The Population Genetic Consequences of Gene Flow during Colonisation and Regeneration of Forest Trees

Samantha Jane Davies

Submitted for the degree of Doctor of Philosophy

Heriot-Watt University

School of Mathematical and Computer Sciences

June 2006

ABSTRACT

This thesis investigates the population genetics of colonisation in the neotropical pioneer tree *Vochysia ferruginea*. A combined approach of empirical molecular analysis and simulation modelling was used to understand how gene flow during colonisation impacts genetic diversity. Samples of adult, seedling and seed from two sites in Costa Rica were analysed to estimate mating system parameters, genetic diversity and gene flow. Spatial modelling was used to investigate the effects of seed and pollen dispersal on the genetic dynamics of recolonisation and to assess potential forest management strategies.

Genetic diversity was higher in primary than in secondary forest indicating that colonisation caused a loss of alleles. However, higher diversity was found in secondary forests trees and seeds than in seedlings suggesting that, with time, high levels of gene flow may counter initial diversity loss due to founder effects. Genetic diversity and spatial genetic structure of individuals contributing to founding populations had significant and population-specific effects. High population densities appeared to decrease gene flow from pollen, impacting fine-scale spatial genetic structure and levels of biparental inbreeding.

Modelling gene flow during colonisation illustrated the importance of long distance seed dispersal and demonstrated how seed from trees in continuous forest with high pollen flow can provide genetically diverse material for regenerating abandoned land.

ACKNOWLEDGMENTS

I would like to sincerely thank my supervisors, Professor Andrew Lowe, Dr. Andrew White and Dr. Stephen Cavers. I feel very lucky to have had such knowledgeable supervisors (and so many of them!). All provided essential advice, guidance and support throughout this PhD for which I am very grateful.

This PhD was funded by European Commission DG RESEARCH INCO-DEV, International role. Contract number ICA4-CT-2001-10094. It was part of a collaborative project with the following organisations; INRA (Bordeaux & French Guiana), IPBO (Belgium), CATIE (Costa Rica), INPA (Brazil), UFRJ (Brazil) Kerala Forest Research Institute (India), Institut Pertanian Bogor (Indonesia), BIOTEC (Thailand), IPB (Belgium), KVL (Denmark) and UQ (Austrailia). I would like to acknowledge those people who I was privileged to meet throughout the course of this project. I would particularly like to thank the staff at CATIE for their assistance with fieldwork. Thanks go to Bryan Finegan for sharing his knowledge of *Vochysia ferruginea* in Costa Rica and his assistance in developing the themes for this study. Thanks and appreciation to Gustavo Hernandez for his help generating molecular data and mapping plots. Thanks too, to Gustavo Hernandez, Vicente Herra and Leonel Coto for providing good company in the field and faultless identification of *Vochysia* and the way back out of the forest.

I would like to thank the staff in the Biosystem Management Research Team, Julia Wilson, Bob Munroe, Jan Dick, Kevin Ingleby, Katy Walker, Nicola Hall and Cecile Bacles for making my stay at CEH an enjoyable one.

Finally, I would like to thank my friends and family with very special thanks to my Dad, Mum and Brand for their support.

(Research Thesis Submission Form should be placed here)

TABLE OF CONTENTS

LISTS OF TABLES AND FIGURES

LIST OF PUBLICATIONS BY THE CANDIDATE

Davies S., White A. and Lowe, A. 2004. *An investigation into effects of long-distance seed dispersal on organelle population genetic structure and* colonisation *rate: a model analysis*. Heredity, **93**, 566-576

Lowe A., Unsworth C.P, Gerber S., **Davies S**., White A., Munro R., Kelleher C., King A., Brewer S. and Cottrell J. *The route, speed and mode of oak postglacial colonisation across the British Isles; Integrating molecular ecology, palaeoecology and modelling approaches*. The Botanical Journal of Scotland, Scheduled for Aug/Sept 2006 (In Press)

Chapter 1 – Introduction

Landscape changes make the process of recolonisation a significant factor affecting the populations of plant species. Deforestation and timber extraction in recent years have meant that many tropical forest habitats, particularly those in Central America, have been substantially disturbed (Kobayashi 1998; Ghazoul and McLeish 2001; Putz *et al.* 2001; Pearce 2001). In addition, in Central America, there has been a recent trend of rural-urban migration leading to increased abandonment of agricultural and marginal grazing land (Aide and Grau 2004). Since many of these abandoned regions become colonised by pioneer tree species, the extent of secondary forest in Latin American countries has seen a concurrent increase. In fact, secondary forest now covers more area than primary forest in many tropical countries (Sezen *et al.* 2005), and is rapidly becoming an important resource for, amongst other things, habitat conservation, improved water quality and as an economic resource in itself (Finegan 1992; Herrera and Finegan 1997; Guariguata *et al.* 1997; Baluarte-Vasquez *et al.* 2000; Sanchez and Tournon 2000; Chokkalingham and Jong 2001; Aide and Grau 2004; Piotto *et al.* 2004).

Conversion of degraded land into secondary forest is a relatively inexpensive method of providing forest products and environmental amelioration that can be maintained by local people (Chokkalingham *et al.* 2001; Piotto *et al.* 2004). Therefore, secondary forest regeneration and management can increasingly be seen as an important focus for international and national commitments to biodiversity conservation, community-based management and forest rehabilitation (Kobayashi 1998; Guariguata *et al.* 1997; Chokkalingham *et al.* 2001). For the long term stability of tropical forests, forestry management is increasingly considering the conservation of genetic diversity as a component of biodiversity conservation.

Research into genetic diversity partitioning and gene flow dynamics has perhaps been most prolific in tropical systems, partially due to the biological interest of such systems, but also due to their increasingly threatened status (Lowe *et al.* 2005). However, it is still not possible to confidently make generalisations about genetic responses to disturbance, habitat change or different management strategies (Lowe *et al.* 2005). For example, variable pollinator foraging behaviour and tolerances to inbreeding mean that species-specific (and sometimes population specific) studies of gene flow are necessary to understand genetic diversity and structure in natural or managed populations (e.g. Murawaski *et al.* 1994; Stacey *et al.* 1996; Aldrich 1998; Dayanandan 1999; Hamilton 1999; Collevatti *et al.* 2001; Dick *et al.* 2003; Ward *et al.* 2005).

As well as molecular investigation, the development of simulation models of gene flow enables extrapolative analysis of a species' genetic response by allowing manipulation of any number of biological and population parameters. Using detailed molecular analysis to parameterise a simulation model offers an opportunity to provide a valuable decision-support tool to aid long-term sustainable management of biologically or commercially important species.

This study investigates the population genetics of colonisation into disturbed habitat by a forest gap pioneer: the neotropical tree *Vochysia ferruginea* Mart.. The species is a fast growing pioneer able to tolerate low fertility soils, which makes it a potentially important species for forest regeneration and timber production on disturbed land. In this study an analysis of genetic diversity, genetic differentiation, mating system (including levels of biparental inbreeding) and gene flow in populations of *V. ferruginea* in both dense secondary forest and in widely distributed primary forest gaps is conducted. The molecular data will be used to inform development and parameterisation of a simulation model, allowing *in silico* exploration of the most important genetic

factors governing diversity maintenance and recolonisation under different management scenarios. The model outputs should help to guide commercially viable and sustainable forestry for this species.

The following serves as a review of the literature that has examined the status of secondary forests, the effects of seed and pollen flow on forest tree populations and how factors relevant to tropical habitats, such as fragmentation, has affected genetic structure.

1.1 Neotropical secondary forest

Generally speaking, the term primary forest relates to the climax forest type for a given region and the term secondary forest relates to successional forests that develop after clearing of the original forest (Chokkalingham *et al.* 2001). Secondary forests are naturally regenerated forests after significant human or natural disturbance, such as (in Central America) hurricanes, and display major differences in forest structure and species composition with respect to nearby primary forest (Chokkalingham *et al.* 2001). Secondary forest is characterised as being dominated by fast growing pioneers and is usually less species diverse than primary forest. Species composition of neotropical secondary forest is highly influenced by the timing of abandonment, land use history and physical factors, e.g. elevation, substrate (Chinea and Helmer 2003). Pioneer species tend to be gap colonists in primary forest, where they will usually grow in canopy breaks created by disturbances, such as tree fall. Pioneer species are then, well adapted to regeneration following disturbance. However, woody regrowth can be rapid after abandonment (Guariguata *et al.* 1997) and mature forest tree species progressively replace initial pioneer species so that seedling populations in secondary forest can be more similar to primary forest than to secondary forest (Guariguata *et al.* 1997; Fuhr *et al.* 2001).

Forest succession on degraded lands has been proven to create more favourable microclimatic conditions that can stimulate surface soil accumulation of organic carbon and attract seed dispersing wildlife (Parotta *et al.* 1997). Farmers in the neotropics utilise secondary forest, primarily for silviculture (Baluarte-Vasquez *et al.* 2000), and increasingly leave a portion of their land as rehabilitated secondary forest as a consequence of government incentives (Piotto *et al.* 2004). In Costa Rica, reforestation has been promoted by the government and incentive programs have encouraged farmers to dedicate some of their land to reforestation by native pioneer species (Piotto *et al.* 2004). In 1996, Costa Rica developed a forestry law which recognises the role of forests in carbon fixation, watershed protection, biodiversity protection and scenic beauty with landowners being paid to provide these environmental services (Pearce 2001). Thus, there is an increasing conservation requirement, that the impact of disturbance on the genetic resources of exploited and impacted tropical species, together with their regeneration and recruitment dynamics in secondary forest contexts receive better understanding. Therefore, analysis of contemporary gene flow dynamics, the key process underlying genetic diversity partitioning, can help advise conservation and forestry management practices, and may then improve forest productivity, for example by identifying fastest routes to forest establishment.

1.2 Gene flow

Gene flow in plants is mediated by pollen and seed dispersal and is influenced by a range of biological and environmental factors, such as breeding system, disperser behaviour, phenology and landscape characteristics. How genetic diversity is partitioned within and between populations of a plant species is largely a consequence of gene flow; hence those factors that influence gene flow will also influence population genetic diversity and differentiation. Using recently developed molecular marker techniques

(e.g. microsatellites) it is possible to directly estimate population-scale gene dispersal dynamics (e.g. using parentage or assignment analyses). Such molecular and statistical methodologies have provided significant insight into how changes in gene flow dynamics affect genetic structure (e.g. Hamrick *et al.* 1993; Boshier *et al.* 1995; Aldrich *et al.* 1998).

1.2.1 Definition of gene flow

Gene dispersal refers to the movement of genes from an individual via reproductive propagules. Gene flow is defined as the proportion of newly immigrant genes moving into a given population (Endler 1977). Gene flow is therefore the effective movement of genes into a population through successful establishment following gene dispersal.

1.2.2 Gene flow in plant populations

Along with mutation, drift and selection, gene flow is one of the most important factors determining the genetic structure of populations (Wright 1931; Ellstrand *et al.* 1989; Ellstrand 1992; Bossart and Powell 1998; Burczyk *et al.* 2004). Where gene flow is restricted, populations can become differentiated from one another (e.g. Dutech *et al.* 2002), whereas extensive gene dispersal within local populations promotes panmixia and reduces the family structuring formed by natural regeneration (e.g. Hardy *et al.* 2004). Estimation of the gene flow of a species is essential for assessing fitness in natural populations, the effective population size of a species, a species' potential for regeneration and its response to natural and human disturbance (Ellstrand 1992; Dawson *et al.* 1997; White *et al.* 1999).

Sexually reproducing plants disperse genes in two ways. Male gametes are dispersed twice; from the pollen parent to the maternal parent via pollen and from the maternal parent as part of the genetic complement of the embryo. Maternal gametes, in contrast, move only once via seeds. Therefore, (and due to very different mobilities) there is a

potential for greater variance in the dispersal distance of pollen relative to seeds. This also means that knowledge of a species' pollination system can offer an insight into the distribution of genetic diversity within and among plant populations (Hamrick *et al.* 1993).

Plants have a nuclear genome, a mitochondrial genome and a chloroplast genome with different inheritance patterns. In angiosperms, chloroplast and mitochondrial genomes are normally maternally inherited whereas, in gymnosperms, chloroplast genomes are usually paternally inherited (Harris and Ingram 1991). The different modes of inheritance mean that over the course of one generation nuclear genes (and paternally inherited chloroplast genomes in gymnosperms) are dispersed by both pollen and seed (Ennos 2001) and maternally inherited chloroplast genes are only dispersed by seed.

Total gene dispersal is: $\sigma^2 = (\sigma_p^2/2) + \sigma_s^2$

where σ_p^2 and σ_s^2 are the variances in pollen and seed movement, respectively and the factor of ½ is required as the contribution of maternal gametes that do not disperse is omitted. Seed dispersal variance therefore contributes twice as much as pollen dispersal variance to the size of genetic neighbourhoods (the breeding area of the population) (Hamilton 1999). These principles can be used to assess relative contributions of pollen and seed dispersal to overall gene flow from genetic structure.

Gene flow can be beneficial or deleterious to the fitness of a population. High gene flow reduces the potential for inbreeding but may reduce fitness if individuals adapted to one habitat disperse genes to a different habitat resulting in maladapted offspring (outbreeding depression). Outbreeding depression can be considered as having two main causes, local adaptation and also intrinsic coadaptation as gene flow into

populations adapted to local conditions increases the proportion of less well adapted offspring (Templeton 1986, Nathan and Muller-Landau 2000).

1.2.3 Factors affecting gene flow in plant populations

The pattern and magnitude of gene transfer can influence the genetic structure of populations, their effective size and the extent to which they may become genetically differentiated (Hamrick *et al.* 1992; Hardy *et al.* 2004).

There are a range of biological and non-biological factors that influence the magnitude and scale of gene flow. The biological characteristics of an organism, particularly vagility of individuals and their propagules, and reproductive system (asexual or sexual, selfing or outcrossing), are important determinants of the magnitude of gene flow both between and within populations (Lowe *et al.* 2004). Other biological factors such as pollination biology and mechanisms of seed dispersal will also have significant impacts on the dynamics of gene flow. Non-biological factors such as geographical features also affect gene flow and can either restrict gene dispersal, e.g. mountain ranges that can provide a barrier to pollinators and/or seed dispersal (e.g. Cavers *et al.* 2005) or aid gene dispersal between populations e.g. rivers that can potentially increase the distance of seed dispersal (e.g. Imbert and Lefèvre 2003).

1.2.4 The effect of gene flow on genetic structuring of populations

A population is an interbreeding group that experiences reduced gene flow with other groups of the same species (Hedrick 2000). Genetic structure develops as populations become differentiated from each other through the process of genetic drift (random fluctuations in allele frequency occurring within populations that ultimately lead to the fixation of alleles), mutation and natural selection. Differentiation among populations may reflect historical impediments to movement and thus to relatively ancient population subdivisions. Differences among populations can also reflect contemporary

patterns of gene flow, provide insights into how natural populations maintain genetic variation, and indicate the impact of anthropogenic fragmentation events on the movement of individuals.

Differences among populations are commonly quantified by the use of one of several statistics, including Wright's statistic, F_{ST} (the proportion of the diversity in the sample that's due to allele frequency differences among populations) and Nei's coefficient of gene variation between populations, G_{ST}. These indices are functions of how heterozygosity is partitioned within and among populations, based on differences in allele frequencies (Wright 1969; Nei 1975; Chai 1976; Wright 1978). These statistics indicate whether the majority of genetic variation is distributed among or within populations. In species with low F_{ST}, G_{ST} (approaching 0), the majority of variation is found within populations; individuals within populations are as likely to be similar to individuals within that population as to individuals in another population. Where F_{ST}, G_{ST} is high (approaching 1), individuals within a population are more likely to be similar to individuals within the population than to individuals from another population.

These statistics are based on the simplest model of gene flow, the island model (Wright 1969). In this model there are a number of subpopulations, each with their own allele frequencies. Averaging across all subpopulations gives the allele frequencies for the whole metapopulation. If a proportion of randomly chosen individuals leave their subpopulation to form a migrant pool, this migrant pool will have the same allele frequencies as the whole metapopulation. If the migrants then randomly immigrate into the subpopulations the process of migration will move the subpopulation allele frequencies towards the whole metapopulation allele frequencies. Therefore, migration reduces population differentiation between subpopulations.

However, the island model, and quantifications of differentiation based on it, makes a number of assumptions that may not be met in real populations (Whitlock and McCauley 1999; Wakeley 2005). It assumes all populations are of equal size with equal contributions to the migrant pool. Real populations vary in size, their separation and in their contribution to the migrant pool over both space and time. The model also assumes there is no spatial structure. In real populations there is usually some structuring, primarily due to distance affected dispersal. The island model assumes that natural populations are at equilibrium, whereas real populations can go extinct and be recolonised and population structure may be a result of historical factors. The model also assumes no selection and no mutation.

In real populations, if gene dispersal between subpopulations is restricted, genetic structure will develop by genetic drift in the absence of selection and the relatedness of individuals increases with their spatial proximity. This is isolation-by-distance (Wright 1943). Analysis of spatial genetic structure of outcrossing species has shown that population differentiation generated by isolation-by-distance is lower than expected and when present appears to be weak, limited to small scales, or only detected in maternally inherited markers that do not disperse in pollen (Ellstrand *et al.* 1989; Ennos 2001). Seed and pollen dispersal patterns can differently shape the genetic composition and structure of plant populations. As seed dispersal is usually more limited than pollen dispersal, maternally inherited genes can be more structured and species with limited seed dispersal will have a greater spatial genetic structuring than those with more extensive dispersal (Schnabel *et al.* 1998; Miyazaki and Isagi 2000). Therefore, isolation-by-distance structuring is predominantly through seed dispersal with more expansive pollen dispersal reducing the degree of structure within the population (Hamrick and Nason 1996; Ennos 2001).

In general, the more restricted the gene flow, the more populations are genetically differentiated and the more closely related are the nearest neighbours, with the magnitude and scale of gene flow predominantly determining the scale of patches of related genotypes within a population (Lim *et al.* 2002). Therefore, high genetic differentiation between populations is usually a result of restricted gene flow. Low genetic differentiation may be the result of high gene flow but may also result if populations are from a previously large population that has been recently subdivided (Lim *et al.* 2002).

Genetic structure within populations also develops as dispersal interacts with population structure, such as adult density, or forest structure, such as gaps (Rousset 2004; Lim *et al.* 2002), microhabitat selection and recruitment ecology (Hamrick *et al.* 1993). Therefore, the dispersal dynamics in a population can involve a number of factors and it may not always be possible to interpret them from genetic structure alone. For example, gene flow via pollen in animal pollinated species reflects the foraging behaviour of the pollinator and it may be difficult to distinguish restricted pollinator foraging from the clustering effect of related individuals following recolonisation of a patch, due to founder effects that differentiate patches from each other (Dutech *et al.* 2002).

Studies of pollen flow in natural populations of tropical trees (Collevatti *et al.* 2001, Dick *et al.* 2003, White *et al.* 2002) have shown that far from being restricted to movement between nearest neighbours, pollination distances for both wind and animal dispersed species can be extensive and realised pollen dispersal distances in animal pollinated species are significantly larger than those inferred directly from pollen tracking or pollinator observation (Ennos 2001). They have also shown a high incidence of interpopulation pollination events.

1.2.5 Gene flow and conservation genetics

Genetic diversity within a population may be lost via four related processes: founder effects, demographic population bottlenecks, genetic drift, and inbreeding. Loss of genetic diversity is believed to have implications for population persistence over various temporal scales. In the short term, inbreeding resulting from small population size may cause inbreeding depression with attendant reductions in mean population fitness. In the long term, reduced genetic diversity within a population or species may lower the evolutionary potential or decrease the probability of future speciation events. It is important to distinguish among measures of genetic diversity. For example, inbreeding within a population will reduce observed heterozygosity but does not alter overall allele frequencies. Therefore, inbreeding reduces one measure of genetic variability (observed heterozygosity) while the other (allelic diversity) remaining similar between populations/species. Evidence for inbreeding within populations can be assessed by comparing observed and expected heterozygosities across loci.

The study of gene flow through estimation of mating distance and number of mating partners is fundamental for understanding the reproductive processes in outcrossing plants and evaluation of the genetic resource within plant populations. In general, restricted gene flow resulting in selfing or mating with relatives, can cause serious inbreeding depression in out-crossing species (Wang *et al.* 1999), therefore, restricted gene flow can become a threat to the viability of populations of out-crossing plants.

Two hypotheses have been proposed for the phenomenon of inbreeding depression (Keller and Waller 2002). The heterozygote advantage hypothesis suggests that individuals with a greater number of heterozygous loci will have a fitness advantage over those that are more homozygous because of the accumulation of rare, deleterious recessive or partially recessive alleles that are masked in the heterozygous condition

(Lande and Schemske 1985; Charlesworth and Charlesworth 1987 1990). The recessive mutation hypothesis proposes that inbreeding increases the probability of individuals having two deleterious recessive alleles at any particular locus. The former implies that individual and population-level fitness will continue to decline until all within-individual diversity is lost, whereas the latter implies that populations may be purged of deleterious alleles as they are exposed to selection (Hedrick 2000). Most models predict that inbreeding depression should decrease with increased inbreeding, as deleterious mutations of large effect are purged from the genome (Lande and Schemske 1985; Charlesworth and Charlesworth 1987 1990; Charlesworth *et al.* 1990b; Barrett and Charlesworth 1991; Husband and Schemske 1995 1996; Crnokrak and Barrett 2002). However, if recessive mutations are only mildly deleterious, or have a low dominance coefficient, selection will be ineffective relative to mutation, and substantial inbreeding depression may be maintained, even in highly inbred populations (Charlesworth *et al.* 1990a 1990b). Therefore, natural populations can have very low levels of within individual and within population genetic diversity with little apparent effects on fecundity or survivorship (Lynch and Walsh 1997; Keller and Waller 2002).

Restricted mating such as sib-mating or selfing, which can be caused by decreased density of flowering trees and restriction of inter-tree movement of pollen vectors, has been found in tropical trees (Konuma *et al.* 2000). For example, in Thai populations of *Shorea siamensis,* a decrease in population density due to selective logging caused an increase in selfing rate, because pollen flow, which is mediated by a small bee, became restricted (Ghazoul *et al.* 1998). Unfortunately it is not possible to make generalisations about plant genetic response to disturbances, such as fragmentation, as pollen and seed dispersal is dependent on specific vectors, and so gene flow can either increase or decrease following disturbance (e.g. Collevatti *et al.* 2001; White *et al.* 2002) . For these reasons species and landscape specific estimates of gene flow are important for

understanding conservation and management issues for commercially important or threatened species (Avise 1999; Konuma *et al.* 2000).

1.3 Molecular marker analyses of contemporary gene flow

1.3.1 Estimates of gene flow using molecular markers

Genetic analysis of natural populations allows investigation of biological questions that previously could only be answered through extensive observational study (Avise 1999; Davies *et al* 1999). It is possible to study effective gene flow through the application of selectively neutral genetic markers. Gene flow, genetic drift, selection, mutational divergence and genetic recombination influence the distribution of alleles in space. Estimating gene flow with genetic markers essentially involves assessing this distribution (Ouburg *et al.* 1999; Le Corre and Kremer 2003). Most markers are located in non-coding DNA regions and therefore it is assumed that they experience very little or no selective pressure, i.e. they are selectively neutral. Under this assumption the potentially significant influence of selection may be disregarded (Ouburg *et al.* 1999; Goldstein *et al.* 1999; Le Corre and Kremer 2003). Different types of genetic marker differ in their mutation rate, the degree of selective neutrality and whether or not they are subject to recombination, although genetic drift and gene flow should affect all markers in a similar way (Ouburg *et al.* 1999).

In plants, molecular markers can be situated in any of the three genomes: nuclear, chloroplast or mitochondrion. A number of genetic markers have proven to be useful, of these allozyme loci, and Variable Number of Tandem Repeat (VNTR) markers are the most effective for studying gene flow. VNTR markers are characterised by a core sequence that consists of a number of identical repeated sequences. They can be divided into two categories based on the repeat length. These are minisatellites, 15 -70 base pairs (bp), and microsatellites, 2-6 bp.

Gene flow in plants has been studied extensively using allozyme markers (Hamrick 1982: Hamrick and Godt 1990). Allozyme markers allow an investigation of protein variation as an estimate of the DNA sequence variation that determines the amino acids of these proteins (Hedrick 2000). The use of allozymes has provided information on pollen and seed mediated gene flow in a number of species (reviewed in Hamrick and Godt 1990 and Gray 1996). However, because of the limited variability revealed by allozyme markers, they are not always suitable for measuring contempory gene flow (Barrett *et al.* 1993; Bossart and Powell 1999). There is also some doubt as to their selective neutrality, at least in some species (e.g. Kreitman and Akashi 1995). Thus, study of gene flow has tended to use microsatellite markers (Jarne and Lagoda 1996; Luikart and England 1999).

Nuclear markers other than microsatellites have a low mutation rate (typically 10^{-5} mutations per generation) and so may have polymorphic loci with only two or three alleles generally present within populations (Ennos 2001). In contrast, the mutation rate for microsatellites may be orders of magnitude higher than allozymes, reaching 10^{-2} per generation for some plant species (Ennos 2001). As a consequence microsatellite markers can show very high degrees of variation, with in some cases, over 20 alleles commonly being maintained per locus within a single population. This makes them particularly suited for investigations of gene flow at fine spatial scales (Luikart and England 1999). The combination of a number of highly polymorphic loci also allows the generation of unique genotypes for individuals within populations, allowing paternity and parentage analysis to determine pollen and seed dispersal distances involved in individual mating and seedling establishment events (Queller *et al.* 1993; Luikart and England 1999; Ennos 2001).

1.3.2 Microsatellite loci

Microsatellite loci are common and relatively evenly spaced throughout the genome (Edwards *et al.* 1992). Most loci are selectively neutral which makes them compatible with the assumptions of population genetic theory (Jarne and Lagoda 1996). Microsatellites are more desirable than the larger VNTR loci because they can be analysed via the Polymerase Chain Reaction (PCR). Finally, microsatellites have been found to be polymorphic even in populations that have low levels of allozyme variation (Estoup *et al.* 1995).

1.3.3 The evolution and mutational processes of microsatellites

The rate of mutation has been estimated for microsatellites but the mechanisms by which mutations occur are not fully understood (Brohede and Ellegren 1999). Two main mechanisms have been proposed 1) unequal crossing over in meiosis and 2) strand-slippage replication (Levinson and Gutman 1987; Ellegren 2004). Of these, strand-slippage replication appears to be the predominant mode of mutation for most microsatellite loci (Wolff *et al.* 1989).

Strand slippage

Strand-slippage is speculated to occur primarily during lagging strand synthesis (Schlotterer and Tautz 1992; Ellegren 2004). For example, it may involve the slippage of the newly synthesised DNA strand upon dissociation of a polymerase complex. This slippage creates a transient bulge that upon DNA repair would be either removed or lead to the elongation of the repeat (Schlotterer and Tautz 1992). Alternatively, the formation of a transient bulge in the template strand may lead to a shortening of the repeat. This is thought to explain small changes in numbers of repeats (adding or subtracting of one or just a few repeats). It also explains how microsatellite loci could be generated in the first place; it is likely that sequences including two or three repeats are randomly

distributed throughout the genome; slippage could then amplify these short repeat sequences into many repeats over successive generations.

Unequal crossing over

Unequal crossing-over during meiosis is thought to explain large changes in numbers of repeats. During crossing over one chromosome obtains too many repeats and the other chromosome obtains too few repeats (Ellegren 2004).

Variability in mutation rate

Whatever the mutational process, there does appear to be some bias in the mutation rate (Whittaker *et al.* 2003). An *in vitro* study has found evidence that repeat length and base composition affect the mutation rate, i.e. dinucleotide repeats mutate faster than trinucleotides, and sequences with a high AT content mutate faster than those with a high GC content (Schlotterer and Tautz 1992; Ellegren 2000; Whittaker *et al.* 2003). This indicates that template stability may affect the mutation rate, perhaps by reducing the frequency of strand-slippage events.

There is strong evidence that mutation rate of microsatellites is length dependent (Ellegren 2000; Harr and Schlotterer 2000; Whittaker *et al.* 2003). Schlotterer and Tautz (1992) found that in the majority of loci they examined, the more complex the repeat structure the lower the likelihood of single step mutational events. This suggests that as the variance in repeat number increases the frequency and/or magnitude (i.e. number of change of repeats) of multi-step mutations increases (Schlotterer and Tautz 1992; Di Rienzo *et al.* 1994: Ellegren 2000; Whittaker *et al.* 2003).

The evolution of microsatellites is a complex mutational process; therefore data from the different classes of microsatellites should not be treated as a single homogeneous

data set, as different mutation events are interpreted somewhat differently. It has been suggested that microsatellites that have more composite repeats are those best suited to study questions such as population subdivision and genetic relationships (Estoup *et al.* 1995) since they will contain the lowest levels of homoplasy (when two alleles are identical in state but not identical by descent).

Models of microsatellite evolution

Most of the current estimators of microsatellite mutation were developed based on interpretation of allozyme variation. In allozymes, it was assumed that most new mutations gave rise to new distinguishable alleles. In addition, the electrophoretic mobility, upon which alleles were assigned, could not be used to assess the mutational relationship amongst the alleles. An Infinite Allele Model (IAM) where every new mutation is assumed to give rise to a new electrophoretically distinguishable allele (or electromorph) models this process (Kimura and Crow 1964). Whilst some homoplasy (i.e. an electromorph which comprises two or more alleles) exists in allozyme markers, the IAM has proven very useful in modelling observed allozyme variation (Nei 1987).

The situation appears to be very different for microsatellite loci. First, since most of the mutations seem to involve the gain or loss of a single repeat unit (Ohta and Kimura 1973; Weber and Wong 1993; Ellegren 2004), it is clear that a relatively high level of homoplasy exists. This is important as homoplasy leads to the underestimation of total variation and genetic distance, and to an overestimation of the similarities among populations. The Stepwise Mutation Model (SMM) has been used to simulate this situation, in which alleles can only mutate by the gain and loss of one repeat unit (Ohta and Kimura 1973). The SMM is generally the preferred model when calculating relatedness between individuals and population sub-structuring, since homoplasy has

little effect on populations over a short period of time (hundreds of generations; Goodman 1998).

1.3.4 Assumptions of microsatellite data

Prior to population genetic analysis, it is essential to pre-test individual locus variation and allelic behaviour to ensure that the basic theoretical assumptions of subsequent analysis are not violated (Beaumont and Nichols 1996). Three main assumptions need to be tested. First, the selective neutrality of each locus should be analysed. Second, the presence of 'null alleles' (alleles that are present but do not produce a product during PCR assay) should be identified. Finally, before the data from various loci are combined, the independent assortment of the loci must be tested.

Neutrality

The assumption of the selective neutrality of microsatellite loci is the key principle behind most population genetic analyses (Ford 2002). Due to this assumption, analyses of data can be based only on the interaction of genetic drift (random change of allele frequency), mutation, and/or migration. Over time, the effects of drift and mutation will lead to the divergence of allele frequencies among subpopulations, while migration will lead to a homogenisation of allele frequencies. If loci are not neutral, strong selection pressures may override these forces (Schlotterer 2002). Selection at a locus can stabilise allele frequencies (e.g. via over-dominance) across all subpopulations and lead to an underestimation of population substructure or genetic distance. Alternatively, differences in selective pressures among regions may cause a fixation of alternate alleles in subpopulations and lead to an overestimation of these parameters. Since the effects of selection can confound population genetic analyses, any loci that are under selective influence should be excluded from analysis. However, the vast majority of microsatellites are believed to be neutral and the comparison of observed genotype

frequencies with those expected under Hardy-Weinberg equilibrium should allow for locus-specific detection of selection (Lewontin and Krackauer 1973; Edwards *et al.* 1992; Brohede and Ellegren 1999; Schlotterer 2002).

Null alleles

Null alleles are any allele at a microsatellite locus that consistently fails to amplify to detectable levels via the polymerase chain reaction (PCR) (Callen *et al.* 1993). They can occur because of poor primer annealing due to nucleotide sequence divergence, low quality DNA template (Dakin and Avise 2004), and differential size variant amplification (Wattier *et al.* 1998), where shorter alleles amplify in preference to larger alleles, that are then undetected (also called partial-null alleles because altered reaction conditions can often make bands detectable). The presence of null alleles leads to an overestimation of observed homozygosity and acts in addition to biological factors, such as inbreeding or the Wahlund effect (also responsible for heterozygote deficits). However, multilocus analysis can usually detect null alleles (and those under selection, see above) since population genetic factors are expected to register more or less concordantly across loci, whereas the effects of null alleles and selection are locus specific (Dakin and Avise 2004).

Independent assortment

It is assumed that the transmission of alleles at any locus is independent of alleles at other loci (independent assortment) and that the fitness of genotypes at one locus is also independent of the fitness of genotypes at other loci (Hedrick 2000). However, there is some evidence that genetic variants may not be independent of each other (linkage disequilibrium). Factors such as multilocus selection, genetic drift, gene flow, mutation and genetic hitchhiking (association between alleles at other genes where selection is operating) can all lead to linkage disequilibrium (Hedrick 2000). To describe genetic

variation, allelic frequencies at a number of loci are required; therefore it is necessary to consider the non-random association of alleles at different loci. With a sufficient number of polymorphic loci the extent of linkage disequilibrium can be estimated (Sloane *et al.* 2001).

1.3.5 Using microsatellite data to analyse gene flow

Indirect measures of gene flow

Molecular markers can be used to infer the level of gene flow indirectly from the amount of genetic drift between populations (e.g. Neigel 1997). The spatial distribution of alleles is quantified and a population model is applied, e.g. spatial autocorrelation analysis or Wrights F statistics to estimate the amount of gene flow that would result in a similar distribution (Ouberg *et al.* 1999, Ennos 2001). The processes giving rise to the structure are then inferred. However, inferences about spatial processes are problematic because they rely on assumptions of drift/migration equilibrium or on the ecological history of the population (Ennos 2001).

Wrights F statistics (Wright 1965) are used to detect and quantify genetic structure within plant populations. The level of genetic differentiation is quantified as F_{ST} - the standardised among-population variance in allele frequency - and is defined as;

$$F_{ST} = \frac{1}{1 + 4N_e m}$$

Where N_e is the effective population size, m is the proportion of individuals that are immigrants, and N_em is the number of migrants. This equation states that, in the absence of selection, and at equilibrium between drift and gene flow, the genetic differentiation between populations is inversely related to the rate of gene flow among them (Wright 1978, Ouberg *et al.* 1999).

F statistics measure structure in terms of inbreeding due to population subdivision. F_{ST} is most straightforwardly applied where subpopulations with potentially different origins can be readily identified. The problem with this analysis is that as subpopulation size decreases, sample size decreases and statistical power to detect population genetic structure also decreases (Bossart and Powell 1999; Ennos 2001).

Indirect measures of gene flow are based on a number of assumptions, e.g. that equilibrium between genetic drift and gene flow has been reached. However such assumptions are often violated in natural populations, specifically those that are in flux, such as species exhibiting metapopulation dynamics (Slatkin 1995; Cockerham and Weir 1993; Bossart and Powell 1999). Although Wright (1943) stated that in populations undergoing isolation-by-distance 'interbreeding is restricted to small distances by the occurrence of only short range means of dispersal', when mutation rates are high, such as in microsatellite loci, mutation may be as significant as gene flow in shaping patterns of genetic differences between populations (Slatkin 1995; Collevatti *et al.* 2001). Therefore, interpretation of genetic differences between populations in terms of limited gene flow may be inappropriate and time since isolation of populations may be a very important factor. Microsatellites are highly variable and subject to high mutation rates so they usually display high levels of within population heterozygosity (Hedrick 1999). Thus statistics that are based on an infinite alleles model and consider alleles to be identical by descent, are likely to underestimate population differentiation (Hedrick 1999).

Direct measures of gene flow

The problems associated with indirect methods of estimating gene flow can be eliminated by directly estimating pollen flow using paternity analysis or seed flow using

maternity analysis (Meagher 1986; Slatkin 1995; Bossart and Powell 1999). Direct methods to obtain pollen flow exploit the fact that seeds can be collected when they are attached to the maternal parent. This allows the genotype of the pollen to be inferred and provides information on the distribution of successful fertilization events (Smouse and Meagher 1994; Gjertson and Morris 1995; Ennos 2001). However, as direct measures do not ascertain which seeds will become successfully established (the proportion of which may be non-randomly determined by fitness differences in ovules) they may differ from the actual level of gene flow via pollen. In other words they measure gene dispersal and not gene flow.

Paternity exclusion analysis

In paternity exclusion analysis the multilocus genotype of a focal maternal plant and a sample of her seeds are determined, together with the multilocus genotypes of all potential fathers (Ennos 2001; Jones and Arden 2003). For each seed, adults that could not have acted as pollen parents are excluded on the basis of their genotype. If the exclusion probability generated by the genetic marker variation is high, a single adult will be left that can be uniquely assigned as the male parent (Chakraborty *et al.* 1988). The more variable the suite of genetic markers used in terms of allelic diversity and evenness of allele frequency, the greater the exclusion probability. The high allelic variation at individual microsatellite loci facilitates the provision of high exclusion probabilities during paternity assignment (Dawson *et al.* 1997). Therefore, the use of microsatellites is effective for quantifying pollen dispersal patterns in natural plant populations (Dawson 1997). With a relatively small number (3 to 6) of highly variable microsatellite markers possessing 20 or more alleles per locus, exclusion probabilities close to 1 can be achieved, and male parents can be identified with confidence (Ennos 2001).

Hall *et al.* (1994) compared paternity exclusion probabilities from allozymes and microsatellites and found that the paternity exclusion probabilities for six allozyme loci ranged from 0.04 to 0.46 and the joint paternity exclusion probability reached 0.73. The paternity exclusion probability using just three microsatellite loci ranged from 0.14 to 0.77 and the overall joint paternity exclusion probability was 0.93.

Using exclusion methods it is possible that all potential parents are excluded, or only one parent can be assigned. In this case the inference is that the missing parent(s) have not been sampled either because they are outside the sampled population or are no longer alive (Ennos 2001).

It is often the case with less variable allozyme loci and some microsatellite loci in angiosperms that multiple potential male parents are identified after paternity exclusion analysis (Ennos 2001). Furthermore, very high exclusion probabilities are not often achievable, either because sufficiently variable markers are not available, or it is not feasible to exhaustively sample the population that may be contributing genes to the offspring. In this instance there are different options for assigning paternity. Complete paternity can be assigned on the basis of the most likely father given the genotype of the mother, maternal offspring and potential fathers (Meagher 1986). Alternatively fractional paternity can be assigned to a number of fathers according to their genetic likelihood of being the true father (Meagher and Thomphson 1987; Devlin *et al.* 1988).

Although feasible in small populations, with a restricted number of potential fathers, paternity exclusion analysis becomes impractical in large continuous populations with a large number of potential pollen donors, which are impossible to sample. In these instances it is more appropriate to use either the neighbourhood approach of Burczyk *et*

al. (1996) or a two-generation (parent-offspring) approach (TwoGener, Austerlitz and Smouse 2001a,b, 2002; Smouse *et al.* 2001).

Using the neighbourhood approach a neighbourhood is delineated around focal maternal parents with pollen flow divided into a within-neighbourhood component and a component representing migration from outside the neighbourhood. A model of male fertility within the neighbourhood can then be fitted to account for the offspring genotype arrays of the focal maternal parents. This model can include pollination distance and direction of pollen flow as factors influencing male success and these parameters can be estimated (Ennos 2001).

The TwoGener strategy for estimating pollen dispersal curves, without exhaustive sampling, is based on the differentiation among inferred pollen pools of a sample of females spread across the landscape (Smouse *et al.* 2001). Comparison of results from a TwoGener analysis and a paternity analysis on a number of species (Austerlitz *et al.* 2004) showed that this approach is capable of approximating the shape of the pollen dispersal curve, although several factors (e.g. a high level of selfing) could bias the dispersal estimate (Austerlitz and Smouse 2001a,b).

1.4 Modelling gene flow

Analysis of gene flow and genetic structure using molecular data and spatial statistical approaches aids understanding of the population genetics of a species but this approach is less amenable to predictive analysis through manipulation of parameters that effect gene flow and therefore genetic structure. Modelling gene flow over a simulated landscape can give further insight into the genetic response of populations to different scenarios. Simulation modelling can therefore be used to investigate possible effects of hypothetical scenarios, seeking to identify thresholds after which there are significant

negative effects on genetic diversity within the simulated populations (Degen *et al.* 1996).

Modelling also allows investigation of the genetic response of a species over much larger time-scales than empirical studies. This is particularly relevant for tree populations where individuals have a potentially very long lifespan. Carbon dating of emergent tropical trees has revealed very ancient individuals persisting in natural populations, including an example of *Cariniana micrantha* which was over 1,400 years old (Chambers *et al.* 1998), although the average maximum ages of tropical emergent trees are unlikely to be much greater than 600 years (Fitchler *et al.* 2003). This means, for example, molecular analysis of tropical tree response to fragmentation is hampered as many of the trees currently alive germinated prior to disturbance.

A spatial simulation model requires a number of inputs such as gene dispersal curves, phenological data (e.g. number of flowering events per year), life history data (e.g. lifespan, age at maturity) and genetic data (e.g. allele frequencies in a population). Once these are determined the model is run for a number of years, this generates an output, which is a simulated population in which each individual has a recorded genotype. This simulated population can then be analysed using the same diversity and differentiation statistics as are used for real populations.

An important consideration when using modelling to inform forestry management decisions is that there are limitations to this approach, as models inevitably use simple relationships to describe complex systems. Before a model can be used for deriving recommendations on the management of a real system, the concordance between the simulation and the real system must be tested. The sensitivity of certain parameters within the model and those that are critical in determining outcomes requires accurate

estimates from and comparison with real populations. As long as there are no critical differences between the simulation and reality, the model can be regarded as valid (Degen *et al.* 1996, Turner *et al.* 2001).

1.4.1 Modelling effects of seed dispersal on organelle genetic structure

Seed dispersal and its impact on maternally derived organelle haplotype structure alone can be modelled relatively simply. A number of seed dispersal models have been developed to look at the relationship between seed dispersal and colonisation speed of an advancing species and also the genetic structure brought about by colonisation (e.g. Le Corre *et al.* 1997; Petit *et al.* 2000).

1.4.2 Modelling effects of pollen and seed dispersal on genetic structure

There have been a number of models developed that spatially simulate both seed and pollen dispersal across a landscape. One approach is to devise mechanistic models of pollen and seed dispersal based on the physical properties of propagules and the environment (Greene and Johnson 2000). A second approach is to use diffusion and logistic growth models to depict the colonisation of habitat (e.g. Andow *et al.* 1990). A third approach is to use the probability density functions of propagule dispersal with reproductive success to determine the probability of dispersal between points in an array of cells (e.g. Higgens and Richardson 1996).

1.4.3 Gene flow modelling as an aid to forest management and conservation

Application of simulation modelling to gene flow analysis in plants has mainly concentrated on evaluation of anthropogenic effects such as climate change (Takenaka 2001), the effect of transgenic plants on natural populations (Waines and Hegde 2003, Burczyk *et al.* 2004) and the effect of logging in tree populations (Degen *et al.* 2002; Degen 2004).

Burczyk *et al.* (2004) developed a spatially explicit model called STEVE (Simulation of Transgenic Effects in Variable Environment). The STEVE model was applied to investigate the potential escape of transgenes from genetically engineered trees in hybrid poplar plantations into native populations of poplar. The model assessed the importance of rare long distance dispersal in determining the impact of transgenic populations and the strong influence of transgenic competitiveness and habitat availability in determining the extent of transgene movement (Difazio *et al.* 2003, Burczyk *et al.* 2004).

Another gene flow model ECOGENE (Degen *et al.* 1996) was developed to analyse the effect of logging on the genetic system of forest tree species. The model employs a database of genetic and ecological data for temperate and neotropical tree species combined with a data generation engine to supply missing data and is validated through comparison with real data collected. Although developed for temperate forests, ECOGENE has been adapted for tropical trees and applied to investigating effects of insect pollination behaviour on pollen dispersal, selfing and effective population size for *Jacaranda copaia* and *Dipteryx odorata* (Degen 2004).

There has been little previous interest in developing long-term sustainable forestry, as it involves a high level of investment. However, if a system is to be developed, then understanding the long-term viability of timber species is very important. Information about the quantity and genetic quality that is likely to result from different planting and harvesting strategies is also useful. Therefore, modelling the long term genetic response of a species under different planting and harvesting scenarios can provide a valuable tool for the development of sustainable tropical forestry (Alvarez-Buylla and Garcia-Barrios 1993).

1.5 Biological systems in tropical trees

The development of molecular methods and statistics has led to a large number of recent studies directly investigating the genetic characteristics of tropical trees and their genetic response to different physical and biological factors (see Table 1.1 for a summary of these studies). In general, genetic analyses of tropical plants have tended to focus on commercially important species that have become vulnerable through over-exploitation. Indeed, most of the literature concerning direct analyses of gene flow in tropical trees has concentrated on the effects of fragmentation on gene flow. Gene flow of secondary forest species or common pioneer species has been largely understudied.

Attempts to characterise genetic variation in tropical trees have been performed on a variety of species and at different spatial scales. Most species have been found to maintain more variation within than among populations (Hamrick *et al.* 1992), indicative of the effective means of gene flow between trees (Schierenbeck *et al.* 1997; Loveless *et al.* 1997; White *et al.* 1999).

1.5.1 Genetic structure in tropical forests

Results from indirect analyses of genetic structure in many tropical species over small scales (1 to a few hectares) have tended to find low levels of population differentiation. The most likely cause of this is a high level of gene flow via long distance dispersal of pollen as this creates greater genetic connectivity between populations and reduces genetic structuring (Ennos 2001; Dutech 2002). Other factors that can lead to low levels of genetic structure in tropical forests include random mortality of seedlings, which in some instances can lead to the vanishing of spatial genetic structure in adults (Hamrick *et al.* 1993). Inbreeding depression can also eliminate progeny from mating between related individuals caused by restricted gene flow (Epperson 1992), and recent

demographic disturbances including variation in local density or extinction-recolonisation processes may also break up genetic structuring (Knowles *et al.* 1992).

Tropical tree species show different degrees of genetic structuring. Boshier *et al.* (1995) showed that near neighbours of *Cordia alliodora* were more genetically related than more distant trees. Using random amplified polymorphic DNA (RAPD) markers on eight tropical tree species, Degen *et al.* (2001) showed significant spatial structure at distances of up to 300 m. However, there was very little structure found in *Carapa procera* populations in French Guiana (Doligez and Joly 1997) and in *Tachigali versicolor* populations in Panama (Loveless *et al.* 1998).

