Population structure and speciation in *Begonia* L.

A thesis submitted to the University of Glasgow for the degree of Doctor of Philosophy

Mark Hughes

Division of Environmental and Evolutionary Biology

June 2002

Declaration

I hereby declare that this thesis is composed of work carried out by myself unless otherwise acknowledged and cited and that this thesis is of my own composition. This research was carried out in the period of October 1998 to June 2002. This dissertation has not in whole or in part been previously presented for any other degree.

ABSTRACT

Biodiversity is unequally distributed between higher taxa; for example a small number of angiosperm genera contain the majority of angiosperm species. *Begonia* is one of the largest angiosperm genera with ca. 1400 species. Studies of such genera can give insights into the processess that cause diversification.

A number of features of the spatial distribution of biodiversity in *Begonia* suggest gene flow between populations is poor and has affected larger-scale patterns of diversity in the genus. These are (i) sporadic distribution of populations, which are usually restricted to a specific micro-habitat, (ii) a high degree of narrow endemism at the species level, (iii) widespread species being rare and also highly morphologically variable unless they show atypical adaptations that promote gene flow, and (iv) geographical restriction of monophyletic groups.

Restricted gene flow between populations allows them to diverge in response to weaker selection pressures than they would be able to respond to in the face of gene flow from other populations. In order to examine population structure (micro-evolution) in *Begonia* and its congruence to higher patterns of diversity (macro-evolution), nuclear microsatellite markers have been isolated and applied to two *Begonia* species, *B. socotrana* and *B. sutherlandii.*

Begonia socotrana is endemic to the Haggeher Mountains of the island of Socotra in the Indian Ocean, where it has a total range of less than 10 x 15 km. Population surveys have highlighted the need for its conservation status to be re-assessed, and it is proposed to reduce its status from 'endangered' to 'least concern'. Population genetic analyses using microsatellite data show a significant degree of population structure (R_{ST}= 0.081, P<0.01; θ=0.096, P<0.01) and significant isolation by distance, even over small spatial scales. The pattern of isolation by distance could be due to restricted gene flow, or the result of small scale vicariance events in the fragmented peaks of the Haggeher Mountains during climate change and resulting altitudinal migration.

Begonia sutherlandii is native to eastern and southern Africa, where it is restricted to shaded, moist banks in indigenous forest. A high degree of population structure was found (θ=0.482, P<0.001; R_{ST}=0.634, P<0.001), which

along with a high number of private alleles reflects the severe isolation of populations in a patchily distributed forest habitat. Population relationships appear to be strongly governed by the history and continuity of forest cover in the region.

The population genetic studies of *B. socotrana* and *B. sutherlandii* show a strong correlation of genetic variation with geography which reflect patterns seen at larger scales. The correlation of micro and macro evolutionary patterns is congruent with a hypothesis of restricted gene flow promoting speciation in *Begonia*.

ACKNOWLEDGEMENTS

Without the philanthropy of the late Mr and Mrs Macintyre in setting up the M.L. MacIntye *Begonia* Trusts, I would not have been able to spend three years studying *Begonia* evolution. I am extremely grateful to them and to the trustees, in particular Prof. Malcolm Wilkins and Dr. Chris Wheeler, for giving me this opportunity.

My supervisor, Pete Hollingsworth, deserves thanks (and no doubt a drink) for writing the initial project proposal and for guiding it through its subsequent changes, and for reading through this entire thesis, giving helpful comments and pointing out mildy amusing *double entendres* where relevant.

This research involved some interesting fieldwork, and I am grateful to Tony Miller and Miranda Morris for organising the expedition to the Socotra archipelago, which was no mean feat; Roger Hyam and Diccon Alexander for being amiable travelling companions; and the people of Socotra for their hospitality and for killing dozens of goats for us. My collecting trip in South Africa was assisted by many people, particularly Tracy McLellan who provided air-conditioned transport and company; Trevor Edwards and Christina Potgeiter who looked after me in Pietermaritsburg when I arrived *sans* luggage and slightly the worse for wear from in-flight tranquilisers; Chris Randle for driving to Tygerskloof; Tony Abbott for a splendid stay in Umtamvuna, and Dirk Bellstedt for being a mine of information and for getting me to some very interesting forest patches. I also received a great deal of help on localities, transport and accomodation from the KZN parks board, particularly Sharon Luow, Barry, Dennis Eckhart (contacts kindly given to me by Michael Moeller), and Neil Crouch. Olive Hilliard and Bill Burtt also provided helpful advice and maps during the planning of the trip.

The large amount of living material resulting from the fieldwork was carefully tended to in quarantine by Fiona Inches and Andrea Fowler, and maintained in excellent condition upon its release by Steve Scott, Neil Watherston and Fred Mobeck among others. Becky Govier patiently went through the mess that

was my collectors book and got most of it on BG BASE. The staff at Glasgow Botanic Garden also deserve a mention for maintaining the national *Begonia* collection, which made any trip through to Glasgow worthwhile.

This project also involved a great deal of labwork, which was made less of a chore due to the excellent lab facilities at RBGE, provided by Michelle Hollingsworth and Alex Ponge, who risked boredom and dermatitis in making up the dozens of polyacrylamide gels I got through. Joanne Russell (SCRI) and Jane Squirrell were both of great assistance in the tricky development of the microsatellite markers. Ruth Hollands patiently dealt with the dozens of orders for oligos and consumables. Thanks also to Jill Harrison for looking after me whilst I invaded her bench space at Edinburgh University, and for being elsewhere when I started a not inconsiderable fire in the PhD office. I would also like to thank everyone in the lab and herbarium at RBGE for making it such a pleasant place to work.

Assistance during the writing up phase came from many sources, including Rod Page and Richard Ennos for providing data analysis advice, and Neil Brummitt for heated debates on species-area relationships. All the RBGE library staff (in particular Graham Hardy) have been very helpful in finding material and sorting out copies and loans with spectacular efficiency. The latter stages of writing were carried out whilst also working on a post doctoral project on *Streptocarpus* at RBGE; Michael Moeller deserves a special mention for (i) employing me and (ii) for being a very understanding boss during a busy time.

This project would not have been half as enjoyable were it not for the two other *Begonia* students Laura Forrest and Vanessa Plana. Thanks especially to Vanessa for knowledge of African *Begonia*, and to Laura for continuous and welcome distraction by email, (including the odd academic one) and both for drinks and good humour. I hope the two current *Begonia* students, Will Goodall-Copestake and Sophie Neale, enjoy their projects as much as I have.

Thanks also, of course, to my family and RG for their continuing support.

PREFACE

This thesis is an investigation of population structure in *Begonia*, and its relevance to evolution and speciation in the genus. There are four introductory chapters. The first is a general discussion of large scale patterns of biodiversity, and possible causes for the unequal distribution of biodiversity between higher taxa. The second is an account of the ecology and systematics of the Begoniaceae, and aspects of *Begonia* biodiversity that may be of relevance to population structure and speciation processes. The third chapter is a discussion of the effects of gene flow on population differentation and speciation, and the fourth is a review of methods of analysing and interpreting population genetic data obtained from microsatellite markers.

The next chunk of the thesis consists of five parts that are written as papers intended for publication. As each is a complete paper in its own right, this inevitably involves a small amount of repetition. The formatting for the submitted papers in terms of tables, figures and references is according to the format of the journal to which they were submitted. The first paper is a monograph of two species of *Begonia* from the Socotra archipelago, and an assessment of their conservation status. This has been accepted for publication by the *Edinburgh Journal of Botany*. Papers two and three are technical papers which describe the isolation and development of nuclear microsatellite markers from two *Begonia* species, *B. socotrana* and *B. sutherlandii*. These have been accepted for publication by *Molecular Ecology Notes*. The fourth paper is an investigation of population differentiation and conservation genetics of *Begonia socotrana*. This has been submitted to *Biological Conservation*. The fifth and final paper is an account of the population structure of *Begonia sutherlandii* in South Africa, and its relevance to broader scale patterns of biodiversity and speciation in the genus as a whole. This paper will be submitted for publication to *Molecular Ecology*. Each paper has its own self-contained reference list, and a bibliography for the rest of the chapters is presented at the end of the thesis.

In the light of the findings of the preceeding chapters, the last chapter (5) discusses the relevance of population structure and evolution in the Datiscaceae (the sister family to the Begoniaceae) and in the angiosperms as a whole.

TABLE OF CONTENTS

CHAPTER 1. THE UNEQUAL DISTRIBUTION OF BIODIVERSITY

CHAPTER 2. THE BEGONIACEAE C. AGARDH.

CHAPTER 3. GENE FLOW AND SPECIATION

CHAPTER 4. MICROSATELLITES AND DATA ANALYSIS

CHAPTER 5. POPULATION STRUCTURE AND SPECIATION: PERSPECTIVES FROM OTHER SPECIES

LIST OF TABLES & FIGURES

CHAPTERS 1-5

PAPERS 1-5

Paper 1

Paper 2

Paper 3

Paper 4

Paper 5

1.1. Some taxa are bigger than others

Lower taxa are unequally distributed between higher taxa. For example, there are many monotypic genera, and far fewer larger ones. Within flowering plants, 50% of species diversity is contained in only 550 (out of over 12,000) genera. This pattern is famously illustrated by Willis' 'hollow curve' (Figure 1.1; Willis, 1922).

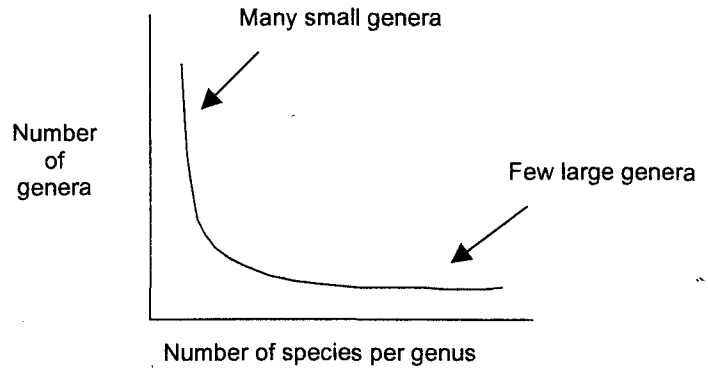

Figure 1.1. The hollow curve of genus size distribution.

The hollow curve is exhibited at all levels in the taxonomic hierarchy (Dial and Marzluff, 1989); for example plotting the number of angiosperm genera per family produces the same pattern (Clayton, 1974). This pervasive phenomenon (Figure 1.2) begs the question: What causes some taxa to become much larger than others? Before tackling this question, though, it is necessary to look at possible artefactual causes of the hollow curve, and assess how likely it is to be due to natural processess.

1.2. Is the hollow curve real or imagined?

The possibility has been raised that the hollow curve has artifactual origins. Walters (1961, 1986) thinks the distribution of plant species among genera of

different sizes is a human construct, caused largely by a Euro-centric view of taxonomy. He suggests that large genera are those first created by Linnaeus and added to by 'chaining', as many taxonomists tend to add to existing taxa rather than create new ones, whilst monotypic genera are those created later for exotic plants which do not fit easily with their European allies. Raikow (1986; oft quoted but rarely agreed with) is of a similar mind, and states that the most likely reason for there being so many kinds of passerine birds compared to a larger number of smaller bird orders is an 'artefact of their classificatory history'. These arguments stem from the subjectivity involved in delimiting higher taxa.

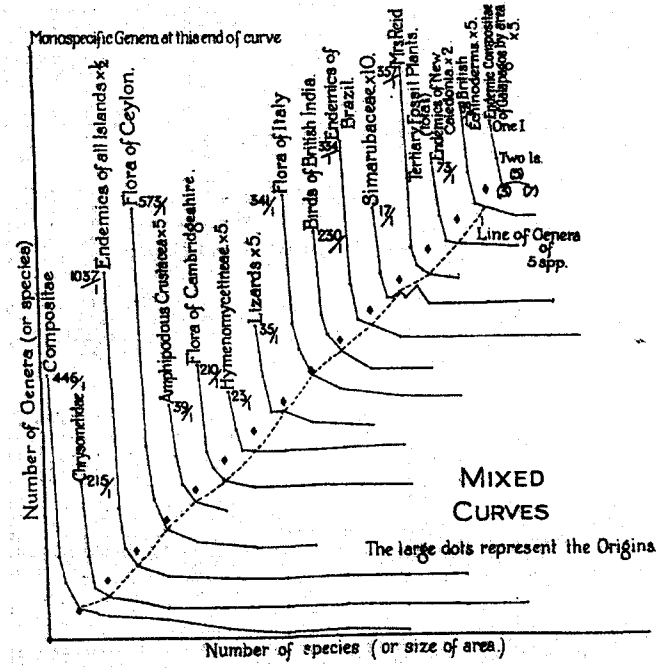

Figure 1.2. Taken from Willis' classic 1922 work (p. 237), this figure shows a range of hollow curves derived from generic sizes of plants and animals at a variety of scales, from local to national.

There is no accepted definition of the genus or family, and although many of the higher taxa defined are likely to have a biological reality (i.e. monophyly), ranking them could be said to be rather arbitrary and hence using them as comparable units questionable. However, this does not completely weaken the case against the pattern being a product of natural causes, as it still appears in within family analyses where the generic concept is likely to be more comparable. Also, given the number of genera, the signal to noise ratio in the data must be quite considerable. Cronk (1989) gives a slightly more moderate view than Walters or

Raikow, and suggests that the pattern of generic size distribution is due to real and artefactual components in approximately equal measure. The pervasiveness of the pattern in the distribution of biodiversity between taxa is a strong case for it being a product of natural phenomena, and it seems an artefactual cause is unlikely. Its reality has been broadly accepted by many authors, from Willis who stated (1922, p. 185); "It is idle to suggest that further work will alter the form of this curve" to Minelli (1991) who regards the pattern as a result of evolutionary processess, and who deemphasises the role of taxonomic bias.

1.3. Causes of the pattern

Although the pattern of unequal distribution of lower taxa between higher taxa is obvious, the cause is not (assuming that the pattern has at least some basis in reality). There are two main schools of thought on why some taxa are more diverse then others; one sees it as the result of stochastic events in a complex system, the other as the result of key innovations and adaptive events.

1.3.1. Stochastic and neutral models

The 'hollow curve' becomes a straight line if presented as a log-log plot. Such a graph would be an expression of the equation $\log N(s) = -\tau \log s$, where s would be the size of a genus and $N(s)$ would be the number of genera with that size; the straight line of gradient τ is the result of the relationship being a power law, as one quantity N can be expressed as some power of another quantity s. Such power law relationships can be found over and over in the natural world and in man made complex systems, such as the distribution of earthquake size ($N(s)$ would be the number of earthquakes of energy s) or fjord lengths (where $N(s)$ would be the number of fjords of length s), or the size of avalanches in a sand pile prodded by a reseärcher at the University of Michigan (Bak, 1997). The distribution of genus size follows the pattern shown by complex or fractal systems, but does it share a common cause? Bak (1997) suggests that there is a common principle underlying all

phenomena which can be expressed as a straight line log-log plot, which is called self organised criticality (SOC). Self-organised critical systems are poised in a state of perpetual instability, and a disproportionate amount of the change that takes place in such a system is due to a few large events rather than gradual change. Although not explicitly applied to the hollow curve distribution by Bak, SOC is thought to be behind much of the complex behaviour exhibited by natural systems, and evolutionary models based on it produce patterns of punctuated equilibrium matching those seen in nature (Bak and Sneppen, 1993). Kauffman (1995, p. 129) suggests that ecosystems and possibly the whole biosphere may exist in a critical state, where evolutionary novelty would come in large bursts and small trickles that would conform to a power law distribution.

Neutral models in which speciation events follow a simple set of underlying principles rather than a more deterministic niche based theory can also produce patterns of biodiversity remarkably similar to those found in the real world (Hubbell, 2001; Solbrig, 1994). These models also predict an underlying fractal nature to biodiversity. This idea is supported by Green (1991) who highlights the strong parallels between systems showing non-linear dynamics and the evolution of life, both of which involve changes over time (i.e. evolution) that are inherently historical, being non reversible and non repeatable. Burlando (1990) also suggests the hollow curve is the result of an underlying fractal nature to biodiversity; the self similarity of the curve at all levels and scales is indicative of its fractal nature (Mandelbrot, 1982) and supports this view.

1.3.2. Key innovations and adaptation

Is there a reason for the existence of disproportionately large taxa or clades other than them being random blips in a complex system? Just because the distribution of biodiversity follows a pattern that suggests it is caused by stochastic events does not mean these events themselves lack an underlying biological reality, and many authors have looked for links between taxon size or diversification rate and underlying biological attributes. The appearance of key innovations or other biological properties in a lineage that may increase diversification rate is likely to be

a random process, so the two schools of thought are perhaps not as exclusive of one another as it might seem.

Willis (1922, p. 193) suggests that the size of genera is accounted for by their age, holding the theory that monotypic and small genera are young and are derived from larger, older genera, and have not yet had time to expand their range or speciate. Cronk (1989) holds the opposite view, that monotypic genera are relics of possibly once larger genera, and that contemporary large genera are the result of recent 'blooming' of a particular lineage. The latter view certainly seems to apply to some cases in the light of phylogenetic investigations, for example Richardson et al. (2001) found that the massive diversity in the legume genus *Inga* had arisen during the late Pleistocene. This 'diversity is recent' view intuitively makes more sense than Willis' argument in that it has an explicit role for extinction in shaping contemporary diversity, rather than having old taxa as ever-growing entities. An empirical investigation into angiosperm diversification based on molecular phylogenetic data (Magallon and Sanderson, 2001) supports Cronk's view, with basal lineages in the angiosperms being significantly species poorer than many of the highly nested clades (crown-groups). The Lamiales, Gentianales, Solanales, Apiales, Asterales and Cyperales were found to have significantly higher diversification rates than other angiosperm clades. The first five groups are core Asterids, and the Cyperales are a highly nested monocot clade; all the orders are relatively young in terms of the age of the angiosperms as a whole, being less than 50 million years old.

What would cause a taxon to enter a 'bloom' phase? Several authors have tried to examine the biology underlying the production of large higher taxa, by comparing their biological attributes with taxon size or diversification rate.

Marzluff and Dial (1991) examined traits that could be responsible for the domination of a taxon by one or few large subtaxa, such as age of first reproduction, longevity and fecundity. They found the age of first reproduction was the life history trait most strongly correlated with high taxonomic diversity within some vertebrate taxa, but found no significant correlations for their limited sampling of vascular plants.

Eriksson and Bremer (1992) calculated the diversification rate, R, for a number of angiosperm families based on the contemporary family size and the first appearance of the family in the fossil record, and correlated this value with the pollination system, dispersal mode and life form that predominates in each family. They found a significant correlation of R with both animal pollination and the herbaceous habit, with insect pollinated families having a significantly higher R than wind pollinated ones, and herbaceous families having a significantly higher R than woody familes. No correlation was found between R and dispersal syndrome.

Rickleffs and Renner (1994) re-examined the data gathered by Eriksson and Bremer, and concluded that the first appearance in the fossil record of a plant family has little bearing on its age, and regard the contemporary species richness of a plant family as a more reliable indicator of its propensity to diversify. They found that species richness is associated with dispersal mode, growth form and pollination mode, in descending order of statistical influence. These single factor correlations were weak when compared to the effects of families having varied dispersal syndromes, growth forms and pollination systems. Rickleffs and Renner conclude that the capacity to diversify morphologically and physiologically has been a major factor responsible for high rates of species proliferation in flowering plants.

Eriksson and Bremer (1991) looked for dispersal mode correlations of species richness in genera of Rubiaceae, and split their data set into herbaceous genera and woody genera. They found abiotically dispersed genera of herbs within the Rubiaceae were tended to be speciose, while biotically dispersed herbaceous genera were relatively depauperate. Conversely, biotically dispersed shrub genera tended to be highly speciose compared to the abiotically dispersed shrub genera. Tiffney and Mazer (1995) further examined this hypothesis that a correlation of dispersal mode with diversification in angiosperms depended on whether growth form was taken into account. Their analysis showed a significant correlation between dispersal syndrome and diversification if growth form was taken into consideration. Herbaceous families with abiotic dispersal had higher taxonomic richness than herbaceous families with biotic dispersal, whilst the converse was true for woody families, where higher taxonomic richness was correlated with biotic dispersal; both of these correlations were significant at the 95% level.

Smith (2001) also looked for a correlation with dispersal mode and species richness, and compared ecologically similar sister clades with differing dispersal modes (small fleshy fruits dispersed by birds, and dry fruits). He found that in tropical understory plants, clades with fleshy fruit were significantly more diverse than clades with dry fruit.

A robust examination of angiosperm diversification was carried out by Barraclough and Savolainen (2001), which used the molecular phylogeny of angiosperms (Soltis et al., 2000) to identify sister familes which by definition diverged at the same time from a common ancestor. This approach is free from the subjectivity involved in defining higher taxa. They found that diversification rates in angiosperms correlated with the neutral substitution rate in both plastid and nuclear genes. However, rates of non-synonymous substitution did not correlate with species numbers, even though they correlate with the neutral substitution rate.

To sum up, the following correlations with increased diversification rate or higher taxon size were found to be significant to some degree: Insect pollination (Rickleffs and Renner, 1994; Eriksson and Bremer, 1992), the herbaceous habit (Eriksson and Bremer, 1992), varied dispersal mode within the taxon (Rickleffs and Renner, 1994), varied habit within the taxon (Rickleffs and Renner, 1994), abiotic dispersal (in herbaceous taxa; Tiffney and Mazer, 1995; Eriksson and Bremer, 1991), biotic dispersal (in woody taxa; Tiffney and Mazer, 1995; Eriksson and Bremer, 1991; in forest understory taxa; Smith, 2001) and the rate of neutral substitution (Barraclough and Savolainen, 2001). What could be the effect of these attributes on increased diversification?

Many authors found correlations with factors affecting gene flow (dispersal and pollination syndromes). The effect of dispersal syndrome on angiosperm diversification could be due to one of two effects. Firstly, an increase in the overall diversification rate could be caused by a reduction in the extinction rate mediated by animal dispersal, which would allow species to disperse to new areas of suitable habitat and so reduce the risk of extinction through range expansion. This hypothesis

has been presented by Tiffney and Mazer (1995) to explain the correlation they found between vertebrate dispersal and the diversity of woody angiosperms.

Conversely, it has been suggested that the correlations found between the various pollination and dispersal syndromes and diversification could be caused by them producing population structures condusive to speciation. Eriksson and Bremer (1992), Crepet (1984) and Smith (2001) support this view, and propose that local seed and pollen movement by animal vectors can produce small genetically isolated populations in which evolution can proceed rapidly.

Barraclough and Savolainen (2001) put forward two main hypotheses to explain the correlations they found. Firstly, they suggest that rates of phyletic change and speciation are increased by smaller generation times or with small population sizes. Secondly, it is argued that increased mutation rates could drive diversification, and so be a cause rather than an effect, with higher mutation rates increasing the rate of divergence of populations and contributing to the development of hybrid incompatability and eventual speciation. Variation of morphology and physiology within a taxon is perhaps the easiest correlation to explain, as the production of novelty is one of the key aspects of speciation and this would permit radiation into new habitats and environments.

Although many hypotheses have been suggested to explain the correlations, most of the authors suggest that the causes of radiations are likely to be complex and multi-factorial. A problem with interpreting many gene-flow related correlations is that the categories used to classify dispersal and pollination syndromes can be said to be somewhat meaningless, as both animal and abiotic dispersal and pollination cover a wide range of overlapping gene flow capabilities; the abiotic dispersal category covers plants with very short range passive dispersal mechanisms to plants with extremely long range wind dispersal mechanisms. This is the case in Smith (2001), in which fleshy fruited clades were compared to 'dry fruited' clades. The fleshy fruited clades were quite uniform in dispersal mode, but the dry fruited clades possessed a range of dispersal mechanisms ranging from hooked fruit to light seeds with papery wings. The different effects of abiotic dispersal on diversification depending on growth form (Erksson and Bremer, 1991, 1992; Tiffney and Mazer, 1995) present a

contradictory picture, as it correlates with increased diversification of herbaceous lineages, but not of woody ones. It is possible to conceive that abiotically dispersed woody plants (e.g. wind dispersed trees) can disperse further than abiotically dispersed herbs (passive dispersal), so the 'population structure conducive to speciation' hypothesis may apply in both cases, with abiotic dispersal hindering the speciation of tree forms but aiding the speciation of herbaceous forms. However, without a better understanding of the dispersal capability of different genera in terms of some directly comparable unit, it is difficult to do anything but speculate.

1.4. *Begonia* as a large genus

The largest angiosperm genera are considered to be *Euphorbia* L., *Piper* L., and *Carex* L., containing about 2000 species each, followed by *Astragalus* L. (1750), *Solanum* L. (1700) and *Psychotria* L. (800-1500) (figures taken from Mabberly, 1997). *Begonia* contains ca. 1400 species (Doorenbos et al. 1998), and so is comfortably placed within the largest ten angiosperm genera. Given the light of phylogenetic information (Wagstaff and Dawson, 2000; Forrest, 2000) it does seem that *Begonia* is a morphologically distinct clade which has a far higher diversification rate than its neighbouring morphologically distinct clades, whether or not one considers genera to be directly comparable units. Figure 1.3 shows the position of *Begonia* in a phylogeny of the Cucurbitales *sensu* the Angiosperm Phylogeny Group (1998). Although there is no bootstrap support for the basal nodes in this tree, which are unresolved in the strict consensus, this topology reflects the accepted sister relationship of the Datiscaceae to the Begoniaceae (Swensen et al. 1998), and places the Cucurbitaceae as sister to the rest of the order which reflects the appearance of the family in the fossil record before Tetramelaceae, Coriaricaeae and Corynocarpaceae (Wagstaff and Dawson, 2000). Given this topology, it does seem that Begoniaceae has a far higher diversification rate than other families in the Cucurbitales, with the possible exception of the Cucurbitaceae (I make family level comparisons here because they reflect monophyletic groups with obvious morphological synapomorphies, although it is perhaps rather a moot point as *Begonia*/Begoniaceae comes out as the largest clade independent of whether one looks at families or genera).

The huge difference in species number between the Begoniaceae and the Datiscaceae is consistent with the hypothesis of insect pollination driven diversification as mentioned previously. Most *Begonia* and *Hillebrandia* are thought to be insect pollinated (mainly by bees, Agren and Schemske, 1991), whilst *Datisca* is wind pollinated (Liston et al., 1990). The presence of insect pollination in *Begonia* could lead to a higher degree of population structure than the wind pollinated *Datisca*, which may have influenced the difference in speciation rates between the two lineages. Although both families are abiotically dispersed, there is likely to be some difference in dispersal ability between them, as *Begonia* occurs in sheltered forest environments whilst *Datisca* occurs in more open riparian habitats. I present here the hypothesis that diversification in the *Begonia* lineage has been aided by a high degree of population isolation, caused by localised pollen transfer by insects and passive seed dispersal mechanisms in a sheltered forest environment. Although it is likely that most radiations are complex in origin and may have several causal factors (Tiffney and Mazer, 1995; Magallon and Sanderson, 2001), and *Begonia* is probably no exception, several aspects of the ecology and biogeography of *Begonia* are congruent with this hypothesis, and a review of these is presented in chapter 2.

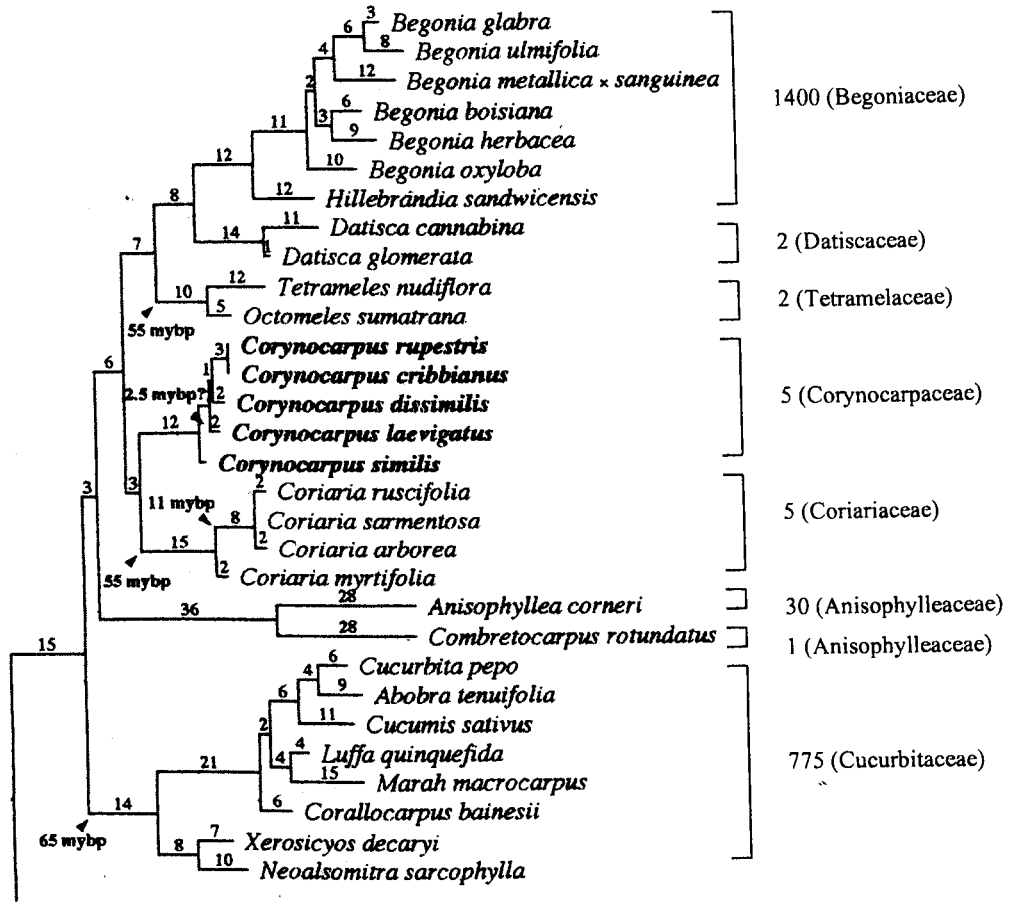

Figure 1.3. Phylogeny of the Cucurbitales inferred from *rbcL* sequences (one of 38 equally most parsimonious trees; from Wagstaff and Dawson, 2000). The number of species in each clade is indicated on the right. This topology (with respect to Begoniaceae, Datiscaceae, Tetramelaceae and Cucurbitaceae) is congruent with an anlaysis using 18S and *rbcL* by Swensen et al. (1998).

1.4.2. Why isn't *Hillebrandia* a large genus?

Perhaps one of the most puzzling things about the Begoniacaeae is why *Hillebrandia* is monotypic. It appears to be ecologically and florally quite similar to *Begonia*, and is endemic to the Hawaiian archipelago which is the home of countless fascinating radiations, yet today exists as a single lineage. The *Hillebrandia* lineage may never have radiated in the Hawaiian islands, although another possibility is that *Hillebrandia* presently represents a taxon in decline, being the lone survivor of a once diverse clade. The relictuality of *Hillebrandia* is evident from the Begoniaceae phylogeny of Forrest (2000), as it is sister to *Begonia* and is likely to have been on

the Hawaiian archipelago for at least 55 million years. It is currently only found on the older islands of the archipelago and probably survived by 'island hopping'. If taxa go through 'bloom and bust' cycles as suggested by Cronk (1989), *Hillebrandia* may the last dying ember of a once spectacular botanical firework.

1.5. Summary.

Angiosperm biodiversity is unequally distributed among taxa (or clades). The observed pattern of the distribution of biodiversity between taxa is fractal in nature, and is possibly an emergent property of the complex nature of evolution.

Investigation of unequal diversification between individual taxa suggests that ecological correlates such as gene flow capability may be one of the causal agents.

It is conceivable that insect pollination, local seed dispersal and patchy population distributions have been important in the diversification of the *Begonia* lineage.

2.1. Introduction to the Begoniaceae

This chapter is an overview of the biogoegraphy, ecology and reproductive biology of the genus *Begonia*, and much of the information presented is background to the hypothesis that localised gene flow has influenced speciation.

2.1.1. Taxonomy and Distribution

The Begoniaceae contains ca. 1400 species in three genera: *Begonia* L., *Symbegonia* Warb. and *Hillebrandia* Oliver. The vast bulk of the species belong to *Begonia,* which has a pan tropical distribution (Figure 2.1) and is classified into 63 currently accepted sections (Doorenbos et al. 1998; all authorities of section names subsequently mentioned are as referenced therein). Each of the sections are limited to one continent and many of them were originally described as separate genera in Klotsch's pioneering treatment of the family (Klotsch, 1854). The name *Begonia* was first published in 1700 by J.P Tournefort, who named the six species he described from the French Antilles in honour of Michel Begon, the Intendent of the islands (Barkley, 1968).

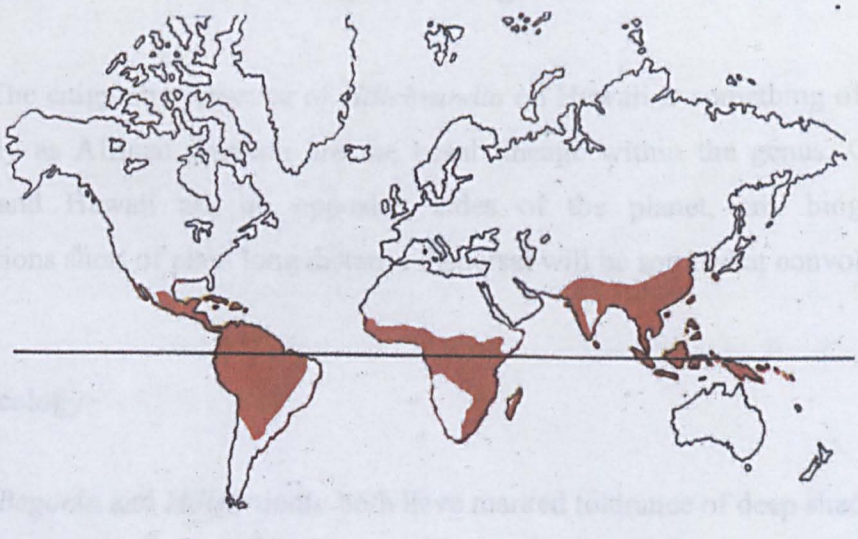

Figure. 2.1. The distribution of *Begonia* (modified from Heywood, 1993).

Of the remaining two genera, *Hillebrandia* is a monotypic Hawaiian endemic, and *Symbegonia* has ca. 12 species in New Guinea. The latter is best regarded as a section of *Begonia*, as it is nested in section *Petermannia* according to a phylogenetic analysis by Forrest (2000) and has no molecular or morphological characteristics that warrant recognition at the genus level over other equally distinct sections of *Begonia*.

The Begoniaceae belongs in the Cucurbitales (*sensu* APG 1998), and has the Datiscaceae Bercht. and J. Presl. as its sister group (Swensen, 1998), which is a ditypic family of perennial herbs with a disjunct distribution in California and central Asia. The affinity of the Begoniaceae with the Cucurbitaceae Juss. ex DC and the Datiscaceae has long been recognised, although the inclusion of previously taxonomically isolated families such as the Corynocarpaceae Engl., Coriariaceae Dumort and Anisophyllaceae Ridley in the Cucurbitales was unexpected.

Begonia follows the distribution of wet tropical forest and tropical montane forest, with Asia and Central and South America having the bulk of the species (ca. 600 spp. each), with diversity hotspots including the equatorial Andes and Malesia. Although Africa (including Madagascar) has less species diversity in terms of number, at ca. 150 species, it has the greatest morphological diversity and the oldest lineages of the genus according to the phylogenetic work of Forrest (2000) and Plana (2002), with most diversity occuring in the Congo basin.

The enigmatic presence of *Hillebrandia* on Hawaii is something of a puzzle, especially as African *Begonia* are the basal lineage within the genus. Given that Africa and Hawaii are on opposing sides of the planet, any biogeographic explanations short of plain long distance dispersal will be somewhat convoluted.

2.1.2. Ecology

Begonia and *Hillebrandia* both have marked tolerance of deep shade, and this can be interpreted as one of the key innovations of the Begoniaceae. The family is also strongly hydrophilous, and throughout the tropics *Begonia* are found in wet and

shaded environments such as stream banks and vertical seep faces under a forest canopy, or the mist zone around forest waterfalls.

Begonia leaves show several features that can be interpreted as adaptations for coping with shade (Sosef, 1994, p.38; Lee and Graham, 1986). Bullate leaves (e.g. *Begonia bullata* and several other species in section *Loasibegonia* and *Scutobegonia*) may be an adaptation to catch light scattered at different angles by the overhead forest canopy, and may also encourage surface water to drain away from the leaf surface after rain and prevent it scattering light. An anthocyanin rich lower leaf surface as seen in *Begonia brevirimosa* (and many other species of *Petermannia*; it is also seen to some degree in most *Begonia* sections) also reflects photosynthetically useful red light back up into the leaf. On the upper leaf surface, some *Begonia* species show a blue iridescence (e.g. *B. johnstonii*) or satiny/glistening sheen (e.g. *B. sutherlandii*). The glistening is caused by lens shaped surfaces of the leaf epidermal cells, which focus light onto the stacked chloroplasts below; the function of the blue iridescence is not known, but it is exhibited by other shade tolerant species (e.g. *Selaginella* sp.) and may reduce reflection of photosynthetically useful light.

The asymmetric leaf is one of the most characteristic qualities of the genus, but its function remains rather enigmatic and no explanation is offered in the literature. However, it seems likely that it serves to create an efficient leaf mosaic, as alternate leaves have the enlarged lobe on the opposite side of the leaf. This has the effect of displacing the point of attachment of the petiole to the side and hence nearer to the main stem, allowing the formation of a leaf mosaic which would require much larger investment in petioles if the point of attachment was at the 'top end' of the leaf (Figure 2.2). Some other shade tolerant plants show this adaptation (e.g. Gesnericeae sp.), but it is not common in plants, perhaps due to developmental constraints which guard against asymmetry.

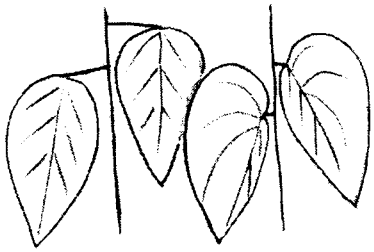

Figure 2.2. A longer petiole is needed to give the same leaf mosaic in symmetrical (left) than asymmetrical leaves (right).

2.1.3. Life cycle

Annual lifecycles are comparatively rare in *Begonia* (limited largely to section *Doratometra* (8 species) and a few species in section *Begonia* and *Rostrobegonia*) with most species being evergreen through several seasons or having perenniating organs such as rhizomes, tubers, tubercles and bulbils (Badcock, 1998). Nineteen sections from all continents contain a greater or lesser number of tuberous species. Tubercles (corm like structures found in the leaf axils) are found in a small number of species, including *B. sutherlandii* and *B. wollastonii* from east Africa, *B. grandis*, *B. notata*, *B. pedunculosa* and *B. gemmipara* from Asia and several species from the Mexican section *Quadriperigonia*. Bulbils are unique to section *Peltaugustia* (*B. socotrana* and *B. samhaensis*), and consist of a compressed shoot in which the leaves are reduced to small fleshy scales, the whole being encased in papery bracts.

2.1.4. Pollination biology

The vast majority of Begonia are monoecious, with a few species being dioecious (e.g. *B. roxburghii* and some other members of sect. *Sphenanthera*, also a few species in sections *Mezierea* and *Tetraphila*; Doorenbos et al. 1998). Monoecious plants show a range of sex separation; many are protandrous, for example *B. glabra* and many of the associated species in the '*Pritzelia*' clade (Forrest, 2000), and some are protogynous, for example section *Petermannia*. Male and female flowers are commonly found on the same inflorescence (e.g. *B. glabra*) or more rarely on separate ones (e.g. *B. herbacea*). The number of male and female

flowers per inflorescence varies enormously between species, from one (e.g. female inflorescence of *B. herbacea*) to over 1000 (*B. luxurians*).

The majority of *Begonia* species are pollinated by deceit (Figure 2.3). The male flowers offer pollen as a reward to insects, whilst the female flowers are rewardless. Vogel (1978) states that in the case of the female flower the "voluminous yellow stylodia mimic the androecium of the male flowers and thus release gathering movements [in visiting insects] which affect pollination."

Deceit pollination in *Begonia* has been confirmed by Agren and Schemske (1991). Their study showed a marked preference of pollinating insects for male flowers (the bee *Trigona grandipennis* was the most frequent visitor to the species under study, *B. involucrata*). Male inflorescences received an average of 7.2 times more visits than female inflorescences, with the visiting insect spending about ten times longer on male flowers than on female ones. *Begonia* pollination has been further investigated by Schemske et al. (1996) and Le Corff et al. (1998). All three studies highlight (i) pollen is the only reward, (ii) visitors show a marked preference for male flowers, (iii) male flower visits last longer than visits to females, and (iv) seed set is pollen limited.

Figure 2.3. Male (upper) and female (lower) flowers of *B. socotrana*, with the yellow stylodia of the female flower mimicking the pollen bearing anthers of the male.

Schemske and Agren (1995) demonstrated a preference of pollinators for larger flowers, by exposing them to different sized artificial flowers modelled on *B. involucrata*. However, the female flowers of the species are actually slightly smaller than the males, so there must be some other unknown ecological or genetic factor which limits flower size. Le Corff et al. (1998) demonstrated the selection pressure for mimicry in female flowers in *B. tonduzii*, whose female and male flowers have six and two tepals respectively. A comparable species, *B. urophylla*, has male and female flowers which have the same number of tepals (two) and achieves a significantly higher level of seed set than *B. tonduzii*. This highlights the selection pressure for mimicry in the female flowers (Le Corff et al., 1998).

The majority of *Begonia* species have flat open faced flowers which are pollinated by small bees and other generalist pollinators. Some sections have developed tubular flowers, which are possibly adaptations to bird pollination; e.g the Andean section *Casparya* and the genus *Symbegonia* from New Guinea. Bird pollinated species represent pollination by double deceit, as neither male nor female flowers offer a nectar reward and pollen is not taken as food by bird pollinators. The exception to this is *B. ferruginea* (sect. *Casparya*) which has been shown to produce nectar in female flowers (Vogel, 1998), the only known case of nectar production in the Begoniaceae.

