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

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

Alexander D. Twyford

Doctor of Philosophy
The University of Edinburgh
Royal Botanic Garden Edinburgh

2012
Declaration

I hereby declare that the work contained in this thesis is my own, unless otherwise acknowledged and cited. This thesis has not in whole or in part been previously presented for any degree.

Alexander D. Twyford

13th August, 2012
Abstract

*Begonia* L. is one of the largest plant genera, comprising over 1500 species. Weak species cohesion, and the rapid evolution of reproductive barriers in allopatry, are two processes that have been postulated to explain the generation of such hyper-diversity of taxa within a single genus of plants. The aim of this thesis is to investigate whether these factors are likely to have been important contributors to the diversity of species found in Central American *Begonia*. Species cohesion was analysed in the widespread Central American species *Begonia heracleifolia* and *B. nelumbiifolia*. Interpopulation seed flow was estimated with seven plastid microsatellites. Breeding system estimates and measures of genetic differentiation at nine nuclear microsatellites were used to infer levels of interpopulation pollen flow. Controlled crosses were employed to assess the strength of reproductive barriers both between populations within species, and between species differing in ecology. The potential for gene flow between species in the wild was assessed in natural hybrid zones using molecular markers. Finally a quantitative trait locus (QTL) approach was employed to investigate the genetic basis of reproductive traits that differ between species. No plastid polymorphisms were found in *B. nelumbiifolia*, suggesting it has been through a recent population bottleneck. In contrast, *B. heracleifolia* possessed many plastid haplotypes that were strongly differentiated between populations (*G'_{ST} = 0.829*). Nuclear microsatellites showed high genetic differentiation within species, and both species were self-compatible and self-fertilize at a moderate rate (*B. heracleifolia* *F'_{ST} = 0.506*, *F_{IS} = 0.249*; *B. nelumbiifolia* *F'_{ST} = 0.439*, *F_{IS} = 0.380*). F1s between ecologically similar *B. heracleifolia* and *B. sericoneura* were partly fertile (2-5% seed set), and F1s and early generation backcrosses were found in a hybrid swarm. F1s between *B. heracleifolia* and the ecologically contrasting *B. nelumbiifolia* were pollen sterile, and 3 hybrid swarms showed no evidence of hybrids beyond the F1 generation. Seven QTL were found for reproductive traits, including: sex ratio, pollen sterility and stamen number. The population biology of *Begonia*, with limited seed and pollen dispersal, small population sizes and frequent self-fertilization predisposes them to genetic isolation, increasing the chances that reproductive barriers evolve. These characteristics may underlie the large number of endemics in *Begonia*. 
Acknowledgements

First and foremost I would like to thank my advisors Catherine Kidner and Richard Ennos for their support and guidance throughout the project. Both of them invested a large amount of time with me and taught me about Begonia and evolutionary genetics, and also gave me space to develop many interesting side projects.

I would like to thank friends and colleagues who have made conducting research at the Royal Botanic Garden Edinburgh such an enjoyable experience. Toby Pennington, Pete Hollingsworth and George Argent particularly deserve a mention for sharing with me their scientific knowledge and guidance. Frieda Christie, Alex Clarke, Rhiannon Crichton, Laura Forrest, Michelle Hollingsworth, Ruth McGregor and Markus Ruhsam provided valuable technical support and advice for the lab work. Begonia students past and present have been a great source of knowledge and support, including: Mobina Ali, Nikki Harrison, Sanjeeta Rajbhandry, Daniel Thomas, Saima Umbreen, Stephen Wrigley, as well as the many undergraduate and masters students who have worked in the lab. I have also had a great time working with many other PhD students and visiting scientists at the botanics, including: Bhaskar Adhikari, Kate Armstrong, Maria-Camila Gomez, Rhiannon Crichton, Oswaldo Cruz Neto, Jane Droop, Allan Elliot, Emma Goodyer, Zoe Goodwin, Tobias Marczweski, Danilio Neves, Carmen ‘The Boss’ Puglisi, Sumudu Rubasinghe, Tiina Sarkinen, Harriet Stone and Eugenio Valderama.

I am indebted to those who have contributed original data and materials that were included in the thesis, particularly for the QTL mapping in Chapter 6. Keith Gardner made the inflorescence measures and flower counts as part of a post-doctoral position at the RBGE. Chris White recorded stamen number in Begonia flowers during an undergraduate project. Adrian Brennan supplied two genetic maps of Begonia prior to publication. Nikki Harrison provided plastome sequence data prior to publication, which was used to design microsatellites in Chapter 2. Andrew Matthews measured genome sizes, which are included in Chapter 4. Two anonymous reviewers and the guest editor of the Botanical Journal of the Linnean Society made considerable improvements to chapter 2.
The fieldwork in Mexico would not have been possible, or anywhere near as successful, without the help of many friends and colleagues. In particular, Keith Gardner did much of the planning, and Mark Hughes came to the rescue and helped me through the first month in the field, especially when we were attacked by snakes. Further help planning the fieldwork came from: Marie-Stéphanie Samain, Carolina Granados Mendoza, Hilda Flores, Silvia Salas and Miguel Angel Pérez-Farrera. My other companions in the field were Héctor Domínguez and Arturo Sanchez, who really helped in collecting. Stephan Helfer advized on importing *Begonia* specimens. Additional material for analysis was provided from Fredy Archila, who sent silica dried samples from Guatemala. Neil Watherston, Fiona Inches and Bruce Robertson looked after the very many plants I bought back from collections.

Financial support for the PhD came from the BBSRC. Additional fieldwork funding was provided by The British Ecological Society, The Genetics Society, The MacIntyre *Begonia* Trust and The Merlin Trust.

Finally I would like to thank my parents and family, and Gunnar, for their love and support throughout the project.
TABLE OF CONTENTS

Abstract ........................................................................................................................................... ii
Acknowledgements ........................................................................................................... iii

CHAPTER 1. INTRODUCTION ........................................................................................................ 1
1.1. Speciation and taxonomy ....................................................................................................... 2
  1.1.1. Species concepts and species delimitation ...................................................................... 2
  1.1.2. Genera and higher levels of taxonomic classification .................................................... 4
  1.1.3. Tracing the speciation process from extant plant species .............................................. 5
  1.1.4. Tools for studying the speciation process ....................................................................... 8
    1.1.4.1. Molecular .................................................................................................................. 8
    1.1.4.2. Non-molecular .......................................................................................................... 9
1.2. Speciation histories in large plant genera .......................................................................... 10
1.3. Evolution of diversity in Begonia ....................................................................................... 13
  1.3.1. Dispersal limitation and population structure ................................................................. 17
  1.3.2. Mating system ............................................................................................................... 18
  1.3.3. Outcomes of secondary contact ..................................................................................... 19
  1.3.4. Genetic architecture of reproductive isolating traits ..................................................... 21
1.4. Aims of the doctoral work and structure of the thesis ...................................................... 23
  1.4.1. Study species ................................................................................................................ 25

CHAPTER 2: POPULATION HISTORY AND SEED DISPERAL IN WIDESPREAD CENTRAL AMERICAN BEGONIA SPECIES (BEGONIACEAE) INFERRED FROM PLASTOME-DERIVED MICROSATELLITE MARKERS ......................................................................................... 27
2.1. Chapter summary .................................................................................................................. 27
2.2. Introduction .......................................................................................................................... 28
2.3. Material and methods ......................................................................................................... 32
  2.3.1. Study species ................................................................................................................ 32
2.3.2. Sampling .................................................................................................................. 33
2.3.3. DNA extraction ........................................................................................................ 36
2.3.4. Plastid microsatellite marker design........................................................................ 36
  2.3.4.1. Plastid marker design .......................................................................................... 36
  2.3.4.2. Plastid marker test ......................................................................................... 37
  2.3.4.3. Multiplex PCR test ............................................................................................ 39
2.3.5. Haplotype distributions and relationship among haplotypes .................................. 39
2.3.6. Population genetic diversity and genetic differentiation ........................................... 39

2.4. Results ........................................................................................................................ 41
  2.4.1. Plastid microsatellite marker design ....................................................................... 41
  2.4.2. Haplotype distributions and the relationship among haplotypes ............................ 42
  2.4.3. Population genetic diversity and genetic differentiation ........................................... 45
  2.4.4. Historical barriers to gene flow ............................................................................. 45

2.5. Discussion ................................................................................................................... 46
  2.5.1. Between-species comparisons of genetic diversity ................................................. 46
  2.5.2. The relative role of historical barriers to gene flow ............................................ 48
  2.5.3. Dispersal limitation in Begonia explains current patterns of genetic structure ...... 49
  2.5.4. Plastid microsatellites as tools to study intraspecific relationships ....................... 50

2.6. Conclusion ................................................................................................................... 52

CHAPTER 3: DEVELOPMENT AND CHARACTERISATION OF MICROsatellite MARKERS FOR CENTRAL AMERICAN BEGonia SECTION GIREOudIA (BEgoniaceae) ........................................................................... 53
3.1. Chapter summary .......................................................................................................... 53
3.2. Introduction .................................................................................................................... 53
3.3. Methods and results ....................................................................................................... 54
3.4. Conclusion ..................................................................................................................... 59

CHAPTER 4: GENETIC DIFFERENTIATION AND SPECIES COHESION IN TWO WIDESPREAD CENTRAL AMERICAN BEGonia SPECIES ........................................................................... 61
5.3.2. Experimental estimates of reproductive barriers ......................................................... 88
  5.3.2.1. Phenology ............................................................................................................ 88
  5.3.2.2. Chromosome numbers ......................................................................................... 89
  5.3.2.3. Crossing barriers and fitness of hybrids .............................................................. 90
5.3.3. Extent of gene exchange in hybrid swarms ............................................................... 92
  5.3.3.1. B. heracleifolia x B. sericoneura hybrid swarm .................................................. 92
  5.3.3.2. B. heracleifolia x B. nelumbiifolia hybrid swarms ................................................ 92
  5.3.3.3. Nuclear genotyping ........................................................................................... 93
  5.3.3.4. Hybrid swarm analyses for nuclear markers ...................................................... 93
  5.3.3.5. Simulated hybrid swarms ................................................................................... 94
  5.3.3.6. Inbreeding .......................................................................................................... 95
  5.3.3.7 Plastid analyses .................................................................................................. 95
5.4. Results ............................................................................................................................ 96
  5.4.1. Barriers to hybridization between B. heracleifolia and B. sericoneura ..................... 96
    5.4.1.1. Experimental assessment of reproductive barriers ............................................... 96
    5.4.1.2. Crossing barriers and fitness of hybrids ............................................................ 96
    5.4.1.3. Analysis of hybridization in nature .................................................................... 99
    5.4.1.4. Accuracy of assignment in simulated hybrid swarms ........................................ 101
    5.4.1.5. Plastid genotyping and direction of hybridization ............................................... 101
  5.4.2. Barriers to hybridization between B. heracleifolia x B. nelumbiifolia ...................... 101
    5.4.2.1. Experimental assessment of reproductive barriers ............................................... 101
    5.4.2.2. Crossing barriers and fitness of hybrids ............................................................ 102
    5.4.2.3. Analysis of hybridization in nature .................................................................... 103
    5.4.2.4. Accuracy of assignment in simulated hybrid swarms ........................................ 106
    5.4.2.5. Plastid genotyping and direction of hybridization ............................................... 106
5.5. Discussion ....................................................................................................................... 107
  5.5.1. Evolution of reproductive isolation in Begonia ......................................................... 107
  5.5.2. Experimental limitations of hybrid swarm comparisons ........................................... 109
  5.5.3. Reproductive barriers in Begonia ........................................................................... 110
  5.5.4. Evolutionary outcomes of hybridization in Begonia ............................................... 111
CHAPTER 6: QUANTITATIVE TRAIT LOCUS (QTL) ANALYSIS OF SEX RATIOS AND INFLORESCENCE ARCHITECTURES IN BACKCROSS BEGONIA POPULATIONS ............................................................ 115

6.1. Chapter summary ................................................................................................................. 113

6.2. Introduction .......................................................................................................................... 113

6.3. Material and methods ........................................................................................................... 115

   6.3.1. Study species and the generation of backcross populations ........................................ 115
   6.3.2. Phenotyping .................................................................................................................. 118
   6.3.3. Segregation patterns and quantitative analyses ........................................................... 120
   6.3.4. QTL analyses ............................................................................................................... 121
   6.3.5. Candidate gene detection .............................................................................................. 122

6.4. Results ................................................................................................................................. 122

   6.4.1. Trait segregation .......................................................................................................... 122
   6.4.2. QTL mapping ............................................................................................................... 125
   6.4.3. Candidate genes .......................................................................................................... 128

6.5. Discussion ............................................................................................................................. 131

   6.5.1. QTL for reproductive traits in Begonia mapping populations ....................................... 131
   6.5.2. Candidate genes for inflorescence development in Begonia ......................................... 133
   6.5.3. Adaptive significance of inflorescence architectures ................................................... 135

CHAPTER 7: GENERAL CONCLUSIONS ......................................................................................... 137

7.1. Evolution of diversity in Begonia ......................................................................................... 137

   7.1.1. Comparison with Senecio .......................................................................................... 139
   7.1.2. Genetic resources for future studies ........................................................................... 143

7.2. Future questions ................................................................................................................. 144

   7.2.1. What is the colonization history of B. heracleifolia and B. nelumbifolia in Central America? .................................................................................................................. 144
   7.2.2. How rapidly has speciation occurred in Central American Begonia? ......................... 144
   7.2.3. Which genes contribute to the early stages of speciation in Begonia? ......................... 145
   7.2.4. Why are the sister genus and related families so species poor? .................................. 145
   7.2.5. Has polyploidy promoted diversification in Begonia? ................................................ 145
7.2.6. Does hybridization play an evolutionary important role in *Begonia*? ....................................... 146
7.2.7. Does ecological divergence promote speciation in *Begonia*? .................................................. 146
7.3. Conclusion ........................................................................................................................................ 147

8. REFERENCES ..................................................................................................................................... 148

9. APPENDICES ..................................................................................................................................... 147
CHAPTER 1: Introduction

There has long been an interest in plant genera and families that are species rich. The many variations in form on a common theme appeal to gardeners and plant collectors, and scientific interest is engaged by the evolutionary processes that underlie such a distinct pattern. Very large groups include the daisy family (Asteraceae), with around 23,000 species, and the orchid family (Orchidaceae), with over 22,000 species (Heywood et al., 2007). Even though these large plant groups are extremely familiar, species rich groups are actually uncommon. A plot of the number of plant species per genus demonstrates the scarcity of species rich groups, as well as the commonality of species poor groups. This pattern has been called the ‘hollow curve’ (Willis, 1922; Fig. 1.1).

![Figure 1.1. The hollow curve of genus sizes.](image)

Expert opinions differ on the way species are delimited. Some experts divide taxa differing in only a few morphological characters (‘splitters’), while others combine broadly similar taxa under one name (‘lumpers’) (Endersby, 2009). This raises the concern that the hollow curve may be a taxonomic artefact, caused by experts delimitating species (and genera) differently. Another
explanation for the hollow curve would be variation in the speciation or extinction rates between lineages.

This thesis examines speciation processes in an exceptionally specious genus. Species delimitation and generic boundaries are discussed first, as these are essential for understanding whether species rich genera are artefacts or are caused by evolution processes. The evolutionary processes that may give rise to species rich genera are then discussed, as well as the ways these can be studied. Examples of species rich genera, and the processes that have shaped them are then considered, before introducing the study system for the thesis.

1.1. Speciation and taxonomy

1.1.1. Species concepts and species delimitation

The use of species as universal biological units in evolutionary research is controversial (Coyne & Orr, 2004). Some researchers have argued that species are meaningful units that can be delimited using simple criteria (Rieseberg et al., 2006), while others agree that species are real but question whether the same criteria can be applied equally to all species groups (Mishler & Donoghue, 1982). A different point of view is that species are human constructs imposed onto the continuum of biological diversity and are not meaningful units (Ehrlich & Raven, 1969; Bachmann, 1998). The ongoing disagreement over species as biological units results in a wide range of ways to delimit species boundaries.

The biological species concept (Mayr, 1942; Mayr, 1969) describes a species as interbreeding populations that are reproductively isolated from other such groups. In this case, reproductive isolation describes any barrier to gene flow between species, whether it is prior to cross-fertilization (prezygotic), or after fertilization has occurred (postzygotic). One strong criticism of the biological species concept is that it ignores the high frequency of interspecific hybridization in nature (Stace, 1975). It also ignores the high rate of interspecific gene flow that occurs during speciation (Machado & Hey, 2003; Nosil, 2008). More recent species concepts address concerns raised with the biological species concept (summarised in Mallet, 1995; Wu, 2001; Coyne &
CHAPTER 1: Introduction

Orr, 2004; Hausdorf, 2011), although none have been universally accepted. One example is the differential fitness species concept (Hausdorf, 2011), which defines species as ‘groups of individuals that are reciprocally characterised by features that would have negative fitness effects in other groups and that cannot be regularly exchanged by groups upon contact’. This is a more accurate approximation of a species, adequately accounts for hybridization, and also covers the early stages of speciation. However, species delimitation from this concept would be far more complex than the biological species concept, as it would require the fitness of many hybrids to be studied.

Morphology alone is usually the sole criteria for species delimitation. Rieseberg et al. (2006) found that morphological groupings inferred from phenotypic clustering generally accord with reproductive isolation inferred from crossing barriers, supporting morphological approaches to species delimitation. Morphology is a useful proxy for species barriers as most species have been isolated for sufficient time for morphological differences to arise, and extinction of related species will further increase morphological distinction. One concern with morphological delimitation is that these traits may not be independent, due to the influence of pleiotropy or linked genes (Doyle, 1992). The study by Rieseberg et al. (2006) shows that non-independence of characters does not typically have a strong impact on species delimitation.

Groups where there is difficulty delimiting species using morphological characters include taxonomically complex groups, recent species radiations, and cryptic species. Taxonomically complex genera are characterised by asexual reproduction and hybridization (Ennos et al., 2005) including Euphrasia L. (Orobanchaceae) and Hieracium L. (Asteraceae). While species can be found that differ in morphology, these often grade into each other making delimitation difficult. Species of recent origin will differ in few morphological characters, such as Mediterranean sexually deceptive orchids (Scopece et al., 2010), and these also pose challenges for species delimitation (Shaffer & Thomson, 2007). Finally, species that are nearly morphologically indistinguishable from each other but differ in other characteristics suggesting they are separate species (cryptic species), pose major difficulties for taxonomists. One example is the liverwort genus Herbertus Gray (Herbertaceae), where DNA barcoding has revealed genetically distinct lineages within what was considered a single species, and morphological characters have now been found to support the genetic groups (Bell et al., 2012). In these cases where choosing
morphological characters to discriminate species is difficult, a more robust approach to species delimitation is to focus on multiple direct and indirect lines of evidence (e.g. crossing barriers, overlap in gene pools, differing morphologies) for reproductive barriers, which are common to all species concepts (Sites & Marshall, 2004).

While species delimitation is difficult in some complex groups, in most cases species can be recognised (Rieseberg et al., 2006). However, the existence of species as evolutionary units is more important than their place in the taxonomic hierarchy. Species are a cohesive unit composed of populations that are connected by gene flow, and where new alleles arise through mutation these will be spread between populations, uniting populations within a species. This differs from higher taxonomic units such as genera, where interspecific gene flow is low compared to intraspecific gene flow (Hey & Pinho, 2012), and this tends to be restricted to secondary contact zones at range margins (Willyard et al., 2009).

1.1.2. Genera and higher levels of taxonomic classification

The higher level organisation of species into genera, and genera into families, is at the core of systematics and taxonomy. However, the delimitation of such groups is controversial (Walters, 1986). Part of this controversy is because no set of criteria define hierarchical taxonomic ranks, such as genera. There are, however, two general conditions that should be met. A genus should be monophyletic, i.e. have a single common origin when the evolutionary relationships of species are assessed with molecular data (Hennig, 1966; Bremer & Wanntorp, 1978). Second, species in a genus should share a common set of morphological characters that differ from those of other genera (Stebbins, 1956). These criteria for generic limits are evidently arbitrary (Stebbins, 1956), and genera will differ in their age, size and genetic diversity. An exceptional contrast in size and age is between the monotypic genus *Ginkgo* Rich. (in the monotypic family Ginkgoaceae) that diverged from cycads ~300Ma (Pryer et al., 2004), and *Astragalus* Medik. with ~2 500 species originating ~12Ma (Sanderson & Wojciechowski, 1996).

The subjectivity of genera does not invalidate them as a unit of comparison. Many genera have been relatively stable over taxonomic history (Frodin, 2004) suggesting that at least morphologically, genera represent cohesiveness in adaptive space, and are therefore biologically
relevant units. Most lineages have diversified within a single biome rather than switching between them (phylogenetic niche conservatism, Pennington et al., 2009), so genera tend to be distinct in their ecology and distribution too. Therefore genera will unite species that have retained ancestral morphology and often ecological adaptations too, and where lineages have diversified extensively from their ancestral phenotype, they are more likely to be put in their own genera.

Uncertainty as to whether a group should be ranked a genus or a family can cause concerns as to whether species richness reflects speciation histories or taxonomic artefacts. However, the hierarchical level of study does not bias patterns of species richness, as similar trends of species richness emerge when the family rather than the genus is used as a unit of comparison (Clayton, 1974). For example, a similar ‘hollow curve’ is obtained when the number of genera per family, rather species per genus, are plotted (Clayton, 1974). Moreover, seven of the ten largest plant genera are contained within the ten largest plant families (e.g. Senecio L. in the Asteraceae, Bulbophyllum Thouars. in the Orchidaceae; www.mobot.org/MOBOT/research/APweb/; discussed below), suggesting many of the properties that have promoted speciation have been present over long periods of time. Therefore, while the genus is a much debated unit of classification, it does represent a biologically meaningful group that can be used as a unit of comparison when the reservations above are taken into account.

1.1.3. Tracing the speciation process from extant plant species

The most direct way to understand evolutionary histories of plants is by studying fossils; but individual fossils have a low information content, and do not tell us about the factors that contributed to speciation (Smith, 2009). Evolutionary biologists more typically study extant groups and use them to infer the forces behind past and current speciation events (Barraclough et al., 1998). A population genetic approach is important for understanding speciation histories, as speciation is one potential outcome when gene flow is reduced between two divided populations (Wright, 1931; Slatkin, 1987; Abbott et al., 2008).

The division of one continuous, randomly breeding (panmictic) population, into subdivided populations, is the first stage in the speciation process. Under these conditions, an individual is
more likely to mate with another in the same population than one from another population. This means alleles are more likely to be shared within rather than between populations, and this division of genetic diversity can be summarised by the statistic $F_{ST}$ (Wright, 1943; Slatkin, 1977; Weir & Cockerham, 1984). Factors that will increase the proportion of genetic diversity found between populations (increase the $F_{ST}$) are genetic drift and natural selection. These differentiation processes are opposed by interpopulation gene flow (migration), which homogenises allele frequencies between subdivided populations, decreasing the $F_{ST}$. When gene flow is reduced to less than one migrant per generation, population-specific mutations will accumulate between ancestrally shared gene copies (Wright, 1931), and genetic differentiation can be measured with genetic distance measures such as $D$ (Nei, 1978).

Characterising the population substructure by calculating $F_{ST}$, and genetic differentiation by calculating Nei’s $D$, are informative ways of studying ongoing and recent evolutionary events. Understanding the strength of interpopulation gene flow can reveal the geographic scale over which speciation can occur (Fig. 1.2). Moreover, if the ongoing factors that affect genetic differentiation are assumed to be the same as those that have promoted divergence in the past, we can learn about past speciation histories from extant species (Barraclough et al., 1998). For example, if related plant species have a genetic signature indicating seed dispersal limitation, one may infer that limited seed dispersal contributed to divergence in the past. In contrast, widespread species with most genetic diversity shared between populations (low $F_{ST}$ values) may indicate gene flow between populations is common, and therefore past speciation events are more likely to be caused by strong geographic barriers than dispersal-limitation between distant populations. However the cause of geographic structuring indicated by $F_{ST}$ values can be interpreted in different ways. For example, the species with low $F_{ST}$ described above may have recently expanded its range, rather than being present for a long time at each site and having widespread gene flow between populations (Hey & Machado, 2003). In this case, calculating genetic diversity in each subpopulation ($h_s$), and analysing the number of unique genetic clusters in the sampled populations (e.g. with STRUCTURE, Pritchard et al., 2000), can help discriminate between causes of structuring within natural populations.
Figure 1.2. Minimum geographic scale that speciation can occur in different organisms. From a meta-analysis of island size and extent of gene flow by Kisel & Barraclough (2010).

Studies of secondary contact zones can also shed light on speciation histories. One may expect few species-specific markers (i.e. markers with fixed differences in allopatric populations) to be shared between species in areas of sympatry (Rieseberg & Ellstrand, 1993) due to reproductive barriers that prevent hybridization. However, hybridization can occur between species when speciation is generally judged to be complete (Hausdorf, 2011). Therefore, rates of introgression may be compared across the genome, which may identify genomic islands of reproductive isolation (Turner et al., 2005; Feder & Nosil, 2010; Nosil & Feder, 2012) where introgression is impeded (Wu & Ting, 2004). Typically these islands include genes for species-specific adaptations (Butlin, 2010). Understanding the size of these islands, and the genes contained within them, may help us understand the evolution of reproductive isolation (Yatabe et al., 2007; Strasburg et al., 2012).
The process of speciation will often leave a clear genetic signature in extant plant populations. If speciation is occurring, strong patterns of population substructure and genetic differentiation would be expected. Studying introgression at species specific markers will show the strength of reproductive isolation, while genome-wide analysis with densely mapped markers will show the parts of the genome that contribute to isolation. These studies can give great insights into speciation processes, as long as suitable molecular markers are available.

1.1.4. Tools for studying the speciation process

1.1.4.1. Molecular

The increasing availability of molecular markers makes the approaches described above, such as genome scans, feasible for all plant species, not just evolutionary model systems that have large-scale genomic resources available (Ekblom & Galindo, 2011). This is largely due to next-generation sequencing (NGS) techniques, which generate a large amount of DNA sequence data at a reduced cost relative to Sanger sequencing (Metzker, 2010). The low costs make it possible to scan genomic DNA for single nucleotide polymorphisms (SNPs) in the study organism, and polymorphic sites can then be sequenced in a large numbers of individuals (Buggs et al., 2010) using sequencing arrays such as Nimblegen AccuSNP (Roche, Madison, USA) or KASP (KBioScience, Hoddesdon, UK), reviewed in Seeb et al. (2011).

Previous to the development of next-generation sequencing, amplified fragment length polymorphisms (AFLPs, Vos et al., 1995) were the principle method for large-scale genotyping. In this technique, DNA is digested with a rare-cutting restriction enzyme, the fragments amplified using the polymerase chain reaction (PCR), and the presence or absence of amplified fragments scored. The markers are dominant, which limits the type of analyses that can be done (such as calculating the inbreeding coefficient) and the information content of individual loci is low (Meudt & Clarke, 2007). However, many loci are typically generated from a single primer-combination, making them well-suited to genetic mapping and other techniques where many loci are more important than levels of polymorphism.
A third common marker type are microsatellites (also known as simple sequence repeats, SSRs, Morgante & Olivieri, 1993). These short tandem repeat sequences are amplified by PCR, and are particularly useful when only a modest number of highly polymorphic loci (5-15) are needed to answer a biological question. Microsatellites developed from enriched libraries tend to be species-specific, and take a long time to develop (Squirrel et al., 2003). Using sequence libraries from next-generation sequencing of expressed genes (transcriptomes) or genomes makes the development of markers cheaper and quicker, and the primer sequences derived from transcriptome sequences are usually conserved between related species (Varshney et al., 2005; Guichoux et al., 2011; Lepais & Bacles, 2011). The continued popularity of microsatellite markers is mostly due to their co-dominance, revealing both alleles at a locus, which can be used to estimate the inbreeding coefficient (F_{IS}). Levels of polymorphism are also high (Zane, 2002). A major issue with microsatellites is null alleles, the non-amplification of an allele caused by mutations in the primer region, causing an individual to appear homozygous when they are heterozygous (Pemberton et al., 1995). Their mutational complexity is also more difficult to model than DNA sequence data. Either each mutation is assumed to increase or decrease the microsatellite by one repeat unit in length (the stepwise mutation model, Ohta & Kimura, 1973; Valdes et al., 1993, used in calculating R_{ST}, Slatkin, 1995), which is not always the case (Micheneau et al., 2010), or the expected relationship between microsatellite length is ignored and only the allelic state is considered (the infinite alleles model, Ohta & Kimura, 1973; Moran, 1975, used in calculating F_{ST}, Slatkin, 1987). In either case, the high mutation rate of microsatellites can make homoplasy an issue (Estoup et al., 2002).

1.1.4.2. Non-molecular

Multiple hypotheses may be proposed to explain the distribution of genetic variation in natural populations, and additional non-molecular experiments can help distinguish between these hypotheses. For example, 100 AFLP markers may show there is little introgression in a hybrid swarm, but cannot distinguish which reproductive barriers are responsible. The reproductive barriers may be investigated by using artificial cross pollinations and measures of fitness in common garden and reciprocal transplant experiments (Burke & Arnold, 2001). Experimental crosses show which genotypes are compatible independent of other reproductive barriers that interact in nature, such as pollinator preferences. The fitness of natural hybrids and those
produced in experimental crosses can be assessed under common garden conditions, allowing hybrid fitness to be analysed under controlled environmental conditions, or using reciprocal transplants, which allows the fitness of hybrids to be measured directly in conditions in which the plants are known to grow (Emms & Arnold, 1997). Another example where such experiments would be useful is to identify the contribution of selection against genetic drift in patterns of genetic differentiation. Reciprocal transplants could show the fitness of genotypes in other environments, to test for local adaptation (Ågren & Schemske, 2012). Local adaptation could also be inferred by comparing morphological trait variation within and between populations with \( Q_{ST} \) analysis (\( F_{ST} \) analogue using trait variation), and comparing this to the genetic structure inferred by \( F_{ST} \) (Spitze, 1993; Waldmann & Andersson, 1999; Steane et al., 2006). A significantly higher \( Q_{ST} \) than \( F_{ST} \) would be consistent with local adaptation, although this approach may be inappropriate when mutation rates at neutral markers are high, or when the adaptive role of traits have not been well characterised (Pujol et al., 2008; Edelaar et al., 2011).

The preceding discussion of species delimitation, generic boundaries, and how the speciation process can be studied, shows many interacting factors that make speciation biology complicated to study. While certain genetic signatures are expected when speciation is occurring, it must also be recognised that speciation is not predictable in its tempo and mode, and that ‘reverse speciation’ can occur (i.e. elevated gene flow merging two species, Taylor et al., 2006). Studies of gene flow (described above) are a powerful way to study speciation histories, but are best set in the context of multiple different types of experiments to aid interpretation when many factors interact. This is particularly the case in large or taxonomically complex groups, where reliance on a single approach may not lead to reliable species delimitation.

1.2. Speciation histories in large plant genera

The ten largest plant genera, listed in order of ascending species number, are: *Astragalus* L. (Leguminosae), *Bulbophyllum* (Orchidaceae), *Psychotria* L. (Rubiaceae), *Euphorbia* L. (Euphorbiaceae), *Carex* L. (Cyperaceae), *Begonia* L. (Begoniaceae), *Dendrobium* Sw. (Orchidaceae), *Acacia* Mill. (Leguminosae), *Solanum* L. (Solanaceae) and *Senecio* (Asteraceae)
CHAPTER 1: Introduction

(Frodin, 2004). Genus-wide studies, such as phylogenies or monographs, are rare for these species rich genera because of their unmanageable sizes. This makes comparisons between them difficult, and these tend to be made on infragenic groups that can be studied in a reasonable amount of time (Bohs & Olmstead, 1997; Pelser et al., 2002). It is even less likely that researchers will study related genera too, which is necessary to test if intrinsic properties of the group underlie their diversification. A rare example is a study of *Astragalus*, where elevated speciation rates were found not to be unique to the genus, but were similar to other species rich genera in the Astragalean clade (Sanderson & Wojciechowski, 1996).

Most of the ten largest plant genera are most species rich in the tropics (Davies & Barraclough, 2007; Mittelbach et al., 2007). A meta-analysis of dated radiations in different plant clades by Linder (2008) showed that species diversification rates are strongly related to geography, where recent rapid radiations (i.e. increased speciation) often explain diversity in areas such as the Andes (e.g. lupins, Hughes & Eastwood, 2006), and persistence of past radiations caused by biome stability (i.e. reduced extinction) explain much of the diversity in areas such as Australia. However, the pattern of high species richness in certain areas does not tell us the factors that have promoted speciation (Weir & Schluter, 2007). The presence of *Carex*, an exclusively temperate genus, in the list of highly diverse plant genera, shows that non-tropical species can also obtain high levels of diversity. In *Carex*, diversification rates have been modest, and low rates of extinction explain its high extant species richness (Escudero et al., 2012). Evidence for rapid species radiations, in addition to low rates of extinction, have also been found in temperate areas. Valente et al. (2010) showed the highest rates of species diversification in plants described so far were not in well known tropical radiations (e.g. *Inga* Mill., Richardson et al., 2001), but in European *Dianthus* L. The relationship between extant species richness and geography is therefore complex, and requires other factors to be considered.

Sanderson & Wojciechowski (1996) proposed 4 main factors common to the largest plant genera, which are likely to have influenced their levels of species richness. These are: strong population substructure, herbaceous habit, rapid chromosome evolution and frequent transitions between ecological specializations. Each of these factors influences speciation in a different way. Many species in these genera have limited gene flow that has caused highly substructured populations (e.g. *Astragalus cremnophylax* Barneby, $F_{ST} = 0.41$, Travis et al., 1996; 4 species of
CHAPTER 1: Introduction

*Euphorbia* $F_{ST} = 0.237 - 0.652$, Park, 2004; *Dendrobium officianale* Kimura & Migo, $F_{ST} = 0.269$, Li *et al*., 2008; *Acacia raddiana* Savi, $F_{ST} = 0.60$, Shrestha *et al*., 2002. Low levels of gene flow between populations leads to reduced species cohesion, and in turn may lead to allopatric divergence. This is supported by the limited distribution ranges of many species in large genera (Domínguez Lozano & Schwartz, 2005). The predominantly herbaceous growth form common in these groups means species have a short generation time relative to long-lived woody species, so mutation rates will be higher (Duminil *et al*., 2009). Frequent chromosomal rearrangements, including whole genome duplication (polyploidy), increases the genetic diversity on which selection may act and may lead to diversification (Levin, 1983; Rieseberg *et al*., 2006; Vamosi & Dickinson, 2006; Rieseberg & Willis, 2007). Changes in chromosome number, or large duplications and insertions, will also lead to strong reproductive barriers between incipient species (Ruhsam, in prep.; Soltis & Soltis, 2009). This is seen in *Carex*, where species have almost every even chromosome number between $2n = 12$ and $2n = 124$, and over 100 species have intraspecific cytological variation (Hipp *et al*., 2009; Chung *et al*., 2011; Escudero *et al*., 2012). Frequent transitions between ecological specializations will cause strong prezygotic reproductive barriers, as species are unlikely to grow in sympatry. Such frequent transitions have been revealed by mapping ecology onto a species-level phylogeny of *Euphorbia*, which revealed sister-species often grow in different habitats (Frajman & Schönswetter, 2011; Horn *et al*., 2012).

Many large genera also have unusual morphological characters, which some have considered ‘key innovations’ leading to the success of a group. For example, *Senna* Mill. (Leguminosae) species often possess extra floral nectaries, and *Senna* have diversified at a greater rate than other legume clades without extra floral nectaries (Marazzi & Sanderson, 2010). Extra floral nectaries may be a key innovation as they attract ants that protect the plants. Moreover, diversification in *Senna* occurred at the same time as diversification in ants (Marazzi & Sanderson, 2010). In many other cases however, the importance of traits that may be key innovations is hard to test, and other factors may better explain extant levels of species richness. For example, *Carex* has a sac-like structure that covers a one-seeded fruit (perigynium or utricle) that may aid dispersal (Escudero *et al*., 2012), but frequent chromosomal rearrangements may play a bigger role in the evolutionary success of the genus. In many cases processes at or below the species level may have caused the break down of species cohesion and contributed to
diversification, and key innovations may have played no significant role (Davies & Barraclough, 2007).

1.3. Evolution of diversity in *Begonia*

Most evolutionary studies in plants are confined to evolutionary model systems, which have been selected to study different evolutionary processes, such as: *Helianthus* L. for studying hybrid speciation and the evolution of reproductive isolation (Rieseberg, 1991; Rieseberg *et al.*, 1999; Rieseberg, 2000; Yatabe *et al.*, 2007; Scascitelli *et al.*, 2010; Whitney *et al.*, 2010), *Iris* L. for studying the adaptive significance of hybridization (Carney *et al.*, 1994; Arnold & Hodges, 1995; Emms & Arnold, 1997; Martin *et al.*, 2007; Arnold *et al.*, 2010), European *Senecio* for studying introgression and the origin of invasive species (Abbott *et al.*, 1992; Hiscock, 2000; Lowe & Abbott, 2004; Hegarty *et al.*, 2008; Kim *et al.*, 2008; Chapman & Abbott, 2010), *Tragopogon* L. for studying polyploidy and its adaptive significance (Soltis *et al.*, 2004; Buggs *et al.*, 2010), *Silene* L. for the evolution of sex chromosomes (Charlesworth, 1991; Charlesworth, 2002; Charlesworth *et al.*, 2005) and *Mimulus* L. for ecological and functional genetics (Schemske & Bradshaw, 1999; Fishman & Willis, 2001; Hall *et al.*, 2006; Sweigart *et al.*, 2006; Fishman & Willis, 2007; Case & Willis, 2008; Wu *et al.*, 2008; Martin & Willis, 2010).

Tropical herbaceous plants are an important species group that are not represented by current model plant species. The tropics are the most species rich region on Earth, with the Neotropics being of particular interest for its species richness (90 000 species); many of these species are in genera and families that are endemic to the region (Gentry, 1992). Between 21 and 47% of tropical plant species are found in the understory (Costa, 2004, and references therein), and developing a tropical herbaceous plant as an evolutionary model system would allow the processes contributing to species richness in these diverse tropical herbaceous genera to be investigated.

In this thesis, speciation processes are investigated in the large genus *Begonia*. We chose *Begonia*, rather than other large plant genera, for the following reasons. First, *Begonia* is
amenable to genetic analysis, with most species being intercrossable, and they produce many seeds per capsule so segregation ratios can be easily tested, and mapping populations made (Neale et al., 2006). Second, Begonia has a large range of morphologies, and there are many transitions in character states between related species, allowing the genetic basis of traits to be assessed with natural independent replicates (Forrest, 2000; Thomas et al., 2011b; Harrison and Kidner, unpubl. data). Third, local monographic revisions and checklists [e.g. checklist of Southeast Asian Begonia, Hughes, 2008; monograph of American section Gireoudia (Klotzsch) A. DC, Burt-Utley, 1985] are available for some groups of Begonia, and many databases (e.g. The Global Biodiversity Information Facility, www.gbif.org; Tropicos, www.Tropicos.org) include occurrence records. Fourth, the distribution of the genus coincides with other large genera, such as Peperomia Ruiz & Pav. (Piperaceae), so evolutionary patterns in Begonia are likely to be of broad consequence for ecologically similar species groups. Finally, there are large research collections at the Royal Botanic Garden Edinburgh (RBGE), making it relatively well-collected and documented considering the size of the genus.