When there is little structure in adult populations there may be genetic structure in seedling populations. This is because limited seed dispersal can lead to aggregation of seedlings, then compensatory mortality due to environmental heterogeneity (e.g. the formation of gaps). Intra and interspecific competition on small-diameter trees can also lead to a reduction in genetic structuring of larger/older trees. Further mortality of adults increases the thinning process so that few individuals survive to become adults potentially causing a more random genetic distribution of adults (Kevin *et al.* 2004). In populations from Panama, Hamrick *et al.* (1993) showed that spatial genetic structure was present in small and intermediate diameter classes for *Platypodium elegans*, *Alseis blackiana* and *Swartzia simplex* when absent from adult populations. Kevin *et al.* (2004) also found significant spatial aggregation (although not necessarily genetic structuring) in medium and small diameter classes for both *Shorea leprosula* and *Shorea ovalis* ssp. *serica* with small diameter trees more clumped than medium diameter trees.

The spatial distribution of adults in a population also impacts on the genetic structure. Simulation studies (Doligez *et al.* 1998) have shown that clustering of individuals can

favour gene flow among individuals in the same patch and produce a stronger genetic structure in the clustered populations than in more continuous populations. Moreover, recolonisation could also enhance spatial genetic structure, if clusters are founded from different maternal origins (Dutech 2002; Bialozyt *et al.* 2004).

Stacey *et al.* (1996) suggests the following generalisations for low-density tropical trees pollinated by diverse small insects. In populations characterised by a high degree of clustering a large proportion (up to 90%) of the matings are with nearest neighbours, with a smaller but substantial fraction occurring over distances of a few to possibly several hundred meters. Conversely, where the spatial distribution of a species is more regular a large proportion of outcrossed pollen moves well beyond the closest reproductive neighbours.

Based on the mating patterns for three insect-pollinated species, *Calophyllum longifolium*, *Spondias mombin* and *Turpinia occidentalis,* Stacey *et al.* (1996) estimated that for a species with an evenly dispersed population and a low density of reproductive adults, the smallest area required for a natural breeding unit, defined as the minimum area within which 95% of the pollen received by a centrally located adult originated, would extend to a minimum of 60 ha. For species characterised by clustering of reproductive trees, a natural breeding unit would occupy at least 40 ha.

This difference in the pattern of genetic structure between uniform and clustered populations can be largely explained by the foraging behaviour of insects. Larger pollinators may 'trapline' (using the air currents above the canopy to move large distances) among widely dispersed conspecific plants. However, smaller insects may only forage under the canopy and movements by these insects between trees may be more haphazard. Where flowering adults occur in clusters, neighbouring conspecifics

would be readily discovered and visited. Where flowering individuals are widely dispersed, locating conspecific individuals could be more difficult and so pollinators may bypass nearest neighbours in their search for pollen and/or nectar (Stacey *et al.* 1996; Degen *et al.* 2004).

Dutech *et al.* (2002) used microsatellites to examine the genetic structure of the tree species *Vouacapoua americana*. They expected strong structuring at different spatial scales due to biological factors limiting gene flow. These factors included low seed dispersal and low adult density (less than 10 individuals per ha), which should produce a high genetic relatedness among individuals separated by a few meters and relatedness should decrease with geographical distance. They found that kinship coefficients were significantly higher than expected (Dutech *et al.* 2002); consistent with previous observations that seed dispersal is restricted. Seeds are dispersed short distances by gravity and rodents and this promotes spatial clustering of parents and their progeny. Furthermore they observed a reduction in the kinship coefficient between 30 m and 150 m. This general pattern seems to confirm that low seed dispersal is the basis of clustering of related individuals in the first 30-60 m. In contrast, the contribution of pollen at this scale probably has little influence as suggested by indirect measures. These show that pollen flow occurs over greater distances and should reduce kinship coefficients in small distance classes.

V. americana is also characterised by relatively low pollen flow, due to the small average foraging distance of its insect pollinators (pollination events are presumed to be restricted to 100 m). In association with low seed dispersal, this produces strong spatial structure among geographically close groups of individuals (Dutech *et al.* 2002). Results from other studies (Epperson and Alvarez-Buylla 1997; Strieff *et al.* 1998) have reported lower kinship coefficients for other species and a sharp decrease of genetic

relatedness among individuals when the distance separating them is more than a few meters.

1.5.2 Breeding system

Most mating systems of tropical trees range from a mixture of outcrossing and selfing (mixed mating) to predominant or complete outcrossing (Murawski 1995). Mixed mating systems can be dynamic and sensitive to ecological factors such that the level of outcrossing can change from one flowering event to another (Degen *et al.* 2004; Kenta *et al.* 2004). The ability to change mating system and tolerate inbreeding is a significant advantage for tropical trees as they often occur at low densities. Thus, the level of inbreeding in a species may be significantly correlated with density (Murawski and Hamrick 1992; Latouche-Halle *et al.*2004). Although high-density plots may contain individuals that are completely outcrossing there still may be high levels of biparental inbreeding (mating between related individuals), as pollinators do not have to move far to gain access to food sources. For example, in high density plots of *Symphonia globulifera,* Degen *et al.* (2004) found low selfing rates but high biparental inbreeding showing limited pollen dispersal in dense plots. In the hummingbird pollinated shrub *Helicteres brevispira*, the density of flowering trees led to either traplining (pollinators flying using currents above the canopy) in low density plots or territorial pollination in high density plots. Both phenomena had the effect of promoting outcrossing where the species was found in lower densities (Franceschinelli and Bawa 2000).

Some species with a mixed mating system show highly adaptable outcrossing rates. For example, the breeding system of *Cavanellesia plantafolia* (Murawaski and Hamrick 1992), a neotropical canopy tree, and *Shorea trapezifolia* (Murawaski *et al.* 1994) can change from completely outcrossing to completely selfing. Other tropical trees with mixed mating systems include *Ceiba pentandra* (Gribel *et al.* 1999; Lobo *et al.* 2005),

Dicorynia guianensis (Latouche Hallé *et al.* 2004), *Shorea leprosula* (Nagamitsu *et al.* 2001), *Shorea magistophylla* (Murawaski *et al.* 1994) and *Stemonoporus oblongifolius* (Murawaski and Bawa 1994).

Outcrossing rate can be influenced by habitat as well as density. Lim *et al.* (2002) found that outcrossing rates in *Dryobalanops aromatica* are greatest in primary forest (0.86 – 0.92) and smaller in logged (0.77 - 0.79) or artificial forest (0.55 – 0.67) with high correlated mating and biparental mating also detected. Species may also respond differently to habitat change. For example Dayanandan (1999) found no increase of inbreeding in the tropical tree, *Carapa guianensis* in fragmented forest. Collevatti *et al.* (2001) looking at recently fragmented populations of the tropical tree *Caryocar brasiliense*, also found similar levels of inbreeding in fragmented and disturbed populations as found in continuous and non-disturbed areas. However, as tree species have a long life cycle, the establishment of the population in these studies predates the start of fragmentation; therefore the observed mating system may reflect an earlier non-disturbed habitat.

1.5.3 Pollen dispersal and pollinator behaviour in tropical tree species

Studies of neotropical tree species indicate that pollen flow can occur over long distances (Nason *et al.* 1998; Bush and Rivera 1998). Estimates for animal pollinated species have reached greater than 4.5 km for *Swietenia humilis*, (White *et al.* 1999) and up to 632 km for the wasp pollinated *Ficus dugandii* (Nason *et al.* 1998).

Pollination in the tropics is predominantly dependent on animal vectors, and insects (especially bees) are the primary pollinators of most tropical trees (Bawa 1990; Fenster *et al.* 2004). Bats and hummingbirds are also important pollinators but, in contrast with northern temperate forests, wind pollination is uncommon. Although wind pollination is

generally effective for promoting pollen mediated gene flow the dense vegetation of tropical forests can lead to restricted gene flow (e.g. for *Araucaria angustifolia* (Sousa and Hattener 2003)). The capacity of insect pollinators to move long distances has been proven (Webb and Bawa 1983; Gribel *et al.* 1999, Dick *et al.* 2003), although the potential to move between stands depends on their behavioural response to open habitat.

The lower canopy is populated with smaller insects that have lower foraging ranges (Bawa *et al.* 1985), whereas emergent trees may be largely pollinated by larger bees and insects that fly above the canopy layer over large distances. Little is known about the foraging behaviour of the diverse array of small insects that occur in neotropical rain forests, although it has been suggested that the capacity for long distance pollen movement by such species is limited because of their restricted foraging ranges (Bawa 1977). Pollen flow analysis of the tropical tree *Gliricidia sepium* (Dawson 1997) is consistent with this type of pollinator behaviour. The species is pollinated by small bees that predominantly show nearest neighbour foraging with occasional long distance intervening flight. However, Stacey *et al.* (1996) identified substantial levels of pollen movement, over 300 m, for two species pollinated by diverse small insects. Thus, even insects with limited foraging ranges may occasionally fly long distances. Larger insect pollinators may contribute more to long distance gene flow (Dick *et al.* 2003; White *et al.* 2002; Dayanandan *et al.* 1990). For example, the emergent tree *Neobalanocarpus heimii* is pollinated by stingless bees and honeybees, the latter are known to be long distance pollinators (Dayanandan *et al.* 1990). Paternity analysis by Konuma *et al.* (2000) on populations of this species found the average mating distance to be 524 m.

Different patterns of pollinator foraging, as detected by variable patterns of gene flow, depend on the spatial structure of sub-populations, and overall patterns of gene flow can be strongly affected by the spatial distribution of reproductive trees (Bawa 1998; Stacey

et al. 1996; Konama *et al.* 2000). In a study of *Spondias mombin, Calophyllum longifolium* and *Turpinia occidentalis* in undisturbed Panamanian forests, Stacey *et al.* (1996) found that most of the pollen movement occurred between nearest neighbours when trees were clustered, whereas pollen flow to isolated trees tended to occur over greater distance. Ghazoul *et al.* (1998) reported that the pollinator (*Trigona fimbriata*) of *Shorea siamensis* flew much further in an area with long distances between reproductive tress than in an area with a higher density of reproductive trees, although there was a limit to interflowering-tree movement by pollinators. Paternity inference for *Swietenia humilis* has also revealed long distances of cross-pollination in addition to a breakdown of nearest neighbour mating in low density populations in disturbed habitats (White *et al.* 2002). Bees and other insects may have difficulty in finding nearest neighbours when flowering conspecific plants are broadly spread. Thus, there may be a threshold of distance over which nearest neighbour mating breaks down.

Other biological factors, such as a species' phenology, can also affect pollinator behaviour. Zucchi *et al.* (2003) found that the restricted gene flow in *Eugenia dysentia,* found in the Brazilian Cerrado, is a consequence of the fast and abundant flowering of the species, which does not promote a large number of flights of the bee pollinator between individuals. Changes in the flowering abundance of a species can have such an effect that there is a pollinator shift. For example the tropical emergent tree *Dipterocarpus tempehes* was pollinated by giant honeybees (*Apis dorsata*) in a year of mass-flowering but by several species of moth in a year of less-intensive flowering (Kenta *et al.* 2004).

Changes in the landscape can also cause pollinator shifts that will then impact on gene flow. Dick *et al.* (2003) studied populations of *Dinizia excelsa* in disturbed habitats, where African honeybees were the primary pollinators, and also populations in

undisturbed habitats, which were visited exclusively by stingless bees and small beetles. In the disturbed habitat there was a very low density of reproductive trees that were either present in fragments or as single trees in pastures. In this landscape the African honeybee acts as a pollinator: they are present in high density and have a preference for agricultural habitats. They found that this bee is not only capable of long distance dispersal but that there was a higher fecundity of trees visited by the African honeybee regardless of the habitat matrix. Dick *et al.* (2003) concluded that in fragmented habitats, African honeybees might mitigate the loss of genetic diversity by expanding genetic neighbourhood areas until rainforest regenerates in the abandoned pastures. Observations indicate that stingless bees and beetles do not cross open pasture so, without pollinator shift, plants pollinated by these species alone are more vulnerable to habitat fragmentation.

Trees with non-specialist pollinators are probably less susceptible to the effects of disturbance on pollen flow compared to trees with a specialist pollinator relationship. However, it is likely that there is a threshold distance after which pollinators will not forage. As this threshold is likely to be pollinator specific, the extent to which landscape changes affects gene flow in trees may be highly variable among species.

1.5.4 Seed dispersal in tropical tree species

Studies of seed dispersal in numerous plant species consistently reveal a rapid decline in seed density with distance from the parent plants (Harper 1977; Howe and Smallwood 1982; Willson 1993). However, patterns of seed survival and the resulting establishment patterns are quite variable (Condit *et al.* 1992; Hammond and Brown 1998). Post-dispersal seed predation is often the major cause of mortality throughout a plants life cycle, and one of the principal processes underlying plant recruitment patterns (Janzen 1971; Nathan and Casagrandi 2004).

Janzen (1970) and Connell (1971) suggested that relatively low recruitment near conspecific adults should be expected because the processes of seed dispersal and seed survival are distance-dependent; dispersal decreases while survival increases with distance from the parent plant due to density-responsive seed predators. Gene flow via seed dispersal can therefore be considered to be consequence of interacting factors, predominantly factors affecting seed dispersal and post-dispersal seed predation but also factors such as germination rate, microhabitat variability (Willson 1988; Kadmon and Shmida 1990), intra- and interspecific competition (Barton 1993; Fowler 1995), and seedling herbivory (Connell 1971; Packer and Clay 2003).

Many tropical tree species have animal dispersed seeds. Therefore, seed-mediated gene flow is significantly affected by the behaviour of disperser species. Restricted seed-mediated gene flow can bring about genetic structuring, so an understanding of seed dispersal is critical in understanding the population genetics of a species. As gene movement through seeds may be limited, so genetic structuring of maternal genotypes may be much greater than that shown by paternity analysis. Limited seed dispersal may also correlate with limited pollen dispersal as limited seed dispersal tends to cause individuals to become more aggregated (Hardy *et al.* 2006). Therefore, trees occur at higher local densities and pollen may disperse less under higher density (Stacey *et al.* 1996; Vekemans and Hardy 2004).

Species with animal dispersed seeds can have higher mean levels of among population genetic heterogeneity than their potential for long distance seed dispersal would indicate. Hamrick *et al.* (1993) found that the magnitude of spatial genetic heterogeneity resulting from seed dispersal depends on several factors. Firstly, the proportion of seeds that immigrate and contribute gametes to the next generation of an established

population versus the proportion that colonise new habitats: if most seed movement is among established populations there will be less genetic structure than if the majority of the seeds colonise newly available habitats. Secondly, the density of reproducing adults: if seeds from several adults colonise an open habitat, genetic structure would be less than if successful colonists had come from one or a few adults. Thirdly, the foraging and deposition behaviour of the seed dispersal agent, if the seed dispersal agent has territorial or limited foraging and deposition the resulting populations should show more structure (Hamrick *et al.* 1993).

Hamilton (1999) has shown that with regard to *Corythophora alta,* a canopy tree, seed dispersal is limited and fragments as large as 10 ha can be composed of a single maternal lineage. In *C. alta* the restricted seed mediated gene flow has led to sharp genetic differentiation of adjacent populations indicating that small forest lots may exchange seeds infrequently, thus limiting the size of genetic neighbourhoods. Population isolation was also found in *Caesalpina echinata* populations, using chloroplast microsatellites Lira *et al.* (2003) found that limited seed dispersal in this species had led to five of seven populations showing no variation. Because gene flow through seeds is responsible for two thirds of the total genetic neighbourhood size, it is essential for estimates of the size of tropical tree breeding populations. Therefore, gene flow estimates from both seed and pollen may better predict genetic changes in tropical forest trees (Hamilton 1999: Hamilton and Miller 2002).

There are tropical trees that show high levels of gene flow via seed dispersal. Frequent long distance seed dispersal was found in the vertebrate dispersed tree *Simarouba amara* (74% of seedlings established greater than 100 m from the maternal parent), with seed-mediated gene flow equivalent to that by pollen (Hardesty *et al.* 2006). Sezen *et al.*

(2005) also found high levels of seed mediated gene flow during colonisation by the palm species *Iriartea deltoidea.*

The extent of seed dispersal is highly affected by the behaviour of seed dispersers and this behaviour may vary in different habitats or after fragmentation. For example, seeds of *Caryocar brasiliense* are mainly gravity dispersed and so tend to remain under the mother tree canopy (Aldrich 1997; Collevatti *et al.* 2001). *Caryocar brasiliense* individuals are spatially distributed in clumps and the territorial bats that pollinate the species and feed on the fruits tend to forage inside the clumps, so restricting gene flow between clusters. Fragmentation may further isolate populations by keeping the bats within isolated clumps and so increasing genetic differentiation between isolated clusters. Aldrich (1997) found that bats will preferentially deposit fruits from pasture trees into remnant forest and this mechanism of concentrating recruitment from pasture trees resulted in remnant forest receiving a broad spectrum of seed genotypes.

Although the majority of seeds are animal dispersed there are still some species that have wind-dispersed seed (e.g. *V. ferruginea* and *Swietenia* spp.). Hamrick *et al.* (1993) found that species with wind dispersed seeds and with lower densities develop more structuring in their seedling population than species with animal dispersed seeds or higher densities. They also found that the effect of seed dispersal on the establishment of genetically related near neighbours (e.g. half and full sibs) had implications for demographic and reproductive processes. In species whose seed dispersal mechanisms promote a strong patch structure, competition for water, nutrients and light will often be among related individuals. In addition, the short distances that may separate susceptible seedling cohorts may facilitate the spread of pathogens. In species where the genetic structure persists into the adult generation, the likelihood of inbreeding is increased (Hamrick *et al.*1993).

It is generally assumed that wind seed dispersal in tropical forests will be limited due to the high density of vegetation. However, the distance may be underestimated by not taking into account tropical storms (Nason *et al.* 1997). Hamrick *et al.* (1993) found that for *Platypodium elegans,* a tree with wind dispersed seeds, there were more near neighbour pairings with higher numbers of alleles in common and less structure in the seedling population than in the bird dispersed species *Swartzia simplex*.

Genetic structure brought about by seed dispersal is also affected by reproductive success of maternal parents. Using a maximum likelihood maternity analysis to estimate female fertilities for maternal trees across seedlings and saplings in populations of *Gleditsa triacotes*, Schnabel *et al.* (1998) found that the three highest fertility females accounted for 58% of the 1313 progeny in one site and 46% of the 651 progeny at another. Their results showed that the combination of seed dispersal mechanisms and fertility variation were sufficient to explain the fine scale genetic structure in the species.

1.5.5 Gene flow response to fragmentation

Fragmentation of tropical forests, because of harvesting for wood products and clearance for agriculture, has been changing the landscape to a mosaic of remnant forest habitat surrounded by open areas (Collevatti and Hay 2001). Habitat fragmentation may reduce genetic variation through genetic bottlenecks, while subsequent founder effects, genetic drift and restricted gene flow may increase population genetic isolation and divergence. Therefore, the genetic response to fragmentation has been well studied in the context of tropical tree populations.

Fragmentation decreases the size and increases the spatial isolation of a population and its effects on the dynamics of gene flow between fragments may ultimately have detrimental consequences for the evolutionary viability of a population, by way of increased levels of inbreeding and random genetic drift. Forest fragmentation impinges on plant reproduction by altering the species composition and behaviour of pollinators. Pollinator limitation can lower the fecundity of host plants (Ghazoul *et al.* 1998) and increase levels of inbreeding (Murawski *et al.* 1994). Evidence from temperate and some sub-tropical plants suggests that not all fragmentation events lead to detectable genetic loss unless a fragmentation threshold is reached. Furthermore, in some cases fragmentation may even increase gene flow among remnant populations as habitat disturbance can enhance pollinator activity and so increase fecundity and gene flow (Collevatti *et al.* 2001; White *et al.* 2002).

A variety of different effects of fragmentation on gene flow of tropical trees have been found. Jennersten (1988) showed that habitat fragmentation resulted in a lower flower variation and seed set in *Dianthus deltoides* when compared to non-fragmented habitats. Aizen and Feinsinger (1994) also found that pollination level and seed output decreased nearly 20% from continuous forest to fragments.

Fragmentation can lead to increased genetic isolation of remnant populations and result in increased inbreeding. Rocha and Aguilar (2001) found that in the predominantly outcrossing neotropical leguminous tree *Enterolobium cylocarpum*, fragmentation led to decreased pollen flow to isolated trees. Consequently, isolated trees had greater levels of selfing and produced seedlings with lower vigour. This effect was in spite of some successful long distance pollen mediated gene flow by hawkmoths with extensive foraging ranges.

Genetic isolation from fragmentation has a greater impact when seed or pollen dispersal is limited. For example, Collevatti *et al.* (2001) found decreased gene flow, and significant inbreeding, in the widely distributed but rare Cerrado tree, *Caryocar brasiliense*. Its seeds are dispersed by gravity and by the large flightless bird species *Rhea americana* and there is evidence of limited seed dispersal and also restricted pollen dispersal within fragment patches due to pollination by territorial bats.

Increased isolation of populations may not necessarily lead to inbreeding if gene flow is extensive. White *et al.* (2002) found that within fragmented populations of *Swietenia humilis* there was an excess of homozygotes. However, this was not a consistent factor and may be a result of the local genetic structure brought about by low seed dispersal. They found that pollen movement to *S. humilis* trees depended on the spatial structure of the particular fragment, with subpopulation size a critical factor, directly influencing the proportion of immigrant pollen. Although the trend of near-neighbour mating was maintained within the fragments, reductions in subpopulation size were paralleled by an increase in the proportion of long distance pollen flow from outside the fragment. A single very isolated tree, separated from the nearest fragment by 1.1 km, experienced direct pollen flow from remote flowering trees, with over 70% coming from the maximum distance category (>4.5km). This contrasts with predictions by Murawski and Hamrick (1992), who suggest that spatially isolated trees are more likely to deviate from random mating and receive pollen from fewer donors. In fact, increase in spatial isolation promoted long distance gene flow between this isolated tree and other fragments (White *et al.* 2002), although different levels of fragmentation may have different effects within the species and certain landscape patterns may be causing greater inbreeding in *S. humilis* (White *et al.* 2002).

Fragmentation may differently affect the genetic diversity and structure of seedlings compared to adult populations. Dayanandan *et al.* (1999) found that a few remnant individuals of a neotropical timber tree, *Carapa guianensis,* present in pastureland surrounding fragments, contributed disproportionately to the sapling cohort present in those fragments. This created a greater genetic distance between the adults (established prior to fragmentation) and the sapling cohort within fragments. Also, unlike other cohort comparisons, the saplings showed reduced allelic richness. They also found greater population structure in the sapling cohorts suggesting that fragmentation has lowered gene flow in this species. Aldrich *et al.* (1998) also reported an increase in genetic divergence in seedling populations of *Symphonia globulifera* following deforestation.

Although genetic impacts of fragmentation are complex and will vary among species, remnant trees and fragments can provide a buffer to the deleterious genetic effects of habitat destruction and may be vital to the long-term viability of a species (White *et al.* 2002). Many pollen flow studies of trees in neotropical mosaic landscapes have found unexpectedly high levels of gene flow (reviewed in Nason and Hamrick 1997). This may be a result of the spatial dispersion of the remnant trees, which may cause pollinators to bypass neighbouring plants (Chase *et al.* 1996), or it may be because of shifts in pollinator composition, as when feral honeybees replace native pollinators in disturbed habitats (Dick *et al.* 2003).

1.5.6 Tropical pioneer species

Most tropical canopy trees occur at a low density in primary forest and so long distance seed and/or pollen dispersal are expected. As a result, they often show little population genetic structure (Latouche-Hallé *et al.* 2004). A similar pattern is expected for pioneer trees colonising tree-fall gaps in primary forest, where they grow at high density but the

gaps themselves are spatially dispersed. Although pioneer species require high levels of dispersal to colonise isolated gaps in primary forest, they are also able to rapidly regenerate in large open areas and genetic structuring may be different in highly disturbed habitats. Wright (1940) expected colonising species to show population genetic structure as new populations are started by a few individuals and are relatively isolated from other populations. More recent theory shows that a regular extinction/recolonisation dynamic, with some migration, can augment gene flow thus reducing divergence among populations. Investigating differences in genetic structure in early and late succession populations of the pioneer *Antirhea borbonica*, Litrico *et al.* (2005) found no evidence that founder events increased differentiation among colonising populations and proposed that multiple colonisation events from different sources limits differentiation.

Gene flow differences between primary and secondary forest populations have been found in *Cecropia obtusifolia* a dioecious, wind pollinated tropical pioneer that colonises tree-fall gaps but is also found in dense stands on artificially disturbed sites. Using indirect measures of gene flow, Alvarez-Buylla and Garay (1994) found little population structure over distances 1 to 107 km and postulated widespread gene flow. However, genetic analyses have shown significant fine scale structure at the intrapopulation level for seedlings and juveniles, interpreted as resulting from the clustering of genetically correlated progeny of individual mother trees because of limited animal-based seed dispersal (Alvarez-Buylla *et al.* 1996; Epperson and Alvarez-Buylla 1997; Kaufman and Smouse 1998). Previous studies of *C. obtusifolia* (Alvarez-Buylla and Martinez-Ramos 1990) show that although the great majority of seeds fall within 30m of the mother tree they can be found as far as 100 km away. Seed dispersal could therefore have a significant impact on gene flow, but because most seeds are dispersed close to the parent plant, populations are expected to show some fine scale

structure (Kaufman and Smouse 1998). As the stands of *C. obtusifolia* mature, they suffer progressive mortality, and there will be successively less genetic structure among surviving saplings, young adults, and mature adults (Kaufman and Smouse 1998). *Cecropia obtusifolia* adults in the disturbed habitat produced viable seeds and male and female flowers at the same time as in the primary forest. Therefore, Kaufman and Smouse (1998) proposed that the density of trees in the disturbed habitat influenced the pollen flow pattern, leading to a greater barrier to pollen flow from outside of the dense stand in disturbed habitat.

Colonisation from primary forest into adjacent abandoned pasture can also lead to a reduction in diversity, due to a low number of adults contributing to the founding population. Greater seed dispersal and clustering of sibs brings about differently structured founder populations, with those founder populations showing more differentiation than primary forest populations.

Investigating the founder populations of *Iriartea deltoidea*, an abundant shade-tolerant canopy palm that freely colonises secondary forest, Sezen *et al.* (2005) found a reproductive dominance of very few individuals from the primary forest. The genetic consequence of this was that founder populations had significantly lower genetic diversity than primary forest. In addition the founder populations consisted largely of half or full sibs and formed patches of similar genotypes that were larger than in primary forest. *Iriartea deltoidea* is pollinated by stingless bees (*Trigona* spp. and *Melipona* spp.) and now by honeybees (*Apis* spp.) and seeds are dispersed by a variety of birds and mammals. Sezen *et al.* (2005) also found that the process of colonisation affected seed and pollen flow distances leading to average seed dispersal distances much greater than those for pollen. The median seed dispersal distance was 270 m with the longest dispersal distance detected more than 875 m whilst maximum pollen

movement was less than 220 m. However, extensive gene flow may mediate diversity loss. In investigating the regeneration ability of *Swietenia macrophylla* in pastures, Cespedes *et al.* (2003) showed that high levels of gene flow from both seed and pollen allowed regeneration into newly abandoned pastures with relatively high levels of diversity. *Elaeocarpus grandis,* an early successional tree of Australian rainforest, also shows high levels of seed-mediated gene flow. Although insect pollination may be restricted, extensive seed dispersal by birds and bats allows colonisation to suitable habitat (Rosetto *et al.* 2004).

Studies to date suggest that the process of colonisation from primary forest to available habitat may lead to significant changes in the genetic properties of those populations with potentially large decreases in genetic diversity, which may not be readily mediated by pollen flow from neighbouring primary forests (Cespedes *et al.* 2003; Sezen *et al.* 2005). To facilitate sustainable utilisation of these increasingly important secondary forest resources, further understanding of the genetic impact of colonisation processes can be an important aid for conservation of a species genetic resource.

1.6 Objectives of study

This study is an investigation of the genetic response of *Vochysia ferruginea*, a pioneer species useful for regeneration and timber in Costa Rica, to colonisation from primary into secondary forest. The main objectives of this study are:

- To assess the differences in genetic diversity and allelic richness in seedling, secondary forest and primary forest poplations and to analyse this against the expected population genetic response of a loss of alleles and a lowering of diversity during colonisation.
- To asses the genetic structure of dense regenerated populations and the impact of structure and density on mating system and the extent of inbreeding.

- To evaluate whether increased homozygosity from greater levels of inbreeding can be correlated to a measure of fitness.
- To develop a simulation model to evaluate how the process of gene flow during colonisation impacts a species genetic diversity and differentiation.
- To use molecular data to parameterise the model and test management scenarios relevant to the sustainable forestry of pioneer species in regenerated forest.

Two sites have been selected to allow examination of adjacent primary and secondary populations: from these, samples of adult, seedling and seed life stages have been collected. The sampling strategy was designed to allow accurate estimation of mating system parameters, levels of genetic diversity and differentiation and gene flow. Populations will be screened for variation using five microsatellite primers.

1.6.1 Genetic diversity and differentiation

Genetic diversity and differentiation will be analyzed at both sites in primary, secondary, seedling and seed (progeny) populations. This information will be used to examine the effects of colonisation on the level and partitioning of genetic variation. By analysing diversity and differentiation at different life stages the temporal variation in patterns of genetic diversity can be assessed. Through genotyping the mature trees in the primary forest, those in the secondary forest and any remnant trees in the pastureland it is possible to estimate the source population that contributes to a regenerating population of *V. ferruginea*. This will aid understanding of the effects of recolonisation on the genetic dynamics of a pioneer species and have practical applications for genetic resource management in regenerated forest.

1.6.2 Breeding system

Vochysia ferruginea is self-compatible and will produce viable offspring in controlled crosses, although it is not known to what extent the species is inbreeding in natural

populations. Data from AFLP analysis (Cavers *et al.* 2005) suggests that whilst inbreeding may be present there is a significant level of outbreeding within populations. Therefore, it is initially important to determine the level of inbreeding in both primary and secondary forest. Estimation of the level of inbreeding using codominant, multilocus markers from both sites should provide a robust estimate. Levels of biparental inbreeding will also be examined in both primary and secondary forest.

In both populations the data on growth rates of mapped adult trees will be correlated with the level of heterozygosity. Growth rate data will be used as a measure of fitness to estimate the possible effect of inbreeding in *V. ferruginea.*

1.6.3 Paternity analysis

Gene flow in *V. ferruginea* populations in both primary and secondary forests will be investigated at both sites. All adult trees (DBH > 10 cm) in a continuous plot will be sampled, mapped and genotyped, then seed will be collected and genotyped from a number of these trees. Paternity and TwoGener analysis of this data will be used to estimate gene flow via pollen.

1.6.4 Regeneration dynamics

Sampling of *V. ferruginea* seedlings in pastureland will provide information that can be used to investigate the genetic characteristics of regeneration in *V. ferruginea*. By sampling 50 adults in pastureland and using knowledge of forest clearance dates and diameter at breast height (DBH), it will be possible to estimate which trees are remnants, originating from primary forest, and those which result from regeneration into the area post clearance. Where trees from earlier regeneration are identified, comparison with the new seedling data can help elaborate conclusions on temporal changes in genetic structure, diversity and gene flow in regenerating populations.

1.6.5 Modelling gene flow

The impact of gene flow on genetic diversity and structure will be simulated in a spatial model. The model will be developed in two stages; initially, a simple model of just the seed dispersal component of gene flow will be developed and tested, then a pollen dispersal component will be included. The seed dispersal model will be validated by comparing it to the post-glacial recolonisation of Europe by oaks, a system for which a great deal of empirical genetic data exists, and for which the speed of colonisation has been estimated from the pollen core record. The effects of seed dispersal on haplotype diversity in modelled populations can also be compared to a nation-wide study of chloroplast DNA diversity in the United Kingdom (Cottrell *et al.* 2002). Modelling seed dispersal during oak colonisation means that the effectiveness and robustness of the model can be tested against this well studied system.

The model will then be modified to include dispersal by both pollen and seed, and used to investigate sustainable management of *V. ferruginea* for timber. Modelling of the *V. ferruginea* system will incorporate two categories of tree: light demanding and shade tolerant. These categories of tree are used so as to model the dynamics of regeneration of a low density pioneer (light demanding species) in closed forest into open areas.

Variation in the genetics of regeneration will be examined by considering the following scenarios: clear felling, partial felling leaving remnant trees and different planting strategies (e.g. using seed collected from one maternal tree or many). Each scenario will be modelled with and without gene flow from adjacent primary forest with low-density *V. ferruginea.* This allows investigation into differences in genetic diversity and density of a pioneer species in naturally regenerating secondary forest relative to primary forest, and the effect gene flow from neighbouring primary forest has on the genetics of that pioneer in secondary forest. The model will also help to determine the impact on

regeneration dynamics of different numbers of remnant *V. ferruginea* seed trees on genetic diversity and numbers of a pioneer in regenerating forest. Other scenarios will investigate the effect of different planting strategies on resulting levels of diversity and how gene flow from neighbouring forest might affect diversity once the planted forest is left to regenerate naturally. Determining how many seed trees are needed to regenerate an area and the consequences of different numbers of source trees for future populations can provide useful information for management of the species for timber and regeneration initiatives. Therefore, these scenarios should help to assess the viability of sustainable forestry of *V. ferruginea.*

Species	**Breeding system**	**Habit**	**Marker**	**Pollen flow distance**	**Pollination vector**	**Seed dispersal distance**	**Seed dispersal vector**	**Inbreeding**	**Mean H_E**	**Mean H_O**	**Genetic structure**	**Reference**
Alseis blackiana		Canopy tree	Allozymes		Bees, butterflies			0.104 (F_{IS})			0.034 (F_{ST})	Hamrick *et al.* 1993
Brosimum alicastrum		Canopy tree	Allozymes		Bats		Bats, arboreal mammals	0.120 (F_{IS})			0.050 (F_{ST})	Hamrick *et al.* 1993
Calophyllum longifolium	Polygamous hermaphrodite t_m = 1.030 t_s = 0.974	Canopy tree of wet forest	Allozymes	Majority >210 m	Small insects		Mammals, primarily bats. Water					Stacey *et al.* 1996
Enterolobium cyclocarpum	Predominantly outcrossed *t_m =0.812-0.913*	Leguminous tree of dry forest	Allozymes		Moths, beetles, other small nocturnal insects							Rocha & Aguilar 2001a,b
Euterpe edulis	Monoecious t_m = 0.94-1.04	Understory palm	Allozymes		Insects				0.278-0.445	0.264-0.423		Conte *et al.* 2003
Spondias mombin	Hermaphrodite t_m = 0.989 t_s = 0.996	Widespread canopy tree of wet forest	Allozymes	300 – >350 m	Small insects		Wide range of frugivores					Stacey *et al.* 1996
Swartzia simplex		Understory tree	Allozymes		Large bees		Birds	0.161 (F_{IS})			0.037 (F_{ST})	Hamrick *et al.* 1993
Turpinia occidentalis	Hermaphrodite t_m = 1.006 t_s = 1.071	Sub canopy tree of wet forest	Allozymes	Majority <130 m	Small insects		Primarily by aboreal frugivores. Secondarily by mammals					Stacey *et al.* 1996
Antirhea borbonica	Functionally dioecious, outcrossing	Pioneer, persists in canopy	Microsatellites		Range of insects		Birds	0.16 to 0. 49 (F_{IS})	0.37 to 0.48	0.2 to 0.36	F_{ST} = 0.04 to 0.19	Litrico *et al.* 2005
Astrocaryum mexicanum		Subcanopy tree of rain forest	Microsatellites	60 m	Beetles							Nason *et al.* 1998
Caesalpinia echinata	Primarily outcrossed	Secondary canopy tree in dry seasonal forest	Chloroplast microsatellites		Probably carpenter bees	4-5m	Gravity (explosive)				(Φ_{ST}) = 0.9106	Lira *et al.* 2003

Species	**Breeding system**	**Habit**	**Marker**	**Pollen flow distance**	**Pollination vector**	**Seed dispersal distance**	**Seed dispersal vector**	**Inbreeding**	**Mean H_E**	**Mean H_O**	**Genetic structure**	**Reference**
Carapa guianensis			Microsatellites					-0.22 - 0.14(F_{IS})		0.12 - 0.89	0.074 R_{ST} (saplings) 0.041R_{ST} (adults)	Dayanandan *et al.* 1999
Caryocar brasiliense	Hermaphrodite Completely outcrossed	Widely distributed tree	Microsatellites	Limited	Small bats	Low	Gravity, *Rhea Americana*	0.11 (f)	0.765	0.765	0.29 (R_{ST})	Collevatti *et al.* 2001
Cordia alliodora		Understorey tree of rain forest	Microsatellites	282 m	Beetle Other small insects							Boshier *et al.* 1995 Nason *et al.* 1998
Dinizia excelsa	Hermaphrodite predominantly outcrossed t_m = 0.63	Undisturbed forest	Microsatellites	212 m	Stingless bees, small beetles		Wind dispersed				0.104 (Φ_{FT})	Dick *et al.* 2003
		Remnant trees		1509 m	*Apis mellifera*						0.002 (Φ_{FT})	
Dipterocarpus tempehes	t_m = 0.93	Emergent	Microsatellites	222.6 m	Moths and *Apis dorsata*				0.808	0.736		Kenta *et al.* 2004
Dryobalanops aromatica	t_m = 0.55 – 0.92	Widely distributed emergent tree	Microsatellites							0.491 ± 0.060	0.062 (G_{ST}) 0.09 (R_{ST})	Lim *et al.* 2002
Eugenia dysenterica		Cerrado tree	Microsatellites		Large bees		Animal mainly humans and monkeys	-0.037 (f)	0.458	0.442	R_{ST} = 0.27 F_{ST} = 0.25	Zucchi *et al.* 2003
Gliricidia sepium	Completely outcrossed	Fast growing leguminous tree	Microsatellites	Majority <75 m some >275 m	Primarily solitary bees							Dawson *et al.* 1997
Neobalanocarpus heimii		Emergent of rain forest	Microsatellites	524 m	Stingless bees *Trigona* spp Honey bees *Apis* spp		Gravity			0.675		Konuma *et al.* 2000
Pithecellobium elegans		Emergent of rain forest	Microsatellites	450 m	Hawkmoth							Chase *et al.* 1996

Species	**Breeding system**	**Habit**	**Marker**	**Pollen flow distance**	**Pollination vector**	**Seed dispersal distance**	**Seed dispersal vector**	**Inbreeding**	**Mean H_E**	**Mean H_O**	**Genetic structure**	**Reference**
Platypodium elegans		Emergent of rain forest	Microsatellites	525 m	Small bees		Wind	0.092 (F_{IS})			0.051 (F_{ST})	Hamrick *et al.* 1993 Nason *et al* 1998 Stacy *et al.* 1996
Swietenia humilis	Monoecious self-incompatible	Small, deciduous tree of dry forest	Microsatellites	Majority up to 600 m	Small insects	Majority dispersed < 50 m	Wind			0.472– 0.90	$\rho = 0.032$	White *et al.* 1999
		Fragmented population		> 4.5 km								
Swietenia macrophylla	Monoecious	Pioneer and canopy emergent	Microsatellites	200 m	Moths and small bees	32 – 80 m	Wind	0.015 (F_{IS}) 0.149 (f) 0.038 (f)	0.45 – 0.82	0.47 –0.93	0.063(F_{ST})	Cespedes *et al.* 2003 Norvick *et al.* 2003 Lowe *et al.* 2003 Lemes *et al.* 2003
Symphonia globulifera	Hermaphrodite primarily outcrossed $t_m = 0.92$	Canopy tree of primary forest	Microsatellites		Hummingbirds		Bats, birds, monkeys and ruminants			0.800 (adults) 0.727 (saplings) 0.639 (seedlings)	$0.056R_{ST}$ (adults) $0.038R_{ST}$ (saplings) $0.238R_{ST}$ (seedlings)	Aldrich *et al.* 1998 Degen *et al.* 2004
Vouacapoua americana	Hermaphrodite with low self fertilization	Shade tolerant tree of primary forest	Microsatellites	51-166 m majority less than 100 m	Small insects	Usually up to 30m	Gravity, occasionally rodents				0.11 (F_{ST}) in Paracou 0.004 ((F_{ST}) in Nourages	Dutech *et al.* 2002

t_m = multilocus outcrossing rate
t_s = single locus outcrossing rate

Table 1.1: Summary of literature on molecular studies of gene flow in neotropical tree species.

Chapter 2: Materials and Methods

2.1 Study species and sites

Vochysia ferruginea Mart. is a fast growing pioneer and an important timber species. It is particularly successful at colonising highly disturbed areas, where it can regenerate to form dense monospecific stands of secondary forest and can also persist as a canopy tree at low density in primary forest. With low nutrient requirements and high tolerance of aluminium, it is a species expected to do well in the low nutrient/high acidity soils typical of land from which tropical forest has been removed for agriculture (Herrera and Finegan 1997). These features also make it a useful plantation species in neotropical areas, particularly suited to planting on land prone to erosion, and an important species for sustainable land use (Herrera *et al.* 1999). In Costa Rica there is a large amount of secondary forest and as a result *V. ferruginea* is becoming an increasingly important timber species (Finegan 1992). Understanding the genetic structure of *V. ferruginea* is also of biological interest, in its role as a pioneer species.

2.1.1 Vochysia ferruginea Mart.

Vochysia is the largest genus of the virtually neotropical family Vochysiaceae (Cronquist 1981), which contains a total of seven genera and over 200 species (Flores 1993). Although most species are found in humid, evergreen forest communities, others occur in seasonal woodland areas and some are characteristic of the extensive cerrado vegetation of central Brazil (Eiton 1972). *Vochysia ferruginea* is widely distributed in the neotropics from Nicaragua to Brazil. In Costa Rica it is limited to the Atlantic lowlands and parts of the southwest and does not occur in the dry northwest (Cavers *et al.* 2005). *V. ferruginea* is an emergent evergreen tree reaching 50 meters and is a gap coloniser in moist lowland forest below 1500 m. It is a fast growing species and in regenerating forest may form dense monospecific stands (see Figure 2.1) or

characteristic secondary forest patches with species such as *Hyeronima alchorneoides, Pentaclethra macroloba* and other *Vochysia* species (Flores 1993; Cavers 2002).

Breeding system

The species is hermaphrodite and, in Costa Rica, flowers between April and June. Stands may flower monosynchronously (see Figure 2.1), the timing of which is probably influenced by changes in annual rainfall.

Figure 2.1: Dense stand of *Vochysia ferruginea* in secondary forest showing synchronous flowering (yellow/orange flowers).

Flowers are visited by hummingbirds and insect pollinators (mainly large bees and butterflies). In six other *Vochysia* species pollination was found to be mediated by a great variety of pollinators: primarily bees, with occasional visits by moths, butterflies and hummingbirds (Flores 1993). In controlled pollinations Bawa *et al.* (1985a) reported *V. ferruginea* as largely self-compatible; however, self-pollinated flowers showed marginally reduced fruit set (proportion of flowers that set fruit was 0.16 in self-pollinated and 0.17 in cross-pollinated, Bawa *et al* 1985b). Despite self-compatibility in controlled pollinations, flowering in natural populations tends to be synchronous and, as pollinators include large bees capable of dispersing pollen over long distances, there is a high potential for outcrossing (Oliveira and Gibbs 1994).

The fruit of *V. ferruginea* consists of small pods, each of which contains up to six winged seeds. These are widely dispersed by both wind and birds, which may help it to establish quickly in forest gaps (Lowe *in prep*). Birds predate on the seeds, but this primarily affects the immature fruit and so influences fecundity rather than seed dispersal (Flores 1993).

Figure 2.2: The leaves, inflorescence and seed pods of *Vochysia ferruginea* (Images copyright of Environmental Sciences Program, Smithsonian Tropical Research institute, used with permission).

Figure 2.3: The seed pods and winged seed of *Vochysia ferruginea* (Images copyright of Environmental Sciences Program, Smithsonian Tropical Research institute, used with permission).

Uses

Vochysia ferruginea is locally important for timber in Central America and is used for general construction as well as fence posts, boxes, veneers and matches. As a fast growing species which tolerates degraded land, it may become of greater commercial importance and replace other timber species in the future. Finegan (1992) identified *V. ferruginea* as having strong potential as a commercially viable crop for degraded sites. As a reforestation species it may also play an important role in maintaining faunal diversity. Its flowers and fruit (see Figures 2.2 and 2.3) are predated on by a wide number of species including nectivorous birds, bees and butterflies and frugivorous

birds such as parrots, toucans and trogons. The fruit is also predated on by peccaries, tapirs, possums, spider and howler monkeys (Flores 1993).

Previous studies

Investigations on the stand dynamics of secondary forests (which include *V. ferruginea* populations) have been undertaken in two long term permanent plots in Costa Rica (Finegan and Delgado 2000; Finegan and Camacho 1999). These show that secondary forest dominated by *V. ferruginea* tends to have a greater abundance of wind dispersed species than similar forest dominated by *Cordia alliodora* and occurs on slopes with soils of high exchangeable acidity. Investigating site productivity of *V. ferruginea* in secondary forest, Herrera *et al.* (1999) found that the following soil variables were correlated with dominant height (the study indicator of productivity): clay (%), organic matter (%), P, Fe (all negatively correlated), and Cu (positively correlated).

After hurricane Joan, which struck the Nicaraguan coast in 1988, there was a study of the regeneration of the rainforest (Boucher *et al.* 1994, Boucher and Mallona 1997). The large 'gap' created was recolonised predominantly by species present in the forest prior to the hurricane, chiefly as a result of high rates of sprouting. *V. ferruginea* experienced 100% mortality of adult trees, after which the rapid growth of its abundant seedlings quickly began to restore the population density. Moreover, it colonised sites where previously it had been absent. Since *V. ferruginea* was an abundant canopy tree in the forest initially, it cannot be considered as simply a pioneer species, yet its response to massive disturbance was unlike the other primary forest species, and its rapid growth in this area allowed it to dominate the stand (Boucher *et al.* 1994).

Using genetic data from AFLPs (Amplified Fragment Length Polymorphism technique of DNA fingerprinting), Cavers *et al.* (2005) investigated levels and distribution of

genetic variation in Costa Rican populations of *V. ferruginea*. In general, diversity was partitioned within populations (80.51%) and the extent of subdivision ($\Phi = 0.195$) was slightly higher than estimates for fully outcrossed species. It was also found that *V. ferruginea* in highly disturbed plots had reduced genetic diversity. Forest blocks that had recently regenerated from cleared sites, or that had been heavily disturbed, had significantly lower population diversities than old growth or lightly disturbed sites. However, due to lack of complete historical data and comparable plots, a causal relationship between disturbance and loss of diversity could not be inferred (Cavers 2002; Cavers *et al* 2005). Within populations, a low level of cpDNA (chloroplast DNA) diversity was found (two haplotypes), and no within-population diversity, which is comparable to other tropical tree species (Cavers *et al.* 2005). However, the distribution of the two haplotypes was geographically structured with different groups to the east and west of Costa Rican mountain ranges, indicating that topography probably represents a contemporary barrier to seed gene flow (Cavers *et al.* 2005). A mantel test identified a significant isolation-by-distance effect using a pair-wise distance measure that took into account gene dispersal around the mountain range. *V. ferruginea* appears to have an effective enough pollen dispersal mechanism to minimize population differentiation in the absence of physical barriers (Cavers *et al.* 2005).

