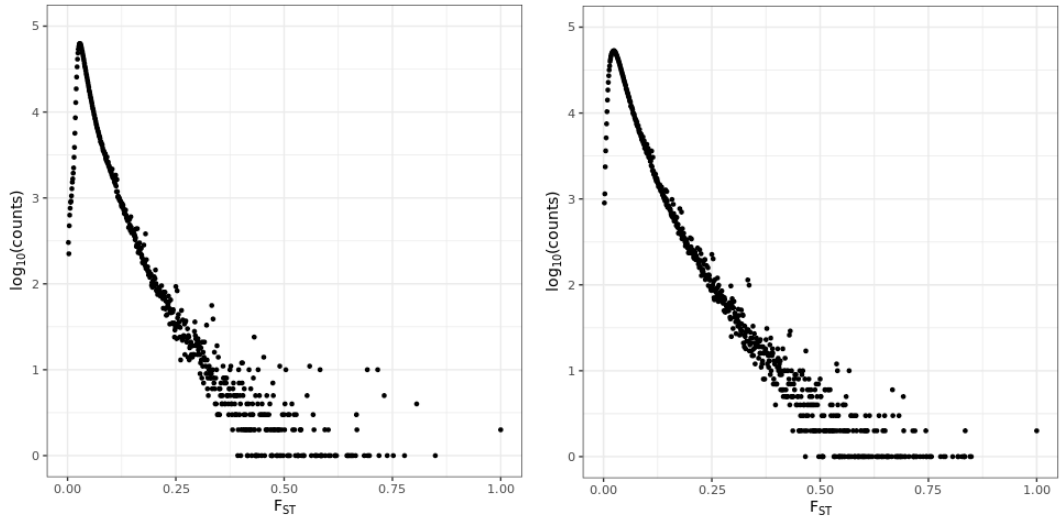

Figure 5.8: F_{ST} values for window-size 10 kbp (left) and 1 kbp (right) with step size 1 kbp from comparison of the pools of the anaflexistylous and cataflexistylous morphs of *Alpinia nigra* generated from Popoolation2 (Kofler et al., 2011). These plots only include non-zero F_{ST} values.

F_{ST} values from the Popoolation2 analysis were used to find regions of allele frequency differences between the two morphs. The distribution of F_{ST} values for the two window-sizes are represented in Figure 5.8, and these show most F_{ST} values are below 0.2. The mean genome-wide F_{ST} for window size 1 kbp was 0.0435 and for 10 kbp was 0.0432. The number of windows greater than F_{ST} >0. 5 and >0.8 was 312 and 13 for window-size 1 kbp and 152 and 7 for 10 kbp respectively.

To visualise the F_{ST} values across the scaffolds, a sliding window analysis was done for window size 10 kbp (Figure 5.9A). As the scaffolds are ordered by length in a Discovar *de novo* assembly starting from the largest, a large accumulation of F_{ST} values was present at the tail end of the genome as the reads don't map so well to the short scaffolds. Overall, no specific region stood out in terms of densely packed high F_{ST} values. Therefore, I decided to zoom in on regions of 500Mbp length to visually check if any moderate genomic differences could be detected (Figure 5.9B-F). Most of the genome shows low differentiation and there is no evidence for clustering of any outliers, which would be indicative of a large inversion. Similar plots were generated for window-size 1 kbp but no outliers were seen for this window size either (Figure 5.10).

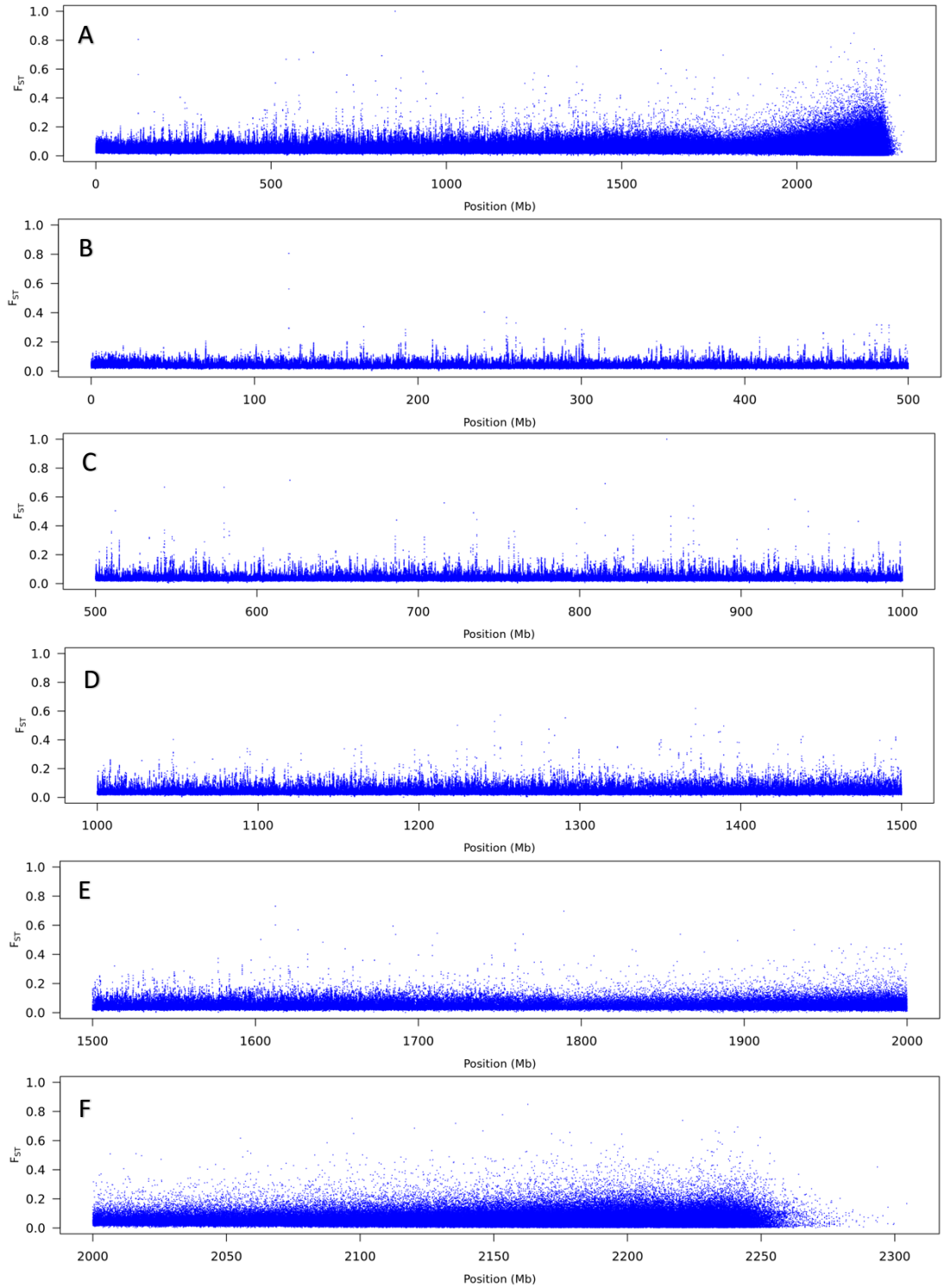

Figure 5.9: F_{ST} summary for window-size 10 kbp and step size 1 kbp from comparison of the pools of the anaflexistylous and cataflexistylous individuals of Alpinia nigra. A- Genome-wide F_{ST} scan, B- scan across the first 500Mbp , C – scan from 500-1000Mbp, D- scan from 1000-1500Mbp, E- scan from 1500-2000Mbp, F- scan from 2000Mbp-end. The scaffolds are ordered by length starting from the largest.

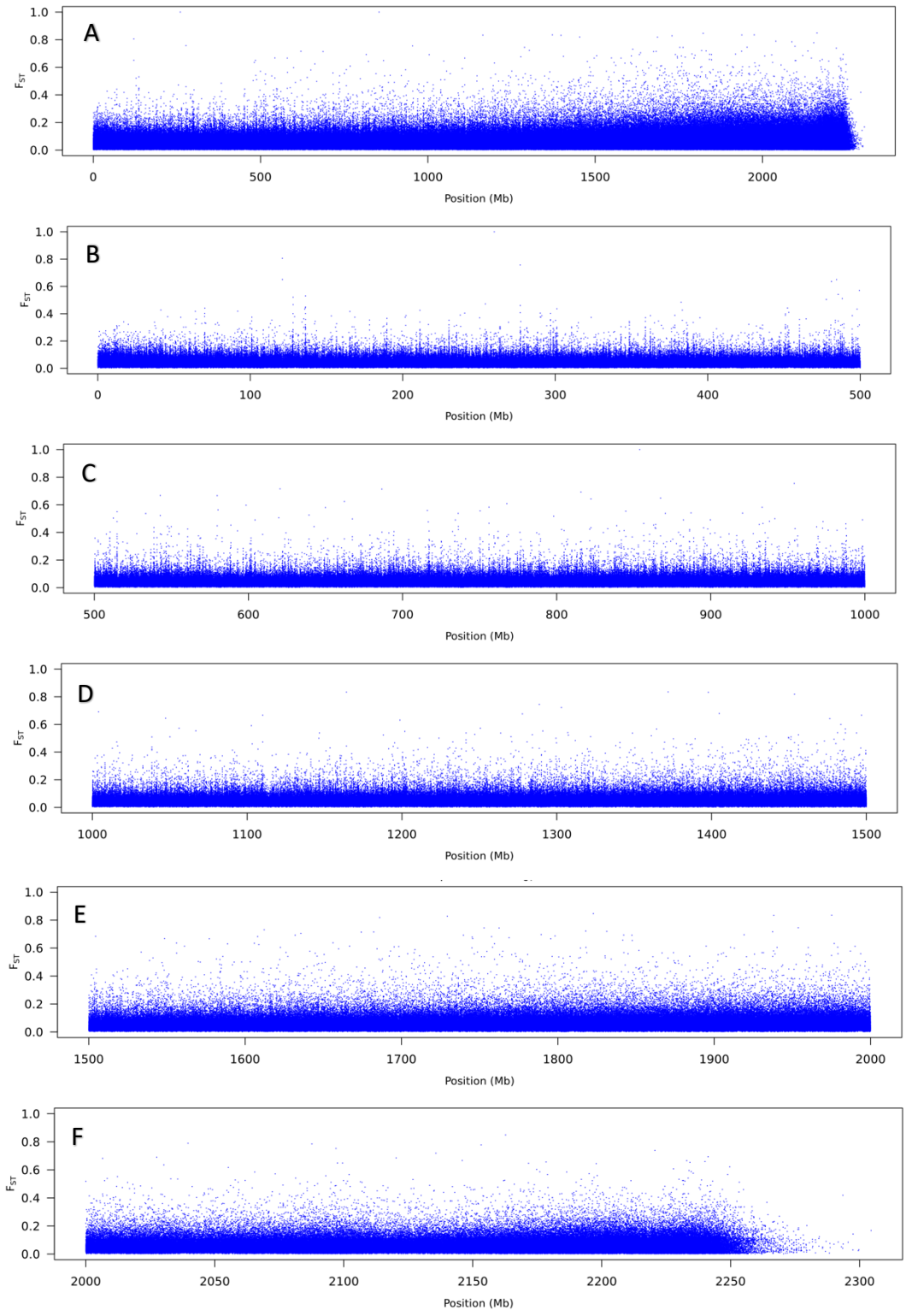

Figure 5.10: F_{ST} summary for window-size 1 kbp and step size 1 kbp from comparison of the pools of the anaflexistylous and cataflexistylous individuals of *Alpinia nigra*. A- Genome-wide F_{ST} scan, B- scan across the first 500Mbp , C – scan from 500-1000Mbp, D- scan from 1000-1500Mbp, E- scan from 1500-2000Mbp, F- scan from 2000Mbp-end. The scaffolds are ordered by length starting from the largest.

5.4 Discussion

5.4.1 Draft genome assembly of *Alpinia nigra*

In this chapter, I have presented a draft genome sequence of *Alpinia nigra*. The assembly metrics indicate that this genome is of high quality. It has a high contiguity (N50 = 48 kbp) for a genome assembled using short-reads that is comparable with other *de novo* assemblies such as those of six heterozygous *Begonia* L. species (N50 ranging from 11.71-58.63 kb) (Campos-Domínguez et al., unpublished), or of *Anopheles arabiensis* (N50=22.3 kb) (Love et al., 2016). It is likely to have a good representation of genic regions (91% BUSCO completeness) and is diploid with a genome size in the range of 2-2.3 Gbp. Within the Zingiberales, only the genomes of *Curcuma longa* (Chakraborty et al., 2020) and banana species such as *Musa acuminata* , *M. balbisiana* , *M. itinerans*, and *M. schizocarpa* have been sequenced so far (D'Hont et al., 2012; Davey et al., 2013; Wu et al., 2016; Belser et al., 2018). As banana is one of the most economically important fruits in the world, these genomes have served as an invaluable resource for breeding strategies for new cultivars with improved yield (Wu et al., 2016). The availability of these high quality genomes has also provided an insight into the evolution of monocot lineages (D'Hont et al., 2012).

Similarly, many species in the Zingiberaceae are of agricultural importance such as ginger (*Zingiber officinale*), cardamom (*Elettaria cardamomum*) and turmeric (*Curcuma longa*). Therefore, availability of a reference genome from the ginger family can help in many downstream analyses such as improving yield, disease resistance, investigating the potential of wild relatives (Edwards and Batley, 2010), and investigating the biosynthetic pathways of essential compounds (Lichman et al., 2020). The draft genome presented here will also be an important resource for ecological, systematic and evolutionary studies in the Zingiberaceae.

5.4.2 The genetic basis and evolution of flexistyly

Here, I have used a Pool-Seq approach to investigate the allele frequency differences between the ana- and cataflexistylous individuals. Pool-seq has been used to identify the genes underlying many traits such as the dwarfism gene in watermelon (Dong et al., 2018), to identify genomic regions that are involved in the maintenance of ecotypic variation in *Mimulus guttatus* (Gould et al., 2017) and to compare genomic diversity in populations of non-model species such as brown trout (Kurland et al., 2019). It is a useful approach for studies that involve detecting the genetic basis of traits with different phenotypes in a cost effective manner (Schlötterer et al., 2014).