Begonia contains c.1500 species spread throughout the tropics, making it the 6th largest plant genus (Frodin, 2004). This diversity is divided between Central America (690 species), Asia (600 species) and Africa (160 species) (Goodall-Copestake et al., 2010). Phylogenetic analysis of Begonia places African Begonia as early diverging, and Central American and Asian Begonia as more recently derived (Goodall-Copestake et al., 2010). The most recent common ancestor of the genus is likely to have occurred between the end of the Cretaceous and the beginning of the Neogene (Goodall-Copestake et al., 2009).

The great species richness in Begonia contrasts with the only other genus in the family Begoniaceae, Hillebrandia Oliv. (Hughes, 2002). The sole species in this genus, H. Sandwicensis Oliv., is restricted to the Hawaiian Islands (Clement et al., 2004). Hillebrandia differs considerably in appearance to Begonia, as it has strongly differentiated petals and sepals, a semi-inferior ovary, fruits that dehisce between the styles, and a different pollen morphology (Clement et al., 2004). In a broader context, the species richness in the Begoniaceae is higher than many related families in the Curcurbitales, such as the Datiscaceae (2 species) and Tetramelaceae (2 species) (Dewitte et al., 2011; Schaefer & Renner, 2011), although Curcurbitaceae is also species rich (c.800, www.curcubit.org).
Begonia is morphologically recognisable by its asymmetric leaves, unisexual monoecious flowers, generally succulent petioles, twisted papillose stigmas, and dry-three winged capsules (Burt-Utley, 1985; Doorenbos et al., 1998; Thomas, 2010). However, within Begonia there is great morphological diversity, particularly in growth form and leaf shape (simple, peltate, dissected, compound, Neale et al., 2006). This morphological diversity goes hand-in-hand with the wide range of habitats Begonia species inhabit (Fig. 1.3). Most commonly, they grow as understory herbaceous plants, and they have become adapted to these shady conditions where less than 1% of canopy light penetrates (Chazdon & Fletcher, 1984). Begonia species have evolved a distinct leaf micromorphology, with both adaxial and abaxial hypodermis, and this may play a role in harvesting diffuse light (Lee et al., 1990) or give structural support to the leaves (Kidner, pers. comm.). In contrast to these forest floor species, a minority of Central America Begonia have become adapted to very dry soils in sunny exposed positions (Burt-Utley, 1985; Fig. 1.3). The Mexican B. peltata Otto & A. Dietr., is one example, where its succulent leaves covered in a thick indumentum allow it to grow in very dry conditions next to cacti (A.D. Twyford, pers. obs.; Fig 1.3).

Figure 1.3. Representative habitats of species from Central American Begonia section Gireoudia (next page). Top row, left: dry canyon with cacti (B. peltata), right: splash zone adjacent to waterfall (B. multistaminea Burt-Utley); 2nd row, left: seasonally dry forest (B. plebeja Liebm.), right: secondary forest (B. heracleifolia Cham & Schltdl.); 3rd row, left: dry sun-exposed cliff face (B. hydrocotylifolia Otto ex Hook.), right: moist roadside (B. nelumbifolia Cham & Schltdl.); Bottom row, left: wet tropical forest (B. thiemei C.DC. ex Donn.Sm.), right: cloud forest (B. fusca Liebm.).
CHAPTER 1: Introduction
Within *Begonia* there is evidence to support all five factors influencing diversity in large plant genera (population structure, herbageous habit, rapid chromosome evolution, frequent ecological specializations, distinct morphological characters of the genus that may be key innovations, Sanderson & Wojciechowski, 1996). Four factors related to population biology (geographic population structure, mating system) and hybridization (evolution of reproductive traits and outcomes of secondary contact) that may promote speciation are considered in more detail below, which sets the background for the research questions that follow.

### 1.3.1. Dispersal limitation and population structure

*Begonia* have a distinct pollination biology and pattern of seed dispersal. Most *Begonia* species produce simple white or pink flowers, which are visited by generalist pollinators (Apidae, Halictidae and *Trigona* bees, Ågren & Schemske, 1991; Wyatt & Sazima, 2011). Pollinators are attracted to male flowers by a pollen reward, and visitation of female flowers is by deceit, as they produce no floral reward but are similar in appearance to the male flowers (Ågren & Schemske, 1991). It may be expected from these floral syndromes that effective pollen dispersal will be low, as insects will not travel far-a-field to collect pollen for low rewards. Interpopulation seed dispersal may also be low. *Begonia* species produce over 10 000 tiny seeds (300-600µm in length) per capsule, which dehisce from slits along the wing attachments at maturity (Thomas, 2010). In most cases, seeds are not effectively dispersed and they fall to the ground, and germinate *in situ*. This can be seen as clusters of plants in the field (Hoover, 1979).

Plants disperse their genes through pollen and seed dispersal (Ennos *et al.*, 1999). Maternally inherited plastid markers are informative of seed flow, while nuclear and plastid markers are informative for the study of pollen flow (Ennos, 1994; Mccauley, 1995). To date, studies in *Begonia* have only used nuclear markers. These studies have shown that *Begonia* populations have strong population structure, supporting the observation of dispersal limitation of pollen. Matolweni *et al.* (2000) found low interpopulation gene flow between South African populations of *B. dregei* Otto & Dietr. ($F_{ST} = 0.882$) and *B. homonyma* Steudl. ($F_{ST} = 0.937$) using allozyme markers. Hughes & Hollingsworth (2008) investigated gene flow at microsatellite loci in *B. sutherlandii* Hook. f., across fragmented forest patches in South Africa. They found significant population structure, with $F_{ST} = 0.485$. However, these species are all restricted to fragmented
habitats and may not be indicative of widespread species that grow in more continuous habitats. Hughes & Hollingsworth (2003) investigated population structure in *B. socotrana* Hook f., using microsatellite markers. This species is endemic to the island of Socotra off the African coast, and a deviation from panmixia was detected even in the small area the species inhabits (< 15 x 10km) (F<sub>ST</sub> = 0.096).

The likelihood of divergence in allopatry is also affected by other demographic factors. One of these is the effective population size, defined as the idealised number of individuals that would give rise to the calculated sampling variance if they breed in the manner of an idealised population (Wright, 1931; Falconer & Mackay, 1996). Typically, the effective population size is much lower than the actual population size (reviewed in Charlesworth, 2009), due to past fluctuations in population sizes. *Begonia* populations are typically small (Hoover, 1979; Hughes & Hollingsworth, 2008), and therefore particularly prone to genetic drift (Wright, 1931; Kimura, 1955; Willi & Määtanen, 2010). For example, the narrow endemic *B. mazae* Ziesenh. is known from a single population of a few hundred individuals (A.D. Twyford, pers. obs.). While we have limited knowledge of the historical survivorship of *Begonia* populations, it is likely that the high levels of genetic differentiation (e.g. *B. dregei*, Nei’s D = 0.640 and *B. homonyma*, D = 0.520) could only have accumulated if populations are long-lived and stable.

### 1.3.2. Mating system

The mating system (degree of outcrossing) and the sex system (hermaphrodite, male or female; *sensu* Charlesworth, 2006) of a plant species affects the structuring of genetic diversity (Hamrick & Godt, 1996), and subsequently whether allopatric divergence will occur. Most *Begonia* species are monoecious (Tebbit, 2005) so their sex system is mostly uniform across the genus, however, they do vary in their mating system.

Self pollination is common in plants as it provides reproductive assurance, particularly for colonizing species where there may be no other conspecific plants with which to mate (Baker, 1955; Lloyd, 1979). Rounds of self-fertilization will decrease both components of genetic diversity, the heterozgosity and the number of alleles per locus, in comparison with their outcrossing relatives (Charlesworth, 2003; Mable & Adam, 2007). The decrease of
heterozygosity in natural plant populations caused by systematic inbreeding can be calculated as the inbreeding coefficient $F_{IS}$. The loss of genetic diversity in populations that systematically inbreed is due to the reduction in the effective population size (Schoen & Brown, 1991; above).

Population genetic studies in *Begonia* (Matolweni et al., 2000; Hughes & Hollingsworth, 2003; Hughes & Hollingsworth, 2008), as well as other experiments (Ågren & Schemske, 1991; Ågren & Schemske, 1993; Schemske et al., 1996; Wyatt & Sazima, 2011), highlight the potential for self-fertilization and inbreeding in the genus. Selfed crosses produce seed in most (16/18) *Begonia* species tested to date (Ågren & Schemske, 1993; Wyatt & Sazima, 2011). The Cucurbitaceae are not thought to have self-incompatibility mechanisms (Rubino & Wehner, 1986), although Brazilian species *B. integerrima* Spreng. and *B. itatinensis* Irmsch. did not produce seed when selfed in 9 and 4 self-pollinations, respectively (Wyatt & Sazima, 2011). Self-compatibility has led to inbreeding in some *Begonia* species. Consistent deviations from Hardy-Weinberg equilibrium (HWE) across loci happen in wild populations of *B. dregei* ($F_{IS} = 0.273, P < 0.001$), *B. homonyma* ($F_{IS} = 0.576, P < 0.001$; Matolweni et al., 2000) and *B. sutherlandii* ($F_{IS} = 0.158, P < 0.001$; Hughes & Hollingsworth, 2008), and this may be a product of systematic inbreeding. A minor deviation from HWE was also detected in the rare *B. socotrana* ($F_{IS} = 0.051, P < 0.05$; Hughes & Hollingsworth, 2003). Ågren & Schemske (1993) demonstrated low outcrossing rates (<5%) for *B. hirsuta* Aubl. and *B. semiovata* Liebm., at a single allozyme locus. Ågren & Schemske (1993) describe a delayed selfing mechanism in these two *Begonia* species, where female flowers have a short window for outcrossing before the adjacent male flowers opens, which selfs in the absence of a pollinator. This inbreeding has a significant fitness consequence for the progeny, as selfed plants have their biomass reduced by 18-31% and seed set by 12% relative to outcrosses (Ågren & Schemske, 1993).

### 1.3.3. Outcomes of secondary contact

Hybridization in *Begonia* has usually been studied by horticulturists, and over 10 000 *Begonia* hybrid cultivars have been described (Tebbit, 2005). F1 hybrids have been formed between all pair-wise crosses of closely related species from section *Gireoudia*, but crosses between species from different *Begonia* sections, with different chromosome numbers, and different genome
sizes, are less successful (27/156 set seed, 17.3%, Dewitte et al., 2011). Despite the ease with which hybrids have been generated in cultivation, hybrids have rarely been reported in nature (Burt-Utley, 1985; Peng & Chen, 1991; Peng & Sue, 1991; Teo & Kiew, 1999; Peng & Shin-Ming, 2009; Peng et al., 2010). Four natural hybrids in Taiwan have been confirmed by chromosome counts, genomic in situ hybridization (GISH), plastid sequencing, morphology, experimental crosses and by measuring fertility (B. × buimontana, Peng & Chen, 1991; B. × breviscarpa, Peng et al., 2010; B. × chungii, Peng & Shin-Ming, 2009; B. × taipeiensis Peng & Sue, 1991). All four of these hybrids are thought to be F1s, and are sterile. Secondary contact between species where reproductive barriers are strong, such as Taiwanese Begonia, can be considered ‘evolutionary stalemate’ situations, which will not yield evolutionary significant results (Barton & Hewitt, 1985). Stalemate conditions can be caused either by strong premating isolation (e.g. non-overlapping flower times) or postmating isolation (e.g. sterile hybrids), or both. A similar situation between Rhododendron ponticum and R. caucasicum (Ericaceae) is likely caused by selection against later generation hybrids that limits hybridization to the F1 generation (Milne et al., 2003).

The only known naturally occurring Begonia hybrid that is fertile is B. decora x B. venusta, from Peninsula Malasia. The hybrid swarm shows continuous morphological variation between that observed in the parents (Teo & Kiew, 1999), and hybrids have admixed AFLP profiles (Kiew et al., 2003). In cases such as this, where reproductive barriers are weak, there are three possible evolutionary outcomes: clinal introgression, complete admixture, or reinforcement.

Widespread hybridization, and backcrossing or segregation in the hybrids, will result in the movement of genes across a species barrier (introgression). This is a common outcome of secondary contact in plants (Arnold, 1997). The direction of hybridization is influenced by many factors such as the mating system (Ruhsam et al., 2010), relative abundances of the two species (Currat et al., 2008; Lepais et al., 2009), or their position in the local area ( Lexer et al., 2005). Introgression is likely to be uneven across the genome, with hybrid zones acting as selective filters only allowing some genes to pass across a species barrier (Barton & Bengtsson, 1986; Martinsen et al., 2001). These genes may provide a selective advantage, such as the CYCLOIDEA-like genes RAY1 and RAY2, responsible for ray floret growth and subsequent levels of pollinator visitation, that have been introgressed from Senecio squalidus L. into S.
vulgaris L. (Kim et al., 2008). Other types of genes, particularly those underlying species differences, are unlikely to be introgressed, as they are integral for the species survival under the specific ecological conditions in which they grow (Turner et al., 2005; Yatabe et al., 2007).

Complete admixture may occur between species if reproductive isolation is weak. This may cause the extinction of one species, particularly if it is rare. In the Canary island endemic genus Argyranthemum Webb., for example, the widespread A. frutescence Sch. Bip. invaded the rare A. coronopifolium Webb. (Levin et al., 1996), likely leading to its extinction through introgression (Brochmann, 2002). An alternative outcome between incipient species with weak reproductive barriers is reinforcement, where selection enhances reproductive barriers between sympatic taxa completing speciation. Sympatric species must preferentially pair with conspecific rather than heterospecific or admixed individuals for reinforcement to occur (Noor, 1999). This has been seen in animals where assortative mating can be influenced by a few loci of large effect (e.g. Ortiz-Barrientos et al., 2012). However, these specific circumstances are thought to be rare in plants (Marshall et al., 2002), and reinforcement likely only plays a minor role in plant evolution.

1.3.4. Genetic architecture of reproductive isolating traits

The genetic architecture of a reproductive isolating trait, defined as the number of genes involved in the phenotype and their interactions, is an important factor that determines the rate at which reproductive isolation can evolve (Rieseberg et al., 1999; Schemske & Bradshaw, 1999; Rieseberg et al., 2003; Martin et al., 2007). Some traits that cause reproductive isolation evolve roughly clock-like, i.e. the strength of reproductive isolation is proportional to genetic divergence. For example, Coyne & Orr (1989) found a significant correlation between the level of genetic divergence between closely related fruit flies (Drosophila spp.) and the strength of reproductive isolating barriers. A similar result was found between populations of arctic Draba L. (Brassicaceae) species, where geographic and genetic distances correlate with the strength of postmating isolation (hybrid sterility) (Grundt et al., 2006). These cases where reproductive isolation evolves as a by-product of genetic divergence, rather than selection for particular traits, are predicted to be influenced by many loci of small effect spread throughout the genome (Coyne & Orr, 1989).
In contrast to a genetic architecture of many loci of small effect, a reproductive isolating barrier can be generated by a small number of genes of large effect. Such genes include ‘speciation genes’, which contribute to the splitting of two lineages by reducing the amount of gene flow between them (Rieseberg & Blackman, 2010). Candidate lists of speciation genes encoded by nuclear DNA in plants include those regulating floral pigments, self-incompatibility genes, and disease resistance genes (Rieseberg & Blackman, 2010). Speciation genes may be also carried by the organellar genomes. Nuclear-organelle interactions are a by-product of the co-evolution of the organelle and the nuclear genome (Rand et al., 2004). Deleterious mutations that occur in the organelle can be masked by coevolving nuclear restorer loci. When an organelle is exposed to a new nuclear background through hybridization, the deleterious mutations will be unmasked and may cause sterility. *Mimulus* is a well studied plant system for nuclear-organelle interactions, where the genetic basis of mitochondrial-nuclear (Case & Willis, 2008) and cytoplasmic-nuclear interactions (Fishman & Willis, 2007) have been investigated. Other species where nuclear-organelle interactions have been studied include maize (Laughnan & Gabay-Laughnan, 1983; Dewey et al., 1987) and sunflowers (Hans Köhler et al., 1991; Laver et al., 1991).

Different genetic architectures can give rise to similar reproductive isolating barriers. For example, hybrid sterility has an important role in isolating species during secondary contact (Barton & Hewitt, 1985; Rieseberg et al., 1999). This infertility can often be explained by genome-wide divergence that prevents homologous chromosomes pairing correctly at meiosis (Tao & Hartl, 2003). Alternatively, it may be caused by genic incompatibilities at few loci, or few large karyotype differences such as deletions and inversion that differ between homologous chromosomes (Rieseberg, 2001; Brown et al., 2004).

The genetic architecture of reproductive isolating traits has not been studied in *Begonia*. However, it may be predicted that reproductive barriers contributing to divergence will be caused by many mutations of small effect, as a by-product of genome-wide divergence in geographically isolated populations (Coyne & Orr, 1989). After divergence, selection on few loci of large effect may explain the frequent transitions between character states between related species. Divergent morphologies and ecologies may act as reproductive isolating barriers, as
species with such different traits are unlikely to come into secondary contact, and if they did hybrids would be maladapted. Forrest (2000) mapped leaf, tepal, and ovary characters onto a phylogenetic tree of *Begonia*, and showed many characters are homoplasious. Thomas *et al.* (2011b) investigated character state evolution in Southeast Asian *Begonia*, and showed traits such as perennation organ (rhizome, tuberous, non-specialised), fruit type (fleshy, dry capsules, rain ballist capsule) and locule number often change states between sister-species. A preliminary phylogeny by Kidner (unpubl. data) shows that morphological traits, such as peltate and compound leaves, have arisen independently in American *Begonia*.

### 1.4. Aims of the doctoral work and structure of the thesis

This thesis addresses why there are so many species of *Begonia*. This involves a detailed investigation of the processes that contribute to speciation, using approaches targeted at different stages in the speciation process. The first part examines the evolution of reproductive barriers within species, and the second looks at the strength of reproductive isolation when species are brought together in secondary contact. Four chapters address specific research questions, and one describes the development of nuclear markers used in the chapters that follow.

**What are the patterns of seed flow in widespread *Begonia* species, and how is this affected by geographic barriers? (Chapter 2)**

This chapter assesses phylogeographical patterns in two widespread *Begonia* species revealed using newly designed plastid microsatellite markers. This plastid data is used to assess whether patterns of haplotype sharing are more strongly influenced by the population biology of *Begonia* (particularly seed dispersal limitation), than Pleisotocene and pre-Pleistocene geographic barriers found to affect the spatial distribution of genetic variation in other species.

**Development of polymorphic nuclear microsatellite markers for population genetic analyses of Central American *Begonia*. (Chapter 3)**
What ongoing population level processes may promote genetic differentiation in incipiently speciating *Begonia*? (Chapter 4)

In this chapter, two factors that contribute to speciation in *Begonia* are assessed: dispersal limitation and mating system. Patterns of pollen flow are estimated for two widespread *Begonia* species using nuclear microsatellite loci, and compared with seed flow estimates from Chapter 2. The nuclear data is used to test whether inbreeding contributes to patterns of genetic differentiation. Experimental crosses between differentiated populations test whether there are early signs of reproductive barriers evolving, and this is related to intraspecific variation in genome size.

How strong is reproductive isolation between related *Begonia* species, and which barriers contribute to isolation? (Chapter 5)

Experimental crosses are performed between ecologically similar (*B. heracleifolia* Cham & Schltdl. x *B. sericoneura* Lieabm.) and different (*B. heracleifolia* x *B. nelumbiifolia* Cham & Schltdl.) species pairs, to test the strength of post-zygotic barriers. Then the frequency of hybrids and their genotypic classes are estimated in multiple natural hybrid swarms between the two species pairs. These data are used to test the strength of reproductive isolation in a recent *Begonia* species radiation.

What is the genetic architecture of reproductive traits that have diverged since speciation? (Chapter 6)

The final research chapter uses trait segregation, and quantitative trait loci (QTL) mapping, to assess the genetic basis of important reproductive traits that differ between species.

Each of the empirical research chapters are written in the style of research articles intended for different scientific journals. This leads to some overlap in the content of the chapters, however for continuity the full list of references is given at the end of the thesis. In chapter 7, the research is summarised, and future directions for genetic studies in *Begonia* are suggested. The appendices give the full data used for genetic analyses in each chapter.
1.4.1. Study species

Five *Begonia* species are studied in the thesis, all from the large Neotropical section *Gireoudia*. This section was chosen as it has many species (c.66) that are diverse in plant form (Fig 1.4), and because a monograph has been completed for the Central American species (Burt-Utley, 1985). Species from section *Gireoudia* are thought to be uniform in chromosome number \(2n = 28\) unlike many other sections in *Begonia* (Legro & Doorenbos, 1969, 1971, 1973), and this removes one factor that will influence patterns of gene flow and hybridization. Preliminary phylogenetic analysis indicates that the group is paraphyletic, with the small section *Weilbachia* nested inside (Harrison and Kidner, unpubl. data).

The species examined are: *B. conchifolia* A. Dietr., *B. heracleifolia*, *B. nelumbiifolia*, *B. plebeja* and *B. sericoneura*. *Begonia heracleifolia* and *B. nelumbiifolia* are the main study species, used for comparative phylogeography (Chapter 2), population genetics (Chapter 4), and a study of hybridization (Chapter 5). These species were selected as they are widespread species, allowing estimates of gene flow over large spatial scales. They also contrast in their ecologies, allowing hypotheses relating to different ecological adaptations to be tested. The patterns of hybridization observed between *B. heracleifolia* and *B. nelumbiifolia* are supplemented with a third species, *B. sericoneura*, which readily hybridizes with *B. heracleifolia*. This allows a comparison in hybrid zone structure between different species pairs (Chapter 5). *Begonia plebeja* and *B. conchifolia* are used as parents for a genetic mapping study described in Chapter 6. They were chosen as their F1 hybrid is fairly fertile, allowing segregating populations to be generated. The species pair also differ in many morphological traits, allowing the genetics of these traits to be investigated. Each species is described in detail in the introduction to each of the chapters.
Figure 1.4. Representative species from *Begonia* section *Gireoudia* and closely related section *Wielbachia*. Labels from left to right: Top row: *B. morrisiorum* P.D.McMillan and *B. mazae* Ziesnh.; 2nd row: *B. fusca* Liebm. and *B. pseudodaedalea* P.D.McMillan & Rekha Morris; 3rd row: *B. hydrocotylifolia* Otto ex Hook. and *B. sousae* Burt-Utley; Bottom row: *B. faustinoi* Burt-Utley & Utley and *B. calderonii* Standl. White bar is approximately 10cm.
CHAPTER 2: Population history and seed dispersal in widespread Central American *Begonia* species (Begoniaceae) inferred from plastome-derived microsatellite markers

*This chapter has been accepted in a special edition of the Botanical Journal of the Linnean Society entitled Neotropical evolution: assembling the big picture.*

**Data contributions:** Nikki Harrison (University of Edinburgh) provided the plastome sequences from which the microsatellite markers were designed. Alexander Twyford designed the markers, performed all genotyping and analyses.

### 2.1. Chapter summary

Seven plastid microsatellite markers derived from plastome sequence data were used to study the population genetic structure in two widespread *Begonia* spp. from Central America. In *B. nelumbiifolia*, no variation was found at any locus. In contrast, significant haplotype diversity was found in *B. heracleifolia* ($h_T = 0.937$, $h_S = 0.444$, 39 haplotypes, mean of 3.3 haplotypes per population), and populations showed high absolute levels of genetic differentiation ($G'_{ST} = 0.829$, $D = 0.407$). The distribution of haplotypes showed strong phylogeographical structure ($G_{ST} = 0.526$, $R_{ST} = 0.737$, $G_{ST} < R_{ST}$, $P < 0.05$), but this pattern was poorly accounted for by commonly studied historical scenarios, such as Pleistocene refugia or Pliocene differentiation at the Isthmus of Tehuantepec. Instead, subdivision into a large number of regions, each containing local populations (e.g. when $K = 9$, $F_{CT} = 0.749$, $P < 0.05$), best explained the haplotype distribution. The lack of haplotype diversity in *B. nelumbiifolia*, a moist adapted species, suggests that it may have been severely restricted in range during dry spells in the Pleistocene, and has subsequently expanded from this recent population bottleneck. The high haplotype diversity in *B. heracleifolia* may indicate that its adaptation to drought enabled it to survive in small, but ecologically suitable, pockets of isolated habitat throughout the Pleistocene. Limited seed exchange between *B. heracleifolia* populations is likely to be responsible for its high population substructure, and provided the opportunity for divergence through genetic drift. This interpretation is consistent with previous population genetic studies in *Begonia*, and suggests a
common pattern of extremely low genetic exchange among a series of small, but long-lived, populations that may predispose the genus to rapid speciation.

2.2. Introduction

The Neotropics are the most species-rich region on Earth, harbouring over 90 000 species of angiosperms (Richardson et al., 2001; Antonelli & Sanmartín, 2011). The northern limit of the Neotropics passes through Mexico, one of the highest ranked countries in terms of its species richness and species endemism (Myers et al., 2000). The assembly of this diversity over time has been complex, with various biotic and abiotic mechanisms contributing to species diversification (reviewed in Antonelli & Sanmartín, 2011). Within this diverse array of species, there is likely to be high genetic diversity as a consequence of long-term survival in stable populations throughout the Pleistocene (Bawa, 1992; Metcalfe et al., 2000). However, the geographical distribution of this genetic variation is dependent on the particular history of the populations involved.

Patterns of genetic diversity in tropical Mexico may reflect pre-Pleistocene events (Ornelas, Ruiz-Sánchez & Sosa, 2010; Bryson, García-Vázquez & Riddle, 2011). In particular, the Isthmus of Tehuantepec forms a narrow strip of lowland linking the south central Mexican highlands to the uplands of Chiapas and Central America (Fig. 2.1). It has been suggested that the area may have been a historical seaway for much of the Pliocene and Pleistocene (Barrier et al., 1998; Morrone, 2006), and therefore a major barrier to gene flow during this period. Such a barrier may have left a genetic signature of differentiation between populations on either side of the Isthmus. This has been detected in phylogeographical studies for many mid- and high-elevation taxa (e.g. pitvipers, Castoe et al., 2009; harvest mice, Sullivan, Arellano & Rogers, 2000; bird-dispersed Rubiaceae, Gutiérrez-Rodríguez, Ornelas & Rodríguez-Gómez, 2011), as well as low-elevation taxa (e.g. lyresnakes, Devitt, 2006; toads, Mulcahy, Morrill & Mendelson, 2006).

A second factor potentially influencing the distribution of genetic variation is the manner in which populations were affected by climatic fluctuations during the Pleistocene. Colínvaux, De
Oliveira & Bush (2000) believed that wet tropical forests were stable during the Pleistocene and, under these circumstances, no effect of Pleistocene climatic fluctuations would be expected on the genetic structure of tropical species. Other authors have proposed that tropical lineages were affected by climatic fluctuations and only survived in areas receiving high rainfall that remained continually warm and therefore acted as refugia during cool, dry Pleistocene periods (Fig. 2.1; Haffer, 1969; Toledo, 1982; Mary & OHara, 1986; Prance, 1987). Neotropical refugia for south Mexico and Guatemala (Toledo, 1982) and elsewhere in Central America (Prance, 1987) have been proposed on the basis of concentrated areas of species diversity and species endemism. If the Neotropical Pleistocene refugia hypothesis (Haffer, 1969) is true, a signature of strong genetic differentiation among these proposed refugial areas would be expected for tropical species. To date, population genetic studies of a variety of taxa have provided no general consensus regarding the importance of Neotropical refugia during the Pleistocene (Gutiérrez-Rodríguez et al., 2011).

In addition to these major geological events and geographical barriers, local patterns of physiographic heterogeneity and habitat fragmentation in more recent times may also have played a role in structuring genetic diversity. Reconstructions of the range of lowland tropical wet forests in the Quaternary suggest that habitat discontinuity was accentuated during cooler drier periods (Toledo, 1982; Metcalfe et al., 2000; Cárdenas et al., 2011). ‘Islands’ of suitable habitat may have been isolated from one another by inhospitable terrain, such that migration between patches was low. Genetic drift will act strongly in small isolated populations, with genetic variation being lost within populations, and genetic differentiation between them being increased (Quinn & Harrison, 1988). The effects of habitat fragmentation are expected to affect short-lived taxa, such as herbaceous plants, more strongly than long-lived trees, as dispersal between populations is more limited, and there are a greater number of generations for mutations to accumulate, and for the action of genetic drift and selection to take place (Duminil, Hardy & Petit, 2009). If there was restricted gene flow between populations within the Quaternary, as outlined above, genetic differentiation among many different regions would be expected, the pattern showing no correspondence to pre-Pleistocene geographical barriers or putative Pleistocene refugial regions.
In order to gain an insight into the history of Neotropical plant species through genetic analysis, suitable genetic markers are required (Avise et al., 1987). Studies using multiple unlinked nuclear markers are needed to resolve patterns of recent gene flow and pollen dispersal, but plastid markers have some merits over nuclear markers for phylogeographical studies. These properties include the absence of recombination, low effective population size and the conservative mutation rate, allowing for primers to be designed over a wide taxonomic range.
(Ennos et al., 1999; Provan et al., 1999; Provan, Powell & Hollingsworth, 2001). However, the main concern with plastid sequencing is the low rate of nucleotide substitution (Whittall et al., 2010), particularly when analysing relationships between groups that have undergone recent radiations (Richardson et al., 2001) or when resolving intraspecific relationships, where sequence divergence is predicted to be low. An alternative to sequence-based markers or single nucleotide polymorphisms (SNPs) are plastid microsatellites, which have a higher mutation rate (Provan et al., 1999, 2001; Jakobsson, 2007). The most direct method for locating suitable plastid microsatellite markers is to sequence the plastid genome (plastome) of individuals and select the most variable loci (Ebert & Peakall, 2009). However, until recently, this would not have been a viable option, because the generation of large quantities of sequence data is expensive. Next-generation sequencing (NGS) technologies are significantly reducing the costs of generating whole plastomes (Cronn et al., 2008), but have not been generated routinely for intraspecific studies to date (Whittall et al., 2010).

In this study, we use plastid microsatellite markers derived from plastome sequence data to investigate population genetic patterns in two widespread Central American species of *Begonia* L.: *B. heracleifolia* Cham. & Schltdl. and *B. nelumbiifolia* Cham. & Schltdl. Previous genetic studies of *Begonia* spp. have shown dispersal limitation between populations in discontinuous habitats (Matolweni, Balkwill & Mclellan, 2000; Hughes, Hollingsworth & Miller, 2003; Hughes & Hollingsworth, 2008), suggesting that *Begonia* is a good study system for understanding how geographical barriers affect patterns of gene flow and genetic differentiation. Here, we ask whether two historical scenarios, dispersal limitation across the Isthmus of Tehuantepec and survival in Pleistocene refugia, have left a detectable genetic signature in extant *Begonia* populations in Central America. If the Isthmus of Tehuantepec has influenced historical gene flow between *Begonia* populations, a clear east–west genetic break would be expected either side of the Isthmus. In contrast, if the Pleistocene glacial refugia proposed by Toledo (1982) have played a role, high genetic diversity would be expected in former glacial refugia, and haplotypes would not be shared between refugia. We also examine whether a third scenario, of dispersal limitation between long-lived isolated *Begonia* populations, fits the pattern of genetic diversity better. If this is the case, high levels of phylogeographical structure, measured using $R_{ST}$, and genetic differentiation, measured using statistics such as $D$ (Jost, 2008), would be expected, which are unrelated to the patterns outlined above. In the light of our
findings, we evaluate the benefits of using microsatellite markers derived from plastome sequences for studying phylogeographical patterns.

2.3. Material and methods

2.3.1. Study species

The mega-diverse genus *Begonia*, with c.1500 species, is one of the ten largest plant genera (Frodin, 2004). Approximately 690 *Begonia* spp. have been recorded in Central America (Goodall-Copestake et al., 2010; Dewitte et al., 2011), and no previous phylogeographical or population genetic studies have been conducted. *Begonia* spp. are typically herbaceous plants or understorey shrubs that grow in wet tropical lowland and montane forests, but some Central American species have a rhizomatous habit and have adapted to seasonally dry forests. They are monoecious, and their flowers are white or light pink with two petaloid sepals, which attract generalist pollinators, and pollen dispersal between populations is predicted to be low (Ågren & Schemske, 1991). The small seeds are not modified to promote dispersal by an animal vector or by wind, and are therefore expected to have limited dispersal. These attributes differ markedly from those of Neotropical trees, which have been the focus of phylogeographical studies (e.g. Cardoso et al., 1998; Cavers, Navarro & Lowe, 2003; Ornelas et al., 2010; Poelchau & Hamrick, 2012), and therefore their responses to past climatic fluctuations in relation to Pleistocene refugia are expected to differ.

The species selected, *B. heracleifolia* and *B. nelumbiifolia*, were chosen for their different ecological preferences (Hoover, 1979), being adapted to dry and moist habitats, respectively. Comparative population genetic studies of related species with different ecologies can potentially distinguish between the influence of past biogeographical barriers, which would affect the dispersal of both species, and the distribution of suitable habitat, which may affect one species more than another (Bermingham & Moritz, 1998). Their widespread distributions (*B. heracleifolia* grows from Mexico to Honduras, *B. nelumbiifolia* from Mexico to Colombia) and their tendency to grow as ruderals differ from most other *Begonia* spp., which are often narrow endemics in primary habitat (Hughes & Hollingsworth, 2008). The species are diploid (2*n* = 28;
Legro & Doorenbos, 1969) nonsister species, based on preliminary analysis using the nuclear ribosomal internal transcribed spacer (nrITS) and plastid \textit{trnL-F} intron (C. A. Kidner, unpubl. data). \textit{Begonia heracleifolia} has repent rhizomes and symmetrical or weakly asymmetrical lobed leaves with serrations, and varying degrees of dark blotches and coloured margins, although leaf phenotype (colour of the leaf blade and degree of leaf lobing) varies across its range (Fig. 2.2). \textit{Begonia nelumbiifolia} has large peltate leaves and densely packed symmetrical inflorescences (Fig. 2.2), and is relatively morphologically uniform throughout its range.

![Begonia heracleifolia and Begonia nelumbiifolia](image)

**Figure 2.2.** Study species. Left: leaf shape variation in \textit{Begonia heracleifolia}, photographs taken from wild collected plants grown in a common glasshouse environment. Right: \textit{Begonia nelumbiifolia} growing on a shady rock near Motzorongo.

### 2.3.2. Sampling

Population samples were collected in the south Mexican states of Chiapas, Oaxaca and Veracruz (Fig. 2.1). Between three and 30 individuals per population were collected for genotyping. Individuals were selected at a minimum sampling distance of 1 m to reduce the chance of collecting clonally reproducing plants (Hughes & Hollingsworth, 2008). Flower or young leaf material was collected in silica gel stored at room temperature prior to DNA extraction (Table 2.1). A single representative of each population is lodged at the herbarium at E, except for the Guatemalan population, for which the specimen is at BIGU.
Table 2.1. Collection details for *Begonia* samples used in this study. For *B. heracleifolia*, the positions relative to the Isthmus of Tehuantepec and the nearest Pleistocene refugium (if applicable) proposed by Toledo (1982) are also indicated. Refugial area refers to those marked in Fig.2.1.