Lowe *et al.* (in prep) investigated fine scale genetic structure of *V. ferruginea* within primary forest, lightly disturbed and clearfelled plots. They found no apparent genetic structure within *V. ferruginea* populations in both primary and lightly disturbed plots. This absence of structure can be expected for a pioneer gap colonist with the potential for long distance pollen transfer. On lightly disturbed sites, there was an increased incidence of gap occurrence but they found that recolonisation in these sites maintained a comparable genetic diversity and structure to that of undisturbed forest. In contrast, a clearfelled plot site had significantly lower genetic diversity, indicating that a genetic

bottleneck had occurred during colonisation. This site has been recorded as regenerating from a neighbouring block of primary forest left after primary forest rich in *V. ferruginea* was cleared (B. Finegan and C. Navarro, pers. comm.). Of the trees that were sampled from this clearfelled site a high proportion (62%) belonged to one of eight genetically similar groups of spatially clustered individuals. Lowe *et al.* (in prep) suggest that these clusters are probably sibling groups derived from seed deposited by trees that were early colonists of the site. Seeds of such individuals would have germinated and established quickly on a relatively sparsely occupied site. *Vochysia ferruginea* does not have a significant seed bank and so survival of seed from pre-clearance individuals can be ruled out (Guariguata 2000). Also evident at the site were a number of genetically very similar individuals that probably represent the progeny of a single tree. Thus, it appears that there are some individuals that are very efficient at dispersing progeny over long distances. One possibility is a large individual on the edge of the primary block. The reduction in genetic diversity at this site is probably due to the limited number of contributing mother trees. Pollen-mediated gene flow from the neighbouring primary block may introduce variation into successive generations (Lowe *et al.* in prep).

These earlier studies offer good background information on the expected extent of gene flow (by both pollen and seed) in natural populations of *V. ferruginea*, which have been used to plan the sampling design for the current study. In addition, where possible, field sites were chosen where existing ecological or genetic information was available.

2.1.2 Study sites and sampling

Two Costa Rican forest sites were selected for study: Tirimbina and Ladrillera. At each site two *V. ferruginea* populations were sampled, one from primary forest and one from neighbouring secondary forest. An ideal population was defined as 100 trees in a

continuous forest block approximating, where possible, a square plot. In each population adult trees over 10 cm DBH were exhaustively sampled. Although there was no prior evidence that each collection of sampled trees was an interbreeding group, each sample set is defined as a population throughout this study.

At Tirimbina and Ladrillera regenerating seedlings were sampled from plantation and pasture respectively. At Ladrillera 100 seedlings in a continuous population were sampled but in Tirimbina, due to the very high number of seedlings, a random sample was taken (Table 2.1).

Each adult tree was mapped by measuring distance and angle from GPS points. DBH was measured and a sample of cambium was taken by punching two 1 cm^2 plugs from the bark and cutting off a thin slice of internal cambial tissue before replacing the plug (Colpaert *et al.* 2005). For seedlings in the regeneration plots, leaves were removed. Seedlings were mapped as described above and the height of each seedling was measured and recorded. Both leaf and cambium tissues were thoroughly dried in silica gel prior to freezing

Seeds were collected in 2003 from four populations: from primary and secondary forest blocks at each of Tirimbina and Ladrillera. In each population 20 adults were randomly selected and 50 seeds collected from different seedpods. Seeds were germinated and grown with the aim of sampling twenty saplings per mother tree. However, severe weather conditions destroyed most of the germinated seedlings from this collection, and so a further collection was carried out in 2004. The second fruiting season of *V. ferruginea* resulted in lower seed set so it was not possible to sample from all 20 target adults in each population. Actual numbers of seeds collected are shown in Table 2.1. As part of the sample at the Tirimbina site, seeds were taken also from the two remnant

trees in the regeneration plot. Seeds were retained from this second collection for direct DNA extraction.

Tirimbina, Costa Rica

Tirimbina Rain Forest, near La Virgen, Sarapiqui Canton, Heredia Province, Costa Rica (10° 34'N, 84° 06'W) is tropical wet forest with mean annual precipitation of 3864 mm and mean annual temperature of 24.5 °C (Finegan and Camacho 1999). It occurs in a mosaic of cultivated land (pepper, cocoa, rubber and ornamental plants), logged mature forest and secondary forest situated on the foothills of the Central Volcanic mountain range of the country at 1660-2200 m asl (Finegan and Camacho 1999).

The study area consists mainly of primary forest surrounded by secondary forest. Within the large primary forest block, is a square plot of mapped forest for which there is long term ecological data from annual surveys by the Centro Agronómico Tropical de Investigación y Enseñanza (CATIE). Within the secondary forest there are two main mapped plots, the first is named Botarrama after the common name for *V. ferruginea* and is dominated by the species; the second is called the Manu plot and has slightly different forest vegetation with sparse incidence of *V. ferruginea*. The nearest continuous primary forest to the mapped secondary plots is the mapped primary forest plot. The primary population used is a continuous block of forest containing 100 adult trees within the mapped area and the secondary population is a continuous block of forest also containing 100 adult trees from within the Botarrama plot.

There are also fragments of primary forest on hilltops within the area of secondary forest. Forest in the surrounding valleys was cut during road building and then the primary forest was clear felled around the roads for timber. Finally, these cleared areas were used for growing small plantations of commercial crops (V. Herrera pers. comm.).

Consequently the hilltop primary fragments have persisted throughout the period of human impact and potentially contain candidates for seeding the mapped secondary plots. The fragments are small: in the first fragment there were only six living mature *V. ferruginea,* in the second hilltop fragment there were four and in the third fragment, two. Within one of the fragments there were also the remains of two very large *V. ferruginea* trees. The secondary forest surrounding the primary forest fragments is a different type of vegetation to that of the mapped secondary. It is a thick, weedy scrub containing some non-native crop species, has few tree species, and very few *V. ferruginea*.

In the Botarrama plot there was previously a large remnant tree in the area which is now secondary forest (B. Finegan pers comm.) so there is a potential that a significant proportion of the current secondary forest was seeded by this single remnant individual. Adjacent to the secondary forest is an area of land that has been cleared and used as a plantation for palms. The plantation is usually cut every year but some areas have not been re-cut after the initial clearance. In the areas that are routinely cut there is very abundant but patchy regeneration of *V. ferruginea,* except where there is dense growth of grass species. In those areas that have not been re-cut *V. ferruginea* is present but has largely been outcompeted by woody weed species. Figure 2.4 shows the locations of the primary forest plot, primary fragments and the secondary plot at the Tirimbina site. See Table 2.1 for the areas of each sample site and the density of *V. ferruginea.*

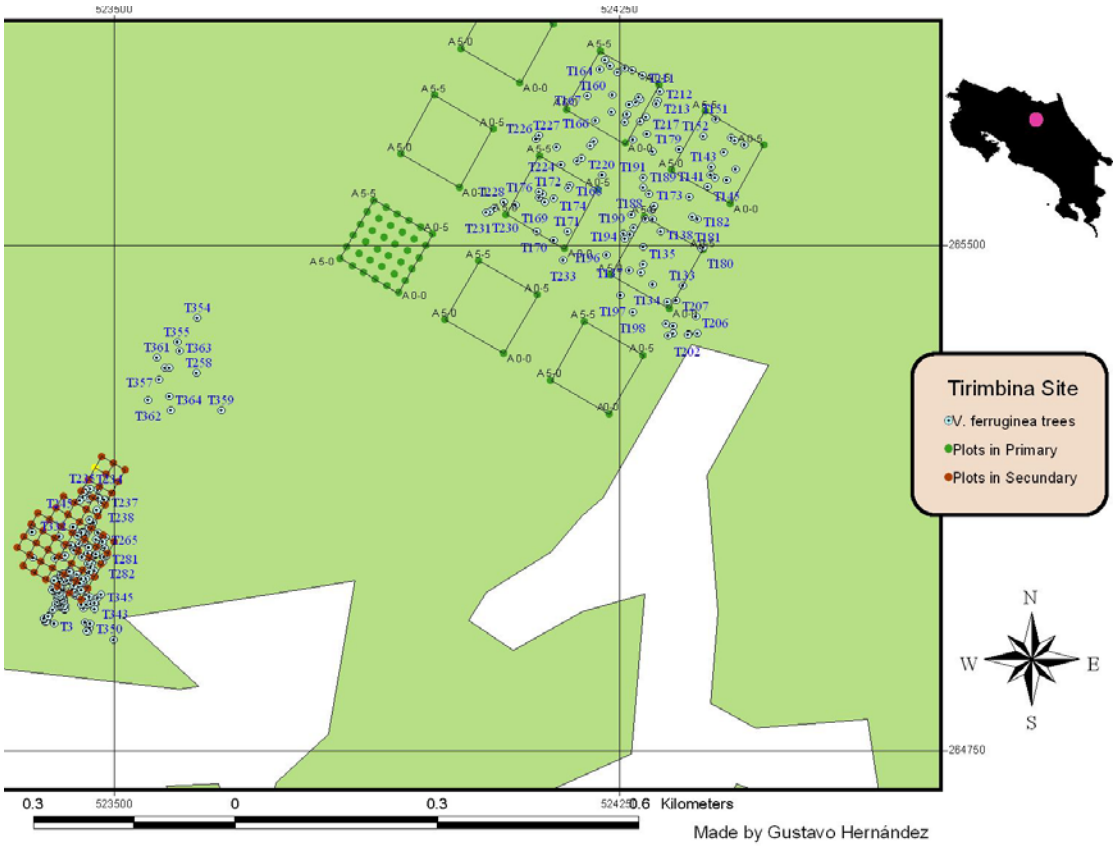


Figure 2.4: Tirimbina site. Grids show plots from previous ecological surveys. The primary forest plot is in the upper right of the map using four of the previously mapped primary forest plots, the secondary forest plot is in the bottom left of the map and the primary fragments are those individuals between the two plots. Inset shows location of site in Costa Rica.

At Tirimbina, during September 2002, 100 adults were sampled from a block of primary forest within the mapped plot. In secondary forest 100 adults were sampled from the mapped plot with a further 15 trees sampled from an unmapped area to extend the plot to the edge of the plantation. In addition, a further 3 remnant trees were sampled from plantation neighbouring the secondary forest and 12 adults were sampled from 3 neighbouring fragments of primary forest (Table 2.2).

La Ladrillera S.A., Costa Rica

La Ladrillera S.A. is a privately owned farmland near Sarapiqui Canton, Heredia Province, Costa Rica (10° 34'N, 84° 06'W), and is tropical wet forest with mean annual precipitation of 3864 mm and mean annual temperature of 24.5 °C (Finegan and Camacho 1999). The site consists of secondary forest with a large primary forest fragment (Figure 2.6). The *V. ferruginea* in the primary forest has been mapped as part of a forest inventory and there is also ecological data from the primary fragment (B. Finegan, CATIE). There is a low density of *V. ferruginea* in the primary fragment with a total of approximately 130 trees: all trees can be considered as 'natural' and are not an artefact of logging. However, within the primary forest there is an area of forest that has experienced a large degree of extraction and consequently has a greater density of *V. ferruginea*.

The primary fragment to the southwest is surrounded by *V. ferruginea* dominated secondary forest (where the *V. ferruginea* is approximately 25 years old) and this extends into pastureland (secondary forest was not been included in the forest inventory). In amongst the secondary forest are primary forest fragments that are either a result of protected zones around watercourses or have been left where the forest is on steep slopes.

The boundary of secondary forest with pastureland is very convoluted as a consequence of protection zones and restricted grazing on steep slopes. The majority of the pastureland is regularly cut to prevent the regeneration of tree species. However, where there is a sufficient degree of slope, cutting is less frequent and there is the potential for *V. ferruginea* to regenerate. In one area where high-density *V. ferruginea* dominated secondary forest is adjacent to sloped pastureland, there is a high incidence of regeneration. These areas of sloping pastureland also had a number of remnant *V.*

ferruginea trees, although the majority have been cut for timber. Figure 2.5 shows remnant trees at the secondary forest boundary with pasture.

Figure 2.5: *Vochysia ferruginea* trees at the boundary between secondary forest and pasture at the Ladrillera site.

The primary forest population sampled consisted of all *V. ferruginea* trees within the forest fragment. Two secondary forest populations were sampled, secondary A was a continuous block of 100 adult trees in secondary forest adjacent to the primary fragment and secondary B was a continuous block of 100 adult trees in secondary forest furthest from primary forest and adjacent to the pastureland. See Table 2.1 for the areas of each sample site and the density of *V. ferruginea.*

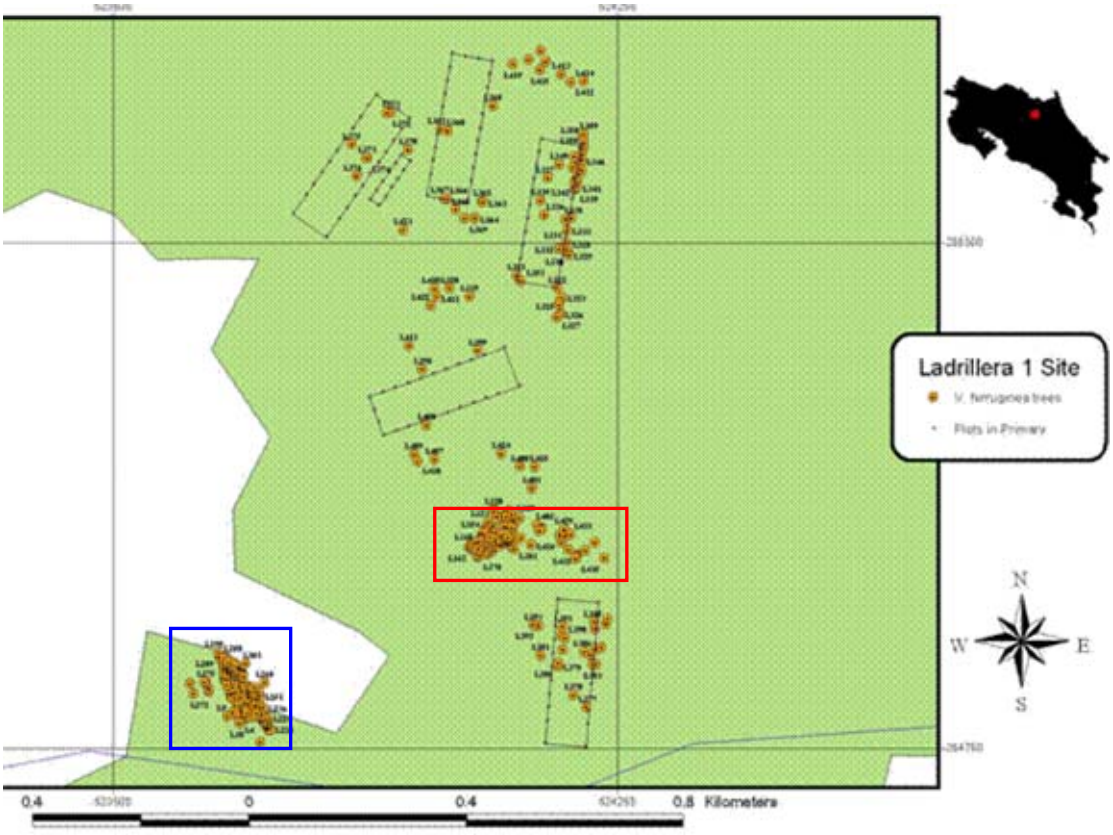


Figure 2.6: The Ladrillera site. Black squares show plots from previous ecological surveys. The secondary A plot is those trees within the red square and the secondary B plot is those within the blue square. Insert shows location of site in Costa Rica.

Table 2.1: Summary of sites showing area and density of *V. ferruginea*

At the Ladrillera site secondary A is near to primary forest and secondary B is further away

Population		Area (hectares)	Density (/ha)
Tirimbina	Primary	15.57	6.4
	Secondary	2.83	40.6
	Seedlings	0.66	1008
Ladrillera	Primary	56.74	2.5
	Secondary A	0.47	241.5
	Secondary B	2.43	224.4
	Seedlings	0.45	223.1

At Ladrillera, during April 2003, 140 trees were sampled from primary forest, 113 in neighbouring secondary forest and 105 from secondary forest adjacent to the area of *V. ferruginea* regeneration in pastureland, with an additional sample of the six remnant trees in this regeneration plot (see Table 2.2).

Table 2.2: Summary of populations showing numbers of adults, seedlings and seeds sampled.

* Numbers of seeds collected are given as (number of seeds) x (number of mother trees)

** At the Ladrillera site secondary A is near to primary forest and secondary B is further away, progeny are taken from secondary B and seedlings are also adjacent to secondary B

	Primary forest		Secondary forest		Regeneration
	Mature trees	Seeds*	Mature trees	Seeds*	
Tirimbina	100 +12 Plot and fragments	20 x 12 240	100 + 15	20 x 15 300	130 seedlings 2 remnants
Ladrillera	140	20 x 16 320	113 secondary A** 105 secondary B**	20 x 19 380 (B**)	100 seedlings 6 remnants

2.2 Molecular analysis

All molecular analyses were performed using a suite of microsatellite markers within nuclear genomic DNA. A set of primers had been previously developed for *V. ferruginea* (Lowe *et al.* 2002) and populations were screened for variation at five of those loci. The allelic state of microsatellites was scored following amplification of the microsatellite region using polymerase chain reaction (PCR) analysis. PCR products were visualised on high resolution polyacrylamide gels and fragment sizes were calculated by comparison with fragment size standards. The number of repeats and hence variation in repeat number between alleles was determined.

2.2.1 Isolation of microsatellite loci and primer selection

Microsatellite markers are found in most genomes and are composed of tandemly repeated 2-, 4-, or 6-base repeat units that can be identified by screening genomic libraries with probes made up of tandemly repeated oligonucleotides (e.g. Schlotterer and Pemberton 1994, Hamilton 1999). Microsatellite probes can be produced from a sample of genomic DNA by creating a recombinant DNA library (Ausubel *et al.* 1987, Sambrook *et al.* 1989, Watson *et al.* 1992). DNA fragments are generated by digesting genomic DNA using restriction enzymes. The fragments are then inserted into the cloning site of vector DNA, restricted with the same enzymes. Screening these fragments allows identification of specific regions of interest, in this case for those regions containing microsatellites.

To amplify a microsatellite region by PCR, it is necessary to develop primers that have a sequence complementary to the areas flanking the microsatellite. For some species, details on microsatellite sequences are available in the public realm, e.g. GenBank. However, for species for which there is no previous data, isolation of microsatellite loci is required. Lowe *et al* (2002) generated a genomic library of *V. ferruginea* enriched for the repeat motifs AC/TG, AG/TC, GAC/CTG and GCT/CGA following a similar procedure to that of Edwards *et al.* (1996), modified according to Butcher *et al.* (2000). More than 70% of the cloned inserts were enriched for selected microsatellite repeat motifs. From this sequence data, primers were developed using the sequences flanking the repeat region.

Lowe *et al.* (2002) found that, of the loci tested, ten were polymorphic and amplified a single locus product without smearing, of these, six were sufficiently polymorphic for population studies and four were highly polymorphic (see Table 2.3 for details). The level of polymorphism for these loci was sufficient to provide high resolution for

assessing fine scale genetic structure and high exclusion probabilities for effective paternity analysis of open pollinated progeny.

2.2.2 Extraction of DNA

DNA extraction from tissue samples was achieved using either CTAB extraction (leaf samples) or commercially available kits (cambium and seed samples). In both cases, the procedure began with mechanical pulverization to separate cells and destroy cell membranes and/or cell walls, while leaving the nuclear DNA intact. Tissue was then immersed in a solution containing a detergent that lyses the nuclear membrane, as well as a proteinase that denatures proteins, especially nucleases, but leaves nucleic acids intact. Proteins were separated from nucleic acids by extraction with organic compounds (phenol and chloroform), and the DNA was purified from the reagents in the extraction buffer by alcohol precipitation.

For the seedlings, leaf material was extracted using a mini-prep protocol modified from Harris (1995). A 1 ml volume of pre-warmed extraction buffer (100 mM Tris-HCl, pH 8.0 20 mM EDTA, 1.4 M NaCl, 1 % PVP-40T, 2 % CTAB, 0.2 % β-mercaptoethanol) was added to the powder; the mixture homogenised and incubated at 65 °C for 30 minutes. Samples were cooled and 5 µl RNAse An enzyme (1 mg / ml) was added and allowed to stand at room temperature for 1 hour. Protein and debris was removed by two washes with 500 µl 24:1 chloroform:octan-1-ol and centrifugation at 13000 rpm for 2 minutes. DNA was precipitated by addition of ice-cold isopropanol and centrifugation at 13000 rpm for 5 minutes. The resulting pellet was washed with aqueous ethanol (76 %) then absolute ethanol (96 %). The pellet was air-dried and resuspended in TE buffer (10 mM Tris-HCl, pH 7.6, 1 mM EDTA) and stored at -20 °C.

Difficulties in extracting good quality DNA from cambium tissue and seeds meant that these were re-extracted using commercial extraction kits (DNeasy 96 Plant Kit, UK QIAGEN LTD 2003 – 2005). A 1 cm^2 leaf / cambium disk was ground to a powder by placing samples in tubes and then inserting tungsten beads and breaking up the material in a mechanical shaker, the extraction then proceeded as described in the DNeasy kit protocol.

2.2.3 PCR protocol

The polymerase chain reaction involves replicating target regions of DNA, which are flanked by regions of known sequence (Ehrlich 1989). Synthetic oligonucleotide primers (usually 20–30 bases long) that are complementary to the flanking regions are required for amplification of the target microsatellite region. These are combined with a small sample, usually nanograms, of genomic DNA, plus free deoxynucleotides, a reaction buffer, and *Taq* DNA polymerase. During a series of heating and cooling cycles, the DNA is denatured into single-stranded molecules, the primers anneal to their complementary sequence flanking the target region, and the DNA polymerase replicates the region downstream from the 5' end to the 3' end between each primer. The amount of target DNA doubles with each cycle, until microgram quantities are present.

The general PCR protocol was as follows: 1-5 ng DNA, polymerase buffer (1x volume, Promega), 0.2 mM dNTPs, 1.5 mM $MgCl_2$, 0.2 μM of each primer, 2 % dimethyl sulfoxide (DMSO, Anachem), ½ unit Taq polymerase (Promega), and made up to a final reaction volume of 15 μl using dH_20. Amplification of PCR products used a touchdown protocol on an MJ PTC100 thermocycler, as follows: initial denaturation at 95 ^oC for 3 minutes; a touchdown program of 95 ^oC for 15 seconds, 65 ^oC for 25 seconds, 72 ^oC for 35 seconds for 11 cycles (where the annealing temperature decreased

by 1 ^{o}C per cycle); followed by a further 25 cycles with the annealing temperature at 55 ^{o}C; and a final extension step at 72 ^{o}C for 15 minutes.

2.2.4 Sizing PCR products

There may only be a small number of base number differences between microsatellite alleles, therefore sizing requires precision and high-resolution gels are essential. For each locus the forward primer of the pair was 5' labelled with infrared dye (IRD 700/800, MWG). Labelled PCR products were then separated using a 6% polyacrylamide gel (25 cm, MWG Biotech), and visualized using a LI-COR 4800 IR2 automated genotyper. PCR products were run out alongside a sizing ladder (50-350 bp Sizing Standard, LI-COR Biosciences) and calculation of size was made manually using SAGA software.

All loci showed the presence of stutter bands. These are thought to be the result of DNA slippage during PCR and show up on the gel as bands, with lower intensity than the main product, that differ in length by multiples of repeat units. An additional band above the allele was also present in some loci, resulting from the terminal transference activity of *Taq* polymerase which adds an A to the product. Loci showed consistent stutter patterns so confidence in manual scoring was possible.

Table 2.3: List of SSR loci indicating repeat motif, Genbank accession number, sequences and size of product (bp) found in 10-20 individuals. For screening microsatellite polymorphism, a population close to Volcan, Costa Rica, was used because this population had been shown to be highly polymorphic by a previous AFLP analysis. Table adapted from Lowe *et al.* (2002).

Locus	SSR repeat	Accession	Forward primer	Reverse primer	bp
A1-5	$(GA)_{16}$	AF436086	TCCCCCACTCCTCTTCATATAGT	GAAAATGGAAGCAAATGGAAAAT	137
A1-10	$(GA)_{13}$	AF436082	CCAAAATAAGCACAAAGTAGCTG	GGGTCCACT CTAATGGCT TGTTA	194
A1-15	$(GA)_{12}$	AF436083	ACAGGAGTAGTCACCTTCATCCA	GGACATGCAATACAATCGACACG	92
A1-20	$(GT)_8GAGT(GA)_{12}$	AF436079	TCTTCAGATACATTCCCTCTCCTC	TCTTCAGATACATTCCCTCTCCTC	109
A1-35	$(GA)_{17}$	AF436087	GGTCTACTCATATCATGCCGAAC	ATGAATGGATTTACCGTCTACCC	170

2.3 Data analysis

Microsatellite allelic variation in adult *V. ferruginea* populations was used to assess differences in genetic diversity and structure in primary and secondary forest. Progeny arrays, taken from both primary and secondary forest, were used to assess mating system and estimate levels of gene flow via pollen. Genetic information from microsatellite variation in regenerated seedlings was used to assess differences in diversity and structure between adult and seedling cohorts.

2.3.1 Genetic diversity

Genetic markers are used to study genetic variation within individuals, and within and between populations by determining what alleles are present in populations. The genetic variation revealed by genetic markers can be quantified in a number of ways, e.g. the proportion of polymorphic loci (*P*), the proportion of heterozygotes over all loci (*H*) and allelic richness. The distribution of genetic variation can be documented at different hierarchical levels within species, within individuals, within populations, among populations and among regions.

Genetic diversity within individuals

In diploid organisms, within-individual genetic diversity is an important component of variability where any particular locus may be heterozygous (with two alleles distinct in DNA sequence) or homozygous (identical alleles on homologous chromosomes). Within-individual patterns of genetic diversity are a focus for some aspects of conservation including the study of mating systems, paternity and inbreeding and inbreeding depression. Where there is inbreeding (defined as mating with self or relatives), homozygosity increases genome-wide and may result in inbreeding depression manifested in possible congenital deformities, reduced fecundity, or reduced survivorship at any life stage (e.g. Ralls *et al.* 1988).

Genetic diversity within populations

Measuring the distribution of genetic diversity within a population makes use of the Hardy-Weinberg model which describes the genetic content in diploid populations in terms of allelic frequencies (Crow 1988). The Hardy-Weinberg principle states that single-locus genotypic frequencies, after one generation of random mating, can be represented by a binomial (with two alleles) or multinomial (with multiple alleles) function of the allelic frequencies (Hedrick 2000). In a random mating population, and in the absence of factors that may change allelic frequencies (e.g. selection, mutation, drift and gene flow), the Hardy-Weinberg proportions will not change over time. The Hardy-Weinberg principle allows for the genetic description of a population based on the allele frequencies at a locus. It states that the frequency of *AA* homozygotes is p^2, the frequency of *aa* homozygotes is q^2 and the frequency of *Aa* heterozygotes is $2pq$, where p is the frequency of the dominant allele in the population and q is the frequency of the recessive allele in the population.

Under Hardy-Weinberg assumptions allele and genotype frequencies remain the same across generations and so when the assumptions are met a population is in Hardy-Weinberg equilibrium (HWE). If the genotypic frequencies within a population are changed without the allele frequencies being changed then the population will return to HWE within one generation of random mating. If the genotypic frequencies in a population depart from HWE, this suggests that one or more of the assumptions of HWE have been violated. However, because HWE is robust to some degree of violation of the assumptions, if no departure from HWE is found this does not mean that none of the assumptions have been violated. A population is likely to deviate from HWE when there is non-random mating within the population such as assortative mating or inbreeding. Assortative mating is non-random mating based on a phenotypic trait and

can be positive when mates are more similar to each other than expected or negative when mates are more dissimilar to each other than expected (Conner and Hartl 2004).

Positive assortative mating will increase the level of homozygosity at any loci linked to the phenotypic trait that is affecting that assortative mating, whereas negative assortative mating will lead to an increase in heterozygosity at those loci linked to the phenotypic trait. The effect of assortative mating does not have a genetic effect at other loci and so can be distinguished from the effect of inbreeding, which affects all loci equally.

Inbreeding does not change the allele frequencies but increases the proportion of homozygotes. Therefore, inbreeding reduces one measure of genetic variability (observed heterozygosity) while the other (allelic diversity) remains the same (Keller and Waller 2002). Evidence for inbreeding within populations is often assessed by comparing observed and expected heterozygosities across loci. Such evidence requires using codominant markers where heterozygotes can be distinguished from homozygotes. By comparing the frequency of heterozygotes in the population to that expected under random mating the increase in homozygotes can be used to measure the extent of inbreeding within a population.

Analysis of genetic diversity

For all microsatellite loci gene diversity was calculated using Nei's F statistics, where H_S is a measure of within sample gene diversity and H_T is a measure of overall gene diversity. The dataset was also tested for Hardy-Weinberg equilibrium using the Markov chain method (Lazzeroni and Lange 1997). This test was carried out for each locus and each population individually with 1000 permutations. These analyses were conducted using the program GENEPOP 3.4 (Raymond and Rousset 1995).

The level of genetic variation within populations was quantified by assessing allelic richness and genetic diversity for each of the five microsatellite loci and averaged over all loci. Allelic richness (Rt) is a measure of the number of alleles made independent of sample size by adaptation of a rarefaction index (Hurlbert 1971, Goudet 2001, El Mousadik and Petit 1996). The inbreeding co-efficient (F_{IS}) was estimated over all loci for all populations according to Weir and Cockerham (1984). These analyses were conducted using the program FSTAT 2.9.3 (Goudet 2001).

2.3.2 Genetic structure

Differentiation among populations

Genetic differentiation between populations indicates how the genetic variation within a species is partitioned in a landscape and can therefore provide insights into how natural populations maintain genetic variation. Differentiation among populations may reflect historical impediments to movement (and thus to relatively ancient population subdivisions) or may reflect contemporary patterns of gene flow. When indirect population level estimates are used, such as F statistics, differentiation is at least the product of multiple generations, so that causes of genetic structure are unknown.

Genetic differentiation among populations is commonly quantified by the use of one of several statistics, including Wright's measure of genetic differentiation (F_{ST}) and Nei's equivalent multilocus coefficient of gene variation (G_{ST}). These indices are functions of the partitioning of heterozygosity within and among populations, based on differences in allele frequencies, and indicate whether the majority of genetic variation is distributed among or within populations (Wright 1965, Wright 1969; Nei 1975; Chai 1976; Wright 1978).

Wright's measure of genetic differentiation (Wright 1951, Wright 1965), on which subsequent statistics are based, uses an approach that measures three different F co-efficients to allocate genetic variation to the total population level (T), subdivisions (S) and individuals (I). The relationship between these co-efficients is described below;

$$1 - F_{IT} = (1 - F_{ST})(1 - F_{IS})$$

$$F_{ST} = \frac{F_{IT} - F_{IS}}{1 - F_{IS}}$$

where F_{ST} is a measure of genetic differentiation over subpopulations and is always positive. F_{IT} and F_{IS} are measures of the deviation from HWE within subpopulations (*S*) and the total population (*T*), where negative values indicate an excess of heterozygotes and positive values indicate a deficiency of heterozygotes.

In species with low F_{ST} (approaching 0), the majority of variation is found within populations. Where F_{ST} is high (approaching 1), the majority of variation is found between populations. Statistics that quantify differentiation among populations vary in the assumptions they make, for example, some measures of differentiation (e.g. F_{ST}) assume that natural populations are at equilibrium with respect to loss and gain of genetic diversity via the processes of genetic drift and gene flow respectively.

Spatial autocorrelation within populations

F statistics are a measure of the differentiation between populations but are not informative about the fine scale genetic structuring that can occur within populations. In estimating fine scale genetic structure, spatial autocorrelation methods can be used to assess the degree to which genotypes tend to cluster together (positive spatial autocorrelation) or are over dispersed (negative spatial autocorrelation) (Shimatani and

Takahashi 2003). There are two different types of spatial autocorrelation statistics that can be used to characterise and quantify spatial structure, joint count statistics and Morans I statistics (Epperson 2000; Shimatani and Takahashi 2003). However, all methods of analysing spatial genetic structure compute pairwise relatedness co-efficients between individuals and relate this to the spatial distance separating individuals (Vekemans and Hardy 2004).

In joint count statistics, distance classes are defined then in all distance classes pairs of individuals are classified (forming a test statistic, SND) according to the mean and variance of their genotypes under the null hypothesis that the genotypes are randomly distributed. Then each distance class has a set of SND statistics, which will increase in size and power with greater sample sizes. If a population has high positive SND statistics at a given distance class, then this shows positive spatial autocorrelation; significant negative SND statistics show negative spatial autocorrelation. Typically there is a distance class where the statistics switch from positive to negative and this distance measures the size of the patches (Epperson 2000).

Morans I statistics are calculated for variables of numerical values (Cliff and Ord 1981), typically by computing allele frequencies among individuals within mapped quadrats. Morans I statistic is the sum of the products of the mean adjusted gene frequencies in pairs of quadrats separated by distances in distance class *k*, normalised by dividing by a measure of the spatial variance in gene frequency. Thus, correlation is measured as a function of distance (Epperson 2000).

Software packages that use spatial autocorrelation statistics to examine genetic structuring in populations, and are appropriate for the data generated in this study, include SGS (Degen 2001) and SpaGeDi (Hardy and Vekemans 2002). All populations

were analysed for fine scale structuring using the program SpaGeDi as this program allows manual manipulation of the spatial distance between each distance class, allowing investigation at a greater detail at smaller distance scales. SpaGeDi estimates genetic distances between populations or relatedness coefficients between individuals using data from co-dominant markers. It computes various statistics describing genetic relatedness or differentiation between individuals or populations by pairwise comparisons and tests their significance by appropriate numerical resampling (Hardy and Vekemans 2002).

2.3.3 Mating system

For organisms capable of both self-fertilization and outbreeding with other individuals (referred to as mixed mating), an important descriptor of the mating system is the relative proportion of selfed versus outcrossed progeny. For mixed mating species, comparing multilocus genotypes of a female and her offspring for the same loci can help to provide estimates of self-fertilization (and across many female-offspring arrays can establish a mean selfing rate). Analysis of mating system is usually an estimate of the frequencies of selfing and outcrossing, but can also include other categories, such as biparental inbreeding (matings between related individuals, e.g. Ritland 1984; Brown 1990).

Estimates of selfing frequency can also be determined at a number of developmental stages (e.g. seeds, seedlings, adults), as fitness effects on selfed offspring may lead to a decrease in selfed individuals between cohorts up to the adult stages. Although often provided as a characteristic of a species, mating system is particular to individuals. This means that ideally a large number of individuals and progeny should be sampled before drawing conclusions. Also, the mating system may vary across a species range and with

environmental conditions. This means that it may be inappropriate to extrapolate from a study of a single population to all populations across a species range.

Analysis of mating system

A number of mixed mating models (e.g. Brown 1990, Ritland 1990) allow for the estimation of selfing frequencies within populations. Ritland's MLTR program (Ritland 1990) generates a number of mating system parameter estimates and uses maximum likelihood analysis to provide minimal variance estimates of these parameters. This program uses a mixed mating model that assumes that a portion of the matings represents selfing events, and that the remaining portion of matings is due to random non-self mating.

The MLTR program estimates the multilocus outcrossing rate (*tm*) and the average single-locus estimate (*ts*). Using these estimates it can also provide an estimate of biparental inbreeding, the probability of mating among relatives (*tm* - *ts*), as true selfing will lead to multilocus similarities between individuals whereas matings between related individuals will lead to an increase in similarity at individual loci. MLTR also provides an estimate of correlated paternity (the likelihood that two randomly selected offspring from a single mother share the same father) (*rp*), which can be translated into an estimate of effective number of pollen donors (*Nep* = 1/*rp*).

Inbreeding effects on fitness

The effects of inbreeding (selfing and biparental inbreeding) are examined in relation to a surrogate measure of fitness. Each individual from the mapped plots (Tirimbina primary and secondary forest, Ladrillera primary forest), for which there were data for 3 or more loci, was given a heterozygosity score (if alleles at one locus were the same it was given a score of -1, if alleles at one locus were different, a score of +1). Thus,

heterozygosity at all loci maximised the score and homozygosity minimised it. The heterozygosity scores were then compared to growth rate, estimated from long term ecological surveys at both sites (B. Finegan, CATIE).

2.3.4 Paternity analysis

Paternity analysis is the determination of the parentage of an offspring, made when maternity is known and paternity is decided from candidate male pollen donors. The aim of paternity analysis is to complete a perfect assignment, that is, every sampled offspring in the population is assigned its true father. The best current methods use cumulative information from multiple polymorphic microsatellite loci to best exclude all non-fathers from assignment. This is not always possible so statistical methods have been developed to cope with multiple potential fathers. These methods can be categorised into three basic approaches; exclusion, categorical allocation and fractional allocation.

Statistical methods of paternity analysis: Exclusion

Exclusion methods of parentage analysis (e.g. Ellstrand 1984), attempt to eliminate all but the true parents from all candidate individuals using incompatibilities between parent and offspring to reject particular parent offspring hypotheses (Jones and Arden 2003). Three alternative conclusions are possible after simple exclusion analysis for paternity analysis. The first outcome is the desired one where only one male cannot be excluded. However, there may be several males that cannot be excluded and potentially all males are excluded.

For exclusion methods to be successful all non-parents need to be eliminated and this is only possible with a large number of highly polymorphic loci, particularly if there are a large number of candidate parents. One further problem of strict exclusion methods is

that if there is an error in determining an individual's genotype, either by mis-scoring or the presence of null alleles, then the true parent will also be excluded. Exclusion methods have been successfully employed (e.g. Ellstrand 1984; Ellstrand and Marshall 1985; Adams *et al.* 1992) yet due to mis-assignment problems, most parentage analysis relies on genetic likelihood methods, either categorical or fractional allocation.

Statistical methods of paternity analysis: Categorical allocation

Categorical allocation uses likelihood-based approaches to select the most likely parent from a pool of non-excluded parents (e.g. Meagher and Thomson 1987; Sanchristobal and Chevalet 1997; Morgan 1998). This method involves calculating a logarithm of the likelihood ratio (LOD score) by determining the likelihood of an individual (or pair of individuals) being the parent (or parents) of a given offspring divided by the likelihood of these individuals being unrelated. After evaluation of all genetically possible parents, offspring are assigned to the parent (or parental pair) with the highest LOD score. When all parent–offspring relationships show zero likelihood, offspring are unassigned. Parentage remains ambiguous when multiple parent–offspring relationships obtain the highest nonzero likelihood. Contrary to strict exclusion methods, likelihood-based allocation methods usually allow for some degree of transmission error due to genotype misreading or mutation. Using a maximum likelihood method also means that the addition of an extra locus compensates for larger increases in the number of parents, so that with a sufficient number of polymorphic loci it is possible to achieve high allocation success even when there is relatively high number of possible parents.

Statistical methods of paternity analysis: Fractional allocation

The fractional allocation (e.g. Devlin *et al.* 1988) method assigns some fraction of each offspring to all non-excluded candidate parents. The portion of an offspring allocated to a particular candidate parent is proportional to its likelihood of parenting the offspring

compared to all other non-excluded candidate parents. Single parent and parent pair likelihoods are calculated in the same way as in the categorical allocation method.

Limitations of likelihood methods

Likelihood methods do not always lead to assignment for all offspring as some progeny may have an ambiguous likelihood profile (Smouse and Meagher 1994), meaning that the data set is reduced. Another problem associated with this method is that homozygotes will always give a higher likelihood score than a heterozygote at any given locus (as a homozygotic male will carry two copies of the compatible allele) so that all offspring with an allele will be allocated homozygotic parents (Devlin *et al.* 1998, Adams *et al.* 1992). Although a homozygote does have a higher likelihood of transmitting the allele the net effect of maximum likelihood methods is still a bias in favour of homozygotes. This can be particularly problematic when assessing inbreeding depression, since inbred individuals are more homozygous and are therefore favoured by the analysis. A reduction in fertilisation success resulting from inbreeding depression may consequently be obscured. The extent of this bias can be reduced using categorical allocation if the number of markers is increased (Smouse and Meagher 1994). Alternatively Devlin *et al.* (1988) show that if offspring are instead fractionally allocated among all non-excluded parents based on their probability of parentage, the bias associated with the categorical assignments can be circumvented.

A further shortcoming of many parentage models is that they assume a uniform prior probability distribution. That is, they assume that each candidate parent is equally likely to be the true parent of a given offspring regardless of proximity of candidate parents to the offspring or reproductive successes of candidates. This might be the best assumption in the absence of any information about the actual distribution of the prior probabilities (Devlin *et al. 19*88). However, it leads to an underestimation of the true variance in

reproductive success as individuals with high reproductive success are underestimated while those with low reproductive success are overestimated (Adams *et al.* 1992). Incorporating biological data into categorical allocation can increase the accuracy of parentage inference (Adams *et al.* 1992).

Statistical methods of paternity analysis: TwoGener analysis

Direct methods of measuring pollen mediated gene flow that rely on paternity exclusion or assignment, require exhaustive sampling of all potential pollen donors in the population studied. However, if all potential fathers have not been genotyped it is possible to obtain an estimate of pollen flow using TwoGener analyses even without having an estimation of adult density (Austerlitz and Smouse 2001a,b, 2002; Smouse *et al.* 2001). The TwoGener method allows for quantification of the heterogeneity among male gamete pools sampled by maternal trees across the landscape and can be used to estimate mean pollination distance and effective neighbourhood size (Smouse *et al* 2001).

Austerlitz *et al.* (2004) found similarity with the estimation of pollen flow distances from TwoGener and paternity-based methods. Burczyk and Koralewski (2005) found that a neighbourhood model (a representative of paternity-based methods) and TwoGener provided comparable estimates of effective number of pollen parents and comparable estimates of mean pollen dispersal distance although the neighbourhood model provided more variances of the parameter estimates. The TwoGener approach assumes that flowering individuals are homogeneously spread over a two dimensional landscape larger than the scale of pollen dispersal. When the population studied does not approximate this, such as when individuals are clumped, then estimates are less reliable.

Problems associated with paternity analysis

Null alleles

When one allele has not been successfully amplified during PCR whilst the other allele has, then the individual will be incorrectly scored as a homozygote at that locus. If the null allele is inherited by the offspring, it can result in the parent of that offspring being wrongly excluded. In studies where one parent is known, null alleles are easier to detect because they result in significant departure from Hardy–Weinberg equilibrium. In studies in which a known parent is sampled with groups of offspring, null alleles are even easier to detect because they result in incompatibilities between the known parent and offspring that invariably involve homozygous genotypes.

Linkage disequilibrium

The effects of linkage disequilibrium and physical linkage between loci are different in regard to parentage analysis. Using loci that are in linkage disequilibrium decreases the expected probability of exclusion and the accuracy of parentage assignments (Chakraborty and Hedrick 1983). This is because non-random associations between loci reduce the amount of useful genetic variation for discriminating parentage (Devlin *et al.* 1988). Conversely, if physically linked loci are examined, and the linkage phase and recombination rate of the candidate parents are known, the accuracy of parentage assignments can be increased, provided that they do not exhibit severe linkage disequilibrium (Devlin *et al.* 1988; Jones *et al.* 1998c). Most parentage analysis techniques consider the data one locus at a time and combine the information over all loci assuming independent assortment. Therefore, it is a requirement that the data is tested for the presence of linkage disequilibrium prior to analysis.

Mutations and scoring error

Mutations and scoring errors can cause offspring to appear incompatible. Even with microsatellite data, mutation rate is slow so that there are unlikely to be many mutations even in a large data set (Jones and Arden 2003). When offspring have a known parent it is possible to assess the level of mutation or scoring error by checking transmission of alleles from parent to offspring.

Family structure

The presence of family members, other than the parents of offspring, in a pool of candidate parents can present a serious challenge to parentage assessment. If sampled adults are related to each other, or if candidate parents are siblings of the offspring, then a higher number of loci are required to determine parentage.

Cryptic gene flow

When there is no genetically compatible parent the parent is assumed to be an un-sampled adult from outside the sampled population. There is also a possibility of a parent being outside the sampled area but the offspring is assigned to a parent within the area, as it is not incompatible with a non-parent within the data set (Burczyk *et al.* 2004). This undetected gene flow from outside the population is cryptic gene flow and will be relative to the level of polymorphism in the suite of molecular markers used. Devlin and Ellstrand (1992) developed a model for estimating total pollen gene flow into a population that takes into account the contribution of cryptic gene flow (that which would go undetected in a conventional paternity analysis). The primary assumption of this model is that gametes can be treated as being drawn at random from a large source population of known allele frequency. Unbiased estimates of relative male reproductive success can be adjusted for cryptic pollen gene flow (Roeder *et al.*

1989). This approach is useful for generating a pollen dispersion curve as well as for estimating gene movement from outside the un-sampled area.

Techniques used in paternity analysis software

There are a number of computer programs designed for reconstructing parentage. Programs that use an exclusion method include PROBMAX (Danzmann 1997), NEWPAT (Worthington Wilmer *et al.* 1999) and KINSHIP (Goodnight and Queller 1999). Programs that use categorical allocation include CERVUS (Marshall *et al.* 1998), PAPA (Duchesne *et al.*2002), FaMoZ (Gerber *et al.* 2000) and PARENTE (Cercuiel *et al.* 2000). Most computer programs use similar methods for assignment but differ in the way they deal with errors in the data set.

For paternity analysis NEWPAT, CERVUS, FaMoZ, PARENTE, PATRI or KINSHIP can be used for exclusion or allocation and are suitable for the data used in this study. In this study paternity analysis was done using the FaMoZ program (Gerber *et al.* 2003). The dataset had a number of missing values and therefore a likelihood method was required that could provide an assessment of confidence in assignment. As a consequence of missing data, not all potential fathers were genotyped, therefore, data was also analysed using the TwoGener method (Smouse *et al* 2001), which is included in the FaMoz program.

Paternity analysis software: null alleles

The most conservative way to handle a locus with null alleles in parentage analysis is to recode all homozygous genotypes as heterozygotes possessing the detected allele and the null allele, thus preventing exclusion on the basis of homozygous genotypes. However, most programs (including FaMoZ) deal with null alleles by treating them as any other mutation or genotyping error. This approach is problematic, because these

programs allow only a single, experiment-wide mistake rate that is constant across loci. Thus, if one locus has a high frequency null allele, then the program can accommodate it only by assuming that all loci have a high rate of mistakes or mutations, thus diminishing the power of the loci that do not display null alleles.