Flexistyly is a complex phenotype so it is most likely to be controlled by multiple loci. These loci are presumably tightly linked as the morph types are always found in a 1:1 ratio within a population (Cui et al., 1996; Li et al., 2001, 2002; Zhang et al., 2003; Ren et al., 2007; Jia et al., 2015). Therefore, it would be predicted that this loci might reside in a region of reduced combination so that they are inherited together, such as a large polymorphic chromosomal inversion or a supergene (Twyford and Friedman, 2015; Charlesworth, 2016; Jay et al., 2018). The genome-wide sliding window F_{ST} scans do not reveal any significant allele frequency differences against the background of low differentiation for most of the genome. There is no clear evidence for a large chromosomal inversion here due to the lack of a large region containing F_{ST} outliers. This lack of a clear genetic signature between the pools could be due to a number of reasons. This trait might be governed by a single small genomic region or a single gene that might be difficult to detect. While I did find a handful of nearly fixed SNP differences between pools, these SNPs appear in isolation, rather than being clustered as would be expected from highly divergent alleles. It is also possible that the region governing this trait belongs to a complex genomic region, which is difficult to assemble and thus not represented in the reference genome. Genome assemblies that use only short read data are often fragmented and lack large contiguous genomic regions (Thomma et al., 2016). There are many short scaffolds (<=1000bp) in the assembly presented here (>913,000). The difficulty in assembling a genome is also exacerbated by the presence of repeats which cause breaks in an assembly. If there is a chromosomal inversion present, it

might be hard to assemble as active transposable elements are often found at the breakpoints (Delprat et al., 2009). As was seen in *Primula* spp., there was a notable coverage difference in the thrum specific region as compared to the pin indicating hemizygosity of the region (Li, Cocker, et al., 2016; Cocker et al., 2018) but this pattern was not seen upon comparing the read depths of the anaflexistylous and cataflexistylous pools.

An alternative route to pooled sequencing for finding the gene(s) underlying stylar movement would be to perform comparative transcriptomics, particularly tissue-specific sequencing of stylar tissue at different time intervals from a few hours prior to anthesis to the first and second curvature. This can reveal the genes that are differentially expressed between the morphs. Utilising long-read sequencing such as PacBio HiFi, combined with long-range scaffolding with technologies such as HiC or BioNano genome mapping, could be used in the future to improve genome contiguity and assembly. This can aid in detecting structural variants between the two morphs.

Understanding the phenotype in more detail can give clues to the genes that underlie flexistyly. All the flowers of the same morph type that open on the same day are synchronous (Li et al., 2001, 2002; Zhang et al., 2003) and the speed of stylar movement depends on the weather (Li et al., 2001). Studies have revealed that this stylar movement is dependent on factors such as light and auxin transport. When exposed to light before anthesis, the style of the anaflexistylous morph curves above the anther instead of downwards. The second curvature was not affected by light in either morph type hence both the curvatures of cata-morph and the second curvature of the ana-morph are controlled endogenously (Luo and Li, 2010; Su et al., 2017). The exogenous IAA (indole-3-acetic acid) treatment increased the stylar curvature in light but not in the dark for both curvatures in both the morph types. Upon treatment with auxin efflux inhibitors, the first curvature was not affected but the second movement of ana-morph was enhanced while the second curvature of the cata-morph did not occur (Luo et al., 2012; Su et al., 2017). In *Alpinia galanga*, neither IAA nor NPA (-N -naphthylphthalamic acid, an auxin efflux inhibitor) had any noticeable

effects on style bending (Su et al., 2017). It is presumed that light might activate auxin transport that promotes stylar movement in *Alpinia* and the effects of these factors are different for different morph types (Luo et al., 2012). Therefore, an annotated genome would be useful to see if any SNP differences between morphs reside in loci involved in light regulation.

There are many reasons as to why this novel mechanism may have evolved in *Alpinia*. *Alpinia* flowers are hermaphroditic and are open only for a day compared to other taxa with similar zygomorphic structure, such as orchids, where the flowers are open for days or even weeks. The risk of selfing is likely to be higher in gingers in which pollination must be achieved in a much shorter period than in most orchids. Therefore, this stylar movement has evolved to ensure cross-pollination and prevent sexual interference within the flower (Zhang et al., 2003; Sun et al., 2007, 2011). As flexistyly has been reported from many genera within Alpinieae, this dimorphism may have evolved in the common ancestor of the tribe or may have evolved independently several times (Kress et al., 2005). As more evolutionary research on the genomic basis of flexistyly is performed, the origins of this trait will become clear.

Chapter 6

General discussion

6.1 Thesis overview

The main aim of this thesis was to recognise the taxonomic challenges present in tropical plants and investigate various aspects of a taxonomically complex tropical group. The main outcomes were that I resolved nomenclatural problems associated with a group of species, investigated the reproductive barriers maintaining species differences, and provided initial characterisation of the genetic basis of a floral trait, all in the genus *Alpinia*.

In chapter 2, I identified the taxonomic challenges that plant biologists face when working on tropical taxa. Based on the survey and literature search, I found that both practical issues and biological factors hinder taxonomic research in the tropics. Practical issues entail various aspects of fieldwork and herbaria-related problems such as inaccessibility of sites, undercollection of specimens, lack of taxonomic expertise, and incomplete herbarium specimens. Biological factors include processes such as rapid radiations, hybridisation, and phenotypic plasticity obscuring species boundaries. These issues have a significant impact when estimating species diversity or making conservation assessments.

In chapter 3, I resolved the nomenclatural confusion associated with the name *Alpinia bracteata* used by both Roscoe and Roxburgh in the early 19th century. I found that the original illustration of *Alpinia bracteata* Rosc. was badly drawn so this name should be rejected. I found that the relative usage of *A. bracteata* Roxb. is low hence its synonym *A. roxburghii* Sweet should be used. I also assigned a lectotype for *A. calcarata* (Haw.) Rosc., a species commonly found in India.

In chapter 4, I investigated the reproductive barriers between and within the different *Alpinia* clades and a few other closely related genera. On carrying out artificial cross-pollination between species, I found that *Alpinia* species show a potential for hybridisation within clades, and in a few cases, between clades and genera as well. Overall, the self-pollinated crosses were the most successful and the seed set decreased with increasing genetic distance.

Finally, in chapter 5, I presented the draft genome of *Alpinia nigra*, a species widespread in Northeast India, and used a Pool-seq approach to investigate the genetic basis of flexistyly. The genome assembly presented here has a high contiguity and completeness and can serve as a reference genome for downstream analyses. On comparing the pools of the cataflexistylous and anaflexistylous morphs, I found no clear genetic difference and very few windows had F_{ST} values greater than 0.8. I demonstrated the utility of the draft genome presented here with its use to investigate the genetic basis of a novel floral trait in a non-model plant.

In light of these results, I will discuss this work in the wider context of tropical plant research. In particular, I highlight the problems working on a non-model tropical plant group. I also highlight the importance of this research in genomics, taxonomy and conservation. Lastly, I discuss proposed future directions of research for evolutionary and taxonomic studies of *Alpinia*.

6.2 Taxonomic complexity is pervasive in tropical plant groups

The problems encountered when working with *Alpinia* such as polyphyly, difficulty assigning names to specimens, problematic classification on the basis of morphological characters, and issues with generic delimitation are common in other tropical genera as well. For example, *Myrcia* is a species-rich tree genus (~770 spp.) found in the Neotropics (Amorim et al., 2019), which until recently was recognized as four distinct genera. These four genera were classically distinguished on the basis of floral characters but molecular phylogenetic studies showed them to be para- or polyphyletic (Lucas et al., 2007, 2011). *Myrcia* and several other large genera in the Myrtaceae such as *Eugenia*, *Eucalyptus*, and *Syzygium* are taxonomically complex as they exhibit convergent as well as parallel evolution of floral traits (Vasconcelos et al., 2015, 2017). *Myrcia* is now divided into nine sections and each of the sections can be recognised on the basis of morphology (Lucas et al., 2018). Thorough integrative systematic analyses have helped in resolving the complex relationships in this group and in identifying threatened species that couldn't be assessed previously

due to data deficiency (Lucas et al., 2018; Amorim et al., 2019; Nic Lughadha et al., 2019).

Similar patterns are also seen in *Echinopsis* (100-150 spp.), one of the largest genera in Cactaceae, which was found to be polyphyletic (Schlumpberger and Renner, 2012). The genus, in its broad circumscription, comprises 18 genera but to ease the problematic generic boundaries, all these were lumped (Arakaki et al., 2011). Based on the molecular results, there is a need to divide the genus into smaller units, or a larger one including a few other closely related genera. Overall this circumscription requires the use of more stable traits (Schlumpberger and Renner, 2012). Genera such as *Miconia* (Melastomataceae, ~1000spp.) (Goldenberg et al., 2008), *Polyalthia* (Annonaceae, ~150 spp) (Xue et al., 2012) are also highly polyphyletic and classification of taxa is difficult due to lack of obvious synapomorphies or parallel evolution. Large scale phylogenetic studies have also revealed the polyphyletic nature of smaller genera such as *Campylospermum* and *Sauvagesia* in the Ochnaceae highlighting the need for extensive taxon sampling (Schneider et al., 2020). As no large scale phylogenetic analysis of *Alpinia* s.l. has been produced in the last 15 years (Kress et al., 2005), it would be extremely informative to perform more taxon sampling and detailed morphological studies in the light of phylogenomic analyses, to distinguish genera and species and find more informative characters.

6.3 Challenges of genomic analyses in a non-model plant

Advances in sequencing technologies have considerably improved the quality of *de novo* genome assemblies (Kersey, 2019; Michael and VanBuren, 2020). The low costs of sequencing, improved bioinformatic algorithms, and the low cost of computation has made it possible to sequence and assemble many genomes of non-model organisms (Schatz et al., 2012). The focus of plant genomics has traditionally been on model species where many high quality, chromosome-level assemblies have been produced using long sequence reads (e.g., *Zea mays*- Jiao et al., 2017; *Arabidopsis thaliana*- Michael et al., 2018). The genomes of these model species have often used highly-inbred homozygous lines that makes them easier to assemble.

Assembling non-model plant genomes can be difficult as they tend to be large, can have high ploidy levels, and are often repeat-rich and heterozygous (Claros et al., 2012; Michael and VanBuren, 2020). Although most *Alpinia* species are self-compatible, generating a homozygous line is difficult as they take about 2-3 years from seed to flowering (Ranavat, pers. obs.). This long generation time was also an issue for investigating the hybrid viability and fertility from the artificial pollinations that I carried out in Chapter 4.

The silica-dried leaf tissue for the *Alpinia* species I used for DNA extraction had several issues for whole genome sequencing. I attempted high molecular weight DNA extraction using various CTAB (cetyltrimethylammonium bromide) based protocols (Fukushima et al., 2017; Harkess, 2017; Hulse-Kemp et al., 2018) that were used for other plant DNA extractions but proved unsuccessful as the samples always had degraded DNA with a low concentration. Therefore, with this degraded DNA I was limited to short-read sequencing. However, the use of long-read sequencing, either on its own or in conjunction with short-read sequences, often leads to more contiguous genome assemblies (e.g., PacBio + Illumina- *Camellia sinensis*, scaffold N50- 1.39Mb (Xia et al., 2019)). Although *Alpinia* accessions successfully amplified by PCR, the Sanger sequencing of nrITS produced low quality sequences for most accessions. *Alpinia* species are rich in secondary metabolites such as terpenoids, phenols, alkaloids and essential oils (reviewed in Ma et al., 2017) that tend to inhibit DNA extraction, amplification, restriction digestion, and sequencing. These compounds can interfere with enzymatic reactions and their oxidising activities can reduce the yield of DNA extracted (Khanuja et al., 1999). As leaf material is rich in secondary metabolites (Healey et al., 2014), using a different part of the plant for extraction might be useful. It is also possible that because the silica-dried material was used between 3-10 months after collection the DNA may have degraded over this period of time (Guo et al., 2018). However, even the freshly silica-dried material showed some degradation. Using methods such as CTAB based extraction can avoid this problem but often yield very low quantities of DNA (Healey et al., 2014). Within *Alpinia*, species may have different concentrations or

compositions of metabolites, which may be why some Sanger sequencing was unsuccessful.

6.4 *Alpinia nigra* genome as a resource

Alpinia nigra is a diploid species with a relatively modest genome size (2.3 Gbp) in comparison to many other angiosperms, which have genome sizes ranging from 61Mb -149Gb (Pellicer et al., 2018). I have presented a good quality draft genome of *Alpinia nigra* using short-read data (N50- 48 kb, BUSCO score- 91.1%), which is comparable to the genome of the polyploid species *Curcuma longa* (turmeric) that has been sequenced recently using 10x Genomics linked reads (N50- 18.8 kb, Complete BUSCOs- 81.8%) (Chakraborty et al., 2020). This assembly can serve as a reference genome that can be used in a wide range of research areas, from conservation genomics to broadscale-phylogenomics (Kersey, 2019). Once the genome is annotated, it could also help in understanding the biosynthetic pathways underlying the production of secondary metabolites (Zhao et al., 2013) . These chemicals not only play an important role in plant defence and attracting pollinators but also in pharmacology, and many studies are under way to characterise and investigate their biological effects. This genome can also play an important role in comparative studies of metabolite production across species in the Zingiberaceae.

As an application of the use of this draft genome as a reference, I have investigated the genetic basis of flexistyly, a novel floral trait in the Alpinioideae. In comparison, the most well-studied stylar polymorphism, heterostyly, has been studied for the past 150 years but its genetic basis (presence of hemizygosity in the pin morph) was only discovered in the last decade (Ushijima et al., 2012; Yasui et al., 2012). Indeed, researchers are still discovering the various genes at the *S* morph locus and their function (Huu et al., 2020). The foundation of this heterostyly research is strengthened by researchers working on the genomics not only in *Primula* (Huu et al., 2016; Li, Cocker, et al., 2016; Cocker et al., 2018) but also other genera such as *Fagopyrum* (Yasui et al., 2012), *Linum* (Ushijima et al., 2012) and *Turnera* (Shore et al., 2019). The availability of high quality, homozygous genomes has also aided in

unravelling the genetic basis of this trait. Therefore, finding the genes underlying a complex trait such as flexistyly will not be straightforward and will take many years. My work has ruled out a simple genetic basis such as a large chromosomal inversion, but more research is required to pin down the gene(s). A similar route adopted for the study of heterostyly could be used to investigate flexistyly. This may involve further development of genomic resources such as genome annotations, transcriptomes, high coverage genome sequencing of the anaflexistylous morph (complementing the cataflexistylous morph that was sequenced in this study), and widening these analyses to other flexistylous genera such as *Amomum* s.l. , *Etlingera*, *Paramomum*, *Plagiostachys*, and *Siliquamomum* (Cui et al., 1996; Kress et al., 2005; Leong-Škorničková et al., 2014; Jia et al., 2015; Poulsen and Phonsena, 2017).