<table>
<thead>
<tr>
<th>Pop. code</th>
<th>Locality</th>
<th>State</th>
<th>Country</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Altitude (m)</th>
<th>W/ E Isthmus</th>
<th>Refugial area</th>
</tr>
</thead>
<tbody>
<tr>
<td>h2</td>
<td>4km NE San Andrés Tuxtla</td>
<td>Veracruz</td>
<td>Mexico</td>
<td>18.47850</td>
<td>-95.17802</td>
<td>630</td>
<td>West</td>
<td>C</td>
</tr>
<tr>
<td>h3</td>
<td>7km NE San Andrés Tuxtla</td>
<td>Veracruz</td>
<td>Mexico</td>
<td>18.50341</td>
<td>-95.16824</td>
<td>883</td>
<td>West</td>
<td>C</td>
</tr>
<tr>
<td>h4</td>
<td>8km NE San Andrés Tuxtla</td>
<td>Veracruz</td>
<td>Mexico</td>
<td>18.50660</td>
<td>-95.16607</td>
<td>910</td>
<td>West</td>
<td>C</td>
</tr>
<tr>
<td>h5</td>
<td>9km NE San Andrés Tuxtla</td>
<td>Veracruz</td>
<td>Mexico</td>
<td>18.52009</td>
<td>-95.16176</td>
<td>1041</td>
<td>West</td>
<td>C</td>
</tr>
<tr>
<td>h8</td>
<td>Agua Azul</td>
<td>Chiapas</td>
<td>Mexico</td>
<td>17.22117</td>
<td>-92.11073</td>
<td>431</td>
<td>East</td>
<td>E</td>
</tr>
<tr>
<td>h9</td>
<td>Unión Juárez</td>
<td>Chiapas</td>
<td>Mexico</td>
<td>15.08881</td>
<td>-92.08448</td>
<td>1718</td>
<td>East</td>
<td>D</td>
</tr>
<tr>
<td>h10</td>
<td>Ocozocautla biosphere reserve</td>
<td>Chiapas</td>
<td>Mexico</td>
<td>16.90533</td>
<td>-93.45153</td>
<td>957</td>
<td>East</td>
<td>n/a</td>
</tr>
<tr>
<td>h12</td>
<td>Ocozocautla biosphere reserve</td>
<td>Chiapas</td>
<td>Mexico</td>
<td>16.94645</td>
<td>-93.45651</td>
<td>625</td>
<td>East</td>
<td>n/a</td>
</tr>
<tr>
<td>h13</td>
<td>Near Berriozábal</td>
<td>Chiapas</td>
<td>Mexico</td>
<td>16.86693</td>
<td>-93.32781</td>
<td>1153</td>
<td>East</td>
<td>n/a</td>
</tr>
<tr>
<td>h14</td>
<td>Roadside to Fortin de las Flores</td>
<td>Veracruz</td>
<td>Mexico</td>
<td>18.89369</td>
<td>-97.01050</td>
<td>902</td>
<td>West</td>
<td>A</td>
</tr>
<tr>
<td>h15</td>
<td>5km south of Orizaba</td>
<td>Veracruz</td>
<td>Mexico</td>
<td>18.78252</td>
<td>-97.08508</td>
<td>1300</td>
<td>West</td>
<td>A</td>
</tr>
<tr>
<td>h16</td>
<td>Valley 30km East of Huatusco</td>
<td>Veracruz</td>
<td>Mexico</td>
<td>19.20111</td>
<td>-96.67139</td>
<td>290</td>
<td>West</td>
<td>A</td>
</tr>
<tr>
<td>h21</td>
<td>Highway 175 between Jasaa and San Juan Bautista</td>
<td>Oaxaca</td>
<td>Mexico</td>
<td>17.74356</td>
<td>-96.32803</td>
<td>1246</td>
<td>West</td>
<td>B</td>
</tr>
<tr>
<td>h23</td>
<td>Highway 175, near Santa Maria Jacatapec</td>
<td>Oaxaca</td>
<td>Mexico</td>
<td>17.86422</td>
<td>-96.21381</td>
<td>60</td>
<td>West</td>
<td>B</td>
</tr>
<tr>
<td>h24</td>
<td>Motzorongo</td>
<td>Oaxaca</td>
<td>Mexico</td>
<td>18.66953</td>
<td>-96.78714</td>
<td>470</td>
<td>West</td>
<td>A</td>
</tr>
<tr>
<td>h26</td>
<td>San Jeronimo Zochein</td>
<td>Oaxaca</td>
<td>Mexico</td>
<td>17.22117</td>
<td>-95.23547</td>
<td>109</td>
<td>West</td>
<td>n/a</td>
</tr>
<tr>
<td>h28</td>
<td>near Santa Maria Xanabi</td>
<td>Veracruz</td>
<td>Mexico</td>
<td>15.98808</td>
<td>-96.11061</td>
<td>1</td>
<td>West</td>
<td>n/a</td>
</tr>
<tr>
<td>h-g1</td>
<td>Cobán</td>
<td>Alta Verapaz Guatemala</td>
<td>Guatemala</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pop. code</td>
<td>Locality</td>
<td>State</td>
<td>Country</td>
<td>Latitude</td>
<td>Longitude</td>
<td>Altitude (m)</td>
<td>W/ E Isthmus</td>
<td>Refugial area</td>
</tr>
<tr>
<td>----------</td>
<td>-----------------------------------------------</td>
<td>---------</td>
<td>---------</td>
<td>-------------</td>
<td>-------------</td>
<td>--------------</td>
<td>--------------</td>
<td>--------------</td>
</tr>
<tr>
<td>n1</td>
<td>Near Los Tuxtlas biological station</td>
<td>Veracruz</td>
<td>Mexico</td>
<td>18.59026</td>
<td>-95.07876</td>
<td>165</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n4</td>
<td>8km NE San Andrés Tuxtla</td>
<td>Veracruz</td>
<td>Mexico</td>
<td>18.50660</td>
<td>-95.16607</td>
<td>910</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n5</td>
<td>9km NE San Andrés s Tuxtla</td>
<td>Veracruz</td>
<td>Mexico</td>
<td>18.52009</td>
<td>-95.16176</td>
<td>1041</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n7</td>
<td>Road to Ocosingo from Palenque</td>
<td>Chiapas</td>
<td>Mexico</td>
<td>17.42477</td>
<td>-91.99712</td>
<td>396</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n11</td>
<td>Ocozocoautla biosphere reserve</td>
<td>Chiapas</td>
<td>Mexico</td>
<td>16.92489</td>
<td>-93.45090</td>
<td>888</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n18</td>
<td>Valley 25km East of Huatusco</td>
<td>Veracruz</td>
<td>Mexico</td>
<td>19.20367</td>
<td>96.742556</td>
<td>635</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n21</td>
<td>Hwy 175 between Jasaa and San Juan Bautista</td>
<td>Oaxaca</td>
<td>Mexico</td>
<td>17.74356</td>
<td>96.328028</td>
<td>1246</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n25</td>
<td>Arroya Zacata</td>
<td>Oaxaca</td>
<td>Mexico</td>
<td>17.73981</td>
<td>95.791333</td>
<td>78</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n26</td>
<td>San Jerónimo Zoochina</td>
<td>Oaxaca</td>
<td>Mexico</td>
<td>17.22117</td>
<td>95.235472</td>
<td>109</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2.3.3. DNA extraction

Total genomic DNA was isolated from 25 mg of silica dried flower or leaf material using a modified protocol for the DNeasy 96-sample kit (Qiagen, Germantown, MD, USA). Dry material was disrupted for two cycles of 2 min (25 Hz), and centrifuged for 10 s (4000 rpm), prior to the recommended heated lysis, to obtain a fine powder. The heated lysis was extended from 30 min to 1 h to increase the yield of DNA. As a result of unknown inhibitors of polymerase chain reaction (PCR), all DNA was diluted 100-fold with Millipore ultrapure water prior to PCR.

2.3.4. Plastid microsatellite marker design

2.3.4.1. Plastid marker design

New plastid microsatellite markers were designed from the draft plastome of *B. nelumbiifolia* (N. Harrison & C. A. Kidner, unpubl. data). The frequency and length of microsatellite repeat motifs in the plastome were calculated using the Phobos Tandem Repeat Finder v3.3.12 plugin for Geneious v5.4 (Mayer, 2006–2010). Primers were designed around perfect (uninterrupted) microsatellite motif repeats using WebSat (Martins *et al.*, 2009). The plastome sequence was uploaded to WebSat in two FASTA files, and primers were designed to amplify PCR products between 100 and 400bp in length, with an optimal primer melting temperature (T<sub>m</sub>) of 60°C and a primer GC content of 40%. Forty potential primer pairs located in WebSat, amplifying mononucleotide or dinucleotide repeats of at least eight repeat units in length, were annotated onto the plastome of *B. nelumbiifolia* in Geneious v5.4 (Drummond *et al.*, 2011). The primer pairs were compared with the plastome sequences of 15 other *Begonia* species (N. Harrison and C. A. Kidner, unpubl. data), and those that were not conserved between species in the same section [*Gireoudia* (Klotzsch) A.DC, 66 species; Burt- Utley, 1985] were discarded. The markers with the most variable microsatellite repeat lengths between species were further tested to determine their suitability for population genetic analysis.
2.3.4.2. Plastid marker test

Sixteen individuals from different populations of *B. heracleifolia* were used to test the amplification success and allelic diversity of the plastid markers. All primers (listed in Table 2.2) were tested individually prior to multiplex testing (below). PCRs were performed using M13-tailed fluorescent primers (Schuelke, 2000). PCRs were performed in a final volume of 10µL, containing 0.5µL of 1mM M13-tailed forward primer (Invitrogen, Grand Island, NY, USA), 1µL of reverse primer (1mM), 1µL of 1mM M13-fluorescently modified primer [6-FAM, VIC, NED, PET (Applied Biosystems, Foster City, CA, USA)], 0.25µL of bovine serum albumin (BSA, 0.4%), 1µL of 10x reaction buffer, 1µL of 2mM deoxynucleoside triphosphates (dNTPs), 0.6µL of 25 mM MgCl2, 3.6µL of double-distilled H2O, 0.05 µL of Biotaq polymerase (Bioline, London, UK) and 1µL of dilute DNA template. PCR was performed in a Peltier Thermal Cycler and consisted of a profile of initial denaturation at 95 °C for 1 min, followed by 40 cycles of denaturation at 95 °C for 1 min, annealing at 57 °C for 1 min, followed by extension at 72 °C for 1 min. PCR products were run on a 2% agarose gel stained with SYBR Safe gel stain (Invitrogen) and visualised under UV light. One microlitre of each PCR product labelled with the four fluorescent dye colours was pooled and diluted 50x in Millipore distilled H2O prior to fragment analysis on an ABI3730 at the GenePool (University of Edinburgh, UK) using GeneScan 500 LIZ internal size standard (Applied Biosystems, Foster City, CA, USA). Fluorescent traces were analysed automatically with manual editing using GeneMapper v4.0 (Applied Biosystems).

PCR products of microsatellite repeats are prone to PCR amplification errors caused by slipped-strand mis-pairing (Fazekas, Steeves & Newmaster, 2010). To check whether changes in the length of the target microsatellite motif correlated directly with the change in PCR product size, rather than other microsatellites in the flanking region or other indels, we sequenced a subset of individuals. Long mononucleotide repeat motifs (>15 bp) can amplify poorly. Therefore, to mitigate this problem, they were amplified using AmpliTaq Gold polymerase, as recommended by Fazekas et al. (2010), employing the protocol recommended by the manufacturer. Excess primers and sequencing reactions were performed using the protocol of Thomas et al. (2011).
Table 2.2. Plastid microsatellite primers generated for this study. Motif refers to the plastome of Begonia nelumbiifolia; N alleles and product size range are the number of alleles found in B. heracleifolia and the polymerase chain reaction (PCR) product size, respectively.

<table>
<thead>
<tr>
<th>Locus</th>
<th>Primer sequences</th>
<th>Multiplex (fluorophore)</th>
<th>Motif</th>
<th>N alleles</th>
<th>Product size Range (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bnc1</td>
<td>F: M13-GGATTCGAGTTGGATTGGACTAR:CGAGAAAGTCTACGGGTTCGAGT</td>
<td>1 (FAM)*</td>
<td>(A)$_9$†</td>
<td>3</td>
<td>343-347</td>
</tr>
<tr>
<td>Bnc2</td>
<td>F: M13-TGTGCTTTAGTGGGCTTAGTTR:TCTGTGTATATGAGTTAGTGCA</td>
<td>5 (PET)</td>
<td>(T)$_{15}$</td>
<td>3</td>
<td>177-183</td>
</tr>
<tr>
<td>Bnc6</td>
<td>F: M13-CCCTCGATAGTTCTTTGTTCGTR:TTTATCCCTTATACAGCCAAC</td>
<td>2 (FAM)*</td>
<td>(AT)$_{14}$</td>
<td>6</td>
<td>398-412</td>
</tr>
<tr>
<td>Bnc7</td>
<td>F: M13-GGGAAGGGAGGATCTACATATAT :R:AACGGAGCACCTAACAACGTAT</td>
<td>3 (VIC)</td>
<td>(T)$_{15}$</td>
<td>3</td>
<td>189-191</td>
</tr>
<tr>
<td>Bnc9</td>
<td>F: M13-CGGCAGAAATAAGGGGATTCTCATATAT:R:TCCCTCCGTTCCTCGTAGTT</td>
<td>5 (PET)</td>
<td>(T)$_{13}$</td>
<td>7</td>
<td>338-350</td>
</tr>
<tr>
<td>Bnc11</td>
<td>F: M13-GCTATGGTAAATCGGTAGACRA:CGGAAGCTATATAGTCTTACCA</td>
<td>3 (VIC)</td>
<td>(C)$_{11}$</td>
<td>n/a‡</td>
<td>350-376</td>
</tr>
<tr>
<td>Bnc13</td>
<td>F: M13-CTTTGCAAAGGAGGAAAGATCGR:CGTCAATTTAGCTTTCTTTCTTTT</td>
<td>4 (NED)</td>
<td>(T)$_{13}$</td>
<td>8</td>
<td>126-137</td>
</tr>
<tr>
<td>Bnc14</td>
<td>F: M13-GATTCCAAAATGGCTTACGAAAR:TCATATCTGTTGATAAAAGTTGAAA</td>
<td>4 (NED)</td>
<td>(T)$_{15}$</td>
<td>6</td>
<td>181-188</td>
</tr>
<tr>
<td>Bnc16</td>
<td>F: M13-TATGACCCGGAGCTAATCR:GGTTGCAATCCTCCTCTTT</td>
<td>4 (NED)</td>
<td>(A)$_{15}$</td>
<td>n/a‡</td>
<td>235-242</td>
</tr>
</tbody>
</table>

*As a result of preferential amplification of one of the products, these loci were amplified individually and pooled in equimolar ratios prior to fragment analysis.
†Mutationally complex region with multiple microsatellites. Longest motif listed.
‡Not used in analyses, see text.

M13 motif, CACGACGTTGAAAACGAC; n/a, not applicable.
2.3.4.3. **Multiplex PCR test**

The plastid microsatellite primers that amplified well individually were tested for multiplex compatibility. To reduce the likelihood of amplifying large nontarget PCR products between adjacent primer sites, the primer pairs were annotated onto the whole plastome of *B. nelumbiifolia* in Geneious v5.4, and primer combinations within 5 kb were excluded. Additional primer–primer interactions that interfere in the amplification of the template DNA were tested for using AutoDimer (Vallone & Butler, 2004). Potentially compatible multiplex partners were then tested empirically by sequentially mixing equimolar concentrations of each primer pair and testing PCR on one individual of each species. Once successful combinations were observed (amplification of both products with similar band intensities on an agarose gel, minimal primer dimers), this procedure was repeated until the optimal multiplexing combinations were derived from the available markers. Multiplex PCR assays were then conducted on population-level samples of *B. heracleifolia* and *B. nelumbiifolia*.

2.3.5. **Haplotype distributions and relationship among haplotypes**

A plastid haplotype can be defined as the unique combination of alleles at multiple loci from the plastid. Plastid haplotypes were defined in a Microsoft Excel (Microsoft Corporation, Washington DC, USA) spreadsheet using the Chloroplast PCR-RFLP Excel macro (French, 2003). To assess the relationships between haplotypes, we calculated haplotype connection lengths in Arlequin v3.0 (Excoffier, Laval & Schneider, 2005), and visualised the minimum-spanning tree in HapStar v0.5 (Teacher & Griffiths, 2011). We also constructed a median-joining network in Network (http://www.fluxus-engineering.com) using default parameters. We visualised the spatial distribution of plastid haplotypes by plotting haplotypes by hand onto a topographical map of Central America (downloaded from http://www.worldclim.org).

2.3.6. **Population genetic diversity and genetic differentiation**

A reduced dataset including only populations for which all loci amplified in at least three individuals (15 of 18 populations) was used for population genetic analyses. The total number of
haplotypes and the mean number of haplotypes per population were calculated, and the intrapopulation haplotype diversity was assessed using the gene diversity index (\(H_e\)) calculated in Arlequin v3.0.

Absolute measures of differentiation (\(G'_{ST}\) and Jost’s \(D\)) were calculated in SMOGD (Crawford, 2010), and their significance was calculated with 1000 bootstrap replicates. \(G'_{ST}\) is a standardised \(G_{ST}\) based on the maximum level of differentiation possible for a given set of gene diversities (Hedrick, 2005). \(D\) (Jost, 2008) is a test statistic based on the effective number of alleles, and is therefore a good measure of allelic differentiation among populations (Meirmans & Hedrick, 2011). We tested for a pattern of isolation by distance, which can indicate dispersal limitation between populations. Pairwise \(F_{ST}\) values between populations were calculated in Arlequin and their significance was calculated by 1000 permutations. Pairwise geographical distances between populations were calculated with the Geographic Distance Matrix Generator v1.2.3 (http://biodiversityinformatics.amnh.org/open_source/gdmg) (Ersts). We regressed pairwise estimates of \(F_{ST}\) / (1 - \(F_{ST}\)) against the natural logarithm of geographical distance (Rousset, 1997) and tested the significance of the relationship using a Mantel test with 10,000 permutations in Isolation by Distance (IBD) Web Service (Jensen, Bohonak & Kelley, 2005).

The relative roles of different historical scenarios on the patterns of haplotype distributions were assessed using analysis of molecular variance (AMOVA), following the barriers tested by Gutiérrez-Rodríguez et al. (2011). First, we evaluated the proportion of \(F_{ST}\) accounted for by the Isthmus of Tehuantepec. Second, we tested the proportion of \(F_{ST}\) accounted for by the glacial refugia presented by Toledo (1982), by placing populations in broad bins surrounding these refugia. AMOVAs were performed in Arlequin with 20,000 permutations per test, and the groups used for each AMOVA are shown in Table 2.1.

We also tested whether other groups of populations form distinct genetic clusters, without defining potential groups a priori, using spatial analysis of molecular variance (SAMOVA) implemented in SAMOVA v1.0 (Dupanloup, Schneider & Excoffier, 2002). This approach groups adjacent populations that are genetically similar to each other, whilst maximizing between-group genetic variance, without making assumptions of ploidy, linkage or Hardy–
Weinberg equilibrium. We tested all K values between $K = 2$ and $K = 15$ using 100 initial conditions per test and the sum of the squared size difference setting recommended for microsatellite data. To test whether related haplotypes are more likely to be found in geographical proximity to each other (phylogeographical structure), we compared patterns of unordered haplotype frequencies ($G_{ST}$) with the squared difference of the number of repeats ($R_{ST}$) in the program cpSSR (Pons & Petit, 1996).

2.4. Results

2.4.1. Plastid microsatellite marker design

The draft *B. nelumbiifolia* plastome contained 232 perfect mononucleotide repeats exceeding eight repeat units in length, 114 of which were longer than 10 repeat units. Repeats longer than eight repeat units in length included: 101 poly(A), 126 poly(T), four poly(C) and one poly(G) repeat. The longest repeat was a mono(A) repeat with 25 repeat units, but this result may be biased as NGS frequently underestimates the length of mononucleotide repeats (Chan, 2009), and long repeat units were often followed by a chain of unknown nucleotides (Ns). The *B. nelumbiifolia* plastome also contained seven dinucleotide repeats with at least six repeat units, six of which were poly(AT) and one poly(TG). Thirty-one of the 40 (77.5%) primer pairs that were compared across the plastome alignment were conserved between species of section *Gireoudia*.

All nine of the plastid microsatellite primers that were conserved across section *Gireoudia* applied in the first round of testing successfully amplified a fragment in both species. A multiplex PCR assay was found to incorporate all nine loci in five PCR reactions without affecting the band profile, producing large primer dimers or greatly reducing the yield of one of the products (Table 2.2). Locus *Bnc16* did not amplify well in some samples, so that these primers were not used for further genotyping. Overall, the full set of eight loci was successfully amplified in 263 individuals from 18 populations for *B. heracleifolia* and 187 individuals from 13 population of *B. nelumbiifolia* (Table 2.1). The full matrix of plastid microsatellite data is in Appendix 1.
Begonia nelumbiifolia was monomorphic for all genotyped loci, and therefore no further population genetic analyses could be conducted on this species. In contrast, B. heracleifolia was polymorphic for all eight loci. One locus, Bnc11, had a 15-bp deletion relative to the second shortest allele at the locus. As some measures of population differentiation (e.g. measures of $R_{ST}$) assume a stepwise mutation model, this locus was excluded from further analyses. The alleles at the remaining seven loci generally differed between the two species, with few overlapping allelic variants, and a 12-bp deletion at locus Bnc6 consistently distinguished the two species. For B. heracleifolia, the mean number of alleles per locus was 5.2; the full lists of the numbers of alleles per locus and gene diversities are given in Tables 2.2 and 2.3, respectively.

### 2.4.2. Haplotype distributions and the relationship among haplotypes

Thirty-nine haplotypes from the allelic combinations of seven plastid microsatellite loci were found in B. heracleifolia (Table 2.3). Only six of the 39 haplotypes (pt1, pt2, pt4, pt6, pt21, pt23) were shared between populations, with the remaining 33 (85%) being private haplotypes (Fig. 2.3). The most widespread haplotypes were pt1 and pt2, which were found in five and seven populations, respectively, and pt5 and pt6 were shared between distant populations within the Mexican Gulf region (sensu Escalante et al., 2007). The only haplotype shared across geographical regions was pt21, which was common to one of the most westerly populations (h15) and the most easterly population from Guatemala (hg), separated by ~800 km. In the haplotype tree, pt21 fell in a haplogroup with four other haplotypes from the Mexican Gulf. The minimum spanning tree (Fig. 2.4), which is a single parsimony tree reconstructing the relationship between haplotypes, showed no large-scale phylogeographical structure or distinct haplogroups. However, local geographical structure was indicated by haplotypes being clustered with others from the same area. The internal haplotypes included pt1 and pt2, which were some of the few widespread haplotypes. The median-joining network also showed consistent results, with the absence of broad-scale phylogeographical structure, but local clustering (results not shown).
Table 2.3. Plastid haplotype frequencies and gene diversities per population. Population codes refer to Table 2.1. N is the number of individuals genotyped. The frequency of each plastid type (pt) is indicated in parentheses.

<table>
<thead>
<tr>
<th>Population Code</th>
<th>N</th>
<th>Haplotypes</th>
<th>Gene Diversity</th>
</tr>
</thead>
<tbody>
<tr>
<td>h2</td>
<td>7</td>
<td>pt1(6), pt2(1)</td>
<td>0.041</td>
</tr>
<tr>
<td>h3</td>
<td>27</td>
<td>pt1(1), pt2(20), pt3(1), pt4(3), pt5(2)</td>
<td>0.107</td>
</tr>
<tr>
<td>h4</td>
<td>3</td>
<td>pt2(2), pt6(1)</td>
<td>0.286</td>
</tr>
<tr>
<td>h5</td>
<td>33</td>
<td>pt2(28), pt6(4), pt7(1)</td>
<td>0.102</td>
</tr>
<tr>
<td>h8</td>
<td>10</td>
<td>pt8(5), pt9(1), pt10(4)</td>
<td>0.137</td>
</tr>
<tr>
<td>h9</td>
<td>1</td>
<td>pt11(1)</td>
<td></td>
</tr>
<tr>
<td>h10</td>
<td>2</td>
<td>pt12(2)</td>
<td></td>
</tr>
<tr>
<td>h12</td>
<td>19</td>
<td>pt13(16), pt14(1), pt15(1), pt16(1)</td>
<td>0.087</td>
</tr>
<tr>
<td>h13</td>
<td>7</td>
<td>pt17(3), pt18(1), pt17(3), pt18(1)</td>
<td>0.218</td>
</tr>
<tr>
<td>h14</td>
<td>3</td>
<td>pt1(3)</td>
<td>0.000</td>
</tr>
<tr>
<td>h15</td>
<td>15</td>
<td>pt1(9), pt19(1), pt20(4), pt21(1)</td>
<td>0.210</td>
</tr>
<tr>
<td>h16</td>
<td>2</td>
<td>pt1(1), pt2(1)</td>
<td></td>
</tr>
<tr>
<td>h21</td>
<td>20</td>
<td>pt4(6), pt22(14)</td>
<td>0.063</td>
</tr>
<tr>
<td>h23</td>
<td>16</td>
<td>pt23(10), pt24(2), pt25(3), pt26(1)</td>
<td>0.364</td>
</tr>
<tr>
<td>h24</td>
<td>35</td>
<td>pt27(29), pt28(3), pt29(1), pt30(1), pt31(1)</td>
<td>0.097</td>
</tr>
<tr>
<td>h26</td>
<td>28</td>
<td>pt2(18), pt5(7), pt32(1), pt33(1), pt34(1)</td>
<td>0.184</td>
</tr>
<tr>
<td>h28</td>
<td>7</td>
<td>pt35(6), pt36(1)</td>
<td>0.041</td>
</tr>
<tr>
<td>h-g1</td>
<td>28</td>
<td>pt21(6), pt37(20), pt38(1), pt39(1)</td>
<td>0.137</td>
</tr>
</tbody>
</table>

Figure 2.3. Spatial distribution of plastid haplotypes in Begonia heracleifolia (next page, top). Pie charts are placed at the site of each collection, and the chart size is proportional to the number of individuals genotyped. The size of each segment corresponds to the number of individuals with a given haplotype. White segments represent widespread haplotypes, and coloured segments are haplotypes limited to that region. Key to colours: pink, Veracruz; red, north Oaxaca; orange, north-east Oaxaca; light blue, Los Tuxtlas; dark blue, south Oaxaca; purple, west Chiapas; dark green, north Chiapas; light green, south Chiapas; yellow, Guatemala.

Figure 2.4. Minimum spanning tree of haplotype relationships (next page, bottom). Numbers refer to the plastid types (listed in Table 2.3). Colours refer to collection sites, and are the same as in Fig. 2.3 (see legend for Fig. 2.3 for explanation). Hypothetical (unsampled) haplotypes are represented by filled black circles.
2.4.3. Population genetic diversity and genetic differentiation

*Begonia heracleifolia* showed significant structuring of genetic diversity ($G_{ST} = 0.526$, $h_T = 0.937$, $h_S = 0.444$). Pairwise comparisons of population structure ($F_{ST}$) were also high, with 89% of pairwise $F_{ST}$ values significant at $P < 0.05$ (data not shown). The mean value of Hedrick’s (2005) standardised estimator of genetic differentiation $G'_{ST}$ was 0.829. The value for Jost’s estimator of genetic differentiation $D$ was 0.407. The Mantel test between genetic distance and the natural logarithm of geographical distance recognised significant isolation by distance ($Z = 153.7$, $r = 0.439$, $P < 0.005$); however, this relationship explained a limited amount of the genetic structure ($r^2 = 19\%$). To test whether plastid mutations had accumulated in each population, we calculated the level of significance between $G_{ST}$ and $R_{ST}$ by a permutation test. An $R_{ST}$ value of 0.737 (0.0872 SE) was significantly greater than the $G_{ST}$ value of 0.526 (0.0577 SE, $P < 0.05$), indicating phylogeographical structuring of related haplotypes.

2.4.4. Historical barriers to gene flow

Most spatial structuring of genetic diversity was not explained by an east–west break at the Isthmus of Tehuantepec ($F_{CT} = 0.08915$, $P = 0.15$) in the hierarchical AMOVA (Table 2.4). A higher and significant level of plastid variation was partitioned between geographical clusters that corresponded to glacial refugia ($F_{CT} = 0.2951$, $P < 0.05$). Patterns of genetic structure influenced by geographical barriers that were not tested in the ANOVAs were assessed with SAMOVA. The highest $F_{CT}$ value in the SAMOVA corresponded roughly to the number of broad geographical areas sampled ($k = 9$, $F_{CT} = 0.749$), or as the level of subdivision approached the actual number of populations collected (e.g. $K = 13$, $F_{CT} = 0.746$). At $K = 9$, a single widespread geographical cluster contained localities in the Mexican Gulf (populations h3, h4, h5, h14, h15, h26).
Table 2.4. Analysis of molecular variance (AMOVA) of Begonia heracleifolia by populations (a), collection groups (inferred Pleistocene refugia) (b) and west–east division at the Isthmus of Tehuantepec (c). DF – degrees of freedom.

<table>
<thead>
<tr>
<th></th>
<th>DF</th>
<th>Sum of squares</th>
<th>Variance components</th>
<th>Percentage of variation</th>
<th>Fixation indices</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Among populations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within populations</td>
<td>14</td>
<td>301.739</td>
<td>1.256</td>
<td>72.77</td>
<td></td>
</tr>
<tr>
<td>(b) Collection groups (inferred Pleistocene refugia)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Among groups</td>
<td>7</td>
<td>222.918</td>
<td>0.524</td>
<td>29.51</td>
<td>F\textsubscript{CT} = 0.296</td>
</tr>
<tr>
<td>Among populations within groups</td>
<td>7</td>
<td>78.821</td>
<td>0.781</td>
<td>43.96</td>
<td>F\textsubscript{SC} = 0.624</td>
</tr>
<tr>
<td>Within populations</td>
<td>243</td>
<td>114.563</td>
<td>0.471</td>
<td>26.53</td>
<td>F\textsubscript{ST} = 0.735</td>
</tr>
<tr>
<td>Total</td>
<td>257</td>
<td>416.302</td>
<td>1.777</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(c) W-E Isthmus Tehuantepec</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Among groups</td>
<td>1</td>
<td>41.641</td>
<td>0.163</td>
<td>8.91</td>
<td>F\textsubscript{CT} = 0.089</td>
</tr>
<tr>
<td>Among populations within groups</td>
<td>13</td>
<td>260.098</td>
<td>1.193</td>
<td>65.28</td>
<td>F\textsubscript{SC} = 0.717</td>
</tr>
<tr>
<td>Within populations</td>
<td>243</td>
<td>114.563</td>
<td>0.471</td>
<td>25.80</td>
<td>F\textsubscript{ST} = 0.742</td>
</tr>
<tr>
<td>Total</td>
<td>257</td>
<td>504.612</td>
<td>2.172</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2.5. Discussion

2.5.1. Between-species comparisons of genetic diversity

Of the two species examined here, B. heracleifolia showed significant spatial structuring of genetic diversity, whereas B. nelumbiifolia showed no polymorphisms at any locus examined. There are a number of possible explanations for the reduced plastid variation within B. nelumbiifolia relative to B. heracleifolia, one of which is that it could be a genotyping artefact. We used seven plastid microsatellite loci and, by chance, the loci selected may show no variation; other loci in the plastome may show variation. However, this seems unlikely, as we used loci with the longest repeat motifs from B. nelumbiifolia; therefore, we do not anticipate
any ascertainment bias. Further evidence for a lack of variation within *B. nelumbiifolia* comes from the absence of any SNP variation in the flanking regions of the microsatellite markers and in seven other plastid microsatellite loci used for genotyping in the marker development phase (data not shown). Therefore, it seems more likely that our results indeed represent reduced plastid diversity in *B. nelumbiifolia*.

The most likely explanation for the absence of plastid variation in *B. nelumbiifolia* is a recent population bottleneck or a selective sweep in a single refugial population prior to range expansion. Allelic diversity is reduced when a population experiences a reduction in its effective population size, and the allelic diversity after a population bottleneck is a function of the population size when it is reduced to its minimum and the rate of population growth (Nei, Maruyama & Chakraborty, 1975; Cornuet & Luikart, 1996). The recent timing of range expansion after a bottleneck in the Pleistocene could explain the lack of any post-expansion accumulation of mutations in *B. nelumbiifolia*. This is consistent with the preference of the species for moist habitats, which are likely to have been geographically more restricted in dry periods during the Pleistocene (Toledo, 1982). The fixation of a single plastid haplotype in response to recent bottleneck events has been found in other plant species over similarly large geographical areas (e.g. Mediterranean *Pinus pinea* L., Vendramin et al., 2008; Australian *Atherosperma moschatum* Labill., Worth et al., 2011).

The absence of plastid variation in *B. nelumbiifolia* contrasts with that in *B. heracleifolia*, where significant genetic diversity was found (average of 3.3 haplotypes per population; Table 2.3). This level of haplotype diversity is surprising, as a single haplotype may be expected to be fixed by genetic drift in the small populations typical of *B. heracleifolia* (Hoover, 1979; A. D. Twyford, pers. observ.). The most likely explanation for the high genetic diversity found in *B. heracleifolia* populations is that this drought-adapted species may have tolerated Pleistocene drought conditions, and populations may have survived *in situ* and not have been confined to refugia. Over a long period of time, there would have been the opportunity to differentiate by drift and selection for local adaptations with limited dispersal of the accumulated new mutations across the range of the species. This is consistent with the wide diversity of leaf morphologies seen in *B. heracleifolia* (Fig. 2.2A), which may be the product of both drift and local adaptation.
Monographic work on the section by Burt-Utley (1985) suggested that morphological variation in *B. heracleifolia* is continuous and does not warrant further division into subspecific taxa, but this question may now be readdressed considering the population level molecular evidence presented here and evidence from future molecular phylogenetic studies.

### 2.5.2. The relative role of historical barriers to gene flow

We evaluated whether two geographical barriers that are known to have reduced dispersal in other groups influenced the spatial structuring of genetic diversity in *B. heracleifolia*: differentiation across the Isthmus of Tehuantepec and differentiation between putative Pleistocene glacial refugia. Bryson *et al.* (2011) compiled a list of 30 phylogenetic and phylogeographical studies of highland and lowland animal species that had genealogical splits between populations within species, or between sister species, that corresponded to the Isthmus of Tehuantepec. Seventeen of these species had exclusively old (Neogene) splits and eight species had more recent (Pleistocene) splits. Gutiérrez-Rodríguez *et al.* (2011) found that an east–west intraspecific split at the Isthmus of Tehuantepec explained more genetic variation ($F_{CT} = 0.737$) than putative Pleistocene glacial refugia ($F_{CT} = 0.556$) for the understorey plant *Palicourea padifolia* (Willd. ex Roemer & Schultes) C.M.Taylor & Lorence (Rubiaceae). The Pleistocene divergence date (309 000 years ago; confidence interval, 136 000–667 000 years ago) between clades in their study suggests that the Isthmus of Tehuantepec has presented a barrier to plant dispersal in recent times, not just pre-Pleistocene. However, we found little evidence for a genetic break in *B. heracleifolia* corresponding to the Isthmus, indicated by the low $F_{CT}$ value in the AMOVA ($F_{CT} = 0.08915$, $P = 0.15$, Table 2.4).

As an alternative scenario, we tested whether the spatial arrangement of genetic diversity could be accounted for by the glacial refugia proposed by Toledo (1982). We found no strong support for past refugia structuring genetic diversity in *Begonia*. Most genetic variation was found within groups that represent refugia ($F_{SC} = 0.62362$), rather than between them ($F_{CT} = 0.29510$). Moreover, much more diversity was explained when the populations were divided at a finer scale, or in the alternative groups supported by the SAMOVAs, as discussed below. These results give little support for the Pleistocene refugia proposed by Toledo (1982), but this may in
part reflect the difficulty in delimiting tropical refugia, especially using methods based on species diversity and endemism. These results show that responses to Pleistocene climatic fluctuations may depend on species-specific adaptations, and therefore the pinpointing of refugia for complex tropical species assemblages will be difficult (Poelchau & Hamrick, 2012b). Some progress in locating common geographical barriers to gene flow that may represent Neotropical refugia has been made by comparing patterns of haplotype diversity in widespread Central American tropical tree species (Poelchau & Hamrick, 2012a), where barriers shared between species with different ecologies were distinguished from haplotype patterns idiosyncratic to individual species. Therefore, more phylogeographical studies of species with different life-history traits are required to understand where potential refugia are likely to be located, although the current data suggest that less clear-cut patterns may be expected than those revealed for European and North American refugia, where congruent geographical patterns of genetic diversity have shed convincing light on their likely locations (Hewitt, 2000; Petit et al., 2003).

2.5.3. Dispersal limitation in *Begonia* explains current patterns of genetic structure

The spatial structuring of plastid variation at a local scale provides further support for the persistence of *B. heracleifolia* populations *in situ* during dry periods in the Pleistocene, rather than preferential survival in the refugia suggested for moisture-loving tropical plants by Toledo (1982). Of the 39 plastid haplotypes found, 33 (85%) were private haplotypes restricted to a single population (Table 2.3 and Fig. 2.3). Although the frequency of plastid haplotypes is strongly influenced by the different sample sizes for each population, additional support for local population genetic structure comes from the high overall $G_{ST}$ value and the many significant pairwise $F_{ST}$ values. The result of the SAMOVAs also supports the subdivision of genetic variation by local geographical groups.

Significant spatial partitioning of genetic diversity and high levels of genetic differentiation (mean value of Jost’s D across loci, $D = 0.422$, $G_{ST}^{*} = 0.829$) in *B. heracleifolia* are best explained by dispersal limitation between long-lived isolated populations. *Begonia* spp. have tiny seeds that are not known to be dispersed by any animal vectors and are poorly wind
dispersed, and therefore effective seed dispersal between populations is expected to be low. This is confirmed by clumps of Begonia seedlings growing directly below adult plants in the field (A. D. Twyford, pers. observ.). Hoover (1979) also observed the clustering of individuals in three Mexican Begonia spp. (including B. heracleifolia). Seed dispersal limitation may explain the isolation of populations for extended periods throughout the Pleistocene, during which time genetic novelties could have arisen that are not shared between populations. The pattern of local geographical structure also shows that human-mediated dispersal, facilitated by road building and translocation for horticultural purposes, has not obscured the natural patterns of geographical variation, as might be expected for a species that tends to grow as a weed in human disturbed environments.

High levels of genetic differentiation between populations have been found in other population genetic surveys of Begonia spp. using nuclear markers. Matolweni et al. (2000) found strong spatial structuring of genetic diversity in a survey of allozymes in the endemic South African B. dregei Otto & A.Dietr. and B. homonyma Steud. Hughes & Hollingsworth (2008) found a high $F_{ST}$ value between forest patches of the endemic B. sutherlandii Hook.f and a strong signal of isolation by distance among populations. They linked this micro-evolutionary pattern of local population structure to the macro-evolutionary process of allopatric species divergence that may have given rise to the large number of species present in the genus (1500 species; Frodin, 2004). These population-level surveys are supported by molecular phylogenetic studies of South-East Asian Begonia spp. using plastid DNA sequences, in which long-distance dispersal events appear to have been rare, with in situ species radiations after dispersal into each of the eastern Malesian islands (Thomas et al., 2011). To allow comparisons of gene flow with these previous studies, and to assess the joint role of interpopulation pollen and seed dispersal, future studies should apply nuclear markers to these widespread species. This will also allow the responses of the two species to Pleistocene refugia to be investigated in more detail.

### 2.5.4. Plastid microsatellites as tools to study intraspecific relationships

The use of plastid markers remains a popular approach for population genetic and phylogeographical studies of Neotropical plants (e.g. Cardoso et al., 1998, Fontaine et al., 2004).
This popularity may partly be attributed to the limited availability of variable nuclear markers; however, there are a number of properties of the plastome that make it a desirable source of data for population genetic and phylogeographical analysis of widespread groups. First, it is typically maternally inherited in angiosperms (Corriveau & Coleman, 1988) and will reflect only seed dispersal patterns, unlike biparentally inherited nuclear markers that are spread in pollen and in seed (Ennos, 1994). Seeds typically disperse less than pollen, which makes seed-specific markers more likely to track geographical barriers. Second, only seed dispersal (not pollen) can result in establishment in a new habitat, so that plastid markers can reveal species distribution changes, without the blurring of ancestral patterns of gene flow caused by subsequent pollen dispersal (Ennos et al., 1999; Petit et al., 2003). Third, plastid markers have a smaller effective population size than nuclear markers (Wright et al., 2008), so that genetic drift is more potent. Moreover, fewer alleles can be maintained at mutation–drift equilibrium, so that fewer individuals need to be genotyped to sample the range of allelic diversity in a population. Finally, plastid genotyping has been widely used in plant phylogeography, and there are typically many other studies in which the distributions of plastid haplotypes have been assessed over a similar geographical area, allowing for comparisons between taxa. Using NGS to generate whole plastomes, and the identification of hypervariable microsatellite markers from these data, is a novel approach to the design of variable plastid markers. This route was used successfully to find variable markers for B. heracleifolia, even when the plastid sequence divergence between populations was anticipated to be low.

Despite the benefits of this approach, there are a number of pitfalls associated with sole reliance on hypervariable plastid markers for population genetic and phylogeographical studies, as highlighted by the results of this study. Population genetic studies typically assume that alleles shared between individuals are the product of gene flow between the populations, and that homoplasy, where the same allele is caused by convergent mutations, violates this assumption. Microsatellite markers are particularly prone to homoplasy as the number of repeat units will not always expand, but can also contract (Provan et al., 2001). Moreover, mutations do not always occur in a stepwise manner (Ceplitis, Su & Lascoux, 2005), and rapid deletion of microsatellite motifs can occur [as demonstrated by Micheneau et al. (2010) in Cephalanthera Rich. (Orchidaceae)]. This may explain the disjunct fragment length seen in this study at locus Bnc11.
The combined influence of the complex mutation model and homoplasy makes coalescent methods for the estimation of divergence age estimates from plastid microsatellite data challenging, as it is difficult to correctly assign suitable priors. This is particularly problematic for phylogeographical studies, as a knowledge of divergence ages can help to discriminate between different hypotheses. Therefore, sequence data should be used in preference to plastid microsatellites when divergence age estimates are required.