Paternity analysis software: mutation and scoring error

For those data sets in which mutations and scoring errors appear to be an important concern, algorithms have been developed to overcome such problems (Sancristobal and Chevalet 1997; Marshall *et al.*1998). The details of these algorithms differ among programs. Sancristobal and Chevalet (1997) developed an algorithm that allows mistakes and mutations to be drawn from a specified distribution that can range from a strict stepwise model to an infinite allele model, depending on parameter values specified by the user. FaMoZ uses a simplified form of this method that restricts its use to the infinite alleles model.

Paternity analysis software: confidence assessment

Assessing the statistical confidence in parentage estimates is an important component of parentage analyses (e.g., Pena and Chakraborty 1994; Evett and Weir 1998; Marshall *et al.* 1998). Most programs use a simulation-based assessment of confidence in assignments. FaMoZ simulates data sets and calculates expected distributions of their test statistic, Δ, which is usually the difference in likelihood ratios between the two males most likely to be the father of the offspring. From the distributions of the test statistic in simulations, FaMoZ can determine a critical value that will produce a desired level of confidence in assignments of parentage. This algorithm is particularly important when mutations and mistakes exist in the data set, because under such circumstances all hypothesised parent–offspring pairings possess a finite probability of being true, and a

measure of confidence provides an objective means by which the validity of pairings possessing some apparent genetic incompatibilities can be evaluated.

The simulation-based approach to confidence assessment implemented by FaMoZ has been criticised by Nielsen *et al.* (2001). The major criticism is that the test statistic Δ only uses information from the two most likely males, discarding the information from other compatible males. In addition, the value of Δ is calculated differently when only one male is compatible with the offspring than when two or more are compatible, presumably resulting in two different distributions of Δ. Statistics developed by Neff *et al.* (2001) now allow a calculation of statistical confidence in fractional allocation tests.

Data validation

Prior to analysis all data was tested for linkage disequilibrium using FSTAT 2.9.3, Goudet (2001). The presence of null alleles, large allele drop-out and mis-scoring was assessed using Micro-Checker (Oosterhout *et al.* 2004). Results of these tests can be found in the appendix. Cryptic gene flow was assessed using FaMoZ (Chapter 5).

2.4 Simulation modelling methods

A simulation modelling component was included in this study to investigate processes affecting the long term genetics of a colonising species. To investigate factors of direct interest to sustainable forestry of a secondary forest species it was necessary to develop both species- and scenario-specific models. The models were written in FORTRAN.

Two distinctly different models were used in this study; one that operated on a landscape scale and only considered seed dispersal and one that operated on a population scale and considered both seed and pollen dispersal. The landscape model assessed cpDNA haplotype structure and used data from a case study of the oak

recolonisation of Britain. The population model assessed variation in diploid multilocus genotypes and used data from analysis of *V. ferruginea* regeneration.

Both models simulated a spatial landscape or array consisting of a grid of cells. Each cell was allowed to contain a single individual that, upon reaching maturity, could disperse pollen and potentially seed within a radius around itself dependent on set dispersal kernels (kernels defined in Chapters 6 and 7). When two or more adult trees were within pollination distance they could produce offspring, which could be dispersed around the maternal parent. Offspring were assigned genotypes from its two parents according to Mendelian inheritance. The spatial array could contain one or more species but only the genotypes of the species of interest were recorded.

2.4.1 Statistical analysis of model outputs

In the landscape scale model, populations of twenty trees were sampled for analysis. Diversity was calculated using Nei's haploid diversity statistics for within and total population diversity and differentiation (Nei 1975). For the population scale model, genotypes for all *V. ferruginea* were recorded and these data were analysed for gene diversities and differentiation using both Nei (1987) and Weir and Cockerham (1984) estimators. The estimates were tested using randomisation methods (analysis was implemented using FSTAT 2.9.3, Goudet (2001)).

Chapter 3: Genetic Diversity and Differentiation in *Vochysia ferruginea* Populations in Primary and Secondary Forest

3.1 Introduction

Attempts to characterise genetic diversity and differentiation in tropical trees within primary forest, have been performed on a variety of species and spatial scales (e.g. Dick *et al.* 2003 Collevatti *et al.* 2001; Lowe *et al.* 2003; Lemes *et al.* 2003, see Table 1.1 in Chapter 1 for further examples). Most species have been found to maintain more variation within than among populations (Hamrick *et al* 1992), indicative of the effective means of gene flow between trees (Schierenbeck *et al.* 1997; White *et al.* 1999). Most tropical canopy layer trees occur at a low density in primary forest and so long distance seed and/or pollen dispersal are expected. As a result, they often show little population genetic structure (Dutech *et al.* 2002; Latouche-Hallé *et al.* 2004; Ward *et al* 2005). A high level of gene flow via long distance dispersal of pollen creates greater genetic connectivity between populations and reduces genetic structuring, although other factors may also bring about low levels of population genetic structuring, e.g. random mortality of seedlings (Hamrick *et al.* 1993); inbreeding depression eliminating related individuals (Epperson 1992) and disturbances effecting local density (Knowles *et al.* 1992).

Previously studies that have investigated genetic variation in tropical tree species within primary forest populations have predominantly investigated the genetic consequences of habitat loss and fragmentation, particularly in species of high economic or conservation value (e.g. *Swietenia* sp, White *et al.* 1999; Cespedes *et al.* 2003; Norvick *et al.* 2003; Lowe *et al.* 2003; Lemes *et al.* 2003; Lowe *et al* 2005). Unlike the majority of tropical canopy trees, pioneer species often occur at high density in disturbance gaps, although the gaps themselves may be spatially distant, and so such species have the potential to

show local-scale structuring. This isolation of spatially distant populations therefore leads to genetic differentiation between gap populations However, as pioneer species require long distance dispersal to utilise gaps in the forest, extensive gene flow may also lead to low levels of structuring in these tropical species (McCauley 1991, Alvarez-Buylla and Garay 1994, Slatkin 1985).

Pioneer species in secondary forest may have critically different genetic characteristics to low density canopy trees as the process of colonisation of abandoned habitat might affect the genetic diversity and structure of a species creating long term impacts in long lived tree populations. The few studies of tropical trees in secondary forest that have been undertaken have suggested that there may be a loss of diversity with colonisation (e.g. in *Iriartea deltoidea*, Sezen *et al.* 2005), although this may be mediated by extensive gene flow (e.g. in *Swietenia macrophylla*, Cespedes *et al.* 2003 and *Elaecarpis grandis*, Rosetto *et al.* 2004).

Vochysia ferruginea is a fast growing pioneer that also retains a presence in primary forest as a canopy tree (Flores 1993). Its tolerance to poor soils and aluminium toxicity mean it can colonise the low fertility conditions of abandoned agricultural land where it can form dense monodominant stands (Flores 1993; Herrera and Finegan 1997; Herrera *et al.* 1999). In fact there is evidence that *V. ferruginea* becomes more abundant in the low nutrient/high acidity soils typical of areas of land from which tropical forest has been removed for agriculture than under more natural situations (Herrera and Finegan 1997; Herrera *et al.* 1999). These properties combine to make *V. ferruginea* an excellent species in which to study genetic diversity and differentiation in primary and secondary forests.

The study in this thesis collects genetic information from two sites (Tirimbina, Ladrillera) in Costa Rica where there is both primary forest and an area of secondary forest. At Tirimbina the secondary forest is approximately 30 years old and at Ladrillera it is 25 years old. At both sites the secondary forest boundary at these locations is adjacent to agricultural land that is either recently abandoned or minimally managed with seedling regeneration found as secondary forest encroaches. For both sites levels of genetic diversity were assessed in primary, secondary and seedling populations. These were compared to progeny arrays, seeds taken from mother trees, in both primary and secondary forest. This allowed investigation into the genetic variation in seed, seedling and adult cohorts. Genetic differentiation was also assessed between these populations allowing comparison of gene flow between populations

This study then evaluated the genetic characteristics of *V. ferruginea* in two differing population structures: low density in gaps in primary forest and in high-density, even-aged secondary forest where it is the dominant tree. By assessing genetic diversity and differentiation of adult trees, seedling regeneration and seeds taken from mother trees the population genetic characteristics of different life stages can be compared to give an insight into the colonisation process from a genetic point of view. For this study, we assumed that secondary forest is formed by colonisation into open land by seeds from individuals in primary forest; and subsequent seedling regeneration from secondary forest represents the next generation of colonisation. Comparison of diversity estimates from progeny arrays and seedling regeneration allowed investigation of the impact of (seed-mediated) colonisation processes versus the impact of (pollen-mediated) gene flow. The genetic characteristics of the progeny array represent the potential genetics of the next generation, whereas the seedling populations represent the actual genetic composition of the regenerating population.

3.2 Methods

3.2.1 Study site and populations

Populations from both Tirimbina and Ladrillera were sampled as described in Chapter 2. At each site, duplicated primary forest, secondary forest tree and seedling populations, and primary and secondary forest progeny arrays were sampled for study. The Tirimbina site included a primary fragment population and the Ladrillera site had two secondary forest populations: one adjacent to primary forest and one at the furthest extent of the secondary forest. It is the latter area of secondary forest from which seeds were collected and where the seedling population was situated.

3.2.2 Sampling, DNA isolation and microsatellite analyses

Cambial tissue was taken from adult trees and leaf material was collected from seedlings. Seeds were collected from randomly selected mother trees in both secondary and primary forest and DNA was isolated directly from the seeds. DNA was isolated as described in Chapter 2. Analyses were made using data obtained from screening using 5 microsatellite primers; primer details, PCR and scoring conditions are also indicated in Chapter 2.

3.2.3 Statistical methods

Locus diversity and Hardy-Weinberg equilibrium

Gene diversity was calculated as H_S, a measure of within sample gene diversity, and H_T, a measure of overall gene diversity (Nei 1973). Hardy-Weinberg equilibrium over all samples was tested for using a Markov chain method (Lazzeroni and Lange 1997), where each locus and each population were sampled individually with 1000 permutations, using the program GENEPOP 3.4 (Raymond and Rousett 1995).

3.2.4 Genetic diversity

The level of genetic variation within populations was quantified by assessing allelic richness and genetic diversity for each of the five microsatellite loci and averaged over all loci. Allelic richness (Rt) is a measure of the number of alleles made independent of sample size by adaptation of a rarefaction index (Hurlbert 1971; Goudet 2001; El Mousadik and Petit 1996). Estimates of gene diversity per locus and over all loci use an unbiased estimator (as described in Nei 1987). These analyses were conducted using the program FSTAT 2.9.3 (Goudet 2001) and tested using bootstrap sampling with 1000 simulations.

3.2.5 Deficit of heterozygotes

The inbreeding co-efficient (F_{IS}) was estimated over all loci for all populations using the program FSTAT 2.9.3 (Goudet 2001) and calculated according to Weir and Cockerham (1984).

3.2.6 Genetic differentiation

Genetic differentiation (F_{ST}) between all pairs of populations was estimated following Weir and Cockerham's (1984) multilocus weighted analysis of variance using the program GENEPOP 3.4 (Raymond and Rousett 1995). This analysis was tested using bootstrap sampling with 1000 simulations.

3.3 Results

3.3.1 Locus diversity and Hardy-Weinberg equilibrium

All five microsatellite loci were highly polymorphic, the mean number of alleles per locus was 16.2, ranging from 13 to 24 (see Table 3.1). The observed proportion of heterozygosity ranged from 0.568 to 0.8, loci with higher levels of polymorphism did not necessarily show higher levels of heterozygosity. Gene diversities within samples were similar to the overall gene diversity. Diversity within samples ranged from 0.583

to 0.79 and overall diversity ranged from 0.719 to 0.838 (Table 3.1). Loci differed greatly in tests for Hardy-Weinberg equilibrium with *P* ranging from 0.000 to 0.998 and all loci could be considered as in HWE except A1-10. Further results on data validation (see appendix) showed that there was no evidence to suggest the presence of null alleles, large allele drop-out or mis-scoring due to stutter bands. There was some evidence for linkage disequilibrium; however, this was most likely an artefact of the data due to similarities in allele dropout in loci and not because of linked loci.

Table 3.1: Characterisation of the five microsatellite loci employed using data from all populations.
N, number of individuals; *A*, total number of alleles; H_O, observed heterozygosity; H_S, within sample gene diversity; H_T, overall gene diversity; *P*, departure from Hardy-Weinberg among all populations using the Markov chain method, standard error in brackets.

Locus	N	A	H_O	H_S	H_T	P
A1-5	884	23	0.758	0.79	0.838	0.000 (0.000)
A1-10	1455	13	0.8	0.746	0.838	0.998 (0.002)
A1-15	1495	14	0.649	0.585	0.751	0.023 (0.005)
A1-20	1236	16	0.568	0.673	0.719	0.000 (0.000)
A1-35	1010	15	0.772	0.717	0.784	0.013 (0.004)
Mean over all loci		16.2	0.709	0.702	0.786	0.004

Within populations, the mean sample size per locus (i.e. the number of individuals scored at any locus) was significantly smaller than the population size, in most cases mean sample size was halved (see Table 3.2). This was a consequence of difficulties in successful amplification of loci, particularly A1-5 (see Table 3.1). Table 3.2 also shows mean number of alleles per locus for each population; however, due to the differing sample sizes allelic richness is a better measure of the number of alleles per population

(Table 3.3). Populations differed significantly regarding deviation from Hardy-Weinberg equilibrium, with *P* values ranging from 0 to 1.

Table 3.2: Sample size, number of alleles and departure from Hardy-Weinberg in all populations averaged over loci.

N, population size; *S*, mean sample size per locus; *A*, mean number of alleles per locus; *P*, departure from Hardy-Weinberg, standard error in brackets.

Site	Population	N	S	A	P
Tirimbina	Progeny primary	15 x 20 = 300	163.8	10.2	0.485 (0.031)
	Progeny secondary	12 x 20 = 240	180.8	9.6	
	Seedlings	132	20	5.8	0.204 (0.012)
	Adult secondary	120	78	8.6	0.980 (0.006)
	Adult primary	100	60.4	9	0.000 (0.000)
	Primary fragments	12	7.6	4.8	0.118 (0.006)
Ladrillera	Progeny primary	20 x 20 = 400	289	10.2	0.000 (0.000)
	Progeny secondary	20 x 20 = 400	223.2	9.8	
	Seedlings	100	49.2	3.4	1.000 (0.000)
	Secondary A	130	49.6	6.8	0.001 (0.000)
	Secondary B	136	37.4	6.4	0.647 (0.019)
	Primary	140	57	7.6	0.081 (0.011)

3.3.2 Genetic diversity

The two sites showed similar patterns in the changes of allelic richness between different populations but varied in the way diversity changed across populations. At Ladrillera, diversity and allelic richness decreased moving from primary to secondary A, then to secondary B, then to seedlings. In Tirimbina, across the same gradient, diversity stayed constant, whilst allelic richness alone dropped from primary forest

through to secondary forest then to seedlings See Table 3.3 for values of genetic diversity (Nei 1987) and allelic richness.

Table 3.3: Genetic diversity, allelic richness and F_{IS} in adult, seedling and progeny arrays, where at the Ladrillera site, secondary A is adjacent to primary and secondary B is adjacent to seedlings.

N, population size; H_E, average genetic diversity over all loci according to Nei (1987); R_T, allelic richness; F_{IS}, deficit of heterozygosity. ns = non significant, * = significant at the 95% confidence level, ** = significant at the 99% confidence level.

Site	Population		N	H_E	R_T	F_{IS}
Tirimbina	Progeny	In secondary	300	0.78	9.90	-0.01^{ns}
		In primary	240	0.75	9.08	-0.007^{ns}
	Seedlings		132	0.74	4.76	0.11^{*}
	Adult secondary		120	0.72	7.95	-0.136^{**}
	Adult primary		100	0.75	8.86	-0.026^{ns}
	Primary fragments		12	0.77	1.82	0.016^{ns}
Ladrillera	Progeny	In secondary	400	0.70	10.07	0.067^{**}
		In primary	400	0.67	9.44	0.042^{**}
	Seedlings		100	0.47	3.48	-0.082^{ns}
	Secondary A		130	0.74	8.20	0.175^{**}
	Secondary B		136	0.60	4.50	-0.157^{**}
	Primary		140	0.68	5.71	-0.069^{*}

At Tirimbina the greatest diversity was found in the primary fragments (H_E =0.77), although this small population showed very low allelic richness (1.82). The primary forest population at Tirimbina had both greater diversity (H_E =0.74) and allelic richness (8.86) than secondary forest (0.72 and 7.95 respectively), although the reduction is small. There was no significant difference in diversity between both the secondary and primary forest populations and the seedling population (seedling population diversity = 0.74); however, there were significantly fewer alleles (allelic richness = 4.76). There

was a high level of diversity and allelic richness found in seeds taken from mother trees in both the primary and secondary populations, with the number of alleles found in the progeny arrays exceeding that found in the adult populations. Seeds taken from trees in primary forest had a diversity of 0.75 and an allelic richness of 9.08, seeds from secondary forest had a diversity of 0.78 and an allelic richness of 9.90. Therefore, at the Tirimbina site genetic diversity was similar across all populations but allelic richness was greatest in primary forest and seeds, decreased in secondary forest and decreased further in seedlings.

A more complex pattern is observed in Ladrillera populations, here primary forest showed a lower level of diversity and allelic richness to that found in Tirimbina (diversity = 0.68 and allelic richness = 5.71). The two secondary forest populations showed very different patterns of diversity. The secondary forest population adjacent to primary forest had greater diversity than the primary forest population (0.74) and showed higher allelic richness (8.2). The other secondary forest population, further from the primary forest, showed a lower diversity and allelic richness than both the secondary population A and that of the primary forest (diversity = 0.6 and allelic richness = 4.5). The seedling population at Ladrillera, adjacent to this lower diversity secondary forest, exhibited a further decrease in diversity (0.47) and again a loss of alleles (allelic richness = 3.48). As in the Tirimbina populations, diversity and allelic richness was high in the progeny arrays with seeds collected from both primary and secondary forest showing a diversity equivalent to that of primary forest (0.7 in progeny from secondary forest and 0.67 in progeny from primary). Progeny from both secondary and primary forest populations also had a greater number of alleles present than in the adult populations. At Ladrillera, seeds taken from trees in primary forest had a diversity of 0.67 and an allelic richness of 9.44, secondary forest had a diversity of 0.7 and an allelic richness of 10.07.

3.3.3 Deficit of heterozygotes (F_{IS})

Most populations had low or negative levels of F_{IS} (see Table 3.3). There was an excess of homozygotes found in the seedling population at Tirimbina (F_{IS} = 0.11) but this was not seen in seedlings at Ladrillera. There was an excess of heterozygotes found in the secondary forest population in Tirimbina (F_{IS} = -0.136) and also the low diversity secondary population at Ladrillera (F_{IS} = -0.157). However, the largest deficit of heterozygotes was also found in secondary forest, at Ladrillera in the forest adjacent to the primary forest block (F_{IS} = 0.175).

3.3.4 Genetic differentiation

The genetic differentiation statistics are shown in Table 3.4. The Tirimbina populations all showed some differentiation from each other, F_{ST} ranged from 0.025 to 0.106. The Ladrillera populations showed a similar level of differentiation between population samples, F_{ST} ranged from 0.009 to 0.112. There were high levels of differentiation between the populations at Tirimbina and Ladrillera, particularly between the seedling populations at both sites where F_{ST} was 0.342. However, the lowest differentiation between a population at Ladrillera and one at Tirimbina was lower (F_{ST} was 0.069 between Ladrillera primary and the progeny from Tirimbina primary) than the greatest within site differentiations at both sites (F_{ST} = 0.106 and 0.112).

At Tirimbina, the primary fragments showed the lowest differentiation from the primary forest (F_{ST} = 0.028) and the greatest differentiation from the secondary forest (F_{ST} = 0.086). The fragments showed a similar level of differentiation from the seedlings (F_{ST} = 0.043) and the progeny, with the fragments less differentiated from progeny in primary forest (F_{ST} = 0.031) than in secondary forest (F_{ST} = 0.069). The primary forest was most genetically similar to the primary fragments, was not largely differentiated from the secondary forest (F_{ST} = 0.036) but showed some differentiation to the

seedlings (F_{ST} = 0.068) and the progeny, where it was differentiated from progeny taken from the secondary forest (F_{ST} = 0.059) and showed some differentiation from progeny taken from within this population (F_{ST} = 0.034).

The secondary forest population was most differentiated from the primary fragments (F_{ST} = 0.086) and was most genetically similar to the primary forest population (F_{ST} = 0.036). The secondary forest showed a similar level of differentiation to the progeny (F_{ST} = 0.04) and seedlings (F_{ST} = 0.041) as did the primary forest (F_{ST} = 0.043 and 0.042). However, the secondary forest showed an equal level of differentiation from the progeny of secondary (F_{ST} = 0.047) and the progeny of primary forest (F_{ST} = 0.048).

Table 3.4: Genetic differentiation (F_{ST}) for each pair of populations (using GENEPOP where F_{ST} is estimated as in Weir and Cockerham 1984). At the Ladrillera site secondary A is near to primary forest and secondary B is further away, progeny are taken from secondary 2 and seedlings are also adjacent to secondary B.

Site		Tirimbina						Ladrillera				
	Population	Progeny	Progeny	Seedlings	Secondary	Primary	Fragments	Progeny	Progeny	Seedlings	SecondaryA	SecondaryB
Tirimbina	Progeny 2º	0.025										
	Seedlings	0.087	0.106									
	Secondary	0.048	0.047	0.041								
	Primary	0.034	0.059	0.068	0.036							
	Fragments	0.031	0.069	0.043	0.086	0.028						
Ladrillera	Progeny 1º	0.079	0.127	0.178	0.155	0.135	0.116					
	Progeny 2º	0.102	0.142	0.214	0.185	0.17	0.154	0.03				
	Seedlings	0.178	0.221	0.342	0.259	0.262	0.28	0.054	0.089			
	Secondary	0.077	0.106	0.149	0.133	0.136	0.122	0.026	0.056	0.084		
	Secondary	0.125	0.161	0.25	0.210	0.175	0.207	0.046	0.01	0.112	0.067	
	Primary	0.069	0.117	0.162	0.138	0.128	0.141	0.009	0.058	0.069	0.028	0.089

The seedling population at Tirimbina showed the highest differentiation to the progeny arrays, F_{ST} was 0.087 for progeny from primary forest and 0.106 for progeny from secondary forest, these two populations were the most differentiated at the Tirimbina site. The progeny array from within the primary forest had similar levels of differentiation from the other populations, with the progeny slightly more differentiated from secondary forest populations and more similar to primary forest and the primary forest fragments, F_{ST} was 0.048 with secondary forest, 0.035 with primary forest and 0.031 with the primary forest fragments. The progeny array from secondary forest was similarly differentiated from the adult population from which it was collected (F_{ST} = 0.047) but showed greater differentiation from the primary forest populations (F_{ST} = 0.059 in primary forest and 0.069 in the fragments).

At Ladrillera, primary forest trees shared the most genetic similarity with the progeny from primary forest (F_{ST} = 0.009), and were the two most similar population samples at this site. The primary forest was moderately differentiated from the secondary B population, furthest away from the primary forest (F_{ST} = 0.089), and from the progeny (F_{ST} = 0.058) and seedling populations (F_{ST} = 0.069) adjacent to that secondary population. The primary population was more similar genetically to the secondary A population, adjacent to the primary forest (F_{ST} = 0.028) and the progeny from primary forest (F_{ST} = 0.026) than to the distant secondary B forest (F_{ST} = 0.067). The secondary A population adjacent to primary forest was also slightly differentiated from the progeny (F_{ST} = 0.056) and seedlings (F_{ST} = 0.084) and from the other, distant secondary forest population (secondary B).

The secondary B population showed a high differentiation to both the primary and the other secondary population. It also showed a greater similarity to progeny within the

primary forest (F_{ST} = 0.046) than it did to the primary forest population itself (F_{ST} = 0.889). This secondary forest population was most similar to the progeny array taken from adults within it (F_{ST} = 0.01) but surprisingly, was most differentiated from the seedling population also presumed to have originated from it (F_{ST} = 0.112).

Seeds from the primary forest progeny array were very similar genetically to the primary forest tree population, but showed some differentiation from the secondary forest adjacent to the primary forest and greater differentiation from the adult, seedling and progeny populations of the further secondary forest. The progeny array in the secondary forest, although similar to the secondary forest population it derived from, showed the highest differentiation (F_{ST} = 0.089) from the seedlings from that site. The progeny array from this secondary forest was equally differentiated from the other secondary (F_{ST} = 0.056) and the primary forest (F_{ST} = 0.058) populations. The seedling population at this site had an unusual pattern of differentiation and was differentiated from all other populations. It was found to be most similar to the progeny in primary forest (F_{ST} = 0.054) and most differentiated from the secondary forest adjacent to the seedlings (F_{ST} = 0.112). Moderate levels of differentiation were also found between the seedlings and primary forest (F_{ST} = 0.069), the other secondary forest population (F_{ST} = 0.084) and progeny within the secondary forest adjacent to the seedlings (F_{ST} = 0.089).

3.4 Discussion

3.4.1 Mature primary and secondary forest diversity, inbreeding and differentiation

Gene diversity in *V. ferruginea* (H_T = 0.79 and A = 16.2) is within the range found in other tropical pioneer tree species. Average gene diversity for microsatellite loci in the pioneer *Elaeocarpus grandis* was found to be 0.62 (Rossetto *et al.* 2004), in *Swietenia*

macrophylla it was found to be 0.66 (H_E) and 13.0 (*A*) in Central American populations (Norvick *et al.* 2003) and in Brazilian populations 0.78 (H_E) and 18.4 (*A*) (Lemes *et al.* 2003).

Genetic diversity (according to Nei 1987) in *V. ferruginea* adult populations ranged from 0.6 to 0.74 in secondary forest and 0.68 to 0.75 in primary forest. This is a similar level of diversity to that found in other tropical tree species (see Table 1.1 in Chapter 1). There is a considerable range of diversity found in other pioneer species, H_E was between 0.42 to 0.64 in different populations of *Elaeocarpus grandis* (Rossetto *et al.* 2004), 0.47 in some populations of *Swietenia macrophylla* (Céspedes *et al.* 2003), 0.81 in other populations of *S. macrophylla* (Lemes *et al.* 2003) and 0.37 in populations of *Antirhea borbonica* (Litrico *et al.* 2005). The range of diversity found within pioneer species may be a consequence of the different densities and colonisation processes that many pioneer species experience. The range of differentiation found between *V. ferruginea* populations at both sites was between F_{ST} = 0.009 and 0.112. These values are within the range of differentiation found in other tropical tree populations, where low differentiation was found for *Dinizia excelsa* (F_{ST} = 0.00167; Dick *et al.* 2003) and high levels of differentiation have been found for *Caryocar brasiliense* (F_{ST} = 0.29; Collevatti *et al.* 2001). In all, levels of differentiation between *V. ferruginea* populations are similar to other pioneer species (e.g. *Swietenia macrophylla* where F_{ST} was 0.005 to 0.238 in populations (Novick *et al.* 2003). Levels of F_{IS} in *V. ferruginea* populations ranged from
-0.157 to 0.175. Many factors can affect levels of heterozygosity so that tropical trees can have an excess or deficit of heterozygotes with F_{IS} values found ranging from -0.22 in *Carapa guianensis* (Dayanandan *et al.* 1999) to 0.5 in *Swartzia simplex* (Hamrick *et al.* 1993).

At the Tirimbina site, the primary fragments consisted of three small isolated hilltop populations each containing a very small number of adult trees. That these trees are remnant from primary forest and are geographically widespread accounts for the high diversity for this group of adults whilst the small population number accounts for the low allelic richness. The primary forest population in Tirimbina had a high diversity, a high allelic richness and no evidence of inbreeding. Although the adjacent secondary forest has lower diversity and allelic richness, it is not significantly different from the primary population. The secondary forest at Tirimbina is over 30 years old and consists of a dense *V. ferruginea* dominated stand. That there is genetic diversity within this population on a scale comparable to primary forest, suggests either that the process of colonisation by *V. ferruginea* has not, in this instance, significantly decreased diversity, or that diversity has increased in secondary forest since colonisation.

A newly colonised population could maintain levels of diversity if, for example, there is long distance gene flow so that many trees contribute to the colonisation process, either directly through seed dispersal or secondarily by pollen dispersal. Dispersal, particularly via pollen from the neighbouring primary forest may also have increased diversity in the 30 years since the area was recolonised by *V. ferruginea*.

Other studies of regeneration by pioneer species have found similar resistance to loss of diversity with colonisation. Investigating the response of regeneration by *Swietenia macrophylla* into pastures, Céspedes *et al.* (2003) found that newly colonised populations had relatively high levels of diversity, presumably in response to high levels of gene flow also found in this species (White *et al.* 1999; Céspedes *et al.*

2003). However, some studies have found significant decreases in diversity due to bottlenecking in founder populations. For example, Sezen *et al.* (2005) found that after colonisation of secondary forest by *Iriartea deltoidea* the founder populations showed significantly lower diversity.

At Ladrillera, one of the secondary forests (secondary A) had a higher diversity and higher allelic richness than the primary forest population. This was unexpected, principally, as the sampled population of primary forest consists of every tree within the large primary forest fragment assumed to be the source of the secondary forest. It also showed higher diversity and allelic richness than the secondary forest at Tirimbina even though it was younger (25 years). It is likely that gene flow from unsampled remnant trees of protected primary forest, bordering the two rivers running through the secondary forest, are responsible for the unusually high diversity of this secondary block. This population also showed a deficit of heterozygosity (F_{IS} = 0.175), which suggests that the population may have been founded by different sources and this deficit is more likely due to this Wahlund effect (Crow and Kimura 1970) than inbreeding. High F_{IS} was also found in populations of the highly outcrossing *Antirhea borbonica*, where Litrico *et al.* (2005) also suggested that this was a result of structure and not inbreeding. The remnant trees that may have contributed to this population consist of a small number of individuals, so very small primary fragments may measurably affect the genetics of secondary forest populations and increase genetic diversity.

The other secondary forest population at Ladrillera (secondary B) also showed deviation from Hardy-Weinberg but with an excess of heterozygotes (-0.157), as did the secondary forest at Tirimbina (-0.136). Both of these populations were some

distance from the most likely primary forest source population. The secondary forest population at Tirimbina was potentially founded by one very large remnant tree that was left in the area after clearing and has since died (B. Finegan pers. comm.). *Vochysia ferruginea* appears to be predominantly outcrossing even as isolated remnants. Therefore, it is possible that at Tirimbina this secondary forest may have been founded by a single, or a few remnant trees. Remnant trees in a cleared area would most likely be pollinated via long distance pollination, thus bringing together spatially and potentially genetically differentiated populations. This population was most highly differentiated from the primary fragment populations and least differentiated from the primary forest that extends beyond the fragment populations further from the secondary forest. This also suggests that the fragments are not the source population and instead the source was the primary forest or remnant, shade trees that were left in the area after the primary forest was logged and have since died.

Long distance gene flow between genetically differentiated individuals may also explain the genetic characteristics of secondary forest at the Ladrillera site as this too has a landscape history of a small number of remnant trees left after clearing with the pollen source from primary forest or other remnant trees some distance away. The genetic character of these secondary forest populations suggests that small differences in the colonisation process, such as distance from source population and presence of remnant trees, may have important consequences for levels of genetic diversity, allelic richness and heterozygosity in the resultant secondary forest.

Genetic differences between populations may not always be due to the difference in population structure and density between primary and secondary forest. At Ladrillera the secondary population adjacent to the primary forest is more genetically similar to

its neighbouring primary forest than more distant secondary forest. Genetic differentiation in populations of the well studied mahogany species show that while low levels of structuring are common in this species, biological and physical factors can both significantly increase (e.g. White *et al.* 1999) or restrict (e.g. Lemes *et al.* 2003) gene flow resulting in a range of structuring found in this species. As with *V. ferruginea,* the tropical pioneer *Cecropia obtusifolia* also occurs within gaps in primary forest and also as monodominant stands of secondary forest. Investigating the genetics of this species in both primary and secondary forest, Alvarez-Buylla and Garay (1994) found significant fine scale structure in newly founded secondary forest that decreased as clusters of related individuals suffered progressive mortality aided by extensive pollen dispersal distances. Sezen *et al.* (2005) also found clustering of related individuals of *Iriartea deltoidea* in newly founded secondary forest with patches of similar genotypes greater than in primary forest.

3.4.2 Seedling and outcrossed progeny diversity, inbreeding and differentiation

Vochysia ferruginea seedling populations at the two sites also showed very different levels of diversity with seedlings in plantation (Tirimbina) showing much higher diversity than those in the pastureland (Ladrillera), even though both show roughly equivalent loss of allelic richness. The lower allelic richness in seedling populations is an expected result of the loss of alleles during colonisation by a minority of seed donors. There was low allelic richness in seedling populations at Tirimbina and Ladrillera. However, the allelic richness, particularly at Ladrillera, is not that greatly diminished from neighbouring secondary forest. The seedling populations were sampled differently, with a subsample taken of 100 seedlings in the abundant Tirimbina population, compared to an exhaustive sampling of 100 seedlings at Ladrillera. However, these populations were sampled over an equivalent spatial area, so the sampling strategy is unlikely to be the cause of the differing levels of diversity.

The pastureland at Ladrillera was cut yearly and most seedlings were of similar height (ranging from 5 to 390 cm in height with the majority under 100 cm), yet in the abandoned plantation at Tirimbina the area was less regularly cut so more than one generation of seedlings may have been present with seedling height ranging from 12 to 640 cm. Therefore, it would appear that a loss of allelic richness is consistent with colonisation, but the loss of diversity in the Ladrillera seedling population is a more likely a consequence of it consisting of seedlings from a single year. The loss of diversity may be partially alleviated by colonisation from successive years' reproductive output.

Colonisation by *V. ferruginea* appears to lead to a loss of alleles and potentially a loss of genetic diversity, however, high levels of gene flow, particularly during colonisation, may counter that effect with time. The differences in diversity found in *V. ferruginea* seedling populations with different age structures suggest that regeneration by successive generations may counter diversity loss in founding populations over a few years through a replacement diversity effect from post colonisation seed and pollen mediated gene flow.

Céspedes *et al.* (2003) found high allelic richness in populations of *Swietenia macrophylla* colonising abandoned pasture, however, they found very few uncommon or rare alleles in the regenerated populations. For *V. ferruginea* populations at both Tirimbina and Ladrillera there were losses in allelic richness in regenerated populations, as expected during colonisation from a limited source (bottlenecking). However, at both sites, seed set in open-pollinated progeny arrays was uniformly more diverse and of higher allelic richness than the trees regardless of habitat, so there appears to be effective sampling of all alleles occurring at reproduction. Therefore,

the observed diversity changes must be due either to post-recruitment processes or to a severely limited source (possibly due to edge effects that would not affect the adults within populations from which the seed was collected).

The progeny arrays at Tirimbina were differentiated from the adults from which they were derived. In both primary and secondary forest, the progeny population is more genetically similar to other populations. However, this is not the case at the Ladrillera site where the progeny arrays from secondary and primary forest are both differentiated to some extent from the other populations except those adult populations from which they were derived (0.01 between secondary forest and progeny, 0.009 between primary forest and progeny). The differentiation between progeny and adults at Tirimbina could be caused by gene flow via pollen from outside the sampled populations. This could be from individuals in unsampled primary forest that adjoins both primary forest and secondary forest. At Ladrillera, the primary forest sample contained all individuals of the primary forest block present at that site, therefore there is less potential for unsampled individuals to be contributing pollen to the progeny. That the progeny from secondary forest at Ladrillera are also genetically similar to the adult population may be a consequence of the greater distance of this sampled block from primary forest. Therefore, gene flow via pollen into the secondary forest block is more likely to be from the adjacent secondary forest, which is potentially more genetically similar.

At Ladrillera, the progeny from secondary forest and the seedling population were highly differentiated from each other, and at this site the highest differentiation was between these populations (0.106). Both the seedling population and the progeny in the secondary forest are presumably descended from the same parents and yet they are

not genetically similar. However, secondary adults were equally differentiated from seedlings and secondary progeny (approximately 0.04) so the strong differentiation between seedlings and secondary progeny could be due to differences in the proportions of paternal gene flow contribution rather than to originating from a different source. A common source for seedlings and secondary progeny would also explain why seedlings are less differentiated from secondary adults than those in primary forest.

An unusual relationship between seedlings and the assumed secondary forest source population was also found at the Ladrillera site. As in Tirimbina the seedlings and the secondary forest progeny array were differentiated (0.089) even though it would be assumed that these two populations share the same parents. However, here the secondary forest showed high differentiation from the seedlings (0.112) but little differentiation from the progeny array taken from mothers within the secondary forest (0.01). This further suggests either different sources for progeny and seedlings at this site or else a dominance of the small subset of remnant trees (6 remnant trees compared to 100 in secondary forest) as a source of the seedling population. The process of colonisation from secondary forest into adjacent vacant habitat, at Tirimbina and Ladrillera, may have led to different genetic consequences than the recruitment into gaps within a population, or the progeny and seedling populations do not share the same ancestry. In both sites, all remnant trees found in the area of regeneration are included as individuals of the secondary population. Therefore, it is likely that the cause of these high levels of differentiation is the processes of colonisation or very high dominance of remnant trees as seed and pollen donors.

3.4.3 Conclusions

Pioneer trees persist in primary forest through the exploitation of gaps in the canopy. This can lead to a population structure of small populations spatially isolated from one another. Therefore, to successfully colonise gaps pioneer species must be able to disperse seeds relatively long distances. Also, maintenance of a genetically viable population requires mechanisms whereby these isolated gap populations can exchange genetic material via long distance pollen flow.

Unlike gaps, clearance of large tracts of primary forest allows highly competitive pioneer species to colonise in very large numbers. As the interest in secondary forest as a resource increases it is becoming important to determine the genetic effects of colonisation by pioneer species that exploit primary forest gaps and whether abundant tree species of secondary forest are sufficiently genetically diverse to be a sustainable forest resource.

Low diversity in seedling populations in newly colonised abandoned pasture (Ladrillera) was potentially the result of a small number of maternal donors contributing to the founding population. In the mature secondary forest, levels of diversity and allelic richness are comparable to those in primary forest, both in Tirimbina where the secondary forest is 30 years old and Ladrillera where the forest is only 25 years old. In both populations diversity and allelic richness were at their highest in the seed collections: for both secondary and primary forest collections. This shows that there is a high level of gene flow to adults in primary forest gaps and secondary forest. It seems likely that it is this high level of gene flow that has countered the founder effect of reduced allelic richness in the initial colonising population, so that after thirty years secondary forests show similar levels of diversity

and allelic richness to those in primary forest. This is most clearly seen in progeny taken from adults in the least diverse secondary forest plot at Ladrillera. Within this population adult diversity and allelic richness are low (0.6, 4.5 respectively), yet progeny from these adults not only show an increase in diversity (0.7) they also show the highest allelic richness (10.1). Presence of neighbouring primary forest has, over a relatively short time, proved effective in increasing the allelic richness in what is a densely populated secondary forest population.

Comparisons of the diversity of seeds collected from secondary forest and primary forest showed that adults in secondary forest had the potential to produce offspring as genetically diverse and with the same extent of allelic richness as those in primary forest. As long as there is the opportunity of gene flow from neighbouring primary fragments secondary forests appear quick to recover from the loss of allelic richness associated with colonisation. However, the high differentiation between seedlings regenerating from secondary forest and progeny taken from adults within the secondary forest showed that, whilst individuals within secondary forest can capture the diversity of the population through pollen flow, actual regeneration into abandoned land may be genetically different as few individuals may founder that population. An excess and deficit of heterozygosity was found in sampled secondary forests, showing that small differences in the number of founding individuals and the genetic distance between them can effect the genetic composition of secondary forests. Therefore, the effects of colonisation on genetic parameters are sensitive to small changes in the population dynamics of a species.

Colonisation from a small number of individuals can result in a loss of diversity within regenerated populations, the extent of which may be unpredictable. However,

unlike predictions from some previous studies (e.g. Sezen *et al.* 2005) these results suggest that in the presence of gene flow from primary forest, genetic diversity in secondary forest tree populations can be restored to levels equivalent to that found in primary forest over relatively short time periods. Although initial colonisation results in a loss of alleles and a dense, closely-related secondary forest, high levels of pollen dispersal lead to effective gene flow from primary fragments and remnant trees into the secondary forest population. This implies that if fragments of primary forest can be maintained as a pollen source, genetic diversity in areas of secondary forest can be restored to levels found in primary forest, and this recovery is possible over relatively short time periods in a pioneer tree species such as *V. ferruginea*.

Chapter 4: Mating System, Fine Scale Genetic Structure and the Effects of Inbreeding on Fitness in Primary and Secondary Forest Populations of *Vochysia ferruginea*

4.1 Introduction

Plant species show a large variety of breeding systems, but surveys of tropical species have noted them to be predominantly outcrossing (Doligez and Joly 1997; Nason and Hamrick 1997; Loveless *et al.* 1998). However, they are not always self incompatible and mating systems of tropical trees appear to be flexible for many species, with the level of outcrossing changing from one flowering event to another (Degen *et al.* 2004). For example, *Cavanillesia platanifolia* (Murawaski and Hamrick 1992), a neotropical canopy tree, and *Shorea trapezifolia* (Murawaski *et al.* 1994) demonstrate a variation in mating system, between individuals and seasons, from completely outcrossing to completely selfing. Flexible mating systems are an advantage in low density tropical populations, this is also true for pioneer species and has been found in the Brazilian pioneer *Senna multijuga* (Ribeiro and Lovato 2004).

Mating system can have a major impact on the genetic structuring and level of genetic variation within populations. Typically, predominantly outcrossing species partition most variation within populations and show little differentiation between populations, whilst predominantly selfing species show higher levels of variation between populations (Brown 1989; Ritland 1989). Biparental inbreeding (matings between related individuals) may reflect spatial structuring within populations (Ennos and Clegg 1982), which can result in genetic similarity among neighbouring individuals, particularly where there is limited seed or pollen dispersal (Vekemans and Hardy 2004). Such characteristics will generate spatial genetic structuring within populations

of a species and clustering of related individuals will lead to increased matings between them.

By sampling individuals of a species at high and low densities, Lebois *et al.* (2004) showed that the occurrence of spatial genetic structure is largely dependent on the density of individuals. Spatial genetic structure is also linked to life stage, particularly in tree species. In many tree species, seedlings are found in much higher densities than adults resulting in a typical L-shaped age distribution with strong spatial genetic structuring often found in seedling populations as these may form clusters of related individuals around the mother tree: this structure decreases in adults due to self thinning (Hamrick *et al.* 1993; Chung *et al.* 2003).

The mating system and spatial genetic structure of pioneer canopy trees in primary forest may differ from that of individuals of the same species in disturbed forest or in dense secondary forest. For example, Murawski *et al.* (1994) noted that for *Shorea* species sampled from logged and unlogged stands, self-fertilization and kin mating were high in the disturbed forest. Murawski and Hamrick (1991, 1992b) also detected a density effect for three Bombacaceae species in Panama (*Cavanillesia platanifolia, Ceiba pentandra*, and *Quararibea asterolepis*). Lim *et al.* (2002) found that outcrossing rates in *Dryobalanops aromatica* were greatest in primary forest (0.86 – 0.92) and lower in logged (0.77-0.79) or artificial forest (0.55 – 0.67) with high correlated mating and biparental mating also detected. However, in their study of *Dryobalanops aromatica*, Kitamura *et al.* (1994) found no significant difference for outcrossing rates in secondary (harvested 20 yr ago, density of flowering trees 6.7 individuals/ha; outcrossing (tm) = 0.79) and primary forest stands (density 14.0 individuals/ha; outcrossing (tm) = 0.86). For *Carapa procera*, (Doligez and Joly

1997) the outcrossing rate and level of biparental inbreeding was lower in logged plots compared to unlogged plots. This study also showed that the differences in the amount of selfing between logged and unlogged sites could not be accounted for by density alone, since densities of *C. procera* were not always significantly different in logged vs. unlogged sites. Therefore, density alone may not be an adequate predictor of changes in outcrossing rates between primary forest and disturbed or secondary forest.

Conservation of intraspecific diversity within tropical forest requires information on genetic structure and reproductive biology of species. However, to devise effective management strategies requires some understanding of species responses to change; therefore, these characteristics must be assessed in different landscape contexts. Responses to logging, for example, should be evaluated by calculating estimates of genetic structure and mating system in both primary and secondary forest (Doligez and Joly 1997). This study investigates the mating system and fine scale genetic structure of *Vochysia ferruginea*, in both low density populations in primary forest and high density populations in secondary forest. *Vochysia ferruginea* is a gap pioneer and will colonise land after disturbance (e.g. anthropogenic or natural, such as hurricanes, Boucher *et al.* 1997, Boucher and Mallona 1997). The colonisation process leads to high density, even aged stands and as a result the density of *V. ferruginea* can be widely different in primary and secondary forest populations.

The process of colonisation involving a restricted number of individuals that establish a highly dense population can lead to aggregation of genotypes and potentially increase matings between related individuals. As genetic structuring increases, so too can the incidence of biparental inbreeding which may in turn, increase the potential

for inbreeding depression within populations. This chapter looks at local genetic structuring within primary, secondary and seedling populations at two sites. The data are used to investigate the process of post-disturbance colonisation on genetic structure within populations and contrast it with an assumed natural state of gap colonisation. In addition, two temporal scales are compared, by analysis of both a seedling population and adult populations after 20 or 35 years of colonisation. The study also investigates the effect of biparental inbreeding on fitness, using long term growth rate data (Finegan and Camacho 1999). The growth data was compared with levels of heterozygosity to determine whether there is an observable fitness effect on growth.

4.2 Methods

4.2.1 Species

Vochysia ferruginea is hermaphrodite and flowers between April and June (Arnáez and Moreira 1995) with an occasional second flowering event in September and October. Stands of *V. ferruginea* may flower monosynchronously and flowers are visited by hummingbirds and insect pollinators, mainly large bees and butterflies (Oliveira and Gibbs 1994). In a study of six *Vochysia* sp. pollination was mediated by a great variety of pollinators: primarily bees, with occasional visits by moths, butterflies and hummingbirds (Flores 1993).

In controlled pollinations Bawa *et al.* (1985b) reported *V. ferruginea* as largely self-compatible and self-pollinated flowers had only marginally reduced fruit set relative to cross-pollinated flowers (proportion of flowers that set fruit was 0.16 and 0.17 respectively, Bawa *et al.*1985a). However, flowering in forest populations tends to be synchronous and, as pollinators include large bees capable of dispersing pollen over

long distances, there is a high potential for outcrossing (Oliveira and Gibbs 1994). The fruit of *V. ferruginea* are mainly dispersed by wind and occasionally birds (Lowe *et al.* 2002).