Su et al., (2017) found that there are a few differences in the mechanism of stylar movement in different *Alpinia* species. When they compared the effects of auxin (Indole-3-acetic acid) enhanced stylar movement and auxin transport inhibitors (1-N -naphthylphthalamic acid) they found that they inhibited the second bending of the cataflexistylous morph in *A. oxyphylla* (clade IV) but this had no obvious effect on *A. galanga* (clade II) (Su et al., 2017). Other clade IV species where this was studied also showed similar responses to auxin and its transport inhibitors (Luo et al., 2012). My hypothesis is that *Alpinia* species from different clades may be affected by chemicals other than IAA that are involved in plant movement, but this needs to be examined across clades and in other flexistylous genera. therefore there is likely to be a different genetic basis in these other groups. The effect of light was similar across all species (Luo and Li, 2010; Luo et al., 2012; Su et al., 2017) therefore the downstream effect of light might be variable for different species.

6.5 Taxonomic challenges of *Alpinia* in India

The number of species of *Alpinia* s.l. in India is contentious as no comprehensive taxonomic study has been carried out for at least 19 years, especially from section *Alpinia* subsection *Catimbium* (sensu Smith, 1990). Based on my fieldwork experience, in the future many new species will be added. On visiting herbaria in

India (ARUN, ASSAM, KFRI, MH, TBGT) and the UK (BM, E, K, LINN and LIV), I found that many specimens from India were wrongly determined. This is a consequence of the species that were described in the early 19th century, where the original material left to study is not ideal due to poorly preserved characters, and the species descriptions are often ambiguous. I have corrected a few of the names, but more effort is required to clarify these problems with nomenclature and species delimitations. This would require careful examination of herbarium specimens, live collections in botanic gardens, and extensive fieldwork. The species collected during my fieldwork in the Western Ghats and the North-East states in India (summer 2017 and 2018) and a few accessions in the RBGE research collections could not be assigned names based on the taxonomic key available (Wu and Larsen, 2000). This highlights the need for an increased collection effort across the geographic range of species and use of molecular tools for species delimitation. Tools such as target capture have been used to resolve species relationships (Hlavatá et al., in press), and could be applied here, too.

As reproductive isolation is critical for speciation, understanding the presence of it is essential. Investigations of reproductive isolation is largely neglected in large tropical groups (but see Pinheiro et al., 2016). I carried out genus-wide crossing experiments (chapter 4) which was a first for *Alpinia*. Carrying out this large scale crossing experiments for this tropical genus with a wide distribution was possible only because of the large collection of live plants at the Royal Botanic Garden Edinburgh. This allowed me to test the range of potential interactions across the genus, and in the future such crossing results could be linked to geographical and ecological isolation of species in the wild. Moreover, the large number of accessions available gave sufficient replication for statistical analyses. This may not be possible for other tropical taxa, but smaller scale studies can be carried out where possible.

6.6 Implications for conservation

As many species are now threatened due to habitat loss and fragmentation, there is an urgent need to conserve them (Corlett, 2016). Nic Lughadha et al. (2020) estimated that about 39.4% of all vascular plants are threatened and the main drivers of biodiversity loss are changes in land and sea use, climate change, and exploitation of organisms. Geographical biases were also visible with tropical Asia having a gap in assessments on the Red List compared to Africa. The average number of species being described every year is about 364 from Southeast Asia (excluding India, Bangladesh and Bhutan) in addition to the 50,000 that have been recorded (Raven et al., 2020). The Environmental Information System (Botanical Survey of India) reports that there are over 18,000 angiosperms that include over 4,000 endemic species in India (Floral Statistics of India 2018, BSI ENVIS). But the Red List assessment coverage of India is low whereas the proportion of species threatened is relatively high (Nic Lughadha et al., 2020). Therefore, the need to explore, collect, and document species is vital in India and tropical Asia in general.

In Indian *Alpinia*, few species have been assessed such as *A. conchigera* and *A. nigra* (least concern), *A. manii* (vulnerable), and *A. malaccensis* (data deficient) (IUCN 2020). Several species in the Zingiberaceae that are now being described are data deficient (Hareesh and Sabu, 2018; Ashokan and Gowda, 2019). The identity of many widespread species such as *A. malaccensis* need to be clarified as the original description was vague and this has caused erroneous application of this name for what are possibly different species (Lim, 2016). Therefore more effort is needed to clarify names and accurately circumscribe species, prior to further conservation assessments. Increased collection effort coupled with detailed morphological studies are critical for species delimitation in Indian *Alpinia*, which will be part of a broader investigation of the problematic subsection *Catimbium* throughout its range from Sri Lanka to New Guinea (Smith, 1990). Most of the species from the *Catimbium* clade that I examined throughout the course of my PhD could not be delimited as there is a lack of a comprehensive taxonomic key. There may be a few new species from this section that need to be described.

6.7 Concluding remarks

This thesis has highlighted the taxonomic challenges in the tropics and emphasised the importance of carrying out taxonomic, ecological, and genomic studies in parallel for understanding the evolutionary history of tropical taxa. There is a big gap in biodiversity of tropical taxa that needs filling (Pimm and Joppa, 2015; Corlett, 2016). The conservation status of species cannot be assessed if they have not been described. This is a common problem in species-rich regions in the tropics, where there is a big conservation gap (Corlett, 2016), and many Zingiberaceae species are found in these regions (Larsen et al., 1998). Therefore, a concentrated taxonomic effort to collect more species for herbarium and live collections, *ex situ* conservation, along with utilisation of genomic resources to protect plant species is essential.

Bibliography

Abbott, R., D. Albach, S. Ansell, J. W. Arntzen, S. J. E. Baird, N. Bierne, J. Boughman, et al. 2013. Hybridization and speciation. *Journal of Evolutionary Biology* 26: 229–246.

Ahrends, A., C. Rahbek, M. T. Bulling, N. D. Burgess, P. J. Platts, J. C. Lovett, V. W. Kindemba, et al. 2011. Conservation and the botanist effect. *Biological Conservation* 144: 131–140.

Alix, K., P. R. Gérard, T. Schwarzacher, and J. S. (Pat) Heslop-Harrison. 2017. Polyploidy and interspecific hybridization: partners for adaptation, speciation and evolution in plants. *Annals of Botany* 120: 183–194.

Amorim, B. S., T. N. C. Vasconcelos, G. Souza, M. Alves, A. Antonelli, and E. Lucas. 2019. Advanced understanding of phylogenetic relationships, morphological evolution and biogeographic history of the mega-diverse plant genus *Myrcia* and its relatives (Myrtaceae: Myrteae). *Molecular Phylogenetics and Evolution* 138: 65–88.

Andermann, T., A. M. Fernandes, U. Olsson, M. Töpel, B. Pfeil, B. Oxelman, A. Aleixo, et al. 2018. Allele phasing greatly improves the phylogenetic utility of ultraconserved elements. *Systematic Biology*.

Andrews, S. 2010. FastQC: a quality control tool for high throughput sequence data.

Anon. IUCN 2020. The IUCN Red List of Threatened Species. Version 2020-3. Website https://www.iucnredlist.org [accessed 29 January 2021].

Anon. 2013. The Plant List version 1.1. *Published on the Internet*. Website http://www.theplantlist.org/ [accessed 1 January 2021].

Antonelli, A., and I. Sanmartín. 2011. Why are there so many plant species in the Neotropics? *Taxon* 60: 403–414.

Arakaki, M., P.-A. Christin, R. Nyffeler, A. Lendel, U. Eggli, R. M. Ogburn, E. Spriggs, et al. 2011. Contemporaneous and recent radiations of the world's major succulent plant lineages. *Proceedings of the National Academy of Sciences* 108: 8379–8384.

Archibald, J. K., M. E. Mort, D. J. Crawford, and J. K. Kelly. 2005. Life history affects the evolution of reproductive isolation among species of *Coreopsis*

(Asteraceae). *Evolution* 59: 2362.

Ashman, T.-L. 2007. The evolution of separate sexes: a focus on the ecological context. *In* L. D. Harder, and S. C. H. Barrett [eds.], Ecology and Evolution of Flowers, 204–222. Oxford University Press.

Ashokan, A., and V. Gowda. 2018. Describing terminologies and discussing records: More discoveries of facultative vivipary in the genus *Hedychium* J.Koenig (Zingiberaceae) from Northeast India. *PhytoKeys* 96.

Ashokan, A., and V. Gowda. 2019. *Hedychium ziroense* (Zingiberaceae), a new species of ginger lily from Northeast India. *PhytoKeys* 117: 73–84.

Aswani, K., and M. Sabu. 2014. Reproductive biology of *Alpinia mutica* Roxb. (Zingiberaceae) with special reference to flexistyly pollination mechanism.

Baack, E., M. C. Melo, L. H. Rieseberg, and D. Ortiz-Barrientos. 2015. The origins of reproductive isolation in plants. *New Phytologist* 207: 968–984.

Bacon, C. D., M. J. McKenna, M. P. Simmons, and W. L. Wagner. 2012. Evaluating multiple criteria for species delimitation: an empirical example using Hawaiian palms (Arecaceae: *Pritchardia*). *BMC Evolutionary Biology* 12: 23.

Barrett, S. C. H. 2010a. Darwin's legacy: the forms, function and sexual diversity of flowers. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365: 351–368.

Barrett, S. C. H. 2003. Mating strategies in flowering plants: the outcrossing-selfing paradigm and beyond. *Philosophical Transactions: Biological Sciences* 358: 991–1004.

Barrett, S. C. H. 2002a. Sexual interference of the floral kind. *Heredity* 88: 154–159.

Barrett, S. C. H. 2002b. The evolution of plant sexual diversity. *Nature Reviews Genetics* 3: 274–284.

Barrett, S. C. H. 2010b. Understanding plant reproductive diversity. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365: 99–109.

Basak, S., and L. Rangan. 2018. New record of nuclear DNA amounts of some Zingiberaceae species from North east India. *Data in Brief* 17: 66–70.

Bates, D., M. Mächler, B. Bolker, and S. Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67.

Bell, C. D., G. Calderon, L. Gonzalez, A. Scholz, and S. Liede-Schumann. 2015.

Resolving relationships within Valerianaceae (Dipsacales): new Insights and hypotheses from low-copy nuclear regions. *Systematic Botany* 40: 327–335.

Belser, C., B. Istace, E. Denis, M. Dubarry, F.-C. Baurens, C. Falentin, M. Genete, et al. 2018. Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. *Nature Plants* 4: 879–887.

Bhaumik, M., T. Langhu, S. Dey, C. R. Deb, and N. S. Jamir. 2017. Two new species of *Monolophus* (Zingiberaceae) from India. *Kew Bulletin* 72: 6.

Bickford, D., D. J. Lohman, N. S. Sodhi, P. K. L. Ng, R. Meier, K. Winker, K. K. Ingram, and I. Das. 2007. Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution* 22: 148–155.

de Boer, H., M. Newman, A. D. Poulsen, A. J. Droop, T. Fér, L. T. Thu Hiền, K. Hlavatá, et al. 2018. Convergent morphology in Alpinieae (Zingiberaceae): Recircumscribing *Amomum* as a monophyletic genus. *Taxon* 67: 6–36.

Brochmann, C., A. K. Brysting, I. G. Alsos, L. Borgen, H. H. Grundt, A.-C. Scheen, and R. Elven. 2004. Polyploidy in arctic plants. *Biological Journal of the Linnean Society* 82: 521–536.

Brown, J. H. 2014. Why are there so many species in the tropics? *Journal of Biogeography* 41: 8–22.

Buchfink, B., C. Xie, and D. H. Huson. 2015. Fast and sensitive protein alignment using DIAMOND. *Nature Methods* 12: 59–60.

Buckley, L. B., T. J. Davies, D. D. Ackerly, N. J. B. Kraft, S. P. Harrison, B. L. Anacker, H. V. Cornell, et al. 2010. Phylogeny, niche conservatism and the latitudinal diversity gradient in mammals. *Proceedings of the Royal Society B: Biological Sciences* 277: 2131–2138.

Buerki, S., M. W. Callmander, F. Schüpfer, M. Ravokatra, P. Küpfer, and N. Alvarez. 2009. Malagasy *Dracaena* Vand. ex L. (Ruscaceae): an investigation of discrepancies between morphological features and spatial genetic structure at a small evolutionary scale. *Plant Systematics and Evolution* 280: 15–28.

Burtt, B. L., and R. M. Smith. 1972. Monoecism in *Alpinia* (Zingiberaceae), with descriptions of six new species. *Notes from the Royal Botanic Garden Edinburgh* 32: 29–43.

Cadena, C. D., K. H. Kozak, J. P. Gómez, J. L. Parra, C. M. McCain, R. C. K.

Bowie, A. C. Carnaval, et al. 2012. Latitude, elevational climatic zonation and speciation in New World vertebrates. *Proceedings of the Royal Society B: Biological Sciences* 279: 194–201.

Cardoso, D., T. Särkinen, S. Alexander, A. M. Amorim, V. Bittrich, M. Celis, D. C. Daly, et al. 2017. Amazon plant diversity revealed by a taxonomically verified species list. *Proceedings of the National Academy of Sciences* 114: 10695–10700.

Carlsen, M. M., T. Fér, R. Schmickl, J. Leong-Škorničková, M. Newman, and W. J. Kress. 2018. Resolving the rapid plant radiation of early diverging lineages in the tropical Zingiberales: Pushing the limits of genomic data. *Molecular Phylogenetics and Evolution* 128: 55–68.

Castro-Insua, A., C. Gómez-Rodríguez, J. J. Wiens, and A. Baselga. 2018. Climatic niche divergence drives patterns of diversification and richness among mammal families. *Scientific Reports* 8: 8781.

Chakraborty, A., S. Mahajan, S. K. Jaiswal, and V. K. Sharma. 2020. Genome sequencing of turmeric provides evolutionary insights into its medicinal properties. *bioRxiv*: 2020.09.07.286245.

Charlesworth, D. 2016. The status of supergenes in the 21st century: recombination suppression in Batesian mimicry and sex chromosomes and other complex adaptations. *Evolutionary Applications* 9: 74–90.

Charlesworth, D., and B. Charlesworth. 1987. Inbreeding depression and its evolutionary consequences. *Annual Review of Ecology and Systematics* 18: 237–268.

Chaudhari, A. 1993. Forest plants of Eastern India. Ashish Publishing House.

Chen, G. F. 2013. Sexual isolation in two bee-pollinated *Costus* (Costaceae). *Plant Reproduction* 26: 3–16.

Chen, G. F., and D. W. Schemske. 2015. Ecological differentiation and local adaptation in two sister species of Neotropical *Costus* (Costaceae). *Ecology* 96: 440–449.