2.6. Conclusion

This study contributes to our growing knowledge of the evolutionary processes occurring at the population level in the genus *Begonia*. The contrasting levels of plastid diversity for the two species tested here suggests the importance of species-specific ecological preferences in terms of their response to climate change in relation to glacial refugia. Dry-adapted tropical lineages, such as *B. heracleifolia*, may have survived *in situ* during dry periods in the Pleistocene, whereas moisture-loving species, such as *B. nelumbiifolia*, may have been more vulnerable and persisted only in refugia. Generalizations about refugia may therefore only hold for species with similar ecological preferences, rather than at broad taxonomic scales. This study also shows that patterns of strong population genetic differentiation and dispersal limitation may be common across the genus. Taxa with low interpopulation gene flow, such as *Begonia*, may thus be especially useful for tracking local geographical variation, as homogenizing gene flow from neighbouring populations is limited. As more data are assembled, *Begonia* is emerging as a powerful study system for testing biogeographical questions.
CHAPTER 3: Development and characterisation of microsatellite markers for Central American *Begonia* section *Gireoudia* (Begoniaceae)

**Data contributions:** Catherine Kidner (University of Edinburgh) provided the transcriptome and genome sequences from which the markers were designed. Alexander Twyford designed the markers, performed all genotyping and analyses.

### 3.1. Chapter summary

**Premise of the study:** Transcriptome sequence data was used to design microsatellite primers for two widespread Central American *Begonia* species, *B. heracleifolia* and *B. nelumbiifolia*, to investigate population structure and hybridization.

**Methods and results:** The transcriptome from vegetative meristem tissue from the related *B. plebeja* was mined for microsatellite loci, and thirty-one primer pairs amplified in the target species. Fourteen primer pairs were combined in two multiplex PCR reactions, which amplified an average of 4 alleles per locus.

**Conclusions:** The markers developed will be a valuable genetic resource for medium-throughput genotyping of Central American *Begonia* species. Their amplification in two divergent species within the section is promising for their transferability to related species.

### 3.2. Introduction

*Begonia* is a diverse tropical genus with over 1500 species. Evolutionary research has focused on the early diverging African species (e.g. Hughes and Hollingsworth, 2008), and the more derived Asian species (e.g. Thomas *et al.*, 2011), with the American species largely overlooked. The most recent common ancestor of Central American *Begonia* is likely to be relatively recent (Miocene, Dewitte *et al.*, 2011), and subsequent speciation has resulted in high species richness (total c.690 species, Copestake *et al.*, 2010). Population studies of Central American *Begonia* spp. will shed light on the evolution of species richness in a morphologically diverse group of
Neotropical herbs; but to date studies have been limited by the availability of suitable nuclear markers to complement plastid microsatellite markers (Chapter 2).

In this study, we describe the development a multiplex PCR assay of nuclear microsatellite markers for two Central American Begonia species: *B. heracleifolia* Cham & Schltdl. and *B. nelumbiifolia* Cham & Schltdl.. These species were chosen as they are two of the most widespread Begonia species in a genus of mostly rare endemics (Hughes and Hollingsworth, 2008). The species are known to hybridize (Burt-Utley, 1985) facilitating studies of species boundaries. The markers are developed from the transcriptome of related *B. plebeja* Liebm., and these genic markers are anticipated to amplify over a broad phylogenetic scope (Lepais and Bacles, 2011). These markers will be used to study patterns of gene flow within and between Central American Begonia species.

### 3.3. Methods and results

Microsatellite markers were designed from the transcriptome sequence of vegetive meristem tissue from *B. plebeja*, a related species from Begonia section Gireoudia (European Nucleotide Archive Sequence Read Archive accession number: ERP001195; Brennan et al., submitted). The QDD bioinformatic pipeline (Meglécz et al., 2010), which integrates microsatellite detection, a redundancy check to avoid amplifying multiple PCR products, and designs primers, was used according to Lepais and Bacles (2011). A FASTA file of the *B. plebeja* transcriptome sequence assemblies were analysed in QDD v1.3 using default parameters: selecting only primers that amplify a PCR product between 90 and 320bp in length, with a repeat motif of 2 - 6bp repeats, and a minimum length of 4 repeat units. To make microsatellite amplification in other species more likely, primers were excluded if they did not have a perfect BLAST match to the transcriptome of *B. conchifolia* A.Dietr. (section Gireoudia).
Table 3.1. Characterisation of nuclear microsatellites for Central American *Begonia* species. The multiplex to which the primer was assigned, primer melting temperature when amplified individually ($T_m$), the motif in *B. plebeja*, the number of alleles ($N_a$) found in *B. heracleifolia* (her) and *B. nelumbiifolia* (nel), and the observed range of PCR product sizes excluding the M13 motif, are listed.

<table>
<thead>
<tr>
<th>Locus name</th>
<th>Primer sequences</th>
<th>Multiplex</th>
<th>Fluorcent dye</th>
<th>$T_m$ (°C)</th>
<th>Motif</th>
<th>$N_a$ her</th>
<th>$N_a$ nel</th>
<th>Allele sizes (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BI4329</td>
<td>F: M13-CAACCAACAATGGCAGCTT R:CATGGAGATAATGGAGCTGG</td>
<td>1</td>
<td>FAM</td>
<td>59</td>
<td>(GGA)$_6$</td>
<td>4</td>
<td>2</td>
<td>89-104</td>
</tr>
<tr>
<td>BI3043</td>
<td>F: M13-CGACATCCAACAAAACCTG R:TTGATAGATGGAAGGTCGC</td>
<td>1</td>
<td>FAM</td>
<td>60</td>
<td>(TC)$_5$</td>
<td>1</td>
<td>2</td>
<td>173-179</td>
</tr>
<tr>
<td>BC432</td>
<td>F: M13-AAACTCCGATGGATCAGCA R:TGAAATAAACACACAAAAAGACA</td>
<td>1</td>
<td>FAM</td>
<td>60</td>
<td>(TG)$_5$</td>
<td>1</td>
<td>1</td>
<td>261-263</td>
</tr>
<tr>
<td>BC344</td>
<td>F: M13-GAGGGAGGGTCCTTTGTTAG R:CCGTCTTCGTGGATCATC</td>
<td>1</td>
<td>VIC</td>
<td>60</td>
<td>(GCA)$_5$</td>
<td>1</td>
<td>1</td>
<td>105-108</td>
</tr>
<tr>
<td>BI6278</td>
<td>F: M13-TGTAGTTGTTGATAGCAGAACTTTG R:CAGATGGGTCGGAGATTTTG</td>
<td>1</td>
<td>VIC</td>
<td>59</td>
<td>(TCC)$_7$</td>
<td>1</td>
<td>3</td>
<td>238-253</td>
</tr>
<tr>
<td>BI5347</td>
<td>F: M13-TCACTCATTCTTAATCAGACC R:CTCTATCATTTCCAAGGATTTC</td>
<td>1</td>
<td>VIC</td>
<td>59</td>
<td>(CTT)$_6$</td>
<td>2</td>
<td>1</td>
<td>171-183</td>
</tr>
<tr>
<td>BC552</td>
<td>F: M13-TGTCTGAGATGGAACTCGCG R:TAGTCGAAGGGATCCGAATG</td>
<td>1</td>
<td>NED</td>
<td>60</td>
<td>(GT)$_5$</td>
<td>2</td>
<td>2</td>
<td>271-273</td>
</tr>
<tr>
<td>BI3348</td>
<td>F: M13-ACCTGTTTCTCGTTGAGGC</td>
<td>1</td>
<td>PET</td>
<td>60</td>
<td>(CT)$_6$</td>
<td>3</td>
<td>3</td>
<td>279-283</td>
</tr>
</tbody>
</table>
BI06534  F: M13-CGTTGCTCTGCTCTAACCCT
          R: AGATAACGACACCCGATTTC
          1   PET  59  (TC)₆  6  2  97-107
BI7112   F: M13-ATCCAATGTAACCTCTCGG
          R: GTGCAATTAGCTCCCGTGT
          2   FAM  60  (TCC)₆  2  2  109-115
BI3820   F: M13-AGGACCAATTTTTGACGGCTA
          R: GAAGCTTTTCTCTTCTGTGA
          2   FAM  59  (CTT)₇  5  2  158-176
BI134    F: M13-ATCAGCTCCTCCCTATCTCTCT
          R: TGCAATCTCCTGTGGTTCTT
          2   VIC  60  (CT)₆  4  2  306-314
BI362    F: M13-CTTCACCTCGGCTGAACAAC
          R: GAGGCCAATATTTATGCGGA
          2   NED  60  (ATG)₆  4  4  147-159
BC332    F: M13-GAACCAGAAGTCAAGGGTTCA
          R: AAAATGATTTTCCTCATCCAA
          2   PET  59  (TCA)₄  4  2  188-200

Additional loci tested

BI4004   F: M13-TCAGGAAATATTCGATTGGGA
          R: GCATTCCTCTGTGTAATGC
          59  (AT)₅  2  3  155-169
BC672    F: M13-CCTGATCGAGAAAGAACCG
          R: AAAAGCCAGCTCCTTCTCTGA
          60  (CTT)₈  3  1  152-158
BI4477   F: M13-GGATCTCCTCTGCTTTGCTG
          R: GGCGAGACCAGAAGAACCCG
          60  (CT)₉  4  2  111-119
BI06604  F: M13-ATTTTTCCACAGAAGAGGCC
          R: GGCAGAACCAGCAGTATATC
          59  (AT)₈  6  1  111-127
BI6294   F: M13-TGCTGGTCTGAATCTTTAATCA
          R: TGCGGTCTTGGTACTCTTTCC
          59  (AT)₁₀  1M  1M  148
BI6701   F: M13-ACTCCACTCTGCTGCAC
          R: GAGATGATGAGGGTTCAGGC
          60  (GA)₆  1M  1M  195
<table>
<thead>
<tr>
<th>Accession</th>
<th>F: M13-Sequence</th>
<th>R: M13-Sequence</th>
<th>Size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BI05710</td>
<td>GAAAGTTTTTGAGGAAGCCC</td>
<td>TGGAAGAGATCAGAAGGTACA</td>
<td>178-184</td>
</tr>
<tr>
<td>BI4848</td>
<td>CGACGCCTCTCAAAAGAAGAA</td>
<td>AGAAGGTACAAGAAGGTACA</td>
<td>71-74</td>
</tr>
<tr>
<td>BC402</td>
<td>TTACTCGAGCTAGAAGCCGC</td>
<td>AGGGCTTGGAGAGCTAGAGG</td>
<td>92</td>
</tr>
<tr>
<td>BC932</td>
<td>GAGTGATGAAGGCGAAGAGG</td>
<td>GAGTGATGAAGGCGAAGAGG</td>
<td>660-662†</td>
</tr>
<tr>
<td>BI3069</td>
<td>TTACTCGAGCTAGAAGCCGC</td>
<td>AGGGCTTGGAGAGCTAGAGG</td>
<td>184-192</td>
</tr>
<tr>
<td>BI3377</td>
<td>AACACAATTCAGCCGACAC</td>
<td>GAAGGAGATGATTATGACGA</td>
<td>MP</td>
</tr>
<tr>
<td>BI5174</td>
<td>GTCGAACGGTTTTTGCTAGGA</td>
<td>GAGTGATGAAGGCGAAGAGG</td>
<td>118-121</td>
</tr>
<tr>
<td>BC42</td>
<td>ACTTACAGGTACTCTATAG</td>
<td>AGGGCTTGGAGAGCTAGAGG</td>
<td>147-173</td>
</tr>
<tr>
<td>BI6984</td>
<td>GGAATACAGGTAGTTGCTGCT</td>
<td>GAGTGATGAAGGCGAAGAGG</td>
<td>148-164</td>
</tr>
<tr>
<td>BI7247</td>
<td>CTTCTTATTCCCGTAAAAAGC</td>
<td>AGGGCTTGGAGAGCTAGAGG</td>
<td>135</td>
</tr>
<tr>
<td>BC312</td>
<td>ATTTCTTTCTCGGAAAGATG</td>
<td>ATCGGAACCTCTGAGCCTGAA</td>
<td>178-180</td>
</tr>
</tbody>
</table>

*M Monomorphic in all individuals tested
† Large product size assumed to be caused by an intron
MP Multiple PCR products amplified

M13 sequence is: CACGACGTGTAAAAACGAC
Thirty-one primer pairs detected in QDD were tested for amplification in 7 individuals of both *B. nelumbiiifolia* and *B. heracleifolia* from different populations. A subset of polymorphic markers that amplified reliably in both species was then tested for multiplex compatibility, by mixing equimolar ratios of each primer. The PCR multiplexes were then tested on a population of each species (20 individuals) to estimate the genetic diversity of the markers. The primers sequences were BLAST searched against the transcriptome sequence of the divergent Asian species *B. venusta* King (section *Platycentrum*), to test for likely cross-amplification of primers in other *Begonia* species.

Approximately 15mg of silica dried leaf material was extracted using DNeasy 96-sample kit (Qiagen, Germantown, MD, USA). To overcome an unknown PCR inhibitor that co-elutes with DNA extractions in *Begonia*, extractions were diluted 100-fold with Millipore dH$_2$O to a final DNA concentration of ~0.1-1.5µg/ml. PCR reactions were performed using the M13-tailed primer method (Schuelke, 2000) in a final reaction volume of 10µl, containing: 0.5µl of 1mM M13-tailed forward primer (Invitrogen, Grand Island, USA), 1µl reverse primer (1mM), 1µl of 1mM M13 fluorescently modified primer (6-FAM,VIC, NED, PET), 0.25µl bovine serum albumin (BSA, 0.4%), 1µl of 10x reaction buffer, 1µl of 2mM,dNTPs, 0.6µl of 25mM MgCl$_2$, 0.05µl Biotaq polymerase (Bioline, UK), 1 µl dilute DNA template, and made up to the final volume using dH$_2$O. PCR cycles consisted of an initial denaturation of 1 minute as 95°C, followed by 40 cycles of denaturation for 1 minute at 95°C, annealing for 1 minute at 57°C followed by extension of 1 minute at 72°C. Five microlitres of each PCR product labelled with the four fluorescent dye colours was pooled and diluted 2x in Millipore dH$_2$O, and the GeneScan 500 LIZ internal size standard (Applied Biosystems, Foster City, Ca) was added prior to fragment analysis on the ABI3730 (GenePool, University of Edinburgh, UK). Fluorescent traces were analysed automatically with manual editing using GeneMapper v4.0 (Applied Biosystems).

A total of 136 primer pairs were located in the *B. plebeja* transcriptome using the QDD bioinformatic pipeline. All thirty one of the subset of primers tested for amplification yielded a PCR product (Table 3.1). Of these loci, 4 loci were monomorphic (BI6701, BC402, BI6294 and BI7247), and one amplified multiple PCR products (BI3377). Two PCR multiplex reactions were designed to amplify a total of 14 polymorphic loci (Table 3.1). All loci were polymorphic
in at least one of the populations tested, and showed moderate genetic diversity with number of alleles per species from 1 to 5 and expected within population heterozgosity between 0 and 0.75 (Table 3.2). Twenty-one of the 62 primers (34%) had perfect BLAST matches in the transcriptome of the divergent \textit{B. venusta}, including both the forward and reverse primers for loci BI3348, BC932, BC552.

**Table 3.2.** Genetic diversity in population samples of \textit{B. heracleifolia} and \textit{B. nelumbiifolia}. The number of alleles per locus ($N_a$), expected heterozgosity ($H_E$) and observed heterozgosity ($H_O$), and the total alleles observed in the two species ($N_T$) are listed.

<table>
<thead>
<tr>
<th>Locus</th>
<th>\textit{B. heracleifolia}</th>
<th>\textit{B. nelumbiifolia}</th>
<th>$N_T$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$N_a$</td>
<td>$H_O$</td>
<td>$H_E$</td>
</tr>
<tr>
<td>BEI4329</td>
<td>3</td>
<td>0.400</td>
<td>0.524</td>
</tr>
<tr>
<td>BEI03043</td>
<td>4</td>
<td>0.000</td>
<td>0.444</td>
</tr>
<tr>
<td>BEC432</td>
<td>2</td>
<td>0.100</td>
<td>0.097</td>
</tr>
<tr>
<td>BEC344</td>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>BEI6278</td>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>BEI5347</td>
<td>3</td>
<td>0.300</td>
<td>0.449</td>
</tr>
<tr>
<td>BEC552</td>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>BEI3348</td>
<td>4</td>
<td>0.579</td>
<td>0.604</td>
</tr>
<tr>
<td>BEI06534</td>
<td>5</td>
<td>0.500</td>
<td>0.750</td>
</tr>
<tr>
<td>BEI7112</td>
<td>2</td>
<td>0.400</td>
<td>0.467</td>
</tr>
<tr>
<td>BEI3820</td>
<td>5</td>
<td>0.600</td>
<td>0.623</td>
</tr>
<tr>
<td>BEC134</td>
<td>4</td>
<td>0.611</td>
<td>0.732</td>
</tr>
<tr>
<td>BEI04004</td>
<td>2</td>
<td>0.059</td>
<td>0.059</td>
</tr>
<tr>
<td>BIC362</td>
<td>2</td>
<td>0.050</td>
<td>0.050</td>
</tr>
<tr>
<td>BEC332</td>
<td>4</td>
<td>0.250</td>
<td>0.483</td>
</tr>
<tr>
<td>Mean</td>
<td>3.333</td>
<td>0.321</td>
<td>0.440</td>
</tr>
<tr>
<td>SD</td>
<td>1.155</td>
<td>0.228</td>
<td>0.246</td>
</tr>
</tbody>
</table>

### 3.4. Conclusions

We have described the development of nuclear microsatellite primers that amplify in two divergent Central American \textit{Begonia} species. Many of the primers have exact BLAST matches in the transcriptome of the South-East Asian species \textit{B. venusta} and may therefore be
transferable widely across the genus. The transferability of markers is important for the study of natural hybrids, and the development of a multiplexed assay of 14 loci should enable accurate assignment to hybrid classes (e.g. F1, backcross). Future studies will use these loci to estimate the genetic structure of populations, and the frequency of hybrids and extent of introgression in hybrid swarms.
CHAPTER 4: Genetic differentiation and species cohesion in two widespread Central American *Begonia* species

**Data contributions:** Andrew Matthews (Queen Mary University London) provided the genome size (C-value) estimates. Alexander Twyford performed all genotyping and experimental crosses, and all the analyses.

### 4.1. Chapter summary

*Begonia* is one of the ten largest plant genera, with over 1500 species. This high species richness may in part be explained by weak species cohesion, which has allowed speciation by divergence in allopatry. In this study, we investigate species cohesion in two widespread *Begonia* species, *B. heracleifolia* and *B. nelumbiifolia*, to test whether the population processes that may have promoted past speciation events, such as dispersal limitation and inbreeding, are ongoing in extant species. Populations from Mexico and Guatemala were genotyped at nine nuclear microsatellite loci. Crosses between divergent populations were used to test whether reproductive barriers have accumulated, and compared with genome size estimates from representative populations. Strong population substructure was found for *B. heracleifolia* ($F_{ST} = 0.364$, $F'_{ST} = 0.506$, $P < 0.05$) and *B. nelumbiifolia* ($F_{ST} = 0.277$, $F'_{ST} = 0.439$, $P < 0.05$), and Bayesian admixture analysis supports the division of most populations into discrete genetic clusters. Comparisons with plastid microsatellite data (Chapter 2) show that pollen flow is particularly limiting between populations, and this dispersal limitation has led to significant genetic differentiation (*B. heracleifolia* Jost’s $D = 0.274$, $P < 0.05$; *B. nelumbiifolia* $D = 0.294$, $P < 0.05$). Moderate levels of selfing (*B. heracleifolia* $s = 0.40$, *B. nelumbiifolia* $s = 0.55$) may further increase genetic differentiation between populations. Intraspecific genome size estimates for *B. heracleifolia* (1C = 0.80pg) and *B. nelumbiifolia* (1C = 0.54pg) were consistent across the species range, except one Oaxacan *B. heracleifolia* population that possessed a genome ~10% larger (1C = 0.88pg) than the species mean. Crosses with this divergent population had a 20% reduction in pollen viability compared to close outcrosses. The population genetic data suggest *Begonia* populations are long lived and stable, but only weakly connected by gene flow. This weak species cohesion has led to reproductive barriers accumulating between the most isolated
populations, supporting the hypothesis that allopatric divergence in situ is the precursor of speciation in *Begonia*.

### 4.2. Introduction

Species cohesion is maintained by gene flow (Slatkin, 1987; Morjan & Rieseberg, 2004). If gene flow between populations is disrupted, populations become genetically isolated from one another, and reproductive barriers subsequently evolve leading to speciation (Coyne & Orr, 2004; Martin & Willis, 2010). Studying this trajectory, from a panmictic population, to a genetically differentiated species with strong reproductive barriers, is one way to get a handle on the processes of speciation (Coyne & Orr, 1989; Hall *et al.*, 2006). Many factors will influence this trajectory, some promoting genetic differentiation between populations, others maintaining species cohesion in spite of reduced gene flow (Wright, 1943).

One major factor influencing cohesion and differentiation processes in plants is the rate of interpopulation gene flow (migration) mediated by pollen and seed dispersal (Ennos, 1994; Linhart & Grant, 1996). Pollen is the main source for gene exchange between populations, as it is normally dispersed over greater distances than seeds (Ennos, 1994; Heuertz *et al.*, 2003; Bacles *et al.*, 2006; Dick *et al.*, 2008, but exceptions include many orchid species and *Anthoxanthum odoratum* L., Freeland *et al.*, 2011). For this reason, pollen flow plays an important role in the cohesion of species (Heywood, 1991). Seed dispersal between populations has a smaller role in gene flow, but is crucial in colonization and subsequent expansion of a species range (Petit *et al.*, 2003). It is therefore important to estimate both pollen and seed flow to best understand population dynamics, and the way that species cohesion is maintained.

Mating systems affect patterns of gene exchange between populations, and therefore contribute to geographic structuring of genetic diversity that is the precursor of reproductive isolation and speciation (Hamrick & Godt, 1996; Lasso *et al.*, 2011). Obligate outbreeding species, such as dioecious or self-incompatible species, usually maintain high levels of genetic diversity in a given population (Charlesworth, 2003). In contrast, populations of species that self-fertilize or inbreed may have reduced allelic diversity (Hamrick & Godt, 1996), and populations with low
genetic diversity are particularly prone to divergence through genetic drift (Vellend & Geber, 2005).

Dispersal limitation and the mating system will contribute to neutral genetic divergence between populations, in combination with other demographic factors such as the effective population size (Charlesworth, 2009). But speciation will only be an outcome if reproductive barriers accumulate between populations (Rieseberg & Blackman, 2010). These barriers may be caused by a small number of genes causing differential adaptation (‘speciation genes’ in the genic species concept, see Wu, 2001) or incompatibilities between many fixed alleles at neutral or adaptive loci between lineages [Dobzhansky-Muller (DM) incompatibilities, Dobzhansky, 1937; Muller, 1942]. Studies of species cohesion must therefore also estimate the strength of reproductive barriers between populations to test for incipient speciation.

High species richness in tropical plant genera can in part be explained by geographically structured populations where inbreeding in common, which promotes allopatric speciation (Baker-Federov hypothesis, Lasso et al., 2011). The link between population patterns of genetic differentiation, and species-level patterns of lineage splitting and speciation, has been made for Begonia (Hughes & Hollingsworth, 2008). Population genetic analyses have been conducted for African Begonia species which are rare or restricted to particular habitats, and these studies have shown high levels of genetic differentiation at nuclear microsatellite (B. sutherlandii, $F_{ST} = 0.485$, $F'_{ST} = 0.896$; Hughes & Hollingsworth, 2008) and allozyme loci (B. dregei, $F_{ST} = 0.882$; B. homonyma, $F_{ST} = 0.937$; Matolweni et al., 2000). At the species level, phylogenetic work shows geographically constrained monophyly of species radiations (Forrest & Hollingsworth, 2003; Thomas et al., 2011a). Micro-evolutionary processes of high genetic differentiation, even at small spatial scales, seem to be correlated with macro-evolutionary patterns of rapid speciation (Hughes & Hollingsworth, 2008). This pattern may in part explain why Begonia is one of the largest angiosperm genera, with over 1500 species.

Here, we test whether there is a strong genetic signature of differentiation between populations of two widespread Begonia species that grow in continuous habitats in Central America. This would be consistent with weak mechanisms of species cohesion that would allow new Begonia
species to form through divergence in allopatry. *Begonia heracleifolia* Cham. & Schltdl. and *B. nelumbiifolia* Cham. & Schltdl. are genotyped at nine nuclear microsatellite markers to estimate genetic differentiation. The data from nuclear markers are compared to plastid data (Chapter 2), to infer the ratio of interpopulation pollen and seed dispersal. The co-dominant data is also used to estimate the level of inbreeding, which may further contribute to genetic differentiation. We then test whether genetic incompatibilities have accumulated between differentiated populations, by looking at the fertility of crosses between populations, and relate the crossing data to an intraspecific survey of genome sizes. The joint genetic, crossing, and genome size data is used to test whether allopatric divergence in situ within *Begonia* species is the precursor of allopatric speciation in this large genus.

### 4.3. Material and methods

#### 4.3.1. Study species

*Begonia nelumbiifolia* and *B. heracleifolia* were chosen as they are two of the most widespread Central American *Begonia* species, found throughout Mexico and into Central America (*B. heracleifolia* to Honduras; *B. nelumbiifolia* to Columbia, Burt-Utley, 1985), in a genus of mostly narrowly distributed endemics (Hughes & Hollingsworth, 2008). These species are easily distinguished from other related *Begonia* species by their leaf and flower morphology (Burt-Utley, 1985). The species pair are an ideal comparison for studying incipient speciation, as *B. heracleifolia* is highly variable in leaf shape and leaf colour throughout its range, whereas *B. nelumbiifolia* is more uniform. They also differ in their ecologies, with *B. nelumbiifolia* growing in moist shaded areas and *B. heracleifolia* in dry sun-exposed areas. These species typically occur in small isolated populations, although they can be locally abundant and form dense stands (Hoover, 1979). Both species can be propagated by splitting rhizomes, allowing them to be easily transported, and grown in cultivation to be used in experimental crosses.
4.3.2. Sampling and genotyping

To test patterns of genetic diversity and differentiation, an average of 25 individuals were genotyped from 13 populations of *B. heracleifolia* and 7 populations of *B. nelumbiifolia* (Table 4.1). Samples of *B. nelumbiifolia* were made in the Mexican Gulf region, while samples of *B. heracleifolia* were made over a broader sampling distance from South Mexico to Guatemala (Fig. 4.1). Population samples of less than 15 individuals were included if this was all the plants at a given sampling locality. A representative specimen from each population was placed in the herbarium in Edinburgh (E) except the Guatemalan population which is in the University of Guatemala herbarium (BIGU). DNA extraction from silica dried material was performed using a modified protocol for the DNeasy 96-sample kit (Qiagen, Germantown, MD, USA), described in Chapter 2. A preliminary test of nuclear microsatellite amplification was made with the 14 nuclear microsatellite loci listed in Chapter 3. The 9 loci which amplified uniformly across species and populations were then used for the full genotyping of the population samples, with the same PCR protocol and amplification program. Input files for genetic analysis were formatted with the Microsatellite Toolkit (Park, 2001) and the file conversion program Create (Coombs *et al.*, 2008). To test whether data from microsatellite loci are independent of each other, linkage disequilibrium between markers was tested in FSTAT 1.2 (Goudet, 1995).
Begonia species can reproduce asexually, and clumps of clonal individuals may arise by rhizomatous growth or from vegetative material being broken off and rooting (A.D. Twyford, pers. obs.). The probability that individuals in a population shared identical genotypes at all 9 loci through random mating was calculated using the approach of Parks & Werth (1993) implemented in GenClone 2.0 (Arnaud-Haond & Belkhir, 2007). Genotypes not considered a product of sexual mating (\(P_{\text{sex}}\)) at \(P < 0.01\) when non-random mating was allowed (positive \(F_{\text{IS}}\) value), were removed from analyses. Individuals from populations where putative hybrids were found were only included if they were confidently considered pure species (i.e. assigned to a

---

**Figure 4.1.** Collection sites of *B. heracleifolia* (orange circles) and *B. nelumbiifolia* (blue squares) from South Mexico and Guatemala.
‘pure’ parental class with at least 95% probability, discussed in Chapter 5). This approach was validated as removing populations where hybrids occurred did not affect values of F-statistics (results not shown).

4.3.3. Measures of genetic diversity

Diversity statistics were calculated per locus and per population in FSTAT v1.2 (Goudet, 1995). The statistics calculated were: number of alleles (A, calculated per locus only), allelic richness corrected for sample sizes (Ae, rarefaction method, Mousadik & Petit, 1996), and gene diversity (hs, Nei, 1987). The number of private alleles per population were scored by eye.

4.3.4. Measures of inbreeding

Inferences of mating systems in natural populations, including self-fertilization and inbreeding, can be estimated from co-dominant genetic data. Weir and Cockerham’s (1984) FIS is a single-locus estimator of heterozygosity deficiency, and the FIS values were calculated and the average across-loci value bootstrapped in FSTAT v1.2. For each species the inferred selfing rate was calculated from the inbreeding coefficient, using the formula of Allard et al. (1969):

\[ S = \frac{2F_{IS}}{1 + F_{IS}} \]

Rates of self-fertilization were also estimated with a multi-locus inbreeding estimator in the program RMES, which considers heterozygote deficiency across all loci simultaneously, making it less sensitive to technical errors at a few loci (David et al., 2007; Jarne & David, 2008).

Self-compatibility was confirmed using greenhouse-grown plants. One cultivated accession of B. nelumbiifolia and five accessions of B. heracleifolia from different populations were self fertilised, and the presence of fertilised seeds (which appear ‘filled’ rather than ‘shrivelled’) checked under a 10x dissecting microscope. Seeds were germinated in 9cm pots of finely sieved bark, which were kept in a propagator at 25°C, and germination was recorded after 6 weeks.
4.3.5. Measures of population structure and genetic differentiation

F-statistics were calculated between sampling sites for each species, as well as between the distinct genetic clusters identified in the BAPS analysis (see below). To allow like-for-like comparisons between the two species over a similar geographic area, F-statistics were also calculated for *B. heracleifolia* excluding populations from south Oaxaca (h28) and Guatemala (h-g1). Weir & Cockerham’s (1984) estimator of F$_{ST}$, which is a measure of the genetic structure in the data, was calculated in FSTAT v1.2. A standardised measure of population structure that takes into account sample sizes and allelic diversity (F’$_{ST}$) was calculated by using RecodeData v0.1 (Meirmans, 2006) and FSTAT. Absolute differentiation was measured with D (Jost, 2008) using SMOGD (Crawford, 2010), and confidence intervals calculated by bootstrapping with 1000 replicates. An F$_{ST}$ analogue which incorporates allele size length (R$_{ST}$) was calculated with SPAGeDI v1.3 (Hardy & Vekemans, 2002), to test for phylogeographical structure in the data. The between species F$_{ST}$ was also calculated, which shows the similarities in allele frequencies between the species.

The ratio of interpopulation pollen to seed dispersal can be estimated using formula 5a of Ennos (1994), which relates the F$_{ST}$ for biparentally inherited nuclear markers (F$_{ST(b)}$) and maternally inherited plastid markers (F$_{ST(m)}$), as well as the level of inbreeding (F$_{IS}$). Maternal inheritance of plastids in *Begonia* have been confirmed by cytological observations (Corriveau & Coleman, 1988) and sequencing plastid DNA in experimental crosses (Peng & Chiang, 2000). No variation was found at 7 plastid microsatellite markers in *B. nelumbiifolia* (Chapter 2), so the pollen to seed ratio could not be calculated for this species. For *B. heracleifolia*, F$_{ST(m)}$ was calculated from the plastid microsatellite data in Chapter 2 using the same populations sampled for the nuclear microsatellites (this chapter), and compared to the F$_{ST(b)}$ and the F$_{IS}$.

To test whether dispersal between populations fits a simple model of dispersal limitation between more distant populations (the stepping stone model of dispersal, Kimura & Weiss, 1964), the relationship between genetic similarity and geographic distance was tested using isolation by distance analysis. Pairwise comparisons of F$_{ST}$ / (1 - F$_{ST}$) for each population were
plotted against the natural logarithm of geographic distance as suggested by Rousset (1997), and implemented in the Isolation by Distance Web Service v3.21 (Jensen et al., 2005).

To visualise the spatial structuring of populations, 3 complementary Bayesian clustering methods were used: STRUCTURE (Pritchard et al., 2000), BAPS (Corander et al., 2008) and InStruct (Gao et al., 2007). A comparison was then made between the genetic clusters assigned in the three different programs, to evaluate whether the results are robust to the different assumptions made each of the programs (Kettle et al., 2007). STRUCTURE assigns individuals to one (or more) of the user defined number of genetic clusters (K value), minimizing Hardy-Weinberg disequilibrium and linkage disequilibrium within each cluster (Pritchard et al., 2000), and maximizing disequilibrium between clusters. Preliminary STRUCTURE results were consistent for short runs (100 00 burn-in iterations, run 100 000 iterations) and long runs (burn-in 300 000 iterations, run 1 000 000 iterations), and summary statistics were stationary for both run lengths suggesting the data had converged (Pritchard et al., 2000), so the shorter run conditions were used. The admixture model was used, with sampling locality as a weak prior (Hubisz et al., 2009). K values between 1 and 13 were evaluated for B. heracleifolia, and 1 and 7 for B. nelumbifolia, with 10 independent replicates per K value. The ad hoc statistic ΔK was calculated across runs, and the greatest value inferred to be the optimal K value (Evanno et al., 2005). For analysis of a given K value, a consensus file correcting for equally optimal solutions (multi-modality) and label switching was produced in CLUMPP (Jakobsson & Rosenberg, 2007), and the results displayed in DISTRUCT (Rosenberg, 2004).

The same data file was used as the input for BAPS 5.4 (Corander et al., 2008), which uses an alternative assignment algorithm to STRUCTURE. The ‘clustering of groups of individuals’ setting was used, and the same number of genetic clusters (K) was evaluated as described above. Five replicates were made for each K value, and the results file was then used as the input for admixture analysis using the ‘mixture clustering option’. The minimum size of each population was set to 3 individuals, and runs were made of 10 000 iterations, and 5 000 reference individuals were used. The optimum value of K is automatically calculated in BAPS using a greedy stochastic optimization algorithm (Corander et al., 2008). The BAPS admixture bar plots produced by the program were used to display the results, with only significant admixture shown (P < 0.05).
As a moderate departure from Hardy-Weinberg equilibrium was detected for both species (see results), the data was also analysed in InStruct (Gao et al., 2007). This program is intended for analysing genetic structure from selfing or inbreeding species, where populations are not at Hardy-Weinberg equilibrium. The same run settings and K values were evaluated as for STRUCTURE, and the convergence of independent runs was confirmed by low Gelman-Rubin scores across runs (Gao, InStruct manual). The optimum number of K was assessed by the lowest value of the deviance information criterion (DIC). Runs were combined in CLUMPP and visualised in DISTRUCT (as above).

To visualise the relationship between populations, a neighbour-joining tree was constructed for each species. Allele frequencies from FSTAT were used as the input for POPTREE2 (Takezaki et al., 2010), where a neighbour-joining tree based on Nei’s (1983) measure of genetic distance was constructed. The tree was edited in FigTree v1.2.2 (available from http://tree.bio.ed.ac.uk/software/figtree/).

4.3.6. Measures of reproductive isolation

To test whether genetic incompatibilities have accumulated between differentiated populations of B. heracleifolia, pollen fertility was scored in artificial interpopulation crosses. It was not possible to grow crosses between every pair of populations due to the large number of populations and the limited greenhouse space, so instead three groups of crosses were made, and pollen sterility compared between the three groups. First, pollen sterility was recorded in selfed individuals, to establish a benchmark for pollen sterility. Second, pollen sterility was recorded in a small subset of crosses between populations across the Mexican gulf (the main area where population samples were collected), selecting the pollen donor at random from those in flower. Finally, pollen sterility was recorded in crosses between the population in Oaxaca (h28) and other populations, selecting the pollen donor at random. The Oaxacan population was chosen as it is the most divergent population with living material available. Seeds were germinated on finely sieved bark, and after 6 weeks seedlings were transferred to 9cm pots in sterilised potting mix (16 bark: 3 peat: 1 perlite plus finely sieved osmocote) and grown at 28°C to flowering. Pollen sterility was measured by acetocarmine staining, which is a reliable method for viability
assessments in *Begonia* (Matthews, 2007; Dewitte et al., 2011), and corresponds well to artificial pollen germination and fluorescent staining (Twyford and Kidner, unpubl. data). Dehiscing pollen from one flower per plant was stained with 1M acetocarmine, visualised under a Leica Microscope, and the proportion of well-stained pollen recorded out of 200 pollen grains.

4.3.7. Measures of genome size

Fully expanded leaf material from four accessions of *B. nelumbiifolia* and five of *B. heracleifolia* were selected for nuclear DNA content (C-value) estimates by flow cytometry. These accessions represent different populations from South Mexico (Table 4.3). The procedure followed Brennan et al. (submitted), summarised here briefly. Approximately 1 cm² of material was chopped with an internal size standard (*Solanum lycopersicum* ‘Stupíké polní rané’ 2C = 1.96 pg, or *Petroselinum crispum* ‘Champion Moss Curled’ 2C = 4.45 pg) in 2 mL of General purpose isolation buffer (GPB; Loureiro et al., 2007) with 3% PVP-40 following the procedure described by Doležel et al. (2007). The nuclear suspension was then filtered through a 30 µm nylon mesh to remove debris, stained with propidium iodide (Sigma-Aldrich) at a final concentration 60 µg·mL⁻¹ and supplemented with 100 mg·mL⁻¹ ribonuclease A (RNase A; Sigma-Aldrich). Samples were kept on ice for 15 min, and then 5 000 particles were recorded using a Partec Cyflow SL3 flow cytometer (Partec GmbH, Canterbury, UK) fitted with a 100-mW green solid state laser (Cobolt Samba, Solna, Sweden). Two to four replicates per accession were processed and the resulting fluorescence histograms were analysed with FlowMax software (Partec GmbH). The mean and standard error for each individual, and per species, were calculated.

4.4. Results

4.4.1. Genetic diversity

A total of 306 individuals from *B. heracleifolia* and 177 from *B. nelumbiifolia* were genotyped (Appendix 2 & 3, respectively). Five individuals of *B. heracleifolia* were identical at 9 polymorphic loci, and are likely to be clones (P < 0.01), so were removed from the data set.
Table 4.1. Collection sites and estimates of genetic diversity per population. \(N_g\), number of individual genotyped; \(N_{adj}\), adjusted number of individuals after suspected clones removed, \(A_e\), allelic richness rarefied to 5 individuals; \(h_S\), gene diversity; \(F_{IS}\), inbreeding coefficient.