4.2.2 Study site and populations

The mating system analysis described in this chapter used the progeny arrays from adjacent secondary and primary forest at both Tirimbina and Ladrillera and included a limited progeny array from two remnant pasture trees situated in abandoned plantation adjacent to the Tirimbina site. 20 mother trees were randomly selected in each population and 20 seeds per mother tree were collected and analysed. Analyses of fine scale genetic structure used primary forest, secondary forest and seedling populations from Tirimbina and Ladrillera. See Chapter 2, Figures 2.4 and 2.6 for detail of each site.

4.2.3 Sampling, DNA isolation and microsatellite analyses

Cambial tissue was taken from adult trees and leaf material was collected from seedlings. Seeds were collected from randomly selected mother trees in both secondary and primary forest and DNA was isolated directly from the seeds. DNA was isolated as described in Chapter 2. Analyses were made using data obtained from screening using 5 microsatellite primers; primer details, PCR and scoring conditions are also indicated in Chapter 2.

4.2.4 Statistical methods

Mating system

Plant mating systems have traditionally been described by a mixed-mating model, which assumes that plants self-fertilize at rate *s,* and the remaining fraction of progeny is derived from outcrosses. Mating system parameters were estimated using maximum likelihood procedures based on the mixed mating model proposed by Ritland and Jain

(1981) and using the multilocus mating system analysis program MLTR (Ritland 1996, Ritland 2002). The MLTR program estimates the fraction of progeny derived from outcrosses from multilocus information (tm) and as the average of all the single locus estimates (ts). The parameters estimated from the progeny array data were the multilocus outcrossing rate (tm), the average singlelocus outcrossing rate (ts), the biparental inbreeding rate (tm-ts), the multilocus correlation of paternity (rp(m)), the single locus correlation of paternity (rp(s)) and the correlation of selfing (rs). These parameters are described in more detail below.

Biparental inbreeding and selfing

Biparental inbreeding or mating between relatives causes apparent selfing or increased homozygosity, relative to random mating. When biparental inbreeding occurs, the multilocus pattern of outcrossing in a progeny array is different from that expected under pure selfing or outcrossing: effective selfing occurs at some loci and apparent outcrossing occurs at others (Ritland 1984). When true selfing is present the difference between multilocus and single locus estimates of outcrossing (tm – ts), can be used to characterize the level of biparental inbreeding. This difference is positive because single locus estimates include all apparent selfing by biparental inbreeding, whereas the multilocus estimate will more often eliminate biparental inbreeding as any difference in one locus will not be categorized as selfed. However, this difference is always an underestimation unless a large number of loci can be used (Ritland 2002). The MTLR program estimates the amount of true selfing by determining the among locus correlation of selfing (rs). This directly approximates the fraction of selfing due to uniparental selfing and is not biased by the number of loci used (Ritland 2002).

Multilocus correlated matings model

In addition to the outcrossing rate, a mating system can be characterised by the degree to which siblings share the same male parent. Correlation can be the outcome of either pollination syndrome or population structure. To characterize this sharing of parentage, Ritland (1988) introduced the 'correlated-matings model', which describes the extent to which siblings share the same father, and the extent of variation among arrays for selfing rate. For any pair of progeny in a mixed mating population, the model allows estimation of the correlation of outcrossed paternity, which is the probability that the two are outcrossed full siblings rather than half-siblings. Multilocus probabilities allow for both greater statistical power and the comparison of single and multilocus estimates of correlated matings, which provides a measure of population substructure (Ritland 2002).

Correlation of paternity

The multilocus correlation of paternity (rp(m)) is the proportion of full sibs among outcrossed progeny. In a randomly distributed population, the greater the correlation of paternity, the lower the number of fathers contributing to the progeny array, however, if population substructure is present this relationship is more complex. By comparing multilocus (rp(m)) and single locus (rp(s)) estimates of the correlation of paternity, it is possible to estimate the degree of population substructuring. If the single locus estimate of correlation of paternity is greater than the multilocus estimate (rp(s)) > rp(m)) then there is an effect of population substructure on male similarity: i.e. although different males are contributing to the progeny array (low rp(m)) they are closely related (high rp(s)).

Spatial genetic structure

Fine scale spatial genetic structure within mapped populations was analysed by relating pairwise genetic distance to geographical distance. The relatedness between a pair of individuals is a continuous quantity defined in terms of probabilities of identity-by-descent and can be estimated using the following relationship co-efficient. A pair of individuals (x and y) in such a population can be genetically correlated in two ways: a single gene at a locus in x is identical by descent with one in y, or both genes in x are identical by descent with those in y. If the probabilities of occurrences of the first and second events are denoted as ϕ and Δ, respectively, the relatedness, r, between individuals x and y can be expressed as

$$r = \phi / 2 + \Delta$$

(Lynch and Ritland 1999). In an outbreeding population, ϕ, Δ, and r are 1, 0, and 0.5, respectively, for parents and offspring; 0.5, 0.25, and 0.5, respectively, for full-sibs; and 0.25, 0, and 0.125, respectively, for half-sibs (Wang 2002). Genetic structure was assessed using a multi-locus pairwise relationship co-efficient (*r*), which jointly estimates two-gene and four-gene coefficients of relatedness between individuals from an outbreeding population (Wang 2002, Hardy and Vekemans 2002). Computation of the statistics was done using SpaGeDi (Hardy and Vekemans 2002). In each site individual locations (populations) were randomised among all individuals with 1000 permutations.

Level of heterozygosity versus growth rate

Level of heterozygosity in individual trees was compared to ecological data on growth rate within the Tirimbina primary and secondary forest plots using data from a long term study undertaken by CATIE (Centro Agronómico Tropical de Investigación y Enseñanza, B. Finegan, unpublished data). Tree condition and mortality did not vary in the sample sets used and so were ignored. Individuals were given a score based on

the level of heterozygosity found at 3 to 5 microsatellite loci. At each locus a score of +1 was given if the alleles differed and -1 if the alleles were the same. Thus, an individual that was homozygote at each of 5 loci would score -5, whereas if it was heterozygous at each locus it would score +5. The sample was split into two groups; trees with high heterozygosity (a score of +3 or greater) and those with low heterozygosity (a score of +2 or less). Sample means of these two groups were tested for a significant difference in growth rate. The difference in the means was calculated as the high heterozygosity mean minus the low heterozygosity mean. The standard error was

$$\sqrt{\sigma_H^2 / n_H + \sigma_L^2 / n_L}$$

where n_H and n_L were the number in the high and low heterozygosity samples respectively.

For secondary forest, growth rate was compared over eight consecutive years. For primary forest, the most complete data was between years 1998 and 2003 so total growth over this period was compared. Other ecological factors that could affect growth rate (light incidence and initial DBH) were tested for their effect on growth rate and whether there was any difference in the effect of these ecological factors between the two groups of high and low heterozygosity trees.

4.3 Results

4.3.1 Mating system

Results from all populations showed that *V. ferruginea* was largely outcrossing in both primary and secondary forest populations and also when found as isolated remnant trees in abandoned plantation (outcrossing rate ranged from 0.85 to 1, see

Table 4.1). The approximate measure of uniparental selfing (the correlation of selfing among loci, rs) showed that there was very little selfing in most of the populations, (rs ranges from 0.001 to 0.217) although estimates suggested that some of the seeds collected from mother trees in the Ladrillera primary forest (rs = 0.217) and the two remnant trees in pasture at Tirimbina (rs = 0.104) may have been a result of true uniparental selfing.

In Tirimbina primary forest, seeds were completely outcrossed (tm = 1) with very little biparental inbreeding found (tm – ts = 0.036). The multilocus correlation of paternity was estimated as 0.281, with an estimated 2.74 pollen donors contributing to the progeny array. In the secondary forest, progeny were predominantly outcrossed, with an increase in the amount of biparental inbreeding (tm – ts = 0.101) and a higher proportion of siblings sharing the same father (rp(m) = 0.425) compared to primary forest. The progeny from the two remnant pasture trees in abandoned plantation at the Tirimbina site showed complete outcrossing (tm = 1.143) but with a high degree of biparental inbreeding (tm – ts = 0.196). An outcrossing rate above 1 may be a consequence of small population size leading to high standard deviations. Most siblings did not share the same father (rp (m) = 0.206) and these trees had a higher number of pollen donors contributing to the progeny arrays (an estimated 4.854 donors) than all other populations.

Table 4.1: Estimates of mating system parameters using the MTLR program. Standard deviation in brackets.

		Tirimbina primary	Tirimbina secondary	Remnant pasture trees	Ladrillera primary	Ladrillera secondary
Progeny sample size		20 seeds x 12 trees	20 seeds x 15 trees	20 seeds x 2 trees	20 seeds x 16 trees	20 seeds x 19 trees
Multilocus outcrossing rate	tm	1.000 (0.014)	0.951 (0.030)	1.143 (0.424)	0.930 (0.027)	0.853 (0.057)
Singlelocus outcrossing rate	ts	0.964 (0.034)	0.850 (0.050)	0.947 (0.347)	0.845 (0.043)	0.744 (0.058)
Level of biparental inbreeding	tm-ts	0.036 (0.032)	0.101 (0.028)	0.196 (0.096)	0.085 (0.026)	0.109 (0.031)
Multilocus correlation of paternity	rp(m)	0.281 (0.165)	0.425 (0.208)	0.206 (0.057)	0.299 (0.081)	0.365 (0.127)
Singlelocus correlation of paternity	rp(s)	0.286 (0.148)	0.339 (0.211)	0.238 (0.063)	0.201 (0.088)	0.175 (0.088)
Number of pollen donors	1/rp(m)	3.559	2.353	4.854	3.344	2.74
Measure of population substructure	rp(m)-rp(s)	-0.005 (0.044)	0.087 (0.090)	-0.032 (0.006)	0.099 (0.039)	0.190 (0.064)
Correlation of selfing	rs	0.001 (0.312)	0.042 (0.024)	0.104 (0.212)	0.217 (0.085)	0.001 (0.040)

A similar pattern was found in the Ladrillera populations. In the primary forest *V. ferruginea* was highly outcrossing (tm = 0.930) with little biparental inbreeding (tm – ts = 0.085) and a small proportion of siblings shared the same father (rp(m) = 0.299). However, there was a higher correlation of selfing in this population suggesting that inbreeding in the progeny array was more likely to be from selfing than biparental inbreeding. Progeny from the Ladrillera secondary forest also showed an increased

level of selfing compared to Tirimbina (tm = 0.853) and a higher difference between multi and single locus measures of outcrossing (tm – ts = 0.109) also suggesting a large component of biparental inbreeding. The proportion of siblings sharing the same father (rp(m) = 0.365) was larger than in the primary forest but not as high as in the secondary forest at Tirimbina.

In general, primary forest showed a greater level of outcrossing, reduced biparental inbreeding and a larger number of pollen donors than secondary forest, and overall Tirimbina trees showed higher outcrossing and lower inbreeding than those at Ladrillera.

4.3.2 Spatial genetic structure

Primary forest

At Tirimbina spatial genetic structuring was evident from 8 m to 30 m (see Figure 4.1), with the observed pairwise relationship coefficient not exceeding the 95 % confidence interval at any other distance. At Ladrillera, no significant spatial genetic structure was found at the smallest spatial scales but there were two small peaks at 50 m and 90 m showing a significant positive relationship, where, for this distance class, individuals were more genetically similar to each other than would be expected at random distribution.

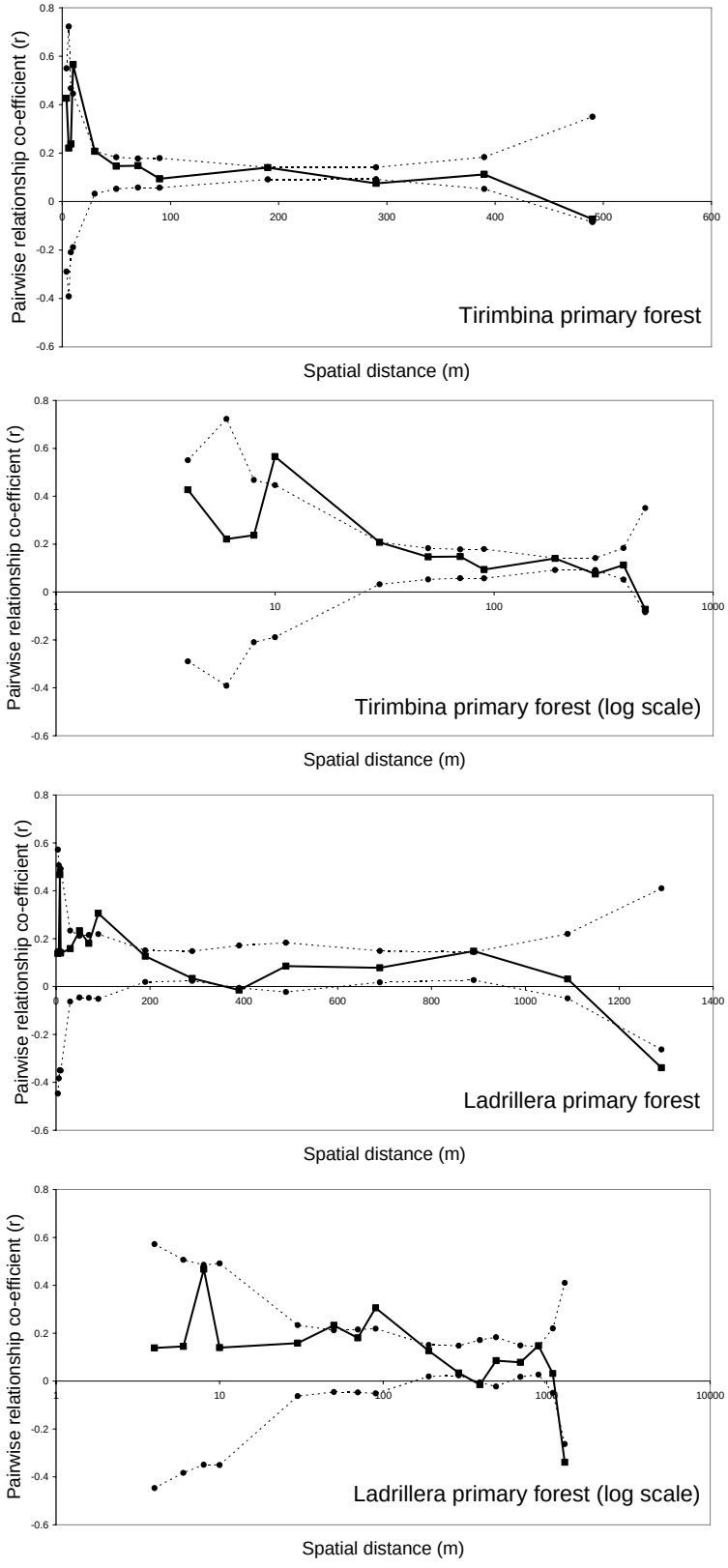


Figure 4.1: Spatial genetic structure of primary forest populations from assessment of the multi-locus pairwise relationship co-efficient, *r* (Wang 2000), dotted lines show the 95% confidence limits. The log scale illustrates detail in the smaller distance classes.

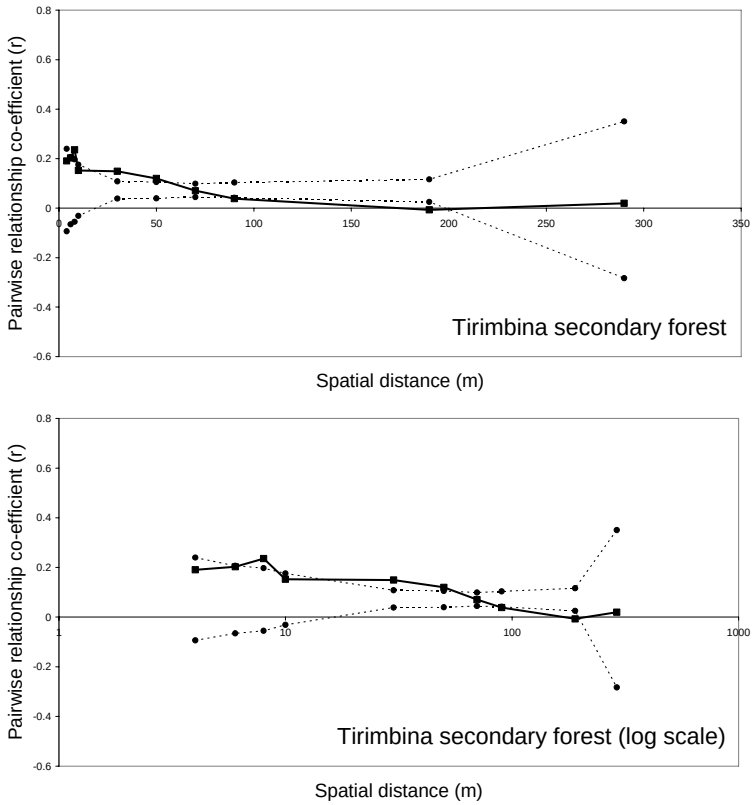


Figure 4.2: Spatial genetic structure of Tirimbina secondary forests from assessment of the multi-locus pairwise relationship co-efficient, r (Wang 2000), dotted lines show the 95% confidence limits. The log scale illustrates detail in the smaller distance classes.

Secondary forest

At Tirimbina there was some structuring found at 8 m, a smaller scale than the two primary forest populations (see Figure 4.2). There was also significant structuring found at 30 m and 50 m. At distances of 90 m and 190 m there was also significant negative relatedness between individuals, where individuals were less genetically similar than would be expected in a random distribution.

At Ladrillera in the secondary forest site A, adjacent to primary forest, strong spatial structure was present at small scales; from 4 m up to 8 m individuals were more genetically similar than would be expected at random (see Figure 4.3). Spatial structure decreased beyond this distance but at 50 m individuals were significantly

less related than would be expected at random. There was a slight second peak of positive relatedness at 70 m.

In the secondary forest site B at Ladrillera, furthest from the primary forest, spatial structuring at the smaller scales was also strong, with the population showing significant relatedness from 4 m up to 30 m (see Figure 4.3). After this the spatial structure decreased. In this population there was no secondary peak of structuring, but after 90 m, continuing to 190 m, individuals showed significant negative relatedness.

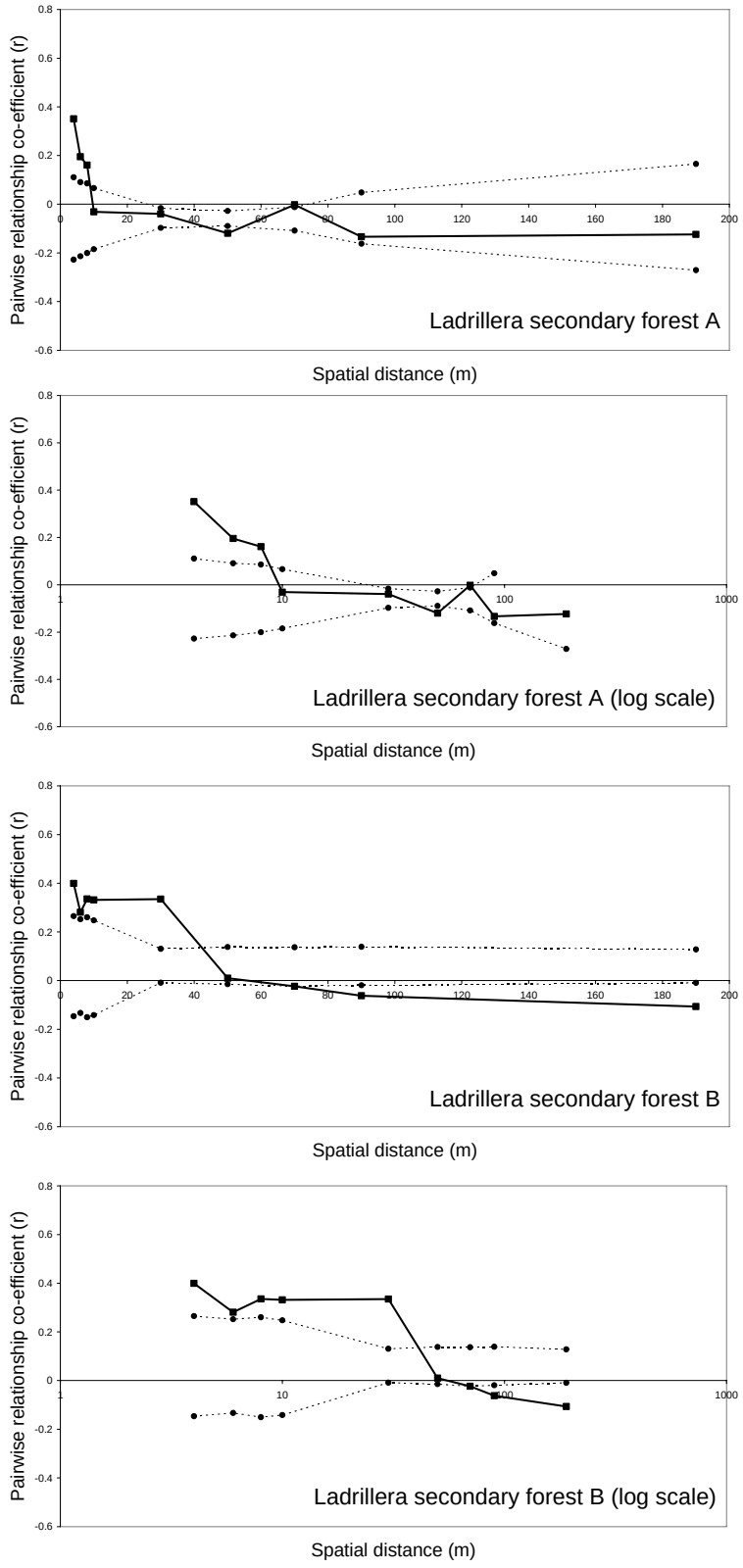


Figure 4.3: Spatial genetic structure of Ladrillera secondary forest populations from assessment of the multi-locus pairwise relationship co-efficient, *r* (Wang 2000), dotted lines show the 95% confidence limits. The log scale illustrates detail in the smaller distance classes.

Seedlings

Seedling populations were sampled at a smaller scale compared to adults. At both sites the values of *r* were higher than in primary and secondary forest populations, showing that seedlings in both populations were more genetically similar than adults. At Tirimbina there was weak structure found at the smallest distance class of 4 m, but with no further structuring found within this population. At Ladrillera the seedling population showed a peak of genetic structuring at 8 m, but after this spatial distribution was not different from that expected at random. At 70 m individuals were significantly less related than expected.

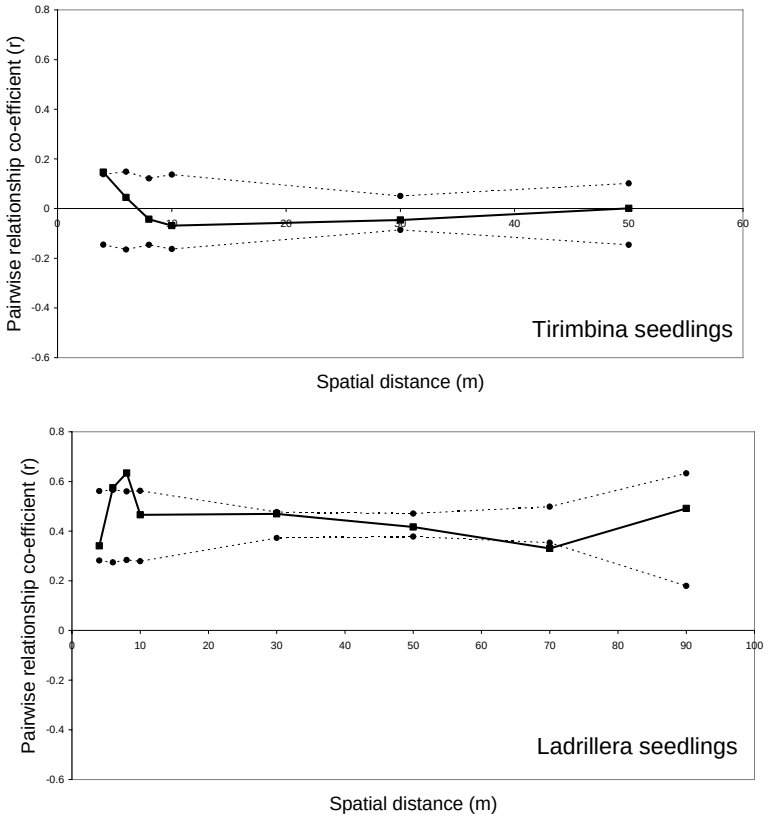


Figure 4.4: Spatial genetic structure of seedling populations from assessment of the multi-locus pair wise relationship co-efficient, *r* (Wang 2000), dotted lines show the 95% confidence limits.

4.3.3 Level of heterozygosity versus growth rate

At Tirimbina, the average growth rate between 1998 and 2003 in the primary forest population was 79.1 cm (standard deviation 28.8) for the sample with low heterozygosity and 64.6 cm (36.3) for high heterozygosity. A statistical test comparing these means showed that the difference was only significant at the 7% level. In Tirimbina secondary forest, high heterozygotic trees had higher growth in all years except between years 1 and 2 with the difference in growth rate increasing from year 2 to year 8 (see Table 4.2 and Figure 4.5). However, even the most significant result (that for years 8-9) showed that the difference between the means was not significant. Thus, although primary and secondary forest display different trends for growth rate against heterozygosity there is no significant evidence to suggest that heterozygosity influences growth rate, at least at the life stage at which adults were analysed in this study.

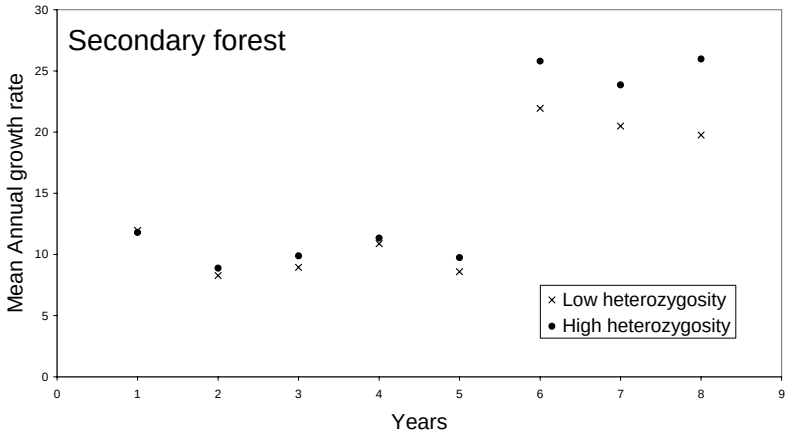


Figure 4.5: Mean growth rate for trees with low and high levels of heterozygosity in secondary forest over eight years.

Table 4.2: The mean growth rate for the high and low heterozygosity samples. The high heterozygosity sample contained 35 individuals and the low heterozygosity sample contained 33 individuals. The first year measurements were taken is 1988. The final column represents the total growth over the nine year period. The % significant row indicated the value at which these samples were significant under the null hypothesis that the samples had the same mean and the alternative hypothesis that the means were different. We would accept the alternative hypothesis if the results were significant at the 5% level (a value of less than 5% in this row).

	Year 1-2	Year 2-3	Year 3-4	Year 4-5	Year 5-6	Year 6-7	Year 7-8	Year 8-9	Year 1-9
High heterozygosity									
Mean growth (cm)	11.8	8.9	9.9	11.3	9.7	25.8	23.9	26.0	127.3
Standard deviation	6.6	4.9	5.3	5.9	4.6	14.6	15.1	18.6	69.7
Low heterozygosity									
Mean growth (cm)	12.0	8.3	8.9	10.9	8.6	21.9	20.5	19.8	110.9
Standard deviation	5.8	3.8	4.5	7.7	4.4	11.0	12.6	16.9	59.7
Difference in means	-0.2	0.6	1.0	0.4	1.1	3.9	3.4	6.2	16.4
Standard Error	1.5	1.1	1.2	1.6	1.1	3.1	3.4	4.3	15.2
% at which significant	91	56	43	79	29	22	32	15	28

Tests were undertaken to see if the results were biased by other environmental factors, in particular, crown illumination and the initial DBH. In the secondary forest there was minimal variation in crown illumination between individuals. There was more variation in the primary forest and Figure 4.6 plots the growth rate over the 5 year period against crown illumination for the high and low heterozygosity samples. There was little correlation between crown illumination and growth rate although the trend of low heterozygosity trees displaying increased growth rate can be seen across all illumination values.

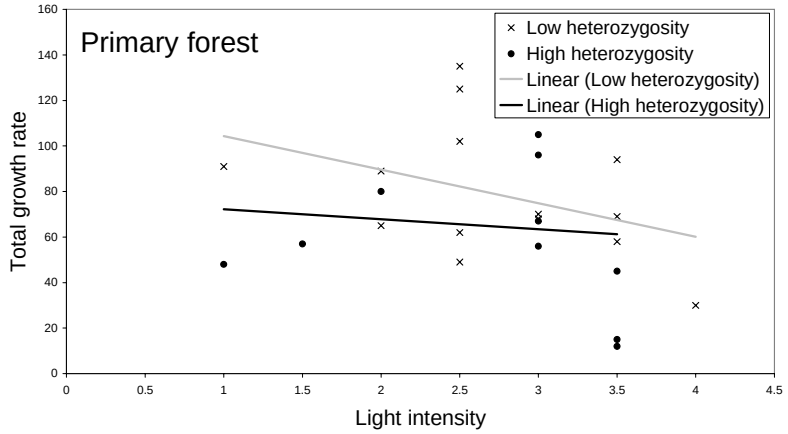


Figure 4.6: The effect of crown illumination on growth rate for trees with low and high levels of heterozygosity in primary forest. Here a crown illumination score of 1 means the crown is fully illuminated and a score of 5 means the crown receives no direct sunlight. Growth rate is calculated for the five year period, between 1998 and 2003, and is measured in cm.

In both primary and secondary forest there was no real difference in initial DBH between individuals with low and high levels of heterozygosity (see Figure 4.7). In secondary forest mean initial DBH was 378.41 for individuals with low levels and 343.69 for those with high levels. In primary forest mean initial DBH in trees with low levels was 269.21 and 296.58 in those with high levels. There was no effect of initial DBH on growth rate found in the primary forest population although there was a positive effect of larger initial DBH on growth rate found in the secondary forest population.

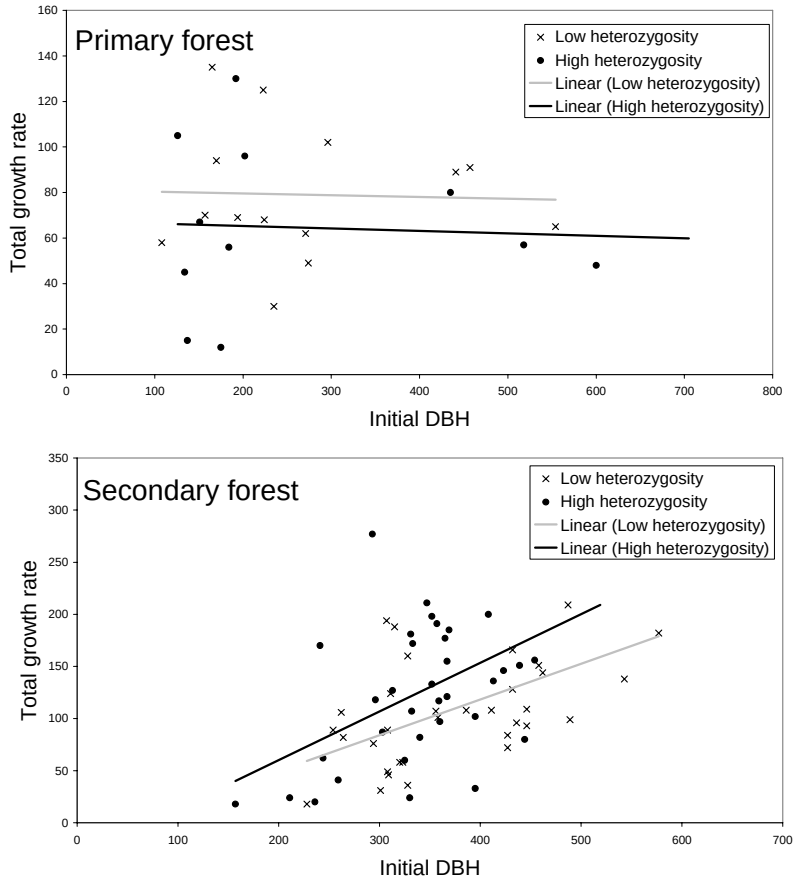


Figure 4.7: The effect of initial DBH on growth rate for trees with low and high levels of heterozygosity in primary and secondary forest.

4.4 Discussion

4.4.1 Mating system

Bawa *et al.* (1985b) reported *V. ferruginea* as largely self-compatible under controlled pollination, showing only minor reduction in fruit set with selfing. However, the data from Tirimbina and Ladrillera showed that in all populations from secondary and primary forest and also in isolated trees, *V. ferruginea* is predominantly outcrossing. However, selfing was observed in the Ladrillera primary forest and in progeny from the two remnant trees (rs = 0.217 and 0.104 respectively).

In primary forest at Tirimbina *V. ferruginea* was exclusively outcrossing, suggesting that there must be extensive pollen flow between spatially separated adult trees. There was also little biparental inbreeding in these primary populations: both Tirimbina and Ladrillera primary forest populations showed very small differences between single and multilocus estimates of outcrossing. In combination, these results suggest that, in primary populations, sparsely distributed adults encounter extensive gene flow via pollen from other unrelated adults within the primary forest. Comparison of rp(m) and rp(s) and analysis of spatial genetic structure indicated little population substructuring, suggesting that those pollen donors nearest to a seed tree were mostly unrelated. With an average of three to four adult trees contributing pollen to each seed tree, it appears that primary populations maintain extensive gene flow within a large effective population of spatially distant adults.

The density of *V. ferruginea* in secondary forest was considerably higher than in the primary forest plots surveyed and therefore there was a greater potential for outcrossing in these populations, yet at both Tirimbina and Ladrillera outcrossing rates were lower. However, estimates of selfing (uniparental inbreeding) were very low at both secondary forest sites so it can be assumed that inbreeding may be accounted for by biparental inbreeding between closely related founder populations. In young, dense secondary forest adjacent trees are likely to be half siblings: a founder effect resulting from the initial colonisation event. In a dense population of synchronously flowering trees insects may fly shorter distances, so are more commonly travelling between related individuals (e.g. in *Symphonia globulifera,* Degen *et al.* (2004) found low selfing rates but high biparental inbreeding (tm – ts = 0.156) in dense secondary populations). This may account for the slightly greater proportion of siblings that shared the same father (36.5-42.5%, compared to 28-30%

in primary forest) and lower average number of pollen donors (between 2 and 3 trees) contributing to the progeny arrays in secondary forest.

The stand of adult trees in secondary forest at both Tirimbina and Ladrillera are predominantly even aged individuals from the initial founding population. It is assumed that this population is composed of patches of siblings from a small number of seed donors. Under these conditions, it could be expected that the majority of progeny would be the result of outcrossing between related individuals. Although biparental inbreeding was found in both secondary populations there was also evidence for outcrossing between unrelated individuals suggesting that pollen flow can limit biparental inbreeding and break down genetic structure.

Despite their isolation, both remnant pasture trees at Tirimbina were also predominantly outcrossing (although values for this and other mating system parameters are less accurate due to a limited sample size of only two adults). These two trees also showed the highest degree of biparental inbreeding (0.196). However, this later value depends on the tm value, which is greater than one and should therefore be viewed with caution. The abandoned plantation is the sample site located the furthest from the primary forest, so pollination events are is more likely to be sourced from the more proximate secondary forest. The two remnant pasture trees are potentially related to the adult trees in the secondary forest and regeneration found in the plantation (either as siblings or parents). Therefore, pollen contributed by possible offspring to those adults may account for the increased level of biparental inbreeding observed. The progeny array of the remnant pasture trees is also the only sample to show evidence of structuring in the pollen donors (the difference between Rp(s) and Rp(m) is negative, -0.032). Therefore, much of the biparental inbreeding may be

accounted for by relatedness between the pollen donors of the more recently colonised adjacent secondary forest. The progeny from these trees had an estimated 4.814 pollen donors, higher than progeny from both primary and secondary forest. This is potentially due to higher frequency of pollinator visits to isolated trees in an open landscape (e.g. as shown in *S. humilis,* White *et al.* 2002).

4.4.2 Spatial genetic structure

In primary forest there was no spatial structure found at the smallest spatial scales at Ladrillera and little at Tirimbina, although there was relatedness between individuals throughout the primary forest. There were however, peaks of positive spatial autocorrelation showing some degree of spatial genetic structuring, at relatively large spatial scales (30 m at Tirimbina and 50 m to 90 m at Ladrillera). This may be a consequence of repeated gap colonisation events, with small-scale structure removed as density declines through competition but with each small gap population showing some similarities in genetic composition to other gap populations.

In secondary forests, a greater amount of spatial structuring was found than in primary forest, with individuals in close proximity being more similar genetically than expected under random distribution. Significant positive spatial autocorrelation at the lower distance scales (1-15 m) suggests a structure of family clusters (Solkan and Oldan 1978). This is consistent with the concept of plant populations being subdivided into neighbourhoods of interbreeding, related individuals (Levin and Kerster 1974). The family clusters were at approximately 8 m at Tirimbina and at 4 m to 8 m in one Ladrillera population and 4 m to 30 m in the other, beyond these spatial scales the structure breaks down. The seedling populations also showed family clustering at the smallest distances (4 m to 8 m), with both seedling populations showing higher within population genetic similarity.

There was little incidence of positive spatial autocorrelation at shorter distances in the primary forest populations of *V. ferruginea*; therefore the strong spatial structure found in the secondary forest populations must be a product of factors relevant only to those populations. The difference is probably either a consequence of the higher population densities found in secondary forest or the structure was established during colonisation.

This contrast, where populations of *V. ferruginea* in primary forest showed little or no spatial genetic structure and those in secondary forest showed structuring that increases with density, was also found using AFLP data (Lowe *et al.* in press). This pattern was similar to those obtained under simulations using an isolation-by-distance model (Sokal and Wartenberg 1983; Sokal and Jacquez 1991), which is based on the expectation that the probability of mating depends on the distance between individuals or the variance in the dispersion of their propagules.

In primary forest there was some structuring at greater distances (30 m to 90 m), yet all secondary forest populations also showed negative spatial autocorrelation at these distances (50 m to 90 m). In dense patchy populations, as pollinators mainly move between nearby plants, mean flight distance decreases and thus pollen flow among patches also diminishes, favouring population subdivision in small patches. Low density has the opposite effect, facilitating pollen dispersal events at longer distances and decreasing the probability of geographical differentiation (Levin *et al.* 1971). The genetic structuring found in populations of *V. ferruginea* suggests that there are different patterns of genetic structure found in primary and secondary forest. Secondary forests may show small scale structuring due to a founder effect with patches of related individuals from limited seed donors or different densities of *V.*

ferruginea in secondary and primary forest may be impacting on the pollinator behaviour. *Vochysia ferruginea* over 50 m apart are genetically similar in primary forest and differentiated in secondary forest, suggesting that in the low density primary forest, pollinators may move a greater distance between trees than they do in secondary forest.

One possible consequence of the spatial genetic structure found at short distances in secondary forest is an increase in matings between spatially proximate individuals, and this may account for the increased biparental inbreeding found in secondary forest populations (as discussed above). The increased incidence of biparental inbreeding may have a detrimental effect on offspring fitness. Several examples of reduced fitness of progeny with decreasing interparental distance have been reported in self-incompatible plants (Heywood 1991; Waser and Price 1994). In contrast, other studies have shown no effect of parental distance on the fitness of self incompatible plants (Newport 1989; Mora´n-Palma and Snow 1997).

4.4.3 Level of heterozygosity versus growth rate

In *V. ferruginea* populations from both primary and secondary forest no significant effect of heterozygosity on growth rate was observed. This suggests that any negative fitness effects suffered by inbred individuals, from selfing or biparental inbreeding, may have already acted at earlier life history stages (pollination, seed set, germination, seedling survival). Thus inbred individuals may already have been purged from the population by the time individuals were surveyed as adults in this study. However, there appeared to be a trend for better growth rate over time for individuals with higher levels of heterozygosity in secondary forest. The secondary forest blocks sampled in this study were dense monospecific stands of regenerated trees of equal age, and individuals can be considered as adults from the original colonisation event.

Therefore, the slight trend for increased growth with increased heterozygosity over time can be directly related to the increasing age of the adults sampled. After colonisation, those individuals with a higher heterozygosity levels may have a growth advantage that is only apparent years after colonisation and may be a response to increased competition from other species. For each year, the mean difference between low and highly heterozygous individuals was not significant, however, the trend of increasing fitness for more heterozygous individuals over time was found throughout all eight years. In contrast, primary forest trees with a lower level of heterozygosity had a higher growth rate over 5 years. However, these primary forest trees had a much lower mean initial DBH and so may be from recent gap colonisation where individuals with greater heterozygosity may prove to be fitter with increasing age. Detrimental recessive genes in the homozygotic state may have a large or small effect on an individual. Genes of a large effect are more likely to be lethal and act at an earlier stage; however, genes of smaller effect are more likely to build up over the life time of an individual and have a progressively detrimental effect over time. The possible post recruitment fitness effects of heterozygosity should be further investigated with greater sample sizes, at different life stages and with a more thorough method of determining within individual heterozygosity to determine whether the trends showed by these results are real. It would also be useful to investigate how other fitness measures, e.g. the fecundity of individuals, might be related to individual heterozygosity.

Chapter 5: Gene Flow in *Vochysia ferruginea* Populations within Primary and Secondary Forest

5.1 Introduction

Assessment of contemporary gene flow in plants is important for understanding various aspects of plant population biology and genetic conservation as gene flow is an important factor determining patterns of variation of both neutral and selected genes. A good estimation of contemporary gene flow also allows evaluation of the genetic impact of recent processes such as fragmentation, re-colonisation and extinction of local populations and prediction of likely future changes within populations (Sork *et al.* 1999).

Tree populations have the potential to disperse genes over a wide area via pollen (Nason *et al.*1998) and recent studies, using microsatellites to directly estimate gene flow, have found evidence of very large pollen flow distances (Collevatti *et al.* 2001; Dick 2001; White *et al.* 2002). Most tropical canopy trees occur in low density and are therefore presumed to be adapted to achieve high levels of gene dispersal. High levels of gene flow are also expected for pioneer trees colonising gaps as these gaps are potentially spatially dispersed (Kaufman *et al.* 1998).

The gene dispersal of tropical trees has been found to be sensitive to ecological factors (Franceschinelli and Bawa 2000) and pollination distances may not be consistent for a species between different populations or flowering events (Degen *et al.* 2004; Kenta *et al.* 2004). *Vochysia ferruginea* is a pioneer tree that occurs in both low and high density populations, previous chapters showing genetic diversity and differentiation have suggested that gene flow may be different in primary and

secondary forest. Increasing density has been found to decrease pollen flow distances in other tropical tree species (Degen *et al.* 2004) and, where populations are fragmented or low density, pollen flow distances may be very large (Dick *et al.* 2003. White *et al.* 1999).

This study uses a likelihood method of paternity analysis to investigate pollen flow in *V. ferruginea* populations in primary and secondary forest with very different densities of reproducing adults. Direct methods of measuring pollen mediated gene flow using paternity assignment require sampling of all potential pollen donors in the population studied and in this study all populations were exhaustively sampled within study plots. However, there was a high level of allelic dropout and so a subset of adults with complete genotypes was used to estimate gene flow using a TwoGener analysis. The TwoGener approach (Smouse *et al.* 2001) estimates the extent of pollen movement based on levels of genetic differentiation among pools of pollen gametes effectively fertilizing ovules of different mother plants. TwoGener analysis allows for the estimation of pollen flow without genotyping all potential fathers in the population and even without having an estimation of adult density (Austerlitz and Smouse 2002). Austerlitz *et al.* (2004) found similarity with the estimation of pollen flow distances from TwoGener and paternity-based methods.

5.2 Methods

5.2.1 Study site and populations

Populations from both Tirimbina and Ladrillera were sampled as described in Chapter 2. For analysis of gene flow, progeny arrays were collected from 20 mother trees in low density primary forest populations and 20 mother trees in high density secondary forest populations. The secondary forest selected at the Ladrillera site was the

secondary plot B, adjacent to pasture land (see Chapter 2, Figures 2.4 and 2.6 for maps).

5.2.2 Paternity analysis

Gene flow via pollen was estimated by a using a likelihood method of paternity analysis. Adults and offspring were genotyped then the most probable fathers were identified by the calculation of LOD scores (log-likelihood ratio). The ratios compare the likelihood of an individual being the parent of a given offspring divided by the likelihood of these two individuals being unrelated (Gerber *et al.* 2000). This was carried out using FaMoz software (Gerber *et al.* 2003). FaMoz estimates: expected gene flow (total reproducing population size – number of genotyped parents)/ total reproducing population size; true gene flow (actual number of times a parent different from the genotyped parents produced one of the offspring); apparent gene flow (number of times no genotyped parent was detected for the simulated offspring according to statistical tests) and cryptic gene flow (number of times a parent had been detected among the genotyped parents according to the statistical tests when the true parent was not one of them). FaMoz determines confidence for assigning paternity by carrying out analyses based on the genotyped dataset versus simulated datasets (comparison of distributions of LOD scores). So that estimates of true gene flow from outside the stand are determined by comparing estimates of paternity drawn from fathers in the simulated population whose genotypes are determined from a pool derived from the allele frequencies based on the genotyped adults.

Gene flow was estimated in all populations with populations assumed to be in Hardy-Weinberg equilibrium and also with a departure from Hardy-Weinberg proportions where homozygous genotypes at any locus were more likely to be obtained by chance. The range for departure from Hardy-Weinberg proportions in FaMoz was from 0 to 1

and gene flow estimates were done at 1 and 0.1 (this reflects the range of F_{IS} values found in primary and secondary populations, see Chapter 3). The effect of scoring errors was assessed by including a probability of error in genotyping into the LOD score calculation where, when microsatellites are used, an allele is arbitrarily replaced with another one. The range for LOD score calculation was also from 0 to 1 and gene flow estimates were done at 1 and 0.5. The effect of simulation error was also tested, at 1 only. The dataset had missing data and so a paternity analysis with replaced alleles based on allele frequencies was made using one population to test the effect of missing data.

5.2.3 TwoGener analysis

In this study a TwoGener analysis (Smouse *et al.* 2001) was used to estimate pollen dispersal. TwoGener uses the difference in allele frequencies among progeny arrays to calculate the parameter Φ_{FT}, which is analogous to Wright's $F_{ST,}$ giving the level of differentiation among pollen clouds. Combined with an estimate of local population density, an estimate of the average distance of pollen flow and the shape of the dispersal curve can be obtained. The principle of this method is to estimate the differentiation of allelic frequencies among the pollen pools sampled by several females in the population. The relationship between Φ_{FT} and dispersal distance has been derived for given dispersal curves allowing for the development of several estimates of pollen dispersal (Austerlitz and Smouse 2002).