Chen, I.-N., C.-C. Chang, C.-C. Ng, C.-Y. Wang, Y.-T. Shyu, and T.-L. Chang. 2008. Antioxidant and antimicrobial activity of Zingiberaceae plants in Taiwan. *Plant Foods for Human Nutrition* 63: 15–20.

Chen, S. H., L. K. Guja, and A. N. Schmidt-Lebuhn. 2019. Conservation implications of widespread polyploidy and apomixis: a case study in the genus *Pomaderris* (Rhamnaceae). *Conservation Genetics* 20: 917–926.

Christenhusz, M. J. M., and J. W. Byng. 2016. The number of known plants species in the world and its annual increase. *Phytotaxa* 261: 201.

Chunco, A. J. 2014. Hybridization in a warmer world. *Ecology and Evolution* 4: 2019–2031.

Claros, M. G., R. Bautista, D. Guerrero-Fernández, H. Benzerki, P. Seoane, and N. Fernández-Pozo. 2012. Why assembling plant genome sequences is so challenging. *Biology* 1: 439–459.

Clugston, J. A. R., G. J. Kenicer, R. Milne, I. Overcast, T. C. Wilson, and N. S. Nagalingum. 2019. RADseq as a valuable tool for plants with large genomes—A case study in cycads. *Molecular Ecology Resources* 19: 1610–1622.

Cocker, J. M., J. Wright, J. Li, D. Swarbreck, S. Dyer, M. Caccamo, and P. M. Gilmartin. 2018. *Primula vulgaris* (primrose) genome assembly, annotation and gene expression, with comparative genomics on the heterostyly supergene. *Scientific Reports* 8: 17942.

Cooper, N., R. P. Freckleton, and W. Jetz. 2011. Phylogenetic conservatism of environmental niches in mammals. *Proceedings of the Royal Society B: Biological Sciences* 278: 2384–2391.

Corlett, R. T. 2016. Plant diversity in a changing world: Status, trends, and conservation needs. *Plant Diversity* 38: 10–16.

Coyne, J. A., and H. A. Orr. 2004. Speciation. Sinauer Associates, Sunderland, MA, USA.

Cui, X., R. Wei, and R. Huang. 1996. A study on the breeding system of Amomum tsao-ko. *In* T. Wu, Q. Wu, and Z. Chen [eds.], Second Symposium on the Family Zingiberaceae, 288–296. South China Institute of Botany, Guangzhou, P. R. China.

Cui, Y. W., and Q. J. Li. 2015. Autonomous self-pollination under dominant flexistylous outcrossing mechanism in *Alpinia galanga* (Zingiberaceae). *Plant Diversity* 37: 793–800.

D'Hont, A., F. Denoeud, J.-M. Aury, F.-C. Baurens, F. Carreel, O. Garsmeur, B.

Noel, et al. 2012. The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature* 488: 213–217.

Dai, C., and L. F. Galloway. 2011. Do dichogamy and herkogamy reduce sexual interference in a self-incompatible species? *Functional Ecology* 25: 271–278.

Daimei, P., and Y. Kumar. 2014. Ethnobotanical uses of gingers in Tamenglong district, Manipur, Northeast India. *Genetic Resources and Crop Evolution* 61: 273–285.

Darwin, C. 1877. The different forms of flowers on plants of the same species. John Murray, London.

Davey, M. W., R. Gudimella, J. A. Harikrishna, L. W. Sin, N. Khalid, and J. Keulemans. 2013. A draft *Musa balbisiana* genome sequence for molecular genetics in polyploid, inter- and intra-specific *Musa* hybrids. *BMC Genomics* 14: 683.

Delprat, A., B. Negre, M. Puig, and A. Ruiz. 2009. The transposon *Galileo* generates natural chromosomal inversions in *Drosophila* by ectopic recombination R. DeSalle [ed.],. *PLoS ONE* 4: e7883.

Docot, R. V. A., C. I. Banag, and A. D. Poulsen. 2019. Reinstatement and revision of the genus *Adelmeria* (Zingiberaceae) endemic to the Philippines. *TAXON* 68: 499–521.

Domingos, F. M. C. B., G. R. Colli, A. Lemmon, E. M. Lemmon, and L. B. Beheregaray. 2017. In the shadows: Phylogenomics and coalescent species delimitation unveil cryptic diversity in a Cerrado endemic lizard (Squamata: *Tropidurus*). *Molecular Phylogenetics and Evolution* 107: 455–465.

Dong, W., D. Wu, G. Li, D. Wu, and Z. Wang. 2018. Next-generation sequencing from bulked segregant analysis identifies a dwarfism gene in watermelon. *Scientific Reports* 8: 2908.

Donkpegan, A. S. L., J.-L. Doucet, J. Migliore, J. Duminil, K. Dainou, R. Piñeiro, J. J. Wieringa, et al. 2017. Evolution in African tropical trees displaying ploidy-habitat association: The genus *Afzelia* (Leguminosae). *Molecular Phylogenetics and Evolution* 107: 270–281.

Donkpegan, A. S. L., O. J. Hardy, P. Lejeune, M. Oumorou, K. Daïnou, and J.-L. Doucet. 2014. On a species complex, *Afzelia*, in African forests of economic

and ecological interest. A review | Un complexe d'espèces d'Afzelia des forêts africaines d'intérêt économique et écologique (synthèse bibliographique). *Biotechnology, Agronomy and Society and Environment* 18: 233–246.

Doyle, J. J., and J. L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19: 11–15.

Drummond, C. S., R. J. Eastwood, S. T. S. Miotto, and C. E. Hughes. 2012. Multiple continental radiations and correlates of diversification in *Lupinus* (Leguminosae): Testing for key innovation with incomplete taxon sampling. *Systematic Biology* 61: 443–460.

Duminil, J., D. Kenfack, V. Viscosi, L. Grumiau, and O. J. Hardy. 2012. Testing species delimitation in sympatric species complexes: The case of an African tropical tree, *Carapa* spp. (Meliaceae). *Molecular Phylogenetics and Evolution* 62: 275–285.

Eckert, C. G., K. E. Samis, and S. Dart. 2007. Reproductive assurance and the evolution of uniparental reproduction in flowering plants. *In* L. D. Harder, and S. C. H. Barrett [eds.], Ecology and Evolution of Flowers, 183–203. Oxford University Press.

Edwards, D., and J. Batley. 2010. Plant genome sequencing: applications for crop improvement. *Plant Biotechnology Journal* 8: 2–9.

Edwards, S. 1816. The Botanical register: consisting of coloured figures of ... Printed for James Ridgway, London .

Ellstrand, N. C., R. Whitkus, and L. H. Rieseberg. 1996. Distribution of spontaneous plant hybrids. *Proceedings of the National Academy of Sciences* 93: 5090–5093.

Endara, M.-J., P. D. Coley, N. L. Wiggins, D. L. Forrister, G. C. Younkin, J. A. Nicholls, R. T. Pennington, et al. 2018. Chemocoding as an identification tool where morphological- and DNA-based methods fall short: *Inga* as a case study. *New Phytologist* 218: 847–858.

Ennos, R. A., G. C. French, and P. M. Hollingsworth. 2005. Conserving taxonomic complexity. *Trends in Ecology & Evolution* 20: 164–168.

Ennos, R. A., R. Whitlock, M. F. Fay, B. Jones, L. E. Neaves, R. Payne, I. Taylor, et al. 2012. Process-Based Species Action Plans: an approach to conserve contemporary evolutionary processes that sustain diversity in taxonomically

complex groups. *Botanical Journal of the Linnean Society* 168: 194–203.

Enquist, B. J., X. Feng, B. Boyle, B. Maitner, E. A. Newman, P. M. Jørgensen, P. R. Roehrdanz, et al. 2019. The commonness of rarity: Global and future distribution of rarity across land plants. *Science Advances* 5: eaaz0414.

Faife-Cabrera, M., V. Ferrero, and L. Navarro. 2018. Relationship between herkogamy, incompatibility and reciprocity with pollen–ovule ratios in *Melochia* (Malvaceae). *Plant Biosystems - An International Journal Dealing with all Aspects of Plant Biology* 152: 80–89.

Feeley, K. J., and M. R. Silman. 2011. The data void in modeling current and future distributions of tropical species. *Global Change Biology* 17: 626–630.

Frodin, D. G. 2004. History and concepts of big plant genera. *Taxon* 53: 753–776.

Fukushima, K., X. Fang, D. Alvarez-Ponce, H. Cai, L. Carretero-Paulet, C. Chen, T.-H. Chang, et al. 2017. Genome of the pitcher plant *Cephalotus* reveals genetic changes associated with carnivory. *Nature Ecology & Evolution* 1: 0059.

Funk, W. C., M. Caminer, and S. R. Ron. 2012. High levels of cryptic species diversity uncovered in Amazonian frogs. *Proceedings of the Royal Society B: Biological Sciences* 279: 1806–1814.

Futschik, A., and C. Schlötterer. 2010. The next generation of molecular markers from massively parallel sequencing of pooled DNA samples. *Genetics* 186: 207–218.

Gagnon, E., C. E. Hughes, G. P. Lewis, and A. Bruneau. 2015. A new cryptic species in a new cryptic genus in the *Caesalpinia*group (Leguminosae) from the seasonally dry inter-Andean valleys of South America. *Taxon* 64: 468–490.

Gale, S. W., S. Duangjai, J. Li, Y. Ito, S. Watthana, P. Termwutthipreecha, M. L. Cheuk, and S. Suddee. 2018. Integrative analyses of *Nervilia* (Orchidaceae) section *Linervia* reveal further undescribed cryptic diversity in Thailand. *Systematics and Biodiversity* 16: 377–396.

Gao, J.-Y., L. Zhang, X.-B. Deng, P.-Y. Ren, J.-J. Kong, and Q.-J. Li. 2004. The floral biology of *Curcumorpha longiflora* (Zingiberaceae): a ginger with two-day flowers. *American Journal of Botany* 91: 289–293.

Genner, M. J., and G. F. Turner. 2012. Ancient hybridization and phenotypic novelty within Lake Malawi's cichlid fish radiation. *Molecular Biology and Evolution*

29: 195–206.

Gleeson, S. K. 1982. Heterodichogamy in walnuts: inheritance and stable ratios. *Evolution* 36: 892.

Goetze, M., C. M. Zanella, C. Palma-Silva, M. V. Büttow, and F. Bered. 2017. Incomplete lineage sorting and hybridization in the evolutionary history of closely related, endemic yellow-flowered *Aechmea* species of subgenus *Ortgiesia* (Bromeliaceae). *American Journal of Botany* 104: 1073–1087.

Goldenberg, R., D. S. Penneys, F. Almeda, W. S. Judd, and F. A. Michelangeli. 2008. Phylogeny of *Miconia* (Melastomataceae): Patterns of Stamen Diversification in a Megadiverse Neotropical Genus. *International Journal of Plant Sciences* 169: 963–979.

Gomes, F. P., C. A. Maggs, and M. B. B. Barros-Barreto. 2020. Integrative approach reveals four new cryptic species in the genus *Gayliella* (Ceramiaceae, Rhodophyta). *Journal of Phycology* 56: 437–457.

Goodwin, Z. A., D. J. Harris, D. Filer, J. R. I. Wood, and R. W. Scotland. 2015. Widespread mistaken identity in tropical plant collections. *Current Biology* 25: R1066–R1067.

Gould, B. A., Y. Chen, and D. B. Lowry. 2017. Pooled ecotype sequencing reveals candidate genetic mechanisms for adaptive differentiation and reproductive isolation. *Molecular Ecology* 26: 163–177.

Goulet, B. E., F. Roda, and R. Hopkins. 2017. Hybridization in plants: Old ideas, new techniques. *Plant Physiology* 173: 65–78.

Guo, Y., G. Yang, Y. Chen, D. Li, and Z. Guo. 2018. A comparison of different methods for preserving plant molecular materials and the effect of degraded DNA on ddRAD sequencing. *Plant Diversity* 40: 106–116.

Gurevich, A., V. Saveliev, N. Vyahhi, and G. Tesler. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29: 1072–1075.

Harder, L. D., and S. D. Johnson. 2009. Darwin's beautiful contrivances: evolutionary and functional evidence for floral adaptation. *New Phytologist* 183: 530–545.

Hareesh, V. S., and M. Sabu. 2018. *Amomum riwatchii* (Zingiberaceae): a new species from northeastern India. *Botany Letters* 165: 223–227.

Harkess, A. 2017. Modified Samberg Phenol:Chloforom HMW DNA prep for (some) plants V.1. Website dx.doi.org/10.17504/protocols.io.kpwcvpe.

Harris, D. J., A. D. Poulsen, C. Frimodt-Møller, J. Preston, and Q. C. B. Cronk. 2000. Rapid radiation in *Aframomum* (Zingiberaceae): evidence from nuclear ribosomal DNA Internal Transcribed Spacer (ITS) sequences. *Edinburgh Journal of Botany* 57: 377–395.

Harris, D. J., and A. H. Wortley. 2018. Monograph of *Aframomum* (Zingiberaceae). *Systematic Botany Monographs* 103: 1–204.

Haworth, A. H. 1805. The botanist's repository, for new and rare plants. Andrews, Henry Charles, London.

Healey, A., A. Furtado, T. Cooper, and R. J. Henry. 2014. Protocol: a simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species. *Plant Methods* 10: 21.

Hebert, P. D. N., E. H. Penton, J. M. Burns, D. H. Janzen, and W. Hallwachs. 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences* 101: 14812–14817.

Henry, A., V. Chithra, and N. Balakrishnan. 1989. Flora of Tamil Nadu, Series 1: Analysis. Botanical Survey of India, Southern Circle, Coimbatore.

Heyduk, K., D. W. Trapnell, C. F. Barrett, and J. Leebens-Mack. 2016. Phylogenomic analyses of species relationships in the genus *Sabal* (Arecaceae) using targeted sequence capture. *Biological Journal of the Linnean Society* 117: 106–120.

Hickey, L. J., and R. K. Peterson. 1978. *Zingiberopsis*, a fossil genus of the ginger family from Late Cretaceous to early Eocene sediments of Western Interior North America. *Canadian Journal of Botany* 56: 1136–1152.

Hillebrand, H. 2004. On the Generality of the Latitudinal Diversity Gradient. *The American Naturalist* 163: 192–211.

Hlavatá, K., J. Leong-Škorničková, E. Záveská, O. Šída, M. Newman, T. Mandáková, M. A. Lysak, et al. Phylogenomics and genome size evolution in *Amomum* s. s. (Zingiberaceae): comparison of traditional and modern sequencing methods.

Hojsgaard, D., S. Klatt, R. Baier, J. G. Carman, and E. Hörandl. 2014. Taxonomy and biogeography of apomixis in angiosperms and associated biodiversity characteristics. *Critical Reviews in Plant Sciences* 33: 414–427.