<table>
<thead>
<tr>
<th>Pop.</th>
<th>Locality</th>
<th>State</th>
<th>Latitude</th>
<th>Longitude</th>
<th>(N_g)</th>
<th>(N_{adj})</th>
<th>(A_e)</th>
<th>(h_S)</th>
<th>(F_{IS})</th>
</tr>
</thead>
<tbody>
<tr>
<td>h2</td>
<td>4km NE San Andrés Tuxtla</td>
<td>Veracruz</td>
<td>18.4785</td>
<td>-95.178</td>
<td>19</td>
<td>19</td>
<td>1.870</td>
<td>0.257</td>
<td>0.01</td>
</tr>
<tr>
<td>h3</td>
<td>7km NE San Andrés Tuxtla</td>
<td>Veracruz</td>
<td>18.50341</td>
<td>-95.1682</td>
<td>34</td>
<td>33</td>
<td>1.970</td>
<td>0.263</td>
<td>0.224</td>
</tr>
<tr>
<td>h5</td>
<td>9km NE San Andrés Tuxtla</td>
<td>Veracruz</td>
<td>18.52009</td>
<td>-95.1618</td>
<td>26</td>
<td>26</td>
<td>1.685</td>
<td>0.209</td>
<td>0.111</td>
</tr>
<tr>
<td>h8</td>
<td>Agua Azul</td>
<td>Chiapas</td>
<td>17.22117</td>
<td>-92.1107</td>
<td>14</td>
<td>14</td>
<td>1.642</td>
<td>0.222</td>
<td>0.514</td>
</tr>
<tr>
<td>h12</td>
<td>Ocozocuatla biosphere reserve</td>
<td>Chiapas</td>
<td>16.94645</td>
<td>-93.4565</td>
<td>28</td>
<td>27</td>
<td>1.860</td>
<td>0.286</td>
<td>0.294</td>
</tr>
<tr>
<td>h13</td>
<td>Near Berriozábal</td>
<td>Chiapas</td>
<td>16.86693</td>
<td>-93.3278</td>
<td>6</td>
<td>6</td>
<td>1.941</td>
<td>0.292</td>
<td>0.684</td>
</tr>
<tr>
<td>h15</td>
<td>5km south of Orizaba</td>
<td>Veracruz</td>
<td>18.78253</td>
<td>-97.0851</td>
<td>27</td>
<td>27</td>
<td>2.004</td>
<td>0.300</td>
<td>0.083</td>
</tr>
<tr>
<td>h21</td>
<td>Highway 175 between Jasaa and San Juan Bautista</td>
<td>Oaxaca</td>
<td>17.74356</td>
<td>-96.328</td>
<td>31</td>
<td>31</td>
<td>2.250</td>
<td>0.390</td>
<td>0.338</td>
</tr>
<tr>
<td>h23</td>
<td>Near Santa Maria Jacatapec</td>
<td>Oaxaca</td>
<td>17.86422</td>
<td>-96.2138</td>
<td>25</td>
<td>25</td>
<td>2.686</td>
<td>0.491</td>
<td>0.316</td>
</tr>
<tr>
<td>h24</td>
<td>Motzorongo</td>
<td>Oaxaca</td>
<td>18.66953</td>
<td>-96.7871</td>
<td>37</td>
<td>37</td>
<td>1.939</td>
<td>0.280</td>
<td>0.274</td>
</tr>
<tr>
<td>h26</td>
<td>San Jeronimo Zoochina</td>
<td>Oaxaca</td>
<td>17.22117</td>
<td>-95.2355</td>
<td>28</td>
<td>28</td>
<td>1.927</td>
<td>0.269</td>
<td>0.171</td>
</tr>
<tr>
<td>h28</td>
<td>Santa Maria Xanabi</td>
<td>Oaxaca</td>
<td>15.98808</td>
<td>-96.1106</td>
<td>12</td>
<td>12</td>
<td>1.436</td>
<td>0.154</td>
<td>0.491</td>
</tr>
<tr>
<td>h-g1</td>
<td>Cobán</td>
<td>Alta Verapaz (Guatemala)</td>
<td>15.47552</td>
<td>-90.3787</td>
<td>26</td>
<td>23</td>
<td>1.709</td>
<td>0.169</td>
<td>0.22</td>
</tr>
</tbody>
</table>

Average across populations

<table>
<thead>
<tr>
<th>Pop.</th>
<th>Locality</th>
<th>State</th>
<th>Latitude</th>
<th>Longitude</th>
<th>(N_g)</th>
<th>(N_{adj})</th>
<th>(A_e)</th>
<th>(h_S)</th>
<th>(F_{IS})</th>
</tr>
</thead>
<tbody>
<tr>
<td>h1</td>
<td></td>
<td></td>
<td>24.1</td>
<td>23.7</td>
<td>1.918</td>
<td>0.276</td>
<td>0.287</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

B. nelumbifolia

<table>
<thead>
<tr>
<th>Pop.</th>
<th>Locality</th>
<th>State</th>
<th>Latitude</th>
<th>Longitude</th>
<th>(N_g)</th>
<th>(N_{adj})</th>
<th>(A_e)</th>
<th>(h_S)</th>
<th>(F_{IS})</th>
</tr>
</thead>
<tbody>
<tr>
<td>n1</td>
<td>Near Los Tuxtlas biological station</td>
<td>Veracruz</td>
<td>18.59026</td>
<td>-95.0788</td>
<td>20</td>
<td>22.21</td>
<td>0.450</td>
<td>0.327</td>
<td></td>
</tr>
<tr>
<td>n3</td>
<td>7km NE San Andrés Tuxtla</td>
<td>Veracruz</td>
<td>18.50341</td>
<td>-95.1682</td>
<td>12</td>
<td>2.257</td>
<td>0.473</td>
<td>0.444</td>
<td></td>
</tr>
<tr>
<td>n5</td>
<td>9km NE San Andrés s Tuxtla</td>
<td>Veracruz</td>
<td>18.52009</td>
<td>-95.1618</td>
<td>30</td>
<td>2.397</td>
<td>0.492</td>
<td>0.328</td>
<td></td>
</tr>
<tr>
<td>n18</td>
<td>valley 25km East of Huastusco</td>
<td>Veracruz</td>
<td>19.20367</td>
<td>-96.7426</td>
<td>16</td>
<td>2.025</td>
<td>0.368</td>
<td>0.535</td>
<td></td>
</tr>
<tr>
<td>n21</td>
<td>Highway 175 between Jasaa and San Juan Bautista</td>
<td>Oaxaca</td>
<td>17.74356</td>
<td>-96.328</td>
<td>41</td>
<td>2.506</td>
<td>0.497</td>
<td>0.407</td>
<td></td>
</tr>
<tr>
<td>n25</td>
<td>Arroya Zacata</td>
<td>Oaxaca</td>
<td>17.73981</td>
<td>-95.7913</td>
<td>27</td>
<td>1.766</td>
<td>0.276</td>
<td>0.471</td>
<td></td>
</tr>
<tr>
<td>n26</td>
<td>San Jerónimo Zoochina</td>
<td>Oaxaca</td>
<td>17.22117</td>
<td>-95.2355</td>
<td>31</td>
<td>2.198</td>
<td>0.367</td>
<td>0.306</td>
<td></td>
</tr>
</tbody>
</table>

Average across populations

<table>
<thead>
<tr>
<th>Pop.</th>
<th>Locality</th>
<th>State</th>
<th>Latitude</th>
<th>Longitude</th>
<th>(N_g)</th>
<th>(N_{adj})</th>
<th>(A_e)</th>
<th>(h_S)</th>
<th>(F_{IS})</th>
</tr>
</thead>
<tbody>
<tr>
<td>n1</td>
<td></td>
<td></td>
<td>25.3</td>
<td>21.96</td>
<td>0.417</td>
<td>0.403</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Descriptive statistics are summarised per population in Table 4.1 and per locus in Table 4.2. All loci were polymorphic, except B5347 which was monomorphic in B. nelumbiifolia. Overall, the levels of genetic diversity were low for both B. heracleifolia (mean values across populations: \( A = 6.6; A_e = 1.972, h_S = 0.276 \)) and B. nelumbiifolia (\( A = 4.7; A_e = 2.197, h_S = 0.417 \)). The most genetically diverse population of B. heracleifolia was h23 (\( A_e = 2.686; h_S = 0.491 \)), and the least was h28 (\( A_e = 1.436; h_S = 0.154 \)). The most diverse population of B. nelumbiifolia was population n21 (\( A_e = 2.51; h_S = 0.497 \)), and the least was n25 (\( A_e = 1.766; h_S = 0.276 \)). Seventeen of the 54 alleles (29.8%) detected in B. heracleifolia were private alleles. The most private alleles were found in populations h15 and h-g1, with 3 each, and the least were in populations h8, h23 and h28, which had no private alleles. Nine of the 44 alleles (20.4%) found in B. nelumbiifolia were private alleles. The most private alleles (4) were found in populations n18 and n21, and one in n26. No private alleles were found in the other populations of B. nelumbiifolia.

Table 4.2. Species-level estimates of genetic diversity and genetic differentiation per locus. \( A \), total number of alleles per locus; \( A_e \), allelic richness rarefied to 5 individuals; \( h_T \) gene diversity. *, \( P < 0.05 \); **\( P < 0.01 \).

<table>
<thead>
<tr>
<th>Species</th>
<th>Locus</th>
<th>A</th>
<th>( A_e )</th>
<th>( h_T )</th>
<th>( F_{ST} )</th>
<th>( F_{IS} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>B. heracleifolia</td>
<td>B4329</td>
<td>4</td>
<td>2.745</td>
<td>0.302769</td>
<td>0.460**</td>
<td>0.118</td>
</tr>
<tr>
<td></td>
<td>B3043</td>
<td>6</td>
<td>2.422</td>
<td>0.181231</td>
<td>0.462*</td>
<td>0.689**</td>
</tr>
<tr>
<td></td>
<td>B6278</td>
<td>2</td>
<td>1.285</td>
<td>0.036385</td>
<td>0.676</td>
<td>0.700</td>
</tr>
<tr>
<td></td>
<td>B5347</td>
<td>7</td>
<td>2.922</td>
<td>0.316231</td>
<td>0.342**</td>
<td>0.324**</td>
</tr>
<tr>
<td></td>
<td>B3348</td>
<td>9</td>
<td>3.895</td>
<td>0.497615</td>
<td>0.312**</td>
<td>0.223**</td>
</tr>
<tr>
<td></td>
<td>B7112</td>
<td>6</td>
<td>1.656</td>
<td>0.116308</td>
<td>0.140</td>
<td>0.320**</td>
</tr>
<tr>
<td></td>
<td>B3820</td>
<td>10</td>
<td>4.115</td>
<td>0.484462</td>
<td>0.350**</td>
<td>0.046</td>
</tr>
<tr>
<td></td>
<td>B134</td>
<td>8</td>
<td>3.045</td>
<td>0.395923</td>
<td>0.269*</td>
<td>0.244*</td>
</tr>
<tr>
<td></td>
<td>B332</td>
<td>5</td>
<td>2.179</td>
<td>0.150308</td>
<td>0.492**</td>
<td>0.340*</td>
</tr>
<tr>
<td>Across loci</td>
<td>6.3</td>
<td></td>
<td>2.696</td>
<td>0.275692</td>
<td>0.364**</td>
<td>0.249**</td>
</tr>
<tr>
<td>B. nelumbiifolia</td>
<td>B4329</td>
<td>5</td>
<td>3.838</td>
<td>0.535714</td>
<td>0.266*</td>
<td>0.060</td>
</tr>
<tr>
<td></td>
<td>B3043</td>
<td>8</td>
<td>3.553</td>
<td>0.456857</td>
<td>0.188**</td>
<td>0.373*</td>
</tr>
<tr>
<td></td>
<td>B6278</td>
<td>5</td>
<td>2.863</td>
<td>0.471857</td>
<td>0.144*</td>
<td>0.421**</td>
</tr>
<tr>
<td></td>
<td>B5347</td>
<td>1</td>
<td>1</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>B3348</td>
<td>4</td>
<td>2.747</td>
<td>0.487571</td>
<td>0.125**</td>
<td>0.316**</td>
</tr>
<tr>
<td></td>
<td>B7112</td>
<td>4</td>
<td>2.128</td>
<td>0.335429</td>
<td>0.383*</td>
<td>0.372**</td>
</tr>
<tr>
<td></td>
<td>B3820</td>
<td>4</td>
<td>2.929</td>
<td>0.301571</td>
<td>0.554**</td>
<td>0.512*</td>
</tr>
<tr>
<td></td>
<td>B134</td>
<td>6</td>
<td>2.594</td>
<td>0.322714</td>
<td>0.142*</td>
<td>0.234**</td>
</tr>
<tr>
<td></td>
<td>B332</td>
<td>5</td>
<td>3.012</td>
<td>0.427714</td>
<td>0.315*</td>
<td>0.868**</td>
</tr>
<tr>
<td>Across loci</td>
<td>4.7</td>
<td></td>
<td>2.7065</td>
<td>0.371048</td>
<td>0.277**</td>
<td>0.454**</td>
</tr>
</tbody>
</table>
4.4.2. Inbreeding

The average $F_{IS}$ value across loci was 0.249 (SE 0.062) and 0.454 (SE 0.095) for *B. heracleifolia* and *B. nelumbiifolia*, respectively. $F_{IS}$ values varied from 0.01 in population h2 to 0.684 in population h13 (*B. heracleifolia*), and from 0.306 in population n26 to 0.535 in population n18 (*B. nelumbiifolia*), see Table 4.2. The values for the inferred selfing rate ($s$) calculated from the inbreeding coefficient $F_{IS}$, were 0.399 for *B. heracleifolia*, and 0.624 for *B. nelumbiifolia*.

The multiple-locus maximum likelihood (ML) estimate of selfing ($s$) averaged across loci for *B. heracleifolia* was 0.154 (non-significant), and only population h23 had a selfing rate significantly different from 0 [ML $s = 0.302$, 95% confidence interval (CI) 0.039 - 0.474]. For *B. nelumbiifolia*, the ML estimate of $s$ averaged across loci was 0.097 (non-significant). Only one population had a selfing rate significantly different from 0 (population n18; ML $s = 0.554$; 95% CI: 0.057 - 0.759).

Seed set was high ( > 90%) in the self-pollination experiment with *B. heracleifolia* ($n = 5$) and *B. nelumbiifolia* ($n = 1$).

4.4.3. Population structure and genetic differentiation

Significant population substructure was found by $F_{ST}$ analysis (average across loci, *B. heracleifolia*, $F_{ST} = 0.364$; SE 0.028, $P < 0.05$; *B. nelumbiifolia*, $F_{ST} = 0.277$; SE 0.055, $P < 0.05$), as well as when the $F_{ST}$ was standardised for the maximum possible value for the loci sampled (*B. heracleifolia* $F'_{ST} = 0.506$, *B. nelumbiifolia* $F'_{ST} = 0.439$). Moderate levels of differentiation were found with Jost’s estimator of absolute differentiation $D$, with average values per locus for *B. heracleifolia* ($D = 0.274$, $P < 0.05$), and for *B. nelumbiifolia* ($D = 0.294$, $P < 0.05$). Recalculating measures of population substructure and genetic differentiation for the genetic clusters identified in the BAPS analysis for *B. heracleifolia* had little effect (results not shown). $R_{ST}$ values, which use allele size lengths and a stepwise mutation model, were significantly different from 0 (*B. heracleifolia* $R_{ST} = 0.212$; *B. heracleifolia* $R_{ST} = 0.257$), but were not significantly larger than the $G_{ST}$ values (not shown), suggesting the absence of
phylogeographical structure. Isolation-by-distance accounted for a modest amount of the genetic variance (\(B. \text{heracleifolia}, r^2 = 0.250, P = 0.001\); \(B. \text{nelumbiifolia}, r^2 = 0.289, P = 0.016\)). The \(F_{ST}\) value between species, which is a measure of genetic similarity between them, was 0.466 (\(P < 0.01\)). When the Oaxacan and Guatemalan populations of \(B. \text{heracleifolia}\) were removed from the analyses, leaving the 11 populations from the Mexican Gulf across a similar range as \(B. \text{nelumbiifolia}\), values of genetic differentiation and substructure were reduced (\(F_{ST} = 0.306, \text{SE} 0.04; F'_{ST} = 0.437; D = 0.219, P < 0.05\)), and more of the variation could be explained by isolation-by-distance (\(r^2 = 0.383, P = 0.001\) Mantel Test \(r = 0.619\)).

Using the \(F_{ST}\) values from the plastid microsatellites (\(F_{ST(m)} = 0.728\)) and the nuclear microsatellites (\(F_{ST(b)} 0.364\)), and also considering inbreeding (\(F_{IS} = 0.249\)), the ratio of pollen to seed dispersal for \(B. \text{heracleifolia}\) was 3.8. Transmission of genes through pollen is therefore c.4 times more effective than through the seed.

The optimal number of genetic clusters for \(B. \text{heracleifolia}\), assessed by comparing the mean DIC value across runs for InStruct, and the maximum \(\Delta K\) value for STRUCTURE, was 10. The most likely number of genetic clusters in the BAPS analysis was \(K=11\) [log likelihood (LL) of \(K = -3428.614\)], closely followed by \(K = 10\) (LL = -3429.484). For \(B. \text{nelumbiifolia}\) the optimal value for \(K\) was 7 for all three approaches.

The genetic clustering results for the three Bayesian methods gave similar major groups, but differed in the amount of admixture between them. For \(B. \text{heracleifolia}\), the genetic clusters identified in BAPS were the same as the sampled populations, with two exceptions (Fig. 4.2a). Populations h3 and h5, which are separated by less than 2km, shared a common gene pool. BAPS assigned population h13, which contained only 6 individuals, as admixed between populations h3, h5, and h8 (270km away). The STRUCTURE analyses (Appendix Figure 4) differed from the BAPS analysis by having greater admixture between less clearly defined genetic clusters, and placing the closely grouped (< 5km apart) populations h2, h3 and h5 together. Population h8 was poorly defined, with a complex pattern of admixture. The InStruct analysis showed even fewer distinct genetic clusters (Appendix Figure 4), with 6 of the 13 populations being recognised as distinct clusters (h12, h21, h23, h24, h28, h-g1).
Figure 4.2. Bayesian assignment to genetic clusters and the relationships between populations. Bayesian admixture results in BAPS for *B. heracleifolia* (a) and *B. nelumbiifolia* (b). Each individual is represented by a vertical bar, and different colours represent the different genetic clusters. Asterisk indicates population h13. Neighbour-joining trees of Nei’s (1983) measure of population divergence for *B. heracleifolia* (c) and *B. nelumbiifolia* (d). Branches are coloured to correspond with the genetic clusters from the BAPS analyses.
Bayesian clustering in BAPS revealed 7 main clusters corresponding to the 7 populations genotyped for *B. nelumbiifolia* (Fig. 4.2b). Levels of admixture were low, with 9 individuals (5%) having admixed ancestry. As with *B. heracleifolia*, the STRUCTURE results showed less resolution (Appendix Figure 4) than BAPS, with only 3 populations being assigned to distinct genetic cluster (populations h18, h25, h26). The InStruct analysis (Appendix Figure 4) detected even fewer distinct genetic clusters than STRUCTURE.

The average value for Nei’s 1983 pair-wise population distance measures, which were used to build the Neighbour-joining trees, were 0.231 (0.011SE) for *B. heracleifolia*, and 0.201 (0.013SE) for *B. nelumbiifolia*. The tree for *B. heracleifolia* included two long branches, connecting populations h28 (mean pair-wise $D_{Nei} = 0.360$) and h-g1 (mean pair-wise $D_{Nei} = 0.313$; Fig. 4.2c). The tree for *B. nelumbiifolia* was roughly star-shaped (Fig. 4.2d).

### 4.4.4. Reproductive isolation

The mean pollen stainability of selfed plants was 97.5% (SE 0.0126, n = 18), and only one plant had a value below 85%. Outcrossed plants were similarly pollen-fertile, with a mean pollen stainability of 98.7% (SE = 0.0197, n = 10). Outcresses involving individuals from Oaxaca had a mean viability of 78.3% (SE = 0.0170, n = 3).

### 4.4.5. Genome size

The average 1C genome size estimated for *B. heracleifolia* and *B. nelumbiifolia* were 0.80pg and 0.54pg, respectively (Table 4.3). Genome size estimates were consistent across individuals within species, except a *B. heracleifolia* individual from population h28 that had an estimated genome size of 0.88pg, which is 10% higher than the parental mean.
Table 4.3. Collection sites and estimates of genome sizes. Standard errors of genome sizes given in brackets.

<table>
<thead>
<tr>
<th>Pop. code</th>
<th>Locality</th>
<th>State</th>
<th>Latitude</th>
<th>Longitude</th>
<th>1C genome size in pg</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>B. heracleifolia</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>h2</td>
<td>4km NE San Andrés Tuxtla</td>
<td>Veracruz</td>
<td>18.47850</td>
<td>-95.17800</td>
<td>0.79 (0.00)</td>
</tr>
<tr>
<td>h13</td>
<td>Near Berriozábal</td>
<td>Chiapas</td>
<td>16.86693</td>
<td>-93.32780</td>
<td>0.78 (0.00)</td>
</tr>
<tr>
<td>h16</td>
<td>30km east of Huatusco</td>
<td>Veracruz</td>
<td>19.2011</td>
<td>-96.67139</td>
<td>0.78 (0.00)</td>
</tr>
<tr>
<td>h21</td>
<td>Highway 175 between Jasaa and San Juan Bautista</td>
<td>Oaxaca</td>
<td>17.74356</td>
<td>-96.32800</td>
<td>0.78 (0.00)</td>
</tr>
<tr>
<td>h28</td>
<td>Santa Maria Xanabi</td>
<td>Oaxaca</td>
<td>15.98808</td>
<td>-96.11060</td>
<td>0.88 (0.00)</td>
</tr>
<tr>
<td><strong>B. nelumbiifolia</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n7</td>
<td>Road to Ocosingo from Palenque</td>
<td>Chiapas</td>
<td>17.42477</td>
<td>-91.99712</td>
<td>0.55 (0.00)</td>
</tr>
<tr>
<td>n11</td>
<td>Ocozocuata biosphere reserve</td>
<td>Chiapas</td>
<td>16.92489</td>
<td>-93.45090</td>
<td>0.54 (0.00)</td>
</tr>
<tr>
<td>n19</td>
<td>3km north of Josaa</td>
<td>Oaxaca</td>
<td>17.64247</td>
<td>-96.337472</td>
<td>0.53 (0.00)</td>
</tr>
<tr>
<td>n21</td>
<td>Motzorongo</td>
<td>Oaxaca</td>
<td>18.66953</td>
<td>-96.787139</td>
<td>0.54 (0.00)</td>
</tr>
</tbody>
</table>

4.5. Discussion

4.5.1. Strong population substructure and genetic differentiation in widespread Begonia species

The two widespread Begonia species showed strong population substructure (B. heracleifolia $F_{ST} = 0.364, P < 0.05$; B. nelumbiifolia, $F_{ST} = 0.277, P < 0.05$), and Bayesian structure analysis
supports the division of most populations into discrete genetic clusters (Fig. 4.2). Moreover, levels of genetic diversity in each population were low \((B.\ heracleifolia\ \text{mean} \ h_S = 0.276; \ B.\ nelumbiifolia\ \text{mean} \ h_S = 0.417)\), and populations showed significant genetic differentiation between them \((B.\ heracleifolia\ \text{Jost’s} \ D = 0.274, \ P < 0.05; \ B.\ nelumbiifolia\ D = 0.294, \ P < 0.05)\). These data suggest *Begonia* populations are isolated, with little homogenizing gene flow between them. Despite the small population sizes of the two species (c.10-300 individuals, A.D. Twyford pers obs.; Hoover, 1979), populations may be long-lived and stable allowing them to become differentiated by drift or local adaptation. These results are consistent with the plastid analysis presented in Chapter 2, where a pattern of strong genetic differentiation was found. In a broader context, long lived and isolated *Begonia* populations have been inferred from population genetic analysis of three other *Begonia* species \((B.\ sutherlandii, \ Hughes & Hollingsworth, 2008; B.\ dregei, B.\ homonyma, \ Matolweni \textit{et al.}, 2000)\), suggesting that a common set of population-level mechanisms (strong genetic drift in isolated populations) play a role in the evolution of intra-specific genetic diversity in *Begonia*.

Pollen-mediated gene flow was found to be low between populations of *B. heracleifolia*. A pollen to seed ratio \((r)\) of 3.8 is at the lower end of the range seen in plants \((\text{e.g.} \ r = 1 - 196, \ Ennos, 1994; \ Squirrell \textit{et al.}, 2001)\). Similarly low pollen to seed ratios \((r = 4.0)\) have been found for other herbaceous plants such as *Dysosma versipellis* (Hance) M.Cheng ex T.S. Ying (Berberidaceae) \((\text{Guan \textit{et al.}, 2010})\), or other plant species that grow below the forest canopy, such as the epiphytic bromeliad *Vriesea gigantea* Lem. (Bromeliaceae) \((r = 3.3, \ Palma-Silva \textit{et al.}, 2009)\). This low level of pollen dispersal is likely related to the pollination biology of *Begonia*. Female *Begonia* flowers do not produce any reward for pollinating insects; instead they are intersexual mimics of the male flowers, which reward pollinators with pollen but not nectar \((\text{Ågren & Schemske, 1991; Schemske \textit{et al.}, 1996})\). This low-reward strategy may only be effective for attracting naive insects, as pollinators may develop an unfavourable search image of flowers with little reward and avoid them in the future \((\text{Goulson, 2000 and references therein})\). By delivering little reward for pollinators, there is little incentive for pollinators to travel far-afield to another *Begonia* plant, making the chances of interpopulation pollen dispersal low. It is however an effective strategy to reduce the high costs of producing floral rewards, and by
attracting generalist pollinators seed set may be relatively high (Wyatt & Sazima, 2011) when the isolated nature of most *Begonia* populations is considered.

### 4.5.2. Inbreeding promotes genetic differentiation

Both *Begonia* species were found to be fully self-compatible, and a moderate amount of self-fertilization occurs in natural populations (40% *B. heracleifolia*, 62% *B. nelumbiifolia*). The occurrence of self-fertilization was supported by the consistent $F_{IS}$ values across loci, although the average across-loci result for the multi-allelic estimator was non-significant. This is likely to be due to the limited diversity at each of the EST-derived microsatellite loci, and more loci, or loci with higher levels of polymorphism, would be required to obtain a significant result. This will also be influenced by the interpopulation variation in selfing rates (Table 4.1). A positive $F_{IS}$ value could also be due to technical artefacts, such as null alleles (Pemberton *et al.*, 1995). The presence of null alleles seems unlikely because there were few individuals that could not be amplified for a given locus (null homozygotes), and where these did occur they did not fit any consistent pattern (e.g. being common in a particular population). Moreover, the primers were designed from transcriptome sequence data and have been shown to be conserved over a broad phylogenetic scope (Chapter 3), suggesting interpopulation polymorphisms in the primer regions are unlikely.

Most *Begonia* species are self compatible (Ågren & Schemske, 1993; Dewitte *et al.*, 2011; Wyatt & Sazima, 2011; Twyford and Kidner, unpubl. data), with only two species studied to date not setting seed in a small number of experimental crosses (Brazilian *B. integerrima* Spreng. and *B. itatinensis* Irmsch. Ex Brade; Wyatt & Sazima, 2011). Levels of inbreeding in *Begonia* species varies, from close to panmictic (*B. sutherlandii*, mean $F_{IS} = 0.158, 7$ microsatellies) to fully selfing (*B. hirsuta* single locus outcrossing rate $0.03 \pm 0.01, 1$ isozyme locus, Ågren & Schemske, 1993). The mechanisms underlying inbreeding in *Begonia* are not currently clear; it may either be autopollination or self-fertilization mediated by insect pollinators (geitonogamy). Most *Begonia* species are monoecious and functionally protandrous, producing male flowers on an inflorescence before females (Forrest & Hollingsworth, 2003). However, as many inflorescences are borne over a flowering season, there are plenty of
opportunities for self-pollination. This would particularly be the case for species such as *B. nelumbiifolia*, which produces many densely-packed inflorescences that can easily become intertwined. Self-pollination in such a way would assure mating success if pollinators are rare, which may be expected in the isolated populations *Begonia* species typically grow.

### 4.5.3. Seed dispersal limitation and self-fertilization promote allopatric divergence

The $F_{ST}$ values between populations for *B. heracleifolia* and *B. nelumbiifolia*, as well as those for other *Begonia* species (Matolweni *et al*., 2000; Hughes & Hollingsworth, 2008), indicate a higher level of population structure than most other plant species (Gitzendanner & Soltis, 2000; Petit *et al*., 2005; Hey & Pinho, 2012). Plant species with higher $F_{ST}$ values are typically distributed across a much more heterogenous environment with large geographic barriers impeding gene flow, or are highly selfing. One example of this is the bromeliad *Pitcairnia geyskesii* L.B.Sm., where there is strong population structure ($F_{ST} = 0.533$) between the large emergent rocky outcrops (inselbergs) where the species grows in French Guiana (Boisselier-Dubayle *et al*., 2010). Selfing species with high $F_{ST}$ values include *Arabidopsis thaliana* L. ($F_{IS} = 0.969$, $F_{ST} = 0.61$, Bomblies *et al*., 2010); *Bromus tectorum* L. ($F_{IS} = 1$, $F_{ST} = 0.53$; Ramakrishnan *et al*., 2006) and *Medicago truncatula* Gaertn. ($F_{IS} = 0.978$, $F_{ST} = 0.3 - 0.75$, Siol *et al*., 2008). While the highest levels of genetic divergence were found between populations separated by a geographic barrier, such as the Southern Oaxacan population h28 (mean pair-wise $F_{ST} = 0.463$, range 0.342 - 0.638) which is isolated from other populations by the Sierra Madre Occidental, relatively high pairwise $F_{ST}$ values were still found between populations in close proximity across semi-continuous habitats. The absence of mechanisms that promote dispersal of pollen and seed in *Begonia* therefore appear to prevent gene flow even over short distances.

To link divergence in allopatry with the potential for speciation, we estimated pollen fertility between crosses from the isolated Oaxacan population and other populations from the Mexican Gulf. We then used flow cytometry to shed light on whether genome size changes have also occurred. We showed a $\sim$20% reduction in pollen stainability in wide outcrosses, and a $\sim$10% increase in C-value of the Oaxacan population relative to the species mean (Table 4.3). This marked intraspecific C-value variation, and reduced fertility of intraspecific crosses, support
populations being isolated from homogenizing gene flow allowing them to diverge on a potential route towards speciation.

4.6. Conclusion

The mean $F_{ST}$ values for these species, and the other *Begonia* species studied to date, approach or exceed the threshold value of $F_{ST} = 0.35$ suggested by Hey & Pinho (2012) for delimiting species. In addition to this strong geographic structure caused by dispersal limitation, inbreeding may further reduce the level of gene flow between populations and promote divergence. The joint role that genetic drift and inbreeding play on levels of genetic differentiation in speciose tropical plants (the Baker-Fedorov hypothesis, Lasso *et al.*, 2011) is now beginning to be appreciated. Moreover, we have shown reduced fertility in crosses between divergent populations within a species, supporting the hypothesis that differentiation leads to the accumulation of genetic incompatibilities that may be involved in reproductive isolation, in a trajectory towards allopatric speciation.
CHAPTER 5: Reproductive isolating barriers revealed by a comparative hybrid swarm analysis in a recent *Begonia* species radiation

**Data contributions:** Alexander Twyford collected all the genotypic and phenotypic measures and performed the analyses.

### 5.1. Chapter summary

*Begonia* is one of the most species-rich angiosperm genera (1500 species), with related species showing striking variation in their morphology and ecology. One hypothesis for this species richness is that reproductive barriers evolve rapidly between incipient species because homogenising gene flow between them is low. Introgression may be reduced further by selection acting on hybrids between ecologically divergent species. Here, we test the strength of reproductive isolation between ecologically similar *Begonia heracleifolia* and *B. sericoneura*, and between ecologically divergent *B. heracleifolia* and *B. nelumbifolia*. Reproductive isolating barriers (hybrid vigour, sterility, phenological differences) were first tested under experimental conditions and the outcomes of natural hybridization then observed by genotyping replicate hybrid swarms with twelve nuclear and seven plastid microsatellite markers. Overall, reproductive barriers were relatively strong, with hybrid sterility, frequent self-fertilization, and differences in flowering time contributing to isolation. Most barriers were more pronounced between ecologically divergent *B. heracleifolia* and *B. nelumbifolia*. In contact zones between *B. heracleifolia* and *B. sericoneura* F1s were common but introgression limited; in the *B. heracleifolia* x *B. nelumbifolia* contact zones F1s were less common and introgression absent. Reproductive barriers appear to evolve rapidly between *Begonia* species, which may be further enhanced by ecological selection against hybrids. These factors, in conjunction with the population biology of *Begonia*, may have promoted speciation in this species-rich genus.

### 5.2 Introduction

Tropical forests harbour half the world’s plant species (Corlett & Primack, 2010), with much of this species richness contributed by a few species-rich genera (or families) (Gentry, 1992;
Frodin, 2004). The cradle model of tropical evolution proposes a recent origin for this species richness (Stebbins, 1974), and this model has gained support from dated phylogenetic analyses of some large plant genera (e.g. Inga, Richardson et al., 2001; Guatteria, Erkens et al., 2007). While much emphasis has been placed on the timing and rates of speciation in the tropics, it is currently unclear how so many new species can arise from a common ancestor over such a short period of time, when reproductive barriers are weak and homogenising gene flow may be strong (Chan & Levin, 2005). A greater understanding of the evolution of reproductive isolation in large genera may shed light on why species diversity has accumulated in biodiversity hotspots such as the tropics.

Ecological selection may play a prominent role in large genera where species richness is recent in origin, because ecological divergence evolves early during speciation and often promotes it (Butlin, 2010). The role ecological selection plays in maintaining species barriers can be tested by comparing reproductive isolation between species that share similar ecologies, and those that differ. One may predict that ecological divergence between hybridizing species will affect pre-pollination barriers more than post-pollination barriers, because ecologically divergent species are less likely to be found in sympatry, but post-pollination barriers will be influenced more by differences in genome structure than by different ecologies (endogenous selection, Jiggins & Mallet, 2000). Alternatively, post-pollination barriers may also be affected by ecological divergence, as hybrids will combine divergent alleles at adaptive loci that may be incompatible (exogenous selection, Jiggins & Mallet, 2000). To compare intrinsic barriers and the role of selection, the strength of each reproductive barrier can be tested under experimental conditions, and the expected frequency of hybrids compared with the number of hybrids in sympatric populations where all reproductive barriers interact.

A particularly suitable system for this combined experimental and field based approach is Begonia, one of the most species-rich plant genera (>1500 species), where species richness has accumulated during recent radiations both in the Americas and in Asia (Miocene, Goodall-Copestake et al., 2009; Dewitte et al., 2011). Analyses of widespread Central American B. heracleifolia Cham. & Schltdl. and B. nelumbiifolia Cham. & Schltdl., have shown gene flow between populations is low (Chapters 2 & 4), and this may have led to a break down in species cohesion with reproductive barriers evolving between divergent populations of B. heracleifolia.
(F1 partial hybrid sterility, Chapter 4). This raises the hypothesis that allopatric speciation occurs in situ (autochthonously) in Begonia, with reproductive barriers evolving quickly due to the fixation of alleles in small founding populations, and such rapidly evolving barriers may promote speciation. However few studies have assessed the reproductive barriers that promote divergence or maintain species barriers during secondary contact between Begonia species.

This study uses estimates of reproductive barriers under experimental conditions, and the frequency of hybrids in nature, to shed light on the evolution of reproductive isolation between two pairs of Central American Begonia species. We predict that reproductive isolation between species will be relatively strong, and in conjunction with the low gene flow typical between populations (Chapter 2; Chapter 4; Matolweni et al., 2000; Hughes & Hollingsworth, 2008), this will facilitate rapid speciation. The second aim is to assess the contribution ecological selection plays in maintaining species barriers, through a comparison of species pairs that differ in their ecologies. We predict a prominent role for selection in a contact zone between species that differ markedly in their ecologies, i.e. that it will contain fewer hybrids and that these will be limited to early generation hybrids. Begonia heracleifolia and B. sericoneura Liebm. represent a pair of species that grow in similarly dry or seasonally dry habitats, while B. heracleifolia and B. nelumbiifolia are two species that grow in very different habitats (dry and permanently moist, respectively). For these two pairs of species we assess the possible reproductive barriers that act in sympatry (flowering phenology, chromosomal number differences, hybrid vigour, hybrid male and female fertility) to determine whether these are greater in the more ecologically divergent species pair. We then analyse the structure of hybrid zones between these pairs of species using nuclear and plastid microsatellites to determine whether differences in experimentally assessed reproductive barriers translate into differences in interspecific gene exchange in the wild.

5.3. Material and method

5.3.1. Study species
Two species pairs that frequently hybridize in the wild were chosen for this study: *B. heracleifolia* x *B. sericoneura*, and *B. heracleifolia* x *B. nelumbiifolia* (Burt-Utley, 1985; Morris, 2008). *Begonia sericoneura* has morphological similarities to *B. heracleifolia*, but can be identified by its persistent stipules and lack of coloured leaf markings (Fig. 5.1). Both species are similar in their ecologies, most often found in dry or seasonally dry habitats near roadsides, as well as in open areas in tropical forests (Burt-Utley, 1985). One hybrid swarm between the two species was analysed (HS4, Table 5.1).

**Figure 5.1.** *Begonia* hybrid triangle. Parental leaf shape and staminate and pistillate flowers are shown in each corner, with F1 hybrids between them. Natural hybrids between *B. nelumbiifolia* and *B. sericoneura* have not been recorded.
Begonia nelumbiifolia is very distinct from *B. heracleifolia*, as it has large unmarked peltate leaves and inflorescences with densely packed small flowers (Burt-Utley, 1985). *Begonia nelumbiifolia* grows in more moist shaded areas (Hoover, 1979; Burt-Utley, 1985). Both species are ruderals and co-occur in sympatry at disturbed roadsides, but within sites they are usually separated into different microhabitats (Hoover, 1979). At least 4 hybrid swarms have been recorded and here we analyse three of them (Table 5.1).