Wind may play a part in the pollination of some *Begonia*, particularly in South American species with large inflorescences such as the lianescent *B. glabra*, which produces large heads of male flowers that shed pollen when shaken, and possesses female flowers with small tepals and very prominent stigmas.

The pollination syndromes of some species remain unknown, such as *B. longirostris* (sect. *Semibegoniella*, an Andean section allied to sect. *Casparya*) which has campanulate male flowers and open faced female flowers, and *B. maurandiae* (sect. *Gobenia*) which has showy male flowers with prominent anthers and inconspicuous female flowers with very reduced tepals. In cases like these where the morphologies of the male and female flowers are so different, pollination is perhaps unlikely to be facilitated by straightforward deceit.

2.1.5. Dispersal biology

2.1.5.1. *Fruit morphology and function*

Tropical forest canopies offer shelter to the ground layer vegetation beneath, and sub-canopy wind dispersed plants are much rarer in closed forest than in more open habitats such as savanna (Hovestadt et al., 1999). This suggests that wind dispersal in forest conditions is not very effective. Killeen et al. (1998) found that true wind dispersal was limited to lianescent angiosperms in a survey of tropical semideciduous forest, with understory shrubs being largely zoochorous and the herbaceous layer being largely autochorous. The most common fruit type in *Begonia* is the three winged dehiscent capsule (Figure 2.4). This is alleged to disperse the seeds in an anemoballistic fashion (de Lange, 1998), although in the sheltered forested environments favoured by most *Begonia* species, wind dispersal is probably not very effective (Burt-Utley, 1985) and some have argued that *Begonia* seeds are passively dispersed, and do not travel far from the parent plant (Agren and Schemske, 1993; de Lange and Bouman, 1999; Matolweni, 2000). Some *Begonia* species show variation away from the 'standard' tri-alate fruit; one of the wings may be considerably enlarged, such as shown by members of sect. *Platycentrum*, in which the two smaller wings form a cup in the strongly recurved mature fruit and which allows splash dispersal of seeds by raindrops. The wings in species of sect. *Casparya* are reduced into three relatively strong hooks (Fig 2.5), forming a rattle-bur type of fruit which releases seeds when knocked by passing animals (de Lange, 1988).

Fleshy fruits are found mainly in Africa, in sections *Baccabegonia*, *Mezierea*, *Squamibegonia* and *Tetraphila*. The fruits may be either dehiscent, often with brightly coloured placentas (e.g. sect. *Tetraphila*) or indehiscent (e.g. *B. oxlyoba*). Although the fruits do not have a strong flavour or high sugar content, they are though to be vertebrate dispersed. Given the fragile nature of *Begonia* seeds, it is probable that the seeds are dispersed ectozoochorously rather than internally and this could be facilitated by having the seeds loosely attached to the surface of the fleshy placentas rather than embedded in it. However, de Lange and Bouman (1991) suggest endozoochory as a possibility, which they say could be facilitated by the

larger seeds with thicker exotesta found in some fleshy fruited species. They also suggest the some species in sect. *Tetraphila* may be myrmecochorous (ant-dispersed), as the seeds have an aril which could function as an elaiosome. The harvesting of seeds by ants has been observed in cultivated specimens of one of the species from this section, *B. rhopalocarpa* (de Wilde, 2002). The only other section to possess fleshy fruit is *Spenanthera* from Asia which has indehiscent green or white fruits lacking markedly fleshy placenatae. Interestingly all the fleshy fruited species also belong to sections with a high level of dioecy, which is a correlation noted among angiosperms in general by Muenchow (1987) and Bawa (1987) among others. Causes of this correlation are little more than speculation, but one possibility is a disproportionate increase in female reproductive fitness with increased female reproductive effort in fleshy-fruited animal-dispersed species, due to the effectiveness of animal dispersal in carrying seeds to suitable habitats away from the parent (Bawa, 1987).

Figure 2.4. *B. dregei* possesses the three winged dehiscent fruit typical of the majority of *Begonia* species.

Figure 2.5. The fruit of *B. urticae* possesses three horns and an extended apical column (scanned from Balls 1939; herbarium specimen, E.).

The African sections *Loasibegonia* and *Scutobegonia* are unusual in that they possess non-fleshy fruit that are also indehiscent. The seeds are released slowly when the fruit wall becomes papery and disintegrates (sect. *Loasibegonia*) or in the case of sect. *Scutobegonia*, the fruit walls remain slightly juicy and release the seeds upon rotting. Seed dispersal is thought to be through contact with animals, with the seeds sticking to mud on their legs, or via water running along the soil surface after heavy

rains (Sosef, 1994). The fruits are often hidden under the leaves, and sometimes the pedicels recurve towards the substrate, placing the fruits at soil level. The species in these two sections contain a high number of narrow endemics, and their poor long distance dispersal capability has led to them being used as likely indicators of Pleistocene forest refuges in Africa (Sosef, 1994).

2.1.5.2. *Seed morphology*

Begonia seeds are small, ranging in size from 220 μm long (*B. iucunda*) to 2240 μm long (*B. elbowensis*), although are more commonly around 300-600 μm long. They are characterised by the presence of longitudinally stretched cells (known as collar cells) which form a ring around the apical end of the seed, splitting open upon germination.

A 'typical' *Begonia* seed is shown in Figure 2.6, and the vast majority of *Begonia* species conform to this type, with some variation in size and ornamentation. In contrast, a small number of species show seed morphologies which can be interpreted as adaptations for more efficient wind dispersal, such as the African epiphyte *B. thomeana* (Figure2.7)

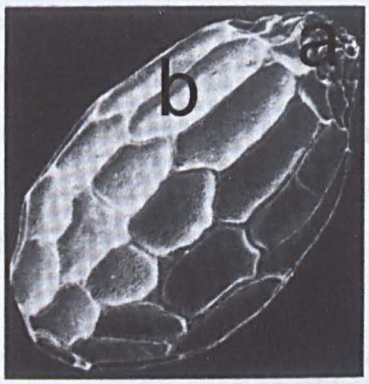

Figure 2.6. A typical *Begonia* seed (*B. palmeri*). (a) operculum; (b) collar cells. From de Lange and Bouman, 1999.

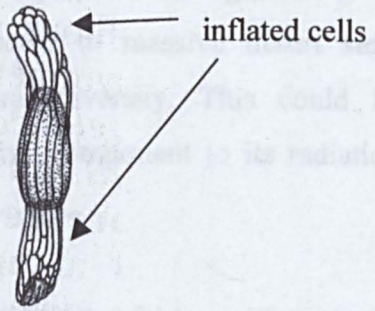

Figure 2.7. Seed of *B. thomeana*, showing the inflated cells at each end. From Doorenbos et al. 1998.

whose seeds have air filled balloon cells at each end. Similar adaptations are shown by members of the neotropical sections *Wagenaria*, *Solananthera*, *Rossmannia* and

Gobenia, (de Lange, 1999) the vast majority of whose species are epiphytes or lianas.

2.2. The distribution of *Begonia* biodiversity and reasons why limited gene flow may be important in *Begonia* speciation

Several aspects of the distribution of *Begonia* biodiversity correlate with the suggestion that capacity for gene flow over more than local scales is poor for most *Begonia* species, and that this may be one of the factors leading to increased diversification of the genus through allopatric speciation.

2.2.1. Species richness

Begonia has over 1400 species. Are there any obvious adaptive features which could account for this? It is relatively easy to speculate on factors affecting speciation in some large groups of angiosperms, depending on how their biodiversity is organised. For example the Orchidaceae has a great deal of floral diversity, whilst having comparatively little vegetative diversity. This could indicate that pollinator driven speciation has been important. The large genus *Euphorbia* has a great range of vegetative forms ranging from temperate forest herbs to massive desert stem succulents, whilst having comparatively little floral diversity. This could be interpreted as indicating a large physiological adaptive component to its radiation (Gill, 1989).

For such a large genus, *Begonia* could be said to be fairly uniform in its ecology. There are a few aberrant species and sections, such as section *Gireoudea* which has some members that are adapted to quite xeric and sunlit habitats in Mexico, and *B. princeae* (sect. *Augustia*) which grows on the shaded side of termite mounds in East Africa. However, the majority of species are shade loving succulent hydrophilus herbs, to be found in a similar 'begonia habitat' throughout the tropics; seep faces, stream banks and waterfall splash zones. Looking at the collection details of *Begonia* herbarium specimens reveals a striking similarity of the habitats they

were collected from across the tropics. There is certainly ecological differentiation between species, but it appears to be of a subtle nature, with many species having apparently very narrow niches.

No *Begonia* species are recorded as having formed specific plant-pollinator relationships. Although pollinator observations are rare, the open faced flowers produced by most species would not seem to offer much opportunity for this, instead attracting a variety of generalist pollinators.

2.2.2. Geographical monophyly.

Monophyletic groups in *Begonia* show strongly restricted geographical distributions, suggesting repeated adjacent allopatric speciation events. Prior to the ITS based phylogeny of the Begoniaceae produced by Forrest (2000), there was little information on the relationships of *Begonia* species and sections, but the phylogeny shows that in some cases geographical proximity is a better indicator of species relationships than morphology. This is exemplified by the monophyly both of the endemic Malagasy taxa and of the South Africa taxa, which were previously thought to have complex relationships with each other and with central African species. This pattern is also shown at much larger scales; American and Asian *Begonia* each form a monophyletic group, albeit on short and unsupported branches (Figure 2.8).

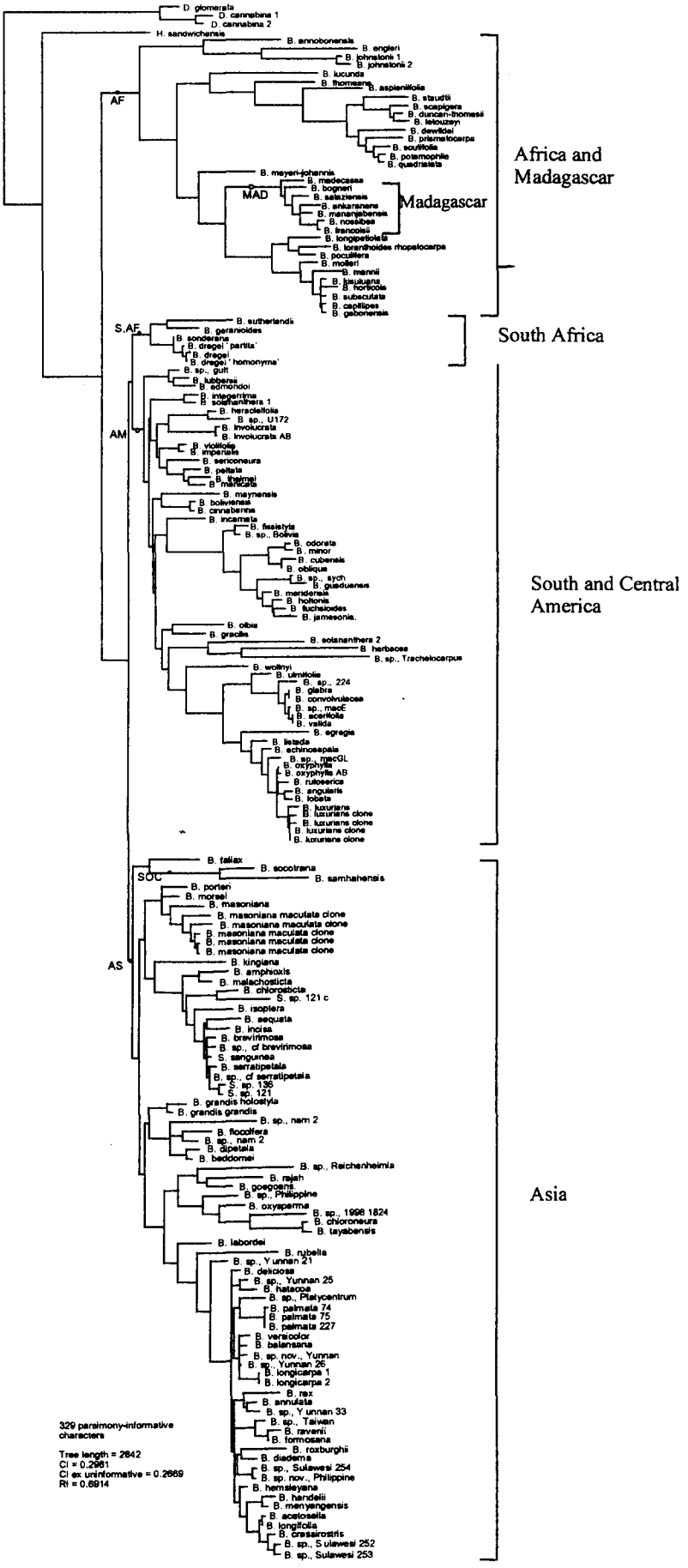

Figure 2.8. Phylogeny of Begoniaceae based on ITS sequence data. One of 10,000 equally parsimonious trees.

2.2.3. Narrow endemism

The level of narrow range endemism of *Begonia* is very high, suggesting species' ranges may be dispersal limited. The extent of narrow endemism in the genus is evident from floristic accounts, and has been noted by many researchers (e.g. Tebbitt, 1997, p.117; Sosef , 1994, p. 116). Kiew's account of the limestone *Begonia* of Sabah (Kiew, 2001) lists 14 species, only one of which is widespread in the study region, with 9 of them being endemic to single hills. In order to get a more objective measure of species endemism in *Begonia* I have plotted log-log species area graphs for the *Begonia*, *Piper* and *Lepanthes* of Ecuador (fig 2.9), obtained by noting the cumulative number of species in increasing nested areas within the country (using data from Jorgensen and Leon-Yanez, 1999). *Lepanthes* is a large (600 spp.) neotropical orchid genus with a high level of endemism in Ecuador (218 out of 288 are endemic). It has been hypothesised that small population size and poor gene flow between dispersed populations has been an important factor in generating species diversity in this and similar orchid genera (Tremblay and Ackerman, 2001). *Piper* (Piperaceae) is a genus of lianas and shrubs with bat dispersed fruits; it has 215 species in Ecuador, with 75 of these being endemic. *Begonia* has 59 species in Ecuador, of which 33 are endemic.

Hubbell (1997) states that the slope (z) and area (x-axis) intercept of nested log-log species area plots contain information on the spatial distribution of biodiversity. A low value of z indicates low β-diversity (regional diversity), whilst a high value indicates high β-diversity, i.e., the slope reflects the amount of diversity at the regional level. The maximum value for the slope of the log-log plot is unity, which means that a doubling of the study area would lead to a doubling of the number of species encountered. The area intercept reflects the level of α-diversity (local diversity), with a low value indicating high α-diversity (the area in which you only find one species is small) and a high value indicating low α-diversity (the area in which you only find one species is larger). Hence, a low value of z and a small area intercept would mean most of the diversity is present at a local level, with most of the species being widespread and having a high level of sympatry (as in Figure 2.10 (b)). A high value of z and a high area intercept would mean local (α) diversity

is poor, but regional (β) diversity is high with most of the species being local endemics (as in fig 2.10 (a)).

The slopes for the *Begonia* and *Lepanthes* graphs are both unity, whilst the slope of the *Piper* graph is 0.64, and this is congruent with poorly dispersed species having higher levels of endemism and smaller ranges than animal dispersed species. It should be pointed out that the level of endemism for neither *Begonia* nor *Lepanthes* in Ecuador is 100 %, and so we would not expect a slope quite as high as 1 for the log-log graph; there may be some skew to the study resulting from having started the nested species count in one of the drier provinces of Ecuador. This skew would have a similar effect in all three genera as they all inhabit a broadly similar habitat of tropical evergreen forest, and the comparison between the three genera used here will hold. The result reflects field observations of *Begonia*, where one to a few species may be encountered in any one local forest area, but where regional species numbers are much higher.

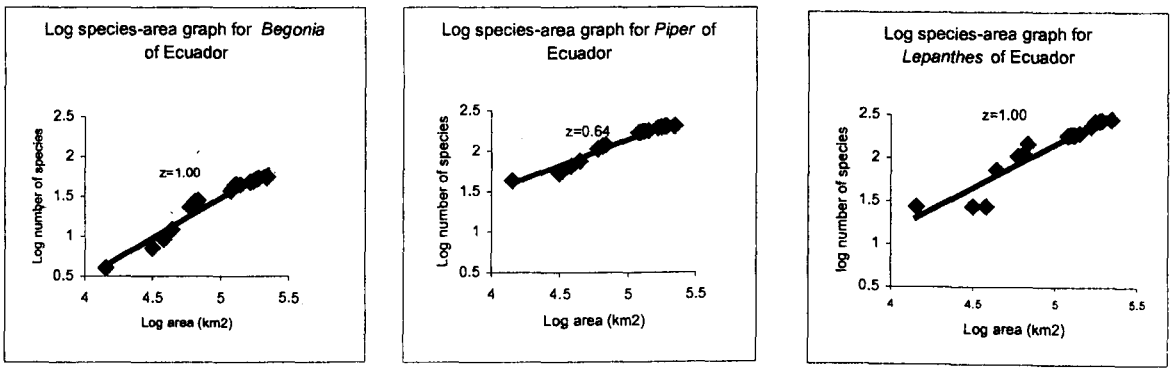

Figure 2.9. Log species area graphs for the *Begonia*, *Piper* and *Lepanthes* species of Ecuador.

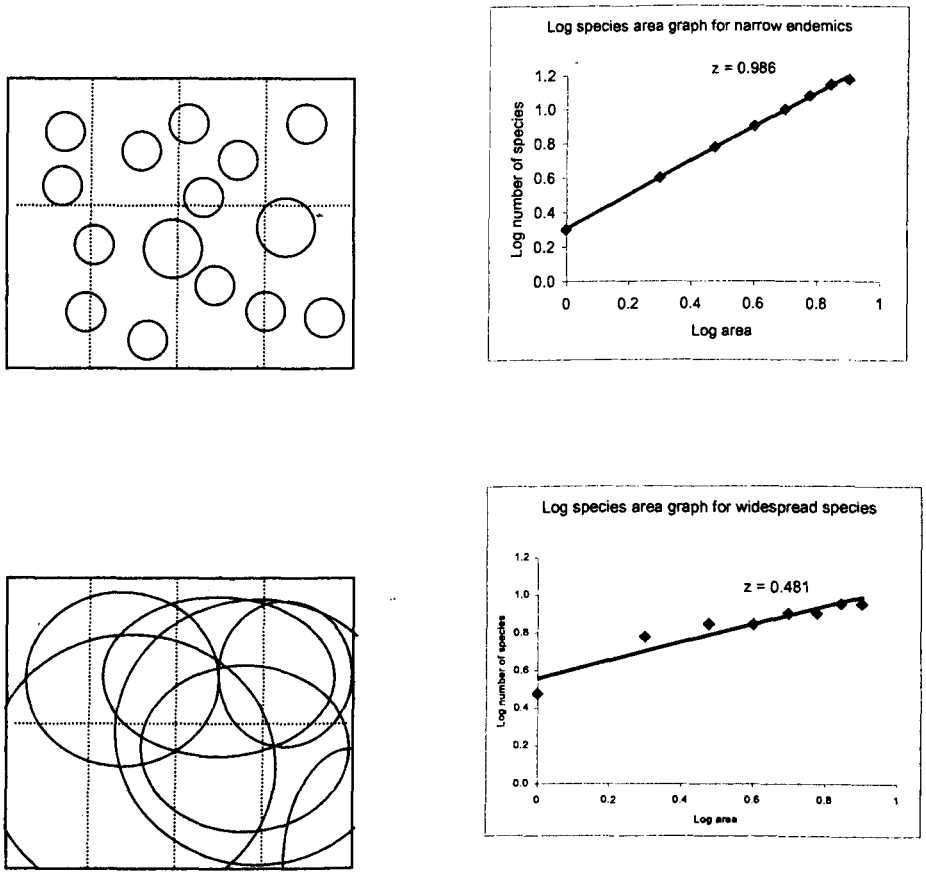

Figure 2.10. Schematic examples of non overlapping (a) and overlapping (b) species distributions.

A further interpretation from log-log species graphs at regional scales can be drawn according to Hubbell's unified theory of biogeography (Hubbell, 1997), who states that the intercepts and slopes of such graphs contain potentially useful information on speciation and dispersal rates. At large scales, the 'standing wave of biodiversity' is defined by rates of speciation, dispersal and extinction rather than simple relative species abundance which would affect biodiversity at more local scales. As one increases the area of study, one will encounter more and more species whose distribution in the study area is dispersal limited, and this factor becomes increasingly important in determining the rate of addition of new species with increased area. Low values for the slope indicate good dispersers, whilst high values such as those found for *Begonia* indicate species dispersing relatively poorly across the continental landscape relative to speciation rates.

2.2.4. Rarity and variability of widespread species

Widespread species are rare in *Begonia*, with most species being narrow range endemics. The relatively small number of species that are widespread tend to show marked morphological variation, suggesting low intra-specific gene flow over moderate to large scales. For example, *B. sutherlandii* has a wide afromontane distribution from Kenya southwards to the northern Transkei in South Africa. Now considered a single variable species, it was split by Irmscher in his 1961 monograph of sections *Augustia* and *Rostrobegonia* into 6 species, 5 varieties and 2 forms. *B. urticae*, which occurs along the Andes from Costa Rica to Peru, is another example of a morphologically highly variable species, as noted in the the Flora of Ecuador account (Smith and Wasshausen, 1986) which also gives 18 synonyms.

There are some examples of widespread *Begonia* species which do not fit this pattern of being differentiated across their range, but these species show atypical adaptations which are likely to increase their capability for gene flow over longer distances. *B. glabra* is the most widespread *Begonia* species in South America, occuring from Mexico to Peru and eastwards into Brazil and the West Indies. It is a liana, which means its winged fruits can be exposed to air currents in the canopy (Figure 2.12), compared to terrestrial species which grow in the still conditions of the forest floor, also the seeds possess extended cells at the distal end of the seed which

may be inflated (de Lange and Bouman, 1999), thus further helping wind dispersal. The protogynous infloresences produced by *B. glabra* are quite large (Figure 2.12) and the species is likely to be pollinated to some extent by wind, as noted in 2.1.4.

B. oxyloba is the most widespread *Begonia* species in Africa, extending from Liberia across tropical Africa to Tanzania and into Madagascar. This terrestrial species has fleshy fruit with orange, melon scented placentae which could be dispersed by a variety of small vertebrates and insects, and produces a very dry pollen freely released from the anthers upon slight movment (pers. obs.), so wind may play a part in its pollination. Other fleshy fruited species in Africa tend to have larger ranges than their counterparts with dry fruits. Figure 2.11 compares the distributions of the range sizes in the mainland species of sections *Tetraphila*, *Mezeirea* and *Squamibegonia* (fleshy fruited) and *Loasibegonia* and *Scutobegonia* (dry fruited).

B. thomeana with its inflated seeds (Figure 2.7) and extended peduncle shows another correlation with dispersal-ability and range, as it occurs both on Sao Tome in the Gulf of Guinea and on the adjacent African mainland. Excell (1973) lists 15 named *Begonia* species for the Gulf of Guinea islands, 11 of which produce fleshy fruits, suggesting that their origins on the islands stem from bird mediated dispersal. Apart from *B. thomeana*, there are three more dry fruited species which co-occur on both mainland and island situations; the remaining five species that show this pattern being fleshy fruited. The dry fruited species are *B. sessilifolia* (*Filicibegonia*), *B. annobonensis* (*Sexalaria*) and *B. prismatocarpa* (*Loasibegonia*); none of these show adaptations for dispersal which would set them aside from other members of their sections and their presence on the islands is probably due to chance long distance dispersal either by wind or animals. *B. prismatocarpa* exists as three subspecies, one of which (subsp. *prismatocarpa*) is endemic to Bioko (Fernando Po), which could indicate there is little of no gene flow between it and its mainland counterparts *B. prismatocarpa* subsp. *delobata* and subsp *petrea*.

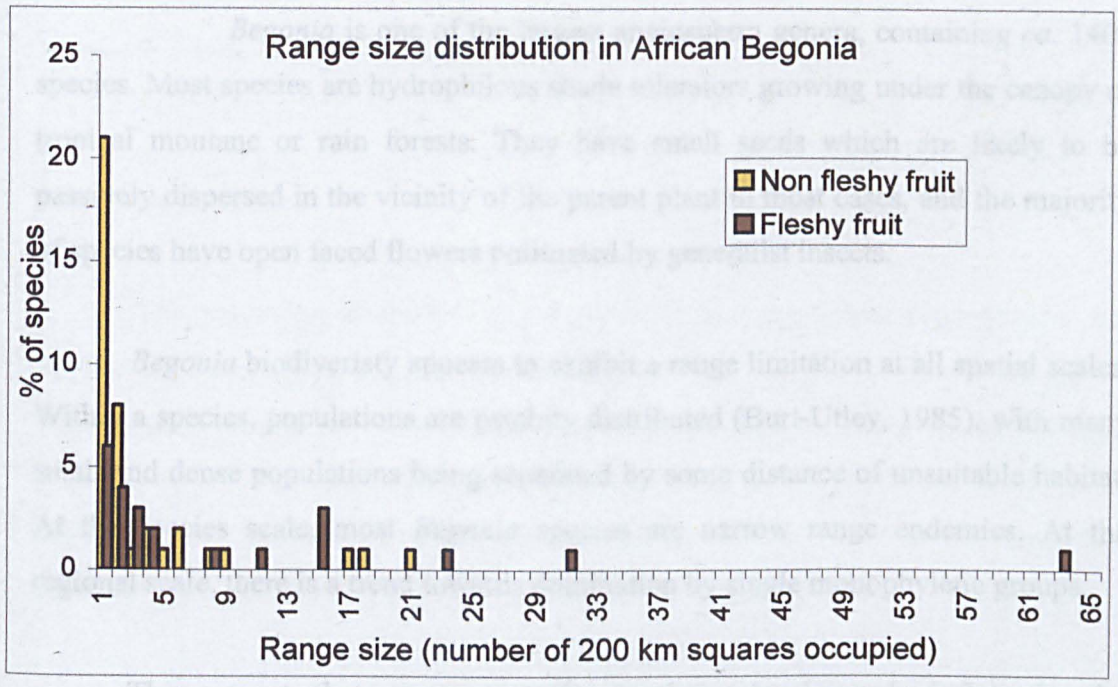

Figure 2.11. Range sizes in mainland African *Begonia*. 74% of fleshy fruited species have range sizes over 200 km^2, whilst only 48% of non fleshy fruited species have a range size over 200 km^2. The largest range size is a fleshy fruited species, *B. oxyloba*, which occupies 63 200 km squares.

Figure 2.12. Fleshy fruit of *B. oxyloba* (left); winged fruits of *B. glabra* (centre); protandrous inflorescence of *B. glabra* in male phase (right).

2.3. Summary

Begonia is one of the largest angiosperm genera, containing *ca.* 1400 species. Most species are hydrophilous shade tolerators growing under the canopy of tropical montane or rain forests. They have small seeds which are likely to be passively dispersed in the vicinity of the parent plant in most cases, and the majority of species have open faced flowers pollinated by generalist insects.

Begonia biodiveristy appears to exhibit a range limitation at all spatial scales. Within a species, populations are patchily distributed (Burt-Utley, 1985), with many small and dense populations being separated by some distance of unsuitable habitat. At the species scale, most *Begonia* species are narrow range endemics. At the regional scale, there is a trend towards domination by single monophyletic groups.

This suggests that processess at the population level may be influencing the spatial distribution of biodiversity at the species level and above, with the lack of connectivity between dispersed populations leading to allopatric speciation occuring over relatively short distances. Little is known about the amount of gene flow between *Begonia* populations and over what distances it is effective. Only a single study has been carried out on *Begonia* population genetics (Matolweni et al., 2000). This allozyme based study of the South African coastal forest endemics *B. dregei* and *B. homonyma* showed that over 90% of genetic variation was partitioned between populations.

In order to speculate in a more informed manner on the role of limited gene flow on speciation patterns in the genus, it is necessary to measure gene flow in *Begonia* species from local scales to regional scales and to find out how connected populations are. Thus the goal of this thesis is to explore whether macroevolutionary patterns in *Begonia* are correlated with population level microevolutionary processess.

3.1 Introduction

The preceding chapter is a summary of biogeographic, phylogenetic and ecological factors that suggest restricted gene flow may have been one of the contributing factors to diversification in *Begonia*. Before going on to look for population level patterns that are congruent with this hypothesis, it is worth addressing the background theory on the influence of gene flow at the population level upon divergence and eventual speciation.

This chapter is a discussion on this aspect of speciation, and although it is very tempting to agree with Bush (1994) who said discussing species concepts before discussing speciation is putting the cart before the horse, I feel it is necessary to give at least some indication of what one means by 'a species'. As the importance of gene flow in the speciation process is the item under discussion, it will suffice for the purposes of this chapter to examine the relationship between gene flow and species unity, rather than disappearing into the semantic murk of a full-blown examination of the species problem. The following discussion, as do the vast majority of any discussions on species, will assume species have been delimited using a taxonomists species concept. This avoids any circularity which would be inherent using some concepts; if one defines species using a cohesion concept akin to Templeton (1989), in which species are united by cohesionary forces such as gene flow, then *ipso facto* species are united by gene flow and the argument becomes empty. In discussing the properties of species, it is helpful to have the reference to the real world that is the taxonomist's species concept.

3.2 Within-species gene flow

There are two schools of thought on the maintenance of species unity. Erlich & Raven (1969) argue that species are not held together by gene flow, and are therefore not fundamentally different from higher taxa in that they are passive products of evolution rather than the largest units that participate in it. They suggest it is selection that is the primary cohesive and disruptive force in evolution, and that "at least in many cases, gene flow is of little or no importance in maintaining many of the phenetic units we call species". They give examples of species with disjunct or widespread distributions over which they assume gene flow to be ineffective, and suggest uniform selection is the reason for morphological cohesion in the species in question. The reason for this view is their belief that "distances of from 50 feet to a few miles may effectively isolate [plant] populations, and there is no evidence of longer range gene flow". Although they cite many examples to support their hypothesis, not all of these are based on empirical studies and in some cases they are purely anecdotal (e.g. the fact that "reef fishes are often remarkably similar throughout tropical seas" (and therefore the product of uniform selection) is stated without a reference).

Reiseberg and Burke (2001) think the conclusion that species are passive end products of evolution is premature, largely because of many of the studies cited by Erlich and Raven grossly underestimate the amount of gene flow between populations. Estimates based on pollinator behaviour are problematic as they do not take into account pollen carry over and are likely to miss occasional long distance events. Direct observations of seed dispersal were also found to be likely to provide underestimates of between population dispersal. Reiseberg and Burke also disagree with the conclusions drawn by Erlich and Raven from widespread species or species with disjunct distributions. They suggest it is possible that species having large ranges or disjunctions could be the result of quite recent events, and that in these cases there may have been insufficient time for populations to diverge. Population ranges are likely to be dynamic in most species, and it seems plausible that populations presently out of contact may not have been in the recent past and may not be so in the near future. Also, some barriers may not be as impermeable as Erlich and Raven believe, especially if one takes into account the effects of sporadic longer

distance dispersal events. A further consequence of the Erlich and Raven view of species unity is that it would render the study of reproductive isolation mechanisms an unfortunate scientific detour, as populations would be free to evolve in splendid isolation without them. Also, populations usually represent a subsample of the genetic variation within a species. This difference in genetic makeup means that populations may respond differently to the same selection pressure due to the difference in available alleles, and so populations not experiencing gene flow are likely to diverge from one another under identical selection pressures (Levin, 2000, p. 63).

Reiseberg and Burke conclude that gene flow between populations of most or all plant species is sufficient to allow the efficient spread of favourable alleles, providing the fitness difference conferred by the alleles is fairly large ($s>0.05$). Such strongly favourable alleles may be the most likely agents of collective evolution. Gene flow may not be high enough, however, to prevent divergence at loci where selection is weak or neutral, and Reiseberg and Burke put forward the idea that we should view species as groups of populations which are collectively evolving at some loci, but which may be diverging at others.

Levin (2000) also takes a more reasoned view of the relationship between gene flow and species unity. He states that answering if species are wedded by contemporary gene flow depends on the continuity of the species distribution, the distance between the populations and the pollen / seed dispersal curves of the species in question.

Despite Bush (1975) stating that Erlich and Ravens paper is "part of a growing body of evidence that suggests the effect of gene flow on differentiation may be small", it seems that this is now a minority view, with the consensus being that gene flow is important in maintaining species unity.

3.3 Speciation concepts

The following will deal with speciation in functionally diploid outcrossing organisms, rather than with the complexities of evolution in selfing or apomictic lineages. It has been suggested that selfing may be an evolutionary dead end (Stebbins, 1957) and that selfing lineages continually go extinct with new lineages being formed from outcrossing progenitors. An analysis of phylogenetic studies of mating system evolution by Takebayashi and Morrell (2001) found most were in accordance with this hypothesis, so it seems likely that selfing lineages are primarily the products of biodiversity rather than producers of it.

When a new species forms, through whatever process, it must possess some kind of mechanism that will prevent it from interbreeding with its progenitor or sister should they occur in the same geographical area, either during or after the speciation process. In plants, reproductive isolation mechanisms can be classified as:

I. Prezygotic
- Ethological – pollinators may discriminate between the two forms
- Genetic – failure of pollen and ovule to produce a zygote

II. Postzygotic
- Zygotic inviability – the embryo may not survive
- Hybrid inviability – the hybrid may be weak or unfit
- Hybrid sterility – the hybrid may fail to set seed either through selfing or crossing with other hybrids or the parents

The way reproductive isolation becomes established may differ depending the mode of speciation.

3.3.1 Sympatric speciation

In sympatric speciation, new species arise and diverge whilst occupying the same range as their progenitor lineage. This requires the evolution of mechanisms

which restrict or prevent gene flow between the diverging entities. Such mechanisms may arise rapidly, in the case of novel gross mutations or polyploid speciation, or more gradually when one or more polymorphic loci are subjected to disruptive selection.

3.3.2 Geographic speciation

Most models of primary speciation have a geographic component, in which the differences in selection, drift and gene flow due to geography are the factors causing divergence. It is generally thought that most speciation involves some degree of allopatry (Mayr, 1963; Lynch, 1989), and this fact has been obvious to observers even in pre-Darwinian times, as eloquently stated by Leopold von Buch (1825; translated in Mayr, 1963, p. 483): "The individuals of a genus strike out over continents, move to far-distance places, form varieties (on account of the differences of the localities, of the food, and the soil), which owing to their segregation cannot interbreed with other varieties and thus be returned to the original main type. Finally these varieties become constant and turn into separate species. Later they may again reach the range of the other varieties which have changed in a like manner, and the two will no longer cross and thus they behave as two very different species".

There are several different models of allopatric or geographic speciation, which vary in terms of the size of the speciating entity and the amount of contact it has with its progenitor or sister.

3.3.2.1 Parapatric speciation

In this model, sister species evolve while adapting to contiguous spatially segregated habitats across a narrow contact zone. Reproductive isolation mechanisms arise simultaneously with the exploitation of a new habitat (Bush, 1975, 1994).

3.3.2.2 Vicariant speciation

This is the classic model of geographic speciation, and relatively simple in its conception. A parent species becomes divided by some kind of barrier, such as a

mountain range, desert or large river. In this model, the resulting two populations are assumed to be large enough for drift and inbreeding to not be a major factor in the speciation process, with divergence being caused mainly by differences in selection.

3.3.2.3 Peripheral isolate speciation

In this model, the speciating entity is a small population at the edge of the parent species range. Bush (1975) and Mayr (1963) regard peripheral isolates as important in the evolution of new species. Founder effect speciation is an extreme form of the peripheral isolate model, in which the founding population goes through a fairly rapid and extreme genetic change leading to reproductive isolation. Forms of founder effect speciation have been proposed by Mayr (1954), Templeton (1980) and Grant (1971) among others. The genetic changes occurring during founder effect, or 'quantum', speciation (Grant, 1971, p.114) are brought about through drift and 'genetic revolution' (Mayr, 1954) rather than the more gradual processes of conventional divergence through selection.

3.4 Frequency of speciation modes

3.4.1 Sympatric speciation

Although thought to be important in speciation through host shifts in some animals, as exemplified by the celebrated *Rhagoletis* work by Bush (Feder et al., 1988), there is little concrete evidence for primary sympatric speciation in other organisms. Some examples of speciation in animals have been suggested to be sympatric due to the size of the area the speciation events have occurred in (e.g., the endemic beetle fauna of St Helena; White, 1978, p. 245). However, it is acknowledged that microgeographic barriers could have provided isolation and aided speciation in this and similar cases; whether sympatric speciation has occurred depends on ones definition of sympatry.

Grant (1949) suggested that a conspicuous mutation affecting flower colour or morphology could result in sympatric reproductive isolation in plants due to a

switch to a different pollinator that avoided the original type. However, there is a lack of evidence for primary sympatric speciation in plants and it is not considered to be of general importance (Grant, 1971; White, 1978; Mayr, 1963).

There are several documented cases of allopolyploid speciation in angiosperms (e.g. *Spartina anglica*, Marchant 1967; *Senecio cambrensis*, Abbott et al., 1983) so it is certain that secondary speciation through hybridisation can occur; this is the only generally accepted form of sympatric speciation in plants. Autopolyploidy is likely to be far less significant than allopolyploidy in speciation, as combining the genomes of two different species affords a greater probability of ecological differentiation (Macnair, 1989). There is more debate about the importance of allopolyploid speciation than its feasibility, with some authors suggesting it has played a major role in the evolution of some plant groups (e.g. ferns and grasses; White 1978, p. 285).

3.4.2 Geographic speciation

There is a consensus that some kind of geographic speciation is the prevailing mode in most animals and plants (Mayr, 1963; Rice & Hostert, 1993). What forms of geographic speciation are possible and what their frequencies are is a matter of greater debate.

3.4.2.1 Paraptric speciation

The importance of parapatric speciation rests on the interpretation of the hybrid zones seen so often in nature. There is some debate over whether these zones are the results of incompletely speciated previously allopatric populations coming into secondary contact, or whether they are of primary origin. Mayr (1963) regards the secondary contact explanation as the most likely, although White (1978) gives several examples of hybrid zones in which there is no evidence for the populations ever having been separated. It is certainly possible for populations that have parapatric distributions to diverge if selection pressure is sufficient. Jain & Bradshaw (1966) showed that selection can cause very localised patterns of microgeographic structure despite high gene flow (although the examples they give involve very high

selection pressures). It seems the parapatric mode of speciation may be feasible, although other things being equal is likely to be less frequent than forms of vicariant speciation, given the higher rate of gene flow between populations that are contiguous.

3.4.2.2. Vicariant speciation

Straightforward vicariant speciation in which a species is split into two large reproductive communities requires, depending on the species range, a substantial barrier. This barrier must also persist for a considerable period of time, especially as this model does not invoke strong differences in selection pressure either side of the barrier as a prerequisite. Speciation is likely to be slower in large populations due to the time it would take for the spread of new favourable alleles throughout the entire range of the reproductive community, so the 'dumb-bell' model of vicariant speciation is likely to be a very gradual process. In a large population, there is also a lack of opportunuity for drift to play any part in divergence.

3.4.2.3 Peripheral isolate speciation

Although not denying the existence of vicariant allopatric speciation, many authors feel that some kind of peripheral isolate model may allow speciation to proceed more rapidly and occur more frequently.

There are several factors that suggest this may be an important mode of speciation:

- The gene pool of a population at the edge of a species range is likely to be different from that at the centre, possibly representing an extreme of any clinal variation.
- Small populations are prone to drift.
- Selection pressures are likely to be different at the edge of a species range.

- The formation of peripheral isolates may be more frequent than larger scale vicariance events.

Within the broader concept of the peripheral isolate model, there is a further debate on the processes that cause the ecological divergence and reproductive isolation of the isolate. There are essentially two extreme views, which centre on the relative importance of gradual divergence through Darwinian natural selection and the more rapid effects of non-adaptive evolution caused by drift and extreme genetic change resulting from founder events.

The latter view has been espoused in various forms by several authors and is encapsulated by Wright's shifting balance theory (SBT) (Wright, 1932). He argued that populations tend to occupy an adaptive peak, and that to move to another adaptive peak requires them to cross a maladaptive valley. Small populations are able to drift off an adaptive peak, with the new allele combinations possibly putting a neighbouring peak within reach. Three phases can be identified in the shifting balance theory: (i) drift causes populations to lose fitness and move towards a maladaptive valley; (ii) selection pushes the population to a new adaptive peak summit; (iii) the adaptations that allow the occupation of the fitter peak spread throughout the entire species.

Mayr (1954) put forward a founder effect model of peripheral isolate speciation, in which the speciating entity is very small at the outset. The resulting bottleneck following a founding event would lead to increased levels of inbreeding, which Mayr argued could lead to selection of unusual combinations leading to a new and different genetically stable combination. Mayr considered chromosomal changes as important in this model, and the evolution of reproductive isolation to be rapid. Models based around founder events and bottlenecking have been proposed by Carson (1975; the flush-crash-founder cycle) and Carson & Templeton (1984; the founder-flush model)

The alternative view is that adaptations occur through relatively simple mass selection, without the need for invoking founder processes and drift. This gradualist position was taken by Fisher (1930), and is supported by Coyne et al. (1997 & 2000).

Coyne et al. (1997) accept that although each of the three stages of the SBT may occur in nature to some degree, many empirical observations are better explained by the more parsimonious theory of simple mass selection. It is suggested that phases one and two of the SBT do not need to be invoked in order to move from one adaptive peak to another, as fitness peaks may be connected by ridges that require no loss of fitness to traverse, or environmental conditions may change the fitness requirements needed and a new peak may be reached by selection alone. The third phase is criticised on the grounds that a fitness adaptation which required isolation and drift for its fixation is unlikely to spread by gene flow.