5.3 Results

5.3.1 Exclusion probabilities and LOD distribution

Differing levels of polymorphism in the five loci meant that exclusion probabilities varied between loci, from 0.363 to 0.781 (see Table 5.1). However, for all four populations the cumulative exclusion probability exceeded 0.95. There was little

discrimination between the LOD score distribution of genotyped individuals and simulated individuals in all populations (see Figure 5.1). The LOD score distribution for simulated individuals was within the curve of the LOD score distribution for genotyped individuals in all but the primary forest population at Ladrillera, where there was a slight separation of the two distribution curves. Including a LOD scoring error did not effect the LOD score distribution.

Table 5.1: Exclusion probabilities at each locus and cumulative over all loci for all populations.

Population	A1-5	A1-10	A1-15	A1-20	A1-35	Cumulative
Tirimbina primary	0.781	0.713	0.543	0.318	0.601	0.992
Tirimbina secondary	0.595	0.743	0.446	0.404	0.577	0.985
Ladrillera primary	0.628	0.390	0.363	0.404	0.519	0.959
Ladrillera secondary	0.609	0.480	0.219	0.504	0.500	0.961

5.3.2 Paternity analysis and gene flow estimates

At Tirimbina the percentage of correct father choice was low, 22.3% in primary forest and 21% in secondary (see Table 5.2a). At Ladrillera the percentage of correct father choice was higher but still quite low, 33.9% in primary forest and 53% in secondary. The number of progeny for which paternity could be correctly assigned was 36 in primary forest and 39 in secondary at Tirimbina, 13 in the Ladrillera primary forest population and 2 in secondary forest. The percentage of paternity correctly assigned ranged from 0.52% in the secondary forest population at Ladrillera to 4.98% in the secondary forest population at Tirimbina.

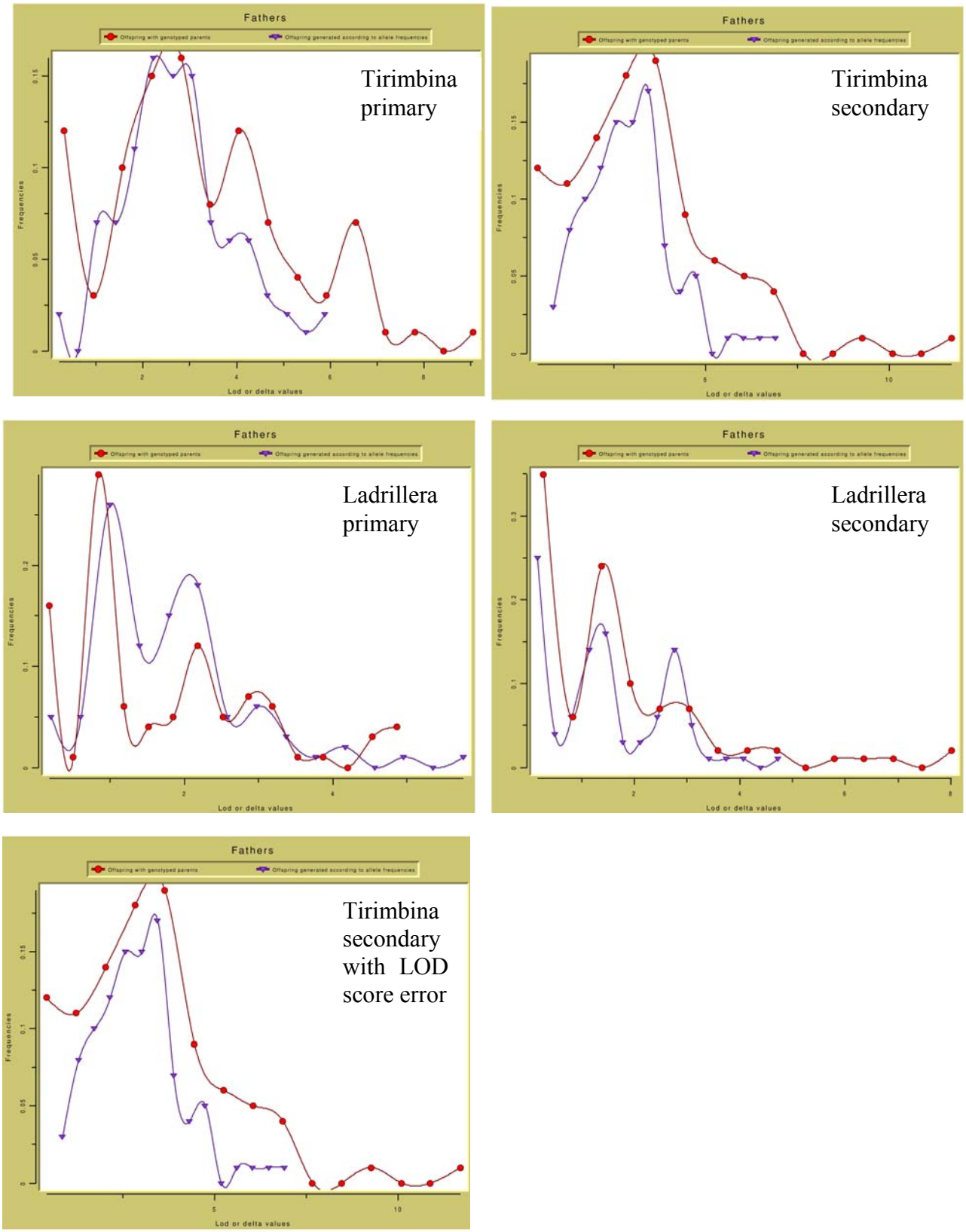


Figure 5.1: LOD score distribution for genotyped individuals (red) and simulated individuals (blue).

Table 5.2a: Percentage and number of paternity correctly assigned. Percentage of correct father choice, Fathers (%); number of paternity correctly assigned, Paternity (*N*); percentage of paternity correctly assigned among the assigned paternity, Paternity (%).

Population	Fathers (%)	Paternity (*N*)	Paternity (%)
Tirimbina primary	22.3	36	4.57
Tirimbina secondary	21	39	4.98
Ladrillera primary	33.9	13	2.14
Ladrillera secondary	53	2	0.52

In all populations the expected gene flow (the proportion of gametes not derived from the genotyped individuals) was very high, ranging from 872 to 899 out of 1000 offspring (see Table 5.2b). The true gene flow (the actual number of times a gamete generated according to allele frequencies (from 'outside the study stand') produced one of the offspring) was also high, therefore simulated individuals outside the population accounted for the gene flow with little difference between expected and true gene flow in all populations. There was no observable difference in expected or true gene flow estimates between both primary and secondary forest and between the two sites.

Table 5.2b Estimates of gene flow from paternity analysis. 20 mothers with 50 offspring, with 1000 reproducing individuals. Simulation error 0, LOD score calculation error 0, Heterozygote deficit 0. Test threshold to choose father as the true one = 2.

Population	Parents	Expected	True	Apparent	Cryptic	% t/a	% c/a
Tirimbina primary	101	899	898	212	711	423.58	335.38
Tirimbina secondary	128	872	861	217	690	396.77	317.97
Ladrillera primary	120	880	878	392	552	223.98	140.82
Ladrillera secondary	105	895	886	617	358	143.6	58.02

In both Tirimbina populations, levels of apparent and cryptic gene flow were similar in primary and secondary forest. Apparent gene flow (the number of times no parent from 'inside the study stand' was detected for the simulated offspring according to statistical tests) was 212 in primary forest and 217 in secondary forest. Cryptic gene flow (the number of times a parent was detected among the genotyped parents according to the statistical tests, whereas the true parent was 'outside the study stand') was high in these populations, 711 in primary forest and 690 in secondary forest. In the two Ladrillera populations there was a difference between primary and secondary forest. In the primary forest population apparent gene flow was 392 and cryptic gene flow was 552; in secondary forest apparent gene flow was greater at 617 and so cryptic gene flow was lower at 358.

Deviation from Hardy-Weinberg equilibrium

With a heterozygote deficit of 0.1 the percentage of correct father choice increased in all populations except Ladrillera secondary forest where it decreased from 53% to 51.4%. The greatest increases were in both primary forest populations; it rose from 22.3% to 40.3% at Tirimbina and from 33.9% to 40.7% at Ladrillera (see Table 5.3a). The number of paternity correctly assigned increased in all populations, ranging from an increase of 48 in Ladrillera primary forest to an increase of 114 in primary forest at Tirimbina. The percentage of paternity correctly assigned among the assigned paternity also increased in all populations; from 4.57% to 21.68% in Tirimbina primary forest, 4.98% to 16.23 in Tirimbina secondary forest, 2.14% to 12.06% in Ladrillera primary forest, 0.52% to 18.81% in Ladrillera secondary forest.

Table 5.3a: Percentage and number of paternity correctly assigned, testing the effect of a heterozygote deficit (heterozygosity = 0.1). Percentage of correct father choice, Fathers (%); number of paternity correctly assigned, Paternity (*N*); percentage of paternity correctly assigned among the assigned paternity, Paternity (%).

Population	Fathers (%)	Paternity (*N*)	Paternity (%)
Tirimbina primary	40.3	150	21.68
Tirimbina secondary	25.1	136	16.23
Ladrillera primary	40.7	61	12.06
Ladrillera secondary	51.4	79	18.81

In all populations, the estimate for true gene flow decreased when heterozygote deficit was 0.1 (see Table 5.3b). This decrease was uniform over all populations; at Tirimbina it decreased by 162 in the primary forest population and by 161 in secondary; at Ladrillera it decreased by 157 in primary forest and by 175 in secondary forest. Increasing the heterozygote deficit also led to differences in apparent gene flow

between primary and secondary forest populations. The number of times no possible father could be found amongst the genotyped individuals increased by 96 in the Tirimbina primary forest population, decreased by 55 in Tirimbina secondary forest population, increased by 102 in Ladrillera primary forest population and decreased by 37 Ladrillera secondary forest population. All populations showed a decrease in the level of cryptic gene flow. Where the number of possible fathers increased in secondary forest; these were more often assigned as parents than when there was no heterozygote deficit due to lower gene flow from outside the population.

Table 5.3b: Estimates of gene flow from paternity analysis, testing the effect of a heterozygote deficit (heterozygosity = 0.1). 20 mothers with 50 offspring, with 1000 reproducing individuals. Simulation error 0, LOD score calculation error 0, Heterozygote deficit 0.1. Test threshold to choose father as the true one = 2.

Population	Parents	Expected	True	Apparent	Cryptic	% t/a	% c/a
Tirimbina primary	101	899	736	308	483	238.96	156.82
Tirimbina secondary	128	872	700	162	585	432.1	361.11
Ladrillera primary	120	880	723	494	377	146.36	76.32
Ladrillera secondary	105	895	720	580	285	124.14	49.14

With a heterozygote deficit of 1, percentage of correct father choice, number of paternity correctly assigned and percentage of paternity correctly assigned among the assigned paternity were all 0 in all populations. No possible father could be found

within the population, or from outside the population so true gene flow was also 0 in all populations.

Effect of scoring and simulation errors

The effect of miscoring was tested with a LOD score calculation error of 0.5 (range is 0-1). In all populations the percentage of correct father choice increased significantly ranging from 86.1% in the Tirimbina secondary forest population to 91.5% in the Ladrillera secondary forest population. There was no difference when the LOD score calculation error was increased to 1.

Table 5.4a: Percentage and number of paternity correctly assigned, testing the effect of LOD score calculation error (0.5). Percentage of correct father choice, Fathers (%); number of paternity correctly assigned, Paternity (*N*); percentage of paternity correctly assigned among the assigned paternity, Paternity (%).

Population	Fathers (%)	Paternity (*N*)	Paternity (%)
Tirimbina primary	90.8	0	NaN
Tirimbina secondary	86.1	0	NaN
Ladrillera primary	88.2	0	NaN
Ladrillera secondary	91.5	0	NaN

When a simulation error was tested the percentage of correct father choice also increased (see Table 5.4b), ranging from 77.1% in the Tirimbina secondary forest population to 83.2% in the Ladrillera secondary forest population. However, in all populations, for either a LOD score calculation error of 1 or a simulation error of 1, the number of paternity correctly assigned dropped to 0 (see Table 5.4a, b).

Table 5.4b: Percentage and number of paternity correctly assigned, testing the effect of simulation error. Percentage of correct father choice, Fathers (%); number of paternity correctly assigned, Paternity (*N*); percentage of paternity correctly assigned among the assigned paternity, Paternity (%).

Population	Fathers (%)	Paternity (*N*)	Paternity (%)
Tirimbina primary	79.6	0	0
Tirimbina secondary	77.1	0	0
Ladrillera primary	78.4	0	0
Ladrillera secondary	83.2	0	0

With a LOD score calculation error of 0.5 to simulate miscoring, levels of true gene flow are not significantly different and remain high. The most significant difference is in apparent gene flow, where no possible father could be found within any of the populations (see Table 5.4c). There was no significant difference when the LOD score calculation error was increased to 1.

Table 5.4c: Estimates of gene flow from paternity analysis, testing the effect of LOD score calculation error (0.5). 20 mothers with 50 offspring, with 1000 reproducing individuals. Simulation error 0, LOD score calculation error 0.5, Heterozygote deficit 0. Test threshold to choose father as the true one = 2.

Population	Parents	Expected	True	Apparent	Cryptic	% t/a	% c/a
Tirimbina primary	101	899	908	1000	0	90.8	0
Tirimbina secondary	128	872	861	1000	0	86.1	0
Ladrillera primary	120	880	882	1000	0	88.2	0
Ladrillera secondary	105	895	915	1000	0	91.5	0

With a simulation error of 1, levels of true gene flow are further increased in all but the Ladrillera primary population. Apparent gene flow is again significantly different, and is higher in all populations (see Table 5.4d). The few instances of there being a potential father found within the population are assumed to be less likely than fathers outside the stand with cryptic gene flow found in all populations.

Table 5.4d: Estimates of gene flow from paternity analysis, testing the effect of simulation error. 20 mothers with 50 offspring, with 1000 reproducing individuals. Simulation error 1, LOD score calculation error 0, Heterozygote deficit 0. Test threshold to choose father as the true one = 2.

Population	Parents	Expected	True	Apparent	Cryptic	% t/a	% c/a
Tirimbina primary	101	899	904	881	108	102.61	12.26
Tirimbina secondary	128	872	875	892	104	98.09	11.66
Ladrillera primary	120	880	876	896	92	97.77	10.27
Ladrillera secondary	105	895	915	910	83	100.55	9.12

Effect of missing data

Conducting a paternity analysis with a complete dataset with missing values replaced with randomly selected alleles based on allele frequencies increased the percentage of correct father choice from 33.9% to 61.1% (see Table 5.5a).

Table 5.5a: Percentage and number of paternity correctly assigned, testing the effect of missing data. Percentage of correct father choice, Fathers (%); number of paternity correctly assigned, Paternity (*N*); percentage of paternity correctly assigned among the assigned paternity, Paternity (%).

Population	Fathers (%)	Paternity (*N*)	Paternity (%)
Ladrillera primary with missing data	33.9	13	2.14
Ladrillera primary with complete data	61.1	9	2.8

Greater accuracy in determining the correct father meant there was a lower ratio of cryptic to apparent gene flow (41.89 with complete data and 140.82 with missing data). However, levels of true gene flow were approximately equal, with 886 out of 1000 parents estimated to be from outside the stand with complete data and 878 out of 1000 with missing data (see Table 5.5b).

Table 5.5b: Estimates of gene flow from paternity analysis, testing the effect of missing data. 20 mothers with 50 offspring, with 1000 reproducing individuals. Simulation error 0, LOD score calculation error 0, Heterozygote deficit 0. Test threshold to choose father as the true one = 2.

Population	Parents	Expected	True	Apparent	Cryptic	% t/a	% c/a
Ladrillera primary with missing data	120	880	878	392	552	223.98	140.82
Ladrillera primary with complete data	105	880	886	678	284	130.68	41.89

5.3.3 TwoGener analysis

In Tirimbina the global estimate of Φ_{ft} was 0.11 in both primary and secondary forest using a normal distribution (see Table 5.6). The average distance between mothers was 215.01 m in primary forest and 74.7 m in secondary forest. Mean pollen dispersal was estimated as between 10.4 m and 29.4 m in primary forest and between 0.9m and 4.4 m in secondary forest. At Ladrillera the global estimate of Φ_{ft} was 0.08 in primary forest and 0.09 in secondary forest with the average distance between mothers was 421.37 m in primary forest and 67.56 m in secondary forest. Mean pollen dispersal was between 5.1 m and 35.8 m in primary forest and between 4.8 m and 7.9 m in secondary forest.

Table 5.6: Estimates of gene flow from TwoGener analysis using a normal dispersal function, where *M* (*m*) is the average distance between mothers in meters and Φ_{ft} is analogous to F_{ST}, and σ is an estimate of the pollen dispersal distance assuming a dispersal curve and a density of reproducing adults (*d*) in the landscape. The joint estimate uses both the density of reproducing adults and the dispersal distance.

Population	Error	Global Φ_{FT}	*M* (*m*)	Global σ	σ joint estimate	Global *d*	*d* joint estimate
Tirimbina primary	0.2	0.11	215.01	29.4	10.4	36.8	13
Tirimbina secondary	0.44	0.11	74.8	4.4	0.9	5.6	1.1
Ladrillera primary	0.53	0.08	421.37	35.8	5.1	44.8	6.4
Ladrillera secondary	1.41	0.09	67.56	4.8	7.9	6.1	9.9

5.4 Discussion

5.4.1 Problems associated with the analysis

The loci used were sufficiently polymorphic for a high cumulative exclusion probability; however, all genotyped individuals had a high number of drop-outs, spread across all loci, which affected the strength of the paternity analysis. Analysis of the LOD score distribution of genotyped individuals and simulated individuals showed that the two distributions could not be discriminated from each other and there was no clear intercept between them. Inclusion of scoring and simulation error altered the LOD score distribution but still failed to separate the two curves. This means there was little differentiation between the genotyped individuals and the simulated individuals so the most likely father would be chosen based on small differences in the LOD score. Therefore, there is little confidence in the father chosen and this is reflected in the low percentage of correct father choice found in most populations (lower than 35% in all populations except Ladrillera secondary). In testing the effect of missing data it would appear that missing data reduces the accuracy of the paternity analysis: the percentage of correct father choice increased from 33.9 % to 61.1% in the tested population. This primarily affects estimates of the level of cryptic gene flow (gene flow from outside the stand incorrectly assumed to be from a possible father inside the stand) compared to apparent gene flow (the number of times no parent could be found inside the stand). However, there was no real difference between estimates of true gene flow (gene flow from outside the stand) with missing values added. Therefore, incomplete data does not necessarily, in itself, constitute a problem in estimating gene flow. However, when FaMoz replaces missing values with randomly chosen alleles based on allele frequencies, it does not decrease the similarity between the genotyped individuals and the simulated population and may actually increase it. Therefore, the input of assumed alleles will not decrease the

error caused by insufficient discrimination between the LOD score distribution of the genotyped population and the simulated population. The dataset is therefore inadequate for accurate paternity assignment using these methods and so results are reported with caution.

Simulating scoring errors with a LOD score calculation error did not alter estimates of gene flow from outside the population. Gerber *et al.* (2000) found that cryptic gene flow was smaller for microsatellites and decreased for nonzero mistyping with a systematic overestimation of true gene flow and apparent gene flow. Without a LOD score calculation error many of the trees in the stand are possible fathers but not the most likely. With a LOD score calculation error of 0.5 no individual in the genotyped population is a potential father and so there is no gene flow within the population. A simulation error of 1 leads to a similar situation, the number of possible fathers is largely decreased. There remain some possible fathers in all populations with most fathers assumed to be from outside the stand.

TwoGener analysis does not require the precise determination of the pollen donor and although greater polymorphic loci yields better estimations it has been shown to reliably estimate genetic structure with low genetic resolution (Smouse *et al.* 2001). The *V. ferruginea* dataset is therefore more amenable to TwoGener analysis. However, in all populations there was a high level of error so results from this analysis should also be treated with caution.

5.4.2 Gene flow in Vochysia ferruginea populations

In all populations of *V. ferruginea* paternity assignment using a maximum likelihood approach estimated high levels of gene flow, with over 80% of gametes assumed to originate from outside the sampled population. Estimates of gene flow from outside

the population ranged from 72%, in Ladrillera secondary (with a heterozygote deficit of 0.1) to 89.8% in Tirimbina primary forest. Assuming a scoring or simulation error increased the accuracy of the tests. However, this meant that there were little or no fathers found for offspring, according to statistical tests, and so this was not useful for estimating gene flow.

In both Tirimbina populations there were few instances where there was no possible father among the genotyped individuals; however, the LOD scores indicated that these possible fathers were less likely to have produced the offspring than other trees from outside the stand. Therefore, there was a high incidence of cryptic gene flow predicted for these populations, with high ratios of both true to apparent and cryptic to apparent gene flow. In the two Ladrillera populations there was an equivalent estimation of high gene flow from outside the sampled population. In contrast to the Tirimbina populations and the Ladrillera primary forest population, the offspring from the secondary forest showed a high level of apparent gene flow and lower cryptic gene flow. In this case there were few possible fathers amongst the genotyped individuals and so cryptic gene flow was less; this was also seen when there was no missing data and reflects a lower number of dropouts and a higher accuracy that excludes more fathers. Although the Ladrillera secondary forest population had significantly fewer possible fathers than the other populations, the level of gene flow from outside the population was estimated as equivalent to the other populations.

With a heterozygote deficit of 0.1, gene flow from outside the population decreased in all populations. The number of possible fathers found within the populations decreased in both primary forest populations and increased in the two secondary populations. With a maximum heterozygote deficit no father could be found for

offspring from within the population or amongst simulated individuals outside the population. In both Tirimbina populations and the Ladrillera primary forest population there was no heterozygote deficit found, however, the Ladrillera secondary forest population had a F_{IS} of 0.175 (see Chapter 3) and may be better analysed by assuming a heterozygotic deficit.

Mean pollen dispersal estimates from TwoGener analysis ranged from 0.9 m to 35.8 m, with a Φ_{FT} of 0.11 in populations at Tirimbina and lower in Ladrillera (0.08 in primary forest and 0.09 in secondary forest). Differences in distance between mothers reflect the densities of the populations, mothers in primary forest populations were over 200 m apart, whilst mothers in secondary populations were less than 75 m apart. The mean pollen flow estimates differ between primary and secondary forest populations, with very low estimates for secondary forest, 1.92 m and 2.47 m increasing to 67.56 m and 47.76 m in the two primary forest populations. The low estimates of mean pollen dispersal contrast with the estimates from paternity analysis; however, dispersal estimates in TwoGener are confounded with adult population structure, and the normal (short-tailed) dispersal function may inadequately represent long-distance gene flow (Lewis 1997; Higgins and Richardson 1999; Burczyk *et al.* 2004; Burczyk *et al.* 2005).

Estimated mean pollen dispersal distances were lower than those found in other tropical trees, particularly in the secondary forest populations. However, it would appear that there is potentially a high level of gene flow from outside the populations. *Vochysia ferruginea* is pollinated by insects, primarily large bees and butterflies (Flores 1983a). Large insects are able to forage large distances. High levels of gene flow are typical for tropical forest trees and can be expected for pioneer species that

exploit gaps in the canopy as they tend to occur in low density in natural forest. The genetic diversity found in these populations and high allelic richness in progeny (Chapter 3) also suggests greater pollen dispersal than estimated by the TwoGener analysis.

5.4.3 Conclusions

The combined results from paternity analysis and TwoGener suggest that whilst there is potentially long distance pollen flow from outside these populations, most pollination events are between nearest neighbours. That mean pollen flow estimates differ between primary and secondary forest is most likely a reflection of the difference in distances between nearest neighbours in high and low density populations.

Other studies of gene flow in tropical trees show higher gene flow estimates than those found in the TwoGener analysis of *V. ferruginea* populations (although TwoGener estimates mean pollen dispersal distance and maximum dispersal distance may be greater). For *Dinizia excelsa,* pollinated by stingless bees, Φ_{FT} was 0.104 with an estimate of the pollen dispersal distance of 188 m assuming a normal dispersal distribution (Dick *et al.* 2003). In fragmented populations pollinated by African bees, *Apis mellifera scutellata*, Φ_{FT} was 0.00167 with an estimate of the pollen dispersal distance of 1264 m assuming a normal dispersal distribution (Dick *et al.* 2003). Studies of *Swietenia macrophylla* in fragmented populations have also shown high levels of gene flow with estimates of 200 m (e.g. Cespedes *et al.* 2003; White and Powell; Novick *et al.* 2003; Lowe *et al.* 2003; Lemes *et al.* 2003). Sezen *et al.* (2005) measured pollen mediated gene flow in the pioneer *Iriartea deltoidea*, a palm pollinated by bees and detected a maximum pollen movement of 220 m during colonisation. Most tropical trees occur in very low densities with density decreasing

as habitats become fragmented. Even in primary forest habitats, increased disturbance had led to increased numbers of *V. ferruginea* and insect pollinators may be able to forage between individuals over short distances in both primary and secondary forest habitats.

The secondary forest populations of *V. ferruginea* had a much higher density of individuals, primary forest populations were approximately 4.1 individuals/ha and secondary forest populations were approximately 192 individuals/ha (C. Navarro pers. comm.). In other species that occur in different densities a density effect has been found on pollinator behaviour and thus, gene flow estimates. In populations of *Symphonia globulifera,* Degen *et al.* (2004) found that dense plots led to less expansive foraging by insect pollinators. Paternity analysis of gene flow in *V. ferruginea* populations in both primary and secondary plots showed an equivalent estimate of extensive gene flow. Populations in primary and secondary forest differed greatly in the density of adult *V. ferruginea* yet this did not appear to affect the amount of pollen coming in from outside the stand. However, gene flow estimates show that the mean pollen dispersal distance significantly decreased in secondary forest populations of *V. ferruginea*, inferring that, within populations, pollinators may respond differently to *V. ferruginea* in secondary and primary forest.

Chapter 6: Using Seed Dispersal Modelling to Investigate the Potential Genetic Consequences of Recolonisation

6.1 Introduction

As illustrated in previous chapters, the processes involved in dispersal and recruitment of a plant species can have a significant impact on the genetic structure of resultant populations, which will be a persistent feature of tree populations where individuals are long lived. For example, small differences in the number and spatial distribution of founding individuals may lead to population level differences, as illustrated by the difference in F_{IS} for secondary forest populations of *Vochysia ferruginea* (Chapter 3). The preceding chapters also show that information from molecular markers can be used to infer the dynamics of founding processes. Another method that can aid such understanding is the use of simulation modelling, where gene flow of a regenerating population can be modelled over a simulated landscape using input parameters that are both known and controlled.

The following two chapters use a modelling approach to investigate the genetic consequences of recolonisation and regeneration on forest tree populations. This chapter focuses on the gene flow via seed dispersal. A simulation model was developed to investigate the recolonisation of a tree species and the effects of gene flow via seed dispersal on maternally inherited, organelle population genetic structure. By simulating the seed dispersal component of gene flow alone the model was simplified and could be more thoroughly tested. This model can be applied at large spatial scales and was used to examine the post-glacial recolonisation of oak across Britain. Although this is a temperate species, access to data on the cpDNA structure of oak populations provided the opportunity to compare model outputs to a real genetic

system. In the next chapter the model was extended to investigate the genetic consequences of combining both pollen and seed dispersal, but was limited to the population and meta-population scale.

Plant species exhibit a variety of mechanisms of seed dispersal. The majority of seeds are dispersed locally around the maternal parent and further secondary dispersal occurs in relation to the speed and direction of wind for wind-dispersed seeds, the crown size for gravity-dispersed seeds and disperser behaviour for animal dispersed seeds (Howe and Smallwood 1982). There is also potential for long distance dispersal through a variety of mechanisms. These include rare climatic events (e.g. hurricanes), long distance animal-dispersal (e.g. migratory birds) or long distance river transport (Janzen 1984; Sorensen 1986; Sauer 1988; Chambers and MacMahon 1994).

Long distance events are not normally measured in empirical studies of seed dispersal. Even with the recent increase in molecular studies targeting maternally inherited, and therefore seed-dispersed, genomes (e.g. chloroplast DNA (cpDNA) in Angiosperms), sample sizes are usually too small to reveal detailed information (i.e. a direct estimate of dispersal) on the dynamics of these potentially very rare events. However, the importance of rare long distance dispersal events to the colonising ability of a species has been highlighted by a number of studies, and has been influential in determining colonisation patterns following glacial periods (Skellam 1951; Hewitt 1996; Ibrahim *et al.*1996; Le Corre *et al.* 1997; Cain *et al.*1998; Clarke 1998;Clarke *et al.* 1999; Cain *et al.* 2000; Petit *et al.* 2001). For forest trees, the timing and direction of postglacial colonisation have been investigated using analysis of fossil pollen data and simulation studies (e.g. Le Corre *et al.* 1997; Birks 1989; Hewitt 1996; Hewitt 1999). These

studies indicate that expansion from refugial centres following glacial retreat is often rapid and must have involved some component of long distance dispersal.

The postglacial colonisation of oaks across Britain provides a good case study to examine the influence of long distance dispersal on migration rates. Within the British Isles there are no recorded refugial oak populations and there is a very good pollen record against which the timing of events can be set (Brewer *et al.* 2002). Colonisation speed has been determined from pollen core analysis and shows that the spread of oak was particularly rapid, reaching speeds up to 500 m yr^{-1} (Huntley and Birks 1983; Birks 1989; Bennett 1997). Oaks, like other forest trees, are long-lived and therefore their present genetic variability is likely to reflect ancient population and colonisation events (Petit *et al.* 2000). Thus the colonisation history of oak can be uncovered by examining cpDNA variation, since this organelle genome is maternally inherited and non-recombinant (Ferris *et al.* 1993; Dumolin *et al.* 1995; Ferris *et al.* 1995). Recent studies have provided detailed spatial information on the distribution of cpDNA variability in European oak populations (e.g. Ferris *et al.* 1993; Ferris *et al.* 1995; Petit *et al.* 2002a; Cottrell *et al.* 2002). The results of phylogeographical analysis of 2613 European populations screened for cpDNA variation show a strong phylogeographic structure with several East-West disjunctions in variation and multiple clines relating to colonisation routes from three refugial populations in the southern peninsulas of Iberia, Italy and the Balkans (Hewitt 1999; Brewer *et al.* 2002; Petit *et al.* 2002b). In addition, most mutations appear to have been generated prior to postglacial expansion (Ferris *et al.* 1993, Ferris *et al.* 1995; Hewitt 1999; Petit *et al.* 2002b) so present cpDNA structure is most probably a consequence of dispersal of pre-existing variation, not mutation. Such detail on the spread of oak makes it a system amenable to mathematical modelling.

Models for the range expansion of oak, and other plant species, historically used normal or short tailed dispersal functions fitted to local data on seed dispersal (Skellam 1951; Cain *et al.* 1988; Birks 1989; Higgins and Richardson 1999). However, these studies fail to capture the rate of postglacial expansion observed in plant populations. More recently, simulation models have been used to evaluate the importance of long distance dispersal in the spread of plants. Comparisons of different dispersal kernels show that leptokurtic dispersal kernels (where the tail of the distribution is positive at large distances) are required to produce the colonisation speeds seen during postglacial expansion of many plant species (Shigesda 1995; Hewitt 1996; Ibrahim *et al.*1996; Le Corre *et al.* 1997; Clarke 1998; Cain *et al.*1998; Clarke *et al.* 1999; Cain *et al.* 2000; Austerlitz and Garnier-Géré 2003). Stratified dispersal kernels use two components to describe dispersal, one represents the distribution of the majority of seed around the maternal parent and the second component represents rare long distance dispersal. It is argued that it is better to represent dispersal by two different mechanisms so that the central tendency and tail can be varied independently (Clarke 1998; Clark *et al.* 1999; Pakeman 2001). Stratified dispersal kernels have been used to realistically simulate the ecto/endozoochorous dispersal of woodland herbaceous species (Pakeman 2001); the rapid colonisation of the forest herb *Asarum canadense* (Cain *et al.* 1998); the behaviour of *Hieracium pilosella* (Winker and Stocklin 2002) and the dispersal of *Calluna vulgaris* and *Erica cinerea* by wind (Bullock and Clarke 2000). Simulation models based on the oak system concluded that long distance dispersal events not only allow for rapid expansion but they are also the most important factor determining the spatial genetic structure of maternally inherited genes (Le Corre *et al.* 1997; Petit *et al.* 2000). What is now required is an understanding of how the dynamics produced

by leptokurtic seed dispersal interact with other physical and environmental factors to influence the genetic structure within populations.

6.1.1 The study and model system

This study uses a spatial simulation model to investigate the effect of long distance dispersal on the spread and genetic structure of plant populations under different dispersal and landscape scenarios. These scenarios have been developed with particular reference to the colonisation of oak in the British Isles, but will also be relevant to other species with leptokurtic seed dispersal patterns. A key difference between this study and previous simulations of plant dispersal (Hewitt 1996; Ibrahim *et al.* 1996; Le Corre *et al.* 1997; Petit *et al.* 2000) is the spatial scale employed. Utilization of spatial models in previous studies simulated large grid areas (100 km by 300 km in Le Corre *et al.* 1997, 100 km by 1320 km in Petit *et al.* 2000), but the number of cells within the study area, chosen for computational efficiency, was relatively small and thus simulated cells represent demes rather than individual trees. These studies therefore made *a priori* assumptions about genetic dynamics within each deme. This study considers a similar landscape area (300 km by 500 km) but divides this area into patches that are 50 m by 50 m in size, and typically occupied by a single tree (Cottrell *et al.* 2003), and so the genetic dynamics and patterning produced by individuals can be modelled directly. Such fine-scale modelling is now possible due to the increased processor power available and by approximating one of the spatial dimensions by using rolled transects (see discussion of Figure 6.1 later).

This study also investigates a number of other potentially important influences on dispersal (and resulting colonisation speed and genetic structure), which remain untested by previous work (e.g. catastrophic climatic events, barriers to dispersal and environmental effects on fecundity and survivorship).

The model is parameterised to approximate the dynamics for oak populations in Britain and a number of scenarios are considered and assessed for genetic structure and colonisation rate, as follows:

1. *Default*: continuous long distance dispersal occurring at low probability.

2. *Hurricanes*: long distance dispersal operates in certain years only, but with the same average probability of long distance dispersal over time as in the default run. This represents freak climatic events that may be capable of dispersal over long distances (e.g. rare storms, hurricanes) and can be used to test the robustness of genetic structure and colonisation rate to the period of long distance dispersal events.

3. *Staggered start*: where a time lag operates between the initial seeding of different haplotypes. Analysis of the pollen core record (Birks 1989; Brewer *et al.* 2002) and current cpDNA distribution (Cottrell *et al.* 2002) indicate that oak with a single haplotype first entered south west Britain around 9500 BP, with the other two haplotypes colonising the south east later.

4. *Barrier*: part of the landscape is made uncolonisable. Many of the mountainous areas in Britain are likely to have presented significant barriers to the dispersal of oak; e.g. those greater than 350m in height such as the Welsh mountains, Pennines, Borders and Grampians. Today the British Isles is a series of islands off the European continent, however, for much of the post-glacial period Britain and Ireland were part of the mainland. It is probable sea-bed depths which are currently less than 100 m were land and could support oak populations during the major wave of colonisation

(between 10 000 and 8000 YBP). However some significant marine dispersal barriers remained, especially the northern and central Irish Seas.

5. *Latitudinal climate change*: to model the effects of climate on fitness, fecundity and/or seed survival is reduced as latitude increases. This simulates a decrease in fitness as environmental conditions become less favourable as a species reaches its range limits. This will allow us to test whether latitudinal climate effects on fitness can account for the lower diversity and lower differentiation observed in the more northerly, slower colonised populations (Cottrell *et al.* 2002). It is acknowledged that climatic conditions may have varied at all latitudes during the colonisations of oak and that this may have impacted on the genetic composition currently observed. This is not considered here; instead the focus is on how the difference in climatic conditions between northerly and southerly latitudes affects fitness.

These experiments allow us the unique opportunity to test the effects of different mechanisms on colonisation speed and population diversity and differentiation. By conducting these tests at a scale roughly equivalent to the size of England, and with dispersal attributes approximating those of oak it will be possible to hypothesize which mechanisms were most influential in producing the distribution of cpDNA diversity observed in British oak populations. However, the results should only be considered qualitatively, to assess trends in the effects of different mechanisms and should not be considered a wholly accurate representation of the oak system.

6. 2 Methods

6.2.1 Population and dispersal parameters

Seed dispersal was assumed to occur annually and was represented by a stratified two-component dispersal kernel (Clarke 1998; Le Corre *et al.* 1997), to approximate local and long distance dispersal. The proportion of local dispersal (p) to long distance dispersal (1-p) can be altered as can the maximum distance for dispersal (d_{max}). Local dispersal was assumed to occur within a 150 m radius and is represented by a linear (decreasing) probability generating function (see equation (6.1)). It is difficult to find estimates for dispersal parameters for colonising oak populations and so the estimates for local dispersal are derived from cpDNA data from established populations (Cottrell *et al.* 2002, J. Cottrell and S. Gerber pers. comm.). Long distance dispersal is represented by a uniform probability generating function between 150 m and d_{max} (see equation (6.1)). There was no data on the maximum distance that an acorn may be dispersed, and so d_{max} is estimated around a maximum of 25-130 km (Petit *et al.* 1997; 2000) and accounts for various mechanisms that may be responsible for rare seed dispersal up to this distance. The probability generating function, y, with distance, x, from the parental plant is represented as follows.

$$y(x) = \begin{cases} \dfrac{2p}{150} - \dfrac{1-p}{d_{\max} - 150} - \dfrac{x}{150}\left(\dfrac{2p}{150} - \dfrac{2(1-p)}{d_{\max} - 150}\right) & x \le 150 \\ \dfrac{1-p}{d_{\max} - 150} & 150 < x \le d_{\max} \end{cases} \qquad (6.1)$$

Maturity of trees was set at 15 years with an average lifespan of 200 years and a maximum life span of 400 years. Each mature tree was assumed to have a mean total number of offspring (N_0) each year (seeds that will successfully germinate), if unaffected by competition from other individuals for free grid cells. The proportion of

seeds dispersed locally was p, therefore the mean number of local offspring was pN_0 and the mean number of long distance offspring was $(1-p)N_0$. The dispersal kernel then determined the distance an offspring was located from its parent, and this, when combined with the dispersal direction (chosen at random between 0 and 360°), determined the grid cell where the offspring potentially grew. If the selected grid cell was vacant then the seed became a tree. If more than one successful seed was predicted to reach an empty cell then the resulting offspring was chosen at random, based on proportions of the seeds present. All offspring had the same chloroplast haplotype as their mother.

6.2.2 Spatial set-up

An individual based, spatial simulation model with a spatial scale relevant to the scale of England (300 km by 500 km) was used. This scale was too large to cover with grid cells so instead cells were modelled along three 10 km wide by 500 km long 'latitudinal transects' (see Figure 6.1), divided into 50 m square grid cells that can be colonised by a single tree.

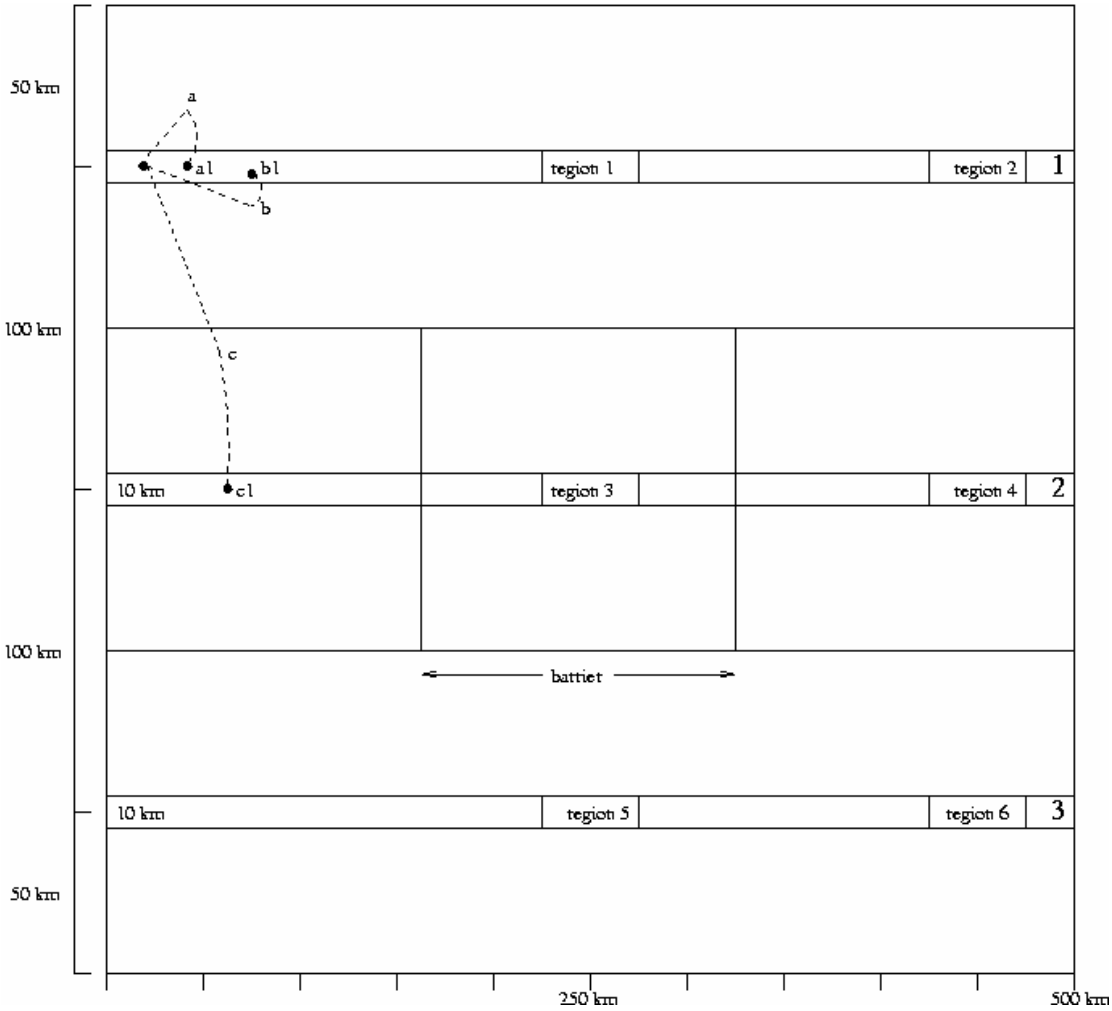


Figure 6.1: Spatial set up and positioning of the three latitudinal transects. The figure shows the effect of rolling the edge of the latitudinal transects on dispersal, the regions used for data extraction (region 1-6) and position of the barrier to colonisation (see later). Rolling the transects allows for representation of dispersal from trees within the 100 km x 500 km sections but outside of the 10 km x 500 km latitudinal transects. For example, if a seed is dispersed to point a, it will be dispersed into the latitudinal transect at point a1 (similarly for b to b1). If a seed is dispersed sufficiently far that it reaches the adjacent section such as point c it will be dispersed to the latitudinal transect in that section at point c1. Seeds that disperse outside of the 300km x 500 km arena are assumed lost. The left hand side of the figure represents south and the right hand side north, thus latitude increases from left to right.

The three latitudinal transects were positioned horizontally in the centre of three adjacent 100 km x 500 km sections. If a dispersal event moves a seed outside of a

latitudinal transect, but not out of the section along which the transect runs, the seed is repositioned into the original latitudinal transect (Figure 6.1, seeds a, b). If a long distance dispersal event occurred that was sufficient to cross the boundary into the adjacent section, the seed was re-positioned into the latitudinal transect in that section (see seed c in Figure. 1). This spatial set up allowed a representation of a very large area while still retaining the detail of an individual based model. The level of competition experienced by individuals when using this method compares favourably to tests undertaken using a 'full' version of the model (where the whole area contains grid cells). This is expected since while migrants are not considered from individuals outside of the three transects, there is a counterbalancing effect by reinserting migrants which leave a transect (i.e. the rolling process). The transect method described here has been compared with a 'full' version of the model (albeit over a smaller 300 km by 150 km scale) and the results for the genetic statistics and colonisation speed in the full model fall within the range of the five replicates using the rolled transect method (with the default parameters) and close to the average values over the five replicates (data not shown).

6.2.3 Initial conditions

In a nationwide study of British oaks (Cottrell *et al.* 2002), 98% of all trees sampled had one of three haplotypes of Iberian origin with the remaining 2% consisting of two different haplotypes of possibly non-natural origin; therefore, three haplotypes were used in the model to simulate the natural oak cpDNA diversity in England. Each latitudinal transect was seeded with one haplotype to randomly fill 2% of the first 10 km approximating the clinal chloroplast distribution in the north of France/south of England prior to colonisation (Cottrell *et al.* 2002; Petit *et al.* 2002b). These trees were assigned a random age up to the maximum of 400 years (see section 6.2.1) and N_0 (number of offspring) was chosen to be 1.

The following terms are used throughout the analysis and are defined below for clarity.

Region: Haplotypes were sampled in six regions as outlined in Figure 6.1.

Population: 20 populations were sampled from different parts of each region.

Latitudinal transect: The three, rolled 10 km by 500 km transects (see Figure 6.1).

Column 1 (Column 2): The combination of regions 1, 3, 5 (2, 4, 6) (see Figure 6.1). Statistics were either averaged over these regions or the populations from these regions were pooled and statistics determined for the combined populations.

6.3 Preliminary study

Initial tests were performed on the default set-up of the model to determine the values of d_{max} and p that result in colonisation speed and genetic structure that best represents British oak populations. Many dispersal factors will have influenced the current genetic structuring of oak populations (not just those in the default set-up) and so it was more important to focus this study on understanding how the results compare with the default scenario rather than how the different scenarios compare to current data. For the initial tests a single run of the model for 2000 years was used with three different lengths for maximum long-distance dispersal (d_{max} = 60, 80 and 100 km) in combination with two different probabilities for long distance dispersal ((1-p) = 0.0001 and 0.001). The speed of colonisation was calculated by determining how long it takes for a column in the grid to become 10% colonised, which approximately corresponds to the likelihood of a pollen grain being present in a pollen core sample. Diversity statistics were determined according to Nei 1973 (see also Pons and Petit 1995) for average population (H_S) and total (H_T) genetic diversity and differentiation (G_{ST}) at a number of patch sizes to determine which was the most appropriate sample scale. Model data were extracted from regions 2, 4 and 6 (Figure 6.1), with 20

populations of 10 individuals randomly sampled for 10 different patch sizes in each region.