**Table 5.1** Collection sites of *Begonia* hybrid swarms used for this study.

<table>
<thead>
<tr>
<th>Hybrid swarm number</th>
<th>Species</th>
<th>Locality details</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Habitat description</th>
</tr>
</thead>
<tbody>
<tr>
<td>HS1</td>
<td><em>B. heracleifolia</em> x <em>B. nelumbiifolia</em></td>
<td>9km SE of San Andres Tuxtlas</td>
<td>18.520090</td>
<td>-95.161760</td>
<td>Forest clearing 20m from roadside. Overgrown grasses and some small shrubs.</td>
</tr>
<tr>
<td>HS2</td>
<td><em>B. heracleifolia</em> x <em>B. nelumbiifolia</em></td>
<td>Highway 175 between Jasaa and San Juan Bautista</td>
<td>17.743560</td>
<td>-96.328028</td>
<td>Roadside cliff. Plants growing at base of cliff, little other vegetation.</td>
</tr>
<tr>
<td>HS3</td>
<td><em>B. heracleifolia</em> x <em>B. nelumbiifolia</em></td>
<td>San Jeronimo Zoochina</td>
<td>17.221170</td>
<td>-95.235472</td>
<td>Roadside bank in woodland, covered in secondary vegetation.</td>
</tr>
<tr>
<td>HS4</td>
<td><em>B. heracleifolia</em> x <em>B. sericoneura</em></td>
<td>Motzorongo</td>
<td>18.669530</td>
<td>-96.787139</td>
<td>Roadside cliff with mixed secondary vegetation of herbs and shrubs</td>
</tr>
</tbody>
</table>

Silica-dried material was collected for genetic analysis from South Mexico (see below), and 1 - 3 rhizomes per species were collected from each hybrid swarm and from other populations (Chapter 4) for use in experimental crosses (Table 5.2). Rhizomes were potted in compost mix (Chapter 4), and grown in a common greenhouse environment (28°C) at the RBGE until flowering size.
5.3.2. Experimental estimates of reproductive barriers

5.3.2.1. Phenology

The flowering times of species were compared, to see if a limited overlap in flowering will prevent cross-pollination. Most *Begonia* species are monoecious (Tebbit, 2005), and female flowers can be identified from males by the presence of an ovary. Nine accessions of *B. heracleifolia*, 6 *B. sericoneura* and 4 *B. nelumbiifolia* were used, collected from populations across South Mexico (Table 5.2). The numbers of open male and female flowers were recorded at weekly intervals from the start of February to the end of April 2012; flowers were counted only if the tepals were still attached to the flower. The number of male and female flowers were averaged across individuals for each species, and plotted against the time interval.

5.3.2.2. Chromosome numbers

Differences in chromosome numbers can render hybrids sterile, which will largely prevent backcrossing and introgression. Previous chromosome counts from multiple accessions of *B. sericoneura* are 2n = 28 (Matthews, 2007). Chromosome counts for *B. heracleifolia* and *B. nelumbiifolia* have only been made from a single cultivated accession, where the count is also 2n = 28 (Legro & Doorenbos, 1969). As these species are widespread and genetically differentiated (Chapters 2 and 4), and because polyploidy is common in the genus (Dewitte *et al.*, 2009; Dewitte *et al.*, 2011), chromosome counts for 3 accession of *B. heracleifolia* and *B. nelumbiifolia* (Table 5.2) were made to determine whether differences in chromosome numbers could contribute to reproductive isolation between them. Roots from 6 week old cuttings were pre-treated in Gammexane (Sigma-Aldrich, Croatia) for 4 hours prior to fixation in 3:1 ethanol to glacial acetic acid (v/v). Roots were softened in enzyme solution (4% pectinase and 4% cellulase) for 20 minutes at 37°C, and squashed in a drop of acetic acid in glycerol. Slides were frozen in liquid nitrogen, and the coverslip removed with a razor blade. The samples were stained with 1µM DAPI in Mcilvaine buffer for 10 minutes in the dark, washed in 1x PBS buffer, mounted in Vectashield (Vector Laboratories, Burlingame, USA) and covered with a glass cover slip. Samples were visualised using the DAPI filter in a Zeis Axioskop microscope.
Table 5.2. Locality details for material used in experimental crosses (X), estimates of flowering time (FT), and chromosome counts (CC). All Material wild collected in Mexico except where indicated.

<table>
<thead>
<tr>
<th>Pop. code</th>
<th>Locality</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Experiment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>B. heracleifolia</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>h2</td>
<td>4km NE San Andrés Tuxtla, Veracruz</td>
<td>18.47850</td>
<td>-95.17802</td>
<td>CC, X</td>
</tr>
<tr>
<td>h8</td>
<td>Agua Azul, Chiapas</td>
<td>17.22117</td>
<td>-92.11073</td>
<td>FT, X</td>
</tr>
<tr>
<td>h13</td>
<td>Near Berriozábal, Chiapas</td>
<td>16.86693</td>
<td>-93.32781</td>
<td>CC, FT, X</td>
</tr>
<tr>
<td>h14</td>
<td>Roadside to Fortin de las Flores, Veracruz</td>
<td>18.89369</td>
<td>-97.01050</td>
<td>X</td>
</tr>
<tr>
<td>h15</td>
<td>5km south of Orizaba, Veracruz</td>
<td>18.78253</td>
<td>-97.08508</td>
<td>FT, X</td>
</tr>
<tr>
<td>h16</td>
<td>Valley 30km East of Huatusco, Veracruz</td>
<td>19.20111</td>
<td>-96.67139</td>
<td>X</td>
</tr>
<tr>
<td>h21</td>
<td>Between Jasaa and San Juan Bautista, Oaxaca</td>
<td>17.74356</td>
<td>-96.32803</td>
<td>CC, FT, X</td>
</tr>
<tr>
<td>h24</td>
<td>Motzorongo, Oaxaca</td>
<td>18.66953</td>
<td>-96.78714</td>
<td>FT, X</td>
</tr>
<tr>
<td>h28</td>
<td>Near Santa Maria Xanabi, Veracruz</td>
<td>15.98808</td>
<td>-96.11061</td>
<td>FT, X</td>
</tr>
<tr>
<td><strong>B. nelumbiifolia</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n3</td>
<td>7km NE San Andrés Tuxtla, Veracruz</td>
<td>18.50341</td>
<td>-95.16824</td>
<td>FT</td>
</tr>
<tr>
<td>n7</td>
<td>Road to Ocosingo from Palenque, Chiapas</td>
<td>17.42477</td>
<td>-91.99712</td>
<td>CC, FT</td>
</tr>
<tr>
<td>n11</td>
<td>Ocozocoautla biosphere reserve, Chiapas</td>
<td>16.92489</td>
<td>-93.45090</td>
<td>CC, FT</td>
</tr>
<tr>
<td>n19</td>
<td>3km north of Josaa, Oaxaca</td>
<td>17.64247</td>
<td>-96.33747</td>
<td>CC, X</td>
</tr>
<tr>
<td>n24</td>
<td>Motzorongo, Oaxaca</td>
<td>18.66953</td>
<td>-96.78714</td>
<td>FT</td>
</tr>
<tr>
<td>n/a</td>
<td>Cultivated accession, unknown provenance</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td><strong>B. sericoneura</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>s24</td>
<td>Motzorongo, Oaxaca</td>
<td>18.66953</td>
<td>-96.787139</td>
<td>FT, X</td>
</tr>
</tbody>
</table>
5.3.2.3. Crossing barriers and fitness of hybrids

Low hybrid fitness will reduce the likelihood of introgression, and four measures of fitness were compared between experimental F1s and their parents: (1) seed set (2) seed germination (3) early vigour (4) pollen viability.

Hybrid seed set. Five *B. heracleifolia* plants were selected at random as maternal parents and used in three different cross types: selfs, outcrosses, and F1 crosses with *B. sericoneura*. The pollen donor for outcrosses and F1 crosses were chosen at random from plants in flower. Crosses were made by rubbing the dehiscing stamens across the stigma, and each cross repeated three times. Other male flowers in the vicinity were removed to prevent accidental cross-pollination. Seed capsules were harvested when the capsule had turned brown, and stored in glassine packets at 4°C. *Begonia* species produce thousands of seed per capsule, so seed set was estimated using the approach of Ågren & Schemske (1993), summarised here briefly. Filled seeds and shrivelled seeds can distinguished under a 10x dissecting microscope; filled seeds may be viable or have aborted late in development, while shrivelled seed never germinate (Ågren & Schemske, 1993). The average mass of an aborted seed (mₐ) or a filled seed (m₇) was estimated by weighing 1000 aborted of filled seeds, pooled from each accession. For each pollinated capsule, the capsule mass (Mₗ), and the proportion filled seeds (p) from a sample of 100 seeds was measured. The number of filled seeds per capsule (N) was estimated as pN using the following formula (Ågren & Schemske, 1993):

\[
N = \frac{M_l}{[pm_f + (1-p)m_a]}
\]

The impact of cross type (self, outcross or F1) and maternal parent (both fixed effects) on seed set was estimated using general linear model analysis of variance (GLM-ANOVA) in Minitab Statistical Software (www.minitab.com). Seed set was also tested in 5 reciprocal F1 crosses (*B. sericoneura* x *B. heracleifolia*). The significance of the differences between means in reciprocal crosses was calculated using Student’s *t*-test in Minitab.

Seed set was also analysed in F1 crosses between *B. heracleifolia* and *B. nelumbiifolia*. Two accessions of *B. nelumbiifolia* flowered in 2011, and these were used as paternal parents to 5
different maternal parents. The reciprocal cross *B. nelumbifolia* x *B. heracleifolia* was also attempted. The significance of the difference between the mean of the reciprocal crosses, and each cross to selfs, were calculated with Student’s *t*-test, as above.

**Hybrid seed germination.** Seed germination was compared between the three crosstypes (selfs, outcrosses, F1 with *B. sericoneura*), to test whether filled seeds are viable or if they have aborted late in development. Twenty-five to forty filled seeds per cross generated from the seed set experiment (above) were germinated on finely-sieved bark in a propagator at 28°C, watered weekly with 1:10 dilute tomato fertilizer, and the proportion germination scored after 8 weeks. The effect of cross type and maternal parent were used in a GLM-ANOVA as above. Seed germination was also attempted in the reciprocal F1 cross to *B. sericoneura*. Four different maternal parents were selected and 11 crosses made, and the significance of the differences between means was calculated using Student’s *t*-test.

F1 *B. heracleifolia* x *B. nelumbifolia* crosses were also germinated and scored, as above. The significance of the difference in the mean F1 seed and that of selfs in this experiment, and the other crossing barriers tested below (hybrid vigour, sterility, and later-generation barriers), was tested using Student’s *t*-test.

**Hybrid vigour.** The vigour of F1s (*B. heracleifolia* x *B. nelumbifolia*, *B. heracleifolia* x *B. sericoneura*) was compared to selfs, using the seedlings from the germination experiment (above). Four eight week old self and F1 hybrids were chosen at random and potted into 9cm pots. After a total 16 weeks, 3 plants per cross type were harvested at soil level, and the number of leaves (>3mm) counted, and the mass of the plants dried at 90°C for 4 days (dry mass herein) recorded. The final plant from each cross type was grown to maturity for pollen viability measures (below).

**Hybrid sterility.** Pollen from a single flower from each of the F1 crosses (from above) was stained with 1M acetocarmine, observed under a Leica microscope, and the number of well-stained pollen recorded from 200 pollen grains. Female fertility was also scored using the wild-collected F1 hybrids that flowered in the greenhouse (*2 B. heracleifolia* x *B. sericoneura* and 3
B. heracleifolia x B. nelumbiifolia). These were used as pollen recipient to both the parental species, and the seed set recorded (as above).

**Later-generation barriers.** Hybrid break down may occur after the F1 generation (Stace, 1989) so the fitness of F1 backcrosses (F1BCs) was recorded. The vigour and pollen stainability of F1BC plants generated in the hybrid sterility experiment (above) were made in the same way as before.

### 5.3.3. Extent of gene exchange in hybrid swarms

#### 5.3.3.1. B. heracleifolia x B. sericoneura hybrid swarm

A total of 100 individuals were analysed at nuclear loci in the hybrid swarm. These were categorised on the basis of morphology as: 42 B. heracleifolia, 28 B. sericoneura, and 30 putative hybrids. Sixty-three of these genotypes were generated for this study (all B. sericoneura, putative hybrids, and 5 B. heracleifolia), and added to the 37 B. heracleifolia genotypes generated in Chapter 4.

A total of 93 individuals were analysed at plastid loci: 35 B. heracleifolia, 28 B. sericoneura, and 30 putative hybrids. Fifty-eight of these genotypes were scored solely for this study (all B. sericoneura and the putative hybrids) and added to the 35 B. heracleifolia genotypes scored in Chapter 2.

#### 5.3.3.2. B. heracleifolia x B. nelumbiifolia hybrid swarms

Morphometric analysis of leaf shape was conducted on hybrid swarm HS1, to test whether hybrid categories from molecular data are consistent with morphology. A large leaf was harvested from each of 57 plants (25 B. heracleifolia, 29 B. nelumbiifolia, 3 putative hybrids) and photographed against a black background with a scale, prior to analysing leaf blade dimensions and leaf shape in LAMINA (Bylesjö et al., 2008). Box plots were made for each measure of leaf shape, comparing individuals assigned to pure parental classes and hybrid classes in the NewHybrids analysis (below).
The number of individuals analysed at nuclear loci was as follows: HS1 total 61 plants, 29 *B. heracleifolia*, 29 *B. nelumbiifolia*, 3 putative hybrids; HS2 total 71 plants, 30 *B. heracleifolia*, 40 *B. nelumbiifolia*, 1 putative hybrid; HS3 total 61 plants, 29 *B. heracleifolia*, 30 *B. nelumbiifolia* and 2 putative hybrids. All hybrids that were found in HS2 and HS3 were genotyped, but only 3 hybrids were collected in HS1 because access to the site was limited. Fifteen of the genotypes were newly generated for this study (6 putative hybrids, 3 *B. heracleifolia* HS1, 1 *B. nelumbiifolia* HS1) and added to the 179 generated in Chapter 4.

A total of 111 plants were analysed for plastid loci from two hybrid swarms: HS1, 33 *B. heracleifolia*, 29 *B. nelumbiifolia*, 3 putative hybrids; HS2, 20 *B. heracleifolia*, 25 *B. nelumbiifolia*, 1 putative hybrid), as well as the two hybrids from HS2. Three of the genotypes were newly generated for this study (6 putative hybrids) and added to the 105 generated in Chapter 2.

### 5.3.3.3. Nuclear genotyping

Plants were genotyped with the 14 nuclear microsatellite loci described in Chapter 3, and scored in the same manner.

### 5.3.3.4. Hybrid swarm analyses for nuclear markers

Three complementary analytical approaches were applied to the nuclear microsatellite data to assess the genetic composition of hybrid swarms. The program NewHybrids 1.1 Beta (Anderson & Thompson, 2002) assigns individuals to one of six genealogical classes of hybrids after 2 generations of crossing: parent A-type, backcross A-type, F1-type, F2-type, backcross B-type, parent B-type. Milne & Abbott (2008) extended the number of hybrid classes to the 45 unique combinations after 4 generations of crosses, and grouped these into the 6 categories described above; individuals not equivocally assigned to a single category are placed in compound categories (e.g. ‘backcross to parent A or parent A’) using a MS Excel spreadsheet (made available by R. Milne, University of Edinburgh, pers. comm.). For the *B. heracleifolia* x *B. nelumbiifolia* hybrid swarms 12 individuals of each species were used as a reference in the
analysis and these were taken from across the species ranges. The analysis for HS4 was run without reference populations, as no suitable populations were found in the field. This has little effect on the accuracy of hybrid assignment (Vähä & Primmer, 2006). Analyses of \textit{B. heracleifolia} x \textit{B. nelumbiifolia} hybrid swarms were repeated without reference populations for comparison. For each hybrid swarm, 2 and 4 generations of crosses were evaluated, using 100 000 sweeps after a burn-in period of 10 000 sweeps. In NewHybrids, individuals are assigned to one or more genealogical classes with a posterior probability score (\textit{q}-value), and to reduce miss-assignment a stringency of \textit{q} > 0.9 was used as recommended by Vähä & Primmer, 2006).

Second, the program FLOCK (Duchesne & Turgeon, 2009) was used, which allocates individuals to one of the user defined numbers of genetic clusters (\textit{K}) with a log-likelihood score (LLOD score). FLOCK is a non-Bayesian approach which operates well even when admixture is high, and when reference populations are not available (Duchesne & Turgeon, 2009). Each hybrid swarm was analysed without reference populations using default parameters and with \textit{K} = 2. A plot of LLOD scores was evaluated to see if there were distinct clusters of parents and putative hybrid classes, or whether there is a continuum of LLOD scores suggesting complete admixture.

Third, Bayesian clustering in BAPS (Corander \textit{et al.}, 2008) was used to measure the contribution of each parental genome to the hybrids. BAPS reports this as a \textit{q}-value. To avoid confusion between NewHybrids and BAPS \textit{q}-values, these are referred to as \textit{q}_{\text{NHZ}} and \textit{q}_{\text{BAPS}} respectively, herein. Preliminary runs with ‘clustering of individuals’ followed by ‘admixture based on mixture clustering’, where no \textit{a priori} information is given about pure individuals, performed poorly (results not shown). Therefore, pure individuals were first detected in NewHybrids (\textit{q}_{\text{NHZ}} > 0.9), and then defined in the input file for ‘admixture based on pre-defined populations’, as recommended in the BAPS manual. Default options were selected, except that 10 000 iterations and 5 000 reference individuals were used.

\textbf{5.3.3.5. Simulated hybrid swarms}

The discriminatory power of the markers and the error rate of assigning individuals to the correct hybrid class were tested using simulated hybrid swarms. Hybrids were simulated from the allele
frequencies of pure individuals in hybrid swarms \( (q_{NHZ} > 0.9) \). This approach allows simulated and empirical datasets to be compared directly, but assumes the parents have not been introgressed by other species (Burgarella et al., 2009). This assumption is supported by their distinct morphologies and their strong differentiation at nuclear and plastid loci (see results). However, if introgression had occurred, this would simply reduce the power of detecting hybrids (Burgarella et al., 2009).

Pure parents from two hybrid swarms (HS1, 29 B. heracleifolia and 28 B. nelumbiifolia; HS4, 39 B. heracleifolia and 26 B. sericoneura) were used as parents for the two simulated hybrid swarms. Random mating between the parents was simulated by drawing alleles at random from the observed parental allele frequencies using HybridLab v1.0 (Nielsen et al., 2006). A total of 90 hybrids (30 F1s and 30 BC to each parent) were simulated for each hybrid swarm, and these were analysed in NewHybrids (as before). The likelihood of assignment to a genetic cluster \( (q_{NHZ}) \) was averaged across each hybrid class and standard errors calculated, and the percentage of individuals assigned to the wrong category recorded.

### 5.3.3.6. Inbreeding

The level of inbreeding can affect the genetic composition of hybrid swarms, as pollen is most likely to be carried from the outcrossing to the selfing species (Ruhsam et al., 2010). The inbreeding coefficient and inferred selfing rates for B. heracleifolia and B. nelumbiifolia were calculated in Chapter 4. For comparison, the \( F_{IS} \) value for B. sericoneura was calculated using individuals that were assigned as parental B. sericoneura in NewHybrids \( (q_{NHZ} > 0.9) \). The same subset of 9 loci were used as Chapter 4, and the \( F_{IS} \) value calculated in FSTAT v1.2 (Goudet, 1995). The inferred selfing rate was calculated with the formula of Allard et al. (1969, see Chapter 4).

### 5.3.3.7. Plastid analyses

Maternal plastid inheritance has been found in cytological observations (Corriveau & Coleman, 1988) and plastid sequencing of experimental crosses (Peng & Chiang, 2000) in other Begonia species. Therefore plastid genotyping was used to detect the direction of hybridization. Plants
were genotyped with 7 plastid microsatellites using the same amplification protocol and method of scoring as Chapter 2. Plastid haplotypes were defined as the unique combination of alleles at all plastid loci. The frequency of plastid haplotypes were compared between the putative hybrids and the parents for each hybrid swarm.

5.4. Results

5.4.1. Barriers to hybridization between *B. heracleifolia* and *B. sericoneura*

5.4.1.1. Experimental assessment of reproductive barriers

**Phenology.** A similar number of flowers were produced at peak flower time for *B. heracleifolia* (40.1, n = 9) and *B. sericoneura* (39.5, n = 6). *Begonia sericoneura* started flowering earlier than *B. heracleifolia*, although peak flowering overlapped (Fig. 5.2).

5.4.1.2. Crossing barriers and fitness of hybrids

**Hybrid seed set.** The number of seeds set when wild-collected *B. heracleifolia* was a pollen recipient to *B. sericoneura* (1080 ± 152; n = 15) was not significantly different to outcrosses (1055 ± 78; n = 15) or selfs (1265 ± 156; n = 15). The seed set between reciprocal F1 crosses were also not significantly different (*B. sericoneura* x *B. heracleifolia*, 1473 ± 418, n = 5). However, the effect of crosstype (self, outcross, F1 to *B. sericoneura*) was marginally significant on seed set in the ANOVA ($P = 0.045$, Table 5.3), with the maternal parent and the interaction effect being more significant ($P < 0.001$; Table 5.3). The interaction plot showed no common pattern between cross type and seed set; some F1 hybrids produced more and others fewer seeds than selfs.
Figure 5.2. Variation in flower time for the three Begonia species. a) B. sericoneura, b) B. heracleifolia, c) B. nelumbiifolia. Number of flowers shown separately for males (solid line) and females (dashed line).

Hybrid seed germination. The germination of B. heracleifolia x B. sericoneura F1 plump seeds (70.0% ± 6.7; n=15) was not significantly different from selfs (64.1% ± 8.7; n=15) or outcrosses (79.5% ± 7.5; n=15). Only the effect of maternal parent was significant on seed germination in the ANOVA (Table 5.3).
Table 5.3. Effect of maternal parent and cross type (self, outcross, or F1) on seed set and germination analyzed with GLM-ANOVA. DF, degrees of freedom; *, P < 0.05; **, P < 0.01; ***, P < 0.001

<table>
<thead>
<tr>
<th>Fitness measure</th>
<th>Source</th>
<th>DF</th>
<th>Mean square</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seed set</td>
<td>Maternal parent</td>
<td>4</td>
<td>1645540</td>
<td>28.58***</td>
</tr>
<tr>
<td></td>
<td>Cross type</td>
<td>2</td>
<td>198144</td>
<td>3.44*</td>
</tr>
<tr>
<td></td>
<td>Maternal parent x cross type</td>
<td>8</td>
<td>377129</td>
<td>6.55***</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>30</td>
<td>57585</td>
<td></td>
</tr>
<tr>
<td>Germination</td>
<td>Maternal parent</td>
<td>4</td>
<td>3785.3</td>
<td>8.25***</td>
</tr>
<tr>
<td></td>
<td>Cross type</td>
<td>2</td>
<td>908.3</td>
<td>1.98</td>
</tr>
<tr>
<td></td>
<td>Maternal parent x cross type</td>
<td>8</td>
<td>10209.5</td>
<td>2.24</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>30</td>
<td>458.7</td>
<td></td>
</tr>
</tbody>
</table>

**Hybrid vigour.** The vigour of *B. heracleifolia* x *B. sericoneura* F1 plants (number of leaves = 5.8 ± 0.2; dry mass = 0.33g ± 0.075; n = 9) was not significantly different to selfed plants (number of leaves = 6.1 ± 0.3; dry mass = 0.42g ± 0.08; n = 18).

**Hybrid sterility.** The 2 wild-collected F1 *B. heracleifolia* x *B. sericoneura* hybrids had low pollen fertility, one plant dropped male flowers, and the other had little stainable pollen (4.8%); none of the experimental F1s flowered in their first year (n = 7). The number of filled seed was not significantly different when the F1 was used as pollen recipients to *B. heracleifolia* (40 ± 10 seeds, n = 8) or *B. sericoneura* (12 ± 3 seeds, n = 6).

**Later-generation barriers.** Germination of seeds in the F1BC to *B. sericoneura* was significantly higher (33.3% ± 5, n = 9) than F1BC to *B. heracleifolia* (11.5 ± 3, n=6). The F1BC seedlings with *B. sericoneura* as a paternal parent (leaves = 6.7 ± 0.5; dry mass = 0.41 ± 0.11g; n = 6) were not significantly more vigorous than when *B. heracleifolia* was used (leaves = 4.8 ± 1.1 leaves; dry mass = 0.18 g ± 0.15, n = 2).
5.4.1.3. Analysis of hybridization in nature

The 14 nuclear microsatellite loci amplified successfully in both species and their putative hybrids (Appendices 5 – 8). Loci BI5347 and BI3348 were excluded due to uneven amplification or unexpected repeat motif lengths. Analyses using 10 loci common to both species pairs gave similar results to the full dataset of 12 loci (results not shown). All loci were polymorphic, except BE332 which was monomorphic in HS4.

Breeding system. Four loci were polymorphic in pure *B. sericoneura* individuals. The F<sub>IS</sub> value across loci was 0.415, which gives an inferred selfing rate of 0.59 (Table 5.4).

Table 5.4. Inbreeding coefficients and inferred selfing rates for the three study species.

<table>
<thead>
<tr>
<th>Species</th>
<th>F&lt;sub&gt;IS&lt;/sub&gt;</th>
<th>Inferred selfing rate</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>B. heracleifolia</em></td>
<td>0.249</td>
<td>0.399</td>
<td>Chapter 4</td>
</tr>
<tr>
<td><em>B. nelumbiifolia</em></td>
<td>0.454</td>
<td>0.624</td>
<td>Chapter 4</td>
</tr>
<tr>
<td><em>B. sericoneura</em></td>
<td>0.415</td>
<td>0.590</td>
<td>This study</td>
</tr>
</tbody>
</table>

Hybrid swarm structure. The NewHybrids analysis using 6 hybrid catagories assigned 27 individuals as pure *B. sericoneura* (mean q<sub>NHZ</sub> = 0.999), 41 pure *B. heracleifolia* (mean q<sub>NHZ</sub> = 0.999), 25 F1s (mean q<sub>NHZ</sub> = 0.980) and 2 BCs to *B. sericoneura* (q<sub>NHZ</sub> = 0.935; Fig. 5.3). Five individuals were not placed in a single category using a stringency of q<sub>NHZ</sub> > 0.9. When stringency was lowered, these were classified as: 3 F1s (q<sub>NHZ</sub> = 0.75), 1 BC to *B. sericoneura* (q<sub>NHZ</sub> = 0.76) and 1 pure *B. sericoneura* (q<sub>NHZ</sub> = 0.55). There was no evidence for any backcrosses to *B. heracleifolia* or F2s. When 4 generations of crosses were considered (Milne & Abbott, 2008), 39 individuals were assigned as *B. heracleifolia*, 29 *B. sericoneura*, 30 F1 or F2s, and 1 BC to *B. heracleifolia* or pure *B. heracleifolia*.

The FLOCK analysis generally supported the NewHybrids results. Three main clusters corresponded to *B. heracleifolia* (mean LLOD 11.28 ± 0.36SE, n = 41), F1s (mean LLOD -11.90 ± 0.44SE, n = 28) and *B. sericoneura* (mean LLOD -26.73 ± 0.22SE, n = 28). Two individuals
placed between the F1s and *B. sericoneura* are likely to be BC (mean LLOD $-21.59 \pm 0.52SE$). The BAPS analysis classified 28 individuals as pure *B. sericoneura* ($P = 1.0$), 41 as pure *B. heracleifolia* ($P = 1.0$) and 31 individuals as admixed (average $q_{\text{BAPS}}$ value across admixed individuals $= 0.40 \pm 0.01$). Five of these admixed individuals (average $q_{\text{BAPS}} = 0.279$, range $0.19$ - $0.33$) are more consistent with being backcrosses than F1s.

**Figure 5.3. Genetic composition and classification of hybrids in hybrid swarm 4 (*B. heracleifolia* x *B. sericoneura*). NewHybrids analysis with 6 hybrid categories (top row): orange, *B. heracleifolia*; purple, *B. sericoneura*, green, F1 hybrid; red, F2; yellow, BC *B. heracleifolia*; dark blue, BC *B. sericoneura*. BAPS (centre row): orange, *B. heracleifolia*, purple, *B. sericoneura*. FLOCK (bottom row) coloured by most likely class in NewHybrids: orange circles, *B. heracleifolia*; green triangles, F1 hybrids; grey crossed squares, putative backcrosses to *B. sericoneura*, purple squares, *B. sericoneura*.
5.4.1.4. Accuracy of assignment in simulated hybrid swarms

The classes that individuals in simulated hybrid swarms were assigned to were compared to their actual genotypic class, to test the accuracy of assignment with the markers. When 6 hybrid classes were considered, 143 plants (91.7%) were correctly assigned, and the other 13 (8.3%) were not assigned to a single category. Assignment was also high when 45 hybrid classes were considered: 140 plants (89.7%) were correctly assigned, and the others were assigned to compound categories (9 plants, 5.8%), remained unassigned (4, 2.6%), or were assigned to the wrong category (3, 1.9%).

5.4.1.5. Plastid genotyping and direction of hybridization

The plastid microsatellites revealed 3 haplotypes in *B. heracleifolia* and 1 in *B. sericconeura* in HS4. No haplotypes were shared between species. The 30 putative hybrids all had the plastid haplotype of *B. sericconeura*.

5.4.2. Barriers to hybridization between *B. heracleifolia* x *B. nelumbiifolia*

5.4.2.1. Experimental assessment of reproductive barriers

**Phenology.** *Begonia nelumbiifolia* produced many more flowers at peak flowering (>150, n = 4) than *B. heracleifolia* (40.1, n = 9; Fig. 5.2). Flowering times for *B. heracleifolia* and *B. nelumbiifolia* overlapped, but peak flowering for *B. nelumbiifolia* was at least 5 weeks later.

**Chromosome numbers.** A diploid chromosome number of 2n=28 was found in both *B. heracleifolia* and *B. nelumbiifolia* from each of 3 accessions (Fig. 5.4).
5.4.2.2. Crossing barriers and fitness of hybrids

Hybrid seed set. The number of seed set in *B. heracleifolia x B. nelumiifolia* F1 crosses (503 ± 86, n = 11), was significantly reduced relative to selfed *B. heracleifolia* (1265 ± 156; n = 15, P < 0.05). The reciprocal cross (*B. nelumiifolia x B. heracleifolia*) also set seed (394 ± 230, n = 3), with the mean not significantly different from the self.

Hybrid germination. Germination success of *B. heracleifolia x B. nelumiifolia* F1 hybrids was 45.2% ± 9% (n = 11), which was not significantly different from selfed *B. heracleifolia* (64.1% ± 9%, n = 15).

Hybrid vigour. The vigour of *B. heracleifolia x B. nelumiifolia* F1s (number of leaves = 5.3 ± 0.6; dry mass = 0.37g ± 0.09; n = 4) was not significantly different from selfs (number of leaves = 6.1 ± 0.3; dry mass = 0.42g ± 0.08; n = 18).

Hybrid sterility. The three wild-collected F1s were largely infertile, producing no viable pollen nor setting any viable seed when selfed. Experimental F1s were also male sterile (4/6 dropped male flowers, 2/6 had 0% stainable pollen). However, wild collected F1s did set a low number of

Figure 5.4. Chromosome spreads of *B. heracleifolia* (left) and *B. nelumiifolia* (right). Scale bar is 10µM.
seeds as a pollen recipient to either of the parents, which did not differ significantly when *B. heracleifolia* (43.5 ± 11.2, n = 9) or *B. nelumbiifolia* (57.4 ± 15.2, n = 14) were the pollen donor.

**Later-generation barriers.** The germination of F1BC seed to *B. heracleifolia* (41.1% ± 6.5, n = 8) was not significantly different from the BC to *B. nelumbiifolia* (47.6% ± 6.1, n=10). No difference was found in the vigour between the backcross plants to *B. heracleifolia* (number of leaves = 6.2 ± 0.8, dry mass = 0.22 ± 0.05g, n = 5) and *B. nelumbiifolia* (number of leaves = 6.5 ± 0.2, dry mass = 0.33 ± 0.11g, n = 4).

### 5.4.2.3. Analysis of hybridization in nature

**Morphometrics.** *Begonia heracleifolia* and *B. nelumbiifolia* were clearly distinguished by morphometric analyses of leaf shape, with the 3 hybrids placed as intermediate (e.g. Fig. 5.5 for leaf circularity index, the percentage of pixels shared between the leaf and a superimposed circle with same width/height as the leaf).

**Hybrid swarm analyses.** The 14 nuclear loci amplified successfully in both species and their hybrids, two loci were excluded (BI6534 and BC332) due to uneven amplification or unexpected repeat motif lengths.

All three analytical methods (BAPS, NewHybrids and FLOCK) identified few hybrids in each hybrid swarm (total of 6: 3 in HS1, 2 in HS2, and 1 in HS3) that are likely to be F1 hybrids, with no other hybrid classes present (Fig. 5.6). NewHybrids analyses strongly supported the 6 hybrid individuals as F1s (mean $q_{NHZ}$ with reference populations = 0.986; without = 0.953). The genotypic classes accorded exactly to their morphology (Fig. 5.5). The mean contribution of each parents’ genome to the 6 hybrids (estimated in BAPS) was 0.48:0.52 (*B. heracleifolia*: *B. nelumbiifolia*, not significantly different from 50:50), and in each case FLOCK identified clusters of *B. heracleifolia* (mean LLOD = -27.06 ± 0.52SE, n = 88) and *B. nelumbiifolia* (mean LLOD = 27.98 ± 0.48SE, n = 99), with putative hybrids intermediate (mean LLOD 3.91 ± 0.73SE, n = 6).
When 4 generations of crosses were considered in NewHybrids (Milne & Abbott, 2008): 84 individuals were considered to be *B. heracleifolia*, 98 *B. nelumbiifolia*, 3 F1 or F2s, 4 BC to *B. heracleifolia* or parental *B. heracleifolia*, 1 BC *B. nelumbiifolia* or parental *B. nelumbiifolia*, and 3 unclassified (any hybrid class). The classification of some hybrids as potential backcrosses was not supported by any of the other analyses.

**Figure 5.5.** Morphometric analysis of leaf shape in hybrid swarm 1 (*B. heracleifolia* x *B. nelumbiifolia*).

**Figure 5.6.** Genetic composition and classification of hybrids in three *B. heracleifolia* x *B. nelumbiifolia* hybrid swarms (next page). NewHybrids analysis (top row): orange, *B. heracleifolia*; blue, *B. nelumbiifolia*; green, F1 hybrid; red, F2; yellow, BC *B. heracleifolia*; dark blue, BC *B. nelumbiifolia*. BAPS (centre row): orange, *B. heracleifolia*; blue, *B. nelumbiifolia*. FLOCK (bottom row) coloured by most likely class in NewHybrids: orange circles, *B. heracleifolia*; blue squares, *B. nelumbiifolia*; green triangles, F1 hybrids.
5.4.2.4. Accuracy of assignment in simulated hybrid swarms

Most individuals were correctly assigned when 2 generations of crossing were considered without reference populations, although some individuals were unassigned (Table 5.5). Increasing the number of hybrid classes to the 45 unique categories after 4 generations of crosses (Milne and Abbott, 2008) marginally increased the assignment error, and many individuals were assigned to pooled categories (Table 5.5).

Table 5.5. NewHybrids assignment for the *B. heracleifolia* x *B. nelumbiifolia* hybrid swarm with different settings. † MCMC analysis consistently ended prematurely.

<table>
<thead>
<tr>
<th>Number of hybrid categories used</th>
<th>Reference populations</th>
<th>Assigned to correct category</th>
<th>Assigned to compound category</th>
<th>Unassigned</th>
<th>Incorrect category</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>Yes</td>
<td>141 (95.3%)</td>
<td>n/a</td>
<td>6 (4.1%)</td>
<td>1 (0.75)</td>
</tr>
<tr>
<td>6</td>
<td>No</td>
<td>143 (96.6%)</td>
<td>n/a</td>
<td>5 (3.4%)</td>
<td>0</td>
</tr>
<tr>
<td>45</td>
<td>Yes</td>
<td>104 (70.3%)</td>
<td>34 (23.0%)</td>
<td>6 (4.1%)</td>
<td>4 (2.7%)</td>
</tr>
<tr>
<td>45†</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5.4.2.5. Plastid genotyping and direction of hybridization

In the three *B. heracleifolia* x *B. nelumbiifolia* hybrid swarms, a single plastid type was found in *B. nelumbiifolia*, and a total of 9 in *B. heracleifolia*. No haplotypes were shared between species. All 6 putative hybrids had the plastid type of *B. nelumbiifolia*. 
5.5. Discussion

5.5.1. Evolution of reproductive isolation in *Begonia*

The hypothesis that relatively strong reproductive barriers accumulate between related *Begonia* species is supported by analyses of two pairs of hybridizing *Begonia* species. In contact zones between both species pairs heterospecific mating will be uncommon due to frequent self-fertilization (inferred selfing rates 40 – 62%) and differences in flowering times (Fig. 5.6), and introgression is unlikely due to low fertility of F1 hybrids (0 – 5% stainable pollen; discussed below). These barriers translated to structured hybrid swarms in the field, composed of parents and mostly early generation hybrids. Other studies of *Begonia* hybrids are mostly consistent with the hypothesis of rapidly evolving reproductive barriers. Wide-outcresses between *B. heracleifolia* populations have a modest reduction in fertility (Chapter 4). Hybridizing *Begonia* species in Taiwan have strong crossing barriers, with sterile F1s and no later generation hybrids, with these reproductive barriers likely to be explained by differences in chromosome number (e.g. *B. x buimontana*, Peng & Chen, 1991; *B. x breviscarpa*, Peng et al., 2010; *B. x chungii*, Peng & Shin-Ming, 2009; *B. x taipeiensis*, Peng & Sue, 1991). Whether hybrids between *B. decora* and *B. venusta*, closely related species from section Platycentrum (Thomas et al., 2011b), fit this pattern is unclear. Hybridization leads to a wide range of hybrid morphologies (Teo & Kiew, 1999) and admixed AFLP profiles (Kiew et al., 2003), but the classes of hybrids have not been estimated from the genetic data.

The strength of reproductive isolation between *Begonia* species contrasts with many other plant hybrid zones between diploid species, where widespread introgression is common (e.g. *Aquilegia*, Hodges & Arnold, 1994; *Helianthus*, Scascitelli et al., 2010; Louisianan *Iris*, Arnold et al., 2010; *Populus*, Lexer et al., 2005; *Silene*, Minder et al., 2007). The scarcity of studies that report limited introgression likely represents publication bias, however structured hybrid swarms, like *Begonia*, give a unique insight into the evolution of reproductive isolation. Similar hybrid swarm structure, where F1s are present but backcrosses rare or absent, have been found between *Rhododendron* species and between some *Quercus* species. In two pairs of hybridizing *Rhododendron* species, F1 hybrids are fertile and the absence of BCs may be caused by extrinsic
habitat selection (Milne et al., 2003; Milne & Abbott, 2008). Hybrids between Quercus kelloggii and Q. wislizenii var. frutescens are also thought to be fertile, and the absence of BCs has been attributed to competition with the parents (Nason et al., 1992).