Hollingsworth, P. M. 2003. Taxonomic complexity, population genetics, and plant conservation in Scotland. *Botanical Journal of Scotland* 55: 55–63.

Holttum, R. E. 1950. The Zingiberaceae of the Malay Peninsula. *The Gardens' Bulletin Singapore* 13: 1–249.

Hooker, J. D. 1894. The flora of British India. L. Reeve, London.

Hörandl, E. 2018. The classification of asexual organisms: Old myths, new facts, and a novel pluralistic approach. *Taxon* 67: 1066–1081.

Hörandl, E., A.-C. Cosendai, and E. M. Temsch. 2008. Understanding the geographic distributions of apomictic plants: a case for a pluralistic approach. *Plant Ecology & Diversity* 1: 309–320.

Hosseini, S., K. Dadkhah, and R. Go. 2016. Molecular systematics of genus *Bulbophyllum* (Orchidaceae) in Peninsular Malaysia based on combined nuclear and plastid DNA sequences. *Biochemical Systematics and Ecology* 65: 40–48.

Hotaling, S., M. E. Foley, N. M. Lawrence, J. Bocanegra, M. B. Blanco, R. Rasoloarison, P. M. Kappeler, et al. 2016. Species discovery and validation in a cryptic radiation of endangered primates: coalescent-based species delimitation in Madagascar's mouse lemurs. *Molecular Ecology* 25: 2029–2045.

Hughes, C., and R. Eastwood. 2006. Island radiation on a continental scale: Exceptional rates of plant diversification after uplift of the Andes. *Proceedings of the National Academy of Sciences* 103: 10334–10339.

Hughes, M., P. W. Moonlight, A. Jara-Muñoz, M. C. Tebbitt, H. P. Wilson, and M. Pullan. 2015. Begonia Resource Centre. Website http://padme.rbge.org.uk/begonia/. [accessed 28 January 2021].

Hulse-Kemp, A. M., S. Maheshwari, K. Stoffel, T. A. Hill, D. Jaffe, S. R. Williams, N. Weisenfeld, et al. 2018. Reference quality assembly of the 3.5-Gb genome of *Capsicum annuum* from a single linked-read library. *Horticulture Research* 5: 4.

Huu, C. N., C. Kappel, B. Keller, A. Sicard, Y. Takebayashi, H. Breuninger, M. D. Nowak, et al. 2016. Presence versus absence of CYP734A50 underlies the style-length dimorphism in primroses. *Elife* 5: e17956.

Huu, C. N., B. Keller, E. Conti, C. Kappel, and M. Lenhard. 2020. Supergene evolution via stepwise duplications and neofunctionalization of a floral-organ identity gene. *Proceedings of the National Academy of Sciences* 117: 23148–23157.

Inoue, K. 1990. Dichogamy, Sex Allocation, and Mating System of *Campanula microdonta* and *C. punctata*. *Plant Species Biology* 5: 197–203.

Ito, Y., N. Tanaka, A. S. Barfod, J. Bogner, J. Li, O. Yano, and S. W. Gale. 2019. Molecular phylogenetic species delimitation in the aquatic genus *Ottelia*(Hydrocharitaceae) reveals cryptic diversity within a widespread species. *Journal of Plant Research* 132: 335–344.

Jablonski, D., K. Roy, and J. W. Valentine. 2006. Out of the Tropics: Evolutionary Dynamics of the Latitudinal Diversity Gradient. *Science* 314: 102–106.

Janzen, D. H., J. M. Burns, Q. Cong, W. Hallwachs, T. Dapkey, R. Manjunath, M. Hajibabaei, et al. 2017. Nuclear genomes distinguish cryptic species suggested by their DNA barcodes and ecology. *Proceedings of the National Academy of Sciences* 114: 8313–8318.

Jaramillo, M. A., R. Callejas, C. Davidson, J. F. Smith, A. C. Stevens, and E. J. Tepe. 2008. A phylogeny of the tropical genus *Piper* using ITS and the chloroplast intron psbJ–petA. *Systematic Botany* 33: 647–660.

Jatoi, S. A., A. Kikuchi, S. A. Gilani, and K. N. Watanabe. 2007. Phytochemical, pharmacological and ethnobotanical studies in mango ginger (*Curcuma amada* Roxb.; Zingiberaceae). *Phytotherapy Research* 21: 507–516.

Jay, P., A. Whibley, L. Frézal, M. Á. Rodríguez de Cara, R. W. Nowell, J. Mallet, K. K. Dasmahapatra, and M. Joron. 2018. Supergene evolution triggered by the introgression of a chromosomal inversion. *Current Biology* 28: 1839-1845.e3.

Jia, X.-C., J. Li, G.-H. Lu, and Y.-Q. Wang. 2015. The concrete evidence of flexistyly in *Plagiostachys*: pollination biology of a wild ginger on Hainan Island, China. *Ecology and Evolution* 5: 5364–5371.

Jiao, Y., P. Peluso, J. Shi, T. Liang, M. C. Stitzer, B. Wang, M. S. Campbell, et al. 2017. Improved maize reference genome with single-molecule technologies. *Nature* 546: 524–527.

Joe, A., T. Jayakrishanan, V. S. Hareesh, and M. Sabu. 2017. *Zingiber*

arunachalensis (Zingiberaceae): a new species from northeastern India. *Phytotaxa* 309: 95.

Jones, G., Z. Aydin, and B. Oxelman. 2015. DISSECT: an assignment-free Bayesian discovery method for species delimitation under the multispecies coalescent. *Bioinformatics* 31: 991–998.

Käfer, J., G. A. B. Marais, and J. R. Pannell. 2017. On the rarity of dioecy in flowering plants. *Molecular Ecology* 26: 1225–1241.

Kaur, A., C. O. Ha, K. Jong, V. E. Sands, H. T. Chan, E. Soepadmo, and P. S. Ashton. 1978. Apomixis may be widespread among trees of the climax rain forest. *Nature* 271: 440–442.

Kenfack, D. 2011. Resurrection in *Carapa* (Meliaceae): a reassessment of morphological variation and species boundaries using multivariate methods in a phylogenetic context. *Botanical Journal of the Linnean Society* 165: 186–221.

Kersey, P. J. 2019. Plant genome sequences: past, present, future. *Current Opinion in Plant Biology* 48: 1–8.

Khanuja, S. P. S., A. K. Shasany, M. P. Darokar, and S. Kumar. 1999. Rapid isolation of DNA from dry and fresh samples of plants producing large amounts of secondary metabolites and essential oils. *Plant Molecular Biology Reporter* 17: 74.

Kim, K. C., and L. B. Byrne. 2006. Biodiversity loss and the taxonomic bottleneck: emerging biodiversity science. *Ecological Research* 21: 794–810.

Knope, M. L., C. W. Morden, V. A. Funk, and T. Fukami. 2012. Area and the rapid radiation of Hawaiian *Bidens* (Asteraceae). *Journal of Biogeography* 39: 1206–1216.

Koenen, E. J. M., J. J. Clarkson, T. D. Pennington, and L. W. Chatrou. 2015. Recently evolved diversity and convergent radiations of rainforest mahoganies (Meliaceae) shed new light on the origins of rainforest hyperdiversity. *New Phytologist* 207: 327–339.

Kofler, R., R. V. Pandey, and C. Schlötterer,. 2011. PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics* 27: 3435–3436.

Košnar, J., J. Košnar, M. Herbstová, P. Macek, E. Rejmánková, and M. Štech. 2010.

Natural hybridization in tropical spikerushes of *Eleocharis* subgenus *Limnochloa* (Cyperaceae): Evidence from morphology and DNA markers. *American Journal of Botany* 97: 1229–1240.

Kovalsky, I. E., J. M. Roggero Luque, G. Elías, S. A. Fernández, and V. G. Solís Neffa. 2018. The role of triploids in the origin and evolution of polyploids of Turnera sidoides complex (Passifloraceae, Turneroideae). *Journal of Plant Research* 131: 77–89.

Kozak, K. H., and J. J. Wiens. 2010. Accelerated rates of climatic-niche evolution underlie rapid species diversification. *Ecology Letters* 13: 1378–1389.

Kozak, K. H., and J. J. Wiens. 2007. Climatic zonation drives latitudinal variation in speciation mechanisms. *Proceedings of the Royal Society B: Biological Sciences* 274: 2995–3003.

Kress, W. J., A.-Z. Liu, M. Newman, and Q. J. Li. 2005. The molecular phylogeny of *Alpinia* (Zingiberaceae): a complex and polyphyletic genus of gingers. *American Journal of Botany* 92: 167–178.

Kress, W. J., M. F. Newman, A. D. Poulsen, and C. D. Specht. 2007. An analysis of generic circumscriptions in tribe Alpinieae (Alpinioideae: Zingiberaceae). *Gardens' Bulletin Singapore* 59: 113–128.

Kress, W. J., and C. D. Specht. 2005. Between Cancer and Capricorn: Phylogeny, evolution and ecology of the primarily tropical Zingiberales. *Kongelige Danske Videnskabernes Selskab Biologiske Skrifter* 55: 459–478.

Kumar, G., L. Karthik, and K. V. B. Rao. 2011. A review on pharmacological and phytochemical properties of *Zingiber officinale* Roscoe (Zingiberaceae). *Journal of Pharmacy Research* 4: 2963–2966.

Kurland, S., C. W. Wheat, M. Paz Celorio Mancera, V. E. Kutschera, J. Hill, A. Andersson, C. Rubin, et al. 2019. Exploring a Pool-seq-only approach for gaining population genomic insights in nonmodel species. *Ecology and Evolution* 9: 11448–11463.

Laetsch, D. R., and M. L. Blaxter. 2017. BlobTools: Interrogation of genome assemblies. *F1000Research* 6: 1287.

Lagomarsino, L. P., F. L. Condamine, A. Antonelli, A. Mulch, and C. C. Davis. 2016. The abiotic and biotic drivers of rapid diversification in Andean

bellflowers (Campanulaceae). *New Phytologist* 210: 1430–1442.

Larcombe, M. J., B. Holland, D. A. Steane, R. C. Jones, D. Nicolle, R. E. Vaillancourt, and B. M. Potts. 2015. Patterns of reproductive isolation in Eucalyptus—A Phylogenetic Perspective. *Molecular Biology and Evolution* 32: 1833–1846.

Larsen, K. 2005. Distribution patterns and diversity centres of Zingiberaceae in SE Asia. *Biol. Skr* 55: 219–228.

Larsen, K., J. M. Lock, H. Maas, and P. J. M. Maas. 1998. Zingiberaceae. *In* K. Kubitzki [ed.], Flowering Plants ·Monocotyledons: Alismatanae and Commelinanae (except Gramineae), 474–495. Springer Berlin Heidelberg, Berlin, Heidelberg.

Leong-Škorničková, J., and M. Newman. 2015. Gingers of Cambodia, Laos and Vietnam. Singapore Botanic Gardens, National Parks Board, Singapore.

Leong-Škorničková, J., O. Šida, V. Jarolímová, M. Sabu, T. Fér, P. Trávniček, and J. Suda. 2007. Chromosome numbers and genome size variation in Indian species of *Curcuma* (Zingiberaceae). *Annals of Botany* 100: 505–526.

Leong-Škorničková, J., O. Šída, M. Sabu, and K. Marhold. 2008. Taxonomic and nomenclatural puzzles in Indian *Curcuma* : the identity and nomenclatural history of *C. zedoaria* (Christm.) Roscoe and *C. zerumbet* Roxb. (*Zingiberaceae*). *TAXON* 57: 949–962.

Leong-Škorničková, J., H. Đ. Tran, Q. B. Nguyen, and O. Šída. 2014. *Siliquamomum alcicorne* (Zingiberaceae: Alpinioideae), a new species from central Vietnam. *Gardens' Bulletin (Singapore)* 66: 39–46.

Lewin, H. A., G. E. Robinson, W. J. Kress, W. J. Baker, J. Coddington, K. A. Crandall, R. Durbin, et al. 2018. Earth BioGenome Project: Sequencing life for the future of life. *Proceedings of the National Academy of Sciences* 115: 4325–4333.

Lewis, D., and D. A. Jones. 1992. The genetics of heterostyly. *In* S. C. H. Barrett [ed.], Evolution and function of heterostyly, 129–150. Springer-Verlag: New York.

Li, F.-W., and A. Harkess. 2018. A guide to sequence your favorite plant genomes. *Applications in Plant Sciences* 6: e1030.

Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25: 1754–1760.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, et al. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25.

Li, J., J. M. Cocker, J. Wright, M. A. Webster, M. McMullan, S. Dyer, D. Swarbreck, et al. 2016. Genetic architecture and evolution of the *S* locus supergene in *Primula vulgaris*. *Nature Plants* 2: 16188.

Li, Q. J., W. J. Kress, Z.-F. Xu, Y.-M. Xia, L. Zhang, X.-B. Deng, and J. Y. Gao. 2002. Mating system and stigmatic behaviour during flowering of *Alpinia kwangsiensis* (Zingiberaceae). *Plant Systematics and Evolution* 232: 123–132.

Li, Q. J., Z. F. Xu, W. J. Kress, Y. M. Xia, L. Zhang, X. B. Deng, J. Y. Gao, and Z. L. Bai. 2001. Flexible style that encourages outcrossing. *Nature* 410: 432.

Li, Y., Y. Tong, and F. Xing. 2016. DNA Barcoding Evaluation and Its Taxonomic Implications in the Recently Evolved Genus *Oberonia* Lindl. (Orchidaceae) in China. *Frontiers in Plant Science* 7.

Lichman, B. R., G. T. Godden, J. P. Hamilton, L. Palmer, M. O. Kamileen, D. Zhao, B. Vaillancourt, et al. 2020. The evolutionary origins of the cat attractant nepetalactone in catnip. *Science Advances* 6: eaba0721.

Lim, C. 2016. The Rediscovery of *Alpinia malaccensis* (Burman f.) Roscoe in Ambon, and the implications for taxa similarly named in Java, Malaysia and India. *Folia Malaysiana* 17: 21–28.

Lissambou, B., T. l. . Couvreur, C. Atteke, T. Stévart, R. Piñeiro, G. Dauby, F. k Monthe, et al. 2019. Species delimitation in the genus *Greenwayodendron* based on morphological and genetic markers reveals new species Y. Brohard [ed.],. *Taxon*.

Liu, J., M. Möller, L.-M. Gao, D.-Q. Zhang, and D.-Z. Li. 2011. DNA barcoding for the discrimination of Eurasian yews (*Taxus* L., Taxaceae) and the discovery of cryptic species. *Molecular Ecology Resources* 11: 89–100.