Gavrilets & Hastings (1996) carried out a model-based study on the plausibility of founder effect models of speciation. The aim was to find out if founder effects could be important in the evolution of reproductive isolation in a neospecies. They found that drift to fixation at one major locus could completely change the selection pressure on other major loci, and this change in selection pressure caused the population to evolve to a new genetic state (similar to the genetic revolution of Mayr (1954)). With a model including change at many minor loci, drift becomes more important than selection and can allow the population to move to a new adaptive state (similar to the founder-flush model of Carson (1984)). Reproductive isolation in these models depended upon the initial variables used, but could theoretically occur in several hundred or even several dozen generations with quite high probability. However, despite the theory, the experimental evidence for the rapid evolution of reproductive isolation in outcrossers through founder effects alone remains equivocal, with some reviews having completely opposing conclusions as to the presence of this phenomena in experimental populations (Rice & Hostert, 1993; Templeton, unpublished manuscript, cited in Gavrilets & Hastings (1996).

Although there is much theoretical debate about the relative importance of the Fisherian and Wrightian schools in evolution, experimental studies have mostly failed to indicate a large role for founder events in the rapid evolution of reproductive isolation (Rice & Hostert, 1993). Evolution and speciation may be better explained by relatively simple mass selection.

3.4.3 Evidence from phylogenetic data

In terms of a human lifespan, speciation is likely to be an imperceptible process (allopolyploidy excepted), and so judging the plausibility and frequency of different speciation modes in a group of taxa is a difficult task that relies mostly on the interpretation of present day biodiversity patterns or the use of experimental populations.

Lynch (1989) examined the phylogenies of a number of vertebrate groups, and estimated the frequencies of speciation modes (vicariant, peripheral isolate and sympatric) by comparing distributional data with phylogenetic relationships. From 66 cases of speciation, Lynch arrived at figures of 71% for vicariance, 15% peripheral isolate and 6% sympatric.

Barraclough et al. (1999) have developed a method for inferring the geography of past speciation events from species level phylogenies. In plots of node height in a phylogenetic tree versus degree of sympatry between taxa, different patterns are expected depending on whether speciation has been allopatric or sympatric, and whether there have been range changes after speciation has occurred. Node height is calculated as the relative distance of a node from the tips of a phylogeny, and the degree of sympatry is calculated as the mean of the proportion of each clade's area overlapped by the other.

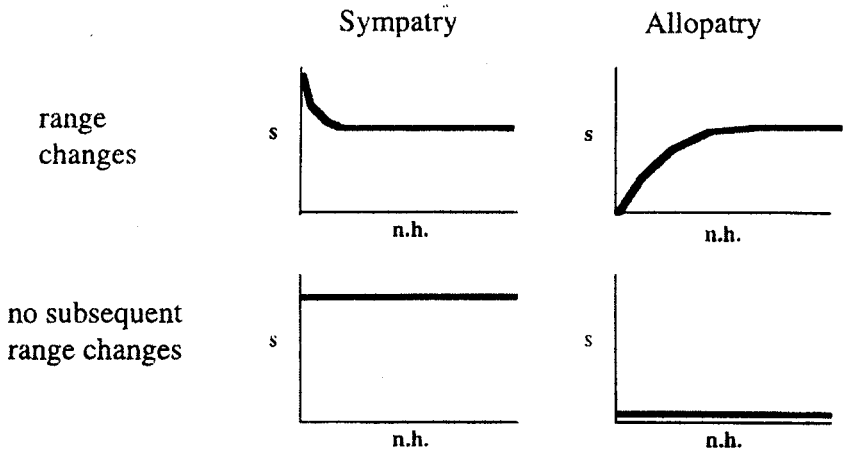

Fig. 3.1. Expected patterns between degree of sympatry (s) and node height (n.h.) under alternative geographical models of speciation.

The expectations of the plots depend on the geography of speciation, and whether range changes have occurred subsequent to speciation. If speciation is allopatric, then we expect an intercept of zero as very recently diverged species should have no overlap in their ranges. This would be followed by an increase in sympatry the deeper one goes down the phylogenetic tree, if species ranges expand and overlap through time. If speciation is entirely sympatric, then we expect an intercept of greater then 0.5, as one of the species of a diverging pair must be completely contained within its progenitor/sister species range at the time of formation. This would be followed by a decline in sympatry with node height if the species expand from their natal range.

Outliers in such plots can expose sympatric events (when a high degree of sympatry exists between two recently diverged species) and historical vicariance events (when complete allopatry persists between taxa despite very long divergence times). An analysis of a species level *Rhagoletis* phylogeny suggests that most of the speciation in the clade examined was allopatric, with only a single sympatric event being identified.

3.5 Gene flow and speciation theory

Section 2.4 was a discussion of largely verbal speciation models, in which there appeared a consensus view that some form of geographic speciation predominates. This indicates that reduction or cessation of between population gene flow is of great importance in the speciation process. In order to further investigate the likelihood of the various allopatric scenarios, we need to know the effects gene flow and population size have on divergence. Mathematical models can be useful in determining the plausibility of speciation theories (Turelli et al., 2001).

Haldane (1930) was among the first to formalise a model including the interaction of gene flow and speciation in the divergence of populations. To prevent local differentiation at a given locus, the fraction of immigrants, m, must exceed the strength of selection, s (the fitness difference between alleles in term of survival

probability). This means that the fewer migrants a population receives from non-local populations, the weaker the selection pressure it can respond to.

Barton (2001) took this further, with more detailed models on the interaction of population adaptation and gene flow. He found gene flow to be an important factor in inhibiting local adaptation, and that gene flow from the centre of a species distribution could effectually limit the range of the species due to the failure of peripheral populations to adapt to conditions to those at the centre.

The same result of gene flow inhibiting peripheral populations from evolving to their local ecological optima was found by Garcia-Ramos & Kirkpatrick (1997). Their models also found that response to local selection pressure can cause rapid and substantial evolution when a peripheral population is isolated from gene flow. A further outcome of the model is that this rapid evolutionary divergence can occur in the absence of drift.

Gavrilets et al. (2000) developed a model of parapatric speciation. If all else is equal, increasing population size should result in greater geographic structure and likelihood of subdivision, increasing the probability of speciation. However, range size can be correlated with dispersal ability, so in general increasing population size (and hence migration rate) significantly decreases the probability of speciation in the model. Gavrilets et al. (2000) suggest species with smaller range sizes (characterised by small population sizes and reduced dispersal ability) should have higher speciation rates.

3.6 Summary

Gene flow can have a creative role in evolution, allowing the spread of advantageous mutations throughout a species range. It can also have a conservative role, with populations being held back from localised adaptive optima by gene flow from neighbouring populations growing under differing conditions.

Disruption of gene flow between diverging entities is crucial to all speciation theories, and this is obvious from verbal models and explicit in mathematical ones.

There is a consensus view that geographic speciation predominates in nature, with some theories favouring a prominent role for small populations at the edge of the parent species range. The relative roles of Darwinian natural selection and non adaptive evolution or extreme genetic change are contested, although many examples of speciation in nature only require selection as an explanation.

Although many authors fiercely defend their own models, it is perhaps best to take a more moderate view, and remember that the different theories are not mutually exclusive. Coyne et al. (1997) concur with this, and state that 'given the multifarious nature of evolution, almost every conceivable scenario must eventually occur'.

4.1 Introduction

For investigations of population structure and recent evolutionary history, microsatellites are now the markers of choice. They are DNA sequences made up of repeats of a motif one to six base pairs long arranged in a tandem fashion. These sequences have been found in the genomes of all organisms that have been analysed so far, and are present in numbers much greater than would be expected by chance (Rose and Falush, 1998). Microsatellites can be simple, where the repeat motif is the sole feature of the microsatellite; compound, where two or more types of repeat motif form adjacent sequences; or interrupted, where the microsatellite region has small sections of non-repetitive DNA.

Microsatellites show high levels of length polymorphism, are co-dominant, are spread throughout genomes and in many cases are thought to be selectively neutral. These factors and the ease with which microsatellite alleles can be scored using PCR has led to them becoming popular as genetic markers.

4.2 Mutation of microsatellite sequences

Microsatellites have some of the highest mutation rates observed at molecular loci. Estimates vary from 10^{-2} in *E. coli* (Levinson and Gutman, 1987*a*), 10^{-3} in humans (Weber and Wong, 1993) and 10^{-4} to 10^{-5} in yeast (Henderson and Petes, 1992), with point mutations for comparison thought to be in the range of 10^{-9} to 10^{-10}. Most of the variability observed at microsatellite loci is due to length changes, hypothesised to be the result of two processes, namely slipped strand mispairing (SSM) and unequal crossing over (UCO). SSM occurs during DNA replication when the template and the newly forming strand dissociate. When the DNA has a repetitive nature, as in microsatellite sequences, the strands may reanneal out of phase, with either the template or nascent strand forming a loop. If replication continues, then the

new strand will be either longer or shorter than the template by the number of repeats contained in the loop. The mechanism of SSM is reviewed by Levinson and Gutman (1987*b*). UCO occurs during recombination, where repetitive areas in the DNA helices become misaligned, which results in a deletion in one chromosome and an insertion in the other.

The relative role of these two mechanisms in microsatellite evolution has been the subject of some debate, although much evidence points towards SSM as being the most important in generating the observed patterns of length variation. UCO is not thought to play a major role, as microsatellite stability is not affected in mutants which show a decreased frequency of recombination events (Levinson and Gutman 1987b; Henderson and Petes 1992), and rates of microsatellite mutation are similar in mitotic and meiotic yeast cells, despite recombination being much more frequent in the latter (Strand et al. 1993). Also, many of the observed mutations in microsatellites involve loss or gain of a single repeat unit, whilst recombinational events would be expected to give rise to a greater range of length differences (Hancock, 1996). There is also more direct evidence in support of the SSM model, mainly from mutants which have defects in their DNA repair mechanisms (Sia et al., 1997). Flores and Engels (1999) showed that *Drosophila* mutant for *spellchecker1*, which promotes the correction of DNA mismatches, had highly increased instability of dinucleotide repeats, with over 90% of observed mutations in these regions being of a single repeat.

The mutation rate of microsatellites has been linked to several variables, including the orientation of the sequence in the genome (Freudenreich et al., 1997), the length of the repeat unit (Chakraborty et al., 1997), the base composition of the repeat unit (Bachtrog et al., 2000) and the total length of the microsatellite (Wierdl, 1997).

Freudenreich et al. (1997) found that the stability of a CTG repeat in yeast was lower when it formed the lagging strand template, which was hypothesised to be due to the higher stability of loops formed from the CTG repeat compared to its reverse complement. Also, the lagging strand is single stranded for longer then the leading strand, and may have more opportunity for forming secondary structures.

(The formation of loops on single stranded template DNA is also likely to be the cause of the stutter banding patterns seen in PCR products generated from microsatellite DNA; stutter bands are usually shorter than the template). There is also evidence from studies in *E. coli* (Veaute and Fuchs, 1993) that leading and lagging strand polymerases have different mutation rates. The dependence of mutation rate on repeat unit length (inversely related) and composition is possibly related to the ability of the sequence to form secondary structures, although the biological basis remains unclear (Estoup and Cornuet, 1996). Bachtrog et al. (2000) suggested the different mutation rates of GT, CT and AT (in order of decreasing rate) microsatellites could be due to DNA mismatch repair enzymes having efficiencies which are sequence dependent.

The increase of mutation rate with microsatellite length has been suggested by Rose and Falush (1998) as a mechanism for the high abundance of longer microsatellites in genomes over that which would be expected by chance alone. They postulate microsatellites are 'born' when a section of repetetive DNA exceeds about 8 nucleotides in total length, when it becomes prone to expansion mutations. There is some evidence for a directional bias in microsatellite mutation, with the tendency being for them to increase rather than decrease in length (Primmer et al., 1996; Amos et al. 1996), although whether this occurs in all types of microsatellite and what the mechanistic basis could be is, again, unclear. If the observed mutation bias is real, then there must be some upper constraints for microsatellite length, otherwise mutation would theoretically drive them towards infinite size; this is also consistent with the observation that most microsatellites are around the order of a few tens of repeats in length. Several hypotheses have been put forward to explain the apparent constraint to microsatellite size, including selection against large alleles (reviewed by Estoup and Cornuet, 1996) and the increasing instability of microsatellite DNA with increasing length, which makes it prone to large deletions (Wierdl et al., 1997). A recombinational mechanism for such deletions has not been ruled out (Hancock, 1996).

4.3 Function of microsatellite sequences

Although microsatellites are usually thought of as useful, polymorphic neutral genetic markers, there is evidence that in some cases at least they do have a functional role in the genome. Perhaps the most obvious role they possess is in coding regions, where they code for homopolymeric stretches in a protein. This is perhaps more likely for three or six base pair repeat motifs (e.g. Hughes et al., in press; paper 2), where stepwise mutations would not cause a frameshift. Two interesting cases of coding region tri-nucleotide microsatellite polymorphism are worthy of comment; one by Sawyer et al. (1997) in which a (Thr-Gly) repeat of either 17 or 20 units was associated with maintaining an accurate circadian rhythm at different temperatures in *Drosophila*, and another by Ebstein et al. (1996), which found a correlation of extroversion and novelty seeking scores in humans with variation in the length of a 16 amino acid region in a dopamine receptor gene.

Microsatellites may also have a role as regulatory elements, as they are often found in upstream promoter regions of coding sequences, and numerous proteins have been found that selectively bind to di and tri nucletide repeats (Kashi and Soller, 1999; Epplen et al., 1993). The presence of microsatellites as functional elements in both promoter and coding regions gives the possibility that they may be a major source of genetic variation and phenotypic novelty, and hence of great importance in evolution.

This does not mean to say that assuming most microsatellites are neutral markers is wrong, however. Rose and Falush's (1998) model explains how microsatellites could arise throughout the genome, and it is possible to see a scenario where microsatellites are 'born' by chance, expand through mutation and are lost either through catastrophic deletions or through gradual accumulation of point mutations. Assuming microsatellites to be neutral debatably requires less of a suspension of disbelief than assuming allozymes to be neutral.

4.4 Analysis of microsatellite allele data

Microsatellite data can be used to make inferences about population structure or differentiation using a number of classical genetic distance measures, or methods developed specifically for microsatellite allele data.

This section will attempt to review some of these methods and discuss their relative merits and shortcomings. In covering such a large topic, it is helpful to set out some kind of taxonomy for these concepts, which can be broadly divided into two main types, namely ***F*-statistics**, which detect deviations from Hardy Weinberg equilibrium and can provide a summary of population substructure in a single figure, and **distance measures**, which compare the similarity (or difference) of individuals or populations, often in some kind of multi dimensional space. Within both of these ways of looking at population structure, there are two main types of model, namely the infinite alleles model (IAM) and the stepwise mutation model (SMM).

The IAM was developed by Kimura and Crow (1964); they derive the expected number of alleles in a population at equilibrium for mutation and drift when each mutation produces a unique allele. This model makes no assumptions about the relatedness of different alleles, and there is assumed to be no homoplasy.

Ohta and Kimura (1973) further developed the model to account for homoplasy in isozyme data, when base changes to not give rise to a distinguishable electromorph. This gives rise to the SMM, in which alleles evolve in a random stepwise fashion. This means that allele identity carries information about allelic relationships, and also that homoplasy is expected, where there is a chance that alleles identical in state may not be identical by descent.

4.4.1 *F-statistics and relatives*

This group of estimators are related through their use of comparing some parameter of a subpopulation ($_S$) to the parameter of a more inclusive population ($_T$); they can provide a measure of population structure as a single figure, but F_{ST} and its

analogues can also be used in population pairwise comparisons and so can also be used as distance measures.

F-statistics themselves are the cornerstones of much of population genetic theory and were orginally developed by Sewall Wright (1951), becoming widely used with the advent of protein electrophoresis in the 1960s. They are derived from measures of deviation from panmixia based on three measures or estimations of heterozygosity, namely

- H_I The average *observed* heterozygosity within subpopulations
- H_S The average *expected* heterozygosity estimated from each subpopulation
- H_T The *expected* heterozygosity estimated from the entire system

From these three measures of heterozygosity, the different *F* statistics (**Fixation** statistics) can be derived.

F_{IS} is the inbreeding coefficient, and is represented by the ratio $F_{IS}=(H_S-H_I)/H_S$. An excess of homozygotes over Hardy Weinberg expectations indicates mating is not random, and this lowering of H_I forces the ratio towards unity. $F_{IS}=1$ indicates no outbreeding, whilst $F_{IS}=0$ indicates mating is random. The value of F_{IS} can range between -1 and 1, with -1 indicating fixed heterozygosity. F_{IS} can be estimated for single subpopulations as well as for the entire system.

F_{ST} is similar to F_{IS} in that it measures inbreeding like effects, but between rather than within subpopulations. It is represented by the ratio $F_{ST}=(H_T-H_S)/H_T$. For a di-allellic locus in which two sub populations are homozygous for different alleles, H_S would be zero and thus F_{ST} would be unity. Increased allele sharing between the two subpopulations would increase H_S, and reduce the value of F_{ST} accordingly; the value for this estimator lies between 0 (no population differentiation) and 1 (maximum population differentiation).

F_{IT} is a measure of deviation from panmixia due to the effects of inbreeeding and subdivision across the entire study system, and is represented by the ratio $F_{IT}=(H_T-H_I)/H_T$. The three coefficients are related by the equation $(1-F_{IT})=(1-F_{IS})(1-F_{ST})$.

Estimates of F_{IS} and F_{ST} provide an indication of how much deviation from panmixia there is within and between populations respectively, and so are biologically relevant figures. F_{ST} is of greater interest from the perspective of population connectivity and history, which is what many population studies aim to investigate. There are several methods used to provide estimates of F_{ST}, and two which are frequently used are Nei's (1987) G_{ST} and Weir and Cockerham's (1984) variance based method θ_{ST}.

G_{ST}, an estimate of F_{ST} for use with multi-allelic loci was developed by Nei through extending Nei's genetic distance between a pair of populations to the case of a hierarchical structure of populations (Excoffier, 2001). It is equivalent to Wright's (1951) F_{ST} when there are only two alleles at a locus; in the case of multiple alleles G_{ST} is equivalent to the weighted average of F_{ST} for all alleles (Culley et al., 2002). Nei's estimate G_{ST} is calculated from $G_{ST}=D_{ST}/H_T$, where D_{ST} is the average gene diversity between populations (Nei, 1987, pp. 188-189).

Weir and Cockerham's variance method uses the observation that, if there is population substructure, then alleles found within a sub population should be found together more often than would be expected given the frequencies of the alleles in the entire population. Slatkin and Barton (1989) describe Weir and Cockerham's estimator as 'algebraically complicated'. The full derivation of their estimator, based on partitioning of variance of allele frequency, can be found in Weir and Cockerham (1984).

Wrights *F*-statistics and the estimators of them discussed above (G_{ST} and θ) are based on the IAM, in which every mutation gives rise to a distinguishable allele (i.e., no homoplasy) and no relationship between alleles is inferred. Although analysing allozymes electrophoretically may lead to hidden homplasy where an 'electromorph' consists of two or more alleles, the IAM has nontheless been widely and successfully

used in the interpretation and explanation of allozyme data. With the use of microsatellites becoming more widespread, a method of estimating F_{ST} using a SMM was developed (Slatkin, 1995) to account for the higher levels of homoplasy which would exist under a scheme of stepwise mutation, and to make use of the information present in the lengths of microsatellite alleles. Slatkin's estimator is $\boldsymbol{R_{ST}}$, and is equal to $R_{ST} = \frac{\overline{S} - S_W}{\overline{S}}$, where S_W and $\overline{S}$ are the average sums of squares of the difference in allele size within a subpopulation and the entire population respectively. This has been shown using coalescent theory to be equivalent to F_{ST} when applied to loci evolving under a SMM (Slatkin, 1995).

The utility and accuracy of these measures depends on the type of data one has and over what timescale the taxa being observed have diverged. This will be discussed in section 4.4.3.

4.4.2 *Genetic distance measures*

There is a diversity of genetic distance measures, based on both the IAM and SMM as well as purely geometric measures not explicitly linked to a model of locus evolution.

4.4.2.1 Geometric distances

The following genetic distances are based solely, in various ways, on the difference in allele frequencies between taxa (e.g., individuals or populations).

A. Proportion of shared alleles

The proportion of shared alleles or P_{SA} is equal to the number of shared alleles summed over loci divided by 2n, where n is the number of diploid loci (Bowcock et al., 1994). P_{SA} is a measure of similarity, and can be converted to a distance measure (D_{SA}) by subtracting it from unity or taking its logarithm. The first use of this measure is usually referenced to Bowcock et al. (1994), although how the 'number of shared alleles' is calculated is unclear as no equation is given in the paper.

Goldstein et al. (1995a) use an allele sharing method called D_{AS} which they refer to Bowcock et al. (1994) and calculate it as

$$D_{AS} = 1 - (1/2N)^{-2} \sum_{i} \sum_{i'} I(i,i')$$

which is equal to 1-(the average number of shared alleles), where the first and second sums are over all alleles in the first and second population, N is the number of diploid individuals, and $I(i,i')$ is an indicator variable that equals 1 if the alleles are the same and 0 if the alleles are not, with i and i' being the alleles sampled from a locus in the first and second populations.

The software package MICROSAT (Minch et al., 1995) calculates P_{SA} as 'the mean of the minima of the relative frequencies of all alleles in the taxonomic units being compared':

$$P_{SA} = 1 - \frac{\sum MIN(x_i, y_i)}{n},$$

where n is the total number of alleles for all loci.

Stephens et al. (1992) calculates the average percentage difference in bandsharing, which is the average between all pairwise comparisons of

$$PD = \frac{(b_x - b_{xy}) + (b_y - b_{xy})}{b_x + b_y}$$

where b_x is the number of bands observed in individual x, b_y is the number of bands observed in individual y, and b_{xy} is the number of bands shared between x and y. (Used for dominant marker data in this case). It is difficult to see if this is equivalent to the D_{AS} attributed to Stephens (discussed below, under chord distances) by Goldstein and Pollock (1997), and how the 'proportion of shared alleles' measure used by Bowcock et al. (1994) is distinct.

B. Fuzzy set similarity

The fuzzy set similarity FS is calculated from the set of alleles found in each of two populations, and dividing the cardinality of their intersection by the cardinality of their union

$$FS = |X \cap Y| \div |X \cup Y|$$

The fuzzy set similarity is related to the other allele sharing measures, but differs in considering the entire set of alleles at once. It can be converted to a distance by subtracting it from unity or taking its logarithm, and the measure 1-FS is equal to the proportion of alleles unique to either of a pair of taxa and hence bounded between 0 and 1. There has been no published use of this measure for microsatellite data, although it is an option in MICROSAT (Minch et al., 1995).

C. Chord distance

The chord distance of Cavalli-Sforza and Edwards (1967) involves transformation of the data into an angular distance θ, so all populations can be conceptualised as points in an m-dimensional Euclidean space, where m is the number of allele frequencies being compared (equal to the total number of alleles in

both populations). From this, the chord distance D_{CH} is calculated as the straight line distance in space between the two points. This is a geometric distance and not based on any biological assumptions, and is similar to Stephens et al. (1992) allele sharing distance D_{AS} and to Nei's (1987) D_A (Goldstein and Pollock, 1997). All three measures are based around the product of allele frequencies between two populations, and take the form

$$D = c\left[1 - \left(\sum x_i y_i\right)^a\right]^b$$

where x_i is the frequency of allele i in population x. For D_{CH}, a and b equal 0.5 and $c = 2/\pi$; for D_A a=0.5 while b and c equal one; for D_{AS} a,b and c are equal to one. For multiple loci, the results are averaged. Nei created D_A in order to scale the chord distance between 0 and 1 (Nei, 1987, p. 216). All these distances vary between some non-zero positive number and c. Goldstein and Pollock (1997; box B) recommend bootstrapping over loci as the between locus variance for these measures can be large.

Also belonging here is Cavalli-Sforza's kinship coefficient (Cavalli-Sforza, 1971), which is defined as the probability that an allele taken at random from a given locus will be identical by descent in the two taxa being examined, although it is calculated in a similar way to the geometric distances above. The probability of drawing an allele i from a locus is equal to its frequency, and the probability of drawing the same allele from two taxa is equal to the product of its frequencies in each population. K_F is a similarity measure, and the distance

$$D_{KF} = -\ln \sum x_i y_i ,$$

which is equivalent to Stephens D_{AS} *sensu* Goldstein and Pollock (1997). It can be calculated over multiple loci and averaged.

D. Rogers' distance

Rogers' distance, D_R, has been widely used with allozyme data. It is a geometric distance, bounded between 0 and 1, estimated by representing the populations or individuals under consideration in an m-dimensional space, where m is the total number of alleles. The calculation is based around the difference in allele frequencies between populations, and is given by

$$D_R = \left[\frac{1}{2}\sum (x_i - y_i)^2\right]^{\frac{1}{2}}$$

This was produced by Rogers (1972) with the division by 2 bounding the distance between 0 and 1. With many loci, the average is used.

It is very similar to a measure used by Provosti et al. (1975), $C_P = \frac{1}{2}\sum |xi - yi|$.

D_R is not proportional to either evolutionary time or to the number of mutations. It also suffers from being mislead by high gene diversity, as two populations fixed for different alleles give a value of less than 1 when the number of alleles is high (for five non-shared alleles, D_R=0.45) (Nei, 1987, p. 211)

E. Likelihood ratio distance

The genotype likelihood ratio distance, D_{LR}, was developed by Paetkau et al. (1997).

$$D_{LR} = \left(\frac{1}{n_X}\sum_i^{n_X} \log\frac{L_{iXX}}{L_{iXY}} + \frac{1}{n_Y}\sum_i^{n_Y} \log\frac{L_{iYY}}{L_{iYX}}\right) \div 2$$

where n_X and n_Y are the number of individuals in populations X and Y, and L_{iXX} and L_{iXY} are the likelihood of finding the an individual with genotype i (from population X) in population X and Y respectively. The likelihoods are derived from the observation that the probability of drawing a single locus genotype from a population is $p=x_ix_j$ (or $p=x_i^2$ in the case of a homozygote), where x_i and x_j are the frequencies of alleles i and j in the population. This measure performed very well in an analysis of data from bear populations by Paetkau et al. (1997), where it had a far lower variance than SMM derived measures, and was very successful in highlighting geographic structure.

4.4.4.2 IAM based measures

In contrast to the geometric distances discussed above, these are based on a model of locus evolution which incorporate the fact that allele distributions are affected by mutation and drift to a greater or lesser extent. They are designed to increase linearly with time when used for loci that are evolving under an IAM.

A. Latter's F_{ST} and Reynold's coancestry

These measures are essentially estimators of Wright's F_{ST}, but derived specifically for population pairwise comparisons. The two are equivalent when sample sizes are large; Latter's estimate is given here (Latter, 1972; see also Takezaki and Nei, 1996):

$$F_{ST} = \frac{(J_X + J_Y)/2 - J_{XY}}{1 - J_{XY}}$$

where J_{XY} is the average over all loci of $j_{xy} = \sum x_i y_i$, with x_i and y_i being the frequencies of allele i in populations x and y respectively, and J_X is the average over all loci of $j_x = \sum x_i^2$. This estimator is based on a drift only model (Weir, 1996; p. 195), and therefore should not be used with loci that have a high mutation rate or over long time scales.

B. Nei's standard distance.

Nei's standard genetic distance (Nei, 1987)

$$D_S = -\ln\left[J_{XY} / \sqrt{J_X J_Y}\right]$$

The expression in brackets is equivalent to I, which is Nei's standard genetic identity (Nei, 1987, p. 221). This is bounded between 0 and 1, hence D_S varies between 0 and ∞.

In contrast to the IAM based coancestry coefficients discussed above, this model used for this distance incorporates both drift and mutation, and so will have a wider applicability.

D_S is expected to be linear with respect to time for evolution under an IAM, as long the same balance between drift and mutation is maintained (Takezaki and Nei, 1996).

4.4.4.3 SMM distance methods

The following methods were all developed within a very short time of one another, with microsatellite data becoming more widely obtainable in the mid 1990s. They were designed to overcome perceived shortcomings of either geometric or IAM based distances when used for loci which are evolving at a high rate under a SMM. All make use of the information that is present in the length difference between two alleles assuming stepwise evolution has occurred, and are designed to increase linearly with time when loci are evolving according to the SMM.

A. Shriver's *Dsw*

Shriver (1995) *s*tepwise *w*eighted genetic distance measure is an extension of Nei's minimum genetic distance (Nei, 1987; p. 219). It weights the components of the measure (the products of allele frequencies) according to the number of steps between the two alleles. Its uses the absolute value function of the difference between allele sizes which is $\delta_{ij} = |i - j|$, where i and i are the number of repeats in the two alleles.

$$D_{SW} = d_{XYW} - (d_{XW} + d_{YW})/2,$$

where $d_{XW} = \sum_{i \neq j} \sum x_i x_j \delta_{ij}$, $d_{YW} = \sum_{i \neq j} \sum y_i y_j \delta_{ij}$ and $d_{XYW} = \sum_{i \neq j} \sum x_i y_j \delta_{ij}$

B. Slatkin's *ASD*

Slatkin's *A*verage *S*quare *D*istance (1995; see also Goldstein et al. (1995a)). This and Goldsteins $(\delta\mu)^2$ discussed below use the square of the number of repeats between alleles rather than the absolute value.

$$ASD = \frac{1}{r} \sum^{r} \sum_{i,j} (i - j)^2 (x_i y_j)$$

for r loci, where i and j are the number of repeats in the two alleles.

Goldstein et al. (1995a) state that this distance is dependent on the data following a SMM quite closely, and will perform progressively worse as the mutation model becomes more like the IAM. This is due to the *ASD* incorporating

the length of alleles and assuming this information is related to time since common ancestry, which under a strict IAM actually bears no relation at all. The linearity of the measure also depends upon whether microsatellite mutation is related to repeat length, for which there is considerable evidence (Wierdl et al. 1997). These limitations also apply to D_{SW} and $(\delta\mu)^2$.

C. Goldstein's $(\delta\mu)^2$

Goldstein et al. (1995b). This measure uses the difference of the mean of the repeat lengths in alleles between taxa.

$$(\delta\mu)^2 = \frac{1}{r}\sum_{j}^{r}(\mu_{X_i} - \mu_{Y_j})^2$$

where $\mu_{X_i} = \sum_i i x_{ij}$ is the average allelic state at the j^{th} locus in population x, and x_{ij} is the frequency of allele i in population x. This measure was developed by Goldstein et al. in order to improve on the average square distance (*ASD*) by removing its dependence on population size and decrease its variance. They show that the *ASD* includes the variance in allele size within populations, and that $ASD = V_A + V_B + (\mu_A + \mu_B)^2$, where V_A, V_A and μ_A, μ_B are the variances and means, respectively, of allele size in populations A and B. With the assumption that populations are at mutation-drift equilibrium this variance does not change over time, and the growth in *ASD* is solely due to the squared difference between the means. $(\delta\mu)^2$ is based only this difference, and is averaged first within populations which standardises the distance with respect to the variation within populations without estimating additional parameters.

D. Range restricted

Developed by Goldstein and Pollock (1997) using a model based on the SMM, but which has reflecting upper and lower boundaries for allele size in an attempt to mimic the actual evolution of microsatellite loci more realistically.

$$D_L = \log\left[1 - \sum_l (\delta\mu)^2{}_l / LM\right]$$

where L is the number of loci and M is the average value of $(\delta\mu)^2$ at maximal divergence. Two further measures were developed from D_L in order to account for differences in range size and mutation rate between loci; see Goldstein and Pollock (1997) for the derivations.

4.4.3 Which measure to use?

F-statistics are useful for generating instantly comparable figures of population structure. The estimators of Nei (1987) and Weir and Cockerham (1984) were derived using an IAM, and assume a primary role for drift, and so would be innaccurate under conditions of high stepwise mutation. They are not strictly applicable to microsatellite data except when divergence times are very short. Another reason for the limited applicablity of IAM estimators of *F*-statistics to microsatellite data is that they bias the estimates downwards when gene diversity is high, even when populations may be fixed for mutually exclusive alleles.

This effect is demonstrated in fig 4.1. R_{ST} performs far better than F_{ST} in estimating the degree of population substructure when there is a high level of gene diversity. Conversly R_{ST}, developed using a SMM is not suitable when drift has played a large role relative to mutation. If population sizes have been small and drift strong, then the resulting allele data is likely to conform to a 'patchy' distribution one might expect to find under an IAM. If one considers two populations, one in Hardy-Weinberg equilibrium for alleles of length 1 and 3, and the other fixed for an allele of length 2, then R_{ST} approaches zero, despite the populations having no alleles in common (whilst estimates for F_{ST} are 0.75). As data departs from a stepwise mutation model (as is often the case with plant microsatellite data) then the expectation of estimators of F_{ST} and R_{ST} converge (Balloux and Goudet, 2002). As estimators of R_{ST} retain a larger variance than estimators of F_{ST} (Excoffier, 2001), the latter remains the statistic of choice for many studies. The shortcomings of these and similar estimators when data deviates from the assumed model is also relevant when they are used as genetic distance measures between two taxa.

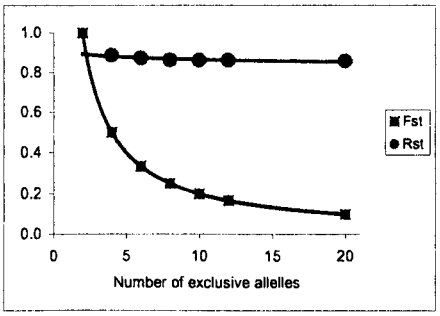

Figure 4.1. The relationship between F_{ST}, R_{ST} and the number of exclusive alleles at a single locus in two populations. (Population pairs with differing numbers of mutually exclusive alleles were created using EASYPOP (Balloux 1999). Each population (500 individuals) has a single locus which has either 2, 4, 6, 8, 10, 12 or 20 alleles which are all private to that population and are at Hardy Weinberg equilibrium. The alleles are are consecutively numbered, e.g. for 10 mutually exclusive alleles, population 1 has alleles 1 to 10 and population 2 has alleles 11 to 20. F_{ST} and R_{ST} were calculated for each population pair using FSTAT (Goudet 2001)).

There is some confusion in the literature about the difference between θ and G_{ST} which are the two commonly used estimates of F_{ST}. Nei's (1973) G_{ST} is calculated from the average gene diversity between sub populations, D_{ST}. D_{ST} is calculated using comparison of sub-populations with themselves, and so is sensitive to the number of sub-populations. A modification to remove this dependence on sub-population number is given by Nei (1973), and the estimator G_{ST}' is calculated using a measure of absolute gene differentiation. This modification is not considered entirely theoretically satisfactory by some (Excoffier, 2001, p. 289; Nei and Kumar, 2000, p. 249). In practice the two estimators G_{ST} and G_{ST}' do not differ greatly as long as the number of sub-populations is greater than five. Weir and Cockerham's θ (1984) is calculated using a method derived from ANOVA. It is independent of the number of sub-populations, and has shown to be a less biased estimator (Ouborg et al., 1999). Excoffier (2001) also states the ANOVA based approach is preferable because of its clear statistical foundations. The two statistics are related by the equation $\theta=G_{ST}/[1-(1- G_{ST})/d]$ where d is the number of locations sampled. Because of the way it is defined, G_{ST} is always >0 for any set of allele frequencies; this is not the case for θ, which is intended as an unbiased estimator (Slatkin and Barton, 1989).

In the real world however, there is little to choose between estimates. In practice both give similar estimates of F_{ST} (as long as the study includes more than 5 populations), and the confidence one has in the result is influenced far more by the number of loci, the number of individuals in a sub-population and the number of sub-populations in ones data set than the minutae of the mathematics.

If one wants to infer a phylogeny of taxa (e.g. populations) then it will be necessary to produce a pairwise distance matrix using one of the many genetic distance measures. A distance that is linear with time is obviously best if any estimates of divergence times are to be made from the tree. The geometric distances such as D_{CH} and its relatives are all based on the overlap of allele frequencies in some form, so cannot measure an increasing divergence between taxa when they have no alleles in common. This means they plateau as time increases, and only reflect divergence times accurately when the taxa are very closely related. With more divergent taxa, the methods developed using the SMM will continue to detect divergence at increasing timescales.

An extensive study into the reliability of reconstructing phylogenetic trees using microsatellite data was carried out by Takezaki and Nei (1996). They modelled the evolution of a single ancestral population into eight derived ones using both a SMM and an IAM, and assessed the accuracy of various genetic distance measures by noting the percentage of times the correct phylogeny was recovered from 200 simulations. They found that Nei's D_A and Cavalli-Sforza's D_{CH} were the most efficient in obtaining the correct tree topology. This is despite the fact that both of the distances are geometric and essentially ignorant of the model of locus evolution. As expected, the values of $(\delta\mu)^2$ and D_S increased linearly with time under the SMM and IAM respectively, although they were poor at reconstructing the correct tree topology. This is because the model derived measures have a much larger variance than than the simple geometric ones. Goldstein et al. (1995a) also compared the accuracy and variance of geometric and SMM based genetic distances. At short divergence times, allele sharing methods outperformed the *ASD* in producing correct tree topologies. At longer divergence times, *ASD* became more accurate than allele sharing despite its larger variance, as the geometric distance could not track the

increasing divergence in allele size. Takezaki and Nei (1996) recommend using either D_A or D_{CH} for reconstructing phylogenetic trees, and using $(\delta\mu)^2$ and D_S to estimate the branch lengths. Allele sharing methods were used by Goldstein et al. (1999) and Bowcock et al. (1994) to produce trees of individuals (foxes and humans respectively) with very strong geographic clustering in which the authors had very high confidence. Goldstein (1995b) re-analysed the human, chimp and gorilla data from Bowcock (1994) using $(\delta\mu)^2$. With the human-only data, bootstrap values were far lower than obtained in the original analysis based on allele sharing, although deep nodes in the tree (between African and non-Africans) had significant bootstrap support. With more divergence, Goldstein's measure became more powerful, as it outperformed both allele sharing and Nei's standard distance in resolving a human-chimp clade with significant bootstrap support.

The most intensive investigation of genetic distance measures using real rather than simulated data was carried out by Paetkau et al. (1997), with the data coming from population samples of three bear species. In line with the findings of other studies, non-model based distances with low variances gave the most realistic results at the intra-specific level, despite their non linearity over time. None of the distances were able to resolve the between species phylogeny, however, which reflects a limitation of microsatellite data at this level rather than limitations of data analysis. SMM based measures such as $(\delta\mu)^2$ should theoretically be linear over timespans of millions of years if microsatellite evolution conforms to the boundless SMM, which in this case it obviously does not. So, it seems that SMM based measures are also non linear with respect to time, which combined with their large variances (especially at small timescales) severely limits their utility.

4.5 Biological interpretation of population structure

F-statistics are widely used because they give comparable figures of population connectivity that are directly linked through theory with the number of migrants exchanged between populations. However, many urge caution when interpreting estimates of *F*-statistics, as they are influenced by several factors which are not

linked to population connectivity (Balloux and Lugon-Moulin, 2002; Bossart and Prowell, 1998).

4.5.1 Interpretation of F_{ST}

Interpreting an extreme value of F_{ST} is relatively straightforward. A value of one means there is no diversity within subpopulations, and that at least two of the sub-populations are fixed for different alleles. A value of zero means the samples are taken from a single panmictic unit. Values between zero and one can be interpreted as representing intermediate levels of population structure, and this is where the trouble starts.

Infinite-allele based estimates of population differentiation depend strongly on the allelic diversity of the loci used. This effect is clearly shown in fig. 4.1, where loci with high mutation rates and gene diversity can give low measures of differentiation. Conversely, differentiation measures based on a stepwise model of mutation can be misleading when drift has been strong relative to mutation. Thus, in interpreting estimates of population structure one must be aware that they are affected, in some cases quite strongly, by factors other than the biology of the study organism.

4.5.2 Calculating number of migrants

Given an estimated F_{ST} value it is tempting to calculate the biologically interesting figure of the number of migrants per generation (Nm). This is fraught with pitfalls. Not only will the result be biased due to the performance of the estimator as described above, but in reality populations never conform to the assumptions of the underlying island model (Whitlock and McCauley, 1999). Furthermore, F_{ST} and R_{ST} are non-linear functions of Nm, and hence estimates of the number of migrants will have very large confidence intervals when the values of population structure estimates are low, rendering the results of little use (Balloux and Lugon-Moulin, 2002). Population differentiation can also be affected by factors other than migration, which can also lead to erroneous results of estimates of Nm.

4.5.3 Inferring isolation by distance

Bossart and Prowell (1998) state that a geographic correlation of genetic structure can arise through two distinct processes. Structure can arise through vicariant events which stop gene flow between populations, and also by gene flow following an isolation by distance model, with nearby populations exchanging more migrants than distance ones. Isolation by distance can be detected by performing a regression of $F_{ST}/(1-F_{ST})$ on the natural log of geographic distance if populations have a two dimensional distribution, or on the non-logged geographic distance if the populations are arranged in a linear fashion (Rousset, 1997). Distinguishing vicariant events from differentiation caused through isolation by distance can be difficult. The two are likely to be confounded as vicariance is more likely to be detected with increasing geographic distance, meaning some samples that contribute to a significant isolation by distance pattern may in fact be completely isolated. Bossart and Prowell (1998) recommend checking whether a single population or group of populations is resonsible for producing a significant pattern, and if this is the case then vicariance should be considered as a possible explanation.

A new endemic species of *Begonia* (Begoniaceae) from the Socotra archipelago.

M. Hughes & A.G. Miller. *Edinburgh Journal of Botany*, in press.

Abstract

The new species *Begonia samhaensis* in section *Peltaugustia* (Warb.) Barkley is described from the island of Samha in the Socotra archipelago. It differs from the other member of the section, *B. socotrana* Hook. f., in a number of gross morphological characters and is likely to be a relict taxon rather than the result of more recent dispersal and divergence. A revision of section *Peltaugustia* is presented.

Detailed surveys have been carried out on both *Begonia*. The new species has a restricted distribution and a total population of less than 1000 individuals, and is recommended to be placed in the IUCN category VU D1, 2. *B. socotrana* has been found in new sites, and is locally common in parts of its range. Its current placing in the IUCN 'Vulnerable' category is considered to be unwarranted, and it is recommended that the species should be listed as 'Least Concern'.

Keywords. *Begonia*, Socotra, conservation.