We tested whether limited introgression in contact zones is caused by ecological selection against hybrids, by comparing reproductive barriers under experimental conditions with the frequency of hybrids in the field between an ecologically similar (both dry-adapted) and more divergent (dry-adapted and moisture-loving) species pair. The frequency and extent of hybridization between ecologically similar B. heracleifolia and B. sericoneura was much greater than between ecologically divergent B. heracleifolia and B. nelumbiifolia. This was the case even though hybrids between both species pairs were relatively vigorous under experimental conditions, although this must be interpreted with caution due to the small sample sizes and the lack of field testing. Strong support for assortative mating and a bimodal distribution of hybrid indices (i.e. hybrid swarm dominated by parental individuals), has been considered evidence for the role of ecology rather than genomic incompatibilities in hybridization (Jiggins & Mallet, 2000). However, we could not estimate the role of selection against hybrids independently of intrinsic reproductive barriers, as intrinsic reproductive barriers were also weaker between B. heracleifolia x B. sericoneura than B. heracleifolia x B. nelumbiifolia. Specifically, B. heracleifolia and B. sericoneura overlap in their flowering time for longer than B. heracleifolia and B. nelumbiifolia (Fig. 5.2), produce more seed in F1 crosses, and some F1 plants produce viable pollen. The stronger reproductive barriers and reduced frequency of hybrids between the ecologically more divergent species may be explained by negative interactions between divergent adaptive loci (Presgraves et al., 2003), or through hybrids having a maladaptive mix of traits ‘blended’ between the parents (Burke & Arnold, 2001), both of which would affect the more ecologically divergent species pair to a greater extent. Alternatively, adaptive loci to particular environments, or prezygotic barriers (flowering time), may have a pleiotropic effect on intrinsic reproductive barriers (e.g. F1 hybrid sterility). One example would be if an adaptive trait is located within an inversion polymorphism that reduces homologous chromosome pairing in meiosis, causing sterility (Hoffmann & Rieseberg, 2008).
A non-mutually exclusive explanation would be that strong reproductive isolation in crosses with *B. nelumbifolia* reflects its unique demographic history, rather than being related to ecological selection. This species frequently self-fertilizes, and may have also been through a population bottleneck (Chapter 2). Both these factors will fix alleles through genetic drift, which may then be incompatible with other *Begonia* genomes when brought together through hybridization. Deleterious alleles fixed through drift have been proposed to explain the low fertility of interspecific crosses between the highly selfing *Glycine falcata* and other *Glycine* species, which does not follow the trend between genetic distance and reproductive isolation found with other *Glycine* crosses (Moyle et al., 2004). It may also be that nuclear-cytoplasmic incompatibilities in interspecific crosses are caused by the fixation of a single plastid haplotype in south Mexican populations of *B. nelumbifolia*. Maternal effects have also been proposed to explain why crosses between another pair of *Begonia* species, *B. formosana* Hayata (Masamune) and *B. aptera* Blume, can only be crossed in one direction (Peng & Chiang, 2000). Future studies to distinguish whether ecological selection or demographic history are the main drivers of strong reproductive isolation in *Begonia* may look to other hybridizing species, or compare patterns of sequence divergence at neutral and putatively adaptively loci underlying ecologically relevant traits.

### 5.5.2. Experimental limitations of hybrid swarm comparisons

An important confounding factor that may influence the results of our study is the phylogenetic relatedness of the species pairs (Funk *et al.*, 2006). Reproductive isolation will usually increase with genetic divergence (Gleason & Ritchie, 1998), as increasingly divergent species are likely to differ for traits that cause prezygotic isolation (e.g. flower colour), and have accumulated genetic incompatibilities and chromosomal differences that cause post-zygotic barriers. This has been shown in fungi (Le Gac *et al.*, 2007) and animals, including: salamanders (Tilley *et al.*, 1990), fruit flies (Gleason & Ritchie, 1998), frogs (Sasa *et al.*, 1998), toads (Malone & Fontenot, 2008), fish (Mendelson, 2003), butterflies and moths (Presgraves, 2002) and birds (Price & Bouvier, 2002). In plants, postmating reproductive isolation correlates with genetic distance in *Silene* (Caryophyllales), *Glycine* (Fabaceae), food-deceptive Mediterranean orchids, but not *Streptanthus* (Brassicaceae) or sexually-deceptive orchids, likely due to insufficient time since
divergence for reproductive barriers to evolve (Moyle et al., 2004; Scopece et al., 2007). Our
current understanding of evolutionary relationships in Central American Begonia is poor
(Dewitte et al., 2011), but preliminary phylogenetic analysis with nuclear and plastid sequence
data place both species pairs as non-sister species that are similar in their genetic distances
(patristic distance: B. heracleifolia x B. sericoneura, 0.0025; B. heracleifolia x B. nelumbiifolia,
0.0016; Harrison, University of Edinburgh, pers comm.). Therefore phylogenetic distance does
not seem to explain the observed differences in hybrid swarm structures. The second concern
with a comparative hybrid swarm approach is that reproductive barriers may not have driven
divergence, but instead accumulated after speciation (Scopece et al., 2010). It can however be
argued that all reproductive isolating barriers are important to the isolation of species upon
secondary contact, even if they have not directly promoted divergence. In Begonia, the
interspecific barriers described in this study have also been found between divergent populations
(e.g. pollen sterility, chapter 4; flower time differences, A.D. Twyford, unpubl. data), therefore it
seems likely that these barriers evolved during divergence, and may have promoted it.

5.5.3. Reproductive barriers in Begonia

The main barriers detected in this study were differences in flowering time, hybrid sterility, and
inbreeding, and the pattern of mating observed in hybrid swarms fitted predictions from the
species population biology and the barriers detected under experimental conditions. For
example, all natural hybrids had the plastid type of the parent that is more highly selfing (B.
nelumbiifolia and B. sericoneura) supporting the more highly outcrossing B. heracleifolia as the
paternal parent as would be expected due to its pollen being more competitive (Ruhsam et al.,
2010). Many other reproductive isolating barriers likely play a role in maintaining species
barriers in Begonia. This study focused on crossing barriers in sympatry, and strong prezygotic
barriers will prevent contact in the first place (Barton & Hewitt, 1985; Widmer et al., 2009). The
narrow ecological niches and limited distribution ranges of Begonia species, such as species in
Begonia section Gireoudia that are adapted to such habitats as dark moist tropical forest (e.g. B.
thiemei C.DC. ex Donn.Sm. and B. sousae Burt-Utley) and sun baked cliffs (B. peltata Otto &
A. Dietr. and B. hydrocotylifolia Otto ex Hook.), will play a critical role in whether species co-
occur. Other post-pollination barriers that were not tested in this study include pollen-stigma
interactions, which have been described in *Iris* (Carney *et al.*, 1994) and *Senecio* (Allen *et al.*, 2011).

Future studies should characterize the genetic basis of the reproductive isolating barriers that we detected. Many loci of small effect would be predicted to influence these barriers if they have evolved incrementally over time, such as those evolving during allopatric speciation (Coyne & Orr, 1989). However, this does not preclude species differences evolving through mutations at loci of large effect after initial divergence. It will be of particular interest to characterize F1 hybrid sterility, as it is common between many pairs of *Begonia* species (Dewitte *et al.*, 2011; Twyford and Kidner, unpubl. data). Strong hybrid sterility in F1 crosses can be caused by: genome-wide divergence (e.g. between *Drosophila* species, Tao & Hartl, 2003), a few sterility loci of large effect (e.g. *Mimulus*, Fishman & Willis, 2007), or chromosome rearrangements such as inversions or deletions (e.g. *Helianthus*, Lai *et al.*, 2005). This study, and others in *Begonia* (Chapter 4), support hybrid sterility evolving more rapidly than hybrid inviability (Malone & Fontenot, 2008). Genetic characterization of hybrid sterility can be achieved through the analysis of populations that segregate for fertility (Chapter 6). This approach would also be useful for other reproductive traits that differ between species, such as flowering time.

### 5.5.4. Evolutionary outcomes of hybridization in *Begonia*

Predicting the long-term evolutionary outcomes of hybridization from young hybrid swarms is difficult. The *Begonia* hybrids swarms were all by roadsides where the surrounding vegetation was well established, suggesting there has been plenty of time for species with a short generation time (1 year; A.D. Twyford pers. obs.) to hybridize. Strong fertility barriers and rare F1 formation in groups such as *Iris* and *Helianthus* has not prohibited introgression and hybrid speciation over long time periods (Arnold, 1993; Cruzan & Arnold, 1993; Ungerer *et al.*, 1998; Yatabe *et al.*, 2007). In irises however, backcrossing may be promoted by preferential pollinator visitation (Emms & Arnold, 2000) and hybrid superiority (Cruzan & Arnold, 1993), which have not been found in *Begonia*. Moreover, dispersal distances of *Begonia* seed are low (Chapter 2), and dispersal away from parental progenitors is important for the establishment of homoploid hybrid species (Gross & Rieseberg, 2005). Whether past hybridization events have occurred in
Begonia could be addressed by comparative phylogenetic analysis with unlinked markers, or through the analysis of introgressed genes in areas of sympatry using densely mapped markers (Twyford & Ennos, 2012).

In conclusion, relatively strong reproductive isolating barriers were found between two pairs of Begonia species, which in conjunction with limited gene flow from neighbouring populations (Chapter 2, Chapter 4, Matolweni et al., 2000; Hughes & Hollingsworth, 2008), makes homogenizing gene flow unlikely to disrupt the speciation process. The wide range of habitats to which species are adapted further contributes to isolation, through reducing the number of species that co-occur in sympatry, as well as the putative role disruptive selection at adaptive loci plays on hybrid fitness. Taken together, the population biology and evolution of reproductive isolation may allow rapid divergence in the early stages of speciation, and may partly explain the high species richness in Begonia. Whether similar evolutionary processes have played a role in other large genera has yet to be tested.
CHAPTER 6: Quantitative trait locus (QTL) analysis of sex ratios and inflorescence architectures in backcross Begonia populations

Data contributions: Keith Gardner (University of Edinburgh) made the measures of flower number, flower time, and inflorescence branch patterns. Chris White (University of Edinburgh) measured stamen number. Alexander Twyford collected measures of pollen and seed viability, and performed all analyses.

6.1. Chapter summary

Inflorescence architecture describes the branching pattern and position of flowers on an inflorescence as determined by the developmental patterns of the inflorescence meristem. These decisions concern the production of branches via axillary meristems, and in monoecious plants, the sex of floral meristems. Variation will affect the sex ratio, the investment in flowers and support structures, and the presentation to pollinators, as well as flower number. Here, we use quantitative trait locus (QTL) analyses to studying the genetics of sex ratios and branching patterns, and other reproductive traits, in backcross populations between Central American B. plebeja and B. conchifolia. Five QTL of moderate effect (13.4 - 18.7% variance explained, VE) are found for variation in sex ratio and related traits, and these map to 2 different linkage groups. A QTL for number of male flowers co-localizes with a QTL for meristem decisions for either axillary branches or female flowers, showing the close link between inflorescence architecture and sex ratio. One QTL is detected in the reciprocal cross suggesting some conserved mechanisms underlie these traits. QTL for other reproductive traits include one of major effect for stamen number (30% VE), and one for pollen sterility (12.3% VE). We propose that biased sex ratios in Begonia are caused by selection on loci responsible for inflorescence branching. The consequences of such changes for inbreeding and pollination are also discussed.

6.2. Introduction

Sex allocation is the relative investment in male and female sexual functions. Most plant species are hermaphrodites; in these species sex allocation refers to the investment in male and female
organs within a flower (Charlesworth, 2006). Sex systems that partition male and female functions between unisexual flowers (dichlony), such as monoecy, are less common, but may confer an advantage as sex ratios can be changed in response to resource availability, or to the optimum value given the mean gender of the population (Charlesworth & Morgan, 1991; Ashman & Diefenderfer, 2001; Dorken & Barrett, 2003). Most often plants will minimize investment in expensive female flowers, and develop a male-biased sex ratio (Delph & Lloyd, 1991). A female-biased sex ratio may be advantageous in less resource limiting environments, or when other individuals have male-biased sex ratios (Ashman & Diefenderfer, 2001; Stehlik et al., 2008).

Flowers are arranged on inflorescences, and these can vary in the branching pattern and position of flowers in a range of forms: panicles, racemes or cymes (Schoen & Dubuc, 1990; Ashman & Hitchens, 2000; Prusinkiewicz et al., 2007; Ortiz et al., 2009). Inflorescence architecture, the branching pattern of inflorescences and position of flowers, is dictated by the frequency and pattern in which axillary meristems and floral meristems are produced by the inflorescence meristem and the pattern of internode lengths (Prusinkiewicz et al., 2007). Understanding inflorescence architecture is particularly important in the study of sex ratios in monoecious plants, as flowers of each sex are often produced only at certain points on an inflorescence (Dorken & Barrett, 2003).

*Begonia* is a large genus (>1500 species), where almost all species are monoecious and functionally protandrous (Tebbit, 2005). *Begonia* have conserved inflorescence architectures, always maturing from the bottom up (acropetal), and usually produce male flowers at each branch point and female flowers at the terminals of each branch (Matzke, 1938). *Begonia* species are pollinated by deceit, as female flowers are non-rewarding intersexual mimics of males, which deliver a pollen reward (Ågren & Schemske, 1991; Castillo et al., 2002). Sex ratios may be under sexual selection as female-biased inflorescences are less attractive to pollinators, while increasingly male-biased inflorescences will set less seed (Castillo et al., 2002). Sex ratios within individual *Begonia* plants can change in response to pollen availability (Lopéz & Domínguez, 2003), showing that sex ratio is a flexible trait of importance in natural populations.
Here we investigate the genetic control of sex ratio determination in *Begonia* species. Sex ratios and inflorescence branch patterns are scored in backcross (BC) mapping populations between Central American *B. plebeja* Liebm. and *B. conchifolia* A. Dietr. (Fig. 6.1; Brennan et al., submitted), which differ in their sex ratios and branching patterns. This allows the number of chromosome segments and their effects to be studied by quantitative trait locus (QTL) mapping, without a priori assumptions of gene homology from model species. We also QTL map other reproductive traits (days to first flower, stamen number, pollen stainability and seed set), to compare the number of QTL and their location between reproductive traits. Understanding the genetic control of sex ratios in *Begonia* species, in particular through changes in inflorescence architecture, will enable us to better understand how monoecious species change their sex ratio in response to selection.

![Figure 6.1. Study species. *Begonia plebeja* (left) and *B. conchifolia* (right). White bar is approximately 10cm.](image)

### 6.3. Material and method

#### 6.3.1. Study species and the generation of backcross populations

*Begonia plebeja* and *B. conchifolia* were chosen as they differ in reproductive strategies (Fig. 6.2). *Begonia plebeja* is widespread in dry forests in Mexico, and produces few large inflorescences, typically 3-12cm in diameter, with approximately 40 flowers (Burt-Utley, 1985;
A.D. Twyford, pers. obs.). These flowers are c.2cm in length (stigma-capsule length), and typically have between 11 - 21 stamens (Burt-Utley, 1985; Table 6.2). *Begonia conchifolia* is restricted to wet tropical forests in Costa Rica, and produces many small inflorescences (2 - 10.5cm diameter), which have around 20 flowers. These flowers are 3 - 7mm in length, and have fewer stamens (6 - 13) (Table 6.2; Burt-Utley, 1985). Morphological (Burt-Utley, 1985) and preliminary molecular phylogenetic analyses (N. Harrison, University of Edinburgh, pers. comm.) using nuclear and plastid sequence data support the species as closely-related, and they produce an F1 that is partly fertile (Brennan *et al.*, submitted).

QTL analyses compare fixed genetic differences between parents of a cross, which may differ from average species values due to intraspecific polymorphism. In this case the parents of the backcross differed in inflorescence branch patterns and symmetry in flower production (Fig. 6.2), which are important traits as they may be mechanisms responsible for modifying the sex ratio. The *B. plebeja* parent was roughly symmetrical in branching, and most branches terminated at the same tier (Fig. 6.2a). The *B. conchifolia* parent was asymmetric in branching, and some branches terminated early in a female flower (Fig. 6.2b). The number of female flowers on each half of the inflorescence (i.e. all branching points that arise from one of the two first branches; grey square in Fig. 6.2) was also asymmetric, with *B. plebeja* producing more female flowers on one side of the inflorescence than the other (Fig. 6.2).

The parents of the cross are non-inbred cultivated accessions of unknown provenance. F1 hybrids were generated by pollinating a plant of *B. plebeja* with pollen from a *B. conchifolia* individual, using the pollination procedure described in Chapter 4. F1 plants were raised to flowering on sterilised compost (16 bark: 3 peat: 1 perlite plus finely sieved osmocote) in a glasshouse at 28°C, and used to generate backcross populations. One F1, labeled CKB137_8, was used as pollen recipient to the *B. plebeja* parent, to generate the *B. plebeja* backcross (hereafter PBC) population, the focus of this study. An additional population, generated using a different F1 plant (CKB137_6), was produced using *B. conchifolia* as the pollen donor (*B. conchifolia* backcross, CBC hereafter). This comparative mapping approach allows QTL to be compared in the different genetic backgrounds.
Figure 6.2. Representative inflorescence architecture for *Begonia plebeja* (top) and *Begonia conchifolia* (bottom). Male and female flowers are shown on inflorescence branches (solid black lines). Tiers of branches are shown by dashed grey lines. The blue arrow shows an unfilled tier. Red circles show where predicted branches are missing (branch asymmetry). The large light grey square represents one half of the inflorescence.
Reproductive and inflorescence architecture traits were scored in the two segregating backcross populations (PBC 225 individuals, CBC 117 individuals) and at least 8 clonal replicates of the parents. The mapping populations and parents were grown in the same glasshouse conditions (described above) throughout the phenotyping. *Begonia* can change their sex ratios in response to pollen loads (Lopéz & Domínguez, 2003), and the closed glasshouse conditions mostly prevented pollinating insects visiting, so sex ratios can be observed under pollen-starved conditions.

### 6.3.2. Phenotyping

Fourteen reproductive traits were scored in the backcross populations, listed in Table 6.2. Traits were scored on the first inflorescence except where noted below, and measured as follows. A diagram was drawn of the inflorescence branching structure and the position of male and female flowers, and this was updated twice weekly through the flowering season. The number of male and female flowers was calculated from the inflorescence diagram once flowering had finished. The sex ratio was calculated as the number of male flowers over the total number of flowers across all inflorescences on a plant. The number of tiers of branches (tiers herein) was counted from the inflorescence diagram (Fig. 6.2), as a measure of inflorescence size. Inflorescence asymmetry was calculated by scoring the proportion of flowers of each sex on one half of the inflorescence (i.e. all branching points arising from one of the two first branches; shaded grey on Fig. 6.2) out of the total produced on the inflorescence. This was calculated separately for male (asymmetry males) and female flowers (asymmetry females), and standardised by recording the frequency on the side of the inflorescence with less of the given sex, giving values between 0 and 0.5.

Additional traits that were scored were stamen number and flower time, and two measures of fertility (pollen stainability and seed set). The stamen number was counted on the first flower on the first inflorescence. The tepals were removed and the stamens counted at 10x magnification. The date the first male and female flower opened was recorded, and used to calculate the days to first flower for each sex. To measure pollen sterility, 200 pollen grains from a subset of 125 PBC plants were stained with acetocarmine, and the percentage well-stained recorded (as in Chapter 3). Well stained pollen grains are likely to be viable (Matthews, 2007). Seed set was measured
using 48 PBC plants as pollen recipients to *B. plebeja*, allowing seed set to be assessed independent of pollen viability (Sweigart *et al.*, 2006). Crosses were performed as in Chapter 4. The number of seeds per capsule was estimated with the formula of Ågren & Schemske (1993; see Chapter 4).

The developmental decisions that give rise to different branch patterns were scored in three different ways in the PBC. Two measures compared the number of male and female flowers from each tier to the model of Cozza (2008), which assumes a male flower and two branches at each internal node, and a female flower at the branch terminals (Table 6.1). The first measure was the number of tiers with fewer male flowers than predicted by the model (unfilled tiers, herein; blue arrow Fig. 6.2). As male flowers occur at branching points, this represents the number of tiers where branches have terminated early. The second measure was the sum of the fraction of missing nodes in each tier (branch asymmetry herein; red circle Fig. 6.1); unlike the measure of unfilled tiers, this incorporates the magnitude of asymmetry at each tier. For these two measures, the final tier was ignored as it is always unfilled. Finally, the developmental decisions that give rise to branch patterns were compared on a tier-by-tier basis. The proportion of female flowers or axillary branches that arise from each node was averaged across the tier; male flowers were ignored because there is always one per internal node (Fig. 6.2).

**Table 6.1.** Predicted branching patterns in a symmetrical *Begonia* inflorescence. Adapted from Cozza (2008)

<table>
<thead>
<tr>
<th>Number of levels of branching (tiers)</th>
<th>Number of male flowers</th>
<th>Number of female flowers</th>
<th>Total flowers</th>
<th>Sex ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>0.33</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>4</td>
<td>7</td>
<td>0.43</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>8</td>
<td>15</td>
<td>0.47</td>
</tr>
<tr>
<td>4</td>
<td>15</td>
<td>16</td>
<td>31</td>
<td>0.48</td>
</tr>
<tr>
<td>5</td>
<td>31</td>
<td>32</td>
<td>63</td>
<td>0.49</td>
</tr>
<tr>
<td>6</td>
<td>63</td>
<td>64</td>
<td>127</td>
<td>0.50</td>
</tr>
</tbody>
</table>
6.3.3. Segregation patterns and quantitative analyses

Quantitative analyses were performed separately for the 14 traits from the 2 populations. The distribution of each trait was assessed by drawing histograms. Thirteen traits were then tested for transgressive segregation with the approach of Kim & Rieseberg (1999). Transgressive segregation for flower time could not be assessed as seeds were not available from the parent plants for comparison. Non-normally distributed traits were first transformed to fit the assumption of normality, count data was square root transformed and proportions arcsine transformed. Traits were considered transgressive if the number of progeny that exceed the higher parental value by 2, 3, or 4 standard deviations, was greater than the number expected by chance given the population size. Trait values for the parents were the average scores from at least 8 clonal replicate plants.

Many traits are not expected to be independent of each other, such as sex ratio and number of male and female flowers, and this was tested by correlation analysis between a subset of reproductive traits. Fifteen correlation analyses were performed, listed in Table 6.3. Pearson’s correlation analyses were performed in Minitab Statistical Software (www.minitab.com), and P-values were adjusted with Bonferonni correction to account for multiple testing. The correlations between traits were compared to the location that QTL mapped to (below), to test whether related traits actually represent a single trait.

Measures of vegetative vigour from a parallel QTL study (Mobina Shaukat Ali, University of Edinburgh, pers. comm.) were compared with reproductive output (number of inflorescences, inflorescence tiers, number of flowers), to test whether reproductive output is constrained by the size of the plant (Schoen & Dubuc, 1990). For each PBC plant, the largest leaf was harvested 6 months after germination, photographed against a black background, and the surface area calculated using LAMINA (Bylesjö et al., 2008). The mass of this leaf was then measured after drying at 60°C for 5 days (dry mass). The number of leaf scars along each rhizome was added to number of leaves, representing the number of leaves produced in the life of the plant (number of leaves here in). The number of leaves was multiplied by the leaf surface area to estimate a fourth proxy of vigour, total leaf surface area. Correlation analyses were performed between the four proxies of vigour and the three proxies of reproductive output using Minitab, with P-values adjusted using Bonferonni correction.
6.3.4. QTL analyses

QTL analyses were performed using the genetic map of Brennan et al. (submitted). The map is constructed using co-dominant SNP markers in candidate developmental genes, and dominant AFLPs. The combined dataset for the PBC was 162 markers (80 AFLPs and 82 SNPs) for 225 individuals, and for the CBC 155 markers (74 AFLPs and 81 SNPs) for 117 individuals. The markers were spread across the 14 distinct linkage groups, and the maps were mostly syntenous in their marker order. The total length of the B. plebeja map was 1099-1110cM, and the B. conchifolia map 1034-1043cM (depending on the approach used; Brennan et al., submitted).

QTL were detected with composite interval mapping, which accounts for co-factors (other QTL that affect a given trait). All QTL analyses was performed in QGene v4.0 (Joehanes & Nelson, 2008). Step-wise cofactors were selected, and the scan interval set to 10cM. The significance of each trait (P < 0.05, < 0.01) was tested using 1000 permutations. Traits that were not normally distributed were transformed prior to QTL analysis (as above).

The data for average proportion of axillary branches or female flowers per node, for each tier, were skewed and could not be transformed to a normal distribution. These traits were first analysed using composite interval mapping, as above, which is usually robust to traits that deviate from normality (Knott, 2005). These QTL results were confirmed with non-parametric analyses; traits were coded as ordinal data and used in single-marker analyses in QGene.

The likelihood of finding significant QTL may be reduced when genotype data are missing. Therefore, a second dataset was prepared, where missing BC genotypes were replaced with the most likely genotype inferred from the surrounding markers. Genotypes were arranged in map order in a MS Excel spreadsheet, and different colours used for homozygote and heterozygote genotypes. Missing genotypes within a colour block were filled in assuming no recombination is more likely than two recombination events. Missing data was not imputed at the ends of linkage groups, or where the surrounding markers show a genotype transition. All analyses where marginally significant QTL (i.e. P = 0.05 - 0.10) were detected, were then repeated using the imputed dataset.
6.3.5. Candidate gene detection

The genetic basis of QTL localizing between AFLP markers were not investigated further. Those located within ~5cM (~2.4Mb; Brennan et al., submitted) of a SNP marker were investigated for potential candidate genes. First, the transcriptome isotig from which the SNP marker was designed was BLAST searched against the draft *B. conchifolia* genome (Kidner, unpubl. data), to retrieve the surrounding sequence. The open reading frames (ORFs) within the genome contig were then located with the EMBOSS tool getorf (Rice et al., 2000). Finally, ORFs were BLAST searched against the annotated protein sequence database for *Arabidopsis* (TAIR; Swarbreck et al., 2008), to find the putative function of the surrounding genes. For each likely match (e < 10^{-40}) the gene name, gene ontology (GO) annotation, expression profile, and mutant phenotypes were recorded.

6.4. Results

6.4.1. Trait segregation

The two species differed for most reproductive traits that were measured (Table 6.2). *B. plebeja* produced fewer inflorescences than *B. conchifolia*, and these inflorescences had more tiers of branching, and double the number of male and female flowers. *Begonia plebeja* had inflorescences with roughly symmetrical branching, while *B. conchifolia* had 23% fewer internal branches than predicted by the model of Cozza (2008), and half the tiers did not have the full complement of branches. These measures of inflorescence branching had large variances in the parents (Table 6.2). The distribution of male flowers was roughly equal, but both species had an unequal distribution of females across the inflorescences, and this was more pronounced in *B. plebeja. Begonia conchifolia* flowers had two-thirds fewer stamens than *B. plebeja.*
Figure 6.3. Trait segregation in the PBC population. Blue arrow: *B. plebeja* average parental value; red arrow, *B. conchifolia* average parental value.

All trait differences segregated in the backcross populations (Fig 6.3). Some PBC individuals exceeded parental values for ten traits (number of inflorescences, male flowers, female flowers, stamen number, number of tiers, branch asymmetry, unfilled tiers, asymmetry males, asymmetry females, sex ratio). Transgressive segregation was found for seven traits, including: sex ratio (P
< 0.001), four measures of asymmetry (unfilled tiers, \( P < 0.01 \); branch asymmetry, \( P < 0.05 \); asymmetry males, \( P < 0.05 \); asymmetry females, \( P < 0.05 \)) and two other traits (number of inflorescences, \( P < 0.001 \); number of tiers, \( P < 0.05 \)). Transgressive segregation was only found for a single trait in the CBC, number of male flowers (\( P < 0.05 \)). Transgressive segregation could not be tested for number of inflorescences as the data could not be transformed to fit the assumption of normality.

**Table 6.2.** Average trait values for the parents of the mapping populations, with standard errors in brackets (\( n = 8 \)).

<table>
<thead>
<tr>
<th>Trait</th>
<th>B. plebeja</th>
<th>B. conchifolia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of inflorescences</td>
<td>2.1 (0.3)</td>
<td>5.2 (0.8)</td>
</tr>
<tr>
<td>Male flowers 1(^{st}) inflorescence</td>
<td>38 (6.4)</td>
<td>19.1 (4.4)</td>
</tr>
<tr>
<td>Female flowers 1(^{st}) inflorescence</td>
<td>16.7 (3.7)</td>
<td>7.9 (3.2)</td>
</tr>
<tr>
<td>Total flowers 1(^{st}) inflorescence</td>
<td>54.7 (9.6)</td>
<td>27.0 (7.5)</td>
</tr>
<tr>
<td>Number of tiers 1(^{st}) inflorescence</td>
<td>5.9 (0.4)</td>
<td>3.3 (0.6)</td>
</tr>
<tr>
<td>Sex ratio</td>
<td>0.68 (0.03)</td>
<td>0.76 (0.03)</td>
</tr>
<tr>
<td>Unfilled tiers</td>
<td>0.13 (0.13)</td>
<td>0.5 (0.34)</td>
</tr>
<tr>
<td>Branch asymmetry</td>
<td>0.05 (0.05)</td>
<td>0.23 (0.17)</td>
</tr>
<tr>
<td>Asymmetry males</td>
<td>0.46 (0.08)</td>
<td>0.45 (0.03)</td>
</tr>
<tr>
<td>Asymmetry females</td>
<td>0.24 (0.09)</td>
<td>0.33 (0.02)</td>
</tr>
<tr>
<td>Days to first male flower</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Days to first female flower</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Pollen viability</td>
<td>&gt; 90%</td>
<td>&gt; 90%</td>
</tr>
<tr>
<td>Seed set</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Stamen number (flower from 1(^{st}) tier 1(^{st}) inflorescence)</td>
<td>25.0 (0.9)</td>
<td>7.9 (0.7)</td>
</tr>
</tbody>
</table>

Only one significant correlation was found in the 12 pairwise comparisons between four proxies of vegetative vigour (area largest leaf, leaf dry mass, number of leaves, total leaf surface area) and three proxies of reproductive output (number of inflorescences, number of tiers and total flowers on the first inflorescence) after correcting for multiple testing. This was between the number of leaves and the number of inflorescences (\( r^2 = 0.296, P < 0.05 \)).
Significant correlations were found between the number of flowers of each sex and the number of tiers on an inflorescence, with a greater proportion of variance explained for male than female flowers (Table 6.3). Similarly, the likelihood of producing axillary branches (as opposed to female flowers) on tiers 4 and 5 explained more variance in the male than female flowers. The number of tiers and number of inflorescences were weakly correlated, but no correlation was found between the number of flowers on an inflorescence and the number of inflorescences, or the sex ratio and number of tiers. Stamen number was not correlated with number of male flowers.

Table 6.3. Pearson’s correlation between reproductive traits. * P < 0.05; ** P < 0.01

<table>
<thead>
<tr>
<th>Traits</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of male flowers</td>
<td>Number of female flowers</td>
</tr>
<tr>
<td>Number of male flowers</td>
<td>Total flowers</td>
</tr>
<tr>
<td>Number of female flowers</td>
<td>Total flowers</td>
</tr>
<tr>
<td>Number of male flowers</td>
<td>Number of tiers</td>
</tr>
<tr>
<td>Number of female flowers</td>
<td>Number of tiers</td>
</tr>
<tr>
<td>Total flowers</td>
<td>Number of tiers</td>
</tr>
<tr>
<td>Sex ratio</td>
<td>Number of tiers</td>
</tr>
<tr>
<td>Sex ratio</td>
<td>Total flowers</td>
</tr>
<tr>
<td>Number of male flowers</td>
<td>Axillary branches tier 4</td>
</tr>
<tr>
<td>Number of female flowers</td>
<td>Axillary branches tier 4</td>
</tr>
<tr>
<td>Number of male flowers</td>
<td>Axillary branches tier 5</td>
</tr>
<tr>
<td>Number of female flowers</td>
<td>Axillary branches tier 5</td>
</tr>
<tr>
<td>Tiers</td>
<td>Number of inflorescences</td>
</tr>
<tr>
<td>Total flowers</td>
<td>Number of inflorescences</td>
</tr>
<tr>
<td>Stamen number</td>
<td>Number of male flowers</td>
</tr>
</tbody>
</table>

6.4.2. QTL mapping

Five significant QTL were detected for reproductive traits in the PBC, and these were for different measures of inflorescence branching or sex ratios (Table 6.4; Fig. 6.4). Two marginally significant QTL (P = 0.05 - 0.10) were significant when missing genotype data were imputed (pollen sterility, stamen number). QTL contributed 12.2 - 30% of the phenotypic variation
explained (PVE), with only a single trait, stamen number, controlled by a locus of large effect (i.e. PVE > 25%, Kim & Rieseberg, 1999, and references therein). The 7 QTL were distributed across 4 linkage groups, with 4 QTL found on linkage group 11. Two of the QTL, for number of flowers and number of female flowers, overlapped in their one-LOD support interval. In addition, a marginally significant QTL (P < 0.1) was found for female flower time in the PBC [linkage group 2, LOD = 3.428, PVE = 19.2%, position 36cM (33.0 - 37.0, 1 LOD support range)]. The full QTL data are given in Appendices 9 and 10.
Figure 6.4. Representative QTL plots for traits in the PBC: number of male flowers (top left), sex ratio (top right), stamen number (bottom left, shown for imputed data), pollen stainability (bottom right). The LOD score (green line) is plotted against each marker (bottom axis) on a given linkage group (top axis). The imputed threshold is indicated by a solid horizontal line, * P < 0.05; ** P < 0.01.
Two QTL were found in the CBC, both for sex ratio, and these explained 19-22% of the phenotypic variation. One QTL for sex ratio in the CBC mapped to the same linkage group, and overlapped in 1-LOD interval, as the male flower QTL in the PBC.

Additional QTL were found for meristem decisions at each tier of branching (Table 6.5). Significant QTL were found with single-marker regression for tiers 4, 5 and 6, with the QTL for axillary branches and female flowers on linkage group 6 overlapping in their 1-LOD interval. Only one of these QTL, for female flowers on tier 4 (linkage group 6), was also found with composite interval mapping. This QTL overlapped in 1-LOD interval with the PBC QTL for number of male flowers, and was on the same linkage group as a sex ratio QTL in the CBC (Table 6.4).

### 6.4.3. Candidate genes

Eleven QTL localised within 5cM of a SNP marker, with 9 of the 11 SNPs being different. Five SNPs had ORF matches to genes in the *Arabidopsis* database, however none of these matched candidate genes directly related to inflorescence or flower development, or transcription factors.
Table 6.4. Putative QTL for reproductive traits in two backcross *Begonia* populations. Imputed values are reported in brackets after the original dataset. PVE: percentage variance explained. Population: PBC, *B. plebeja* backcross; CBC, *B. conchifolia* backcross. The direction of additive effects are shown in square brackets after the value; C, *B. conchifolia* alleles overrepresented; P, *B. plebeja* alleles overrepresented.  

<table>
<thead>
<tr>
<th>Population</th>
<th>Trait</th>
<th>Linkage group</th>
<th>LOD score</th>
<th>PVE (%)</th>
<th>Additive effect</th>
<th>Position cM (1 LOD interval)</th>
<th>Nearest SNP</th>
</tr>
</thead>
<tbody>
<tr>
<td>PBC</td>
<td>Pollen sterility</td>
<td>4</td>
<td>3.061&lt;sup&gt;ns&lt;/sup&gt; (3.569*)</td>
<td>10.6</td>
<td>10.039 [C]</td>
<td>11.0 (10.0-16.0)</td>
<td>BLP1</td>
</tr>
<tr>
<td>PBC</td>
<td>Number of male flowers</td>
<td>6</td>
<td>3.434*</td>
<td>14.5</td>
<td>13.004 [P]</td>
<td>41.0 (38.0-58.0)</td>
<td>n/a</td>
</tr>
<tr>
<td>PBC</td>
<td>Number of female flowers</td>
<td>11</td>
<td>3.396*</td>
<td>14.8</td>
<td>6.491 [P]</td>
<td>67.6 (62.6-71.6)</td>
<td>BMYBTF5</td>
</tr>
<tr>
<td>PBC</td>
<td>Total number of flowers</td>
<td>11</td>
<td>4.007*</td>
<td>16.7</td>
<td>18.494 [P]</td>
<td>71.6 (69.7-76.6)</td>
<td>BMYBTF5</td>
</tr>
<tr>
<td>PBC</td>
<td>Asymmetry males</td>
<td>11</td>
<td>4.133*</td>
<td>18.7</td>
<td>0.11 [C]</td>
<td>59.6 (58.6-62.6)</td>
<td>BSBP1</td>
</tr>
<tr>
<td>PBC</td>
<td>Sex ratio</td>
<td>11</td>
<td>4.575**</td>
<td>13.4</td>
<td>0.077 [C]</td>
<td>27.6 (18.6-29.6)</td>
<td>BP2</td>
</tr>
<tr>
<td>PBC</td>
<td>Stamen number</td>
<td>8</td>
<td>3.311&lt;sup&gt;ns&lt;/sup&gt; (4.4*)</td>
<td>23.8</td>
<td>5.264 [C]</td>
<td>27.7 (19.7-39.7)</td>
<td>BDFL2</td>
</tr>
<tr>
<td>CBC</td>
<td>Sex ratio</td>
<td>6</td>
<td>3.557*</td>
<td>19.9</td>
<td>4.704 [C]</td>
<td>58.0 (57.0-60.0)</td>
<td>BEIN1, BMYBTF6</td>
</tr>
<tr>
<td>CBC</td>
<td>Sex ratio</td>
<td>12</td>
<td>3.987*</td>
<td>22.0</td>
<td>0.08 [C]</td>
<td>21.0 (13.0-34.0)</td>
<td>BHP1</td>
</tr>
</tbody>
</table>
Table 6.5. QTL mapping of meristem decisions and inflorescence branching in the PBC. Phenotypic scores are the average probability of producing female flowers or axillary meristems across meristems at the given tier of branching. Analysis: SMR, single-marker regression; CIM, composite interval mapping.