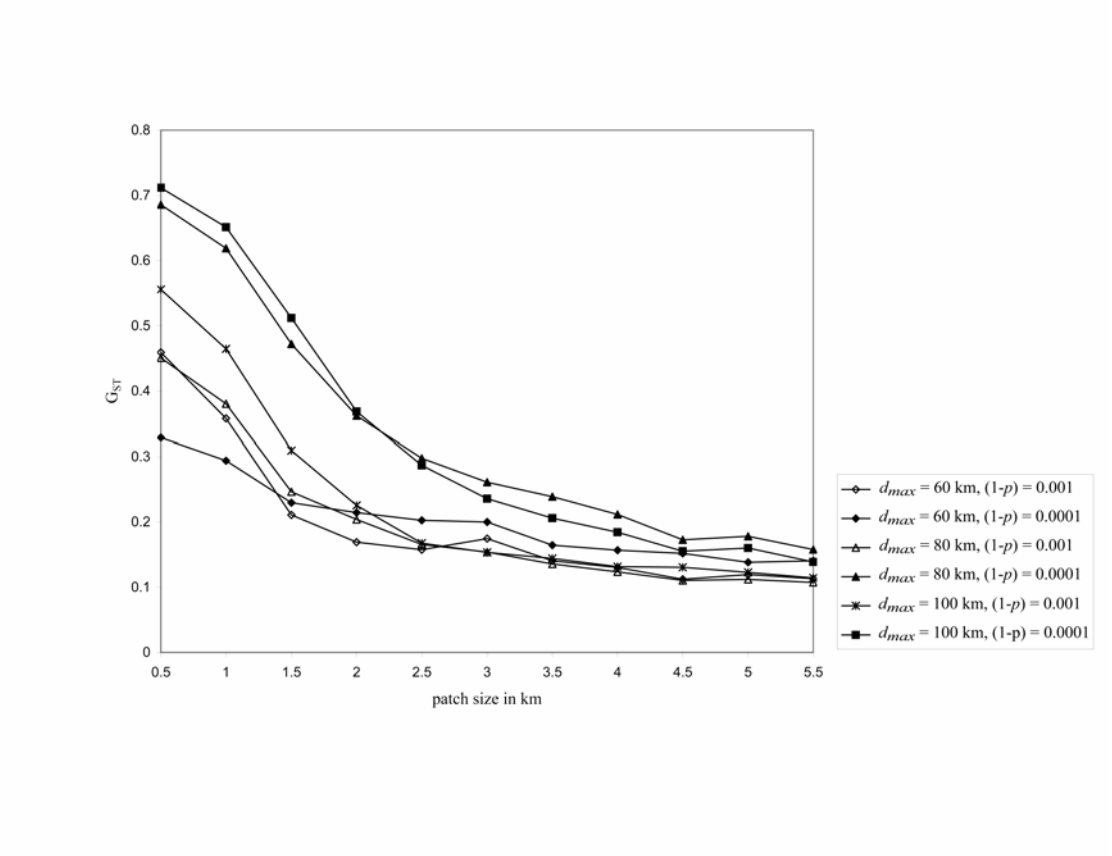


Figure 6.2: Changes in G_{ST} with patch size at different lengths of long distance dispersal (dmax = 60, 80 or 100km) and with a probability of 0.001 or 0.0001.

Changes in differentiation for varying patch size indicate the scale of haplotype aggregation. An ideal sampling strategy is one that picks up the differences between the diversity and differentiation measures, by sampling on a scale within haplotype patches and comparing samples on a scale that is greater than the distance between haplotype patches (an area containing trees of a single haplotype). Changes in G_{ST} for different d_{max} and p values (Figure 6.2) show that G_{ST} was still increasing at the smallest patch size in all runs, therefore haplotype patch size was approximately 0.5 km^2 or below (sampling from a population in a patch size smaller than this would have been problematic due to the small number of trees) and differentiation levelled

out for a patch size greater than 3 km. Therefore, results were sampled using a patch size of 0.5 x 0.5 km with greater than 3 km between populations sampled. This is also roughly equivalent to the scale on which experimental populations were sampled (Cottrell *et al.* 2002).

Table 6.1: Speed, diversity and differentiation for different probabilities of long distance dispersal and at different lengths of maximum dispersal. Populations sampled at regions 2, 4 and 6 (between 400 – 450 km) after 2000 years. Colonisation to Britain had an average speed of approximately 500m/year and current oak populations have an average H_T of 0.6 with a G_{ST} of 0.75 (Cottrell *et al.* 2002).

Maximum length of dispersal		Probability of long distance dispersal (1-*p*)			
		0.001		0.0001	
		Average	Pooled	Average	Pooled
60 km	Speed (m/year)	523		278	
	H_S	0.18	0.18	0.05	0.05
	H_T	0.27	0.65	0.14	0.65
	G_{ST}	0.27	0.71	0.22	0.92
80 km	Speed (m/year)	713		348	
	H_S	0.24	0.24	0.13	0.13
	H_T	0.41	0.65	0.39	0.65
	G_{ST}	0.40	0.63	0.69	0.80
100 km	Speed (m/year)	937		450	
	H_S	0.32	0.32	0.20	0.20
	H_T	0.55	0.63	0.58	0.61
	G_{ST}	0.41	0.49	0.65	0.67

Comparing the results in Table 6.1 to those for the British oak population (Cottrell *et al.* 2002), values of d_{max} = 100 km and (1-p) = 0.0001 provide the most concordance for colonisation speed and diversity measures between experimental and modelled populations, and the parameters were used for the remaining simulations (unless otherwise stated).

6.4 Results

For each scenario detailed below, five replicates of the model were ran for 2000 years.

6.4.1 Default scenario

Figure 6.3a shows that the haplotypes form a patchy structure with dominance of the initially seeded haplotype decreasing with latitude (over the five runs the average H_S = 0.16 in column 1 and 0.17 in column 2, with the range over the five replicates around these average values is shown in Figure 6.4a). Within-population diversity is low, indicating that a single haplotype dominates within each 0.5 km x 0.5 km sample patch. Average regional diversity is relatively high indicating that the regions themselves contain a mixture of the three haplotypes. There is a slight trend of increasing within-population and average regional diversity with latitude (average regional diversity in column 1 is 0.5 and 0.56 in column 2, Figure 6.4a).

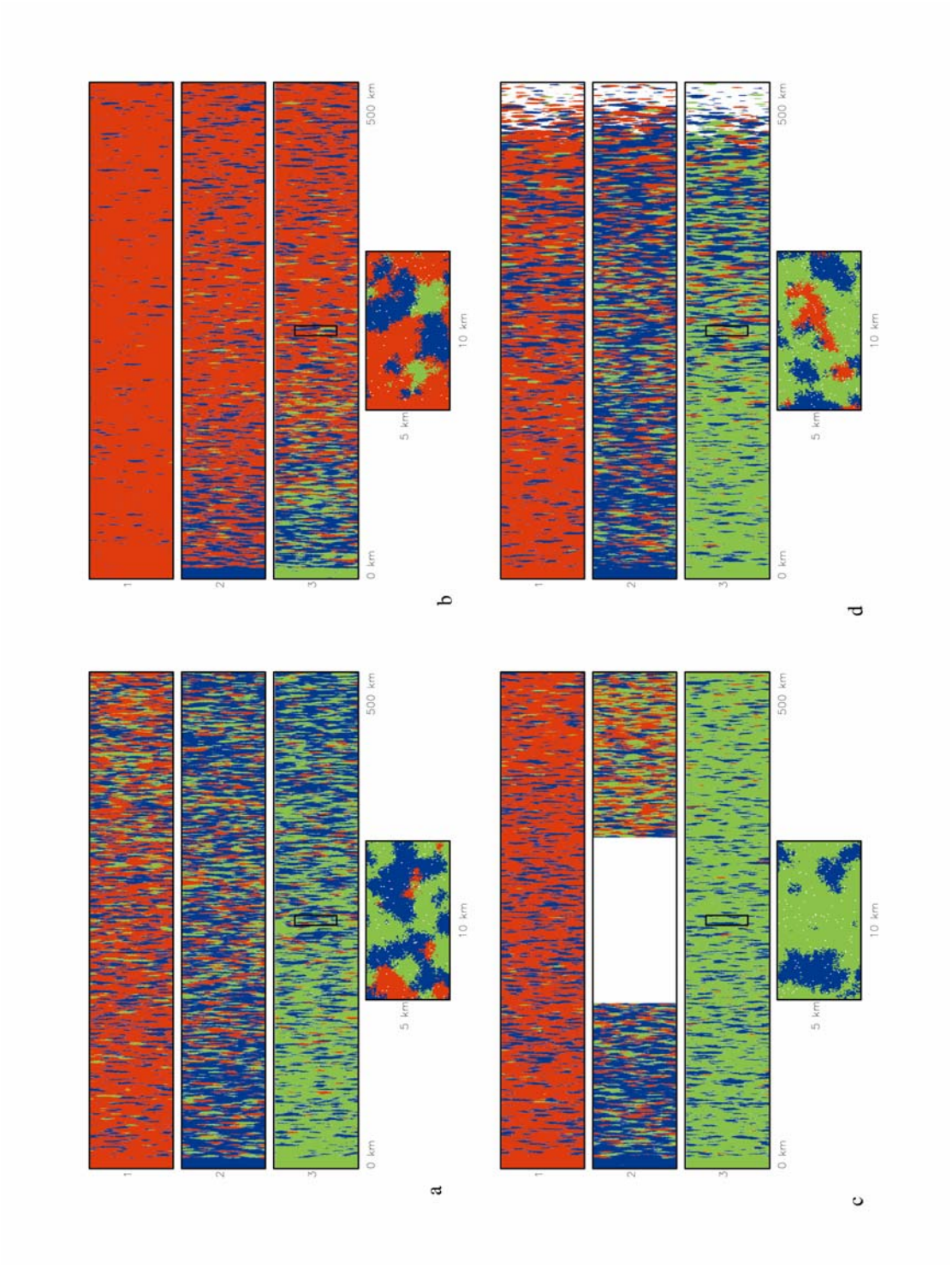


Figure 6.3: The spatial arrangement of the three haplotypes in the latitudinal transects after 2000 years. The relative scale of the transects (10 km x 500 km) gives a compressed appearance and so a small area is enlarged with the true scale on each axis showing the patchy genetic structure. Here (a): default scenario; (b): 100 year staggered start scenario; (c): barrier scenario; (d): decreased seed survival with latitude.

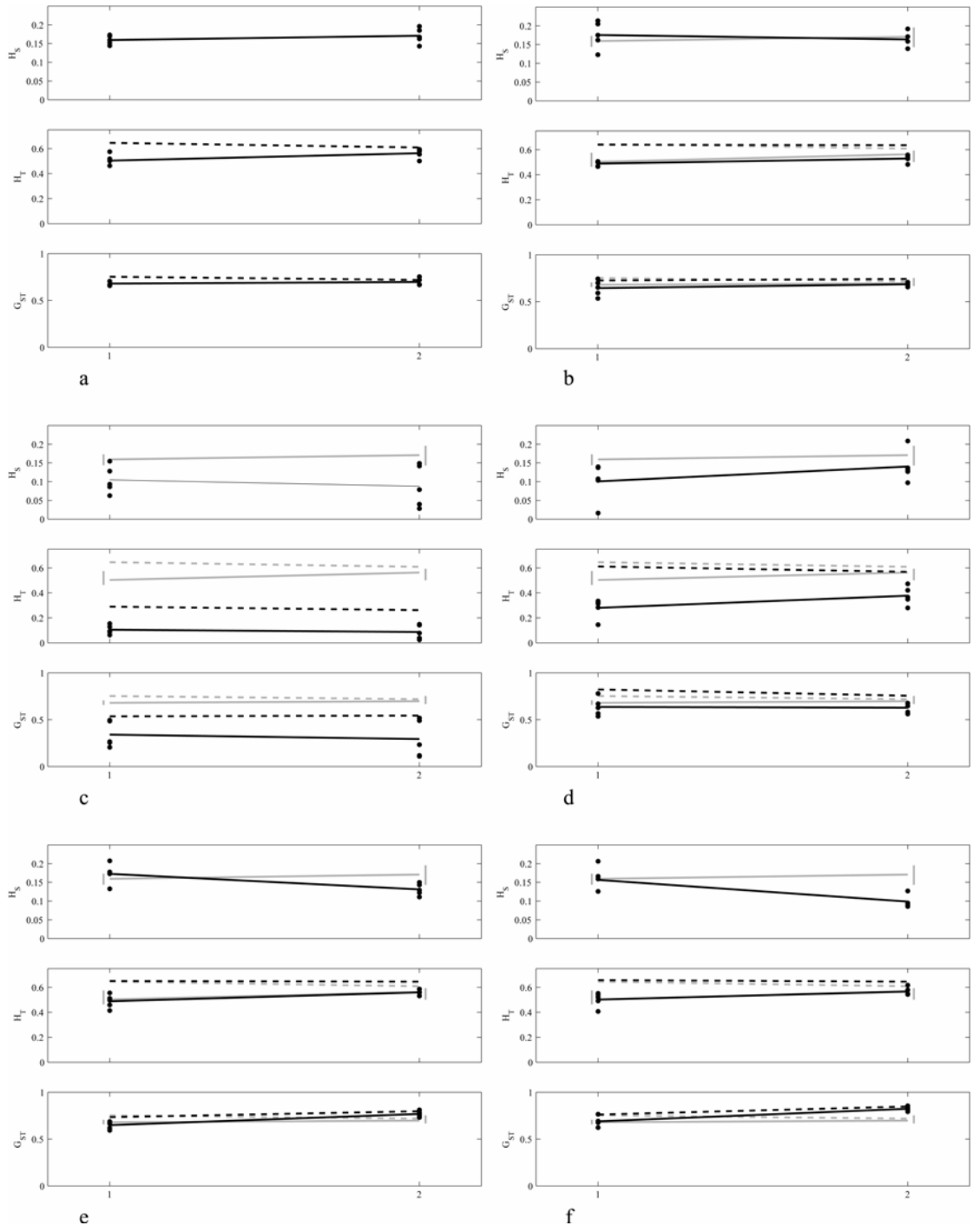


Figure 6.4: H_S, H_T, and G_{ST} values for the average within-region and pooled populations from all three regions in column 1 and column 2. The regional average results for each of the five replicated simulations are represented by a point (•) and the average is shown by a solid line (—). The average of the five pooled statistics is shown by a dotted line (--). Here (a): default; (b): hurricane; (c): 100 year stagger; (d): barrier; (e): decreased fecundity with latitude; (f): increased seed mortality with latitude. In (b) to (f) the default results are added in grey for comparison. The sidebars show the range of

variation in the default runs. For each individual replicate (in all scenarios) the 95% confidence range was small (and within ± 0.01 for H_S, ± 0.005 for H_T and ± 0.02 for G_{ST}).

Total diversity for the pooled populations in columns 1 and 2 is close to its maximum value indicating a near equal proportion of the three haplotypes over the three transects. The individual regions and pooled populations both show genetic structure due to the spatially patchy nature of the populations with the average within-region $G_{ST} = 0.68$ and pooled $G_{ST} = 0.75$. The sample populations in each region are dominated by one haplotype but the regions themselves contain a mixture of all haplotypes, so differentiation remains high within regions.

6.4.2 Hurricane scenario

The hurricane scenario uses the same parameters as the default scenario except that the proportion of long distance dispersal (1-p) = 0 for 99 years and 0.01 for 1 year in every 100. Thus, the amount of long distance dispersal over a 100-year period is the same as in the default scenario but the actual event occurs at periodicity of 100 years. The diversity and differentiation statistics for runs with this hurricane effect were in close agreement with those of the default scenario (Figure 6.4b). This indicates that rare, periodic instances of long distance dispersal of the same additive total magnitude have a similar effect on genetic structure as the continuous low-probability of long distance dispersal events of the default scenario.

6.4.3 Staggered start scenario

The set-up is the same as in the default runs except that, in the first year only the uppermost latitudinal transect is seeded (2% of the first 10 km of the transect are randomly seeded with haplotype 1, as in the default run). After 100 years the central latitudinal transect is seeded with haplotype 2 and after a further 100 years the bottom latitudinal transect is seeded with haplotype 3. Figure 6.3b shows the spatial

arrangement of the three haplotypes with this 100-year stagger. Clearly, the first haplotype given a 'head start' can dominate in all transects and the last haplotype to be seeded exists only in a few isolated patches. A staggered start results in a reduction of within-population and total diversity well below that exhibited in the default runs (Figure 6.4c). There is also a trend of decreasing diversity with latitude (a reversal of the default scenario trend), as the haplotype seeded first becomes more dominant with latitude (Figure 6.4c). Tests were also undertaken with a reduced stagger time of 50 years. As the stagger was reduced the dominance of the initially seeded haplotype was also reduced and genetic structure tends to that of the default runs.

6.4.4 Barrier scenario

A barrier to dispersal was simulated by making one area in the array uncolonisable (the central third of the central transect, see Figure 6.1). In this area trees could not grow but if seeds were dispersed far enough and could pass over the barrier they could establish on the other side (i.e. similar to mountain ranges, seas). This scenario shows genetic structure similar to the default runs up until the barrier (Figure 6.3c). The barrier then greatly reduced the dispersal of haplotype 2 (blue) and prevented colonisation of haplotype 1 into transect 3 (and vice versa) so that transects 1 and 3 were largely dominated by the original seeded haplotype. Haplotype 2 was maintained, in isolated patches, in transects 1 and 3. After the barrier, the central transect was dominated by a mixture of haplotypes 1 and 3. The statistics highlight this with greatly reduced diversity in column 1 (since populations were sampled from transects 1 and 3 and both were dominated by a single haplotype), with an increase in diversity in column 2, due to the mixing of haplotypes in the central transect after the barrier (H_T in column 1 was 0.28 and 0.38 in column 2). Throughout, the diversity is greatly reduced from the default scenario (Figure 6.4c, in default runs H_T was 0.5 in column 1 and 0.56 in column 2).

6.4.5 Latitudinal scenarios

Two potential effects of climate change with latitude are modelled. In the first, fecundity decreases with latitude, with average number of offspring decreasing linearly from 1 to 0.1 over the latitude range. In the second effect, survival of seeds decreases with latitude. Here an additional survival probability for successfully dispersed seeds was applied and varied linearly from 1 to 0.1 over the latitude range. Genetic diversity and structure for these scenarios are in close agreement with each other and the default scenario (Figure 6.3d, Figures 6.4e and 6.4f). There is however, a key difference for within-population diversity; the two scenarios compare closely with the default run in column 1 (H_S in default runs was 0.16 and was 0.16 with the fecundity effect and 0.17 with the seed survival effect) but both show a clear decreasing trend with latitude and a reduction of diversity in column 2 (in default runs within-population diversity increased to 0.17 but decreased to 0.13 with the fecundity effect and to 0.10 with the seed survival effect).

6.4.6 Speed of colonisation

Figure 6.5 shows the position of the leading edge of the dispersing tree population over time. Dispersal speed is similar for the default, hurricane, staggered start and barrier scenarios, and in all of these scenarios speed of advancement is constant and approximately 500 m yr^{-1}. Thus these factors do not affect the rate of spread. The rate of population spread is reduced in the scenarios where latitude affects fecundity and seed survival. Here the colonisation speed decreases with latitude. The decrease in speed is most severe when latitude affects seed survival.

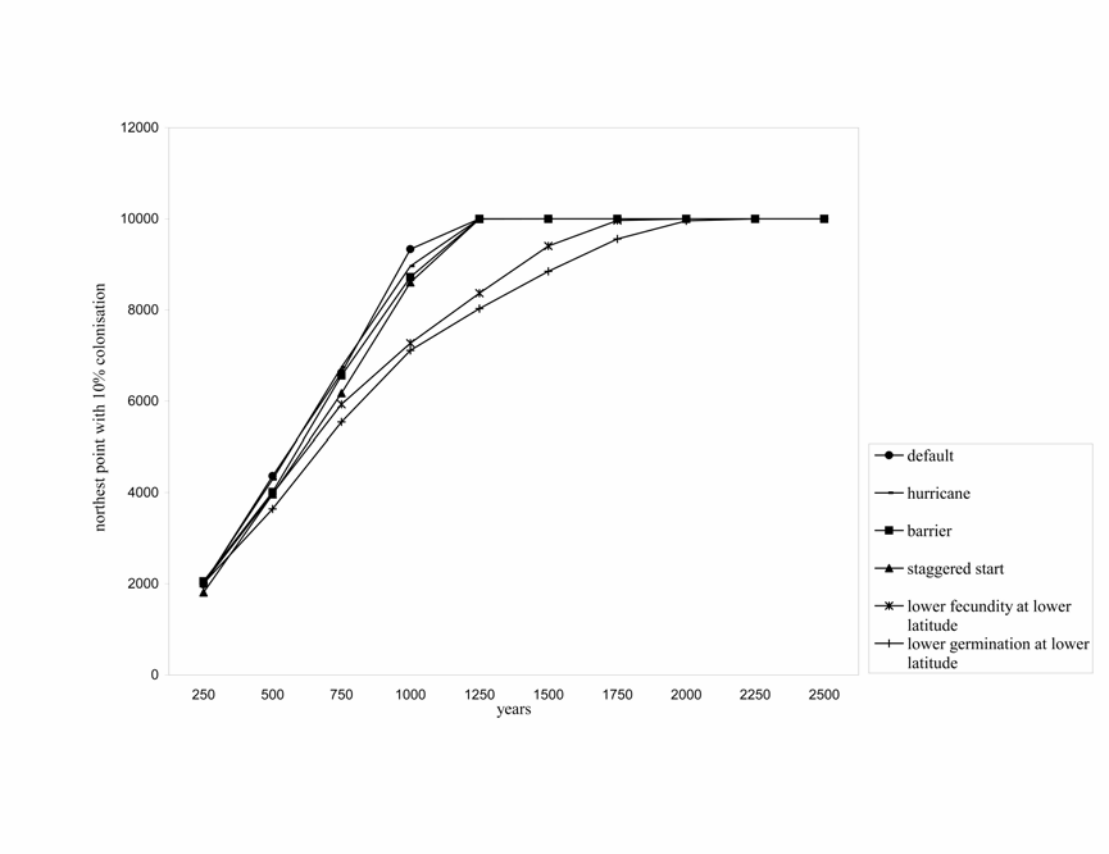


Figure 6.5: Position of the leading edge of the dispersing tree population over time for the different scenarios.

6.5 Discussion

6.5.1 Changes in colonisation speed and genetic structure with varying long distance dispersal parameters

The individual based model allowed assessment of how rare long distance dispersal events affect genetic structure and colonisation rates in plant populations. This study showed that increasing the maximum length of dispersal and the probability of a long distance dispersal event substantially increased colonisation speed (Table 6.1). These effects allow more dispersal between latitudinal transects and so also increase average diversity within a population (H_S) and the diversity of the region (H_T) (Table 6.1). When within-population diversity is low and differentiation within a region is high it indicates that the genetic structure of the region contains patches dominated by a

single haplotype and this patch size is greater than the area of the population sampled. Le Corre *et al.* (1997) proposed that patch size is indicative of the distance of long distance dispersal and Ibrahim *et al.* (1996) proposed that this was a result of a greater lag between populations formed by longer dispersed migrants and the advancing front of colonisation. Our results confirm this but also suggest that patch size is most sensitive to the proportion of long distance dispersal. A lower proportion of long distance dispersal generated a cpDNA structure with larger scale patchiness since successful long distance migrants will dominate local populations with less chance of competition from other long distance dispersers.

6.5.2 Scenario effects on colonisation speed

The analysis in this study shows that long distance dispersal is critical for producing speeds of colonisation on the same scale of oak recolonisation to Britain (see Table 6.1). The speed of colonisation remains constant and is determined by the frequency and length of rare long distance dispersal events. This is in agreement with other models of seed dispersal (Hewitt 1996, Ibrahim *et al.*1996, Le Corre *et al.* 1997, Clarke 1998, Clarke *et al.* 1999, Bullock and Clark 2000, Clarke *et al.* 2001, Pakeman 2001, Petit *et al.* 2001). The results from this study additionally show that the speed of colonisation is robust to changes in the periodicity of dispersal, to a barrier of seedling establishment and to a staggered start of colonisation. Colonisation speed is also reduced when a latitudinal gradient of decreasing fecundity or decreasing seedling survival is applied. Both decreased fecundity and decreased seedling survival effectively reduce the number of successful long distance dispersal events and further demonstrate the strong relationship of rare long distance dispersal to colonisation speed. The pollen core data shows a slowing in the speed of recolonisation of oak to Britain (Birks 1989). Such a pattern could result from climate conditions that become

less favourable with latitude or due to climate change slowing during the period of colonisation.

6.5.3 Scenario effects on genetic structure

In model populations, the genetic structure resulting from long distance dispersal consists of mosaics of patches of a single haplotype. This occurs as long distance dispersal allows the establishment of individuals ahead of the main colonisation front, followed by secondary local dispersal resulting in permanent patches of migrant haplotypes. Other simulated populations show patches of highly inbred demes that persist for hundreds of generations in spite of gene flow into patches (Ibrahim *et al.*1996, Le Corre *et al.* 1997, Petit *et al.* 2002). The stability of these migrant patches results in highly differentiated populations. High levels of differentiation are observed in oak populations (Petit *et al.* 2002b) and are therefore in agreement with a genetic structure dominated by the successive founder events of long distance dispersal.

At more northerly latitudes there was less dominance of the initially seeded haplotype and the genetic structure of the array became more influenced by long distance dispersal, either directly or secondarily from the stable patches of migrant haplotypes. Therefore, distance between refugial populations and established colonised populations may be important in determining genetic characteristics in post colonisation populations. Cottrell *et al.* (2002) found notable clumping of British populations with 66% of mixed species woods fixed for a single haplotype. This clumped pattern of haplotypes has also been found in populations in western France where areas of up to 50 km across are fixed for a single haplotype (Petit *et al.* 1997, Petit *et al.* 2002b). The genetic structure in model populations also remained stable over time, where populations in southerly latitudes maintain significant differentiation between transects after 2000 years (Figure 6.4).

Simulations changing the periodicity of long distance dispersal showed no real change in genetic structure. This implies that the genetic structuring resulting from long distance dispersal is robust to the frequency of dispersal events, providing the overall amount of dispersal remains the same, and allows a greater number of vectors of long distance dispersal: birds, animal and also rare events such as hurricanes, to be considered as important for rapid colonisation.

Staggering the start time of different haplotypes colonising the arena caused a dramatic change in genetic composition, with a decrease in population diversity as the earliest colonising haplotype dominated in all transects. The haplotype given a 'head start' will usually remain dominant in it's transect and had a greater opportunity to disperse to, and dominate, free colonisable space in the neighbouring latitudinal transect. However, late colonisers still maintained a presence after 2000 years through formation of stable fixed patches occurring from long distance dispersal events, even with a 200-year lag after initial colonisation. Haplotype structure in Britain shows a dominance of one haplotype along the western edge and in patches across the country increasing its dominance in northern England and Scotland (Cottrell *et al.* 2002). It may be that this haplotype, which dominates Cornish populations, may be that of the initial colonisation to Britain and has had an advantage during colonisation.

A barrier to colonisation restricted the spread of the haplotype seeded in that transect (Figure 6.1), but individuals possessing that haplotype managed to establish small stable patches through long distance dispersal (Figure 6.3c). Haplotypes seeded in the transects at either side of the barrier were also more restricted to their own transects than the default runs, as dispersal across the barrier is limited. In British populations,

one haplotype is more abundant in English populations than in Scottish populations. In the area around the Pennines the other two haplotypes dominate with one haplotype to the west and one to the east (Cottrell *et al.* 2002). It is possible that this land barrier also acts as a significant barrier to establishment and may partially account for the decrease in one haplotype in northerly populations and the predominantly east/west division of the two other haplotypes.

In model populations with a latitudinal climate effect, lower diversity resulted. Two potential reasons for this are that, either lower fecundity lessened the frequency of long distance dispersal, or the extra mortality means that long distance seed dispersal was less successful. Haplotype patch size was also larger, which is probably an effect of lower long distance dispersal success but maintenance of limited local dispersal. It was notable that patches of a single haplotype tend to become larger, particularly in northerly regions where the mortality effect is most severe. The average and pooled regional diversity is similar to the default runs reflecting the fact that although the local haplotype patch sizes has increased there is still a mixture of the three haplotypes across the transects. Therefore, latitudinal effects on seed fecundity or mortality do not alter regional diversity but do affect within-population diversity through the increased size of local patches. This could explain the results for British oak populations, which show that all haplotypes are present in Scottish populations, but there is a lowering of genetic diversity within populations with increasing latitude (Cottrell *et al.* 2002).

6.5.4 Leptokurtic seed dispersal effects on genetic diversity

During colonisation it is predicted that multiple founding events, a characteristic of leptokurtic dispersal, will show a strong decrease in allelic richness with distance from source that may not be mitigated by pollen dispersal in dense, fixed populations

(Nei *et al.* 1975, Hewitt 1996). These founder effects may be particularly severe in oak, as after the successful establishment of a colonising individual, most acorns dispersed by the tree germinate within its neighbourhood (König *et al.* 2002). In our model, population genetic structure was the product of multiple founder events from long distance dispersal with local dispersal confined to those cells surrounding the mother tree, yet within-population diversity increased as the proportion of long distance seeds dispersal increased (in contrast to other studies, e.g. Austerlitz and Garnier-Géré 2003). This study indicates that the creation of a patch structure from long distance dispersal of individual trees can maintain genetic diversity of populations because although the founder events of long distance colonisation may result in low within-population diversity, the stability of these patches allows maintenance of migrant haplotypes within the total population. Presence of long distance dispersal in model populations allowed all haplotypes to be present to some degree in the most northerly regions, even when there was some disadvantage to a haplotype due to late colonisation or a landscape barrier. European oak populations also show this maintenance of cpDNA diversity where haplotypes are not lost during spread of a few thousand km (Petit *et al.* 2002b) and all three haplotypes found in Britain are present to some degree in the most northerly populations (Cottrell *et al.* 2002).

Decreases in haplotype diversity have however, been reported for populations of forest trees following postglacial colonisation. For oaks, whilst 3 haplotypes of Iberian origin are found in Britain, there are 12 haplotypes in the potential source area (Cottrell *et al.* 2002, Petit *et al.* 2002a). The phylogeographic work on oaks (Cottrell *et al.* 2002, Petit *et al.* 2002) indicates that there has been relatively little loss of genetic diversity with northward migration of oaks. Most diversity is lost within the

first stages of colonisation, which coincides with negotiation of large physical barriers (Pyrenees, Alps, Carpathians) but once past these, almost all cpDNA types are then transferred to northern latitudes. Studies of cpDNA variation in other European trees have found 13 haplotypes in refugial populations of *Alnus glutinosa* with only 2 found in northern Europe (King and Ferris 1998), and samples from *Fagus sylvatica* consisted of 11 haplotypes in refugia but only 1 in northern populations (Demesure *et al.* 1996). Therefore, long distance dispersal may not be able to mitigate loss of alleles and diversity in certain circumstances. For instance, if environmental factors that can decrease diversity, such as a staggered start or a barrier, are severe (e.g. the Pyrenean mountain range, Petit *et al.* 2002b). Alternatively if southern populations are not able to spread beyond the northern leading edge (Hewitt 1999), as may be the case for *Fagus sylvatica,* then there is less opportunity for these long distance dispersal effects (Birks 1989; Hewitt 1999). There could also potentially be decreased diversity as a species colonises new habitat due to differences in colonisation ability or fitness differences associated with certain haplotypes. Therefore, the dispersal of oak to Britain illustrates that long distance dispersal has the potential to alleviate loss of alleles and therefore loss of diversity during colonisation; however, this is not evident in all postglacial colonising species so must be considered in the context of biological, physical and environmental factors relevant to any given species.

In summary, this study has confirmed findings of previous studies that long distance seed dispersal is crucial for producing the colonisation rates and the genetic structure observed in postglacial populations of European deciduous oaks. However, in addition, it is clear how combinations of physical and environmental mechanisms may explain the patterns of genetic diversity and haplotype structuring found in oak

populations in Britain. The results can be generalised to other plant systems in which rare long distance dispersal may be a feature.

Chapter 7: Using Seed and Pollen Dispersal Modelling to Investigate Genetic Consequences of Secondary Forest Regeneration - A Preliminary Investigation

7.1 Introduction

The value of secondary and logged forests as a conservation resource is becoming increasingly recognised, following a recent expansion in area of this habitat (Cannon *et al.* 1998; Chokkalingam and Jong 2001; Aide & Grau 2004). Secondary forests provide habitat for a range of native species, have proven to be important in reducing soil erosion and floods (Aide & Grau 2004), and are increasingly being recognised for their commercial potential, as a source of valuable timber products (Finegan 1992; Baluarte-Vasquez *et al.* 2000). However, for these forests to continue to perform their ecological functions whilst realizing their economic potential, they must be managed appropriately (FAO, FLD, IPGRI 2004). Thus, developing sustainable forestry techniques for tropical tree species in regenerated forest is an important consideration for forestry decision makers (Wright 2005; Montagnini & Jordan 2005).

One important aspect of successful sustainable forestry is consideration of genetic resources. To maintain species for long term commercial or conservation purposes, it is necessary to consider population genetics, since the level of genetic variation within populations is critical to how well that species will respond to environmental stress, competitors and diseases (Sherwin & Moritz 2000; FAO, FLD, IPGRI 2004). Management strategies to aid the maintenance of sufficient levels of variation within populations also require a consideration of the biological processes (e.g. survivorship, reproduction and dispersal) that determine the partitioning of diversity within populations (Loveless & Hamrick 1984; Loveless 2002; FAO, FLD, IPGRI 2004).

In Costa Rica, conservation is already an important part of forestry management, with about 25% of its land area protected, most of it covered by forest (Kleinn *et al.* 2005). However, further conservation of forest resources requires initiatives that are profitable and desirable to landowners. FUNDECOR (Foundation for the Development of the Central Volcanic Mountain Range), a non-profit foundation working with MIRENEM (Ministerio de Recursos Naturales, Energía y Minas) and the National Parks Service, develops regional forest management plans that aim to regenerate damaged forests and reforest areas that have previously been clear-cut. These management plans include selective harvest and natural regeneration as a means of sustainable and profitable forestry. In a recent report (Church *et al.* 1994), FUNDECOR selected eight native species, including *Vochysia ferruginea*, as effective species for reforestation of humid tropical areas of the central valley.

The increased interest in sustainable forestry is recent and little is known about the impacts of proposed sustainable forestry techniques on the demography and genetic resources of target species. There are a number of factors (e.g. level of inbreeding, fluctuations in gene flow) that may impact on the genetic characteristics of a population, but are difficult to determine and isolate in real systems. It is in such circumstances that mathematical modelling can be a useful tool (Alvarez-Buylla and Garcia-Barrios 1993). Models can be readily manipulated to determine the most important factors relating to genetic diversity and allow different scenarios to be tested.

In this chapter a preliminary model investigation is conducted to assess genetic diversity and structure, under different management strategies. By comparing results for different management strategies, we can assess their genetic resource impact and

determine the sensitivity of populations to different management regimes. By using *V. ferruginea* as a test species, the results will also be relevant for assessing genetic impacts in Costa Rican forest restoration schemes.

7.2 Modelling overview

The model considered here extends the seed dispersal model framework described in Chapter 6 to include pollen flow, and represents population and genetic dynamics at an individual scale. Similar individual-based approaches of seed and pollen dispersal are rare, most probably because the computational power required for such simulations has only recently become available. An exception is the spatial model ECO-GENE (Degen 1996) that investigates the impact of selective logging management alternatives on the genetics of tree populations over specified timescales. The ECO-GENE model combines population genetic processes and a population dynamic forest growth model. It has been validated using empirical observations (Degen 1996) and used to assess, for example, the effect of insect behaviour on pollen dispersal (Degen 2004) and the effect of selective logging on seed cohorts (Degen 2002). The ECO-GENE model has been developed to produce specific predictions when provided with detailed information about spatial location and age-structure for particular tree species.

This study is focused on understanding how forest management strategies affect genetic dynamics and specifically for secondary forest regeneration. The approach taken will be considerably simpler than that used in the ECO-GENE model (Degen 1996). In particular the model will not include a complex, species-specific population level framework, and neither population level dynamics nor comparison with empirical spatial observations will be attempted. The main benefit of this approach is that it does not require knowledge of the biological characteristics of a species, which

are largely unknown. This approach is also more flexible, allowing manipulation of specific initial conditions and is therefore more appropriate for the investigation of the population genetics of regenerating cohorts.

The strategy in this study will be to conduct a preliminary investigation into the 'natural' dynamics of forest systems and then compare the consequences of including different mechanisms, such as small scale natural disturbance and long distance components to seed and pollen dispersal. Using the model, the impact of different management scenarios on the population genetics of a regenerating species will also be investigated.

Initially the differences in genetic diversity and species composition in primary forest and naturally regenerating secondary forest with and without long distance dispersal via seeds and pollen are assessed. The results from the initial investigations will be used to inform the development of a default forest system to be used to test different management scenarios. The first management scenario represents clear-cutting a region of primary forest and regularly harvesting from naturally regenerated forest. The second management scenario is as the first but considers leaving remnant seed trees from which to regenerate the region. The patterns of genetic structure of the regenerated population when different numbers of seed trees are left after clear cutting will be compared. The third management scenario represents the regeneration of a region from seeds taken from trees within primary forest. The genetic structure of the naturally regenerated population will be assessed when seeds are taken from different numbers of seed trees. This will uncover the key effects of different management scenarios on genetic dynamics and allow for hypotheses testing about the most effective way to maintain diversity in natural systems. By using five replications for each scenario the model will also assess the variability of the outcome.

In this investigation the potential use of a modelling approach in considering the long term genetic resource effects that can arise from the management of long lived tree species is highlighted. Such models may ultimately be used to guide sustainable forestry techniques to maximise yield and genetic diversity.

7.3 Methods

7.3.1 Spatial set-up

The spatial set-up of the model consisted of a 1 km x 1 km array containing cells (10 m x 10 m) that could be occupied by a single tree. Initially the array was assumed to contain primary forest; this was then harvested or cleared and the array was allowed to regenerate. Scenarios were tested in which the array regenerated in isolation or with a second spatial 1 km x 1 km array of adjacent primary forest. The boundary conditions were set such that any seed attempting to disperse beyond the 1 km x 1 km array would be unsuccessful, unless it dispersed into the defined neighbouring array.

7.3.2 Population dynamics and dispersal

The model had two categorisations of tree, light demanding pioneers (LD) and shade tolerant species (ST). In this study we considered twenty species of tree – two LD species (one 'target LD' species which we considered to be representative of the dynamics of *V. ferruginea*, and a non-target LD species which had similar dynamics but for which genetic data was not recorded) and eighteen ST species. This gave an initial primary forest composed of 10% pioneer species and 90% shade tolerant species with each species at an equal starting density of approximately 5%. Tropical forests have a higher number of shade tolerant species than pioneer species (Wright 2003) and the abundance, and number, of pioneer species can vary due to ecological differences between sites (Condit *et al.* 1999; Hubbell *et al.* 1999). Estimates of the

density of *V. ferruginea* in primary forest were from Costa Rican sites in this study (C. Navarro pers. com.); where only two pioneer tree species dominated plots. The number of ST species was determined so that each species initially had equal density so there was no advantage to any species.

Each *Vochysia ferruginea* tree had a multilocus genotype of 5 codominant loci. There was no genetic information included for the non-target LD species or the ST species. All species had the same dispersal dynamics. In each generation (2 generations per year due to two fruiting seasons), mature adults could potentially disperse seed into all surrounding empty cells up to a seed dispersal maximum. In each empty cell the maternal parent was chosen based on a linear decreasing seed dispersal kernel, so that an adult mature tree near to an empty cell had a greater probability of becoming the maternal parent than one further away. Local dispersal decreased linearly up to 30 m. It was also possible to enforce a long distance component to seed dispersal. When long distance dispersal was included there was a probability of 0.01 that the maternal parent was chosen from a uniform distribution of 30m-100m from an empty cell (local dispersal as described above occurred with a probability of 0.99). The probability of becoming a maternal parent was also scaled for the competitive advantage of either LD or ST species and was based on the number of trees surrounding an empty cell (high numbers represented high shading from the canopy of surrounding trees/cells). A non-linear functional response represented shading such that when an empty cell was surrounded by 6 trees (out of a possible 8) there was no advantage to either type, therefore, light demanding trees had a colonisation advantage when there was lower numbers of surrounding trees and shade tolerant trees had an advantage when the number of surrounding trees increased.

If the maternal parent was a *V. ferruginea* tree then the paternal parent was determined based on a decreasing linear pollen dispersal curve up to 100m centred on the mother tree. If there were no target LD trees within this dispersal kernel then the mature *V. ferruginea* tree that was nearest to the maternal tree, within a distance of 1 km, was located with the probability of this tree becoming the paternal parent decreasing with distance from the maternal tree. *Vochysia ferruginea* is self compatible (Bawa 1985; Bawa & Beach 1985) although predominantly outcrossing (see Chapter 4), therefore, the model included the possibility of selfing, but only when there were no possible pollen donors within a 1 km radius. Thus, if no paternal tree was selected then the maternal tree was assumed to 'self' (i.e. provided both maternal and paternal genes). Selfed trees also had an initial 50 % chance of mortality (but if they survived they were treated as 'normal' target LD trees). Therefore, if there was a potential pollen donor within a 100 m radius, pollen movement was restricted to local dispersal. However, if there was no potential donor there was a decreasing possibility of long distance dispersal up to 1 km. The model was also developed to allow enforcement of a long distance dispersal component. When included, there was a probability of 0.01 for long distance dispersal from a randomly chosen parent in a radius of 100 to 1000m.

For *V. ferruginea* trees, once both parents have been assigned then, at each locus there was a 50 % probability that offspring would inherit one of 2 maternal alleles and also a 50 % probability of inheriting one of 2 paternal alleles. The genotype for each *V. ferruginea* tree in the array was recorded at intervals throughout the model run. If the maternal parent was not *V. ferruginea* (i.e. another LD or ST) then the paternal parent was assessed using the same pollen dispersal dynamics as for *V. ferruginea*. If mature

maternal and paternal parents were successfully selected, an offspring of that tree type was created but genetic information was not recorded.

After two generations each tree in the array became a year older. Each tree had an associated probability of mortality which was equal to the reciprocal of half the maximum age (hence trees survived, on average, for half their maximum age). If an individual tree reached its maximum age it was assumed to die, leaving an empty cell. Light demanding pioneers are fast growing with relatively short life spans, whilst shade tolerant species are typically slower growing, reaching maturity later with longer life spans (Hubbell *et al.* 1999). Therefore, in the model, light demanders had a maximum age of 200 generations and shade tolerant species a maximum age of 400 generations. However, light demanders will reproduce earlier: *V. ferruginea* has a mean age at first reproduction of just 7 years (Boucher and Mallona 1997). In the model, light demanders reached maturity at 20 generations and shade tolerant species reached maturity at 43 generations. Thus, light demanders could dominate in gaps and clear felled areas but eventually would be replaced by the long lived shade tolerant species once the canopy closed.

7.3.3 Initial conditions

The 1 km x 1 km array was randomly seeded with a 5% representation of each tree species. Each tree was given an initial age chosen at random up to the maximum age. For *V. ferruginea* trees, the genetic composition was chosen randomly (for each allele at each locus). Each individual was assigned a genotype at five codominant loci, with alleles at each locus selected randomly. Each locus was given 10 possible alleles with even initial frequencies.

7.3.4 Natural disturbance

When the model ran with the above dynamics to represent a closed primary forest, the shade tolerant species outcompete the two pioneer species. To maintain the presence of these species a natural disturbance factor was added, all primary forests will encounter disturbance due to tree falls, but hurricanes and fire are also broader scale disturbance factors in Costa Rican forests. Thus in the model, disturbance events ranged from removing a single tree up to removing a block of 25 trees (5 x 5 array of cells). The probability of a disturbance event was set to 20% per year for each 1 km x 1 km area.

7.3.5 Model scenarios

Initial study– Primary forest and regeneration dynamics

Initial tests were made using five replications of the following four scenarios (A-D), each of which were simulated for 2000 generations.

Scenario A. *Primary forest with and without natural disturbance*: Genetic data and number of individuals was recorded for the primary forest block. This scenario was assessed with and without the natural disturbance regime as outlined above.

The following scenarios all included a disturbance regime.

Scenario B. *Effects of long distance dispersal in primary forest*: The primary forest set-up of scenario A with a long distance component of seed and pollen dispersal included.

Scenario C. Natural regeneration following pasture abandonment (without long distance dispersal): A 1 km x 1 km primary forest block was clear felled and left to regenerate from individuals in an adjacent primary forest block.

Scenario D. Natural regeneration following pasture abandonment (with long distance dispersal): Scenario C but with the inclusion of the long distance component of seed and pollen dispersal.

Management study - Harvest of natural, regenerated and planted systems

Scenario B and D from these initial studies (i.e. with the natural disturbance regime and the long distance component of seed and pollen dispersal) was used to test a variety of management scenarios that examined, for a regenerating pioneer tree species, the genetic consequences of timber harvesting. Each scenario had five replications and was simulated for 960 generations.

Management Scenario 1. Natural regeneration following pasture abandonment: After initial set-up of primary forest (as in scenario B) a 1 km x 1 km forest block was felled and allowed to regenerate from an adjacent 1 km x 1 km primary forest block. Unlike scenario D above, this scenario included the harvest of all *V. ferruginea* over 40 generations of age every 30 years. This harvest period is typical of that in the field, with the current minimum felling cycle set at 25 years (GFC 2002). In the model it was assumed that 40 generations was equivalent to 60 cm DBH, which is a typical harvestable size in the field and recommended for sustainable timber harvest (GFC 2002). All non *V. ferruginea* tree types were removed at harvest regardless of size.

Management Scenario 2. Natural regeneration from remnant V. ferruginea trees: After initial primary forest set-up (as in scenario B), a 1 km x 1 km forest block was felled but with 1, 5, 50 or 200 mature *V. ferruginea* left as remnant seed trees. Thereafter the 30 year harvest of all *V. ferruginea* over 40 generations was included. All non-*V. ferruginea* tree types were removed at harvest regardless of size. This scenario was modelled with and without an adjacent primary forest block

Management Scenario 3. Planting V. ferruginea from collected seed: A clear felled 1 km x 1 km block was randomly and completely seeded with progeny sampled from either 1, 5, 50 or 200 mother trees in primary forest. The 30 year harvest of all *V. ferruginea* over 40 generations in age was included. All non *V. ferruginea* tree types were removed at the harvest regardless of their size. This scenario was modelled with and without an adjacent primary forest block.