Introduction

The first *Begonia* to be described from the Socotra archipelago was *B. socotrana* Hook. f., discovered by Isaac Bayley Balfour during a British Association expedition to the island of Socotra in 1880 whilst he was Regius Professor of Botany at the University of Glasgow. Upon his return, living material of the plant was donated to the Royal Botanic Gardens, Kew, where it was described by Hooker (1881: 8), who noted that Socotra was 'one of the last places in the world in which a *Begonia* could have been expected to occur', as the island suffers a prolonged and severe dry season during the summer months. Upon its introduction to cultivation the plant was an immediate horticultural success (Gleed, 1961) because, being a strictly short day plant, it made possible the production of the first winter-flowering cultivars. It is endemic to the granitic Haggier mountains and adjacent high limestone plateaus of eastern Socotra.

Hooker considered *B. socotrana* to be closely allied to *B. geranioides* Hook. f. of South Africa and placed it in the same section, *Augustia* (Klotzsch) A. DC.: 'From the geographical position of the island, the affinity of this discovery may be considered to be either Asiatic or African, and, upon the whole, though referable to none of the sixty sections of the genus founded by Klotsch [1854] and De Candolle [1864], it must, I think, be placed in the African one of *Augustia*, from the character of which it differs chiefly in the male perianth having four segments, in the shorter filaments, rounded top of the anther, the six lobes of the female perianth (instead of five), and the intwisted arms of the style, characters all of which, except the last, occur in the Natal *B. geranioides*, to which *B. socotrana* is unquestionably closely allied' (Hooker, 1881: 8).

Warburg (1894: 140) considered the species to be distinct enough to warrant the creation of a subsection within *Augustia* to accommodate it: (translated from German) '*B. socotrana* has been placed in a separate subsection *Peltaugustia* due to it having peltate leaves, one winged fruit and bulbils on a swollen rootstock. This subsection is transitional with section *Reichenheimia*'. Although *Peltaugustia* was not recognized by Irmscher in his monograph of *Augustia* (Irmscher, 1961), it was

elevated to sectional status by Barkley (1972), reflecting its unusual anatomy and isolated position within the genus.

A second species of *Begonia* from the archipelago was discovered on the island of Samha by an expedition from the Royal Botanic Garden, Edinburgh in 1996. Until more recently (1999) the species was known only from a single plant found on the northern side of the highest point of the island, which is a limestone outcrop approximately 50 m^2. Only bulbils were seen as the plant had died back for the dry season, and it was thought to be *B. socotrana*. The bulbils were cultivated at the Royal Botanic Garden, Edinburgh, where it became apparent the collection represented a new species. Samha could be considered an even less likely place than Socotra in which to find a *Begonia*, as it only reaches an altitude of 779 m. It therefore attracts a reduced amount of moisture in the form of mist and lacks the lush montane vegetation associated with *B. socotrana*. The new species (*B. samhaensis* M. Hughes & A.G. Miller, described below) has been placed in sect. *Peltaugustia* with *B. socotrana* as it possesses bulbils and peltate leaves which are the definitive characters of the section, although it is distinct from *B. socotrana* in a number of gross morphological characters, summarized in Table 1.

	B. socotrana	*B. samhaensis*
Tuber	absent	present
Leaf shape	orbicular	ovate
Male tepals	subequal	unequal
Male bud	conical	purse-shaped
Stigmatic surface	helical	irregular
Capsule wings	one enlarged	equal

TABLE 1. Morphological differences between *B. socotrana* and *B. samhaensis*.

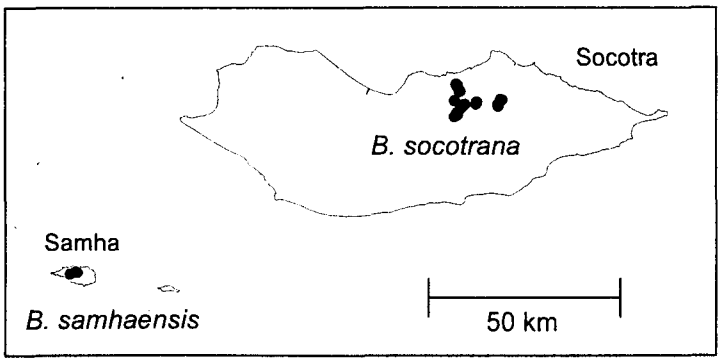

FIG. 1. The distribution of *Begonia* sect. *Peltaugustia* on the Socotra archipelago.

The morphological differences and the marked divergence in nuclear ribosomal ITS sequences (M. Hughes, unpublished data) suggest the species is relict on the island rather than the result of more recent dispersal and divergence. The distribution of both species is shown in Figure 1.

Conservation status

Begonia socotrana has been rumoured to be very rare in the wild since an expedition to Socotra in 1967 by Lavranos and Radcliffe-Smith. In their list (Lavranos & Radcliffe-Smith, 1969: 3) the plant was recorded only from two of the highest peaks of the Haggier mountains, and was described as 'not common, but not in immediate danger owing to virtual inaccessibility to man and goat.' In the 1978 IUCN Plant Red Data Book (Lucas & Synge, 1978: 79) it is listed as 'endangered', based largely upon the information from the 1967 expedition but stating 'the population of this island endemic has reached critically low levels' and citing grazing as the probable cause of its decline. The fact that it was found on only two high peaks in 1967 is highlighted, although reference is made to an earlier expedition by Popov (1957) who found *B. socotrana* on the Reiged limestone plateau to the west of the Haggier. The listing of *B. socotrana* in the IUCN red data book (1978) has caused it to be highlighted in other publications on threatened plants. Koopowitz & Kaye (1983: 63) suggested that 'the population of begonias has steadily eroded' and Belousova & Denisova (1992: 323), described the plant's populations as at a

'critically low level.' Both cite overgrazing by goats as the main threat. The latest IUCN Red list of threatened plants (Walter & Gilliet, 1998: 73) lists the species as 'vulnerable', a category one step below the 'endangered' status that the species was awarded in 1978.

Observations during RBGE expeditions in 1989, 1990, 1992, 1993, 1996, and 1998 suggested that *B. socotrana* was more common than the previous publications state, and this prompted detailed surveys of both *Begonia* species in Spring 1999 and 2000. These surveys confirmed that *B. socotrana* is locally common and has a far wider distribution in the Haggier than stated by Lavranos and Radcliffe-Smith (1969), and still occurs in sizeable populations on the limestone plateaus of Reiged and Rewged, as found by Popov (1957). Part of the reason for the apparent scarcity of the species in 1967 is likely to be the timing of the trip, which occurred in March and therefore coincided with the start of the dry season and the die-back of low altitude populations of *Begonia*. The threat of grazing seems to have been overstated, given the large size of some of the populations found growing within potential reach of livestock. Goats will eat the leaves, but they are quite acidic and during the wet season there is plenty of other more palatable fodder, which is eaten in preference. Even if grazing pressure were to increase, many of the populations of *B. socotrana* grow on inaccessible cliffs and outcrops, making it less vulnerable than many other Socotran endemics. Although the area of occupancy for the species is less than 50 km^2, given the negligible impact of livestock and of collecting by locals and the fact that many populations are inaccessible it is not especially 'prone to the effects of human activities (or stochastic events whose impact is increased by human activities) within a very short period of time in an unforseeable future' (IUCN 1994). Thus, *B. socotrana* does not meet the criteria for 'Vulnerable' as defined in either the current (IUCN, 1994) or the recommended changed version (IUCN/SSC Criteria Review Working Group, 1999) of the IUCN red list criteria and we recommend that it should therefore be placed in the Least Concern category.

B. samhaensis was also surveyed in detail during Spring 1999, and was found to be growing in quite dense groups where conditions were suitable at the original collection site (i.e., north to north-east facing vertical limestone faces or more southern aspects with shading overhangs), and possibly numbering up to 200

individuals. The increase in the number of plants compared to the single specimen seen in 1996 is due in part to the earlier timing of the 1999 trip, which managed to catch the end of the wet season. Approximately 30 more plants were found growing in two new sites on the northern edge of the island's limestone plateau at an altitude of 650 m, over 100 m lower than the original collection site. This raised the possibility that the vertical cliffs on the northern side of the island might also harbour *B. samhaensis*, though these are very difficult to survey. However, during the January 2000 expedition, an examination of these cliffs using binoculars failed to reveal any new sites, and it now seems likely that the total area of occupancy is restricted to the three known sites, in an area of 2 km by 500 m, which probably harbour less than 1000 plants. This small total population size and the fact that *B. samhaensis* exists only in a specific microclimate at the very highest parts of Samha do make the plant prone to the effects of human activities (e.g. livestock herders chewing the leaves) and stochastic events such as those due to climate change. This species should therefore be listed under the IUCN red list criterion VU D1, 2.

A revision of sect. *Peltaugustia* is presented here, in order to include a modified description of the section and to allow comparison to be made between the two species.

Sect. **Peltaugustia** (Warb.) Barkley, Phytologia 24: 156 (1972). *Begonia* sect. *Augustia* subsect. *Peltaugustia* Warb., in Engler & Prantl, Nat. Pflanzenfam., ed. 1, 3 (6a): 140 (1894). Type: *B. socotrana* Hook. f.

Perennial herbs. *Tuber* present or absent; bulbils crowded around stem bases, encased in papery bracts, inner scales fleshy. *Stipules* boat-shaped, persistent. *Leaves* peltate, ovate to orbicular, crenate-dentate, funnel-shaped around the insertion of the petiole, edges recurved, hypodermal layer present, stomata in clusters of 2—15. *Infloresence* a dichasial cyme, bracts boat-shaped, tepals pink. *Male flower*: tepals four, subequal to unequal; anthers distinctly hooded; filaments free. *Female flower*: bracteolate, tepals (5) 6, persistent, subequal; styles 3, bifid, stigmatic surface papillose and helically twisted or irregularly lobed; ovary 3-locular, 3-ribbed, one rib sometimes developed into a beak; placentae entire, triangular. Endemic to the Socotra archipelago.

Key to Begonia *section* Peltaugustia.

1a. Leaves ovate, male tepals unequal_____**1. B. samhaensis**

1b. Leaves orbicular, male tepals subequal_____**2. B. socotrana**

1. Begonia samhaensis M. Hughes & A.G. Mill., **sp. nov. Figs. 2 & 3.**

B. socotranae Hook. f. similis sed foliis late ovatis; floribus masculis tepalis non aequalibus; capsula alis aequalibus haud rostratis.

Type: Samha. Highest point of the island, shady north-facing cliffs, frequently mist covered, c. 750m, 16 ii 1999, *Miller* 17092 (E).

Perennial caulescent herb to 30 cm tall. *Tuber* irregular, pink in cross section, upper surface covered with bulbils encased in papery bracts. *Stipules* boat-shaped, not keeled, persistent, c. 13 x 13 mm, tip retuse to rounded, entire, with scattered short glandular hairs and longer (c. 1.5 mm) simple hairs present around the margin, papery when old. *Leaves* peltate; petiole centrally inserted perpendicular to the leaf blade, up to 6 cm long, fleshy, deep pink, with scattered short glandular hairs; leaf blade fleshy and succulent, brittle, asymmetric, ovate, base rounded, apex acute, up to 8 cm wide x 12 cm long, more commonly c. 5 cm wide x 7 cm long, with 6—8 palmate main nerves, funnel-shaped near the insertion of the petiole, hypodermal layer present; margin recurved, slightly undulate, crenate; the upper surface uniformly green, matt, glabrous, primary and secondary nerves distinctly sunken; the under surface paler green with scattered short glandular hairs, primary and secondary nerves prominent, the stomata in clusters of 5—15. *Inflorescence* a dichasial cyme; bracts persistent, in pairs, subtending each branching point, boat-shaped, not keeled, retuse to rounded, entire, with scattered short glandular hairs. *Male flowers*: buds purse-shaped; tepals 4, unequal, pink, glabrous; outer broadly orbicular, rounded at base, the edges slightly recurved, apressed, 15—22 mm long x 17—25 mm wide; inner obovate elliptic, 14—20 mm long x 8—14 mm wide, cuneate at base; stamens 30—45 in a globose cluster; anthers c. 1.5 mm long, hooded, narrowing towards their bases; filaments c. 1.5 mm long, free. *Female flowers*: tepals (5) 6, persistent, subequal, pink, glabrous, obovate, 10-18 mm long x 10—17 mm wide; styles 3; stigmatic surface irregularly lobed, bright yellow; ovary triangular in cross section,

3-locular, bracteolate, the bracteoles linear; wings 3, reduced, fleshy, subequal, semicircular, cordate at apex and base; placentae entire, thickened, triangular. *Fruit* pendulous, dehiscent either side of the wings.

Additional specimens examined. SAMHA. Summit of limestone plateau, sheltered cliffs c. 700m, 16 iii 1996, *Miller & Plana* 14208, (spirit material from cultivated specimen, E).

Notes. Endemic to Samha. *B samhaensis* has a restricted distribution, its entire range being the north western part of the high plateau on Samha in an area no more than 2 km by 500 m. It occurs in shaded cracks or pockets in north-facing vertical limestone faces from altitudes of 650 m up to the highest point on the island at 779 m. Its local name is 'seberbeher'. Local uses are as listed under *B. socotrana.*

FIG. 2. *Begonia samhaensis* M. Hughes & A.G. Miller., x 0.7. Photograph of a painting by Lizzie Sanders.

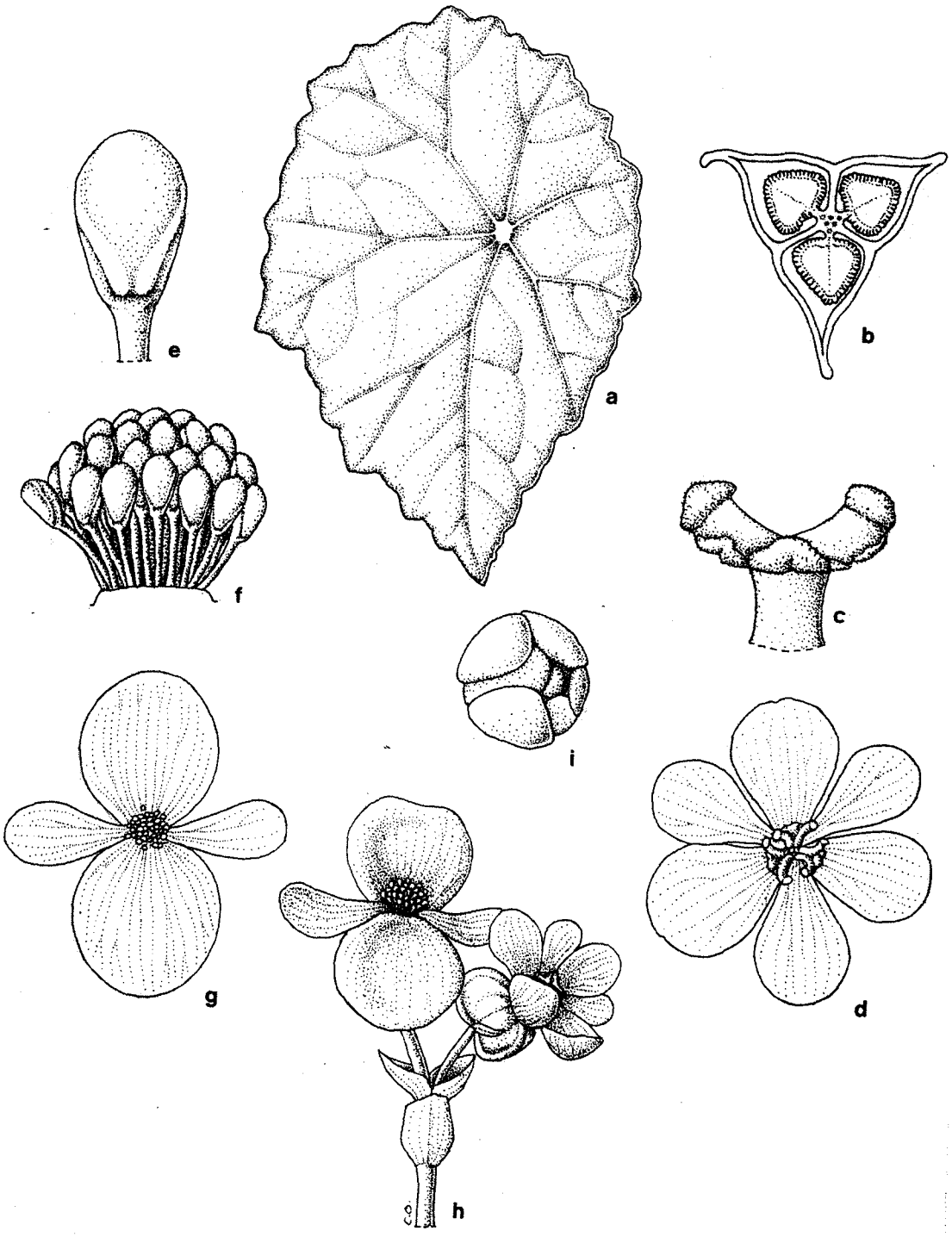

FIG. 3. *Begonia samhaensis* M. Hughes & A.G. Miller. (a), leaf, x 1; (b), cross-section of ovary, x 2; (c), stigma, x 4; (d), female flower, x 1; (e), anther, x 6; (f), androecium, x 4; (g), male flower, x 1; (h), part of inflorescence, x 1; (i), bulbil, x 2.

2. B. socotrana, Hook. f., Gard. Chron. 15: 8, fig. 1. (1881) & Bot. Mag. CVII, t. 6555 (1881); Irmscher Bot. Jahrb. 81: 123 (1962).

Type: Socotra, *I.B. Balfour* B.C.S. 419 (K).

Perennial caulescent herb, with contracted internodes at the base of the stem, to 45 cm tall. *Tuber* absent, bulbils encased in papery bracts crowded around stem base. *Stipules* shallowly boat-shaped, not keeled, persistent, c. 9 x 9 mm, tip rounded, with scattered short glandular hairs and longer (c.1.5 mm) simple hairs present around the margins, papery when old. *Leaves* peltate, basal ones appearing pseudo-rosulate; petiole centrally inserted perpendicular to the leaf blade, up to 20 cm long, fleshy, green, covered in short glandular hairs; leaf blade fleshy, orbicular, up to 20 cm diameter, most commonly around 10 cm diameter, sometimes shallowly lobed, 7—8 palmate main nerves, funnel-shaped near the insertion of the petiole, hypodermal layer present; margin recurved, crenate to crenate-dentate; the upper surface uniformly green, matt to slightly glossy, glabrous; the under surface paler green, with short glandular hairs, longer hairs present on the veins, the primary and secondary nerves distinctly sunken, the stomata in clusters of 2—8. *Inflorescence* a dichasial cyme; bracts persistent, in pairs, subtending each branching point; bracts shallowly boat-shaped, not keeled, apex rounded, margin shallowly denticulate to entire, covered in short glandular hairs with longer hairs present at tip. *Male flowers*: buds conical; tepals 4, imbricate, subequal to equal, deep pink, obovate to broadly obovate-orbicular, cuneate at base, 18—20 x 13—19 mm; stamens 25—35 in a globose cluster; anthers c. 1.5 mm long, hooded; filaments c. 1.5 mm long. *Female flowers*: tepals 6, persistent, subequal, obovate, 14—17 x 7—10 mm, deep pink, glabrous; styles 3, forked; stigmatic band helically twisted, bright yellow; ovary triangular trilobed in cross section, 3 locular, bracteolate, the bracteoles linear; wings 3, not fleshy, cordate to rounded at base with dorsal wing beaked, the beak sometimes reduced; placentae entire, thickened, triangular. *Fruit* pendulous, dehiscent either side of the wings.

Additional specimens examined. SOCOTRA. Reiged plateau, 4km SW of Hadiboh, thickets with grassland clearings on slightly north-dipping limestone plateau, dominated by *Boswellia ameero*, *Commiphora* sp., *Dracaena*, *Trichocalyx* sp., and *Croton socotranus*, shady cracks

in cliffs, flowers pink, bulbils at base, 740m, 21 ii 1989, *A.G. Miller et al.* M 8335 (E); Aduno Pass, small cliffs by spring, scrub dominated by *Cephalocroton & Hypericum* spp., flowers pink, leaves fleshy, 775m, 6 iii 1989, *A.G. Miller et al.* M 8667 (E); Muqadrihon Pass, c. 10 km SW of Hadiboh, granite slopes south of pass, deciduous woodland with *Buxus*, *Boswellia elongata*, *Commiphora elongata*, *Dracaena*, *Acacia pennivenia*, growing in shady damp cracks by spring, flowers pink, stems bulbiferous at base, very common in the area, 700m, 26 i 1990, *A.G. Miller et al.* M 10061 (E).

Notes. Endemic to the Haggier mountains and adjacent high limestone plateaus in the north east of Socotra. It occurs at altitudes from c. 700 m to 1500 m, growing mainly in shaded north-facing sites around the bottom of boulders and in crevices in rock faces, but also occurs terrestrially under the cover of montane shrubland. Its local name is 'seberbeher', with two variants, 'seberbeher sa'alhul' and 'seberbeher kikehe', for large-leaved and small-leaved plants respectively. The leaves and succulent petioles are eaten for their acidic taste and are considered a good tonic and stomach cleanser. The crushed leaves are used to make sour milk in the absence of a starter culture from a previous batch.

Acknowledgments.

The authors would like to thank Vanessa Plana for discovering *B. samhaensis*, Steve Scott, Fiona Inches and Andrea Fowler for their expertise in maintaining the living collections, Lizzie Sanders and Christina Oliver for providing the illustrations, Dr Miranda Morris for ethnobotanical information, Dr Robert Mill for assisting with the Latin diagnosis and Dr Pete Hollingsworth for comments on the manuscript. MH is particularly grateful to the M.L. MacIntyre *Begonia* Trust for providing funding for this study and its associated fieldwork. AGM would also like to thank the British Government Darwin Initiative, the Yemen Environmental Protection Council and GEF/UNOPS for providing support for research on Socotra.

References.

BARKLEY, F.A. (1972). Nomenclatural notes on *Begonia*. *Phytologia* 24: 156.

BELOUSOVA, L.S., DENISOVA, L.V. (1992). *Rare Plants of the World.* Rotterdam: A.A.Balkema

DE CANDOLLE, A. (1864). Begoniaceae. *Prodromus systematis naturalis regni vegetabilis* 15, 1: 266—408. Paris: V. Masson & Sons.

GLEED, C.J. (1961). Hybrid winter flowering Begonias. *J. Roy. Hort. Soc.* 80: 319—322.

HOOKER, J.D. (1881). New Garden Plants. *Gard. Chron.* n. s., 15: 8.

IRMSCHER, E. (1961). Monographische revision der Begoniaceen Afrikas 1. Sekt. Augustia und Rostrobegonia sowie einige neue Sippen aus anderen Sektionen. *Bot. Jahrb. Syst.* 81: 106—188.

IUCN, (1994). *IUCN Red List Categories*. As Approved by the 40th Meeting of the IUCN Council. Gland, Switzerland: IUCN.

IUCN/SSC Criteria Review Working Group. (1999). IUCN Red List Criteria Review Provisional Report: Draft of the Proposed Changes and Recommendations. *Species*, 31—32: 43—57.

KLOTSCH, J.F. (1854). Begoniaceen – Gattungen und Arten. *Abh. Konigl. Akad. Wiss. Berlin*, 1—135.

KOOPOWITZ, H. & KAYE, H. (1983). *Plant Extinction. A Global Crisis.* Washington: Stone Wall Press Inc.

LAVRANOS, J. & RADCLIFFE-SMITH A. (1969). List of rare and endangered Socotran and Abd al Kuri endemics, with an account of their present status, as far as could be determined from observations made in 1967. Middle East Command Expedition to Socotra, 1967. Unpublished.

LUCAS, G. & SYNGE, H. (eds). (1978). *The IUCN Plant Red Data Book.* Morges, Switzerland: IUCN.

POPOV, G.B. (1957). The vegetation of Socotra. *J. Linn. Soc., Bot.* 55: 706—720.

WALTER, K.S. & GILLIET, H.J. (eds). (1998). *1997 IUCN Red List of Threatened Plants*. Morges, Switzerland & Cambridge, U.K: IUCN.

WARBURG, O. (1894). Begoniaceae. In ENGLER, A. & PRANTL, K. (eds) *Die Naturlichen Pflanzenfamilien,* ed. 1, 3 (6a): 121—150. Leipzig: Wilhelm Engelmann.

Polymorphic microsatellite markers for the Socotran endemic herb *Begonia socotrana.*

M. Hughes[1], J. Russell[2] and P. M. Hollingsworth[1].

Molecular Ecology Notes, in press.

1. Royal Botanic Garden Edinburgh, 20a Inverleith Row, Edinburgh, EH3 5LR, UK.

2. Scottish Crop Research Institute, Invergowrie, Dundee, DD2 5DA, UK.

Abstract

Six polymorphic microsatellite markers have been developed to examine population structure and outcrossing rates in the narrow range endemic *Begonia socotrana*. Only two of the markers amplify products in its recently discovered sister species *B. samhaensis*. All of the loci amplify in winter flowering *Begonia* hybrids derived from *B. socotrana*, revealing little polymorphism and demonstrating the narrow genetic base of the material used in their production.

Keywords: *Begonia*, microsatellites, Socotra

Begonia socotrana (sect. *Peltaugustia*) is a bulbiferous herb endemic to the island of Socotra in the northern Indian Ocean, where it grows in sheltered north facing crevices in the Haggier mountains. It is listed by the IUCN as 'vulnerable' (Walter & Gilliet 1998). To investigate patterns of population genetic structure and outcrossing rates in this species, we have developed 6 polymorphic microsatellite markers.

DNA enriched for microsatellites hybridising to an $(AC)_{13}$ oligomer was isolated from *B. socotrana* following the method of White & Powell (1997), with the modification that no size selection was performed on the initial genomic DNA digest or on the post enrichment PCR fragments.

The enriched DNA was ligated into a ZAP Express *Eco*RI vector, followed by packaging using a ZAP Express Predigested Gigapack III Gold cloning kit (Stratagene). Plaques were lifted from a plating of the library using pasteur pipettes, placed in 500 µl of SM buffer (30 mM NaCl, 1 mM $MgCl_2$, 50 mM Tris pH 7.5, 0.01 % gelatin) with 20 µl of chloroform and left to diffuse overnight at 4°C. The resulting recombinant phage suspensions were screened for microsatellites using a three primer PCR as follows: 1 µl of the phage suspension was combined in a total volume of 10 µl with 1 µM M13 F primer, 1 µM M13 R primer, 1 µM of $(AC)_{13}$ oligomer, 1x PCR buffer (16 mM $(NH_4)_2SO_4$, 67 mM Tris-HCl (pH 8.8), 0.01% Tween-20), 2.5 mM $MgCl_2$, 0.2 mM dNTPs and 0.5 units of BioTaq (Bioline). An initial denaturing step at 94 °C for 5 minutes was followed by 35 cycles of 94°C for 30 seconds, 55°C for 30 seconds and 72°C for 1 minute, carried out in a GeneAmp 9600 thermocycler (Perkin Elmer). The PCR products were run on a 2% agarose gel; positive clones were indentified by the presence of more than one band. Plasmids were extracted from positive plaques following the manufatcurers protocol, and the sequence of the insert determined using a Thermosequenase II dye terminator cycle sequencing kit (Amersham) and an ABI 377 DNA sequencer.

Approximately 25% of the colonies screened contained microsatellites, and 24 (out of 75) were chosen for primer design, which was carried out using Primer-3 (Rozen & Skaletsky 1998) with the modified parameters of Beasley *et al.* (1999) as

the starting point. Three clones were discarded due to the possibility of them being chimeric, as they showed varying combinations of identical flanking regions around different microsatellites; these products may be the result of premature strand termination and subsequent priming off other microsatellites during the PCR bulking of the enriched DNA.

The primers were tested in 10 µl PCR reactions containing 10 ng genomic DNA, 1x PCR buffer (16 mM $(NH_4)_2SO_4$, 67 mM Tris-HCl (pH 8.8), 0.01% Tween-20), 2.5 mM $MgCl_2$, 0.2 mM dNTPs, 0.5 µM of each primer and 0.5 units of BioTaq (Bioline). The reactions were denatured at 95 °C for 7 minutes, followed by 30 cycles of 95°C for 30 seconds, 55 °C for 15 seconds, 72°C for 30 seconds, using a GeneAmp 9600 thermocycler (Perkin Elmer).

Of 24 primers pairs tested, 6 gave no product, the remaining 18 amplified products which were close to the size of the cloned sequence when run on a 2% agarose gel. Initial population screens of these 18 primer pairs were undertaken by including 4µM TAMRA labelled dCTP (PE Applied Biosystems) in the PCR and analysing the products on an ABI 377 DNA sequencer. Twelve primer pairs gave banding patterns which could not be interpreted as single loci, and one was monomorphic. The remaining 6 loci produced one or two bands per individual and were polymorphic; one primer from each pair was fluorescently labelled for use in further population genetic surveys (Table 1).

Although an $(AC)_{13}$ oligomer was used for the pre-cloning enrichment, other repeat motifs were found in the cloned products, such as $(CT)_n$ repeats (e.g. locus B17b) and a $(CTCACA)_6$ repeat located in an open reading frame (locus B128). Primers for one of the loci (B125) amplified products *ca* 120 bases shorter than expected, and sequencing of the products obtained from population samples showed this was due to a drastic shortening of the microsatellite region.

Of the six loci, only two amplified products (both monomorphic) from the sister species of *Begonia socotrana* which occurs on the neighbouring island of Samha (*Begonia samhaensis*; M Hughes & A.G Miller, in press). All loci amplified

products in the winter flowering *Begonia* Hiemalis cultivars, which are derived from a cross between *B. socotrana* and *Begonia* x Tuberhybrida. Only two of the loci were polymorphic in the cultivars (B17b and B226), both of which showed two alleles, present only in the homozygous state (Fig. 1). No products were amplified from the other *Begonia* species tested (*B. sutherlandii, B. geranioides, B. dregei* and *B. fallax*).

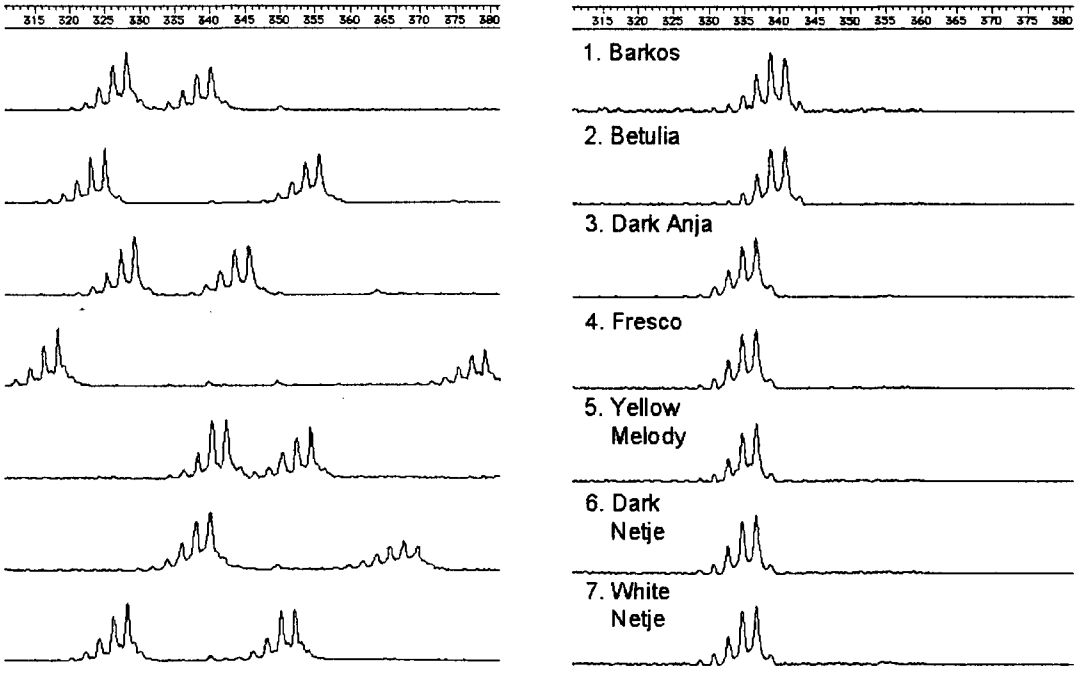

Figure 1. Locus B17b showing allellic variation among seven wild individuals of *Begonia socotrana* (left) and among seven *Begonia* Hiemalis cultivars (right), representing the main four groups in cultivation (1 & 2, Barkos group; 3, Rosanna group; 4 & 5 Schwabenland group; 6 & 7 Ilona group).

Table 1. Characterisation of six microsatellite primers in *Begonia socotrana*. H_O, observed heterozygosity; H_E, expected heterozygosity; *n*, number of individuals genotyped.

Locus	Repeat in clone	Primer sequence (5'-3')	Size range (bp)	No. alleles	H_O	H_E	(*n*)	GenBank no.
B17b	$(CT)_{26}..(CT)_{10}..(CT)_{13}$	TCCCCGATATTCCAACATATCAC ATGATTGGACCCCGTATCACAT*	300-386	31	0.809	0.895	133	AF403057
B63	$(AT)_3(AC)_8$	CTTAAGCTTCATACTCCAATCAC* GTTTTGAACTTGAGAATACTAGTGAG	176-190	4	0.099	0.137	94	AF403054
B128	$(CTCACA)_6$	TTCCCTTTGACAGTTTGTTGTT* AATTTCGGTAATCAGCAGACAGG	148-172	6	0.502	0.524	143	AF403052
B130	$(AC)_{24}$	GCACCTCCTTTTGATGATACACC* CCTAGTCTCTTCACTTATCACAAGGT	105-125	11	0.680	0.704	140	AF403053
B215	$(AC)_{25}(AT)_6$	CGCGTTAAAAATATGTGAAGCAC TACTATGTGGCAAGCCTCAAACA*	73-81	6	0.520	0.547	143	AF403055
B226	$(AC)_9$	GGACGGTGTTTAGGCCTTTCTAT* CAATAGTTGTGGATGCAAGGTGA	163-181	7	0.688	0.685	140	AF403056

*Fluorescently labelled primer.

Acknowledgements

The authors would like to thank S. Scott for his expertise in maintaining the living collections, M. L. Hollingsworth, A. Ponge and R. Hollands for technical support, T. J. H. Hoogkamp for supplying cultivar material and A. G. Miller for organising fieldwork. This work was funded by the M. L. MacIntyre Begonia trust, to whom MH is particularly grateful.

References

Beasley EM, Myers RM, Cox DR & Lazzeroni LC (1999) Statistical refinement of primer design parameters. In *PCR Applications* (ed. Innis MA), pp. 55-71. Academic Press, San Diego.

Hughes M & Miller AG (In press) A new endemic species of *Begonia* (Begoniaceae) from the Socotra archipelago. *Edinburgh Journal of Botany.*

Rozen S & Skaletsky HJ (1998) *Primer 3.* Code available at http://www-genome.wi.mit.edu/cgi-bin/primer/primer3.cgi

Walter KS & Gilliet HJ (eds) (1998) *1997 IUCN Red List of Threatened Plants*, p. 73. IUCN, Morges, Switzerland & Cambridge.

White G & Powell M (1997) Isolation and chracterization of microsatellite loci in *Swietenia humilis* (Meliaceae): an endangered tropical hardwood species. *Molecular Ecology*, **6**, 851-860.

Isolation of polymorphic microsatellite markers for *Begonia sutherlandii* Hook. f.

M. Hughes, P. M. Hollingsworth & J. Squirrell

Molecular Ecology Notes, in press.

Royal Botanic Garden Edinburgh, 20a Inverleith Row, Edinburgh, EH3 5LR, UK.

Abstract

Seven polymorphic microsatellite loci have been characterised for investigating population structure in the patchily distributed herb *Begonia sutherlandii*. Two loci (BSU3 & BSU4) exhibited population specific null alleles; primer redesign and allele sequencing for one of these loci showed two transition mutations in the original primer site. Two loci exhibited imperfect repeat polymorphisms due to single base pair indels in the flanking region (locus BSU6) and in the microsatellite region itself (BSU7). Transversion mutations were also found in the microsatellite region of locus BSU7. The remaining three loci amplified in all individuals tested and appeared to conform to a simple stepwise mutation pattern.

Keywords: *Begonia*, microsatellites, Kwazulu-Natal

Begonia sutherlandii is one of the most widespread *Begonia* species in Africa, with an east Afromontane distribution from Tanzania to the northern part of the Transkei in South Africa. It occurs in a narrow habitat range, on wet and shaded steep slopes, especially in proximity to rivers or waterfalls in forested areas. Given its habitat preference and the naturally fragmented distribution of montane forest in eastern and southern Africa (Eeley *et al.* 1999; Lawes 1990), *B. sutherlandii* occurs sporadically, often with considerable distances of unsuitable semi-arid habitat separating populations.

The amount of gene flow between disjunct populations is an important factor in determining to what extent and how quickly populations can become adapted to local conditions, as gene flow from 'foreign' populations can retard differentiation (Barton, 2001). Species of *Begonia* appear to have poor powers of pollen and seed dispersal, leading to the expectation that populations may become isolated over relatively short distances. This may be one of the factors leading to the evolution of high species diversity in the genus (*Begonia* contains about 1400 species). Microsatellite markers have been developed to examine population genetic structure in *B. sutherlandii*, to establish over what scales population differentiation occurs.
DNA enriched for microsatellite sequences was obtained following a method based on Edwards *et al.* (1995) and Squirrell & Wolff (2001), with the modifications that *Tsp509*I restriction enzyme (AATT; New England Biolabs Inc.) and the *Tsp509*I PCR adapters of White & Powell (1997) were used. The nylon membranes to which the genomic DNA was hybridised were prepared with 6 μg each of $(GA)_{13}$, $(CA)_{13}$ and $(AAG)_8$ per single piece of 8 x 8 mm Hybond® N+ membrane.

The enriched DNA was cloned using a PCR-Script™ Amp Cloning Kit (Stratagene), and the sequence of the insert in recombinant clones was determined using a Thermosequenase II dye terminator cycle sequencing kit (Amersham) and an ABI 377 sequencer. 110 clones were sequenced, and ca. 70 % contained microsatellites of 8 or more repeat units. $(AC)_n$ was the most common motif, accounting for approximately 70% of the microsatellites found. 33 sequences were chosen for primer design, which was carried out using Primer-3

(http://www.path.cam.ac.uk/cgi-bin/primer3.cgi) with the modified parameters of Beasley *et al.* (1999) as the starting point.

The primers were tested in 10 µl PCR reactions containing: 10 ng genomic DNA, 1x PCR buffer (16 mM $(NH_4)_2SO_4$, 67 mM Tris-HCl (pH 8.8), 0.01% Tween-20), 2.5 mM $MgCl_2$, 0.2 mM dNTPs, 0.5 µM of each primer and 0.5 units of BioTaq (Bioline). The reactions were denatured at 95°C for 7 minutes, followed by 30 cycles of 95°C for 30 seconds, $T_{annealing}$ °C for 20 seconds, 72°C for 30 seconds, finishing with a final extension step at 72°C for 15 minutes, using a GeneAmp 9600 thermocycler (Perkin Elmer). The products were visualised on a 2% agarose gel. Out of 33 primer pairs, 5 gave no product, 1 gave several bands and 27 gave a product consisting of one or two bands of size similar to that in the clone. Of these 27 loci, only 7 produced one or two bands per individual and were polymorphic when the profiles were examined in detail by including 4µM TAMRA labelled dCTP (PE Applied Biosystems) in the PCR and analysing the products on an ABI 377 sequencer. One of each of these primer pairs was labelled using either FAM, TAMRA or JOE (PE Applied Biosystems) for use in screening populations.

Loci BSU4 and BSU3 exhibited population specific null alleles, where none of the individuals in one (BSU4) or two (BSU3) populations gave a PCR product. The primers for BSU4 were redesigned (5'-3': forward AATCTCTTGAGATG-GAGGAAACA, reverse GTTGGTAACTTGGTATGGTGGA; original primer sequences are shown in table 1) to anneal outside the original primers, and these successfully amplified products in the population previously exhibiting nulls. Sequencing the product showed two transition (C-T) mutations 2 and 12 bases from the 3' end of the original reverse primer binding site. The clone from which the primers for BSU3 were designed was too short to permit the design of further primers. The products of two loci (BSU6 and BSU7) showed alleles with single base pair length differences; sequencing alleles from homozygous individuals showed this was due to an extension of one base pair in a T_8 region adjacent to the $(TC)_{18}$ microsatellite in locus BSU6, and the loss of one cytosine from the CC motif in locus BSU7; this locus also showed two transversion mutations (C-A) in the microsatellite region in some individuals.

Table 1. Characterisation of seven microsatellite primers in *Begonia sutherlandii*. . H_O, observed heterozygosity; H_E, expected heterozygosity; *n*, number of individuals genotyped.

Locus	Repeat in clone	Primer sequence (5'-3')	T_{an}	Size range (bp)	No. alleles	H_E	H_O	(*n*)	GenBank no.
BSU1	$(CT)_{10}$	AAAAGCCTTACTATATAATGACAA CGACCAAGAAAATAAATGAAAT	55	100-122	10	0.531	0.428	245	AF467454
BSU2	$(AG)_{14}$	CCCTTTCTCTTACCCGTTTCCTT TCATAACCAAACCCAATCTCACC	55	114-140	14	0.539	0.458	256	AF467455
BSU3	$(CT)_{18}$	CATGGCTCTAGTAGTTTCTTCCATTT GTAGTGCAACGGCAATGATGAC	55	79-105	11	0.441	0.403	156	AF467456
BSU4	$(CT)_{14}$	TGGAGGAAACATATCACGAAGAAA CCAAGTCTTATGGAAGGATGAACA	55	120-144	14	0.230	0.141	224	AF467457
BSU5	$(AG)_{12}$	GTCTTTCTCAACCCCACAGACAA GACCTGTCCATTTGCAAAATCTC	55	148-199	22	0.382	0.356	257	AF467458
BSU6	$(TC)_{18}$	CTCTGGGCTAATAACCATACC CTAGTAAGATCATTTACAGATACGA	53	162-198	21	0.623	0.588	237	AF467460
BSU7	$(CT)_8CC(TC)_8$	TGTCTCTGCAGAATATGTTCACT TTTAACCAGGCCATGAATGTT	53	134-163	9	0.231	0.186	210	AF467459

Acknowledgements

The authors would like to thank M. L. Hollingsworth, A. Ponge, G. French and R. Hollands for technical support and N. Watherston for maintaining living material. The Royal Botanic Garden Edinburgh is supported by the Scottish Executive Environment and Rural Affairs Department. Mark Hughes is funded by the M.L. MacIntyre Begonia Trust.