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Trait</th>
<th>Linkage group</th>
<th>LOD score</th>
<th>PVE (%)</th>
<th>Additive effect</th>
<th>Position cM (1 LOD interval)</th>
<th>Nearest SNP</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIM</td>
<td>Tier 4 female flowers</td>
<td>6</td>
<td>3.86*</td>
<td>15.9</td>
<td>0.209 [C]</td>
<td>51.0 (37.0-64.0)</td>
<td>n/a</td>
</tr>
<tr>
<td>SMR</td>
<td>Tier 4 female flowers</td>
<td>6</td>
<td>4.146**</td>
<td>16.9</td>
<td>7.246 [C]</td>
<td>41.5 (37.6-68.6)</td>
<td>n/a</td>
</tr>
<tr>
<td>SMR</td>
<td>Tier 4 axillary branches</td>
<td>6</td>
<td>3.661**</td>
<td>15.1</td>
<td>7.232 [P]</td>
<td>41.5 (37.6-68.6)</td>
<td>n/a</td>
</tr>
<tr>
<td>SMR</td>
<td>Tier 5 axillary branches</td>
<td>5</td>
<td>2.581*</td>
<td>11.5</td>
<td>5.045 [P]</td>
<td>13.2 (0-36.4)</td>
<td>BTF33</td>
</tr>
<tr>
<td>SMR</td>
<td>Tier 6 axillary branches</td>
<td>5</td>
<td>2.16*</td>
<td>9.2</td>
<td>0.859 [P]</td>
<td>0 (0-19.2)</td>
<td>BTF33</td>
</tr>
</tbody>
</table>
6.5. Discussion

6.5.1. QTL for reproductive traits in *Begonia* mapping populations

The two study species shared the same inflorescence architectures (i.e. branch pattern and positional differences in flower sex), but differed in their sex ratios, with *B. conchifolia* having a more male-biased sex ratio than *B. plebeja*. Differences in sex ratios could not be explained by the number of tiers on an inflorescence (Table 6.1, 6.3), nor by developmental constraint of investment in reproductive structures as more vigorous plants did not produce bigger inflorescences (or have more female-biased sex ratios). Correlation analyses in the PBC showed the number of tiers strongly dictates the number of male flowers per inflorescence ($r^2 = 0.970$, $P < 0.01$), but the number of female flowers is less closely linked to architecture ($r^2 = 0.874$, $P < 0.01$), as found in *Begonia urophylla* Hook. (Cozza, 2008). This raises the prospect that genes underlying inflorescence architecture affect both male and female flower number, with additional genes further modifying the number of female flowers.

The differences in sex ratios between parents (Fig. 6.2) allowed genetic analysis through QTL mapping. Five QTL were found for traits related to sex ratios in the PBC (number of males, number of females, total flowers, asymmetry in males, sex ratio; Table 6.4). However, the strong correlation between many of these traits in the PBC (Table 6.3), and the co-localization of 4 QTL to linkage group 11, suggests these traits are not independent. This is most obvious for number of female and total number of flowers, where QTL overlapped in their 1-LOD interval and had similar additive effects. Whether the other QTL represent different genes is hard to tell with the sparse marker coverage on linkage group 11. Correlations between traits in natural populations would help distinguish the independence of traits.

To investigate the way these QTL influence the phenotype, we investigated meristem decisions (female flower or axillary branches) at each tier. The most striking result was that a QTL for number of male flowers mapped to the same position on linkage group 6, as a QTL for meristem decisions on tier 4 in the PBC. This meristem decision QTL also co-localised with a sex ratio QTL in the CBC. This supports *Begonia* inflorescences differing in their sex ratios by
terminating some inflorescence branches in a female flower earlier than others. This gains further support from the strong correlation between meristem decisions at tier four and the number of both male and female flowers (Table 6.3). An alternative mechanism that sex ratios could differ, independent of branching, is by selectively aborting male or female flowers. Cozza (2008) suggested that *Begonia urophylla* selectively aborts female flowers, as shown by the weaker correlation between inflorescence architecture and female flowers, than with male flowers. Moreover, two small female buds are visible at each branch terminal, not all of which develop. It also seems likely selective abortion of female flowers occurs in the PBC, as the correlation between sex ratio and number of tiers of branching or total flowers was not significant after Bonferroni correction, and the correlation only explained a modest amount of variance in the sex ratio.

Some QTL for sex ratio may be shared between species, as shown by sex ratio QTL in the CBC mapping to syntenous markers to a QTL for number of male flowers in the PBC. Whether the second QTL for sex ratio in the CBC (linkage group 12) represents a novel mutation, as it had no similar QTL in the PBC, is currently unclear, as a different F1 parent was used to generate the two backcross populations. Moreover, the CBC had a smaller population size and so the ability to detect QTL is reduced.

Comparisons of QTL for inflorescence traits with QTL for vegetative traits will be necessary to more fully understand the interacting mechanisms that give rise to differences in sex ratios. The non-independence of vegetative and inflorescence traits is suggested by the moderate amount of variance explained in the number of inflorescences by the number of leaves on a plant ($r^2 = 0.296$, $P < 0.05$). It will also be important to understand phenotypic plasticity in reproductive traits under different environments with different levels of resources available, or with different rates of pollinator visitation, as these factors have been found to affect sex ratios in *Begonia* (Lopéz & Domínguez, 2003).

In addition to QTL relating to sex ratios, a QTL of large effect was found for stamen number (30% VE, imputed data), which differs greatly between the two mapping parents (Table 6.2). This shows the different reproductive strategies of the species, with *B. plebeja* making few large inflorescences with expensive flowers (large tepals, many stamens), and *B. conchifolia* more
inflorescences with less expensive flowers (smaller tepals, fewer stamens). The genetics of stamen number are not well understood in model systems, in part due to the conserved stamen number of 6 found in all Brassicaceae species, preventing mapping of interspecific differences between Arabidopsis species (Edwards & Weinig, 2011). However, candidate genes for differences in stamen number would include those that control meristem activity such as AGAMOUS (Yanofsky et al., 1990) and CLAVATA (Clark et al., 1993). Having a single locus of large effect controlling stamen number has important evolutionary implications, as an allele at this locus could be quickly fixed by strong selection (Louthan & Kay, 2011), or may be lost by genetic drift in small populations. Future fine mapping and comparisons of sequence variation at this locus in other Begonia species may elucidate more about this important reproductive trait.

A single QTL of modest effect (12.3% VE, imputed data) was found for pollen sterility. The position of the sterility QTL does not correspond to markers that segregate in a non-Mendelian manner (transmission ratio distortion), nor map close to recombination breakpoints (Brennan et al., submitted). Phenotyping more individuals, or having more markers on the map, will help resolve the genetic basis of pollen sterility in the Begonia BC population. This warrants further work as pollen sterility is common in F1 crosses in Begonia (Matthews, 2007; Dewitte et al., 2011), and plays an important role in maintaining species identities in secondary contact (Chapter 5). Different genetic architectures have been found for pollen sterility in other organisms. Two pollen sterility QTL of large effect were found in a BC1 Helianthus mapping population (H. annuus x H. debilis ssp. cucumerifolius; 38.2-38.8% VE; Kim & Rieseberg, 1999), whereas pollen sterility was polygenic in a higher resolution BC2 map between different species (H. annuus x H. petiolaris; 11 QTL, 5-26% VE; Lai et al., 2005).

6.5.2. Candidate genes for inflorescence development in Begonia

Begonia species have a distinct inflorescence architecture (Fig. 6.2; Cozza, 2008), and this will be determined by the position of floral and inflorescence meristems (Prusinkiewicz et al., 2007). In Begonia the developing inflorescence meristem rapidly undergoes a transition to flower meristem, terminating the inflorescence in a male flower. After this, two new axillary inflorescence meristems develop, which give rise to subsequent branches and flowers. This sympodial pattern of inflorescence architecture bears some resemblance to tomato (Solanum
lycopersicum L.), except that the axillary meristems in tomatoes first produce leaves (i.e. they start as vegetative meristems before becoming inflorescence meristems) (Park et al., 2012), whereas in Begonia they produce inflorescences. A gene that may play a role in inducing axillary inflorescence meristems is TERMINAL FLOWER 1 (TFL1; Alvarez et al., 1992; Liljegren et al., 1999; Conti & Bradley, 2007), as TFL1 is involved in shoot meristem activity, and tfl1 mutants produce flowers instead of shoots on the inflorescence (Bradley et al., 1996). FLOWERING LOCUS F (FT) has the opposite function (Hanano & Goto, 2011), and could be responsible for terminating inflorescences in a flower.

This study investigated the ways that monoecious plants can change their sex ratios; however, the molecular genetic mechanisms for determining unisexual flowers are still unclear. Understanding sex determination in Begonia inflorescences will help shed light on how monoecy has evolved and is maintained. Studies of sex determination in cucumbers, which are also monoecious, and are in the same order as Begonia (Cucurbitales), have shed some light on this question, and some mechanisms may be conserved. One hypothesis proposes that sex determination is hormonally regulated. Ethylene, stimulated by auxin, arrests anther development in female flowers, and the positional sex effects in Begonia inflorescence may be due to lower expression of ethylene receptors (e.g. CsETR1) at more distal nodes as is the case in cucumber (Wang et al., 2010). Alternatively, differential expression of AGAMOUS (AG, Yanofsky et al., 1990) in different flower meristems may cause sex differences. Two homologs of AGAMOUS (CUM1 and CUM10) have been cloned in cucumber (Cucumis sativus), and these genes regulate floral organ identity, with mutants converting sepals into carpel-like structures and petals into stamens (CUM1), and petals into anther-like structure (CUM10) (Kater et al., 1998). Studying hormonal effects on sex ratios, as well as candidate genes such as AG, will be important for understanding the evolution of monoecy. However, hundreds of genes are differentially expressed between unisexual cucumber flowers and gynodioecious mutants, including hormone signaling genes and transcription factors, suggesting the process of sex determination is complex (Guo et al., 2010; Wu et al., 2010).
6.5.3. Adaptive significance of inflorescence architectures

The distinct inflorescence architecture (bisexual, protandry, acropetaly) described here is common to most *Begonia* species. The large Asian section Petermannia (> 270 species) is an exception, as species are protogynous, producing a two flowered female inflorescence distal to a many-flowered male inflorescence (D. C. Thomas, University of Hong Kong, pers. comm.). There are also rare reports of dioecy (Tebbit, 2005) or plants with unisexual inflorescences (Forrest, 2000). It does however appear that most *Begonia* species share an inflorescence architecture with sex ratios related to the arrangement of flowers and the pattern of inflorescence branching. What are the reasons that inflorescence architectures are conserved with only minor modification across *Begonia*? Many other families have a conserved inflorescence type, and these include: the compound umbel of the Apiaceae (Umbelliferae), cymose capitulum of the Asteraceae (Compositae), ultimate spike of the Poaceae, spadix of the Araceae, and the helicoid cyme of the Boraginaceae (Tucker, 1999). Inflorescence types are even shared between related families (Endress & Doyle, 2009). One hypothesis is that a single inflorescence type needs only minor modifications to take a large range of forms. In this case, terminating inflorescence branches early can change the sex ratio in response to a variety of selection pressures, without changing the whole inflorescence form.

*Begonia* species tend have the joint syndromes of acropetaly, protandry, and bee pollination (Wyatt & Sazima, 2011). This combination of characteristics is predicted to be maladaptive (Darwin, 1877) as bees typically visit inflorescences from the bottom up (Harder & Barrett, 2000; Jordan & Harder, 2006), and this would cause frequent self-pollination and reduce pollen movement between plants (pollen discounting, Harder & Barrett, 1995). This assumes that the foraging paths of solitary bees follow those bee species that have been studied to date (Jordan & Harder, 2006), and it may be that temporal separation of sexes reduces self-pollination in these fully compatible species (Ågren & Schemske, 1993; Wyatt & Sazima, 2011). However, this inflorescence architecture may be advantageous in other ways. Most notably, producing cheap male flowers early in the flowering season may provide ample opportunities for cross-pollination, with the number of expensive female flowers determined later depending on the available resources.
The inflorescence architecture may also bring fitness benefits when the phase that flowers open across the inflorescence, which will alters the floral display as seen by a pollinator at a given time (the display architecture), is considered (Jordan & Harder, 2006). Producing rewarding male flowers first may be necessary to attract pollinators to visit the deceptive female flowers that follow (Ågren & Schemske, 1991; Wyatt & Sazima, 2011; Castillo et al., 2012). Differences in display architecture that may be related to pollinator visitation can be seen between the study species, with *B. plebeja* producing a large display of female flowers on the terminal tier, whereas *B. conchifolia* has a longer period of time where a similar number of male and female flowers are open together (Fig. 6.2). It may therefore be that species differences in sex ratios could be a by-product of pollinator selection for display architecture. *Begonia conchifolia* inflorescences with female flowers only at the terminal tier may be selected against as the female-biased display would not be as attractive to pollinators, and instead inflorescences with females dispersed over multiple tiers next to male flowers may be favoured. In contrast, the larger terminal tier on a *B. plebeja* inflorescence has more positions available for female flowers to occur with male flowers. The smaller display of female flowers at a given time on *B. conchifolia* may also promote selfing (Harder & Barrett, 1995), which is supported by the more homozygous transcriptome of *B. conchifolia* relative to *B. plebeja* (Brennan et al., submitted). Finally, inflorescence architectures will affect other factors not considered in this study, such as the transport of resources (Schoen & Dubuc, 1990), and the structural support of fruit (Stebbins, 1973), therefore this architecture may have evolved in response to other selection pressures.
CHAPTER 7: General conclusions

7.1. Evolution of diversity in *Begonia*

Studying the evolutionary mechanisms that promote diversification in speciose genera is difficult because of their unmanageable sizes, and because many different processes interact to shape patterns of species richness. For these reasons, evolutionary histories may be investigated in a subset of species where patterns are more broadly applicable to the group of interest. This thesis used genetic markers from two genomes (nuclear and plastid), in different contexts (phylogeography, population genetic, hybridization), to study population processes in two widespread *Begonia* species. The results have shed light on the processes that promote speciation in *Begonia*, and the main findings are outlined below.

First, *B. heracleifolia* had strong population genetic structure at plastid loci (Chapter 2). This shows that without mechanisms to promote seed dispersal, or any known seed dispersal vector, inferred seed dispersal distances are low and populations are weakly connected by seed-mediated gene flow. While seed dispersal is usually low relative to pollen dispersal (Ennos, 1994), and plays a limited role in maintaining species cohesion, it is essential for colonizing new habitats. Species with poor dispersal are less likely to invade other congeneric species, which can cause species collapse through introgression, especially in the early stages of speciation when reproductive barriers are weak.

Second, low pollen-mediated gene flow in these extremely widespread species leads to a breakdown in species cohesion (Chapter 4). *Begonia heracleifolia* has a distribution range of ~3200km, from West Mexico to Panama, and *B. nelumbiifolia* ~2000km, from Central Mexico to Honduras; high gene flow may be expected to maintain species cohesion over such distances. However, interpopulation pollen flow inferred from nuclear markers was low, with populations showing strong geographic structure. This result is supported by populations of *B. heracleifolia* being divergent in their morphology (leaf shape), as well as accumulating genetic incompatibilities causing interpopulation crosses to have a reduced fertility (Chapter 4). Genetic divergence may be ongoing in these species, with speciation a possible outcome. This is
supported by narrow endemic species being present at the edge of the range of *B. heracleifolia*, with these species distinguished from *B. heracleifolia* by few morphological characters (e.g. *B. philodendroides* Zeisenh.). While a break down in species cohesion is consistent with the data, it may actually be that weak species cohesion is maintained by strong selection for advantageous alleles at loci for adaptive traits (Rieseberg & Burke, 2001; Morjan & Rieseberg, 2004). In this way genome wide divergence can occur (as observed with high mean $F_{ST}$ at 9 nuclear microsatellites), and cohesion is maintained by low gene flow and strong selection at specific loci. This seems most possible for *B. nelumbiifolia*, given its wide distribution and morphological uniformity throughout its range. To definitively understand mechanisms of species cohesion, it will be necessary to understand levels of divergence across the genome, including at loci that may contribute to adaptive differences.

Third, both species were found to be self-compatible and have moderate levels of inbreeding (Chapter 4), similar to other *Begonia* species (Ågren & Schemske, 1993; Matolweni et al., 2000). Inbreeding is common in herbaceous plants, which on average have a four-fold higher average $F_{IS}$ value than woody plants (Duminil et al., 2009). There is clear potential for inbreeding to promote the genetic isolation of populations by reducing interpopulation gene flow, and population structure measured with $F_{ST}$ is correlated with breeding system estimates in plants (Duminil et al., 2009). Inbreeding may therefore promote speciation in tropical herbs (Lasso et al., 2011). Some *Begonia* species may have evolved to promote inbreeding by densely packing self-compatible flowers on inflorescences, while others discourage it by more strongly separating male and female flowers both spatially and temporally. Levels of inbreeding in *Begonia* are comparable with other understory tropical plants such as bromeliads, and like bromeliads outcrossing rates and morphologies to promote inbreeding appears to differ between species (Palma-Silva et al., 2009).

Fourth, reproductive isolating barriers were relatively strong in sites of secondary contact between two pairs of species (Chapter 5). Species barriers were maintained by a number of pre- and post-zygotic reproductive isolating barriers, including differences in flowering time and F1 hybrid sterility. Such differences have also been found between divergent populations within *B. heracleifolia*, suggesting some of these barriers may have evolved during divergence. The complexity of reproductive isolation between diploid species in conjunction with strong
population structure points to allopatric speciation as an important mode of speciation in the group. The genetic mechanisms that underlie traits that differ between species were then tested by QTL analysis, focusing on reproductive traits (Chapter 6). This analysis, as well as a parallel study of ecophysiological and micromorphological traits in the same BC populations (Mobina Shaukat Ali, unpubl. data) show that some QTL of moderate to large effect influence important species differences. Rapid evolution for such traits with a simple genetic architecture may reduce the likelihood of introgression between species.

Taken in concert, poor pollen and seed flow between populations, frequent self-fertilization, and reproductive barriers that evolve quickly, may pre-dispose the genus to rapid speciation. These processes are unlikely to be unique to Begonia, and the findings in this thesis are consistent with the predictions made for other large plant genera (Sanderson & Wojciechowski, 1996). In particular, support is given for species in large genera having a geographic population structure with restricted gene flow (Sanderson & Wojciechowski, 1996). An explanation for a second factor, that large genera are typically composed of many ecologically specialised species, can also be proposed from the data presented here. Rapid shifts in ecology may occur when selection acts on a trait with a simple genetic architecture, and low gene flow between incipient species will allow divergence for these adaptive traits. Moreover, ecological isolation may influence the outcomes of secondary contact, with selection against hybrids or low intrinsic fitness of hybrids caused by divergence of adaptive loci causing conflicts.

7.1.1. Comparison with Senecio

Common characteristics that promote speciation in phylogenetically divergent lineages can be identified by comparing Begonia with other species-rich genera. These common characteristics will influence speciation histories, even if genera are found in different geographic areas and have diversified over different time scales. Little is known about the population biology and evolutionary genetics of most species-rich plant genera, with Senecio being a notable exception. Many broad parallels can be drawn between Senecio and Begonia, such as their wide variety of vegetative and floral forms (Senecio species are very variable in: leaf shape, indumentums type, inflorescence type, flower colour, Barkley, 1978, cited by Pelser et al., 2007) and the habitats species are adapted to, the frequent occurrence of hybridization and polyploidy (Lowe & Abbott,
2000; Kirk et al., 2004; Lowe & Abbott, 2004; James & Abbott, 2005; Chapman & Abbott, 2010), and the tendency for some species to grow as invasive ruderals over large distribution ranges (e.g. *S. vulgaris* L.). Other life-history traits are quite dissimilar. For example self-incompatibility is common in *Senecio* but not know from any *Begonia* species. Here, the four factors proposed to promote diversification in *Begonia* (above) are compared to *Senecio*, to see whether similar characteristics promote diversification in both groups.

The first factor to consider is whether seed dispersal limitation affects patterns of genetic diversity in *Senecio*. *Senecio* fruits (achenes) have a pappus to promote wind dispersal, and trichromes to aid animal attachment and subsequent dispersal (Schmitt, 1980; Mcevoy & Cox, 1987), so high seed dispersal between populations would be expected. No correlation was found between geographic distance and genetic divergence at plastid loci in the Mediterranean *Senecio gallicus* Chaix, suggesting long-distance seed dispersal and Pleistocene range expansion have obscured geographic structuring of genetic diversity, as would be expected if seed dispersal was limiting (Comes & Abbott, 1998). Studies of Israeli *Senecio glaucus* L. and *S. vernalis* Walst. and Kit. show plastid haplotypes are widespread and this is likely caused by long-distance seed dispersal (Comes & Abbott, 1999). Overall, with seeds showing adaptations to promote dispersal, and plastid haplotypes often being widespread, seed dispersal appears to be an important mode of gene dispersal in *Senecio*. This contrasts with *Begonia* where seed dispersal is limiting (Chapter 2). However, species in *Senecio* section *Senecio* shed their pappus at maturity, so may not spread their seeds as far as other *Senecio* sections. Significant geographic structure of plastid haplotypes have been found in Moroccan *Senecio leucanthemifolius* Poiret var. *Casablancae* Alexander (Coleman & Abbott, 2003), supporting seed dispersal limitation in section *Senecio*.

The second factor proposed to promote speciation in *Begonia* is limited pollen dispersal. Most *Senecio* species are generalists in the pollinators they attract, like *Begonia*, with pollinators including: solitary bees, syrphid flies, and other dipterids (Schmitt, 1980; Comes & Abbott, 1998). Most of these pollinators are likely to travel short-distances between plants (Schmitt, 1980), promoting the geographic structuring of genetic diversity. This is supported by high $F_{ST}$ values for some *Senecio* species (Swiss *Senecio vulgaris*, $F_{ST} = 0.39$ with RAPD’s, Müller-Schärer & Fischer, 2001; $F_{ST} = 0.49$ with AFLP’s, Steinger et al., 2002; *S. leucanthemifolius*, $F_{ST}$
Interpopulation gene flow does not prevent drift or selection occurring in populations of *Senecio gallicus* and *S. glaucus* (Comes & Abbott, 1998; Comes & Abbott, 1999), and this has in turn lead to morphological differentiation between populations, as seen in *S. leucanthenemifolius* (Coleman & Abbott, 2003). Therefore, pollen dispersal between populations does appear to be low. Exceptions include those North American *Senecio* species that are pollinated by butterflies, which often travel between geographically isolated populations (Schmitt, 1980).

The third factor proposed to promote genetic differentiation between *Begonia* populations is frequent selfing. *Senecio* are more varied in their mating systems than *Begonia*, with some *Senecio* species being fully self-compatible and others self-incompatible. Ferrer & Good-Avila (2007) estimated that ~63% of Asteraceae species are self-incompatible, including many *Senecio* species. Self-incompatability in the Asteraceae is sporophytic, and pseudo-self incompatibility in *Senecio squalidus* and potentially other *Senecio* species is caused by unlinked modifiers of the self-incompatibility locus (Hiscock, 2000). Some colonizing *Senecio* species have a mutation at the *RAY* locus (Kim *et al.*, 2008) stopping them producing showy ray florets that attract pollinators, subsequently increasing their selfing rates (Marshall & Abbott, 1982). The impact of selfing on the genetic structure of *Senecio* populations is not clear, as most population genetic studies have used dominant markers and therefore have not calculated the inbreeding coefficient or inferred selfing rate. *Senecio vulgaris* is a commonly selfing species, and selfing may have contributed to the highly structured populations (described above). Other population genetic studies have been conducted on the largely outcrossing section *Senecio*, although Comes & Abbott (1998) reported population $F_{IS}$ values up to 0.21 at allozyme loci in *Senecio gallicus*, and this may be explained by non-random mating which may hence influence the genetic structure of populations (discussed above).

The final factor proposed to promote speciation in *Begonia* is strong reproductive isolating barriers. Hybridization in *Senecio* has been studied in a number of species; here two pairs of hybrids are considered where experimental crossing barriers and observations of hybrids have been made under natural conditions. *Senecio eboracensis* Abbott & Lowe is a recent allotetraploid with strong reproductive barriers to its parental progenitors *S. squalidus* and *S. vulgaris* (Lowe & Abbott, 2004). Artificial F1 hybrids between *S. squalidus* and *S. vulgaris* are...
relatively fertile, with pollen stainability and seed viability when open-pollinated greater than 60% of the parental means. Hybrids are extremely rare in the wild, likely due to a combination of pre- and post-zygotic barriers (Lowe & Abbott, 2004). A second experimental system for hybridization in *Senecio* is between closely related self incompatible *S. jacobaea* and *S. aquaticus* Hill. F1 crosses between *S. jacobaea* and *S. aquaticus* produce significantly more seeds, and the offspring are more vegetatively vigorous, than intraspecific *S. jacobaea* crosses. The ease of producing F1 hybrids is similar to F1 *Begonia heracleifolia* x *B. sericoneura* crosses. The F1 *Senecio* hybrids also perform well in both parental habitats (Kirk et al., 2005a). Seed production is then reduced in F2s and BCs (Kirk et al., 2005b). Hybrid swarms have admixed AFLP profiles, and likely include later-generation backcrosses (Kirk et al., 2004). Much like *Begonia*, the outcomes of hybridization in *Senecio* are dependent on the species being crossed. Many of the same barriers contribute to reproductive isolation between sympatric *Begonia* and *Senecio* species, such as the high frequency of self-fertilization and differences in flowering time (Chapter 5). There is also evidence that reproductive barriers are evolving within *Senecio* species, like seen in *Begonia heracleifolia* populations (Chapter 4). Seasonal cohorts within *S. vulgaris* flower at three different times of year, and may co-occur in sympathy, although only a small amount of genetic diversity is partitioned between cohorts (Haldimann et al., 2003).

Hybridization has also been shown to have evolutionary significant consequences in *Senecio*, generating new homoploid and alloplloid taxa, and through the introgression of important phenotypic traits. *Senecio squalidus* is a homoploid hybrid between *S. aethnensis* and *S. chrysanthemifolius*, which has formed within the last 300 years in a hybrid swarm in Sicily (James & Abbott, 2005). Since being introduced to the United Kingdom, *Senecio squalidus* has gone on to hybridize with *S. vulgaris* to give rise to new hybrid derivatives *S. vulgaris* var. *hibernicus* Syme. and the York radiate groundsel (Lowe & Abbott, 2000). Introgession of the *RAY* locus from *S. squalidus* into non-radiate *S. vulgaris* shows the importance of introgression for mediating rapid phenotypic changes (Kim et al., 2008). Past introgression has also been supported in Mediterranean *Senecio* species, with *S. flavus* susp. *breviflorus* Kaderit and *S. rupestris* Waldst. & Kit. found to acquire ITS sequences and plastid haplotypes from related species in sympathy (Comes & Abbott, 2001). Similarly, *Senecio massaicus* Maire has either
received genes through introgression, or may be a species of reticulate origin, as it contains genetic material from two divergent clades in a *Senecio* phylogenetic study (Pelser et al., 2012).

At the very broad scale, the evolutionary processes proposed to promote diversification in large genera Sanderson & Wojciechowski (1996) are present in both *Begonia* and *Senecio*, including their predominantly herbaceous growth form, frequent polyploidy and multiple transitions between ecological states. However, of the four characteristics proposed to promote divergence in *Begonia* some characteristics directly contrast *Senecio*, such as seed dispersal distances, while others are generally in agreement, such as low pollen dispersal. Geographic structuring of populations is an important precursor to allopatric speciation, and this occurs in different ways in the two genera, through low seed and pollen dispersal in *Begonia*, and in most *Senecio* species just through restricted pollen flow. This geographic structuring of genetic diversity may be promoted by selfing in *Begonia* and those *Senecio* species that are self-compatible.

### 7.1.2. Genetic resources for future studies

In addition to empirical findings for *Begonia*, the molecular markers developed for this thesis will be valuable for future genetic studies in *Begonia*. The nuclear microsatellites were designed from transcriptome sequence data, and unlike the species-specific markers currently available for *Begonia* (Hughes et al., 2002; Nakamura et al., 2012), they amplify over a wide phylogenetic scope (Chapter 3). These will facilitate population genetic studies in a range of *Begonia* species, allowing direct comparisons of levels of diversity and genetic differentiation using a common set of markers. The plastid markers will be valuable for tracing the parentage of experimental crosses, and as a seed specific marker for future population biology studies. These markers add to the suit of genetic resources that are currently being developed for *Begonia*, which include: transcriptome sequences of three species (*B. conchifolia*, *B. plebeja*, Asian *B. venusta* King, Brennan et al., submitted), plastid genomes of 16 species (Harrison, in prep.), BC genetic linkage maps (*B. plebeja* x *B. conchifolia*, Brennan et al., submitted; Chapter 5), and the whole genome sequence of *B. conchifolia* (Kidner, ubpubl. data). These genetic markers will be used to answer some of the outstanding questions about speciation in *Begonia* that have been raised by this study, which are outlined below.
7.2. Future questions

7.2.1. What is the colonization history of *B. heracleifolia* and *B. nelumbiifolia* in Central America?

The phylogeographic scenario in Chapter 2, where *B. heracleifolia* populations persisted in situ during the Pleistocene, whereas *B. nelumbiifolia* was restricted to refugia of unknown location, was based on samples only from south Mexico and west Guatemala. Genotyping samples from across the species ranges with the same markers would shed light on their colonization histories. This would be of particular interest if the intraspecific patterns could be related to geographic structuring of a phylogeny of Central American *Begonia* (see below).

7.2.2. How rapidly has speciation occurred in Central American *Begonia*?

The lack of species cohesion, and the rapid accumulation of reproductive barriers (Chapters 4 & 5), raises the hypothesis that Central American *Begonia* has undergone rapid speciation. A species-level phylogeny, calibrated with a relative dimension of time, would be the ideal way to address this hypothesis. The main difficulty would be obtaining a comprehensive sample of the large number of species that occur in remote localities (Thomas, 2010). If species representative of major *Begonia* clades were sequenced, new Bayesian analytical approaches could be applied, where diversification rates are estimated with incomplete species-level samples (Moore & Donoghue, 2009; Silvestro *et al.*, 2011; Wertheim & Sanderson, 2011; Drummond *et al.*, in press). The second barrier to this approach is the scarcity of fossil *Begonia* material (Stults & Axsmith, 2011) for calibrating a dated phylogeny. Alternative approaches include the use of a calibration point from a broader phylogenetic study (i.e. a secondary calibration point, e.g. Thomas *et al.*, 2011a), calibration from a geographic event of known age (Plana *et al.*, 2003), or using mutation rates from other species (Richardson *et al.*, 2001).
7.2.3. Which genes contribute to the early stages of speciation in *Begonia*?

The search for speciation genes is an increasingly popular research direction for evolutionary biology (Rieseberg & Blackman, 2010). *Begonia* is an ideal system to study speciation genes, as populations within species differ markedly in their levels of divergence, and may represent different stages in the speciation process (Chapter 4). Genome scans for pairwise $F_{ST}$ outliers between differentiated populations would be one way to identify genes under selection, and these would be candidate genes for adaptive differences (Scotti-Saintagne *et al*., 2004; Hohenlohe *et al*., 2010; Scascitelli *et al*., 2010). Such an experiment is increasingly feasible with next-generation sequencing technologies. Alternatively, if differentiated populations have clearly distinguishable phenotypes that may contribute to isolation, a QTL analysis would allow the genomic location of these genes to be mapped, and would be the first stage towards fine-mapping and more detailed genetic characterization.

7.2.4. Why is the sister genus and related families so species poor?

Hughes (2002) raised the question: why is the sister genus, *Hillebrandia*, and related families such as Datiscaceae, so species poor in comparison to *Begonia*? Explanations could include any combination of high extinction/low speciation in related groups, or low extinction/high speciation in *Begonia*. An analysis of diversification rates that accounts for extinction (reviewed in Purvis, 2008) could be applied to samples of *Begonia* and related groups to test this hypothesis. This approach would have to use DNA sequences that resolve both the backbone and species-level relationships with confidence, over a broad phylogenetic scope. Another approach, along the lines explored in this thesis, would be to investigate ongoing population processes that predispose these groups to low species diversity. These may include, high intraspecific gene flow preventing allopatric speciation, or low effective population sizes preventing adaptation to changing conditions, which may cause frequent local extinction.

7.2.5. Has polyploidy promoted diversification in *Begonia*?

Genome duplication is common in large plant genera (Sanderson & Wojciechowski, 1996), and *Begonia* species vary ten-fold in their chromosome numbers (Legro & Doorenbos, 1969, 1971,
1973; Dewitte et al., 2011). Recent transcriptome analysis suggests a whole genome duplication has occurred prior to diversification of Begonia, and polyploidy may have been an evolutionary stimulus in the group (Brennan et al., submitted). The species investigated here were diploid, so the role that polyploidy has played in diversification could not be assessed. This could be addressed by comparing rates of diversification between diploid and polyploid clades in a phylogeny. However, this approach cannot identify polyploidy as the actual cause of elevated divergence rates. To understand the functional role duplicated genes play, expression levels of homeologous loci could be compared between recent polyploids and their diploid progenitors (Buggs et al., 2011). This could use synthetic Begonia polyploids as a comparison (Dewitte et al., 2009, 2010).

7.2.6. Does hybridization play an evolutionary important role in Begonia?

The hybrid swarm analysis (Chapter 5) shows introgression in young hybrid swarms is limited, but this may not prevent important evolutionary outcomes over longer time periods. A study of past hybridization could test whether introgression, or hybrid speciation (Buerkle et al., 2000; Gross & Rieseberg, 2005; Chapman & Burke, 2007; Mallet, 2007), contributes to diversity in Begonia. Many techniques can be used to detect ancient hybrids, and these include comparing the topologies of species-level phylogenies constructed with unlinked markers, or comparative genome scans to identify shared alleles (reviewed in Twyford & Ennos, 2012, see Appendix 11).

7.2.7. Does ecological divergence promote speciation in Begonia?

The attributes of Begonia discussed in this thesis, particularly those which lead to geographic isolation of populations, go some way to explain why such species richness has accumulated in the genus. However, these attributes are found in many other taxa too, including genera that are not species-rich. While extant species richness reflects the stochastic nature of evolution, as well as the factors discussed above, many other factors are also likely to be involved in promoting divergence in Begonia. In particular, the ease at which colonizing species can adapt to new ecological conditions will be important. One explanation for rapid adaptation to new environments is pre-existing genetic variation (standing genetic variation), which will be a source for selection to act on in new habitats. The genetic signature of standing genetic variation
can be distinguished from new mutations using population genomic approaches (review in Barrett & Schluter, 2008). QTL analyses of ecologically important traits will also be useful in understanding the genetics of adaptation in *Begonia*.

### 7.3. Conclusion

*Begonia* is an emerging model system to study the mechanisms that promote diversification in large plant genera. Through genetic analysis, the population biology of *Begonia* has been shown to have profound consequences for speciation. Future genetic and genomic studies of *Begonia*, when integrated in a broader phylogenetic context, will continue to teach us much more about the fascinating and intricate evolutionary history of speciose plant genera.
8. REFERENCES


REFERENCES


consensus of ancient Middle American highland biogeography. *Journal of Biogeography*, 36, 88-103


REFERENCES


REFERENCES


REFERENCES

taxonomy and systematics of species rich taxa. (ed. by T. Hodkinson, Parnell, J. And Waldren, S), pp. 149-164. CRC Press.


REFERENCES


Edelaar, P.M., Burraco, P. & Gomez-Mestre, I. (2011) Comparisons between QST and FST — how wrong have we been? Molecular Ecology, 20, 4830-4839


REFERENCES


fragmentation and demographic expansion in Mexican cloud forests. *Molecular Phenylogenetics and Evolution*, 61, 603-615


Kisel, Y. & Barraclough, T.G. (2010) Speciation has a spetial scale that depends on levels of gene flow. The American Naturalist, 175, 316-334


REFERENCES


Marazzi, B. & Sanderson, M.J. (2010) Large-scale patterns of diversification in the widespread legume genus Senna and the evolutionary role of extrafloral nectaries. Evolution, 64, 3570-3592


Muller, H.J. (1942) Isolating mechanisms, evolution, and temperature. *Biological Symposia, 6*, 71-125


Nei, M. (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics, 89*, 583-590


Rambaut, A. Figtree: Tree figure drawing tool, version 1.2.2. available from: http://tree.bio.ed.ac.uk/software/figtree/


REFERENCES


REFERENCES


REFERENCES


incompatibilities, degree of dominance, and implications for Haldane’s rule. *Evolution*, **57**, 2580-2598


REFERENCES


REFERENCES


Wright, I.J. (1931) Evolution in Mendelian populations. *Genetics*, **16**


9. **APPENDICES**

The CD-ROM that accompanies the thesis includes the content listed in the table below. Full description of the data is included in each file.

<table>
<thead>
<tr>
<th>File name</th>
<th>Chapter</th>
<th>Contents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appendix 1 – <em>Begonia heracleifolia</em> plastid haplotypes.txt</td>
<td>2</td>
<td>Plastid microsatellite genotypes at 7 loci for 18 populations of <em>B. heracleifolia</em> (263 individuals)</td>
</tr>
<tr>
<td>Appendix 2 - <em>Begonia heracleifolia</em> nuclear microsatellite data.txt</td>
<td>4</td>
<td>Nuclear microsatellite genotypes at 9 loci for 13 populations of <em>B. heracleifolia</em> (306 individuals)</td>
</tr>
<tr>
<td>Appendix 3 - <em>Begonia nelumbiifolia</em> nuclear microsatellite data.txt</td>
<td>4</td>
<td>Nuclear microsatellite genotypes at 9 loci for 7 populations of <em>B. nelumbiifolia</em> (177 individuals)</td>
</tr>
<tr>
<td>Appendix 4 – Genetic clustering of <em>B. heracleifolia</em> and <em>B. nelumbiifolia</em> individuals.pdf</td>
<td>4</td>
<td>Genetic clustering results from InStruct and STRUCTURE for <em>B. heracleifolia</em> and <em>B. nelumbiifolia</em>, over all K values tested. See text for details</td>
</tr>
<tr>
<td>Appendix 5 - Nuclear microsatellite genotypes for hybrid swarm 1 (HS1).txt</td>
<td>5</td>
<td>Nuclear microsatellite genotypes at 12 loci for 61 individuals in a hybrid swarm between <em>B. heracleifolia</em> and <em>B. nelumbiifolia</em></td>
</tr>
<tr>
<td>Appendix 6 - Nuclear microsatellite genotypes for HS2.txt</td>
<td>5</td>
<td>Nuclear microsatellite genotypes at 12 loci for 71 individuals in a hybrid swarm between <em>B. heracleifolia</em> and <em>B. nelumbiifolia</em></td>
</tr>
<tr>
<td>Appendix 7 - Nuclear microsatellite genotypes for HS3.txt</td>
<td>5</td>
<td>Nuclear microsatellite genotypes at 12 loci for 61 individuals in a hybrid swarm between <em>B. heracleifolia</em> and <em>B. Nelumbiifolia</em></td>
</tr>
<tr>
<td>Appendix 8 - Nuclear microsatellite genotypes for HS4.txt</td>
<td>5</td>
<td>Nuclear microsatellite genotypes at 12 loci for 100 individuals in a hybrid swarm between <em>B. heracleifolia</em> and <em>B. sericoneura</em></td>
</tr>
<tr>
<td>Appendix 9 – PBC QTL.txt</td>
<td>6</td>
<td>Genetic map and trait variation for QTL analysis in the <em>Begonia plebeja</em> backcross mapping family. (Genotype data from Brennan <em>et al.</em>, submitted)</td>
</tr>
<tr>
<td>Appendix 10 – CBC QTL.txt</td>
<td>6</td>
<td>Genetic map and trait variation for QTL analysis in the <em>Begonia conchifolia</em> backcross mapping family. (Genotype data from Brennan <em>et al.</em>, submitted)</td>
</tr>
<tr>
<td>Appendix 11 - Twyford and Ennos (2012) Next-generation hybridization and introgression.pdf</td>
<td>n/a</td>
<td>Publication in Heredity during the PhD</td>
</tr>
</tbody>
</table>