Liu, S.-C., C.-T. Lu, and J.-C. Wang. 2009. Reticulate hybridization of *Alpinia* (Zingiberaceae) in Taiwan. *Journal of Plant Research* 122: 305–316.

Liu, S. C., and J. C. Wang. 2009. New natural hybrid, *Alpinia× ilanensis* (Zingiberaceae) in Taiwan. *Taiwania* 54: 134–139.

Lloyd, D. G. 1982. Selection of combined versus separate sexes in seed plants. *The American Naturalist* 120: 571–585.

Lohman, D. J., K. K. Ingram, D. M. Prawiradilaga, K. Winker, F. H. Sheldon, R. G. Moyle, P. K. L. Ng, et al. 2010. Cryptic genetic diversity in "widespread" Southeast Asian bird species suggests that Philippine avian endemism is gravely underestimated. *Biological Conservation* 143: 1885–1890.

Love, R. R., N. I. Weisenfeld, D. B. Jaffe, N. J. Besansky, and D. E. Neafsey. 2016. Evaluation of DISCOVAR *de novo* using a mosquito sample for cost-effective short-read genome assembly. *BMC Genomics* 17: 187.

Lowry, D. B., J. L. Modliszewski, K. M. Wright, C. A. Wu, and J. H. Willis. 2008. The strength and genetic basis of reproductive isolating barriers in flowering plants. *Philosophical Transactions of the Royal Society B: Biological Sciences* 363: 3009–3021.

Luc-Cayol, F., and L. Fereol. 1997. X Alpingera martinica (Zingiberaceae): An intergeneric hybrid between *Alpinia purpurata* and *Etlingera elatior*. *HortScience* 32: 914–915.

Lucas, E. J., B. S. Amorim, D. F. Lima, A. R. Lima-Lourenço, E. M. Nic Lughadha, C. E. B. Proença, P. O. Rosa, et al. 2018. A new infra-generic classification of the species-rich Neotropical genus *Myrcia* s.l. *Kew Bulletin* 73: 9.

Lucas, E. J., S. A. Harris, F. F. Mazine, S. R. Belsham, E. M. Nic Lughadha, A. Telford, P. E. Gasson, and M. W. Chase. 2007. Suprageneric phylogenetics of Myrteae, the generically richest tribe in Myrtaceae (Myrtales). *TAXON* 56: 1105–1128.

Lucas, E. J., K. Matsumoto, S. A. Harris, E. M. Nic Lughadha, B. Benardini, and M. W. Chase. 2011. Phylogenetics, Morphology, and Evolution of the Large Genus *Myrcia* s.l. (Myrtaceae). *International Journal of Plant Sciences* 172: 915–934.

Luo, Y.-L., T.-J. Bi, D. Li, Z.-L. Su, C. Tao, Y.-J. Luo, and Q.-J. Li. 2012. Effects of indole-3-acetic acid and auxin transport inhibitors on the style curvature of three *Alpinia* species (Zingiberaceae). *Acta Physiologiae Plantarum* 34: 2019–2025.

Luo, Y.-L., and Q.-J. Li. 2010. Effects of light and low temperature on the reciprocal style curvature of flexistylous *Alpinia* species (Zingiberaceae). *Acta Physiologiae Plantarum* 32.

Ma, X.-N., C.-L. Xie, Z. Miao, Q. Yang, and X.-W. Yang. 2017. An overview of chemical constituents from *Alpinia* species in the last six decades. *RSC Advances* 7: 14114–14144.

Majeský, L., F. Krahulec, and R. J. Vašut. 2017. How apomictic taxa are treated in current taxonomy: A review. *Taxon* 66: 1017–1040.

Mallet, J. 2007. Hybrid speciation. *Nature* 446: 279–283.

Mallet, J. 2005. Hybridization as an invasion of the genome. *Trends in Ecology & Evolution* 20: 229–237.

Mangaly, J., and M. Sabu. 1992. A taxonomic revision of South Indian *Alpinia* Roxb. (Zingiberaceae) . *Rheedea* 2: 38–51.

Mannion, P. D., P. Upchurch, R. B. J. Benson, and A. Goswami. 2014. The latitudinal biodiversity gradient through deep time. *Trends in Ecology & Evolution* 29: 42–50.

Mapleson, D., G. Garcia Accinelli, G. Kettleborough, J. Wright, and B. J. Clavijo. 2016. KAT: a K-mer analysis toolkit to quality control NGS datasets and genome assemblies. *Bioinformatics*: btw663.

Marques, I., D. Draper, L. Riofrío, and C. Naranjo. 2014. Multiple hybridization events, polyploidy and low postmating isolation entangle the evolution of neotropical species of *Epidendrum* (Orchidaceae). *BMC Evolutionary Biology* 14: 20.

Matute, D. R., I. A. Butler, D. A. Turissini, and J. A. Coyne. 2010. A test of the snowball theory for the rate of evolution of hybrid incompatibilities. *Science* 329.

McCain, C. M. 2009. Vertebrate range sizes indicate that mountains may be 'higher' in the tropics. *Ecology Letters* 12: 550–560.

Medrano, M., P. Guitián, and J. Guitián. 2000. Patterns of fruit and seed set within inflorescences of Pancratium maritimum (Amaryllidaceae): nonuniform pollination, resource limitation, or architectural effects? *American Journal of Botany* 87.

Merrill, E. D. 1922. An enumeration of Philippine flowering plants. Bureau of Printing, Manila.

Meyer, C., P. Weigelt, and H. Kreft. 2016. Multidimensional biases, gaps and

uncertainties in global plant occurrence information. *Ecology Letters* 19: 992–1006.

Michael, T. P., F. Jupe, F. Bemm, S. T. Motley, J. P. Sandoval, C. Lanz, O. Loudet, et al. 2018. High contiguity *Arabidopsis thaliana* genome assembly with a single nanopore flow cell. *Nature Communications* 9: 541.

Michael, T. P., and R. VanBuren. 2020. Building near-complete plant genomes. *Current Opinion in Plant Biology* 54: 26–33.

Mittelbach, G. G., D. W. Schemske, H. V. Cornell, A. P. Allen, J. M. Brown, M. B. Bush, S. P. Harrison, et al. 2007. Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. *Ecology Letters* 10: 315–331.

Moonlight, P. W., W. H. Ardi, L. A. Padilla, K.-F. Chung, D. Fuller, D. Girmansyah, R. Hollands, et al. 2018. Dividing and conquering the fastest-growing genus: Towards a natural sectional classification of the mega-diverse genus *Begonia* (Begoniaceae). *Taxon* 67: 267–323.

Moraes, A. P., M. Chinaglia, C. Palma-Silva, and F. Pinheiro. 2013. Interploidy hybridization in sympatric zones: the formation of *Epidendrum fulgens* × *E. puniceoluteum* hybrids (Epidendroideae, Orchidaceae). *Ecology and Evolution* 3: 3824–3837.

Mounce, R., P. Smith, and S. Brockington. 2017. Ex situ conservation of plant diversity in the world's botanic gardens. *Nature Plants* 3: 795–802.

Moyle, L. C., and T. Nakazato. 2010. Hybrid incompatibility 'snowballs' between *Solanum* species. *Science* 329: 1521–1523.

Moyle, L. C., M. S. Olson, and P. Tiffin. 2004. Patterns of reproductive isolation in three angiosperm genera. *Evolution* 58: 1195–1208.

Mráz, P., J. Chrtek, and J. Fehrer. 2011. Interspecific hybridization in the genus Hieracium s. str.: evidence for bidirectional gene flow and spontaneous allopolyploidization. *Plant Systematics and Evolution* 293: 237–245.

Muñoz-Rodríguez, P., T. Carruthers, J. R. I. Wood, B. R. M. Williams, K. Weitemier, B. Kronmiller, Z. Goodwin, et al. 2019. A taxonomic monograph of *Ipomoea* integrated across phylogenetic scales. *Nature Plants* 5: 1136–1144.

Naiki, A. 2012. Heterostyly and the possibility of its breakdown by polyploidization. *Plant Species Biology* 27: 3–29.

Nayar, T. S., A. Rasiya Beegam, N. Mohanan, and G. Rajkumar. 2006. Flowering Plants of Kerala: A Handbook. Tropical Botanic Garden and Research Institute, Kerala, Thiruvananthapuram.

Neri, J., T. Wendt, and C. Palma-Silva. 2018. Natural hybridization and genetic and morphological variation between two epiphytic bromeliads. *AoB PLANTS* 10.

Newman, M., S. Ketphanh, B. Svengsuksa, P. Thomas, K. Sengdala, V. Lamxay, and K. Armstrong. 2007. A checklist of the vascular plants of Lao PDR. Royal Botanic Garden Edinburgh.

Nic Lughadha, E. M., S. P. Bachman, T. C. C. Leão, F. Forest, J. M. Halley, J. Moat, C. Acedo, et al. 2020. Extinction risk and threats to plants and fungi. *Plants, People, Planet* 2: 389–408.

Nic Lughadha, E. M., V. G. Staggemeier, T. N. C. Vasconcelos, B. E. Walker, C. Canteiro, and E. J. Lucas. 2019. Harnessing the potential of integrated systematics for conservation of taxonomically complex, megadiverse plant groups. *Conservation Biology* 33: 511–522.

Nicholls, J., R. Pennington, E. Koenen, C. Hughes, J. Hearn, L. Bunnefeld, K. Dexter, et al. 2015. Using targeted enrichment of nuclear genes to increase phylogenetic resolution in the neotropical rain forest genus *Inga* (Leguminosae: Mimosoideae). *Frontiers in Plant Science* 6: 710.

Van der Niet, T., R. Peakall, and S. D. Johnson. 2014. Pollinator-driven ecological speciation in plants: new evidence and future perspectives. *Annals of Botany* 113: 199–212.

Nopporncharoenkul, N., J. Chanmai, T. Jenjittikul, K. Anamthawat-Jónsson, and P. Soontornchainaksaeng. 2017. Chromosome number variation and polyploidy in 19 *Kaempferia* (Zingiberaceae) taxa from Thailand and one species from Laos. *Journal of Systematics and Evolution* 55: 466–476.

Odyuo, N., D. K. Roy, and A. A. Mao. 2019. *Zingiber dimapurense* (Zingiberaceae), a new species from Nagaland, India. *NeBIO* 10: 59–65.

Ornduff, R. 1992. Historical Perspectives on Heterostyly. *In* S. C. H. Barrett [ed.], Evolution and Function of Heterostyly. Monographs on Theoretical and Applied Genetics, vol 15, 31–39. Springer-Verlag, Berlin, Heidelberg.

Orr, H. A. 1995. The population genetics of speciation: the evolution of hybrid

incompatibilities. *Genetics* 139: 1805–1813.

Orr, H. A., and M. Turelli. 2001. The evolution of postzygotic isolation: accumulating Dobzhansky-Muller incompatibilities. *Evolution* 55: 1085–1094.

Owens, G. L., and L. H. Rieseberg. 2014. Hybrid incompatibility is acquired faster in annual than in perennial species of sunflower and tarweed. *Evolution* 68: 893–900.

Padilla-García, N., B. M. Rojas-Andrés, N. López-González, M. Castro, S. Castro, J. Loureiro, D. C. Albach, et al. 2018. The challenge of species delimitation in the diploid-polyploid complex *Veronica* subsection *Pentasepalae*. *Molecular Phylogenetics and Evolution* 119: 196–209.

Palma-Silva, C., T. Wendt, F. Pinheiro, T. Barbará, M. F. Fay, S. Cozzolino, and C. Lexer. 2011. Sympatric bromeliad species (*Pitcairnia* spp.) facilitate tests of mechanisms involved in species cohesion and reproductive isolation in Neotropical inselbergs. *Molecular Ecology* 20: 3185–3201.

Pellicer, J., O. Hidalgo, S. Dodsworth, and I. J. Leitch. 2018. Genome size diversity and its impact on the evolution of land plants. *Genes* 9: 88.

Pennington, R. T., M. Lavin, D. E. Prado, C. A. Pendry, S. K. Pell, and C. A. Butterworth. 2004. Historical climate change and speciation: neotropical seasonally dry forest plants show patterns of both Tertiary and Quaternary diversification Y. Malhi, and O. L. Phillips [eds.],. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 359: 515–538.

Pérez-Escobar, O. A., D. Bogarín, R. Schley, R. M. Bateman, G. Gerlach, D. Harpke, J. Brassac, et al. 2020. Resolving relationships in an exceedingly young Neotropical orchid lineage using Genotyping-by-sequencing data. *Molecular Phylogenetics and Evolution* 144: 106672.

Pessoa, E. M., M. Alves, A. Alves-Araújo, C. Palma-Silva, and F. Pinheiro. 2012. Integrating different tools to disentangle species complexes: A case study in *Epidendrum* (Orchidaceae). *Taxon* 61: 721–734.

Phuong Hanh, N., and N. Quoc Binh. 2014. Distribution of *Alpinia* (Zingiberaceae) and their use pattern in Vietnam. *Journal of Biodiversity & Endangered Species* 02.

Pianka, E. R. 1966. Latitudinal gradients in species diversity: a review of concepts.

The American Naturalist 100: 33–46.

Pimm, S. L., and L. N. Joppa. 2015. How many plant species are there, where are they, and at what rate are they going extinct? *Annals of the Missouri Botanical Garden* 100: 170–176.

Pinheiro, F., F. De Barros, C. Palma-Silva, D. Meyer, M. F. Fay, R. M. Suzuki, C. Lexer, and S. Cozzolino. 2010. Hybridization and introgression across different ploidy levels in the Neotropical orchids *Epidendrum fulgens* and *E. puniceoluteum* (Orchidaceae). *Molecular Ecology* 19: 3981–3994.

Pinheiro, F., and S. Cozzolino. 2013. *Epidendrum* (Orchidaceae) as a model system for ecological and evolutionary studies in the Neotropics. *TAXON* 62: 77–88.

Pinheiro, F., S. Cozzolino, F. de Barros, T. M. Z. M. Gouveia, R. M. Suzuki, M. F. Fay, and C. Palma-Silva. 2013. Pylogeographic structure and outbreeding depression reveal early stages of reproductive isolation in the Neotropical orchid *Epidendrum denticulatum*. *Evolution* 67: 2024–2039.

Pinheiro, F., M. V. Dantas-Queiroz, and C. Palma-Silva. 2018. Plant species complexes as models to understand speciation and evolution: A review of South American studies. *Critical Reviews in Plant Sciences* 37: 54–80.