References

Barton NH (2001) The evolutionary consequences of gene flow and local adaptation: future approaches. In *Dispersal* (eds. Clobert J, Danchin E, Dhondt AA & Nichols JD), pp.329-340. Oxford University Press, Oxford.

Beasley EM, Myers RM, Cox DR & Lazzeroni LC (1999) Statistical refinement of primer design parameters. In *PCR Applications* (ed. Innis MA), pp. 55-71. Academic Press, San Diego.

Edwards KJ, Barker JHA, Daly A, Jones C & Karp A (1995) Microsatellite libraries enriched for several microsatellite sequences in plants. *Biotechniques*, **20**, 758-760.

Eeley HAC, Lawes MJ & Piper SE (1999) The influence of climate change on the distribution of indigenous forest in KwaZulu-Natal, South Africa. *Journal of Biogeography*, **26**, 595-617.

Lawes MJ (1990) The distribution of the samango monkey (*Cercopithecus mitis erytharchus* Peters, 1852 and *Cercopithecus mitis labiatus* I. Geoffroy, 1843) and forest history in southern Africa. *Journal of Biogeography*, **17**, 669-680.

Squirrell J & Wolff K (2001) Isolation of polymorphic microsatellite loci in *Plantago major* and *P. intermedia*. *Molecular Ecology Notes*, **1**, 179-181.

White G & Powell M (1997) Isolation and chracterization of microsatellite loci in *Swietenia humilis* (Meliaceae): an endangered tropical hardwood species. *Molecular Ecology*, **6**, 851-860.

Population genetic structure in the island endemic *Begonia socotrana*

M. Hughes, P.M. Hollingsworth & A.G. Miller

Abstract

Begonia socotrana is a bulbiferous perennial herb endemic to the Haggeher Mountain region on the island of Socotra in the Arabian Sea. Its total distribution falls within an area of only *ca.* 15 km x 10 km. Population genetic surveys were carried out to investigate patterns of reproduction and dispersal to provide baseline data for conservation programmes. A total of 158 individuals was sampled from 10 populations and screened for variability at 5 nuclear microsatellite loci. The species is not panmictic across its range (R_{ST}=0.081, P<0.01; θ=0.096, P<0.01) and there was evidence for significant isolation-by-distance. Despite the plants producing prolific bulbils, the vast majority of samples had distinct multilocus genotypes suggesting that sexual rather than asexual reproduction is the major means of reproduction. Estimates of outcrossing rates from progeny arrays (multilocus t=0.83 ± 0.10), coupled with estimates of the inbreeding coefficient from field populations (f = 0.051±0.017) suggest that the species is predominantly outcrossing.

Keywords: *Begonia*; Socotra; microsatellites; SSRs; conservation genetics

1. Introduction

The Socotra archipelago is a group of four islands of continental origin in the Arabian Sea that possess a remarkable relictual flora with high levels of endemism. Approximately 850 species of vascular plant occur on the islands of which 297 (35%) are endemic (Miller and Morris, 2001). This degree of endemism is comparable to that found on the Canaries (26%; Nieves *et al.*, 1986) and the Galapagos (39%; Wiggins and Porter, 1971), and the Socotra archipelago has been declared a 'Special Protected Area' by the Yemeni government. It is also in the process of being declared a 'Man and Biosphere Reserve' and 'World Heritage Site' by UNESCO (United Nations Educational, Scientific and Cultural Organisation).

The main island in the group is Socotra, which lies some 380 km south of mainland Arabia and 210 km east of Somalia (Fig. 1). The island is approximately 135 km long by 42 km wide and has an area of some 3799 km^2, with three main physiographic zones being recognisable: coastal plains, limestone plateaux (ranging from 300-800m in altitude) and the Haggeher mountains (750-1550m). The granitic Haggeher are thought to represent land which has been above sea level since the Cretaceous, with the surrounding limestone areas dating to the mid-Tertiary (Beydoun and Bichan, 1970). The climate is strongly monsoonal, with the north-easterly winter monsoon bringing most of the islands rainfall. The summer monsoon brings rain to the Haggeher and higher south-west facing slopes, and these areas also receive moisture during the summer in the form of mist and heavy dews. The vegetation on Socotra ranges from sparse scrub in coastal areas through succulent shrublands to more mesic montane vegetation in the Haggeher and high limestone plateaux.

Island floras are notoriously fragile and can be decimated by the introduction of grazing animals. An expedition to Socotra by Lavranos and Radcliffe Smith in 1967 (Lavranos and Radcliffe-Smith, 1969) suggested that this was an immediate threat with many plant species considered in danger of extinction due to vast uncontrolled grazing by goats which were being allowed to roam freely over the island.

Begonia socotrana Hook. f. is one of the flagship plants of the archipelago and serves as a symbol for conservation awareness. It is a monoecious perennial herb endemic to Socotra, where it has been recorded from the higher regions of the Haggeher mountains (Balfour, 1888). The species is also known from the adjacent limestone plateau of Reiged, where it was recorded as being among the commonest of the plants restricted to soil pockets in rock crevices (Popov, 1957). Since these early records, however, *B. socotrana* has been considered to be very rare in the wild. The 1967 expedition of Lavranos and Radcliffe-Smith recorded the plant from only two of the highest peaks of the Haggier mountains (Lavranos and Radcliffe-Smith, 1969 p3). Based largely on the reports from this expedition, the 1978 IUCN Plant Red Data Book (Lucas and Synge, 1978) lists *B. socotrana* as 'endangered'. Lucas and Synge (1978, p. 79) noted that "the population of this island endemic has reached critically low levels...it is now confined to high altitude, mountain pinnacles virtually inaccessible to man and goat". Koopowitz and Kaye (1983, p. 63) stated that "the population of begonias has steadily eroded" and Belousova and Denisova (1992, p. 323), described the plant's populations as at a "critically low level." Lucas and Synge (1978), Koopowitz and Kaye (1983) and Belousova and Denisova (1992) all cite overgrazing by goats as the main threat to the species. The latest IUCN Red list of threatened plants (Walter and Gilliet, 1998, p. 73) lists *B. socotrana* as 'vulnerable'.

In 1999 the Socotran islands were the subject of a GEF (Global Environment Facility)/UNOPS (United Nations Office for Project Services) funded biodiversity inventory exercise in preparation for the delimitation of conservation areas. As part of this biodiversity inventory, a detailed study of *Begonia socotrana* was undertaken. One of the goals of this work was to undertake general population surveys to determine its current distributional range and abundance (Hughes and Miller, 2002). An additional goal, and the subject of this paper, was to gain some insights into the reproductive biology and population structure of *Begonia socotrana.* An understanding of how populations reproduce and maintain themselves, and the spatial scales over which gene flow occurs, is important for both *in situ* and *ex situ* conservation programmes. *Begonia socotrana* is monoecious, although fully self-fertile in cultivation. The species also produces bulbils at the base of the plant, which represent a potential vegetative mechanism for perpetuation and dispersal. Our goal

was to establish whether the plants are predominantly inbreeding or outbreeding, whether they reproduce predominantly sexually or asexually, and to what extent populations are connected with one another.

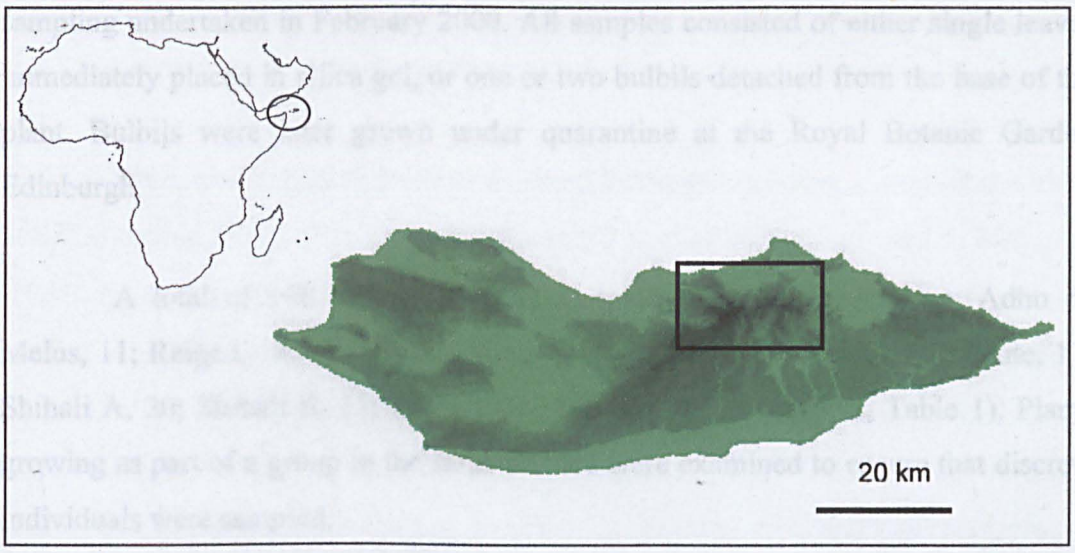

Fig. 1. The location of the Socotra archipelago (left). Topographic map of Socotra (right). The area within the box is the location of the Haggeher mountains and adjacent plateaux Reiged and Rewged, which are shown in detail in Fig. 2.

2. Materials and methods

2.1. Population sampling

Populations of *B. socotrana* were sampled in February 1999, with further sampling undertaken in February 2000. All samples consisted of either single leaves immediately placed in silica gel, or one or two bulbils detached from the base of the plant. Bulbils were later grown under quarantine at the Royal Botanic Garden Edinburgh.

A total of 158 individuals was sampled from 10 populations: Adho de Melus, 11; Reiged, 19; Rewged, 7; Mugudrihon, 12; Dicksam, 16; Skand route, 17; Shihali A, 20; Shihali B, 17; Shihali C, 18; Shihali D, 21 (Fig. 2, Table 1). Plants growing as part of a group in the same crevice were examined to ensure that discrete individuals were sampled.

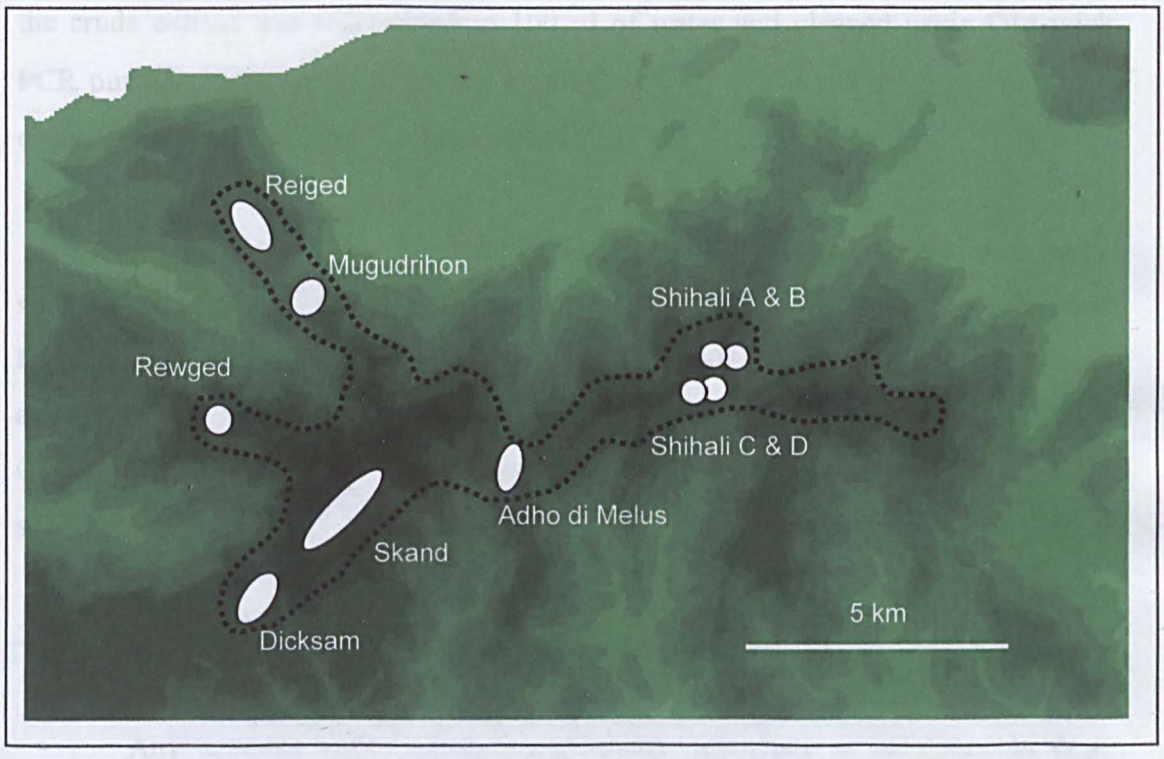

Fig. 2. Detailed map of the location of sampled populations of *B. socotrana* shown in Fig. 1. The population pairs from Shihali (A+B and C+D) are 500 m apart in altitude. The estimated entire range of *Begonia socotrana* is indicated by a dashed line. Shading indicates topography, ranging from sea level to 1500 meters in altitude with contours at approximately 200 m intervals.

Single ripe fruits were collected from 5 selected plants to provide progeny arrays for estimates of outcrossing rates (mean of 12.8 progeny per maternal plant; range 10-15). The only populations that had fruits at the right stage of development during the fieldwork were Reiged (4 maternal plants) and Rewged (1 maternal plant).

2.2. DNA extraction and microsatellite analysis

DNA was extracted from silica dried leaf material using a method modified from Doyle and Doyle (1987). Approximately 1 cm^2 of leaf was ground in 800 µl of CTAB buffer with 2.5 % β-mercaptoethanol, 0.5 % PVPP and a pinch of acid washed sand, and then incubated at 65°C for 30 minutes. After washing with 500 µl of 24:1 chloroform:isoamylalcohol, the aqueous layer was removed and DNA precipitated with 600 µl of freezer-cold isopropanol. DNA isolated from *Begonia* species can contain PCR-inhibiting levels of co-isolated salts and carbohydrate, so the crude extract was redissolved in 100 µl of water and cleaned using QIAquick PCR purification columns (Qiagen), eluting in 50 µl of warm (65°C) EB buffer to ensure recovery of high molecular weight DNA.

Five polymorphic microsatellite loci were used for the population genetic surveys, using fluorescently labelled primers (B17b, B128, B130, B215, B226) and PCR conditions as detailed by Hughes *et al.* (2002). The PCR products were analysed on an ABI 377 DNA sequencer, followed by gel image analysis using ABI Genescan® software (version 3.1.2) and allele size scoring using ABI Genotyper® software (version 2.0).

2.3. Data analysis

Any samples with identical multi-locus genotypes at all five loci (e.g. samples that may have been ramets of the same genet) were only represented by a single sample in all population genetic analyses. Microsatellite data were formatted for input into various population genetic software programmes using the Microsatellite Toolkit (Park, 2001). Basic descriptive population genetic statistics (*A*

= mean number of alleles per locus; H_E = expected heterozygosity (or gene diversity) and H_O = observed heterozygosity) were calculated using GDA (Lewis and Zaykin, 2001).

Population genetic structure was assessed using Weir and Cockerham's (1984) estimates of Wright's *F*-statistics. Estimates of F_{IS} (*f*; deviation from panmixia attributable to non-random mating within populations), F_{ST} (θ; deviation from panmixia attributable to non-random mating among populations), and the inclusive measure F_{IT} (*F*; deviation from panmixia attributable to non-random mating within and between populations) were estimated using FSTAT (Goudet, 2001). Permutation tests (>10 000 permutations, randomising alleles) were used to test whether the estimates of F_{IS} and F_{IT} were significantly different from the null hypothesis of panmixia. Per locus estimates of F_{ST} were tested for significance by jackknifing across populations; the global estimate of F_{ST} was tested for significance by jackknifing across loci. Per locus and species wide estimates of R_{ST} (Slatkin, 1995), an analogue of F_{ST}, were carried out using RST CALC (Goodman, 1997). R_{ST} incorporates information on allele size differences into estimates of population differentiation and was derived for loci such as microsatellites that evolve under a stepwise mutation model. Permutation tests (10 000 permutations, randomising alleles) were used to test whether the global and per locus estimates of R_{ST} were significant.

Significance values were corrected for multiple tests using the sequential Bonferroni test (Rice, 1989).

Arlequin v2.000 (Schneider *et al.*, 2000) was used to carry out a Mantel test to test for isolation by distance, with the distance matrices being natural log of direct geographic distance (calculated from GPS data) and $F_{ST}/1\text{-}F_{ST}$ (Rousset, 1997) or $R_{ST}/1\text{-}R_{ST}$. Microsat (Minch *et al.*, 1995) was used to create a population pairwise distance matrix for the genetic distance measure 1-*Dps* (Bowcock *et al.*, 1994); a neighbour joining tree was constructed from this matrix using PAUP* 4.0 (Swofford, 1998).

Outcrossing rates were estimated from progeny array genotypes using MLTR version 1.1. (Ritland, 1996) using the expectation-maximisation option. The outcrossing rate was also estimated from the global estimate of F_{IS} using the equation $t=(1-F_{IS})/(1+F_{IS})$ (Allard *et al.*, 1969)

3. Results

3.1. Descriptive statistics

All microsatellite loci were polymorphic in all populations. At the species level the number of alleles at each locus ranged from A = 5-32, with the mean gene diversity per locus ranging from H_E = 0.61-0.93 (Table 1). Within populations the mean number of alleles per locus ranged from A = 4-7.2, and the mean gene diversity ranged from H_E = 0.6-0.84 (Table 2).

3.2. Sexual versus asexual reproduction

Individuals with identical multi-locus genotypes were detected at three sites (Adho, 5 out of 11 individuals; Mugudrihon, 2 out of 13 individuals; Rewged, 2 out of 8 individuals). However, the vast majority of individuals possessed distinct multilocus genotypes (149 out of 158).

3.3. Estimation of breeding behaviour

A species wide estimate of F_{IS} (f = 0.051 $\pm$ 0.017, P<0.05) revealed a small but significant deficit of heterozygotes due to deviations from random mating within populations. Within individual populations, estimates of the inbreeding coefficient range from f = -0.185 to 0.313 (Table 2). Of these, only one, (Shihali B) showed an inbreeding coefficient estimate that was significantly different from zero.

Estimates of the multilocus outcrossing rate, t, ranged from 0.48 to 1.00, with a mean of 0.86 (Table 3). These values are similar to the value of t estimated from F_{IS} , t = 0.90.

3.4. Population genetic structure

There was significant population structure in *B. socotrana*, with a moderate amount of diversity being partitioned between populations (θ = 0.096, R_{ST} = 0.081; Table 1). A Mantel test revealed significant isolation by distance (P<0.01, R^2 = 0.12

for the $F_{ST}/1\text{-}F_{ST}$ matrix; $P<0.05$, $R^2 = 0.04$ for the $R_{ST}/1\text{-}R_{ST}$ matrix) (Fig. 4). Evidence for a geographical component to the distribution of genetic variation is also evident from a Neighbour joining analyses of inter-population genetic distances; geographically proximal populations tend to cluster together in the tree (Fig. 3).

Table 1.

Per-locus population genetic statistics of *Begonia socotrana*. *A*, mean number of alleles per locus; H_E, expected heterozygosity; H_O, observed heterozygosity; *f* and *F*, Weir and Cockerham's (1984) estimates of the inbreeding coefficients F_{IS} and F_{IT}; R_{ST}, Slatkin's (1995) stepwise mutation model estimate of F_{ST}; θ, Weir and Cockerham's (1984) estimate of F_{ST}.

Locus	*A*	H_E	H_O	*f*	R_{ST}	θ	*F*
B128	5	0.68	0.53	0.044	0.111**	0.202**	0.238**
B130	12	0.73	0.66	0.050	0.102*	0.062*	0.109*
B226	7	0.72	0.68	0.008	0.046*	0.066*	0.073
B17b	32	0.93	0.81	0.094*	0.025	0.050*	0.139**
B215	8	0.61	0.52	0.037	0.120**	0.132**	0.164**
All loci	12.8	0.74	0.64	0.051*	0.081**	0.096**	0.143**

** $P<0.01$

* $P<0.05$

Table 3

Estimates of the outcrossing rate ($t \pm$ S.E.) in *B. socotrana*.

Array number	Multilocus outcrossing rate (± S.E.)
Family 1	$t = 0.48 \pm 0.13$ ($n = 13$)
Family 2	$t = 0.98 \pm 0.08$ ($n = 12$)
Family 3	$t = 0.84 \pm 0.11$ ($n = 10$)
Family 4	$t = 1.00 \pm 0.00$ ($n = 14$)
Family 5	$t = 1.00 \pm 0.00$ ($n = 15$)
Over all families	$t = 0.83 \pm 0.10$

Table 2

Population level statistics of *Begonia socotrana*. *n*, mean number of individuals sampled per locus after removing individuals with identical genotypes; *A*, mean number of alleles per locus; H_E, expected heterozygosity; H_O, obeserved heterozygosity; *f*, Weir and Cockerham's (1984) estimate of the inbreeding coeffiecient F_{IS}.

Population	Latitude	Longitude	Altitude	*n*	*A*	H_E	H_O	*f*
Adho	12.573	54.048	950	7	4.0	0.60	0.66	-0.095
Reiged	12.619	54.001	750	19	7.2	0.73	0.78	-0.064
Dicksam	12.543	53.994	1050	10.4	6.8	0.84	0.77	0.093
Skand	12.573	54.020	1300	16.8	6.6	0.76	0.71	0.072
Mugud	12.606	54.006	700	11.6	5.6	0.66	0.63	0.049
Rewged	12.582	53.994	700	6.6	4.2	0.62	0.74	-0.185
ShihaliA	12.572	54.104	800	17.4	5.0	0.59	0.55	0.096
ShihaliB	12.569	54.105	1300	14.6	4.4	0.49	0.35	0.313*
ShihaliC	12.584	54.114	800	11	4.2	0.66	0.58	0.130
ShihaliD	12.584	54.111	1300	16	5.8	0.67	0.66	0.020
Mean				13.7	5.38	0.66	0.64	

(*P<0.05)

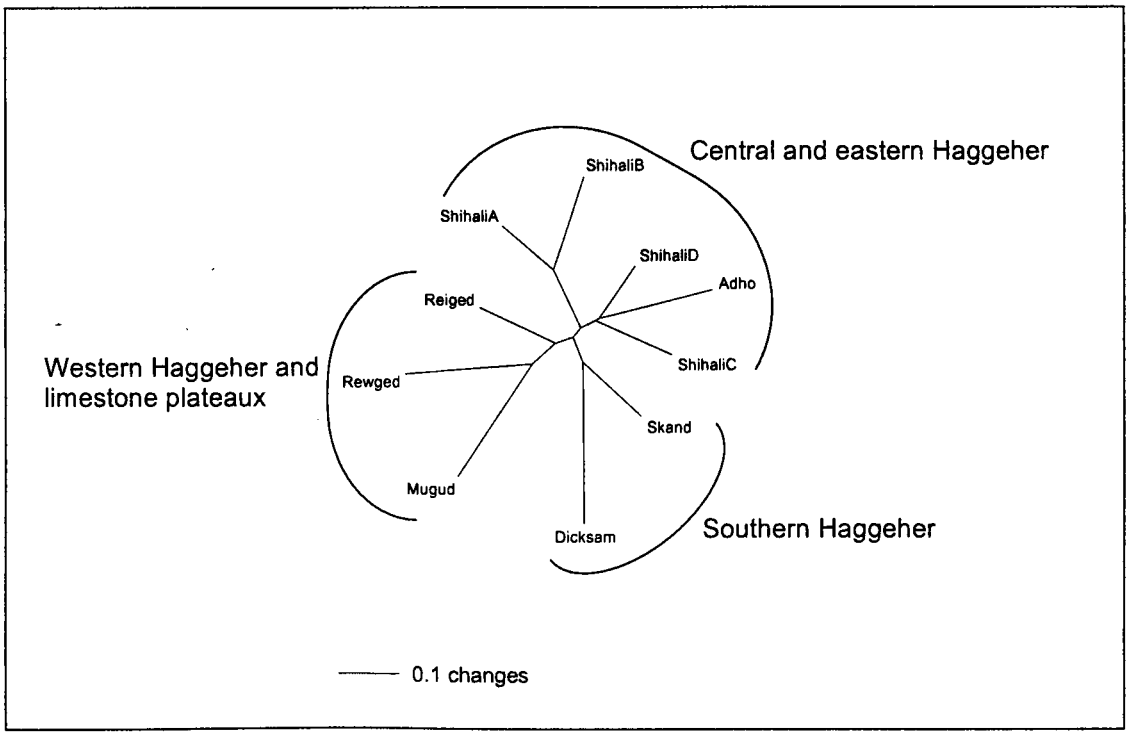

Fig. 3. Neighbour joining tree generated from pairwise 1-D_{PS} distances of *Begonia socotrana* microsatellite data.

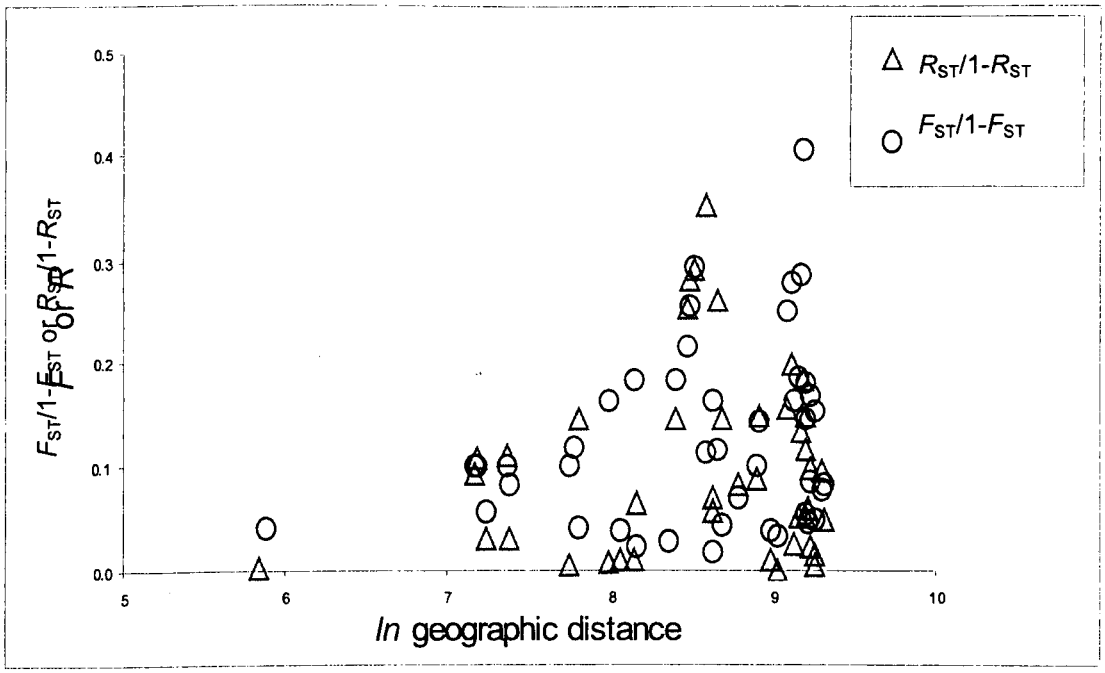

Fig. 4. Evidence for isolation by distance in *B. socotrana*. A Mantel test shows the relationship to be significant (P<0.01 for F_{ST}/1-F_{ST}; P<0.05 for R_{ST}/1-R_{ST})

4. Discussion

4.1. Population structure

During the field work for the current study our population surveys revealed that *B. socotrana* is more abundant within its range than previous records suggested (Hughes and Miller, in press). The apparent rarity of the species appears to be attributable to the 1967 Lavranos and Radcliffe Smith expedition taking place during the dry season when the plants die back to inconspicuous bulbils. In the course of our work the plant was found at a new site (Dicksam), and in healthy populations at previously recorded localities. Within the area of its range (*ca.* 15 km x 10 km) the plant is relatively continuously distributed (at least in the sense that the maximum gaps between populations are about 1 km). It is largely restricted to north facing crevices between boulders and on cliffs, although the aspect becomes less critical with altitude and at *ca.* 1400 m plants grow in quite exposed south facing sites. It also occurs terrestrially under the cover of montane shrubland, and forms patches where in some cases it is the dominant ground cover. Despite this relatively continuous distribution within a small total range, the estimates of population genetic structure indicate a significant deviation from panmixia (θ= 0.096, R_{ST}= 0.081; Table 1). There is a degree of genetic/geographic clustering in the neighbour joining tree (Fig. 3), and evidence for significant isolation-by-distance (Fig. 4).

There are three possible contributing factors that might explain these patterns of genetic variability. Firstly, *B. socotrana* produces seeds *ca.* 0.4 mm long in a dehiscent capsule, and in this respect is similar to most other *Begonia* species which are though to release their seeds passively, leading to localised dispersal patterns (Agren and Schemske, 1993; Matolweni *et al.*, 2000). Restricted gene flow due to local seed dispersal is likely to contribute towards population differentiation. Secondly, the concept of 'continuous distribution' in *B. socotrana* is of course open for discussion. The plants often occur in crevices and cracks; these microhabitats will naturally restrict dispersal abilities compared to growth in open exposed habitats. Thirdly, it is possible there is an historical component to the distribution of the genetic variability. While long term historical climatic data is sparse for Socotra,

there is some evidence that suggests the island may have previously been through periods of increased aridity. Studies of the southern Arabian highlands suggests there was a markedly dry period during the mid Holocene between 5000 and 2500 years ago (Cole *et al.*, 2001). *Begonia socotrana* occurs in a high-altitude zone of mountains and inter-connecting plateaus. These currently attract mist and cloud cover and receive rainfall in the winter and summer monsoons (Miller and Morris, 2001). However, in periods of increased aridity the moisture belt will migrate upwards, potentially creating 'mountain peak islands' fragmenting species ranges that are more continuously distributed during wetter periods. Under these conditions differentiation is likely to occur, and this may leave a genetic signature that persists when species' ranges expand as aridity decreases. It is thus possible that the current distribution of population genetic variability in *Begonia socotrana* reflects a combination of contemporary dispersal patterns overlaid onto historical localised vicariance.

4.2. Sexual versus asexual reproduction

Despite the proliferous production of bulbils, there was no evidence for clonal growth in most of the populations of *B. socotrana*. The only site where more than two individuals shared the same genotype was terrestrial rather than a cliff, and the plants occurred in a space of about 1 metre growing in loose soil in a gully. They are likely to be the result of a single cluster of bulbils becoming washed apart during heavy rain. In the other two sites that had individuals with identical genotypes (Mugudrihon and Rewged) the plants were growing adjacently in horizontal rock clefts and probably represent clonal growth over a very limited area. The bulbils appear to be quite firmly attached to the base of the plant, and their major role appears to be perenniation rather than dispersal. The major means of dispersal thus appears to be via sexually derived seeds.

4.3. Inbreeding versus outbreeding

Most *Begonia* species are thought to be deceit pollinated. The female flowers are rewardless and attract pollinator visits by having bright yellow stylodia which mimic the pollen bearing anthers of the male (Agren and Schemske, 1991).

Pollinator observations are rare in *Begonia*, but they are thought to be visited by generalist pollinators such as bees. Published data on outcrossing rates in *Begonia* are available from only two out of the *ca.* 1400 species in the genus. Both of these species (*B. hirsuta* and *B. semiovata*) were found to have very low outcrossing rates, consistent with reproduction being predominantly via self-pollination. *B. hirsuta* and *B. semiovata* are both annuals, and have many-flowered infloresences with proximal male and female flowers which promotes self-pollination (Agren and Schemske, 1993). In contrast our estimates of breeding behaviour in *B. socotrana* suggest that cross-pollination is important in this species.

B. socotrana has fewer flowered infloresences than *B. hirsuta* and *B. semiovata.* Its infloresences are cymose, with 1-3 female flowers being subtended by several male flowers. This spatial separation of the flowers, coupled with some degree of protandry, presumably contributes towards the observed evidence for outcrossing. Nevertheless it is interesting to note that the species is fully self fertile in cultivation, and that the temporal separation of male and female flower maturity is not absolute; male and female flowering times overlap considerably. The lack of inbreeding in the wild could be facilitated if the male flowers are harvested of their pollen by the time the female flowers open; genetic factors such pollen competition and/or early acting inbreeding depression may also contribute.

4.4. Future conservation of Begonia socotrana

The conservation status of *B. socotrana* has been reduced to the IUCN category of 'Least Concern' based on our recent surveys (Hughes and Miller, in press). Nevertheless, the species has been recognised as a 'Flagship' species - a symbol of conservation awareness and also an indicator of the health of cliff vegetation in Socotra's mountains (Miller and Morris, 2001). The genetic surveys undertaken here represent a baseline from which any changes in this flagship species can be monitored. Furthermore, the basic insights gained into the species' reproduction and population structure can facilitate the formulation of informed management plans in the event of an increasing level of threat, and also contribute towards sampling strategies and the management of *ex-situ* collections. Despite the remarkable level of endemism and the global importance of the Socotran flora, this

study represents the first investigation into natural population genetic structure of any Socotran plant.

Acknowledgements

The authors would like to thank M. L. Hollingsworth, A. Ponge and R. Hollands for technical support, and S. Scott, F. Inches and A. Fowler for looking after the living material. This work was funded by the M. L. MacIntyre *Begonia* trust, to whom MH is particularly grateful, and carried out using the facilties of the Royal Botanic Garden Edinburgh (supported by the Scottish Executive Environment and Rural Affairs Department). AGM would also like to thank the British Government Darwin Initiative, the Yemen Environmental Protection Council and GEF/UNOPS for providing support for research on Socotra.

References

Agren, J., Schemske, D.W., 1991. Pollination by deceit in a neotropical monoecious herb, *Begonia involucrata*. Biotropica 23. 235-241.

Agren, J., Schemske, D.W., 1993. Outcrossing rate and inbreeding depression in two annual monoecious herbs, *Begonia hirsuta* and *B. semiovata*. Evolution 47. 125-135.

Allard, R.W., Jain, S.K., Workman, P.L., 1969. Genetics of inbreeding populations. Advances in Genetics 14. 55-131.

Balfour, I.B., 1888. The botany of Socotra. Transactions of the Royal Society of Edinburgh 31.

Belousova, L.S., Denisova, L.V., 1992. Rare Plants of the World. A.A.Balkema, Rotterdam.

Beydoun, Z.R., Bichan, H.R., 1970. The geology of Socotra, Gulf of Aden. Quarterly Journal of the Geological Society of London 125. 413-466.

Bowcock, A.M., Ruiz-Linares, A., Tomfohrde. J., Minch, E., Kidd, J.R., Cavalli-Sforza, L.L., 1994. High resolution of human evolutionary trees with polymorphic microsatellites. Nature 368. 455-457.

Cole, K., McCorriston, J., Miller, A.G., 2001. Holocene paleoenvironments of the souther Arabian highlands reconstructed using fossil hyrax middens. Abstract of presentation given at the 86th annual meeting of the Ecological Society of America, Scientific Program p. 86. http://www.usgs.nau.edu/global_change/ESAAbstract.html

Doyle, J.J. Doyle, J.L., 1987. A rapid DNA isolation procedure for small quantities of fresh leaf material. Phytochemical Bulletin 19. 11-15.

Goodman, S.J., 1997. RST CALC: A collection of computer programs for calculating unbiased estimates of genetic differentiation and determining their significance for microsatellite data. Molecular Ecology 6. 881-885.

Goudet, J., 2001. FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available from http://www.unil.ch/izea/softwares/fstat.html.

Hughes, M., Miller, A.G., in press. A new endemic species of *Begonia* (Begoniaceae) from the Socotra archipelago. Edinburgh Journal of Botany 59.

Hughes, M., Russell, J. Hollingsworth, P.M., in press. Polymorphic microsatellite markers for the Socotran endemic herb *Begonia socotrana.* Molecular Ecology Notes 2.

Koopowitz, H., Kaye, H., 1983. Plant Extinction. A Global Crisis. Stone Wall Press Inc., Washington.

Lavranos, J., Radcliffe-Smith, A., 1969. List of rare and endangered Socotran and Abd al Kuri endemics, with an account of their present status, as far as could be determined from observations made in 1967. Middle East Command Expedition to Socotra, 1967. Unpublished.

Lewis, P.O., Zaykin, D., 2001. Genetic Data Analysis: Computer program for the analysis of allelic data. Version 1.0 (d16c). Available from http://lewis.eeb.uconn.edu/lewishome/software.html

Lucas, G. Synge, H. (eds), 1978. *The IUCN Plant Red Data Book.* IUCN: Morges, Switzerland.

Matolweni, L.O., Balkwill, K. McLellan, T., 2000. Genetic diversity and gene flow in the morphologically variable, rare endemics *Begonia dregei* and *Begonia homonyma* (Begoniaceae). American Journal of Botany 87. 431-439.

Miller, A.G., Morris, M., 2001. Conservation and Sustainable use of the Biodiversity of the Soqotra Archipelago. UNDP-GEF report no. YEM/96/G32

Minch, E., Ruiz-Linares, A., Goldstein, D.B., 1995. MICROSAT. http://hpgl.stanford.edu/projects/microsat/

Nieves, M., Henriquez, G., Rodrigo Perez, J.D., Rodriguez, C.S. 1986. Flora y Vegetacion del Archipelago Canario. Edirca, Gran Canaria.

Park, S.D.E., 2001. Microsatellite toolkit. http://acer.gen.tcd.ie/~sdepark/ms-toolkit/

Popov, G.B., 1957. The vegetation of Socotra. Journal of the Linnaean Society, Botany 55. 706-720.

Rice, W.R., 1989. Analyzing tables of statistical tests. Evolution 43. 223-225.

Ritland, K., 1996. MLTR. http://genetics.forestry.ubc.ca/ritland/programs.html

Rousset, F., 1997. Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. Genetics 145. 1219-1228.

Schneider, S., Roessli, D., Excoffier, L., 2000 ARLEQUIN ver 2.000. A software for population genetics data analysis. Genetics and Biometry Lab, Dept. of Anthropology, University of Geneva. http://lgb.unige.ch/arlequin/index.php3

Slatkin, M., 1995. A measure of population subdivision based on microsatellite allele frequencies. Genetics 139. 457-462.

Swofford, D.L., 1998. PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Walter, K.S., Gilliet, H.J. (eds), 1998. 1997 IUCN Red List of Threatened Plants. IUCN, Morges, Switzerland and Cambridge.

Weir, B.S., Cockerham, C.C., 1984. Estimating *F*-statistics for the analysis of population structure. Evolution 38. 1358-1370.

Wiggins, I.L., Porter, D.M. 1971. Flora of the Galapagos Isands. Stanford University Press, Stanford, California.

Addendum to **Population genetic structure in the island endemic *Begonia socotrana***. M. Hughes, P.M. Hollingsworth & A.G. Miller.

Begonia socotrana has a sister species *Begonia samhaensis* M. Hughes & A.G. Miller on the nearby island of Samha in the Socotra archipleago. *B. samhaensis* is morphologically very distinct from *B. socotrana*, and is endemic to the highest altitude parts of the limestone plateau at the summit of Samha.

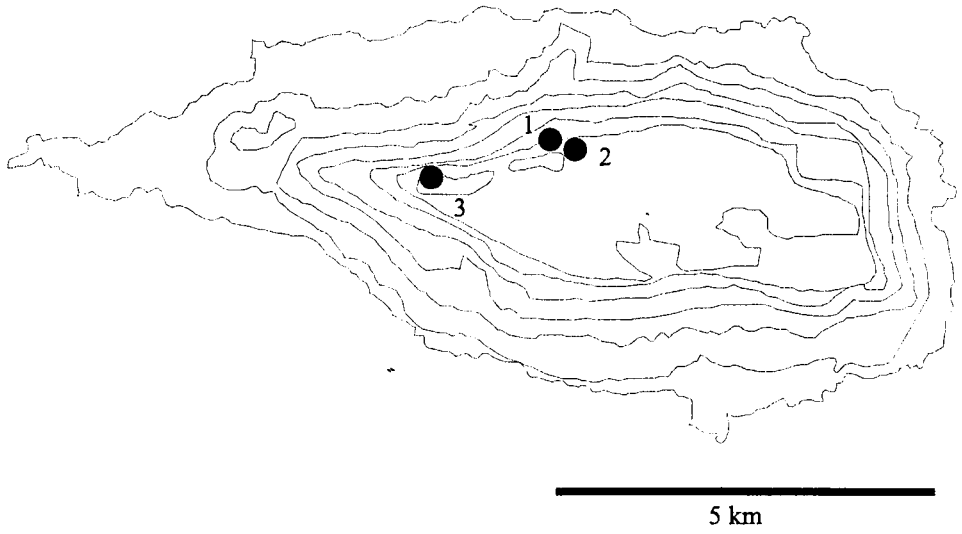

Figure 1. Contour map of Samaha island showing the collection sites of *B. samhahensis* (•) which also represent its entire distribution. The contours are at 100m intervals, with the collection sites being adjacent to or above the 700m contour which delimits the two highest parts of the island.

Leaf samples dried in silica gel were taken from three populations of *B. samhaensis* in February 1999. Forty individuals were sampled, nine from the two slightly lower altitude sites (site numbers 1 and 2; ca. 690m) and the remainder from the highest point of the island which is a limestone torr at (site number 3; ca. 750m). The torr is the main stronghold of the species, with a population of around 200 plants. The majority of these are in a single sheltered rock cleft about 1 metre wide and six metres long. A flowering specimen of *B. samhaensis* growing at the summit is shown in Figure2.

Microsatellite analysis was carried out on the samples of *B. samhaensis* using the primers and protocols as listed in Hughes et al. (2002). Only two of the microsatellite loci isolated from *B. socotrana* amplified products from *B. samhaensis* (B215 and B226).