Pinheiro, F., T. M. Zanfra de Melo e Gouveia, S. Cozzolino, D. Cafasso, P. Cardoso-Gustavson, R. M. Suzuki, and C. Palma-Silva. 2016. Strong but permeable barriers to gene exchange between sister species of *Epidendrum*. *American Journal of Botany* 103: 1472–1482.

Portillo, F., W. R. Branch, W. Conradie, M.-O. Rödel, J. Penner, M. F. Barej, C. Kusamba, et al. 2018. Phylogeny and biogeography of the African burrowing snake subfamily Aparallactinae (Squamata: Lamprophiidae). *Molecular Phylogenetics and Evolution* 127: 288–303.

Poulsen, A. D., and P. Phonsena. 2017. Morphological variation and distribution of the useful ginger *Etlingera pavieana* (Zingiberaceae). *Nordic Journal of Botany* 35: 467–475.

POWO. 2019. _. *Plants of the World Online. Facilitated by the Royal Botanic Gardens, Kew. Published on the Internet*. Website http://www.plantsoftheworldonline.org/ [accessed 28 January 2021].

Prata, E. M. B., C. Sass, D. P. Rodrigues, F. M. C. B. Domingos, C. D. Specht, G.

Damasco, C. C. Ribas, et al. 2018. Towards integrative taxonomy in Neotropical botany: disentangling the *Pagamea guianensis* species complex (Rubiaceae). *Botanical Journal of the Linnean Society* 188: 213–231.

Ramachandran, K. 1969. Chromosome Numbers in Zingiberaceae. *Cytologia* 34: 213–221.

Ranallo-Benavidez, T. R., K. S. Jaron, and M. C. Schatz. 2020. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nature Communications* 11: 1432.

Rangsiruji, A. 1999. A study of the infrageneric classification of *Alpinia* Roxb. (Zingiberaceae) using molecular data. University of Edinburgh.

Rangsiruji, A., M. F. Newman, and Q. C. B. Cronk. 2000. A study of the infrageneric classification of *Alpinia* (Zingiberaceae) based on the ITS region of nuclear rDNA and the trnL-F spacer of chloroplast DNA. *In* K. L. Wilson, and D. A. Morrison [eds.], Monocots—systematics and evolution, 695–709. CSIRO Publishing, Collingwood, Australia.

Raven, P. H., R. E. Gereau, P. B. Phillipson, C. Chatelain, C. N. Jenkins, and C. Ulloa Ulloa. 2020. The distribution of biodiversity richness in the tropics. *Science Advances* 6: eabc6228.

Rehder, A., C. A. Weatherby, R. Mansfeld, and M. L. Green. 1935. Conservation of Later Generic Homonyms. *Bulletin of Miscellaneous Information (Royal Gardens, Kew)* 1935: 351–352.

Ren, P. Y., M. Liu, and Q. J. Li. 2007. An example of flexistyly in a wild cardamom species (*Amomum maximum*, (Zingiberaceae)). *Plant Systematics and Evolution* 267: 147–154.

Renner, S. S. 2001. How common is heterodichogamy? *Trends in Ecology & Evolution* 16: 595–597.

Renner, S. S. 2014. The relative and absolute frequencies of angiosperm sexual systems: Dioecy, monoecy, gynodioecy, and an updated online database. *American Journal of Botany* 101: 1588–1596.

Rhymer, J. M., and D. Simberloff. 1996. Extinction by hybridization and introgression. *Annual Review of Ecology and Systematics* 27: 83–109.

Rice, A., L. Glick, S. Abadi, M. Einhorn, N. M. Kopelman, A. Salman-Minkov, J.

Mayzel, et al. 2015. The Chromosome Counts Database (CCDB)– a community resource of plant chromosome numbers. *New Phytologist* 206: 19–26.

Rice, A., P. Šmarda, M. Novosolov, M. Drori, L. Glick, N. Sabath, S. Meiri, et al. 2019. The global biogeography of polyploid plants. *Nature Ecology & Evolution* 3: 265–273.

Richardson, J. E., R. T. Pennington, T. D. Pennington, and P. M. Hollingsworth. 2001. Rapid diversification of a species-rich genus of Neotropical rain forest trees. *Science* 293: 2242.

Rieseberg, L. H., and J. H. Willis. 2007. Plant Speciation. *Science* 317: 910–914.

Rivera, P. C., R. E. González-Ittig, A. Robainas Barcia, L. I. Trimarchi, S. Levis, G. E. Calderón, and C. N. Gardenal. 2018. Molecular phylogenetics and environmental niche modeling reveal a cryptic species in the *Oligoryzomys flavescens* complex (Rodentia, Cricetidae). *Journal of Mammalogy* 99: 363–376.

Robertson, A., T. C. G. Rich, A. M. Allen, L. Houston, C. Roberts, J. R. Bridle, S. A. Harris, and S. J. Hiscock. 2010. Hybridization and polyploidy as drivers of continuing evolution and speciation in *Sorbus*. *Molecular Ecology* 19: 1675–1690.

Rolland, J., F. L. Condamine, F. Jiguet, and H. Morlon. 2014. Faster speciation and reduced extinction in the tropics contribute to the mammalian latitudinal diversity gradient C. Moritz [ed.],. *PLoS Biology* 12: e1001775.

Romdal, T. S., M. B. Araújo, and C. Rahbek. 2013. Life on a tropical planet: niche conservatism and the global diversity gradient. *Global Ecology and Biogeography* 22: 344–350.

Roscoe, W. 1807. A new arrangement of the plants of the Monandrian class usually called Scitamineae. *Transactions of the Linnean Society of London* v.8: 347-.

Roscoe, W. 1828. Monandrian plants of the order Scitamineae. George Smith, Liverpool .

Roscoe, W. 1815. Remarks on Dr. Roxburgh's description of the Monandrous Plants of India. Part 2. *Transactions of the Linnean Society of London* v.11: 281-.

Roxburgh, W. 1810. Descriptions of several of the monandrous plants of India. *Asiatic Researches* 11: 350–356.

Roxburgh, W. 1820. Flora Indica or descriptions of Indian plants. Mission Press, Serampore.

Sabu, M. 2006. Zingiberaceae and Costaceae of south India. Indian Association for Angiosperm Taxonomy.

Sabu, M., and V. S. Hareesh. 2020. *Hedychium mechukanum* (Zingiberaceae), a new species from the eastern Himalayas, India. *Gardens' Bulletin Singapore* 72: 291–297.

Sabu, M., and J. K. Mangaly. 1991. *Alpinia smithiae* (Zingiberaceae): a new species from South India. *Edinburgh Journal of Botany* 48: 69–71.

Sabu, M., E. Sanoj, and A. Pradeep. 2008. *Alpinia ovoidocarpa* (Zingiberaceae): A new record for India. *Acta Botanica Hungarica* 50: 391–395.

Salisbury, C. L., N. Seddon, C. R. Cooney, and J. A. Tobias. 2012. The latitudinal gradient in dispersal constraints: ecological specialisation drives diversification in tropical birds A. Mooers [ed.],. *Ecology Letters* 15: 847–855.

Sampaio, D. S., N. S. Bittencourt Júnior, and P. E. Oliveira. 2013. Sporophytic apomixis in polyploid *Anemopaegma* species (Bignoniaceae) from central Brazil. *Botanical Journal of the Linnean Society* 173: 77–91.

Sanderson, M. J., and M. J. Donoghue. 1994. Shifts in diversification rate with the origin of Angiosperms. *Science* 264: 1590–1593.

Sangvirotjanapat, S., J. Denduangboriphant, and M. F. Newman. 2019. A taxonomic revision of *Globba* subsect. *Nudae* (Zingiberaceae). *European Journal of Taxonomy* 503.

Sangvirotjanapat, S., H. Đ. Trần, and M. F. Newman. 2020. Ten new species of *Globba* section *Globba* from continental South-East Asia. *Thai Forest Bulletin (Botany)* 48: 212–233.

dos Santos, A. P. M., C. M. Fracasso, M. Luciene dos Santos, R. Romero, M. Sazima, and P. E. Oliveira. 2012. Reproductive biology and species geographical distribution in the Melastomataceae: a survey based on New World taxa. *Annals of Botany* 110: 667–679.

Särkinen, T. E., J. L. Marcelo-Peña, A. D. Yomona, M. F. Simon, R. T. Pennington, and C. E. Hughes. 2011. Underestimated endemic species diversity in the dry inter-Andean valley of the Río Marañón, northern Peru: An example from

Mimosa (Leguminosae, Mimosoideae). *Taxon* 60: 139–150.

Särkinen, T., M. Staats, J. E. Richardson, R. S. Cowan, and F. T. Bakker. 2012. How to open the treasure chest? Optimising DNA extraction from herbarium specimens. *PLoS ONE* 7: e43808.

Saryan, P., S. Gupta, and V. Gowda. 2020. Species complex delimitations in the genus *Hedychium* (Zingiberaceae): A machine learning approach for discovery of clusters. *Applications in Plant Sciences* 8: in press-undefined.

Sasidharan, N. 2004. Biodiversity documentation for Kerala. Part 6: Flowering Plants. Kerala Forest Research Institute.

Schatz, M. C., J. Witkowski, and W. R. McCombie. 2012. Current challenges in *de novo* plant genome sequencing and assembly. *Genome Biology* 13: 243.

Schley, R. J., R. T. Pennington, O. A. Pérez-Escobar, A. J. Helmstetter, M. de la Estrella, I. Larridon, I. A. B. Sabino Kikuchi, et al. 2020. Introgression across evolutionary scales suggests reticulation contributes to Amazonian tree diversity. *bioRxiv*: 2019.12.12.873927.

Schlötterer, C., R. Tobler, R. Kofler, and V. Nolte. 2014. Sequencing pools of individuals — mining genome-wide polymorphism data without big funding. *Nature Reviews Genetics* 15: 749–763.

Schlumpberger, B. O., and S. S. Renner. 2012. Molecular phylogenetics of *Echinopsis* (Cactaceae): Polyphyly at all levels and convergent evolution of pollination modes and growth forms. *American Journal of Botany* 99: 1335–1349.

Schneider, J. V., T. Jungcurt, D. Cardoso, A. M. Amorim, M. Töpel, T. Andermann, O. Poncy, et al. 2020. Phylogenomics of the tropical plant family Ochnaceae using targeted enrichment of nuclear genes and 250+ taxa. *TAXON*: tax.12421.

Schumann, K. 1904. Zingiberaceae. Heft 20 Pp. 1-458 In: Engler A, Das Pflanzenreich IV, 46. Berlin.

Scopece, G., C. Lexer, A. Widmer, and S. Cozzolino. 2010. Polymorphism of postmating reproductive isolation within plant species. *Taxon* 59: 1367–1374.

Scopece, G., A. Musacchio, A. Widmer, and S. Cozzolino. 2007. Patterns of reproductive isolation in Mediterranean deceptive orchids. *Evolution* 61: 2623–2642.

Seehausen, O. 2004. Hybridization and adaptive radiation. *Trends in Ecology & Evolution* 19: 198–207.

Seppey, M., M. Manni, and E. M. Zdobnov. 2019. BUSCO: Assessing genome assembly and annotation completeness BT - Gene Prediction: Methods and Protocols. *In* M. Kollmar [ed.], 227–245. Springer New York, New York, NY.

Shaffer, H. B., and R. C. Thomson. 2007. Delimiting species in recent radiations. *Systematic Biology* 56: 896–906.

Shore, J. S., H. J. Hamam, P. D. J. Chafe, J. D. J. Labonne, P. M. Henning, and A. G. McCubbin. 2019. The long and short of the *S*-locus in *Turnera* (Passifloraceae). *New Phytologist* 224.

Silvestro, D., G. Zizka, and K. Schulte. 2014. Disentangling the effects of key innovations on the diversification of Bromelioideae (Bromeliaceae). *Evolution* 68: 163–175.

Smith, C. E. 1971. Preparing herbarium specimens of vascular plants. USDA Agriculture Information Bulletin No. 348, Washington, DC.

Smith, R. M. 1990. *Alpinia* (Zingiberaceae): A proposed new infrageneric classification. *Edinburgh Journal of Botany* 47: 1–75.

Solís-Lemus, C., L. L. Knowles, and C. Ané. 2015. Bayesian species delimitation combining multiple genes and traits in a unified framework. *Evolution* 69: 492–507.

Soltis, D. E., P. S. Soltis, D. W. Schemske, J. F. Hancock, J. N. Thompson, B. C. Husband, and W. S. Judd. 2007. Autopolyploidy in Angiosperms: Have We Grossly Underestimated the Number of Species? *Taxon* 56: 13–30.

Soltis, D. E., C. J. Visger, and P. S. Soltis. 2014. The polyploidy revolution then...and now: Stebbins revisited. *American Journal of Botany* 101: 1057–1078.

Soltis, P. S., and D. E. Soltis. 2000. The role of genetic and genomic attributes in the success of polyploids. *Proceedings of the National Academy of Sciences* 97: 7051–7057.

Soltis, P. S., and D. E. Soltis. 2009. The role of hybridization in plant speciation. *Annual Review of Plant Biology* 60: 561–588.

Stace, C. A., C. D. Preston, and D. A. Pearman. 2015. Hybrid flora of the British

Isles. Botanical Society of Britain and Ireland.

Stebbins, G. L. 1970. Adaptive Radiation of Reproductive Characteristics in Angiosperms, I: Pollination Mechanisms. *Annual Review of Ecology and Systematics* 1: 307–326.

ter Steege, H., N. C. A. Pitman, D. Sabatier, C. Baraloto, R. P. Salomao, J. E. Guevara, O. L. Phillips, et al. 2013. Hyperdominance in the Amazonian tree flora. *Science* 342: 1243092–1243092.

Stephenson, A. G. 1981. Flower and fruit abortion: Proximate causes and ultimate functions. *Annual Review of Ecology and Systematics* 12: 253–279.

Su, Z.-L., X.-L. Cui, L. Li, Q.-J. Li, and Y.-L. Luo. 2017. Similar behavior, different mechanisms: the research of the style bending of *Alpinia* species. *Acta Physiologiae Plantarum* 39: 143.

Sultan, S. E. 2000. Phenotypic plasticity for plant development, function and life history. *Trends in Plant Science* 5: 537–542.

Sun, S., J. Y. Gao, W. J. Liao, Q. J. Li, and D. Y. Zhang. 2007. Adaptive significance of flexistyly in *Alpinia blepharocalyx* (Zingiberaceae): A hand-pollination experiment. *Annals of Botany* 99: 661–666.

Sun, S., D. Y. Zhang, A. R. Ives, and Q. J. Li. 2011. Why do stigmas move in a flexistylous plant? *Journal of Evolutionary Biology* 24: 497–504.

Surveswaran, S., V. Gowda, and M. Sun. 2018. Using an integrated approach to identify cryptic species, divergence patterns and hybrid species in Asian ladies' tresses orchids (*Spiranthes*, Orchidaceae). *Molecular Phylogenetics and Evolution* 124: 106–121.

Sweet, R. 1826. Sweet's Hortus britannicus : or a catalogue of plants cultivated in the gardens of Great Britain, arranged in natural orders. J. Ridgway, London .

Takano, A., J. Gisil, and M. Suleiman. 2013. Floral size variation causes differentiation of pollinators and genetic parameters in *Alpinia nieuwenhuizii*, a flexistylous ginger (Zingiberaceae). *Plant Systematics and Evolution* 299: 865–871.

Takano, A., J. Gisil, M. Yusoff, and T. Tachi. 2005. Floral and pollinator behaviour of flexistylous Bornean ginger, *Alpinia nieuwenhuizii* (Zingiberaceae). *Plant Systematics and Evolution* 252: 167–173.

Takano, A., and H. Okada. 2002. Multiple occurrences of triploid formation in *Globba* (Zingiberaceae) from molecular evidence. *Plant Systematics and Evolution* 230: 143–159.

The Legume Phylogeny Working Group, A. Bruneau, J. J. Doyle, P. Herendeen, C. Hughes, G. Kenicer, G. Lewis, et al. 2013. Legume phylogeny and classification in the 21st century: Progress, prospects and lessons for other species–rich clades. *Taxon* 62: 217–248.

Thomas, V. P., M. Sabu, and H. Lalramnghinglova. 2013. *Amomum dampuianum* and *A. mizoramense* spp. nov. (Zingiberaceae) from Mizoram, northeast India. *Nordic Journal of Botany* 31: 561–568.

Thomma, B. P. H. J., M. F. Seidl, X. Shi-Kunne, D. E. Cook, M. D. Bolton, J. A. L. van Kan, and L. Faino. 2016. Mind the gap; seven reasons to close fragmented genome assemblies. *Fungal Genetics and Biology* 90: 24–30.

Thompson, T. E., and L. D. Romberg. 1985. Inheritance of heterodichogamy in pecan. *Journal of Heredity* 76: 456–458.

Thongam, B., and B. Konsam. 2017. A New Species of *Caulokaempferia* (Zingiberaceae) from Manipur, India. *J. Jpn. Bot* 92: 82–86.

Todesco, M., M. A. Pascual, G. L. Owens, K. L. Ostevik, B. T. Moyers, S. Hübner, S. M. Heredia, et al. 2016. Hybridization and extinction. *Evolutionary Applications* 9: 892–908.

Toussaint, E. F. A., J. Morinière, C. J. Müller, K. Kunte, B. Turlin, A. Hausmann, and M. Balke. 2015. Comparative molecular species delimitation in the charismatic Nawab butterflies (Nymphalidae, Charaxinae, *Polyura*). *Molecular Phylogenetics and Evolution* 91: 194–209.

Tripathi, S. 2002. Taxonomic revision of the genus *Alpinia* Roxb. (Zingiberaceae) in North-East India. *Phytotaxonomy* 2: 171–184.

Tripp, E. A., Y.-H. E. Tsai, Y. Zhuang, and K. G. Dexter. 2017. RADseq dataset with 90% missing data fully resolves recent radiation of *Petalidium* (Acanthaceae) in the ultra-arid deserts of Namibia. *Ecology and Evolution* 7: 7920–7936.

Turland, N. J., J. H. Wiersema, F. R. Barrie, W. Greuter, D. L. Hawksworth, P. S. Herendeen, S. Knapp, et al. 2018. International Code of Nomenclature for

algae, fungi, and plants (Shenzhen Code) adopted by the Nineteenth International Botanical Congress Shenzhen, China, July 2017. N. Turland, J. Wiersema, F. Barrie, Greuter W, D. Hawksworth, P. Herendeen, S. Knapp, et al. [eds.],. Koeltz Botanical Books, Glashütten.

Turner, B., O. Paun, J. Munzinger, S. Duangjai, M. W. Chase, and R. Samuel. 2013. Analyses of amplified fragment length polymorphisms (AFLP) indicate rapid radiation of *Diospyros* species (Ebenaceae) endemic to New Caledonia. *BMC Evolutionary Biology* 13: 269.

Twyford, A. D., and R. A. Ennos. 2012. Next-generation hybridization and introgression. *Heredity* 108: 179–189.

Twyford, A. D., and J. Friedman. 2015. Adaptive divergence in the monkey flower *Mimulus guttatus* is maintained by a chromosomal inversion. *Evolution* 69: 1476–1486.

Twyford, A. D., C. A. Kidner, and R. A. Ennos. 2015. Maintenance of species boundaries in a Neotropical radiation of *Begonia*. *Molecular Ecology* 24: 4982–4993.

Uribe-Convers, S., and D. C. Tank. 2015. Shifts in diversification rates linked to biogeographic movement into new areas: An example of a recent radiation in the Andes. *American Journal of Botany* 102: 1854–1869.

Ushijima, K., R. Nakano, M. Bando, Y. Shigezane, K. Ikeda, Y. Namba, S. Kume, et al. 2012. Isolation of the floral morph-related genes in heterostylous flax (Linum grandiflorum): The genetic polymorphism and the transcriptional and post-transcriptional regulations of the S locus. *The Plant Journal* 69.

Valderrama, E., J. E. Richardson, C. A. Kidner, S. Madriñán, and G. N. Stone. 2018. Transcriptome mining for phylogenetic markers in a recently radiated genus of tropical plants (*Renealmia* L.f., Zingiberaceae). *Molecular Phylogenetics and Evolution* 119: 13–24.

Vamosi, J. C., and S. M. Vamosi. 2004. The role of diversification in causing the correlates of dioecy. *Evolution* 58: 723–731.

Vasconcelos, T. N. C., G. Prenner, M. O. Bünger, P. S. De-Carvalho, A. Wingler, and E. J. Lucas. 2015. Systematic and evolutionary implications of stamen position in Myrteae (Myrtaceae). *Botanical Journal of the Linnean Society* 179:

388–402.

Vasconcelos, T. N. C., G. Prenner, M. F. Santos, A. Wingler, and E. J. Lucas. 2017. Links between parallel evolution and systematic complexity in angiosperms—A case study of floral development in *Myrcia* s.l. (Myrtaceae). *Perspectives in Plant Ecology, Evolution and Systematics* 24: 11–24.

Vega, Y., I. Marques, S. Castro, and J. Loureiro. 2013. Outcomes of extensive hybridization and introgression in *Epidendrum*(Orchidaceae): Can we rely on species boundaries? *PLoS ONE* 8: e80662.

Wäldchen, J., and P. Mäder. 2018. Machine learning for image based species identification. *Methods in Ecology and Evolution* 9: 2216–2225.

Walter, G. M., R. J. Abbott, A. C. Brennan, J. R. Bridle, M. Chapman, J. Clark, D. Filatov, et al. 2020. Senecio as a model system for integrating studies of genotype, phenotype and fitness. *New Phytologist* 226: 326–344.

Wan, D., Y. Sun, X. Zhang, X. Bai, J. Wang, A. Wang, and R. Milne. 2014. Multiple ITS copies reveal extensive hybridization within *Rheum* (Polygonaceae), a genus that has undergone rapid radiation. *PLoS ONE* 9: e89769.

Wang, L., H. Schneider, X.-C. Zhang, and Q.-P. Xiang. 2012. The rise of the Himalaya enforced the diversification of SE Asian ferns by altering the monsoon regimes. *BMC Plant Biology* 12: 210.

Wang, X., G. Gussarova, M. Ruhsam, N. de Vere, C. Metherell, P. M. Hollingsworth, and A. D. Twyford. 2018. DNA barcoding a taxonomically complex hemiparasitic genus reveals deep divergence between ploidy levels but lack of species-level resolution. *AoB PLANTS* 10.

Wang, Y.-Q., D.-X. Zhang, and Z.-Y. Chen. 2005. Pollination biology of *Alpinia hainanensis* (Zingiberaceae). *Journal of Systematics and Evolution* 43: 37–49.

Waser, N. M., and J. Ollerton. 2006. Geographical variation in diversity and specificity of pollination systems. *In* J. Ollerton, S. D. Johnson, and A. B. Hingston [eds.], Plant-Pollinator Interactions: from specialization to generalization, 283–308. University of Chicago Press.

Weisenfeld, N. I., S. Yin, T. Sharpe, B. Lau, R. Hegarty, L. Holmes, B. Sogoloff, et al. 2014. Comprehensive variation discovery in single human genomes. *Nature Genetics* 46: 1350–1355.

White, T., T. Bruns, S. Lee, and J. Taylor. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *In* M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White [eds.], PCR protocols: a guide to methods and applications, 315–322. Academic Press, San Diego, USA.

Wickham, H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Wiens, J. J., and M. J. Donoghue. 2004. Historical biogeography, ecology and species richness. *Trends in Ecology & Evolution* 19: 639–644.

Wood, T. E., N. Takebayashi, M. S. Barker, I. Mayrose, P. B. Greenspoon, and L. H. Rieseberg. 2009. The frequency of polyploid speciation in vascular plants. *Proceedings of the National Academy of Sciences* 106: 13875–13879.

Wright, S. I., S. Kalisz, and T. Slotte. 2013. Evolutionary consequences of self-fertilization in plants. *Proceedings. Biological sciences* 280: 20130133.

Wu, D., and K. Larsen. 2000. Zingiberaceae. *In* Z. Y. Wu, and P. H. Raven [eds.], Flora of China, 333–346.

Wu, W., Y.-L. Yang, W.-M. He, M. Rouard, W.-M. Li, M. Xu, N. Roux, and X.-J. Ge. 2016. Whole genome sequencing of a banana wild relative *Musa itinerans* provides insights into lineage-specific diversification of the *Musa* genus. *Scientific Reports* 6: 31586.

Wyse, S. V., J. B. Dickie, and K. J. Willis. 2018. Seed banking not an option for many threatened plants. *Nature Plants* 4: 848–850.

Xia, E., F. Li, W. Tong, H. Yang, S. Wang, J. Zhao, C. Liu, et al. 2019. The tea plant reference genome and improved gene annotation using long-read and paired-end sequencing data. *Scientific Data* 6: 122.

Xia, Y.-M., W. J. Kress, and L. M. Prince. 2004. Phylogenetic analyses of *Amomum* (Alpinioideae: Zingiberaceae) using ITS and matK DNA sequence data. *Systematic Botany* 29: 334–344.

Xue, B., Y. C. F. Su, D. C. Thomas, and R. M. K. Saunders. 2012. Pruning the polyphyletic genus *Polyalthia* (Annonaceae) and resurrecting the genus Monoon. *TAXON* 61: 1021–1039.

Yang, Y.-W., Z.-G. Qian, A.-R. Li, C.-X. Pu, X.-L. Liu, and K.-Y. Guan. 2016. Differentiation in fructification percentage between two morphs of *Amomum*

tsaoko (Zingiberaceae). *Breeding science* 66: 391–395.

Yasui, Y., M. Mori, J. Aii, T. Abe, D. Matsumoto, S. Sato, Y. Hayashi, et al. 2012. S-LOCUS EARLY FLOWERING 3 is exclusively present in the genomes of short-styled buckwheat plants that exhibit heteromorphic self-incompatibility. *PLoS ONE* 7.

Yost, J. M., and K. M. Kay. 2009. The evolution of postpollination reproductive isolation in *Costus*. *Sexual Plant Reproduction* 22: 247–255.

Younger, J. L., L. Strozier, J. D. Maddox, Á. S. Nyári, M. T. Bonfitto, M. J. Raherilalao, S. M. Goodman, and S. Reddy. 2018. Hidden diversity of forest birds in Madagascar revealed using integrative taxonomy. *Molecular Phylogenetics and Evolution* 124: 16–26.

Zanella, C. M., C. Palma-Silva, M. Goetze, and F. Bered. 2016. Hybridization between two sister species of Bromeliaceae: *Vriesea carinata* and *V. incurvata*. *Botanical Journal of the Linnean Society* 181: 491–504.

Záveská, E., T. Fér, O. Šída, K. Krak, K. Marhold, and J. Leong-Škorničková. 2012. Phylogeny of *Curcuma* (Zingiberaceae) based on plastid and nuclear sequences: Proposal of the new subgenus *Ecomata*. *TAXON* 61: 747–763.

Záveská, E., T. Fér, O. Šída, K. Marhold, and J. Leong-Škorničková. 2016. Hybridization among distantly related species: Examples from the polyploid genus *Curcuma* (Zingiberaceae). *Molecular Phylogenetics and Evolution*.

Zhang, L., Q. J. Li, X. B. Deng, P. Y. Ren, and J. Y. Gao. 2003. Reproductive biology of *Alpinia blepharocalyx* (Zingiberaceae): Another example of flexistyly. *Plant Systematics and Evolution* 241: 67–76.

Zhao, N., G. Wang, A. Norris, X. Chen, and F. Chen. 2013. Studying plant secondary metabolism in the age of genomics. *Critical Reviews in Plant Sciences* 32: 369–382.

Zizka, A., A. Antonelli, and D. Silvestro. 2020. sampbias, a method for quantifying geographic sampling biases in species distribution data. *bioRxiv*: 2020.01.13.903757.

Appendices

The files listed below have been uploaded separately on DataSync.

File name	Contents
Appendix2.1_Full_responses.xlsx	Individual responses for the 'Taxonomic complexity in the tropics' survey
Appendix2.2_Survey_questions_summary	Appendix 2.2.1: A list of questions from the survey titled 'Taxonomic complexity in the tropics.' Appendix 2.2.2: Summary of results for Q1-Q7 from the survey
Appendix4.1_crosses_accession_details.xlsx	List of all the accessions from the RBGE used for crosses.
Appendix4.2_Crosses_RBGE_all.xlsx	A list of all the *Alpinia* crosses done in 2017-2018.
Appendix4.3_seedset_all.xlsx	A list of all the successful *Alpinia* crosses with their seed set.
Appendix4.4_Seeds_germinated_proporton.xlsx	A list of all the accessions from 2017 with their germination proportion.
Appendix4.5_genetic_distance_ITS.xlsx	A list of all the pairwise genetic distances based on the neighbour joining tree in in Geneious v. 8.1.9
Appendix5.1_Flexistyly_tissues_Pakke_2018.xlsx	Collection details of the anaflexistylous and cataflexistylous tissues from Pakke Tiger Reserve, Arunachal Pradesh, India.
Appendix5.2_mapping_depth_comparison_flexistyly	A comparison of the paired mapping depth per scaffold of the anaflexistylous and cataflexistylous pool of the 200,000 longest scaffolds.