Figure 2. *Begonia samhaensis* growing on a sheltered limestone cliff at the summit of Samha.

The poor transference of loci developed from *B. socotrana* to *B. samhaensis* indicate the species have a considerable genetic distance between them and that they are likely to have been evolving separately for some time; this is consistent with a hypothesis of both species being relics on their respective islands rather than having a more recent progenitor-derivative relationship. This is supported by the relatively long branches the species have in the nuclear ribosomal ITS phylogeny by Forrest (2000). The sequences of the two species differ by 72 transition mutations and 27 transversion mutations, with an overall uncorrected pairwise sequence divergence of 11% (as estimated using PAUP 4*). ITS has been shown not to conform to a molecular clock in *Begonia* (Forrest, 2000), so a divergence time cannot be estimated from this data. However, this

degree of divergence is remarkable and indicative that both species have been separate for a timescale measured in millions of years.

The two species can be hybridised successfully, with full seed set and rapid germination, which indicates a degree of hybrid vigour. The F_1 offspring are morphologically intermediate between the parents with respect to both vegetative and floral morphology, although several times larger. The rapid germination and much larger size indicate a high level of heterosis in the F_1 generation, which may reflect a high genetic load in one or both the parent species. Attempts to self the F_1 plants did not produce any seed set.

There is a trend of microsatellite loci that are transferable tending to be less variable in the species one transfers them to (e.g. Glenn *et al.* 1996). One of the possible causes of this is that primer sites are more likely to be maintained if they are under some functional constraint, and this constraint may also apply to the microsatellite region itself. There may also be ascertainment bias, as studies suggest the median allele length of microsatellites is longest in the species from which the markers were derived (Crawford *et al.*, 1998). In this case however, the results are likely to reflect a true lack of diversity because preliminary investigations using ISSR and AFLP loci shown to be polymorphic in *B. socotrana* were completely monomorphic in *B. samhaensis* (data not shown). Also, the allele sizes observed in *B. samhaensis* were of a similar range to that observed in *B. socotrana*, so ascertainment bias does not seem to be operating.

The 40 individuals analysed (from all three populations) were homozygous for allele 3 at locus B215 and allele 9 at locus B226. *B. socotrana* posesses an allele equivalent in length to the allele (3) observed in *B. samhaensis* at locus B215. An allele corresponding to length 9 at locus B226 is not found in *B. socotrana*, but the allele falls within the size range observed in the species at that locus. The fact that allele sizes are of a similar range in the two species despite their very long isolation is congruent with size limitation acting on the loci. Thus the microsatellite data, with the allele lengths found in *B. samhaensis* being nested inside the size range found in *B. socotrana*, fails to reflect the divergence shown by the ITS sequence data. This highlights the limit of the

utility of microsatellite markers, which are only applicable to more recently diverged taxa.

The fixation for single alleles at the two amplifiable loci in *B. samhaensis* is likely to be the result of inbreeding and drift due to the small total population size of this species, which is estimated as no more than a few hundred individuals. In dryer periods, altitudinal migration is not an option as it is for *Begonia socotrana* as the mist zone will simply reduce in size due to it already being limited to the highest 50-60m of the island. It is in fact quite surprising *B. samhensis* is not extinct, as it is possible that its population size has been reduced even further in the past given the narrow range of the mist zone habitat the species is restricted to and the susceptibility of this habitat to climatic changes.

References

Crawford, A.M., Kappes, S.M., Paterson, K.A., de Gortari, M.J., Dodds, K.G., Freking, B.A., Stone, R.T. and Beattie C.W., 1998. Microsatellite evolution: Testing the ascertainment bias hypothesis. Journal of Molecular Evolution 46. 256-260.

Forrest, L.L., 2000. A phylogeny of Begoniaceae. PhD thesis, University of Glasgow.

Glenn, T.C., Stephan, W., Dessauer, H.C., Braun, M.J., 1996. Allelic diversity in alligator microsatellite loci is negatively correlated with GC content of flanking sequences and evolutionary conservation of PCR amplifiability. Molecular Biology and Evolution 13. 1151-1154.

Hughes, M., Russell, J. Hollingsworth, P.M. (2002). Polymorphic microsatellite markers for the Socotran endemic herb *Begonia socotrana.* Molecular Ecology Notes 2, in press.

Extreme population differentiation in the sporadically distributed herb, *Begonia sutherlandii* Hook. f.

M. Hughes & P.M. Hollingsworth

Abstract

Begonia sutherlandii is native to eastern and southern Africa, where it is restricted to shaded, moist banks in indigenous forest. Nine populations were sampled from the forests of Kwazulu-Natal, South Africa and analysed for variation at seven microsatellite loci in order to investigate population structure and relationships. A high level of population differentiation was found (θ=0.482, P<0.001; R_{ST}=0.634, P<0.001), which along with a high number of private alleles reflects the severe isolation of populations in a patchily distributed forest habitat.

A pattern of significant isolation by distance was found within the populations from the mist belt forests in the Kwazulu-Natal midlands. This correlation between genetic and geographic distance was primarily due to comparisons between northern and southern populations, and it seems likely that vicariance events rather than contemporary gene flow are reponsible for the patterns of divergence seen. Population relationships appear to be strongly governed by the history and continuity of forest cover in the region.

The high degree of population isolation and divergence in *B. sutherlandii* is discussed in the light of macro evolutionary patterns in the genus as a whole. It seems probable that restricted gene flow between *Begonia* populations has been a contributing factor to the large radiation of the genus, which contains ca. 1400 species.

Keywords: *Begonia*; microsatellites; forest distribution; Kwazulu-Natal.

Introduction

Begonia L. is a genus containing ca. 1400 species (Doorenbos et al. 1998). It occurs throughout the tropics, with the greatest diversity occurring in the montane and evergreen forests of the Andes and Malesia. The majority of *Begonia* species are shade tolerant hydrophilic herbs, growing in damp and sheltered environments such as seep faces and waterfall mist zones under a forest canopy. The Begoniaceae (comprising *Begonia*, *Symbegonia* Warb. and *Hillebrandia* Oliver) is classified in the Cucurbitales (Angiosperm Phylogeny Group, 1998), where it is the largest family. *Begonia* itself accounts for nearly 99% of the species in the Begoniaceae; in terms of species number, it is one of the ten largest angiosperm genera.

Large genera or clades are likely to possess some property that make them prone to speciation, so their study may give insights into the causes of diversification. Biodiversity in *Begonia* exhibits a range limitation at all spatial scales. This is congruent with gene flow within most *Begonia* species being of an extremely local nature, and suggests they may be prone to forming isolated populations and speciating by some kind of allopatric model. Several aspects of the distribution of *Begonia* biodiversity are congruent with a hypothesis of speciation being aided by restricted gene flow. These are (i) populations having restricted distributions, often being restricted to sporadic patches of suitable microhabitat; (ii) a high level of narrow endemism; (iii) widespread species being rare and highly morphologically variable unless they have atypical dispersal adaptations; (iv) monophyletic groups of species showing geographic clustering.

Patchy occurrence of populations

Begonia species usually form populations with a very small range (Burt-Utley, 1985). They are limited to areas of specific microhabitat, which is due to their demand for deep shade and moist conditions, often coupled with a requirement for a steeply banked substrate. This means populations are often separated by considerable distances of unsuitable habitat. Dispersal in most *Begonia* species is thought to be passive, with seeds being dispersed by gravity a short distance from the parent plant

(Matolweni et al., 2000). The patchy population distribution arising from the narrow niche of *Begonia* species means that seed dispersal via gravity dispersed seeds between distant populations is likely to be sporadic. Gene flow through pollen may also be limited if the between population distances exceed pollinator movement distances.

Narrow endemism

Many *Begonia* species are narrow endemics. Sosef (1994) found the species of the African *Begonia* sections *Loasibegonia* and *Scutobegonia* to be useful as indicators of Pleistocene forest refuges due to their severely restricted ranges and poor dispersal mechanisms. A monograph by Kiew (2001) of the limestone *Begonia* of Sabah lists fourteen species, only one of which is widespread in the study region with nine of the species being endemic to single hills. Narrow endemism could indicate dispersal limitation of species ranges, and local gene flow could favour the evolution of a larger number of locally adapted species rather than fewer widespread species with broader ecological tolerance.

Variability of widespread species

Although most *Begonia* species have quite narrow distributions a few are relatively widespread, and these species (e.g. *B. urticae*, Costa Rica to Peru; *B. sutherlandii*, Tanzania to South Africa) are usually highly morphologically variable. This variability is often reflected in extensive taxonomic synonomy, and in the case of *B. urticae* and *B. sutherlandii* many single populations were originally described as unique species. The variability of widespread *Begonia* species could be due to a lack of gene flow at the regional scale; this would permit individual populations to evolve to their local ecological optima without being swamped by alleles from 'foreign' populations adapted to slightly differing conditions (Barton & Clarke, 1990).

In rare cases, widespread species of *Begonia* do show a relative morphological uniformity across their range. These species tend to have unusual adapations which may allow populations to remain in genetic contact over longer distances than most *Begonia* species. For example, *B. oxyloba* occurs across tropical

Africa from Liberia to Tanzania and into Madagascar. This species has relatively large sweetly scented fleshy fruit, and is likely to be animal dispersed. As it occurs on Madagascar as well as the African continent, it seems likely that birds are one of the dispersal vectors. The lianescent *B. glabra* has the widest range of any *Begonia* species in the neotropics, and occurs from Costa Rica south to Peru and east into Brazil. This species is likely to be wind pollinated to a large extent as it produces flowers with small tepals; the female flowers have elongated and protruding stigmas, and the male flowers are borne in large infloresences and produce a very dry pollen that is freely released from their anthers. Wind dispersal of pollen and seed at the height of the forest canopy is likely to favour longer distance gene flow than insect pollination and passive seed dispersal in the the sheltered conditions of the forest floor.

Geographical restriction of monophyletic species groups

The restriction in the range of diversity is also seen above the species level in *Begonia*, as the members of monophyletic groups within the genus tend to occur in the same region. A phylogeny of *Begonia* by Forrest (2000) based on nuclear ribosomal ITS sequence data and including 160 species from throughout the tropics showed that in some cases geographical proximity was a better indicator of species relationships than the accepted taxonomy. Previously unsuspected monophyly of South African and endemic Malagasy *Begonia* was uncovered, and at a larger scale Asian and Neotropic species each formed separate clades. A phylogeny of *Begonia* by Plana (2002) using the chloroplast trn*L* intron also recovered the monophyly of the endemic Malagasy species and South African species. The correlation of phylogeny and geography is congruent with a hypothesis of restricted long distance dispersal.

A recurrent pattern of geographical and genetic correlation at all levels in a phylogeny indicates that the same evolutionary process (restricted gene flow) is the cause (Templeton, 1998). Despite *Begonia* being one of the largest angiosperm genera, there is little known about the geographical distribution of population genetic variation within *Begonia* species. In order to investigate intra-specific patterns of diversity in *Begonia* in detail, this paper presents a study of the widespread African

species *B. sutherlandii* using nuclear microsatellite markers.

The study species

B. sutherlandii has an east Afromontane distribution, from northern Tanzania to the Transkei in South Africa. It occurs in primary forest on steep substrates, usually shaded banks or boulders, and always in wet conditions. Given the preference of *B. sutherlandii* for a specific microhabitat with a favourable aspect (facing south to south-easterly in South Africa), the species has a sporadic distribution within its main primary forest habitat. Its forms discrete populations which can number up to several hundred plants in favourable conditions. Typically for the genus, *B. sutherlandii* has asymmetric leaves, which are lanceolate and vary in size from 2 to 30 cm long, though are most commonly around 5-10 cm long. The leaves show a range of outlines, ranging from shallowly dentate to being dissected almost to the midrib. In common with many other east African *Begonia*, it is a perennial which survives the mild seasonality of the Afromontane climate as a tuber. It also produces small bulbils in the leaf axils, which are derived from a compressed shoot and provide futher means of perenniation and possibly limited dispersal. It is monoecious, with bisexual axillary or terminal dichasial inflorescences, bearing from a few to around twenty flowers, and is unusual in *Begonia* in having orange flowers (occasionally brick red or yellow). The female flowers in *Begonia* are rewardless, and attract pollinator visits by deceit, with the yellow stylodia mimicking the pollen bearing anthers of the male flower (Agren & Schemske, 1991). The seeds of *B. sutherlandii* are around 400-500 μm long, which are contained in a dehiscent fruit with three wings. In common with many *Begonia* species the seeds are likely to be passively dispersed in the vicinity of the parent plant (Agren & Schemske, 1993; de Lange & Bouman, 1999; Matolweni et al., 2000).

The study area

The indigenous forest which is the main habitat of *Begonia sutherlandii* covers only 0.56% of the land area in South Africa, and is the country's smallest biome. It has a highly fragmented distribution, with most of the forest occurring in patches of less than 1 km^2 (Eeley et al., 1999). One-sixth of this forest cover is found within Kwazulu-Natal (Figure 1, page 129), where two main forest types can be distinguished, Afromontane forest and Indian Ocean coastal belt forest. These can be

divided in a number of subtypes according to Eeley et al. (1999), with the Afromontane forest comprising montane and mist belt forests. The montane forests are largely restricted to the Drakensberg escarpment, with the mist belt forests being the dominant forest type over most of the Kwazulu-Natal midlands. Both types are restricted to south and south-eastern facing hills and slopes. The Indian Ocean coastal belt forests comprise dune forest, swamp forest, sand forest, riverine forest, coastal lowland forest and coastal scarp forest. Indian Ocean coastal belt forests are restricted to the flat coastal plain, with swamp, sand and riverine forest being confined to the north of the province. Dune and lowland forests form a fragmented belt along the coast to the south of Durban. Scarp forests are restricted to gorges and south to south-eastern facing slopes on the escarpment to the west of the coastal plain.

The highly fragmented nature of the forest cover in Kwazulu-Natal could be the result of either human clearance (Acocks, 1998) or environmental factors (Geldenhuys, 1992). Although anthropogenic clearance has undoubtedly occurred, there is some evidence that the current forest distribution is related to climate. Rather than having a random distribution as might be expected from wholesale clearances, forest is restricted to south and south easten facing slopes and sheltered gorges. This is due to the spread of grassland at the expense of forest during dryer and colder periods, leading to a fire dominated landscape which restricts forest to sheletered pockets. Forest cover is likely to have expanded and contracted during the climatic cycles experienced by South Africa throughout the Pleistocene, and the current distribution probably reflects the latest contraction of forest range since the warmer and wetter conditions during the Holocene altithermal *ca.* 7000 years ago (Eeley et al., 1999)

Aims of the study

The fragmented distribution of forest produces a corresponding fragmented distribution of *B. sutherlandii*, and represents an ideal system for investigating the connectivity of populations over a range of distances. The aims of this study are to examine the geography of the distribution of genetic diversity within *B. sutherlandii*,

and look for congruence between the microevolutionary patterns seen in this species and the macroevolutionary patterns seen in the genus *Begonia*. Populations of *B. sutherlandii* have been sampled from throughout Kwzulu-Natal in order to investigate the geographic patterns of genetic structure and their cause.

Specifically, we aim to (i) to examine the amount of genetic diversity and its partitioning within and between populations, (ii) to look for evidence of genetic isolation by distance through restricted dispersal, and (iii) to examine the relatedness of populations in relation to the forest cover in Kwazulu-Natal and its recent history.

Materials and methods

Sampling

Leaf samples (dried in silica gel) and voucher specimens were collected from 9 populations of *Begonia sutherlandii* in February 2001 from forest patches in Kwazulu Natal (KZN), South Africa (Fig. 1). Three main forest types were visited for collection, which are classified as (i) mist belt, (ii) montane and (iii) scarp following Eeley et al. (1999). The locality details and the number of samples collected is shown in Table 1. All individuals in a population were collected within 500m of each other, with the exception of the Kokstaad population, which consists of four sub samples collected within 3.8km of one another.

The mist belt forests of Ferncliffe, Hoha, Dulini and Kokstaad are in the southern KZN midlands. This is an area of relatively high altitude and varied topology, which contains a network of numerous small forest patches. The mist belt forests of Tygerskloof, Nkandla and Qudeni are in northern KZN, and are restricted to peaks which rise to 1600m in altitude. These forests are more isolated than their counterparts in the southern KZN midlands, being separated by the hotter and dryer lowland areas of the Tugela basin.

Rainbow gorge is a patch of montane forest near Cathedral Peak in the Drakensberg mountains. This forest patch is relatively isolated from other Drakensberg escarpment forests, and also from the KZN midlands mist belt forests. The region is surrounded by extensive grasslands which are subjected to frequent fires (Geldenhuys, 1992).

Umtamvuna is a forested gorge in southern coastal KZN, which lies near the middle of the Pondoland centre of endemism (White, 1983). The scarp forest which covers the sheltered ravine has a disproportionately rich flora (Geldenhuys, 1992), and is among the sheltered coastal forest patches that may have been important refugia during glacial maxima (Lawes, 1990).

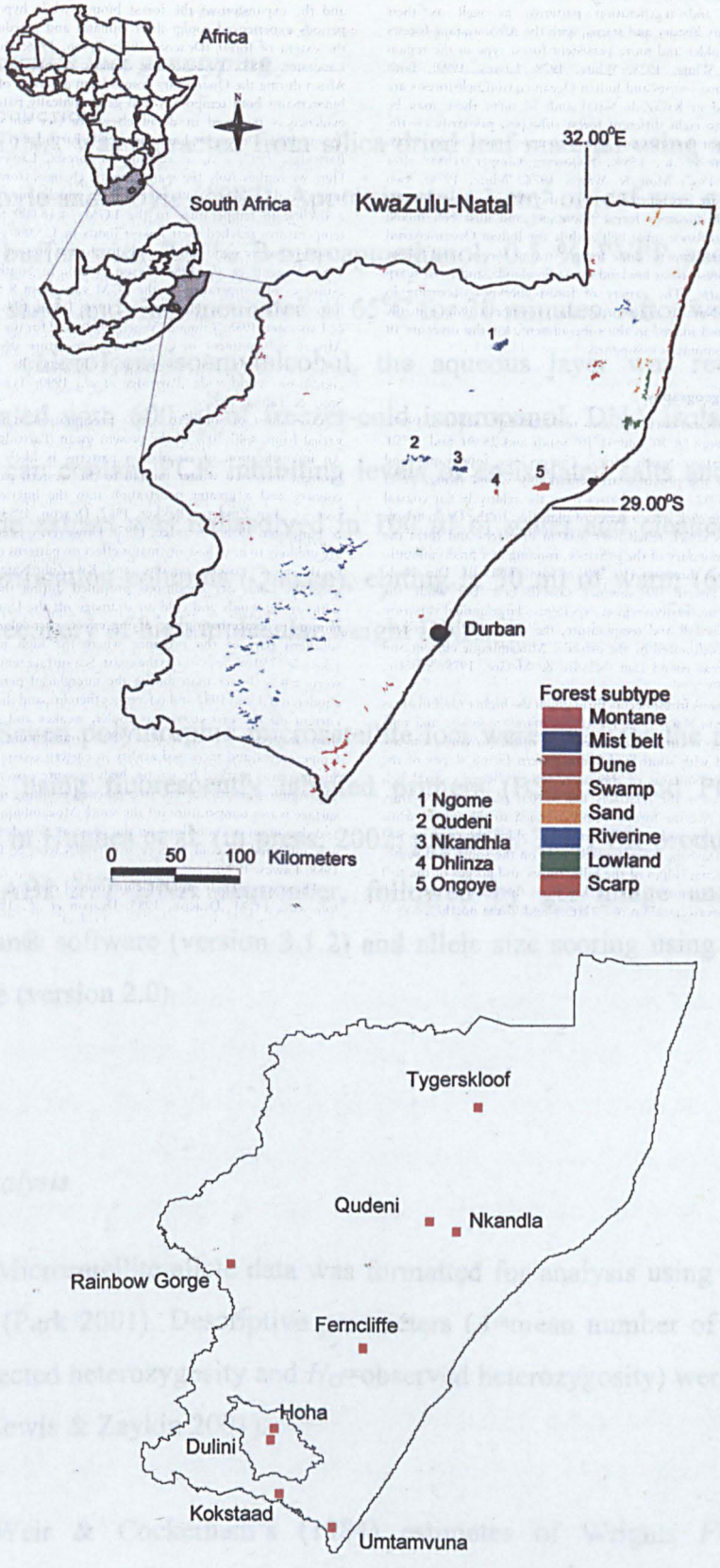

Figure 1. Top, distribution of indigenous forest in Kwazulu-Natal (taken from Eeley et al., 1999). The montane forests (pink) are restricted to the eastern border of Kwazulu-Natal; the mist belt forests form a broken distribution running from south west to north east through the midlands; the scarp forests are limited to the coastal strip running south of Durban. Below, collection localities for *B. sutherlandii*.

Microsatellite loci genotyping

DNA was extracted from silica dried leaf material using a method modified from Doyle and Doyle (1987). Approximately 1 cm^2 of leaf was ground in 800 µl of CTAB buffer with 2.5 % β-mercaptoethanol, 0.5 % PVPP and a pinch of acid washed sand, and then incubated at 65°C for 30 minutes. After washing with 500 µl of 24:1 chloroform:isoamylalcohol, the aqueous layer was removed and DNA precipitated with 600 µl of freezer-cold isopropanol. DNA isolated from *Begonia* species can contain PCR inhibiting levels of co-isolated salts and carbohydrate, so the crude extract was redissolved in 100 µl of water and cleaned using QIAquick PCR purification columns (Qiagen), eluting in 50 ml of warm (65°C) EB buffer to ensure recovery of high molecular weight DNA.

Seven polymorphic microsatellite loci were used for the population genetic surveys, using fluorescently labelled primers (BSU1-7) and PCR conditions as detailed in Hughes et al. (in press, 2002; paper 3). The PCR products were analysed on an ABI 377 DNA sequencer, followed by gel image analysis using ABI Genescan® software (version 3.1.2) and allele size scoring using ABI Genotyper® software (version 2.0).

Data analysis

Microsatellite allele data was formatted for analysis using the Microsatellite Toolkit (Park 2001). Descriptive parameters (A=mean number of alleles per locus; H_E=expected heterozygosity and H_O=observed heterozygosity) were calculated using GDA (Lewis & Zaykin 2001).

Weir & Cockerham's (1984) estimates of Wrights F-statistics F_{IS} (f; deviation from panmixia attributable to non-random mating within populations), F_{ST} (θ; deviation from panmixia attributable to non-random mating among populations), and the inclusive measure F_{IT} (F; deviation from panmixia attributable to non-random mating within and between populations) were estimated using FSTAT (Goudet, 2001). The per locus estimates of F_{ST} were tested for sigificant difference

from the null hypothesis of panmixia by jackknifing over populations. The global estimate of F_{ST} was tested for significance by jackknifing across loci. Permutation tests (1000 permutations, randomising alleles) were used to assess the significance of F_{IS} and F_{IT} estimates. The outcrossing rate, t, was estimated from the global estimate of F_{IS} using the equation $t=(1-F_{IS})/(1+F_{IS})$ (Allard et al., 1969). Per locus estimates of the stepwise-mutation model analogue of F_{ST}, R_{ST} (Slatkin, 1995) were calcuated using RST CALC (Goodman, 1997), testing for significance using permutation tests (1000 permutations). Due to missing data for all individuals in one or two populations for two of the loci (locus BSU3 was scored as missing from all individuals in Nkandla and Qudeni; locus BSU7 was scored as missing from all individuals in Umtamvuna), permutation tests to assess the significance of a global estimate of R_{ST} using all loci could not be completed using RST CALC. A global estimate of R_{ST} using a reduced data set of five loci was instead used to assess significance. A global estimate using all 7 loci was carried out using FSTAT, which can accommodate missing data.

A Mantel test to assess the significance of isolation by distance patterns was carried out using Arlequin version 2.000 (Schneider et al., 2001), with the two matrices being the natural log of straight-line geographic distances or straight line geographic distances (calculated from GPS data) and $F_{ST}/1-F_{ST}$ or $R_{ST}/1-R_{ST}$ (Rousset 1997). The natural log of geographic distance was used when comparing all populations. The direct distances in kilometres were used for analysing a subsample of populations which approximated a linear distribution; Rousset (1997) showed that for elongated habitats, population differentiation is a linear function of geographic distance.

Microsat (Minch et al., 1995) was used to create pairwise genetic distance matrices for both individual plants and populations using chord distance (Cavalli-Sforza, 1967). This geometric distance measure was found to be more accurate in tree reconstruction than higher variance stepwise-mutation model based measures by Takezaki & Nei (1996). Population pairwise matrices were also created for Nei's standard genetic distance (Nei, 1987) and $1-P_{SA}$ (Bowcock et al., 1994) using Microsat. PAUP* (Swofford, 1998) was used to create neighbour joining trees from these distance matrices. Bootstrap values (exhaustive sampling over loci) were

calculated for the population level chord distance neighbour joining tree using Populations 1.2.24 (Langella, 2002)

As *B. sutherlandii* produces bulbils, it has the potential to grow clonally. Where two leaf samples from separate individuals within a population provided identical genotypes at all loci, the probability of the genotypes being the result of clonal growth rather than random sexual mating was calculated following Parks & Werth (1993). The probability that a zygote aquires a given multilocus genotype (p_{gen}) can be calculated from the product of its allele frequencies in the source population as

$$p_{gen} = \left(\prod p_i\right) 2^h$$

where p_i is the frequency of each allele in the population (two per locus) and h is the number of loci that are heterozygous in that particular genotype. Allele frequencies at each locus for each population were calculated using only individuals in the same population that had different multilocus genotypes at the other loci, in order to avoid circularity and upward bias of the frequency of rare alleles (Parks & Werth, 1993). For a given sample size G (estimated as the number of different multilocus genotypes following Parks & Werth, 1993) the probability of coming across the same genotype an n^{th} time as a product of random mating can be calculated as

$$\sum_{x=n}^{G} \frac{G!}{x!(G-x)!} \left(p_{gen}\right)^x \left(1 - p_{gen}\right)^{G-x}$$

The above expression was calculated for increasing values of n up to the number of individuals sharing the same multilocus genotype. If the probability of n encounters through random mating was found to be unlikely at the 0.05 level, then for that particular genoptype (n-1) individuals were assumed to be the product of random mating and any further encounters were assumed to be the result of clonal growth and removed from further population genetic analyses.

Microsatellite scoring

Microsatellite null alleles can be caused by mutations in the PCR priming sites, and can be identified by a failure to amplify a PCR product in individuals that are homozygous for the null allele. Nulls were found for locus BSU3 in the populations sampled from Nkandla and Qudeni, and for BSU4 in Tygerskloof. No PCR product could be reliably amplified using the PCR primers for this locus in any of the individuals from these populations, whilst the remaing 6 loci amplified without difficulty. Data for Nkandla and Qudeni at locus BSU3 was scored as missing. Locus BSU4 could be amplified in individuals from Tygerskloof with a set of redesigned primers that annealed outside two transition mutations (unique to this population) in the primer site of the original primers (Hughes et al., 2002).

Locus BSU7 gave profiles that could not be interpreted as diploid from individuals sampled from Umtamvuna (i.e. more than two bands per individual were present), and all data was scored as missing for this population at this locus.

Despite being di-nucleotide loci, three of the loci (BSU5, BSU6 and BSU7) showed single base pair length polymorphisms in populations from Tygerskloof (BSU5 and BSU6), Qudeni and Kokstaad (BSU7). The cause of this in the case of BSU6 and BSU7 was determined by allele sequencing (Hughes et al., 2002). The one base-pair shift in some BSU6 alleles in the Tygerskloof population was due to an expansion mutation in a T_8 region adjacent to the di-nucleotide repeat. The presence of single nucleotide step mutations in individuals from the Qudeni population was due to a loss of a cytosine base in a CC motif in the middle of the microsatellite; no individuals homozygous for odd-numbered alleles (and hence easily sequenceable) were found from the Kokstaad population and so the homology of the mutation with respect to the Qudeni population is unknown.

Descriptive statistics

Within populations the mean number of alleles per locus ranged from A=1.9 to 5.0 (mean 3.7), with the proportion of polymorphic loci ranging from P=0.57 to 1.00. The gene diversity within populations ranged from H_E=0.254 to H_E=0.603 (mean 0.459). The number of private alleles in a population ranged from 1 (Hoha) to 13 (Tygerskloof) (Table 1). At the species level the number of alleles per locus ranged from A=9 to 22 (mean 14.4), with the mean gene diversity per locus ranging from H_E=0.752 to 0.928 (mean 0.833) (Table 2).

Clonal growth

All populations except Dulini contained at least one individual which had a significant probability of being the result of clonal growth (P<0.05). The percentage of clonal individuals in the population samples ranged from zero in Dulini up to 11% in Qudeni, with the mean across all populations being 7%. The population sample sizes adjusted for the removal of clonal individuals are shown in Table 1.

In six out of the nine populations, after the removal of all probable clonal individuals, every plant could be identified with a unique multi-locus genotype. The three populations with the lowest gene diversity, Umtamvuna, Rainbow Gorge and Kokstaad, each contained a small number of individuals that shared a multi-locus genotype but which were likely to be the result of random sexual mating rather than clonal growth (P<0.05).

Population structure

There is significant population structure in *B. sutherlandii*, with a high degree of differentiation between populations (R_{ST}=0.634, P<0.001; θ =0.482, P<0.001; Table 2). The estimate of R_{ST} was calculated using a reduced data set of 5 loci, as two loci with missing data for at least one population were removed. A global estimate using all 7 loci gives a value of R_{ST}=0.689. There was significant differentiation even between closely situated populations; the Kokstaad population

can be subdivided into 4 sub populations which were sampled within 3.8 km of each other; these sub populations gave values of R_{ST}=0.238 (P<0.01) and θ =0.220 (P<0.01).

Estimates of F_{IS}, the inbreeding coefficient, ranged from f=0.060 to f=0.271 within populations, with six out of the nine estimates being significantly different from zero. The species wide estimate was f=0.154 (P<0.01) indicating a small but significant deficit of heterozygotes. Calculating the outcrossing rate from the F_{IS} estimate according to Allard et al. (1969) gave a value of t=0.73.

Isolation by distance

The neighbour joining tree constructed using chord distance between individuals (Fig. 2) shows individuals clustering together with all the other individuals from the same population (with the exception of one individual which clusters between its source population, Dulini, and the nearest neighbouring population, Hoha). The neighbour joining tree constructed using chord distance between populations has an identical population level topology (Fig. 3) and indicates a geographic correlation with genetic distance, with spatially proximal populations tending to cluster together. The exceptions to this are Rainbow Gorge and Umtamvuna, which do not cluster with their geographic nearest neighbours.

Umtamvuna is geographically closest to Kokstaad in southern KZN, but in the neighbour joining tree takes up an isolated position and shows no obvious affinity; it clusters next to Tygerskloof but on a very short internal branch. Although the clustering of the Umtamvuna and Tygerskloof populations has no bootstrap support, the population in Umtamvuna certainly differs from the nearby mist belt forest populations of *B. sutherlandii*, which together form a group with 75% bootstrap support (comprising Ferncliffe, Dulini, Hoha and Kokstaad).

Rainbow Gorge clusters with Nkandla, although this pairing has no bootstrap support. The clustering of the Nkandla, Rainbow Gorge and Qudeni populations has 55% bootstrap support. The relationship is also found in neighbour joining trees constructed from Nei's standard distance and allele sharing. Despite the apparent

close relationship of the Rainbow Gorge and Nkandla populations, it should be noted that the null allele fixed at locus BSU3 in Nkandla and Qudeni is not present in Rainbow Gorge, where the locus amplifies normally.

Using all the populations, a Mantel test based on Slatkin's linearised F_{ST} and the natural log of geographic distance showed there to be no significant pattern of isolation by distance (P=0.063, R^2=0.06; Fig. 4). The plot shows a scattered distribution, with all the extreme outlying comparisons involving comparisons with Umtamvuna and Rainbow Gorge. These population comparisons are highlighted in Fig. 4. Removing the Umtamvuna and Rainbow Gorge populations from the dataset leaves only the populations from the mist belt forests. A Mantel test on this reduced data set was significant at the P<0.001 level (R^2=0.62) when natural log distance was used as the geographic matrix. As the populations in this reduced dataset approximate a linear distribution, a Mantel test was repeated out using non-logged geographic distance. This was also highly significant (P<0.001, R^2=0.79; Fig. 5).

Using R_{ST}/1-R_{ST} and natural log of geographic distance as the matrices in a Mantel test involving all populations gives a result which is just significant at the P<0.05 level (P=0.047, R^2=0.07; Fig. 6). Using this stepwise mutation model based measure of genetic distance, only Umtamvuna appears to provide outlying points in the plot of genetic and geographic distance (Fig. 6), with comparisons between it and neighbouring populations giving higher degrees of genetic difference than the main trend. The pairwise comparisons of Rainbow Gorge to other populations do not provide markedly abberrant results as with the F_{ST} based plot, and appear to fit within the main trend.

Removing the Umtamvuna population and repeating the Mantel test (using natural log of distance as the geographic matrix) increases the significance (P=0.009) and decreases the scatter (R^2=0.21). Removing both the Umtamvuna and Rainbow Gorge populations (leaving the mist belt populations only) further increases the significance and decreases the scatter, although less markedly (P=0.003, R^2=0.30). This is similar to the result obtained using non-logged distance as the geographic matrix for this reduced population data set (P=0.003, R^2=0.30).

Bossart & Prowell (1998) and Slatkin (1993) advocate examining the effect of different clusters of populations on significant patterns of isolation by distance. In the plot of Slatkins linear F_{ST} against geographic distance for the mist belt populations (Fig. 5), the comparisons within the northern populations (Tygerskloof, Nkandla and Qudeni) and within the southern populations (Ferncliffe, Hoha, Dulini and Kokstaad) are highlighted. None of the within-northern or within-southern groups of pairwise comparsons give a significant pattern of isolation by distance when considered alone or combined.

Table 1. Descriptive statistics by population. N_g, number of individuals genotyped; N_{adj}, number of individuals included in the population genetic analysis after the removal of probable clonal individuals; N_d, number of distinct multilocus genotypes; n, mean sample size over all polymorphic loci; P, number of polymorphic loci/number of loci applicable; A, mean number of alleles per locus; A_p, number of private alleles; H_E, expected heterozygosity; H_O, observed heterozygosity

Population	Forest type	Latitude	longitude	N_g	N_{adj}	N_d	n	P	A	A_p	H_E	H_O	f
Tygerskloof	Mist belt, N	31.314	27.846	21	19	19	16.7	6/7	5.0	13	0.603	0.555	0.082^{ns}
Qudeni	Mist belt, N	30.904	28.649	56	50	50	44.8	6/6	4.7	3	0.551	0.492	0.108**
Nkandla	Mist belt, N	31.135	28.729	43	41	41	32.8	6/6	4.3	9	0.561	0.457	0.193**
Ferncliffe	Mist belt, S	30.340	29.547	28	27	27	27.0	7/7	4.3	5	0.572	0.471	0.179**
Hoha	Mist belt, S	29.575	30.128	23	21	21	19.6	5/7	3.1	1	0.415	0.437	-0.054^{ns}
Dulini	Mist belt, S	29.556	30.187	13	13	13	13.0	5/7	3.4	3	0.455	0.429	0.060^{ns}
Kokstaad	Mist belt, S	29.647	30.585	39	36	34	30.3	7/7	3.7	6	0.401	0.294	0.271**
Rainbow Gorge	Montane	29.226	28.960	26	24	22	21.9	4/7	1.9	3	0.254	0.208	0.186*
Umtamvuna	Coastal scarp	30.173	31.002	45	43	36	34.7	4/6	2.5	6	0.323	0.241	0.265**
Mean				33	30	29	27		3.7	5.4	0.459	0.398	0.143

** $P<0.01$

* $P<0.05$

Table 2. Descriptive statistics by locus. The sample size for locus BSU3 is reduced due to null alleles present in all individuals of populations from Nkandla and Qudeni, and the sample size for locus BSU7 is reduced due to unscorable profiles in all individuals from Umtamvuna. (n=number of individuals after removal of probable clonal individuals; N_p=number of populations; A=number of alleles per locus).

Locus	n	N_p	A	H_E	H_O	R_{ST}	θ	f
BSU1	241	9	10	0.832	0.386	0.629**	0.382**	0.275*
BSU2	242	9	14	0.869	0.467	0.656**	0.454**	0.091^{ns}
BSU3	164	7	11	0.847	0.317	0.784**	0.562**	0.260*
BSU4	237	9	14	0.795	0.215	0.932**	0.667**	0.202^{ns}
BSU5	243	9	22	0.812	0.412	0.622**	0.480**	0.093^{ns}
BSU6	238	9	21	0.928	0.630	0.334**	0.294*	0.073^{ns}
BSU7	208	8	9	0.752	0.293	0.810**	0.569**	0.128^{ns}
All			14.4	0.833	0.389	0.634**[1]	0.482**	0.154**

**P<0.001

*P<0.01

ns – not significant

1. This estimate of R_{ST} was calculated using a reduced data set of 5 loci that had no missing data.

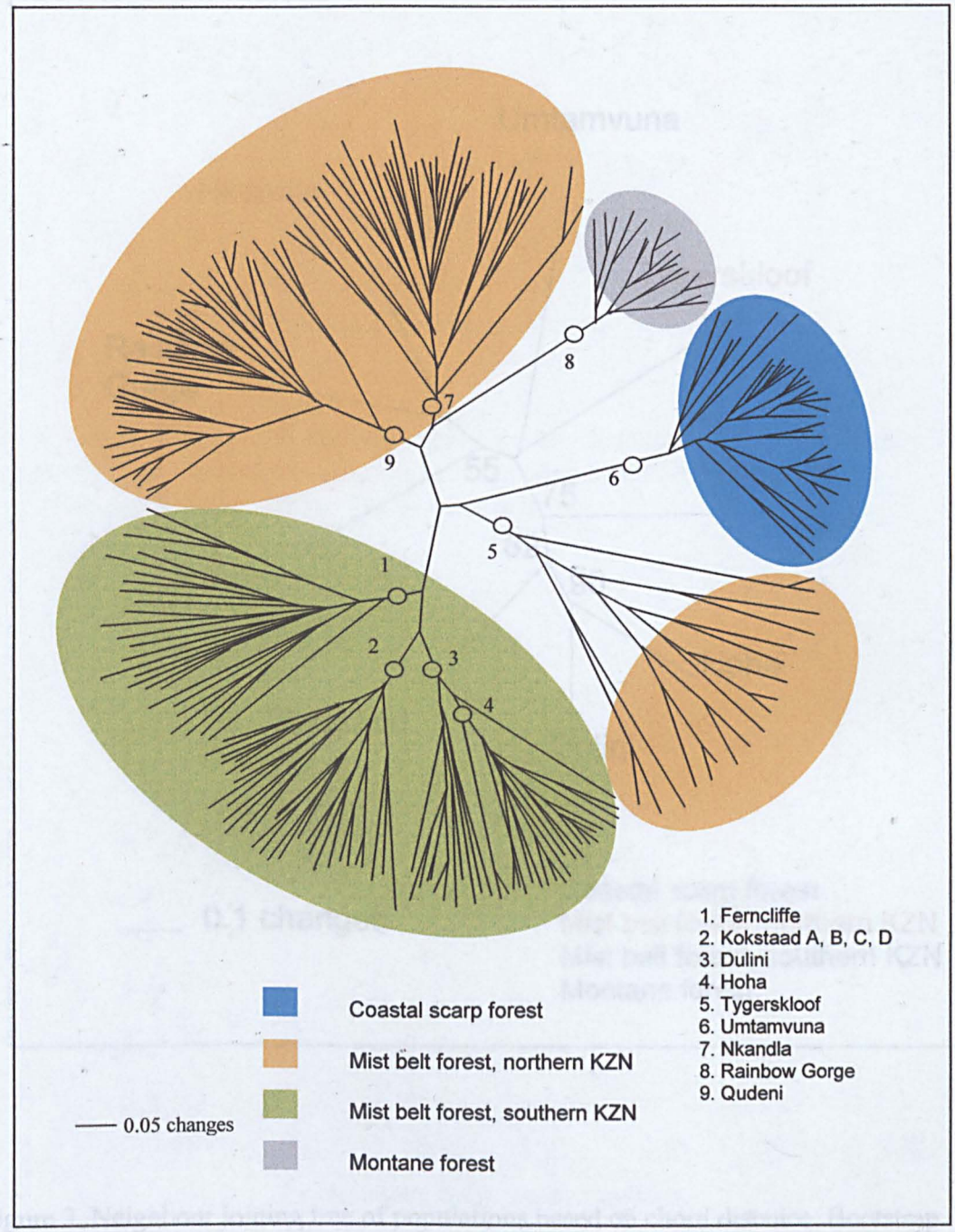

Figure 2. Neighbour joining tree of individuals based on chord distance.

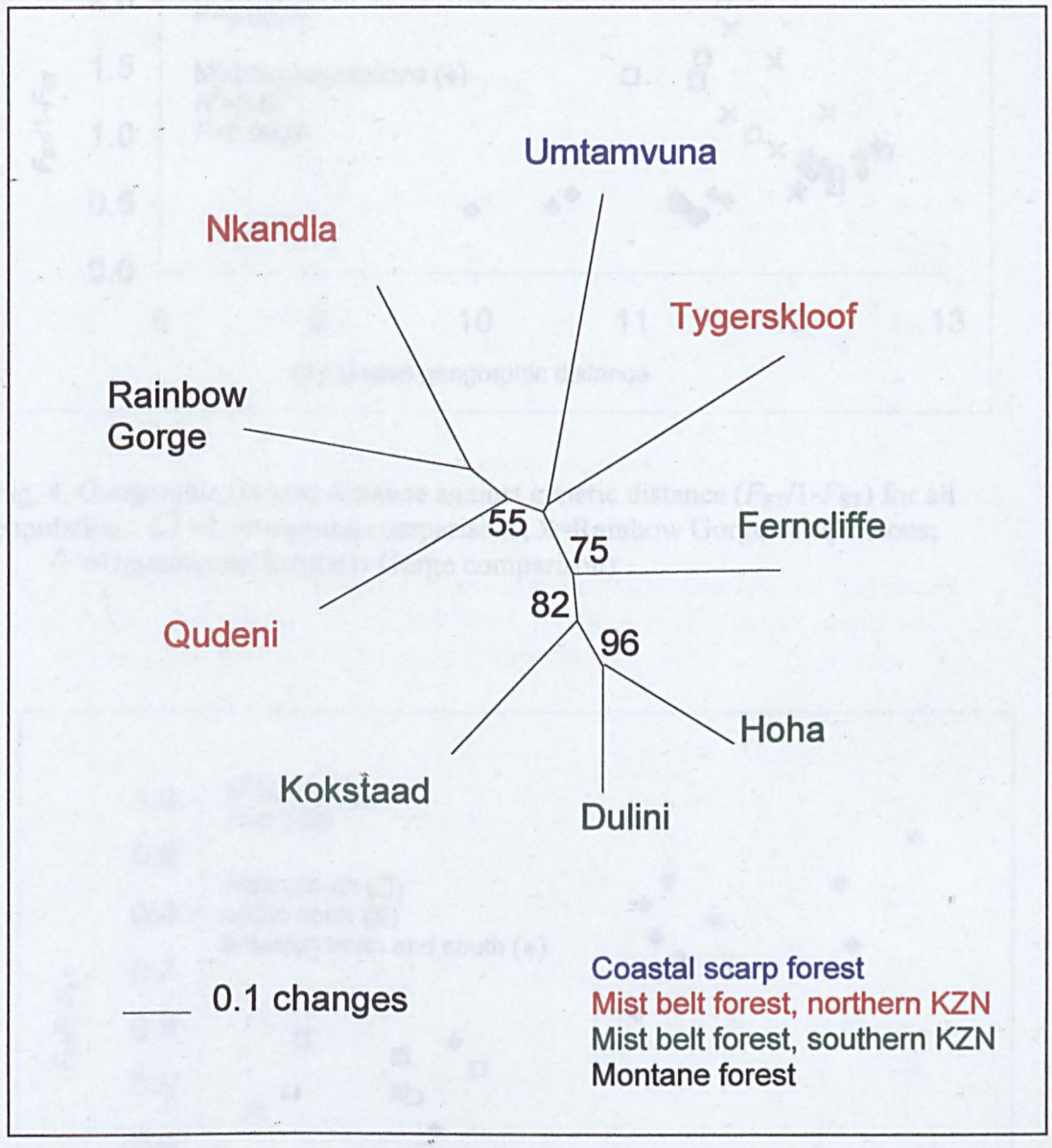

Figure 3. Neighbour joining tree of populations based on chord distance. Bootstrap values of greater than 50 are shown.

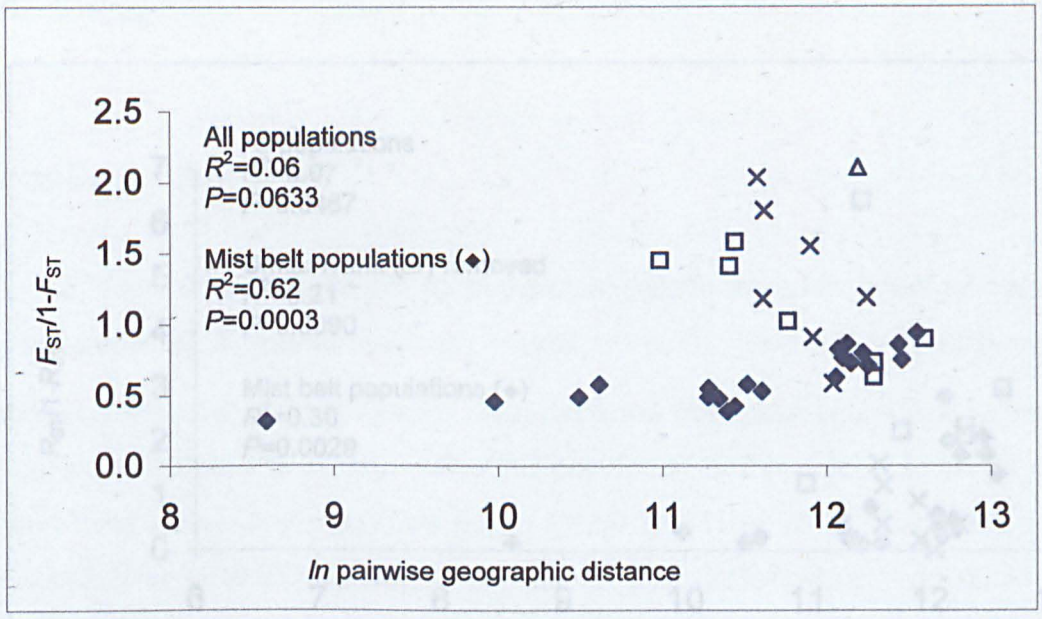

Fig. 4. Geographic (*ln* km) distance against genetic distance ($F_{ST}/1\text{-}F_{ST}$) for all populations. (□ =Umtamvuna comparisons; X=Rainbow Gorge comparisons; △ =Umtamvuna/Rainbow Gorge comparison)

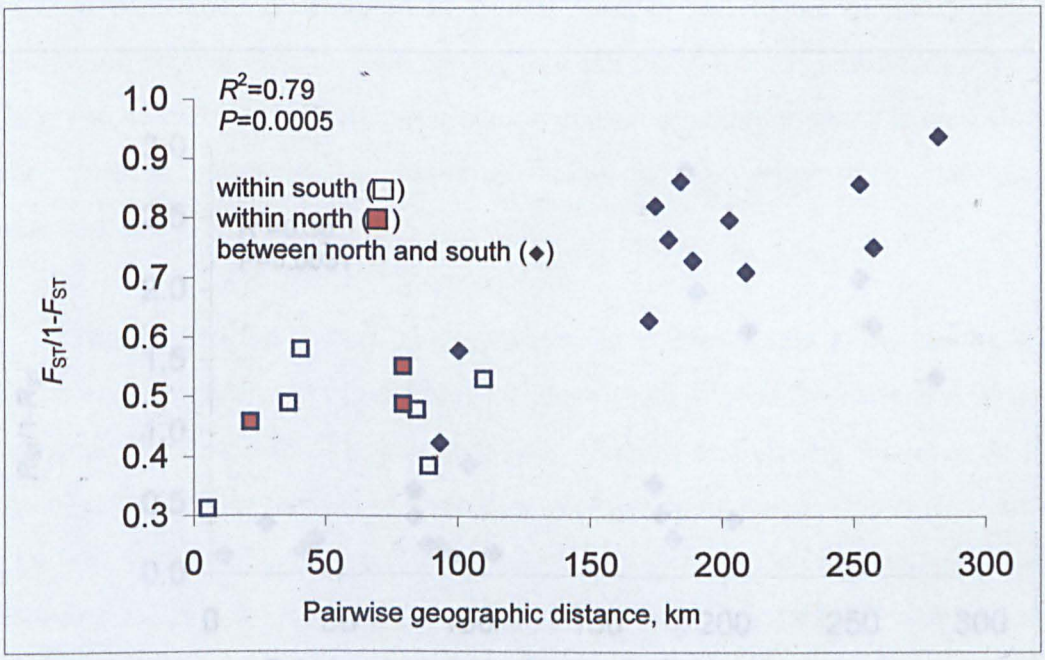

Fig. 5. Geographic (km) distance against genetic distance ($F_{ST}/1\text{-}F_{ST}$) excluding the populations from Rainbow Gorge and Umtamvuna (mist belt populations only).

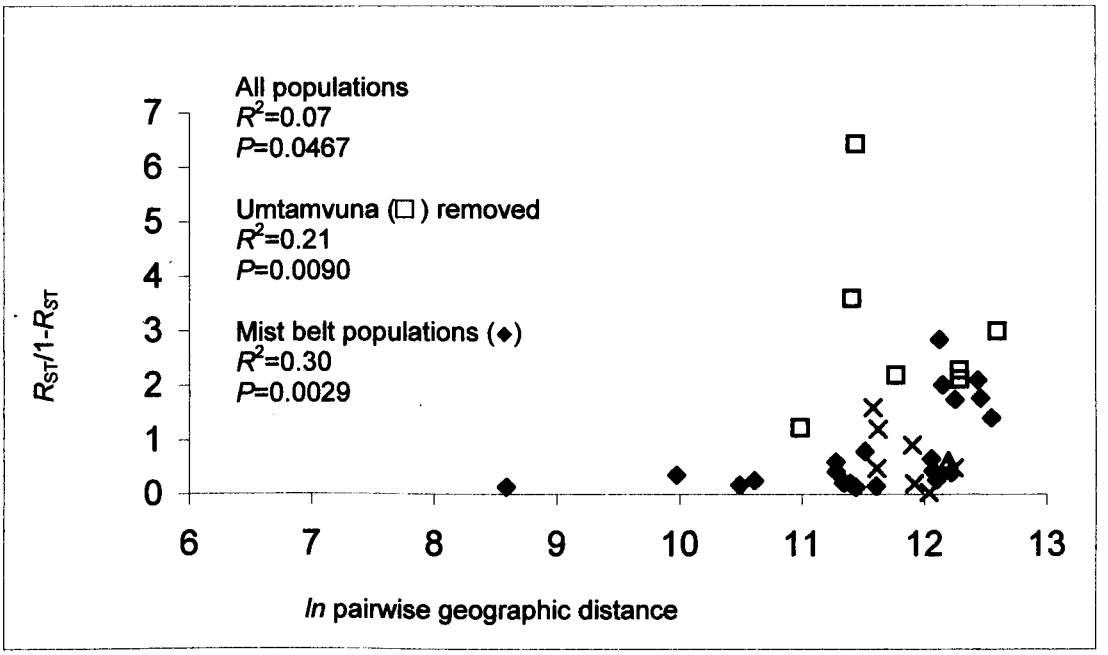

Fig. 6. Geographic (*ln* km) distance against genetic distance ($R_{ST}/1\text{-}R_{ST}$) for all populations. (□=Umtamvuna comparisons; x=Rainbow Gorge comparisons; △ =Umtamvuna/Rainbow Gorge comparison)

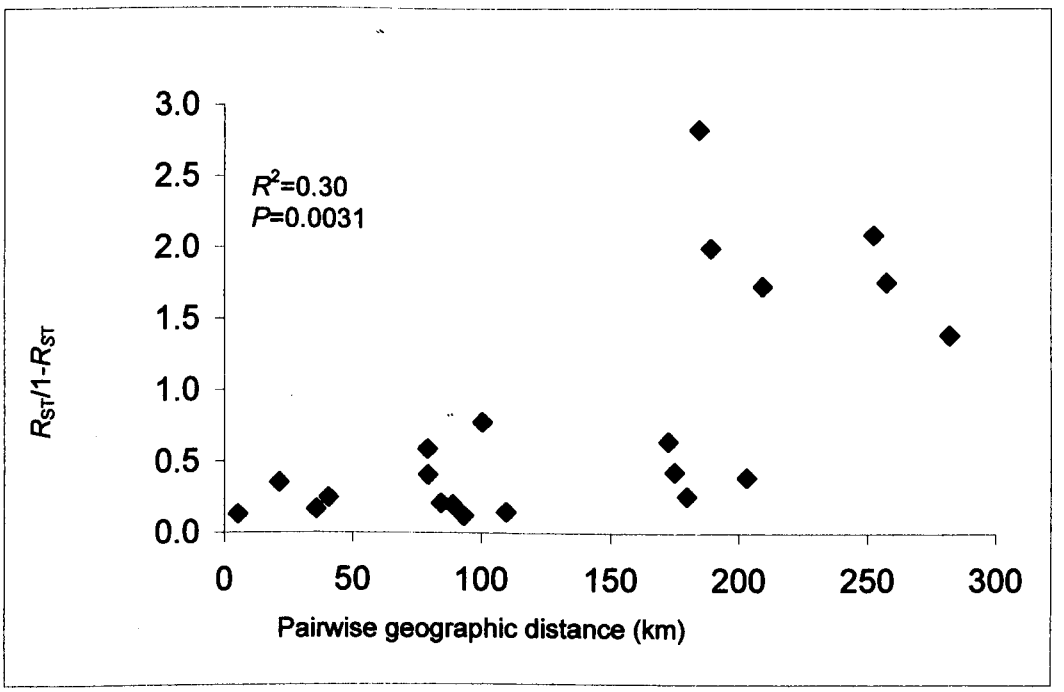

Fig. 7. Geographic (km) distance against genetic distance ($R_{ST}/1\text{-}R_{ST}$) excluding the populations from Rainbow Gorge and Umtamvuna (mist belt populations only).

Population differentiation

There is a high degree of population genetic structure observed between populations of *B. sutherlandii* (R_{ST}=0.634, P<0.001; θ=0.482, P<0.001) which indicates a marked and highly significant deviation from panmixia. These values are high compared to the average values of population differentiation found in other short-lived outcrossing perennials (mean G_{ST}=0.218; Hamrick & Godt, 1996). The marked divergence of the four sub-populations from Kokstaad (R_{ST}=0.238, P<0.01 θ=0.220, P<0.01) indicates that distances of the order of a few kilometres can represent substantial barriers to gene flow. Estimates of the inbreeding coefficient F_{IS} showed a deviation from random mating within populations, with a small but significant deficit of heterozygotes (f =0.154, P<0.001) Within population values ranged from f =-0.054 (Hoha) to f =0.271 (Kokstaad). The high f value for the Kokstaad population is likely to be due at least in part to the sampling of four separate sub populations. The outcrossing rate calculated from F_{IS} estimates (t =0.73) is likely to be deflated by within-population genetic structure to some extent, so this figure perhaps represents a minimum value for the outcrossing rate in *B. sutherlandii*.

The distribution of null alleles, which have been found to be restricted to either a single population (Tygerskloof for the original primers for locus BSU4) or to geographically close pair of populations (i.e., Nkandla and Qudeni for locus BSU3) is further evidence for the lack of gene flow between populations. This is also backed up by the presence of single-nucleotide length mutations, which in the case of locus BSU5 and BSU6 were restricted to a single population (Tygerskloof). Limited gene flow is also supported by the distribution of private alleles, which were present in all populations. Hoha and Dulini have 1 and 3 private alleles respectively, despite being only 5.4 km apart. Tygerskloof has 13 private alleles, which reflects the isolated geographical position of this population. Eight of the private alleles in this population were due to the two loci that showed single base pair polymorphisms; BSU5 contributed 7 of these and BSU6 contributed 1.

It is obvious from these results that the distribution of genetic variation within *B. sutherlandii* in Kwazulu-Natal is strongly linked to geography. Templeton (1998) lists three major biological factors that can cause spatial association with genetic variation. These are (i) restricted gene flow leading to isolation by distance; (ii) range expansion (dispersal) and (iii) range fragmentation.

The first pattern is due to contemporary gene flow, whilst the latter two patterns are effects of population history rather than population structure. The estimators of F_{ST} and R_{ST} indicate that there is a high degree of correlation between geography and genetics, but further analysis is needed to identify the cause of this correlation.

Plots of pairwise geographic distance against pairwise genetic distance can detect patterns of isolation by distance, using a the natural log of distance for a 2-dimensional array of populations or the non-logged geographic distance for a 1-dimensional (linear) array of populations (Rousset, 1997). Outliers on such plots show larger or smaller genetic distances than would be expected from their geographic proximity to other populations. Such outliers can be useful in highlighting populations that are the result of either dispersal or vicariant events, i.e., historical events as opposed to contemporary population structure.

The plots of the mantel tests based on F_{ST} show two populations that deviate from an otherwise highly significant correlation of genetic and geographic distance, namely Umtamvuna and Rainbow Gorge. Both show higher genetic differentiation from their nearest population neighbours than would be expected under a pattern of isolation by distance. Removing these two populations from the analysis reveals a highly significant pattern of isolation by distance in the remaining mist belt populations.

The plots of the mantel tests based on R_{ST} show some of the genetic-geographic comparisons of the Umtamvuna population as conspicuous outliers, but in contrast to the plot using F_{ST} all the pairwise comparisons of Rainbow Gorge appear to be within the main trend of isolation by distance. Although Umtamvuna

appears as aberrant on both plots irrespective of the genetic distance measure used, Rainbow Gorge only lies outwith the main trend if F_{ST} is used.

This discrepancy may be due at least in part to the behaviour of F_{ST} under conditions of varying gene diversity (Charlesworth, 1998; Balloux & Lugon-Moulin, 2002). Under high gene diversity, heterozygosity within populations is more likely to be high. Hence, the value of F_{ST} can be reduced due to the high value of H_S, even if populations have very few alleles in common. Conversely, under conditions of low gene diversity, the value of H_S can be very small and hence F_{ST} can be very high if the populations have few alleles in common. Rainbow Gorge has the lowest gene diversity of any of the populations of *B. sutherlandii* in this study, and this has possibly inflated the values of F_{ST} obtained during population pairwise comparisons relative to the other populations which have higher gene diversity. The highest pairwise F_{ST} value between populations was between Rainbow Gorge and Umtamvuna (Fig 4); Umtamvuna also has markedly low gene diversity compared to the rest of the populations in this study and this could account in part for the high value observed. R_{ST} is unaffected by gene diversity and relies instead on the comparing the size of alleles; using this stepwise mutation based measure Rainbow Gorge does not appear as markedly outwith a trend of isolation by distance with the other populations of *B. sutherlandii* from the mist belt forests.

In common with the mantel test plots, the neighbour joining trees of populations based on a geometric measure of genetic distance (Cavalli-Sforza's & Edward's Chord distance, 1967) hint at a strong geographical-genetic relationship, and the same two populations that are outliers on the mantel plots (Umtamvuna and Rainbow Gorge) also show non-geographic clustering on the neighbour joining tree, clustering instead with populations that are not their nearest neighbour. The Umtamvuna population clusters with Tygerskloof, although this is on very short and unsupported branch. Both Tygerskloof and Umtamvuna show the largest measures of pairwise distance to other populations on the neighbour joining tree. Rainbow Gorge clusters next to Nkandla in a group including Qudeni; the grouping of these three populations has 55% bootstrap support. Aside from Nkandla, which is 170km away from Rainbow Gorge, the other mist belt populations of *B. sutherlandii* are only marginally nearer (Qudeni is 150km away; Ferncliffe 110km; Hoha 107km; Dulini 110km), so there is not a great deal to choose between the prospective geographic

nearest neighbours of Rainbow Gorge. However, given the proximity of the southern mist belt forests to Rainbow Gorge relative to the more isolated northerly patches, (Fig. 1, top) one might expect a relationship to populations from these southern forests to be more likely.

Of the three possible causes of genetic and geographical patterns being correlated, only the central populations of *B. sutherlandii* from the mist belt forests are differentiated according to a pattern that corresponds to one of isolation by distance. Umtamvuna shows no genetic similarity to its nearest populations, and indeed shows no obvious affinity for any other of the populations sampled. Rainbow Gorge shows more differentiation than would be expected under a pattern of isolation by distance from its nearest neighbours when compared using F_{ST}, but not with R_{ST}. Using a geometric measure of genetic distance, Rainbow Gorge appears most similar to a population from the Kwazulu-Natal midlands mist belt forests, Nkandla.

Although the remaining populations show a highly significant pattern of isolation by distance when considered together, this is no longer the case when the northern and southern mist belt populations are considered alone (Fig. 5). In such cases where only a sub set of comparisons is responsible for a pattern of isolation by distance, then vicariance should be considered as an alternative explanation (Bossart & Prowell, 1998). This is likely to be the case in this study, especially as one is likely to encounter vicariant events at larger sampling scales (Bossart & Prowell, 1998), such as between the northern and southern mist belt forests. It should be noted that the population groups within either the northern or southern forests are quite small, and probably not large enough to test for isolation by distance within these areas; further sampling is required to test for isolation by distance at this scale.

Tygerskloof especially seems to be markedly divergent from the other mist belt population samples, as it possesses the highest number of private alleles, unique mutations in the primer sites for BSU4, and unique single-step mutations at BSU5 and BSU6. It also shows a lack of similarity to any other population in the neighbour joining tree based on chord distance. Nkandla and Qudeni also possess null alleles that are not found in any other population.

It seems that the majority of the population differentiation in Kwazulu-Natal populations of *Begonia sutherlandii* is due to historical isolation rather than isolation by distance. This is likely to be linked to the history of forest cover in the province, as *B. sutherlandii* is largely confined to this habitat.

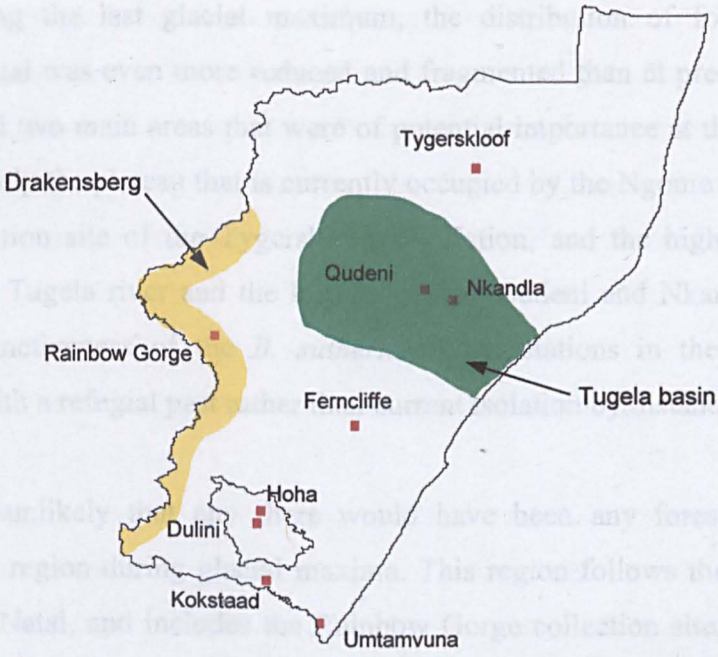

Fig. 8. Schematic map of Kwazulu Natal.

The population from Umtamvuna shows strong genetic divergence from all other populations. The Umtamvuna nature reserve is a forested ravine located on the coast of Southern Kwazulu-Natal near Port Edward. Such coastal forests have been suggested as being important as refugia during the last glaciation (Lawes, 1990), and this is likely to be the case with the forests in Umtamvuna. The Umtamvuna reserve lies in the Natal Pondoland sandstone complex, and this area has a remarkably high number of endemic species (Van Wyk, 1981). In a study of 14 forests spread throughout South Africa, the flora of Umtamvuna was found to be disproportionately rich, and contains several taxa for which Umtamvuna is their only locality in Kwazulu-Natal (Geldenhuys, 1992). This opens the possibility that the *B. sutherlandii* populations in the reserve are relics older than the populations from the scarp forests further inland, and this is congruent with them being genetically more

divergent from their geographically closest populations than other population pairs the same distance apart. The Umtamvuna plants are also distinct morphologically, as they have leaves with a distinct oblong shape which are also densely pubescent. All the other plants sampled for this study were sub-glabrous; the only known locality of pubescent *B. sutherlandii* is in the Transvaal to the north of Kwazulu-Natal (Hilliard, 1967).

During the last glacial maximum, the distribution of forest over all of Kwazulu-Natal was even more reduced and fragmented than at present. Eeley et al. (1999) found two main areas that were of potential importance at this time as forest refugia, namely the plateau that is currently occupied by the Ngome forest and which is the collection site of the Tygerskloof population, and the highland region that parallels the Tugela river and the location of the Qudeni and Nkandla forests. The genetic distinctiveness of the *B. sutherlandii* populations in these localities are congruent with a refugial past rather than current isolation by distance equilibrium.

It is unlikely that any there would have been any forest refugia in the Drakensberg region during glacial maxima. This region follows the western border of Kwazulu-Natal, and includes the Rainbow Gorge collection site. Montane forest currently exists along this border in small fragments in sheltered valleys, but would probably have been eliminated during the last glacial due to the intense cooling at higher altitudes. Cold and desiccating winds emanating from the Drakensberg are likely to have also cause the retreat of forest from the Tugela drainage basin (Lawes, 1990), which is an area of fairly flat topography to the south and west of the higher altitude Nkandla and Qudeni (Fig. 8). This suggests that there may have been no suitable habitat for *B. sutherlandii* in the Drakensberg region during the last glacial, and hence the lack of correlation between genetic and geographic distance cannot be explained in terms of relicuality. It is more likely that the population of *B. sutherlandii* in Rainbow Gorge is the result of a more recent long distance dispersal event, and it is tempting to speculate that the founder may have come from a population in the Kwazulu-Natal midlands near the Nkandla area. This is congruent with the genetic similarity between Nkandla and Rainbow Gorge as shown by results from the neighbour joining analysis of chord distances, and also with the low gene diversity of the population which could indicate a bottleneck during the founding of the population. Further sampling is needed from Drakensberg populations north of

Rainbow Gorge to ascertain whether the montane forests extending along the Drakensberg escarpment have been colonised once or many times from different sources.

The forest cover in the Southern Kwazulu-Natal midlands was reduce to a small number of tiny fragments during the last glacial maximum, although this region currently has a high density of forest compared to the rest of the province. The forest cover was even more extensive and less fragmented under the conditions of the Holocene altithermal ca. 7000 years ago (Eeley et al., 1999). The strong relationship of all the populations of *B. sutherlandii* sampled from the southern mist belt forests (Fig. 3) reflects the relative continuity of the forest in this area relative to the rest of Kwazulu-Natal, and a more recent common ancestry of these populations.

Population isolation in Begonia and implications for speciation

The high degree of population structure and the complete isolation of *B. sutherlandii* in many of the forest patches in Kwazulu-Natal suggests successful long distance dispersal is rare. This is likely to be due to the passive dispersal mechanism of the species. Most *Begonia* species, including *B. sutherlandii*, have a dehiscent, tri-alate fruit. The wings on the fruit have been suggested as assisting in anemochory, with patterning on the surface of the tiny seeds causing micro-turbulence which helps the seeds stay airborne (de Lange & Bouman, 1999). However, may authors on *Begonia* suggest that dispersal is largely passive, with the seeds being dispersed by gravity a short distance from the parent plant, (Burt-Utley, 1985; Agren & Schemske, 1993; de Lange & Bouman, 1999; Matolweni et al., 2000) and this study would seem to confirm this. Also, wind dispersal is not effective in the sheltered conditions of the forest floor habitat that *Begonia* species grow in; true wind dispersal mechanisms are very rarely found in ground-layer forest plants (Hovestadt et al., 1999; Killeen et al., 1998).

The narrow niche of *B. sutherlandii* will also hamper successful colonisation, as suitable habitat exists as very small and highly sporadic patches, even within a forested area. The fragmented distribution of populations is also likely to prevent effective population connection by pollinators. As a result of the lack of connectivity

between populations, within species genetic diversity shows a very strong correlation with geography. This is matched by the distribution of monophyletic groups of species in *Begonia* which tend to have regional distributions (e.g., south Africa, Madagascar) and deeper clades of monophyletic groups, which are bound within continents according to Forrest (2000). Templeton (1998, p.135) suggests the same pattern of restricted geographical clustering on the tips relative to the interiors of clades occurring repeatedly at many clade levels is strong evidence for a recurrent evolutionary force, such as restricted gene flow.

Geographic structure in phylogenetic trees can also be caused through hybridisation of plants that grow in the same region. This is perhaps unlikely to be the case in *Begonia*, where hybrids are rarely encountered under natural conditions (Teo & Kiew, 1999). Two hybrids have been documented from Taiwan (Peng & Chen, 1991; Peng & Chiang, 2000), *Begonia* x *taipensis* and *Begonia* x *buimontana*. Both exist in the wild only as sterile F_1 hybrids. Teo & Kiew (1999) found five hybrid populations between *B. decora* and *B. venusta* in the Cameron highlands in Malaysia, which are both in section *Platycentrum*. In this case, the hybrid was fertile and there was evidence for introgression. The only other well documented case of a fertile hybrid in a natural *Begonia* population is by Sosef (1994), which involved two sister species (*B. susaniae* and *B. vittarifolia*) from section *Scutobegonia* in the Crystal Mountains, Gabon. Hybridisation is possible in cultivation even between quite distantly related species, although such crosses can be of very low fertility (e.g. Gleed, 1961) and are usually only successful if the two parents have the same chromosome number (McGregor, 1969). Despite the crosses achieved with cultivated material, *Begonia* species in nature seem to successfully maintain their integrity. How much of this is due to genetic barriers to hybridisation and how much is due to other factors such as niche differentiation or temporal differences in flowering is open to speculation.

The same patterns of population isolation as found in *B. sutherlandii* have been found in two recent studies on *Begonia* population genetics. Matolweni et al. (2000) estimated F_{ST} to be 0.901 in the *B. dregei* species complex which is endemic to the coastal forests of eastern South Africa. A study of *B. socotrana* by Hughes et al. (in press; paper 4) found significant population structure within the species, despite its range being only ca. 10 x 15 km. *B. socotrana* showed a significant

pattern of isolation by distance, due to either restricted dispersal or local vicariance events in its small native range of the Haggeher mountains on Socotra.

Poor dispersal can be a highly adaptive trait. Cody and Overton (1996) found very high pressures for the evolution of reduced dispersal in *Lactuca* (Asteraceae) on small islands off the Pacific coast of Canada. Island populations quickly evolved seeds with a reduced pappus in response to the fatal consequences for propagules that were dispersed into the sea. Such pressures are likely to operate on any species which occupies an archipelago-like habitat outside which it cannot survive, such as *B. sutherlandii* which is restricted to a specific micro-habitat in indigenous forest patches. Although highly localised dispersal can be advantageous, it can increase the risk of extinction in changing conditions if new patches of suitable habitat cannot be colonised (Cain et al., 2000). Given the wide range of *B. sutherlandii*, the species is undoubtedly capable of long distance dispersal and colonisation, although this does not appear to be frequent enough to prevent the genetic isolation of populations 60 km apart.

The narrow endemism exhibited by most *Begonia* species could be a result of responses to local selection pressures in the absence of gene flow from neighbouring populations growing under different conditions. Gene flow between populations can homogenise allele frequencies and prevent local adaptation (Barton, 2001). The strength of selection a population can respond to is dependent on the number of migrants it receives from populations not experiencing the same environmental conditions. The less the gene flow between populations the weaker the selection pressure they can respond to (Barton & Clarke, 1990). Widespread *Begonia* species that show atypical adaptations for dispersal or pollination that would enhance gene flow between populations tend to show a higher degree of uniformity across their range than other widespread species of *Begonia*. Gene flow may be preventing local adaptation in these cases, and if one considers range size a reflection of niche width (Brown & Lomolino, 1998), then these can be thought of as more uniform, ecologically tolerant species rather than a collection of a number of locally adapted isolates as is the case for *B.sutherlandii*.

The current distribution of forest in Kwazulu-Natal (Eeley et al., 1999) means that other angiosperms with a similar ecology to *Begonia* will also have a highly

fragmented distribution. *Streptocarpus* species are largely limited to forest patches, and like *Begonia* have a preference for shaded and damp conditions. In many cases they show a high degree of morphological differentiation between populations (Hilliard & Burtt, 1971). Present day forest refugia in Kwazulu-Natal and the resulting vicariant events may be important in driving differentiation in other angiosperm taxa.

Gene flow, population differentiation and speciation are undoubtedly linked. Although species radiations are complex in origin and due to the interplay of many factors, it seems plausible that the high degree of population structure seen in *Begonia* has been instrumental in allowing the genus to produce over 1400 species.

References

Acocks, J.P.H. (1988). Veld types of Southern Africa. *Memoirs of the Botanical Survey of South Africa*, **28**: 1-146.

Agren, J., Schemske, D.W. (1991). Pollination by deceit in a Neotropical monoecious herb *Begonia involucrata*. *Biotropica*, **23**: 235-241.

Agren, J., Schemske, D.W. (1993). Outcrossing rate and inbreeding depression in two annual monoeciuos herbs, *Begonia hirsuta* and *B. semiovata*. *Evolution*, **47**: 125-135

Allard, R.W., Jain, S.K., Workman, P.L., (1969). Genetics of inbreeding populations. Advances in Genetics 14. 55-131.

The Angiosperm Phylogeny Group. (1998). An ordinal classification for the families of flowering plants. *Annals of the Missouri Botanical Garden*, **85**: 531-553.

Balloux, F. and Lugon-Moulin, N. (2002). The estimation of population differentiation with microsatellite markers. *Molecular Ecology*, **11**: 155-165.

Barton, N.H. (2001). The evolutionary consequences of gene flow and local adaptation: future prospects. In *Dispersal* (eds. J. Clobert, E. Danchin, A.A. Dhont and J.D. Nichols), pp. 329-340. Oxford University Press, Oxford.

Barton, N.H. & Clarke, A. (1990). Population structure and process in evolution. In *Population biology: ecological and evolutionary viewpoints*. (eds. K. Wohrmann and S. Jain). Springer-Verlag, New York. 456 pp.

Brown, J.H. & Lomolino, M.V. (1998). *Biogeography*. Sinauer Associates, Massachusetts.

Bossart, J.L. & Prowell, D.P. (1998). Genetic estimates of population structure and gene flow: limitations, lessons and new directions. *Trends in Ecology and Evolution*, **13**: 202-206.

Bowcock, A.M., Ruiz-Linares, A., Tomfohrde, J., Minch, E., Kidd, J.R & Cavalli-Sforza, L.L. (1994). High resolution of human evolutionary trees with polymorphic microsatellites. *Nature*, **368**: 455-457.

Burt-Utley K (1985) A revision of the Central-American species of *Begonia* section Gireoudia (*Begoniaceae*). *Tulane Studies in Zoology and Botany* **25**, 1-131.

Cain, M., Milligan, B.G. & Strand, A.E. (2000). Long-distance dispersal in plant populations. *American Journal of Botany*, **87**: 1217-1227

Cavalli-Sforza, L.L. & Edwards, W.F. (1967). Phylogenetic analysis: models and estimation procedures. *Evolution*, **21**: 550-570.

Charlesworth, B. (1998). Measures of divergence between populations and the effect of forces that reduce variability. *Molecular Biology and Evolution*, **15**: 538-453.

Cody, M.L. & Overton, J.M. (1996). Short term evolution of reduced dispersal in island plant populations. *Journal of Ecology*, **84**: 53-61.

Doorenbos J, Sosef, MSM & de Wilde JJFE (1998) The sections of *Begonia*. *Wageningen Agricultural University Papers*, **98**-2, 1-226.

Doyle, J.J. Doyle, J.L., (1987). A rapid DNA isolation procedure for small quantities of fresh leaf material. *Phytochemical Bulletin*, **19**: 11-15.

Eeley, H.A.C., Lawes, M.J. & Piper, S.E. (1999). The influence of climate change on the distribution of indigenous forest in Kwazulu-Natal, South Africa. *Journal of Biogeography*, **26**: 595-617.

Forrest, L.L. (2000). A Phylogeny of Begoniaceae Bercht. & J. Presl. Unpublished PhD thesis, University of Glasgow.

Geldenhuys, C.J. (1992). Richness, composition and relationships of the floras of selected forests in Southern Africa. *Bothalia*, **22**: 205-233.

Gleed, C.J. (1961). Hybrid winter flowering Begonias. *Journal of the Royal Horticultural Society*, **80**: 319-322.

Goodman, S.J. (1997). RST CALC: A collection of computer programs for calculating unbiased estimates of genetic differentiation and determining their significance for microsatellite data. *Molecular Ecology*, **6**: 881-885.

Goudet, J. (2001). FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available from http://www.unil.ch/izea/softwares/fstat.html.

Hilliard, O.M. (1976). Begoniaceae. *Flora of Southern Africa*, **22**: 136-144.

Hovestadt, T., Yao, P. & Linsenmair, E. (1999). Seed dispersal mechanisms and the vegetation of forest islands in a West African forest-savanna mosaic. *Plant Ecology*, **144**: 1-25.

Hamrick, J.L. & Godt, M.J. (1996). Effects of life history traits on genetic diversity in plant species. *Philosophical Transactions of the Royal Society of London series B*, **351**: 1291-1298.

Hilliard, O.M. (1976). Begoniaceae. *Flora of Southern Africa*, **22**: 136-144.

Hilliard, O.M. and B.L. Burtt (1971). *Streptocarpus*, an African Plant Study. University of Natal Press, Pietermaritzburg.

Hughes, M., Hollingsworth, P.M. & Squirrell, J. (2002). Isolation of polymorphic microsatellite markers for *Begonia sutherlandii* Hook. f. *Molecular Ecology Notes*, in press.

Kiew, R. (2001). The limestone Begonias of Sabah, Borneo - flagship species for conservation. *Gardens Bulletin Singapore*, **53**: 241-286.

Killeen, T.J., Jardin, A., Mamani, F. & Rojas, N. (1998). Diversity, composition and structure of a tropical semideciduous forest in the Chiquitania region of Santa Cruz, Bolivia. *Journal of Tropical Biology*, **14**: 803-827.

Lange, A. de, & Bouman, F. (1999). Seed micromorphology of neotropical *Begonias*. *Smithsonian Contributions to Botany*, Number 90.

Langella, O. (2002). Populations. Computer program available from http://www.cnrs-gif.fr/pge/index.php?lang=en

Lawes, M.J. (1990). The distribution of the samango monkey (*Cercopithecus mitis erytharchus* Peters, 1852 and *Cercopithecus mitis labiatus* I. Geoffroy, 1843) and forest history in southern Africa. *Journal of Biogeography*, **17**, 669-680.

Lewis, P.O., Zaykin, D. (2001). Genetic Data Analysis: Computer program for the analysis of allelic data. Version 1.0 (d16c). Available from http://lewis.eeb.uconn.edu/lewishome/software.html

Matolweni, L.O., Balkwill, K. & McLellan, T. (2000). Genetic diversity and gene flow in the morphologically variable, rare endemics *Begonia dregei* and *Begonia homonyma* (*Begonia*ceae). *American Journal of Botany*, **87** (3): 431-439.

McGregor, T. (1969). Chromosomes in *Begonia*. *The Begonian* **36**: 230-232.

Minch, E., Ruix-Linares, A. and Goldstein, D.B. (1995). MICROSAT. Computer program available from http://hpgl.stanford.edu/projects/microsat/

Nei, M. (1987). *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Park, S.D.E., 2001. Microsatellite toolkit. Computer program available from http://acer.gen.tcd.ie/~sdepark/ms-toolkit/

Parks, J.C. & Werth, C.R. (1993). A study of spatial features of clones in a population of bracken fern, Pteridium aquilinium (Dennstaedtiaceae). *American Journal of Botany*, **80**: 537-544.

Peng, C. & Chen, Y. (1991). Hybridity and parentage of *Begonia buimontana* Yamamoto (Begoniaceae) from Taiwan. *Annals of the Missouri Botanical Garden*, **78**: 995-1001.

Peng, C.I. & Chiang, Y.S. (2000). *Begonia* x *taipensis*, a new natural hybrid in Taiwan. *Botanical Bulletin Academica Sinica*, **41**: 151-158.

Plana, V. (2002). Phylogeny and Biogeography of African *Begonia*. Unpublished PhD thesis, University of Glasgow.

Rousset, F. (1997). Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics*, **145**: 1219-1228.

Schneider, S., Roessli, D., Excoffier, L. (2000). ARLEQUIN ver 2.000. A software for population genetics data analysis. Genetics and Biometry Lab, Dept. of Anthropology, University of Geneva. http://lgb.unige.ch/arlequin/index.php3

Sosef, M.S.M. (1994). Refuge Begonias. taxonomy, phylogeny and historical biogeography of *Begonia* sect. *Loasibegonia* and sect. *Scutobegonia* in relation to glacial forest refuges in Africa. (Studies in Begoniaceae 5). Wageningen Agricultural University papers.

Slatkin, M. (1993). Isolation by distance in equilibrium and nonequilibrium populations. *Evolution* **47**: 264-279.

Slatkin, M. (1995). A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**: 457-462.

Swofford, D.L. (1998). PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Takezaki, N. and Nei, M. (1996). Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics*, **144**: 389-399.

Templeton, A.R. (1998). The role of molecular genetics in speciation studies. In *Molecular Approaches to Ecology and Evolution* (eds. R. Desalle and B. Schierwater). Birkhauser, Switzerland. pp. 131-156.

Teo, L. & Kiew, R. (1999). First record of a natural *Begonia* hybrid in Malaysia. *Gardens Bulletin Singapore*, **51**: 103-118.

Van Wyk, A.E. (1981). Vir 'n groter Umtamvuna-natuurreservaat. *Journal of Dendrology*, **1**: 106-108.

Weir, B.S. and Cockerham, C.C. (1984). Estimating *F*-statistics for the analysis of population structure. *Evolution*, **38**: 1358-1370.

White, F. (1983). The vegetation of Africa; a descriptive memoir to accompany the UNESCO/AETFAT/UNSA vegetation map of Africa. Natural Resources Research No. 20. UNESCO, Paris.

CHAPTER 5. Population structure and speciation: perspectives from other species.

5.1 Introduction

This thesis has investigated microevolutionary patterns at the population level in *Begonia*, which reveal a high degree of correlation between genetic variation and geography. These patterns are congruent with those seen at the macroevolutionary scale, and thus it seems reasonable to suggest that restricted gene flow between populations has contributed to the high number of species and the high degree of narrow endemism seen in the genus (Hughes et al., 2002, paper 4; Hughes and Hollingsworth, 2002, paper 5). This chapter will review data on population differentiation for the sister family to Begoniaceae, the Datiscaceae, and for the angiosperms as a whole, to examine whether similar forces have been relevant to the evolution of these groups as well as *Begonia*.

5.2 Population structure and the evolution of *Datisca*.

The Datiscaceae is a ditypic family of wind-pollinated, tall, long-lived perennials which occur in riparian habitats. *D. cannabina* is native to south-western and central Asia, whilst *D. glomerata* is distributed from northern California to Baja California in northern Mexico. The seeds of both *Datisca* species are broadly similar to those of *Begonia* in size and ornamentation, and are shed from dehiscent capsules. Both species are wind pollinated (Liston et al., 1989),with *D. cannabina* being dioecious and *D. glomerata* being androdioecious.

Despite being wind pollinated, there is a very high degree of population isolation in both *Datisca* species, as determined in an allozyme study by Liston et al. (1989); G_{ST}=0.975 for *D. cannabina* and G_{ST}=0.896 for *D. glomerata*. These high levels of differentiation may be due in part to the wide range over which the species were sampled, which included some very disjunct populations. This lack of

gene flow between populations, however, has not promoted differentiation or high levels of speciation in *Datisca*, and contrasts markedly with the situation seen in *Begonia*. Both *Datisca* species are morphologically and anatomically quite similar to each other (Davidson, 1973) with young plants of the two species being virtually indistinguishable and differing as adults only in breeding system; *D. cannabina* is strictly dioecious, whilst *D. glomerata* is androdioecious and exists as hermaphrodite and male individuals. This represents a degree of morphological stasis, as there is evidence that the two species have been separate for at least 10 million years (Liston et al., 1989). They also show a high degree of intra-specific morphological uniformity (Liston et al., 1989). Such evolutionary stasis can be considered to be the result of either stabilising selection or genetic constraints (Williamson, 1987). Liston et al. suggest that stabilising selection has been important in maintaining uniformity within and between *Datisca* species, as there is a high degree of differentiation at neutral allozyme loci which they would not expect if there were strong genetic constraints. However, a more objective and comparable measure of genetic change over time can be obtained from molecular phylogenetic trees, and within the Cucurbitales, *Datisca* has the shortest branches (57 changes from the basal node) and *Begonia* the longest (136 changes from the basal node; APG, 1998). This suggests that there is also a degree of genetic stasis in the Datiscaceae, which may be a contributing factor to the lack of evolutionary change observed in the family.

5.3 Population structure and the evolution of angiosperm biodiversity

There is a considerable amount of speculation in the literature on the effect of dispersal and pollination syndromes have on population structure and speciation (Crepet, 1984; Sytsma and Schaal, 1985; Eriksson and Bremer, 1991, 1992; Bawa, 1992; Oakwood et al., 1993; Rickleffs and Renner, 1994; Tiffney and Mazer, 1995; Goldblatt, 1997; Smith, 2001). If a high degree of population isolation promotes speciation as a general rule, then this should be visible in a correlation of F_{ST} and speciation rate. Fig 5.1 represents a comparison between the estimates of F_{ST} for clades of angiosperms that were defined as being extremely species rich or having expected diversity (compared to the background diversification rate) by Magallon

and Sanderson (2001); this comparison is more objective than using family of genus size as a measure of diversification rate. The F_{ST} estimates were obtained from Hamrick and Godt (1996); for each clade type n=10 (including several estimates obtained from the world wide web). No F_{ST} estimates could be found for any species in the clades that were determined to be extremely species poor.

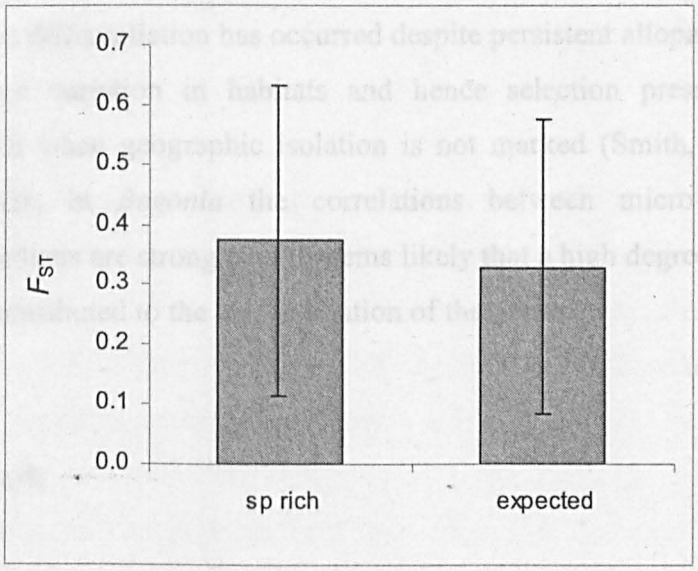

Figure 5.1. Estimates of F_{ST} in angiosperm clades, defined as 'extremely species rich' or 'expected diversity' in Magallon and Sanderson, (2001).

There is no evidence for a correlation between population structure and diversification rate in the angiosperms as a whole. The average values for F_{ST} in the two clade types are similar, and as can be seen from the standard errors, there is a wide range of values within each clade type. The clades are as defined by the APG (1998), and many of these clades contain families and genera that are physiologically very different. It is conceivable (although perhaps unlikely) that this broad analysis of diversification may be concealing a pattern that would become evident at a finer scale.

5.4 Conclusions

If one thing is evident from nature, it is that every radiation is different. The lack of an indication that population structure and speciation are linked either in *Datisca* or in the angiosperms as a whole emphasises there is more to evolution than gene flow. Selection is also of prime importance, and can lead to cases where little phenotypic differentiation has occurred despite persistent allopatry (Schneider, 1999) or where variation in habitats and hence selection pressures leads to divergence even when geographic isolation is not marked (Smith, 2001; Moritz, 2000). However, in *Begonia* the correlations between micro- and macro-evolutionary patterns are strong, and it seems likely that a high degree of population structure has contributed to the large radiation of the genus.

5.5 Further work

5.5.1 Transplant experiments

Given the strong evidence for population isolation having played a large role in the evolution of *Begonia* biodiversity, it would be interesting to examine population evolution and divergence in more detail. *Begonia sutherlandii* shows a great range of leaf size and shape. Each population is uniform for a given leaf morphology, with the majority of the variation being partitioned between populations. Fig. 5.2 shows representative leaf shapes and sizes from some of the populations sampled in Kwazulu-Natal. Initial common-garden experiments with plants grown from wild-collected seed show that much of this variation is genetic, although leaf size shows a degree of plasticity. This is in agreement with observations by Hilliard (1967) in her account of *Begonia* for the Flora of South Africa. It is interesting to note some convergences in the leaf shapes of South African *Begonia*. *B. dregei* also shows a high degree of variation in leaf shape between populations, with one population mimicking very closely the leaf shape of a fern which co-occurs in the same habitat (T. McLellan, pers. com., 2000). *B. sutherlandii* has some populations with highly dissected, feathery leaves which were originally described as different species (*B. dissecta* Irmsch., *B. buttonii*

Irmsch.). Some such populations in coastal forest patches just south of Durban have a leaf shape which is remarkably similar to *B. dregei* from the same area. This suggests that leaf shape is a highly adaptive trait in South African *Begonia*, with local selection favouring different leaf forms in different areas. This hypothesis of local adaptation in *Begonia* populations could be tested with reciprocal transplant experiments.

5.5.2 Pollination ecology

Accounts of pollination in *Begonia* are rare, and nothing is known about the influence of pollinator behaviour on between-population gene flow. It would be useful to identify the pollinators of *B. sutherlandii* and study their ecology and range size; this would give some idea of the distance that would isolate two populations from pollen-mediated gene flow.

Given the self compatibility of *B. socotrana* and *B. sutherlandii*, the high degree of outcrossing achieved by both these species is perhaps surprising. The effect of pollen competition could be tested by applying a self/non-self mixture of pollen to receptive female flowers, and paternity testing the offspring to see which, if any, was more successful. The opening times of male and female flowers on individuals plants and the development of stigma receptivity could be examined to see if temporal events are instrumental in reducing the amount of selfing.

5.5.3 Reproductive isolation

How do *Begonia* species maintain their identity in nature? The relative influences of genetic, ecological and ethological reproductive barriers are unknown. There is some evidence for incipient genetic isolation in the *B. dregei* complex (T.McLellan, pers. com., 2000), with crosses between some *B. dregei* populations and populations that were originally described as *B. rudatisii* showing reduced seed set. Also, crosses between *B. dregei* populations with markedly different leaf shapes produce offspring with malformed leaves. Hybridisation experiments were attempted with *B. sutherlandii* plants grown from seed from different populations in Kwazulu-Natal, in order to look for reduced seed set between distantly related

populations which had not exchanged genes for a considerable time. Unfortunately, an outbreak of mildew caused the plants to die back and no results could be obtained. It would be worth repeating these crossing experiments, and following them through to an F_2 generation, as genetic reproductive isolation may take more then one generation to reveal itself.

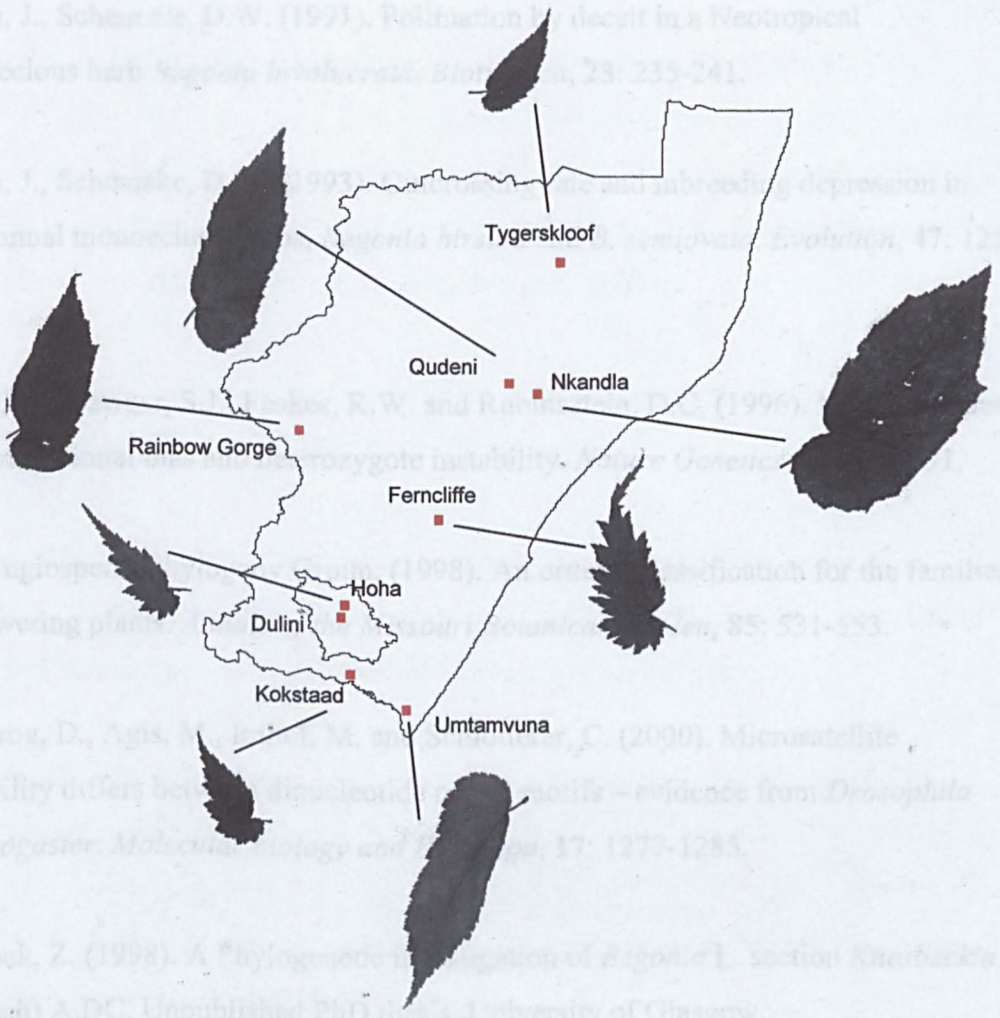

Figure 5.2. Leaf shapes of wild-collected material of *B. sutherlandii* in Kwazulu-Natal.

BIBLIOGRAPHY

Abbott, R.J., Noltie, H.J. and Ingram, R. (1983). The origin and distribution of *Senecio cambrensis* in Edinburgh. *Transactions of the Botanical Society of Edinburgh*, **44**: 103-106.

Agren, J., Schemske, D.W. (1991). Pollination by deceit in a Neotropical monoecious herb *Begonia involucrata*. *Biotropica*, **23**: 235-241.

Agren, J., Schemske, D.W. (1993). Outcrossing rate and inbreeding depression in two annual monoeciuos herbs, *Begonia hirsuta* and *B. semiovata*. *Evolution*, **47**: 125-135.

Amos, W., Sawcer, S.J., Feakes, R.W. and Rubinsztein, D.C. (1996). Microsatellites show mutational bias and heterozygote instability. *Nature Genetics*, **13**: 390-391.

The Angiosperm Phylogeny Group. (1998). An ordinal classification for the families of flowering plants. *Annals of the Missouri Botanical Garden*, **85**: 531-553.

Bachtrog, D., Agis, M., Imhof, M. and Schlotterer, C. (2000). Microsatellite variability differs between dinucleotide repeat motifs – evidence from *Drosophila melanogaster*. *Molecular biology and Evolution*, **17**: 1277-1285.

Badcock, Z. (1998). A Phylogenetic investigation of *Begonia* L. section *Knesbeckia* (Klotsch) A.DC. Unpublished PhD thesis, University of Glasgow.

Bak, P. and Sneppen, K. (1993). Punctuated equilibrium and criticality in a simple model of evolution. *Physical Review Letters* **71**: 4083-4086.

Bak, P. (1997). *How Nature Works: The science of self organised criticality*. Oxford University Press, Oxford.

Balloux, F. and Goudet, J. (2002). Statistical properties of population differentation estimators under stepwise mutation in a finite island model. *Molecular Ecology*, **11**: 771-783.

Balloux, F. and Lugon-Moulin, N. (2002). The estimation of population differentiation with microsatellite markers. *Molecular Ecology*, **11**: 155-165.

Barkley, F.A. (1968). *Begoniaceae to 1763*. Hopkins Press, Providence.

Barraclough, T.G., Vogler, A.P. and Harvey, P.H. (1999). Revealing the factors that promote speciation. In *Evolution of Biological Diversity* (eds. A.E. Magurran and R.M. May), pp. 202-216. Oxford University Press, Oxford.

Barraclough, V. and Savolainen, T.G. (2001). Evolutionary rates and species diversity in flowering plants. *Evolution* **55**: 677-683.

Barton, N.H. (2001). The evolutionary consequences of gene flow and local adaptation: future prospects. In *Dispersal* (eds. J. Clobert, E. Danchin, A.A. Dhont and J.D. Nichols), pp. 329-340. Oxford University Press, Oxford.

Bawa, K.S. (1980). Evolution of dioecy in flowering plants. *Annual Review of Ecology and Systematics*, **11**: 15-40.

Bawa, K.S. (1992). Mating systems, genetic differentiation and speciation in tropical rainforest plants. *Biotropica*, **24**: 250-255.

Bossart, J.L. and Prowell, D.P. (1998). Genetic estimates of population structure and gene flow: limitations, lessons and new directions. *Trends in Ecology and Evolution*, **13**: 202-206.

Bowcock, A.M., Ruiz-Linares, A., Tomfohrde, J., Minch, E., Kidd, J.R and Cavalli-Sforza, L.L. (1994). High resolution of human evolutionary trees with polymorphic microsatellites. *Nature*, **368**: 455-457.

Brown, J.H. and Lomolino, M.V. (1998). *Biogeography*. Sinauer Associates, Massachusetts.

Burlando, B. (1990). The fractal dimension of taxonomic systems. *Journal of Theoretical Biology*, **146**: 99-146.

Burt-Utley, K. (1985). A revision of the Central-American species of *Begonia* section *Gireoudia* (Begoniaceae). *Tulane Studies in Zoology and Botany* **25**, 1-131.

Bush, G.L. (1975). Modes of animal speciation. *Annual Review of Ecology and Systematics*, **6**: 339-364.

Bush, G.L. (1994). Sympatric speciation: new wine in old bottles. *Trends in Ecology and Evolution*, **9**: 285-288.

Carson, H. L. (1975). The genetics of speciation at the diploid level. *American Naturalist* **109**: 83-92.

Carson, H. L. and A. R. Templeton, (1984). Genetic revolutions in relation to speciation phenomena: The founding of new populations. *Annual Review of Ecology and Systematics*, **15**: 97-131.

Cavalli-Sforza, L.L. and Edwards, W.F. (1967). Phylogenetic analysis: models and estimation procedures. *Evolution*, **21**: 550-570.

Charlesworth, B. (1998). Measures of divergence between populations and the effect of forces that reduce variability. *Molecular Biology and Evolution*, **15**: 538-453.

Chakraborty, R., Kimmel, M., Stivers, D.N., Davison, L.J. and Deka, R. (1997). Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. *Proceedings of the National Academy of Sciences USA*, **94**: 1041-1046.

Chown, S.L. and Gaston, K.J. (2000). Areas, cradles and museums: the latitudinal gradient in species richness. *Trends in Ecology and Evolution*, **15**: 311-315.

Clayton, W.D. (1974). The logarithmic distribution of angiosperm families. *Kew Bulletin*, **29**: 271-279.

Cody, M.L. and Overton, J.M. (1996). Short term evolution of reduced dispersal in island plant populations. *Journal of Ecology*, **84**: 53-61.

Corff, J. Le, Agren, J., Schemske, D.W. (1998). Floral display, pollinator discrimination, and female reproductive success in two monoecious *Begonia* species. *Ecology*, **79**: 1610-1619.

Coyne, J.A., Barton, N.H. and Turelli, M. (1997). Perspective: a critique of Sewall Wright's shifting balance theory of evolution. *Evolution*, **51**: 643-671.

Coyne, J.A., Barton, N.H. and Turelli, M. (2000). Is Wright's shifting balance process important in evolution? *Evolution*, **54**: 306-317.

Crepet, W.L. (1984). Advanced (constant) insect pollination mechanisms: pattern of evolution and implications *vis-à-vis* angiosperm diversity. *Annals of the Missouri Botanical Garden*, **71**: 607-630.

Cronk, Q.C.B. (1989). Measurement of historical and biological influences on plant classifications. *Taxon* **38**: 357-370.

Culley, T.M., Wallace, L.E. Gengler-Nowak, K. and Crawford, D.J. (2002). A comparison of two methods of calculating G_{ST}, a genetic measure of population differentiation. *American Journal of Botany* **89**: 460-465.

Davidson, C. (1973). An anatomical and morphological study of the Datiscaeae. *Aliso*, **8**: 49-110.

Dial, K.P. and Marzluff, J.M. (1989). Nonrandom diversification within taxonomic assemblages. *Systematic Zoology*, **38**: 26-37.

Dodd, M.E., Silvertown, J., Chase, M.W. (1999). Phylogenetic analysis of trait evolution and species diversity variation among angiosperm families. *Evolution*, **53**: 732-744.

Doorenbos, J., Sosef, M.S.M. and de Wilde, J.J.F.E. (1998). The sections of *Begonia*. *Wageningen Agricultural University Papers*, **98-2**: 1-226.

Ebstein, R.P., Novick, O., Umansky, R., Priel, B., Osher, Y. (1996). Dopamine D4 receptor (*D4DR*) exon III polymorphism associated with the human personality trait of Novelty Seeking. *Nature Genetics*, **12**: 78-80.

Eeley, H.A.C., Lawes, M.J. and Piper, S.E. (1999). The influence of climate change on the distribution of indigenous forest in Kwazulu-Natal, South Africa. *Journal of Biogeography*, **26**: 595-617.

Epplen, C., Melmer, G., Siedlaczck, Scwaiger, F.W., Maueler, W. and Epplen, J.T. (1993). On the essence of 'meaningless' simple repetitive DNA in eukaryote genomes. In *DNA fingerprinting. State of the science* (eds. Pena, S.D.J., Chakraborty, R., Epplen, J.T. and Jeffreys, A.J.), pp. 29-45. Birkhauser, Berlin.

Eriksson, O. and Bremer, B. (1991). Fruit characteristics, life forms, and species richness in the plant family Rubiaceae. *American Naturalist*, **138**: 751-761.

Eriksson, O. and Bremer, B. (1992). Pollination systems, dispersal modes, life forms, and diversification rates in angiosperm families. *Evolution*, **46**: 258-266.

Erlich, P.R., and Raven, P.H. (1969). Differentiation of populations. *Science*, **165**: 1228-1232.

Estoup, A. and Cornuet, J. (1996). Microsatellite evolution: inferences from population data. In *Microsatellites: evolution and applications* (eds. D.B. Goldstein and C. Schlotterer), pp. 49-65. Oxford University Press, Oxford.

Excoffier, L. (2000). Analysis of population subdivision. In *Handbook of Statistical Genetics* (eds. D.J. Balding, M. Bishop and C. Cannings), pp. 271-307. John Wiley & Sons, New York.

Exell, A.W. (1973). Angiosperms of the islands in the Gulf of Guinea (Fernando Po, Principe, Sao Tome and Annobon). *Bulletin of the Bristish Museum (Natural History), Botany*, **4**: 327-411.

Feder, J.L., Chilcote, C.A. and Bush, G.L. (1988). Genetic differentiation between sympatric host races of *Rhagoletis pomonella*. *Nature*, **336**: 61-64.

Fisher, R.A. (1930). *The Genetical Theory of Natural Selection*. Oxford University Press, Oxford.

Flores, C. and Engels, W. (1999). Microsatellite instability in *Drosophila spellchecker1* (MutS homolog) mutants. *Proceedings of the National Academey of Sciences USA*, **96**: 2964-2969.

Forrest, L.L. (2000). A Phylogeny of Begoniaceae Bercht. and J. Presl. Unpublished PhD thesis, University of Glasgow.

Freudenreich, C.H., Stavenhagen, J.B. and Zakian, V. (1997). Stability of a CTG/CAG trinucleotide repeat in yeast is dependent on its orientation in the genome. *Molecular and Cellular Biology*, **17**: 2090-2098.

Garcia-Ramos, G. and Kirkpatrick, M. (1997). Genetic models of adaptation and gene flow in peripheral populations. *Evolution*, **51**: 21-28.

Gaston K.J. and Chown, S.L. (1999). Geographic range size and speciation. In *Evolution of biological diversity* (eds. A.E. Magurran and R.M. May), pp.236-259. Oxford University Press, Oxford.

Gavrilets, S. and Hastings, A. (1996). Founder effect speciation: A theoretical reassesment. *The American Naturalist*, **147**: 466-491.

Gavrilets, S., Li, H., and Vose, M. (2000). Patterns of parapatric speciation. *Evolution* **54**: 1126-1134.

Gill, D.E. (1989). Fruiting failure, pollinator inefficiency, and speciation in orchids. In *Speciation and its consequences* (eds. D. Otte and J.A. Endler), pp. 458-481. Sinauer Associates Inc., Massachusetts.

Givnish, T.J., Evans, T.M., Pires, J.C. and Sytsma, K.J. (1999). Polyphyly and convergent morphological evolution in Commelinales and Commelinidae: evidence from *rbcL* sequence data. *Molecular Phylogenetics and Evolution*, **12**: 360-385.

Goldblatt, P. (1997). Floristic diversity in the Cape Flora of South Africa. *Biodiversity and Conservation*, **6**: 359-377.

Goldstein, D.B., Linares, A.R., Cavalla-Sforza, L.L. and Feldman, M.W. (1995a). An evaluation of genetic distances for use with microsatellite loci. *Genetics*, **139**: 463-471.

Goldstein, D.B., Linares, A.R., Cavalla-Sforza, L.L. and Feldman, M.W. (1995b). Genetic absolute dating based on microsatellites and the origin of modern humans. *Proceedings of the National Academy of Sciences USA*, **92**: 6723-6727.

Goldstein, D.B. and Pollock, D.D. (1997). Launching microsatellites: a review of mutation processes and methods of phylogenetic inference. *Journal of Heredity*, **88**: 335-342.

Grant, V. (1949). Pollination systems as isolating mechanisms. *Evolution*, **3**: 82-97.

Grant, V. (1971). *Plant Speciation*. Columbia University Press, New York.

Green, D.M. (1991). Chaos, fractals and nonlinear dymanics in evolution and phylogeny. *Trends in Ecology and Evolution*, **6**: 333-337.

Hamrick, J.L., Linhart, Y.B. and Mitton, J.B. (1979), Relationships between life history characteristics and electrophoretically detectable genetic variation in plants. *Annual review of Ecology and Systematics*, **10**: 173-200.

Hamrick, J. L. and Godt, M. J. W. (1996). Effects of life history traits on genetic diversity in plant species. *Philosophical Transactions of theRoyal Society of London* **351**: 1291-1298.

Hancock, J.M. (1996). Microsatellites and other simple sequences: genomic context and mutational mechanisms. In *Microsatellites: Evolution and Applications* (eds. D.B. Goldstein and C. Schlotterer), pp. 1-9. Oxford University Press, Oxford.

Heale, S.M., and Petes, T.D. (1995). The stabilization of repetetive tracts of DNA by variant repeats requires a functional DNA mismatch repair system. *Cell*, **83**: 539-545.

Henderson, S.T. and Petes, T.D. (1992). Instability of simple sequence DNA in *Saccharomyces cerevisiae. Molecular and Cell Biology*, **12**: 2749-2757.

Heywood, E. (1993). *The Flowering Plants of the World*. B.T. Batsford, London.

Hilliard, O.M. (1976). Begoniaceae. *Flora of Southern Africa*, **22**: 136-144.

Hovestadt, T., Yao, P. and Linsenmair, E. (1999). Seed dispersal mechanisms and the vegetation of forest islands in a West African forest-savanna mosaic. *Plant Ecology*, **144**: 1-25.

Hubbell, S.P. (1997). A unified theory of biogeography and relative species abundance and its application to tropical rain forests and coral reefs. *Coral Reefs*, **16**, Suppl: S9-S21.

Hubbell, S.P. (2001). *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press, Princeton.

Hughes, M., Hollingsworth, P.M. and Miller, A.G. (2002). Population structure in the island endemic *Begonia socotrana*. *Biological Conservation*. Submitted.

Hughes, M. and Hollingsworth, P.M. (2002). Marked population differentiation in the sporadically distributed herb *Begonia sutherlandii*. In prep.

Irmscher, E. (1961). Monographische revision der Begoniaceen Afrikas 1. Sekt. Augustia und Rostrobegonia sowie einige neue Sippen aus anderen Sektionen. *Botanische Jahrbuche,* **81**: 106-188.

Jain, S.K. and Bradshaw, A.D. (1966). Evolutionary divergence among adjacent plant populations. I. The evidence and its theoretical analysis. *Heredity*, **21**: 407-441.

Jorgensen, P.M. and Leon-Yanez, S. (eds.) (1999). *Catalogue of the Vascular Plants of Ecuador*. Missouri Botanical Garden Press, Missouri.

Kashi, Y. and Soller, M. (1999). Functional roles of microsatellites and minisatellites. In *Microsatellites: evolution and applications* (eds. Goldstein, D.B. and Schlotterer, C.), pp 10-23. Oxford University Press, Oxford.

Kauffman, S. (1995). *At home in the Universe. The search for the laws of self organisation and complexity*. Penguin, London.

Kiew, R. (2001). The limestone Begonias of Sabah, Borneo – flagship species for conservation. *Gardens Bulletin Singapore*, **53**: 241-286.

Killeen, T.J., Jardin, A., Mamani, F. and Rojas, N. (1998). Diversity, composition and structure of a tropical semideciduous forest in the Chiquitania region of Santa Cruz, Bolivia. *Journal of Tropical Biology*, **14**: 803-827.

Kimura, M. and Crow, J.F. (1964). The number of alleles that can be maintained in a finite population. *Genetics*, **49**: 725-738.

Klotsch, F.K. (1854). Begoniaceen – Gattungen und Arten. *Abhandlungen der Konigl Akademie der Wissenschraft zu Berlin*, 1-135.

Lange, A. de. (1988). *Begonia* seeds and their adaptations to dispersal. *Acta Botanica Neerlandica* **37**: 322.

Lange, A. de, and Bouman, F. (1991). Seed micromorphology of the genus *Begonia* in Africa: taxonomic and ecological implications. *Wageningen Agricultural University Papers* **91-4**: 1-82.

Lange, A. de, and Bouman, F. (1986). Micromorphology of the seeds in *Begonia* section *Solananthera* A.DC. *Acta Botanica Neerlandica* 35: 489-495.

Lange, A. de, and Bouman, F. (1999). Seed micromorphology of neotropical begonias. *Smithsonian Contributions to Botany*, Number 90.

Latter, B.D.H. (1972). Selection in finite populations with multiple alleles. *Genetics* **70**: 475-490.

Lee, D.W. and Graham, R. (1986). Leaf optical properties of rainforest sun and extreme shade plants. *American Journal of Botany*, **73**: 1100-1108.

Levin, D.A. (2000). *The origin, expansion and demise of plant species*. Oxford University Press, New York.

Levinson, G. and Gutman, G.A. (1987*a*). Slipped-strand mispairing: A major mechanism for DNA sequence evolution. *Molecular Biology and Evolution* **4**: 203-221.

Levinson, G. and Gutman, G.A. (1987*b*). High frequencies of short frameshifts in poly-CA/TG tandem repeats borne by bacteriophage M13 in Escherichia coli K-12. *Nucleic Acids Research*, **15**: 5323-5328.

Liston, A., Rieseberg L.H. and Elias, T.S. (1989). Morphological stasis and molecular divergence in the intercontinental disjunct genus *Datisca* (Datiscacaea). *Aliso*, **12**: 525-542.

Liston, A., Rieseberg L.H. and Elias, T.S. (1990). *Datisca glomerata* is functionally androdioecious. *Nature*, **343**: 641-642.

Lynch, J.D. (1989). The gauge of speciation: on the frequencies of modes of speciation. In *Speciation and its consequences*, (eds.D. Otte and J. A. Endler), pp 527-553. Sunderland, Mass.: Sinauer Associates.

Mabberly, D.J. (1997). *The plant-book: a portable dictionary of the higher plants*, 2nd edition. Cambridge University Press, Cambridge.

Magallon, S. and Sanderson, M.J. (2001). Absolute diversification rates in angiosperm clades. *Evolution*, **55**: 1762-1780.

Mandelbrot, B. (1982). *The Fractal Geometry of Nature*. Freeman, San Fransisco.

Marchant, C.J. (1967). Evolution in *Spartina* (Graminae). 1. The history and morphology of the genus in Britain. *Journal of the Linnaean Society of London (Botany)*, **60**: 1-24.

Matolweni, L.O., Balkwill, K. and McLellan, T. (2000). Genetic diversity and gene flow in the morphologically variable, rare endemics *Begonia dregei* and *Begonia homonyma* (Begoniaceae). *American Journal of Botany*, **87**: 431-439.

Marzluff, J.M. and Dial, K.P. (1991). Life history correlates of taxonomic diversity. *Ecology*, **72**: 428-439.

Mayr, E. (1954). Change of genetic environment and evolution. In *Evolution as a Process* (eds. Huxley, J.S., Hardy, A.C., and Ford, E.B.), pp. 157-180. Allen and Unwin, London.

Mayr, E. (1963). *Animal species and evolution*. Belknap Press, Cambridge, Massachusetts

Mayr, E. (1982). *The growth of biological thought; diversity, evolution and inheritance*. Harvard University Press, Cambridge, Massachusetts.

Minch, E., Ruix-Linares, A. and Goldstein, D.B. (1995). MICROSAT. http://hpgl.stanford.edu/projects/microsat/

Minelli, A., Fuscos, G., and Sartori, S. (1995). Self similarity in biological classifications. *Biosystems*, **26**: 89-97.

Minelli, A. (1993). The unequal distribution of taxonomic diversity. In *Biological Sytematics – The State of the Art*, pp. 185-197. Chapman and Hall, London.

Moritz, C. (2000). A molecular perspective on the conservation of diversity. In *The Biology of Biodiversity* (ed. M. Kato), pp. 21-34. Tokyo: Springer-Verlag.

Muenchow, G.E. (1987). Is dioecy associated with fleshy fruit? *American Journal of Botany*, **74**: 287-293.

Nei, M. (1987). *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Nei, M. and Kumar, S. (2000). *Molecular Evolution and Phylogenetics*. Oxford University Press, Oxford.

Oakwood, M., Jurado, E., Leishman, M. and Westoby, M. (1993). Geographic ranges of plant species in relation to dispersal morphology, growth form and diaspore weight. *Journal of Biogeography*, **20**: 563-572.

Ohta, T. and Kimura, M. (1973). A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genetical Research*, **22**: 201-204.

Ouberg, N.J., Piquot, Y. and Van Groenendael, J.M. (1999). Population genetics, molecular markers and the study of dispersal in plants. *Journal of Ecology*, **87**: 551-568.

Paetkau, D., Waits, L.P., Clarkson, P.L., Craighead, L. and Strobeck, C. (1997). An empirical evaluation of genetic distance statistics using microsatallite data from bear (*Ursidae*) populations. *Genetics*, **147**: 1943-1957.

Plana, V. (2002). Phylogenetics and biogeography of African *Begonia*. Unpublished PhD thesis, University of Glasgow.

Provosti, A.J., Ocana, J. and Alonso, G. (1975). Distances between populations of *Drosophila subobscura* based on chromosome arrangement frequencies. *Theoretical and Applied Genetics*, **45**: 231-241.

Primmer, C.R., Ellegren, H., Saino, N. and Moller, A.P. (1996). Directional evolution in germline microsatellite mutations. *Nature Genetics*, **13**: 391-393.

Raikow, R.J. (1986). Why are there so many kinds of passerine birds? *Systematic Zoology*, **37**: 76-79.

Rice, W.R. and Hostert, E.E. (1993). Laboratory experiments on speciation: what have we learned in 40 years? *Evolution*, **47**: 1637-1653.

Richardson, J.E., Pennington, R.T., Pennington, T.D. and Hollingsworth, P.M. (2001). Rapid diversification of a species-rich genus of neotropical rain forest trees. *Science*, **293**: 2242-2244.

Rickleffs, R.E. and Renner, S.S. (1994). Species richness within families of flowering plants. *Evolution*, **48**: 1619-1636.

Rieseberg, L.H., and Burke, J.J. (2001). The biological reality of species: gene flow, selection, and collective evolution. *Taxon*, **50**: 47-67.

Rogers, J.S. (1972). Measures of genetic similarity and genetic distance. In *Studies in Genetics VII*, pp. 145-453. Univeristy of Texas, Austin.

Rose, O. and Falush, D. (1998). A threshold size for microsatellite expansion. *Molecular Biology and Evolution*, **15**: 613-615.

Rousset, F. (1997). Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. *Genetics*, **145**: 1219-1228.

Sanderson, M.J. and Wojciechowski, M.F. (1996). Diversification rates in a temperate legume clade: are there "so many species" of *Astragalus* (Fabaceae)? *American Journal of Botany*, **83**: 1488-1502.

Sawyer, L.A., Hennessy, J.M., Peixoto, A.A., Rosato, E., and Parkinson, H. (1997). Natural variation in a *Drosphila* clock gene and temperature compensation. *Science*, **278**: 2117-2120.

Schemske, D.W., and Agrem, J. (1995). Deceit pollination and selection on female flower size in *Begonia involucrata*. An experimental approach. *Evolution*, **49**: 207-214.

Schemske, D.W., Agren, J., Corff, J. Le. (1996). Deceit pollination in a tropical monoecious herb *Begonia oxoacana*. In *Floral biology: studies on floral evolution in animal-pollinated plants* (eds. D.G. Lloyd and S.C.H Barrett), pp. 292-318. Chapman and Hall, New York.

Shriver, M.D., Jin, L., Boerwinkle, E., Deka, R., Ferrel, R.E. and Chakraborty, R. (1995). A novel measure of genetic distance for highly polymorphic tandem repeat loci. *Molecular Biology and Evolution*, **12**: 914-920.

Sia, E.A., Jinks-Robertson, S. and Petes, T.D. (1997). Genetic control of microsatellite stability. *Mutation Research*, **383**: 61-70.

Simpson. G.G. (1961). *Principles of animal taxonomy. The species and lower categories*. Columbia University Press, New York.

Slatkin, M. (1995). A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**: 457-462.

Slatkin, M. and Barton, N.H. (1989). A comparison of three indirect methods for estimating average levels of gene flow. *Evolution*, **43**: 1349-1368.

Smith, L.B. and Wasshausen, D.C. (1986). 133. Begoniaceae. In *Flora of Ecudaor 25* (eds. Harling, G. and Andersson, L.). Swedish Research Council, Stockholm.

Smith, J.F. (2001). High species diversity in fleshy-fruited tropical understory plants. *American Naturalist*, **157**: 646-653.

Solbrig, O.T. (1994). The complex structure of the taxonomic system. *In Biodiversity and Terretrial Ecosystems* (eds. Peng, C.I. and Chou, C.H.), pp. 7-24. Academia Sinica Monograph Series No. 14, Taipei.

Soltis, D.E., Soltis, P.S., Chase, M.W., Mort, M.E., Albach, D.C., Zanis, M., Savolainen, V., Hahn, W.H., Boot, S.B., Fay, M.F., Axtell, M., Swensen, S.M., Nixon, K.C. and Farris, J.S. (2000). Angiosperm phylogeny inferred from a combined data set of 18S rDNA, *rbcL*, and *atpB* sequences. *Botanical Journal of the Linnaean Society*, **133**: 381-461.

Sosef, M.S.M. (1994). *Refuge* Begonias. *taxonomy, phylogeny and historical biogeography of begonia sect.* Loasibegonia *and sect.* Scutobegonia *in relation to glacial forest refuges in Africa. (Studies in Begoniaceae V).* Wageningen Agricultural University papers.

Stebbins, G.L. (1957). Self fertilization and population variability in higher plants. *American Naturalist*, **91**: 337-354.

Stephens, J.C., Gilbert, D.A., Yuhki, N. and O'Brien, S.J. (1992). Estimation of heterozygosity for single-probe multilocus DNA fingerprints. *Molecular Biology and Evolution*, **9**: 729-743.

Strand, M., Prolla, T.A., Liskey, R.M. and Petes, T.D. (1993). Destabilisation of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature*, **365**: 274-276.

Swensen, S.M., Luthi, J.N. and Reiseberg, L.H. (1998). Datiscaceae revisited: Monophyly and the sequence of breeding system evolution. *Systematic Botany*, **23**: 157-169.

Sytsma, K.J. and Schaal, B.A. (1985). Genetic variation, differentiation, and evolution in a species complex of tropical shrubs based on isozymic data. *Evolution*, **39**: 582-593.

Takebayashi, N. and Morrell, P.L. (2001). Is self fertilization an evolutionary dead-end? Revisiting an old hypothesis with genetic theories and a macroevolutionary approach. *American Journal of Botany*, **88**: 1143-1150.

Takezaki, N. and Nei, M. (1996). Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics*, **144**: 389-399.

Tebbit, M.C. (1994). A systematic investigation of *Begonia* section *Sphenanthera* (Hassk.) Benth. & Hook. f. Unpublished PhD thesis, University of Glasgow.

Templeton, A.R. (1989), The meaning of species and speciation: a genetic perspective. In *Speciation and its Consequences* (eds. D. Otte and J.A. Endler), pp. 3-27. Sinauer, Sunderland, MA.

Templeton, A.R. (1998). The role of molecular genetics in speciation studies. In *Molecular Approaches to Ecology and Evolution* (eds. R. Desalle and B. Schierwater), pp. 131-156. Birkhauser, Switzerland.

Tiffney, B.H. and Mazer, S.J. (1995). Angiosperm growth habit, dispersal and diversification reconsidered. *Evolutionary Ecology*, **9**: 93-117.

Tremblay, R.L. and Ackerman, J.D. (2001). Gene flow and effective population size in *Lepanthes* (Orchidaceae): a case for genetic drift. *Biological Journal of the Linnean Society*, **72**: 47-62.

Turelli, M., Barton, N.H. and Coyne, J. A. (2001). Theory and speciation. *Trends in Ecology and Evolution*, **16**: 330-343.

Veute, X. and Fuchs, R.P.P. (1993). Greater susceptibility to mutations in lagging strand of DNA replication in *Escherichia coli* than in leading strand. *Science*, **261**: 598-600.

Vogel, S. (1978) Evolutionary shifts from reward to deception in pollen flowers In *The Pollination of Flowers by Insects* (ed. A. J. Richards). Academic Press, New York.

Vogel, S. (1998). Remarkable nectaries: structure, ecology, organophyletic perspectives IV. Miscellaneous cases. *Flora* **193**: 225-248

Wagstaff, S.J. and Dawson, M.I. (2000). Classification, origin, and patterns of diversification of *Corynocarpus* (Corynocarpaceae) inferred from DNA sequences. *Sysematic Botany*, **25**: 134-149.

Walters, S.M. (1981). The shaping of angiosperm taxonomy. *New Phytologist*, **60**: 74-84.

Walters, S.M. (1986). The name of the rose: A review of ideas on the European bias in angiosperm classification. *New Phytologist*, **104**: 527-546.

Weber, J.L. and Wong, C. (1993). Mutation of human short tandem repeats. *Human Molecular Genetics*, **2**: 1123-1128.

Weir, B.S. and Cockerham, C.C. (1984). Estimating *F*-statistics for the analysis of population structure. *Evolution*, **38**: 1358-1370.

White, M.J.D. (1978). *Modes of Speciation*. W.H. Freeman and Co., San Fransisco.

White, F. (1983). The vegetation of Africa; a descriptive memoir to accompany the UNESCO/AETFAT/UNSA vegetation map of Africa. Natural Resources Research No. 20. UNESCO, Paris.

Whitlock, M.C. and McCauley, D.E. (1999). Indirect measures of gene flow and migration: $F_{ST} \neq 1/(4Nm+1)$. *Heredity*, **82**: 117-125.

Wierdl, M., Dominska, M. and Petes, T.D. (1997). Microsatellite instability in yeast: dependance on the length of the microsatellite. *Genetics*, **146**: 769-779.

Wilde, J.J.F.E. de, (2002). *Begonia* section *Tetraphila* A.DC., a taxonomic revision. Studies in Begoniacaeae VII. Wageningen Agricultural University papers

Williams, C.F. and Guries, R.P. (1994). Genetic consequences of seed dispersal in three sympatric forest herbs. I. Hierarchical population-genetic structure. *Evolution*, **48**: 791-805.

Williamson, P.G. (1987). Selection or constraint?: a proposal on the mechanism for stasis. In *Rates of Evolution* (eds. K.S.W. Campbell and M.F. Day), pp. 212-225. Allen and Unwin, London.

Willis, (1922). *Age and Area*. Cambridge University Press, Cambridge.

Wright, S. (1932). The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proceedings of the 6th International congress of Genetics*, **1**: 356-66.

Wright, S. (1951). The genetical structure of populations *Annals of Eugenics*, **15**: 323-354.