7.2.6 Statistical analysis

Genotypes for all *V. ferruginea* trees in the regeneration area (or primary area for the initial primary forest scenarios) were recorded. For primary forest study blocks, data was collected at 60, 200 and 2000 generations (30, 100 and 1000 years). Here, 30 years is a realistic timescale for harvesting timber, 1000 years shows the long-term behaviour of the model and 100 years is an intermediate value relevant to long term conservation. In the management scenarios study, data was collected at 60, 120, 180, 240 and 960 generations (30, 60, 90, 120 and 480 years), representing the first four harvests and long term state of the model system. Data was analysed using the software package FSTAT 2.9.3 (Goudet 2001), which estimates gene diversities and differentiation statistics for co-dominant genetic markers. The level of genetic variation within populations was quantified by using an unbiased estimator of genetic

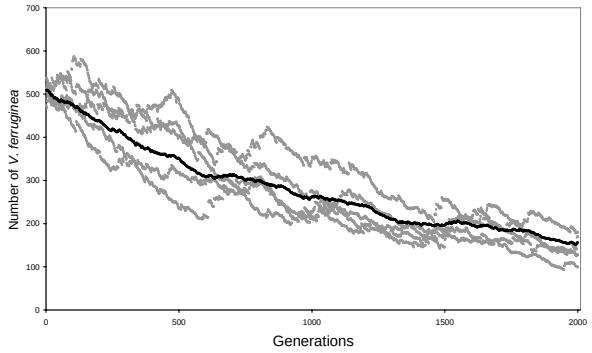
diversity (according to Nei 1987) for each of the five microsatellite loci and averaged over all loci.

7.3 Results

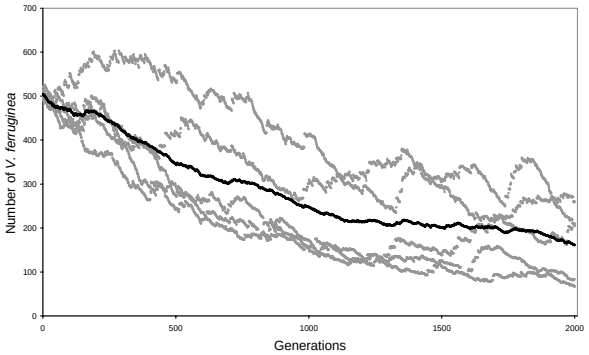
7.3.1 Initial study-Primary forest and regeneration dynamics

Scenario A. Primary forest with and without natural disturbance

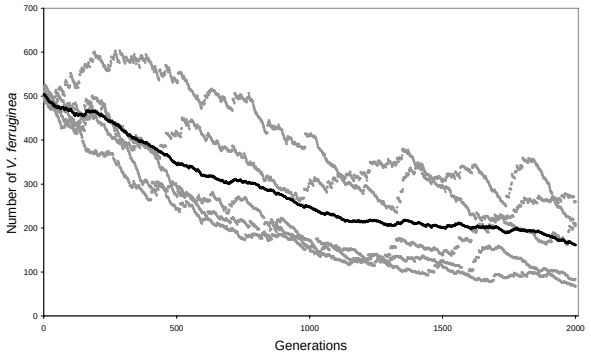
When there was no natural disturbance, numbers of *V. ferruginea* severely decreased during the first 500 generations and *V. ferruginea* was largely out-competed by shade tolerant species by 2000 generations (Figure 7.1a).



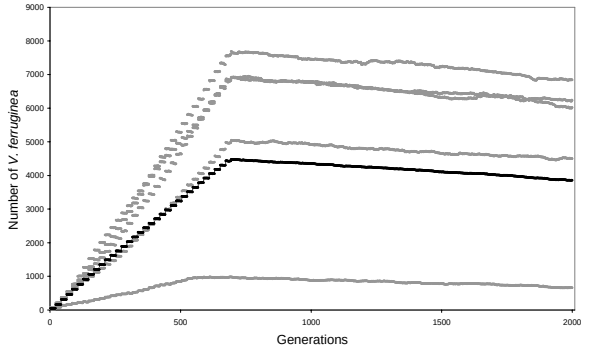
Scenario A. Primary forest with disturbance



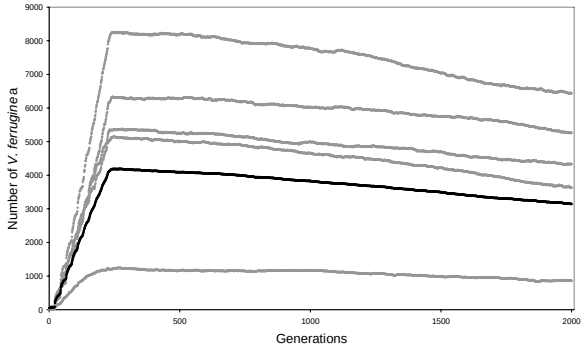
Scenario A. Primary forest without natural disturbance



Scenario B. Primary forest with long distance dispersal



Scenario C. Regeneration following pasture abandonment (without long distance dispersal)



Scenario D. Regeneration following pasture abandonment (with long distance dispersal)

Figure7.1. Numbers of *V. ferruginea* for all preliminary scenarios. Results from each replicated run are shown in grey with average values in black.

With natural disturbance the number of pioneer species also decreased but *V. ferruginea* retained a presence after 2000 generations in all replicated runs (Figure 7.1b).

In general, as numbers decreased so did genetic diversity. However, this decrease was not severe when adequate numbers of *V. ferruginea* were retained in the population. In the model runs including natural disturbance, average population numbers decreased from 487.6 to 154.6 and diversity only decreased from 0.9 to 0.827 (Table 7.1). Without disturbance, numbers of *V. ferruginea* in primary forest fell to levels that were too low to produce useful genetic estimates.

The very low numbers of *V. ferruginea* after 2000 generations in primary forest without disturbance are not sufficient to produce reliable estimates of inbreeding. However, after only 100 generations, average F_{IS} increased from 0.0004 to 0.0196. In primary forest with disturbance, average F_{IS} also increased over time, from -0.0028 at 30 generations to 0.4092 after 2000 generations. However, in this scenario F_{IS} was unchanged after 100 generations. Since the long-term viability of *V. ferruginea* is only possible with a natural disturbance regime it was included in all further model simulations.

Scenario B. Effects of long distance dispersal in primary forest

Figure 7.1 shows the numbers of *V. ferruginea* when a long distance seed and pollen component is assumed to occur with a 1% probability. The number of *V. ferruginea* over time is similar to that in scenario A, although there is a greater range of values (Figure 7.1c) over the replicate simulations. In scenario B there was a slight decrease in genetic diversity (average diversity was 0.9 at 30 generations and 0.848 at 2000 generations) and an increase in inbreeding (average F_{IS} was -0.0016 at 30 generations and 0.229 at 2000

generations) over time, but these effects were not as severe as scenario A, above (see Table 7.1).

Scenario C. Natural regeneration following pasture abandonment (without long distance dispersal)

Numbers of *V. ferruginea* in forest regenerated after pasture abandonment were much greater than those found in primary forest. In scenario C there was a large range in the numbers of *V. ferruginea* at the end of the simulations (Figure 7.1d). This resulted from small differences in the initial invasion success of *V. ferruginea* into the regeneration area. Although there were a great many *V. ferruginea* in the regeneration area the populations showed less average diversity than in scenario A (but with similar decrease over time). Here genetic diversity was 0.861 at 30 generations, compared to 0.9 in primary forest and 0.789 at 2000 generations, compared to 0.8268 in primary forest (Table 7.1). In regenerated populations, there was initially a higher level of inbreeding than in scenario A (0.0232 in the regenerated population and -0.0028 in primary forest). However, the level of inbreeding did not increase over time and was lower than scenario A by the end of the simulation (0.1022 in the regenerated population and 0.4092 in primary forest after 2000 generations).

Scenario D. Natural regeneration following pasture abandonment (with long distance dispersal)

When a long distance dispersal component was included, numbers of *V. ferruginea* increased more rapidly in the regeneration area than was observed in scenario C (Figure 7.1). However, there were fewer *V. ferruginea* present after 2000 generations and there was a higher level of variability between runs.

The genetic diversity found in scenario D was slightly lower than that in scenario C and showed a similar trend of decrease over time (0.869 at 30 generations and 0.735 at 2000 generations, Table 7.1). Levels of inbreeding were slightly higher than in scenario C (average F_{IS} was 0.048 at 30 generations and 0.141 at 2000 generations) but still showed a similar trend of an initially moderate level of F_{IS} after 30 generations, which remained approximately constant over time.

7.3.2 Management study - Harvest of natural, regenerated and planted systems

Management Scenario 1. Natural regeneration following pasture abandonment

Numbers of *V. ferruginea* trees were low at first harvest. After the first harvest numbers increased rapidly, since competing individuals were removed at harvest. Full coverage of the regeneration area occurred by the fourth harvest (Figure 7.2a). Genetic diversity was high at first harvest (0.866) and declined gradually to 0.829 by generation 960 (Figure 7.2a). There was a deficit of heterozygotes found in the regenerated population. Initially average F_{IS} was 0.1 and this increased to a maximum of 0.150 at the third harvest and then gradually declined to 0.115 after 480 years (Figure 7.2a).

Table 7.1: Average genetic diversity (according to Nei 1987) and inbreeding (F_{IS}) in *V. ferruginea* populations for 5 replicated runs for all preliminary scenarios after 30, 100 and 2000 generations. Numbers in brackets show the range of values in replicated runs.

	30 generations		100 generations		2000 generations	
	Diversity	F_{IS}	Diversity	F_{IS}	Diversity	F_{IS}
A. Primary forest without natural disturbance	0.899 (0.899 - 0.9)	0.0004 (-0.01 – 0.006)	0.899 (0.899 - 0.9)	0.0196 (-0.005 – 0.1)	n/a	n/a
A. Primary forest with natural disturbance	0.9 (0.899 - 0.9)	-0.0028 (-0.008 – 0)	0.899 (0.898 - 0.9)	0.003 (-0.001 – 0.021)	0.827 (0.802 - 0.85)	0.4092 (0.362 – 0.493)
B. Primary forest with long distance dispersal	0.9 (0.899 - 0.9)	-0.0016 (-0.009 – 0.016)	0.899 (0.898 - 0.899)	0.0094 (0.002 – 0.017)	0.848 (0.842 - 0.856)	0.299 (0.268 – 0.364)
C. Regeneration following pasture abandonment (without long distance dispersal)	0.861 (0.837 - 0.878)	0.0232 (-0.007 – 0.051)	0.848 (0.816 - 0.866)	0.0894 (0.075 – 0.124)	0.789 (0.743 - 0.835)	0.1022 (0.064 – 0.134)
D. Regeneration following pasture abandonment (with long distance dispersal)	0.869 (0.862 - 0.875)	0.048 (0.032 – 0.07)	0.835 (0.81 - 0.857)	0.1088 (0.09 – 0.127)	0.735 (0.69 - 0.806)	0.141 (0.114 – 0.169)

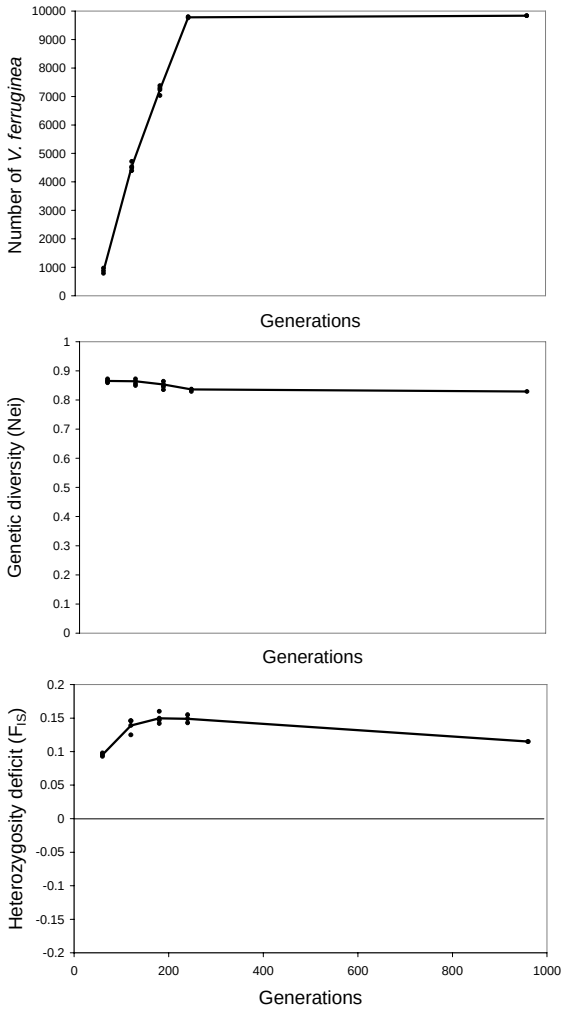


Figure 7.2: Management 1; natural regeneration from primary forest with no remnant trees after harvest at 30, 60, 90, 120 and 480 years. Number of trees, diversity (according to Nei 1987) and heterozygosity deficit (F_{IS}). Points represent the five replicated runs and the line shows the average values. There is no data at year 0 as there are no trees of harvestable age.

Management Scenario 2. Natural regeneration with remnant *V. ferruginea* trees

Natural regeneration from a single remnant tree, with no neighbouring primary forest, led to very low diversity at 30 years (average = 0.41) with high variance between runs, and reflects the level and range of genetic variation typically observed in single remnant trees (Figure 7.3a). The number of trees at harvest increased approximately linearly from low numbers at first harvest to full colonisation of the regeneration area by the fourth harvest. When there was adjacent primary forest, seed dispersal from this area increased diversity after 30 years

(average = 0.78) and the regeneration area became fully colonised by the third harvest (Figure 7.3a).

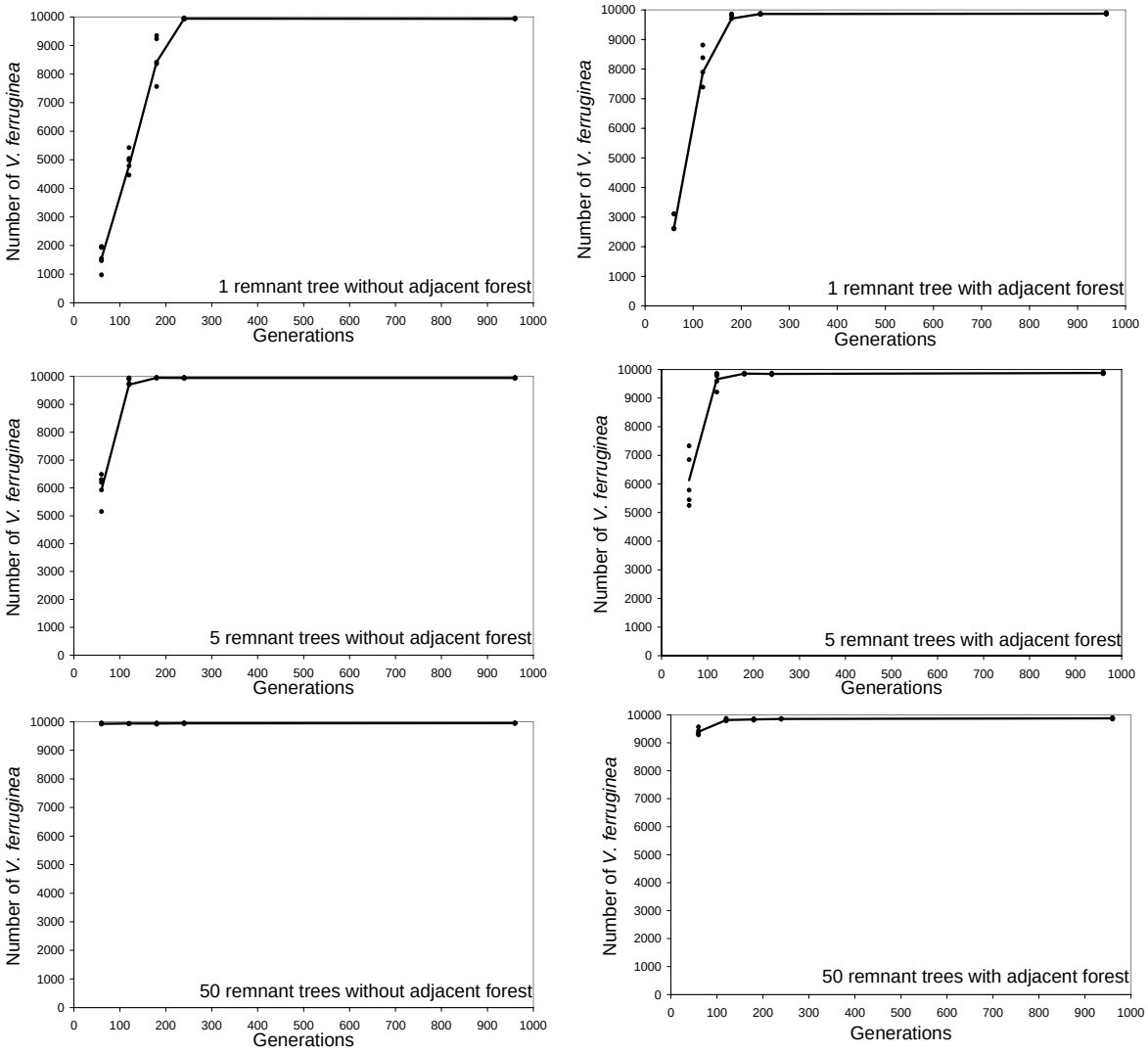


Figure 7.3a: Management 2; natural regeneration from remnant trees with a harvest at 30, 60, 90, 120 and 480 years. Number of trees. Points represent the five replicated runs and the line shows the average values. There is no data at year 0 as there are no trees of harvestable age.

When natural regeneration was from 5 seed trees genetic diversity was higher at 30 years (average = 0.79) than the single remnant tree scenarios, and the regeneration area was fully colonised after approximately 60 years (Figure 7.3a). With adjacent primary forest, diversity was slightly increased at 30 years (average = 0.83), but there was a comparable speed of regeneration to the non-adjacent forest scenarios (Figure7.3a). When regeneration was from 50 seed trees the regeneration area was completely colonised by first harvest. Simulations of this scenario showed high levels of diversity and low variation between simulations. Results,

with and without adjacent primary forest, were comparable (Figure 7.3b) and adding further seed trees (200 trees) did not change the results.

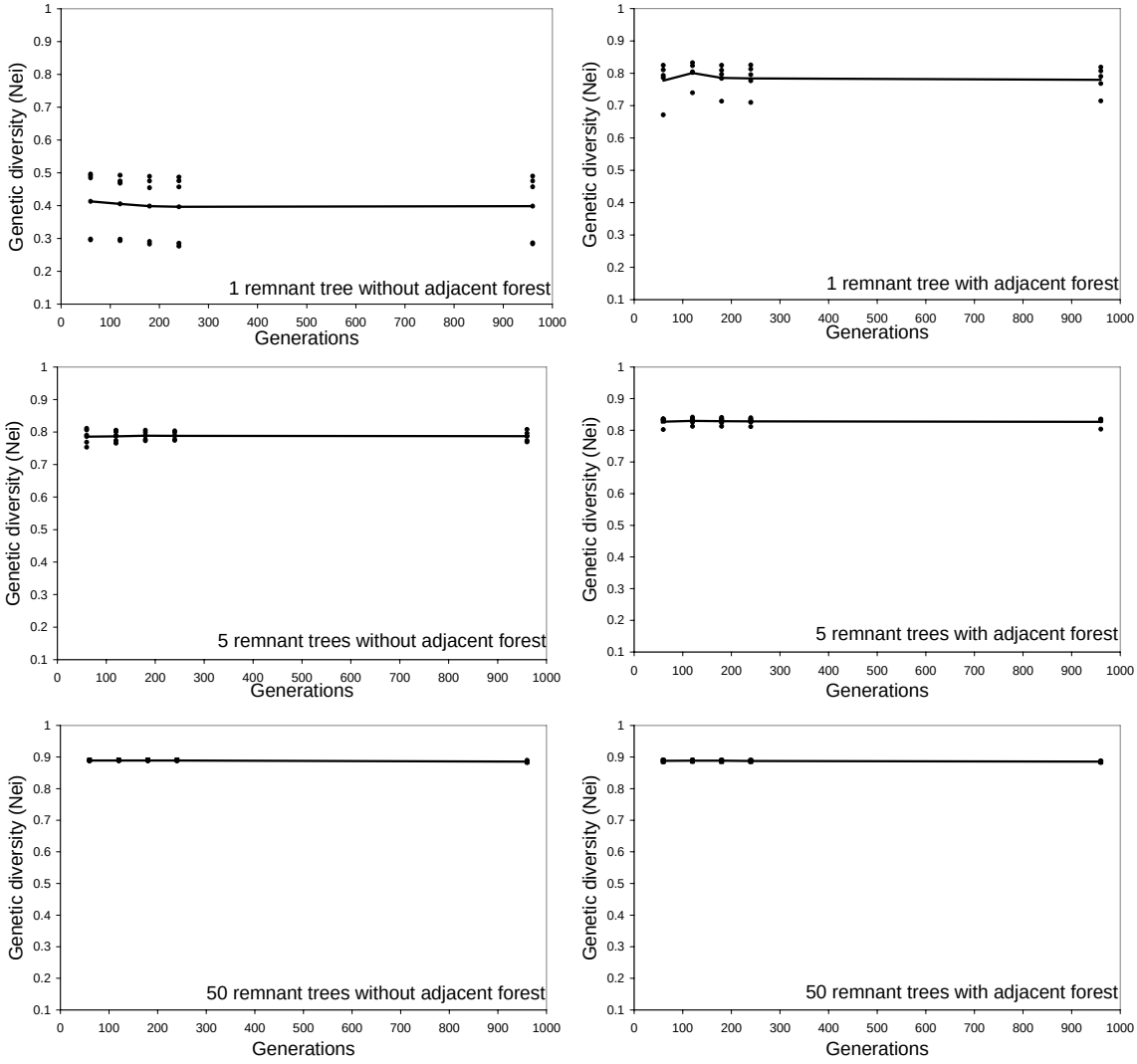


Figure 7.3b: Management 2; natural regeneration from remnant trees with a harvest at 30, 60, 90, 120 and 480 years. Diversity (according to Nei 1987). Points represent the five replicated runs and the line shows the average values. There is no data at year 0 as there are no trees of harvestable age.

In all runs where regeneration was from remnant trees, there was some degree of inbreeding (see Figure 7.3c), where positive values of F_{IS} show there were a greater number of homozygotes than expected under random mating. When regeneration was from a single remnant tree there was a low level of inbreeding found in trees at the first harvest (since

most offspring are derived (randomly) from the single remnant tree). The level of inbreeding then increased reaching a maximum at the fourth harvest and gradually decreased until the end of the simulation. The level of inbreeding followed a similar trend when there was adjacent primary forest. When regeneration was from 5 trees, average F_{IS} was 0.15 at the first harvest, increased to a maximum of 0.18 at the second harvest and declined to 0.14 after 480 years. The results were similar with adjacent primary forest. Natural regeneration from 50 or 200 trees (with and without adjacent primary forest) displayed a similar trend to that of 5 remnant trees but the level of inbreeding was slightly reduced for 50 trees and reduced further for 200 trees.

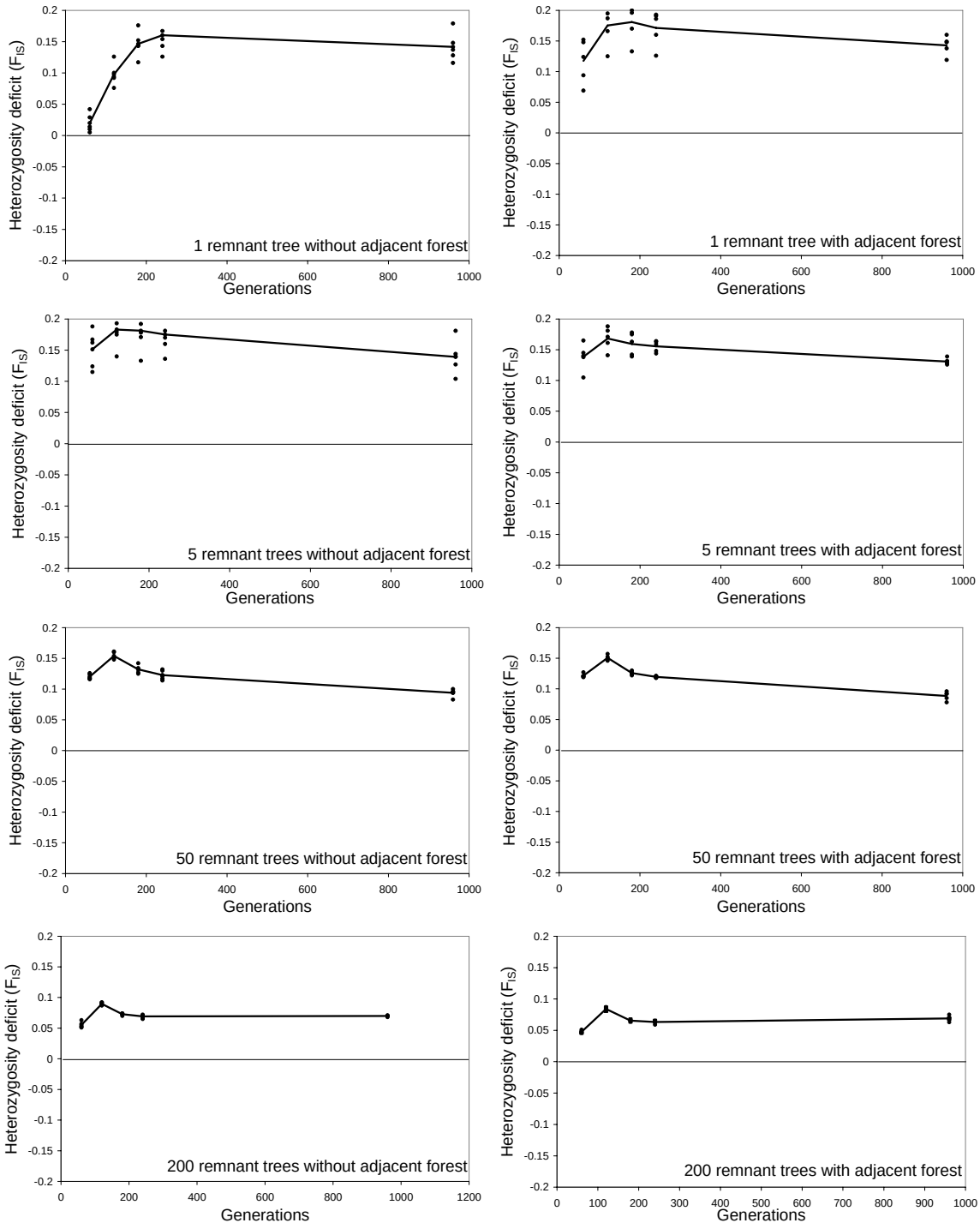


Figure 7.3c: Management 2; natural regeneration from remnant trees with a harvest at 30, 60, 90, 120 and 480 years. Heterozygosity deficit (F_{IS}). Points represent the five replicated runs and the line shows the average values. There is no data at year 0 as there are no trees of harvestable age.

Management Scenario 3: Planting *V. ferruginea* from collected seed

For scenario 3 runs, average genetic diversity remained approximately constant over time. Therefore the diversity numbers quoted are the values at the first harvest (30 years). There was also little difference between model runs using different numbers of seed trees (see Figure 7.4a). The regeneration area was completely planted so that all cells in the array contained offspring, so there was a maximum number of *V. ferruginea* at first and all subsequent harvests.

When seed was collected from 1 mother tree, average diversity was 0.78 and similar to results with adjacent primary forest (average = 0.76; Figure 7.4b).There was an increase in diversity with seeds taken from 5 mother trees (average = 0.87), again results were similar to those with adjacent primary forest (average = 0.87). Diversity was maximised when seeds were collected from 50 trees (average = 0.9; Figure 7.4b) and there was no difference between these runs and those with seeds taken from 200 mother trees (data not shown). In both cases adjacent forest had no effect on average genetic diversity.

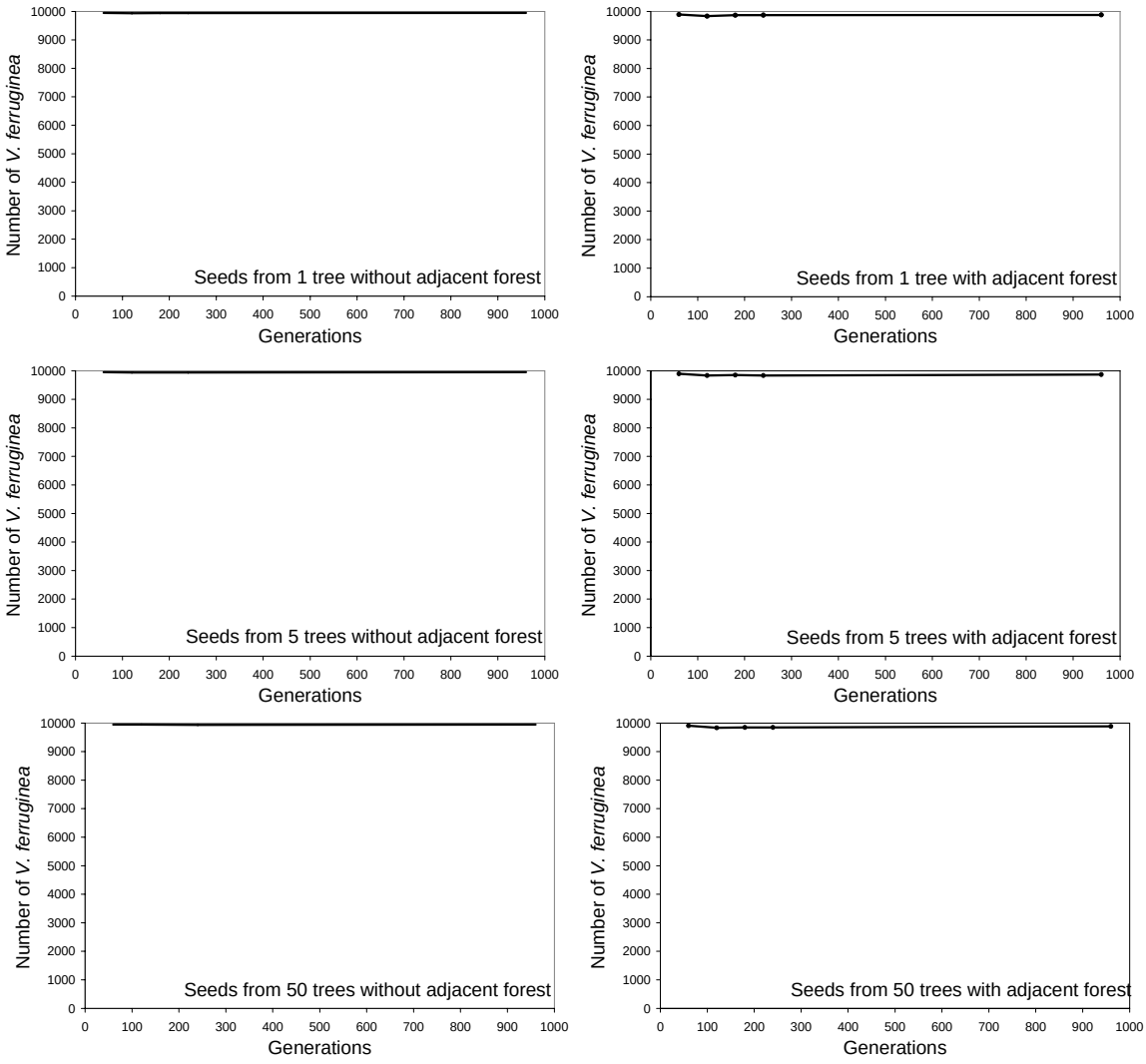


Figure 7.4a: Management 3; regeneration from collected seeds with a harvest at 30, 60, 90, 120 and 480 years. Number of trees. Points represent the five replicated runs and the line shows the average values. There is no data at year 0 as there are no trees of harvestable age.

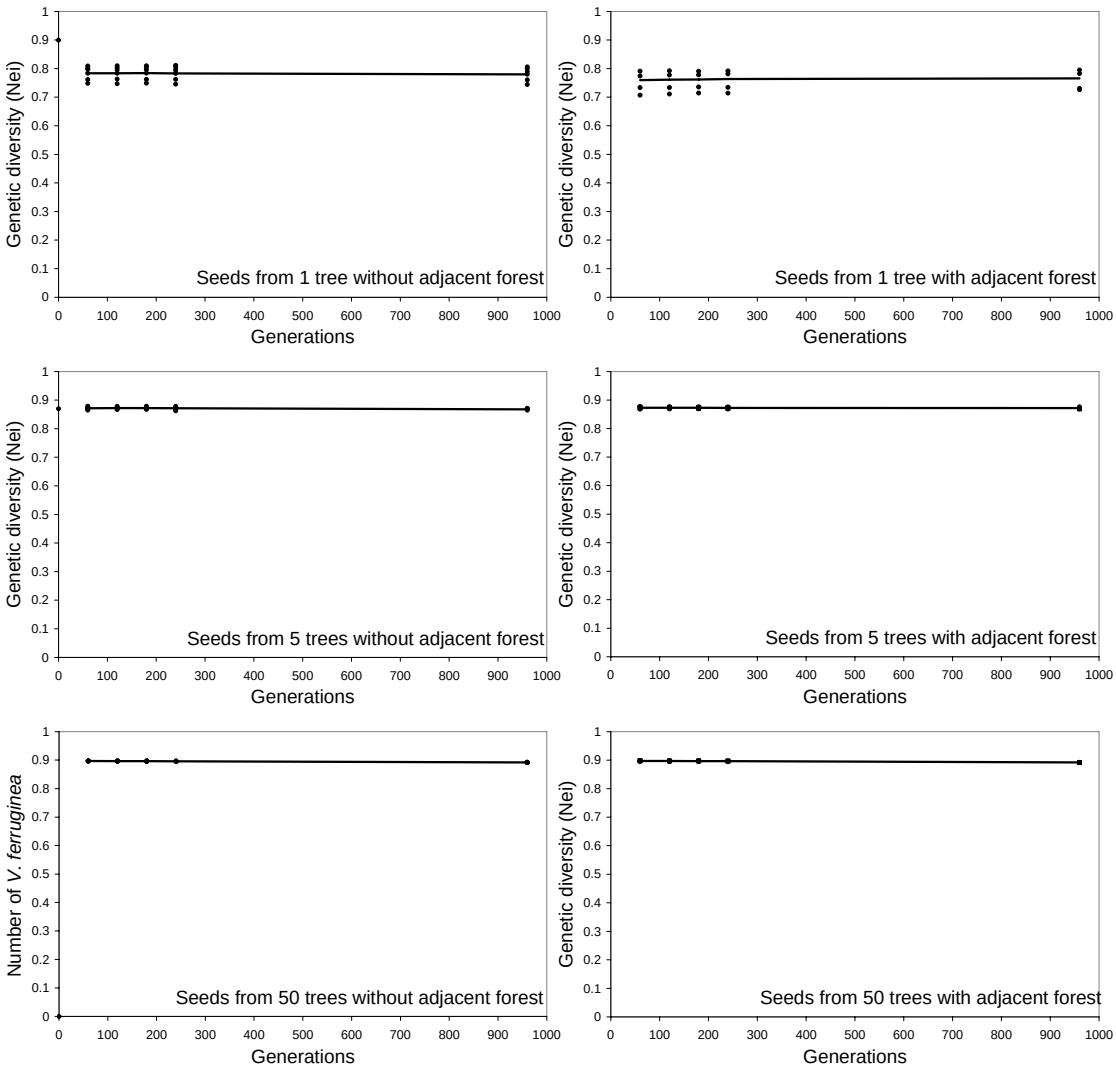


Figure 7.4b: Management 3; regeneration from collected seeds with a harvest at 30, 60, 90, 120 and 480 years. Diversity (according to Nei 1987). Points represent the five replicated runs and the line shows the average values. There is no data at year 0 as there are no trees of harvestable age.

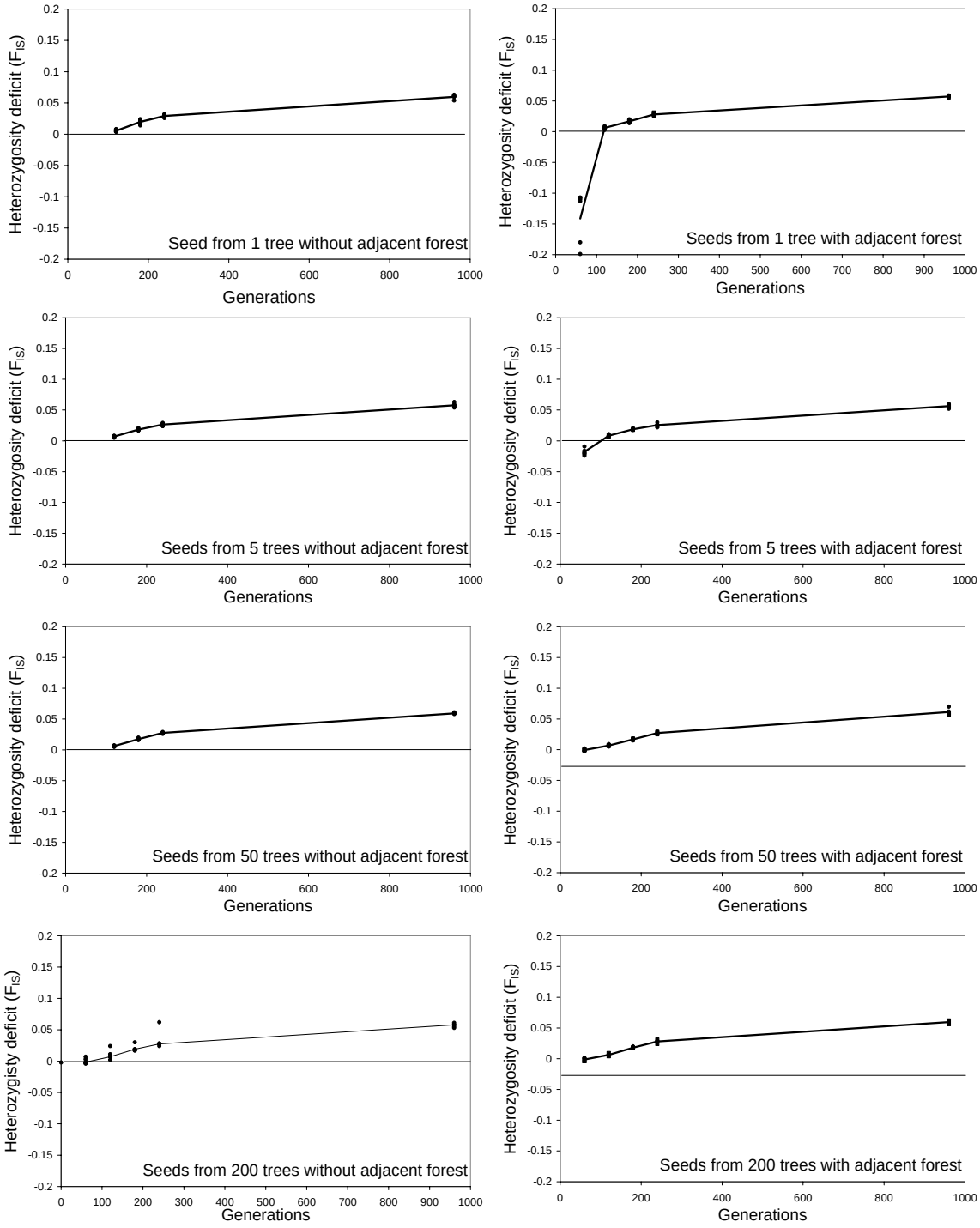


Figure 7.4c: Management 3; regeneration from collected seeds with a harvest at 30, 60, 90, 120 and 480 years. Heterozygosity deficit (F_{IS}). Points represent the five replicated runs and the line shows the average values. There is no data at year 0 as there are no trees of harvestable age.

In scenario 3 simulations (see Figure 7.4c), where 1 or 5 mother trees were used there was an excess of heterozygotes found in trees at the first harvest. For all simulations (regeneration from collected seeds from 1 to 200 trees) the average level of inbreeding was

approximately zero at the second harvest and increased to approximately 0.06 by the end of the 960 generation runs. Thus, the level of inbreeding did not appear to be significantly affected by increasing the number of mothers that seeds were collected from or whether there was an adjacent primary forest.

7.4 Discussion

7.4.1 Initial study: Primary forest

When disturbance was included in the primary forest model, light demanding species could maintain population numbers as they were able to quickly colonise the gaps in the array brought about by disturbance. Without disturbance, succession led to elimination of the two pioneer species over the time period modelled. In an established forest, including a long distance dispersal component did not lead to an appreciable increase in numbers of *V. ferruginea*, showing that the pioneer species were able to colonise the gaps left by disturbance even when they are restricted by limited pollen and seed dispersal

When natural disturbance was included there was a slight decrease of diversity with time and an increase in inbreeding due to changes in number of trees. With long distance dispersal, diversity within the array was not affected, however, this increased dispersal led to a decrease in inbreeding over the long term. As *V. ferruginea* trees were relatively sparse in the primary forest there may be few individuals contributing to any disturbance gap (which can be 1 to 25 cells area) and so individual patches of *V. ferruginea* may be derived from few parents. However, gaps created by disturbance occurred randomly throughout the array and so individual patches may consist of seedlings descended from different parents leading to an array-wide population that retains diversity. Long distance dispersal did not appear to alter the diversity in the whole array but may allow for gene flow from more parents into disturbance gaps. When there was only local dispersal subsequent generations in gap populations are likely to be offspring from individuals that initially colonised the gap,

thereby leading to inbreeding. However, when long distance dispersal is allowed, trees from surrounding areas may also contribute to successive generations and therefore the level of inbreeding may be mediated.

7.4.2 Initial study, natural regeneration in a cleared area

After clear felling, trees from adjacent primary forest colonised into the empty array through seed dispersal. In these open areas, *V. ferruginea* and the other pioneer species had a competitive advantage over the shade tolerant species. The light demanding species also reached maturity sooner, so were able to quickly establish populations following colonisation. This resulted in very high numbers of both pioneer species in the regeneration area. If there was regular harvesting, with all non-target species being removed, there was a long term, array-wide domination by the single target pioneer species. Zagt *et al.* (2005) using models of logging scenarios, found that altering the minimum cutting diameter will affect yields in relation to the reproductive age of the species. Early maturing species such as *V. ferruginea* will have a greater opportunity to set seed in-between logging cycles

When there was long distance dispersal present, the clear felled area was colonised more rapidly. Any seed dispersed via long distance dispersal into the empty array will be free from competition as illustrated in the landscape model of oak colonisation (Chapter 6). In this scenario, long distance dispersal by shade tolerant species also occurred, which created greater opportunity for abundant shade tolerant species in the primary forest to colonise the array. However, any empty cells around the dispersed individual may become colonised by neighbouring pioneer species due to faster maturity in the pioneer species. Thus, the number of *V. ferruginea* that have an advantage due to long distance dispersal varies between runs and may account for the greater variation in numbers (Figure 7.1). Without long distance dispersal, it was possible for the pioneer species to expand in numbers into the colonisable space and effectively form a barrier to colonisation by locally dispersed seed from shade

tolerant species. Long distance dispersal from shade tolerant species also resulted in quicker succession in the regenerated area as these species were present in high density in the primary forest and could invade single cell gaps left by pioneer trees dying or small disturbances. All runs with long distance dispersal showed succession towards a forest containing more shade tolerant species occurring more rapidly, with declining numbers of the two pioneer species. However, there is evidence that pioneer species tend to have greater dispersal ability than shade tolerant species (Hubbell and Foster 1992). A future investigation could examine a scenario with long distance seed and pollen dispersal restricted to pioneer species.

Although there was a much higher number of *V. ferruginea* after clear felling, these populations showed slightly lower diversity than those in primary forest (Table 7.1). Colonisation into the cleared area was only possible from trees from one primary forest edge. This limited the number of *V. ferruginea* that could contribute seed to the regenerating area. Although gene flow via pollen from *V. ferruginea* in the adjacent primary forest will have mediated the loss of diversity after this population bottleneck, there was still less diversity than within primary forest. Long distance dispersal potentially increased the number of *V. ferruginea* that could contribute seed to the regenerating area and also allowed for greater gene flow via pollen dispersal to *V. ferruginea* in the clear felled area. With long distance dispersal, there was slightly more diversity after 30 years and slightly less diversity at 100 years and 1000 years. Long distance dispersal allowed for a more rapid colonisation, with the array largely colonised by 200 generations as opposed to 700 generations without it. However, long distance dispersal also provided an advantage for shade tolerant trees and so numbers of *V. ferruginea* started to decline earlier than without long distance dispersal. This population decline could lead to a reduction of diversity. The rapid colonisation also means that in runs with long distance dispersal the array will be completely colonised by individuals from fewer generations thereby potentially experiencing less gene flow from

primary forest before numbers decline. Long distance dispersal is expected to increase diversity as there is greater gene flow to populations; however, these results show that this must be considered in context with other, potentially complex, demographic and genetic changes affected by long distance dispersal.

In regenerating populations the level of inbreeding initially increased but then stabilised (Table 7.1). Initial populations in the clear cut area were from very few individuals and the subsequent populations are all likely to show biparental inbreeding. Later populations will have experienced some gene flow from less related pollen donors and the increased succession will break down the sibling cohorts so there is less pollen swamping from neighbouring siblings. The long-term level of inbreeding in regenerating populations was lower than in primary forest, probably caused by the much higher numbers of *V. ferruginea*. Long distance dispersal increases gene flow from unrelated trees so leads to less inbreeding even when there are fewer individuals present. Therefore, regenerated forest from neighbouring primary forest can lead to highly variable numbers of dominant pioneer species that show a lower diversity than primary forest but with less inbreeding.

In all runs more expansive pollen and seed dispersal did not appear to increase diversity but decreased inbreeding. The level of extra dispersal in the model did not counter any genetic drift that may be occurring in either the primary forest or the regenerated population. Increased seed dispersal did not create a significant advantage for colonisation by *V. ferruginea*; however, where long distance dispersal was present it had a large effect on the colonisation speed from primary forest into an empty array. The succession of shade tolerant species into the pioneer dominated forest was also more rapid.

7.4.3 Management study, natural regeneration with V. ferruginea as a remnant tree

Results from the runs testing different management strategies with a 30 year harvest showed that when natural regeneration was from a low number of remnant trees there was consequently lower genetic diversity. Where there was no gene flow from adjacent forest the diversity in seedling populations was limited to the maternal alleles of the parent(s) of the remnant trees. When regeneration was from 1 remnant tree without adjacent primary forest the tree was forced to revert to selfing and all genotypes in the population in the regenerated area were composed of alleles present in the original remnant tree. However, in populations regenerated from a small number of remnant trees (1 to 5), gene flow from adjacent primary forest increased genetic diversity over time.

When the regeneration was from 5 trees or less, the number of trees at first harvest was restricted by the colonisation speed from these trees. In these instances, adjacent primary forest increased the number of *V. ferruginea* at harvest as any *V. ferruginea* in primary forest able to disperse into the regeneration area contributed to numbers. The pioneer trees also had a competitive advantage in the initial stages of regeneration. When regeneration was from a larger number of remnant trees (e.g. over 50) then the regeneration area could be completely colonised rapidly. Here adjacent primary forest had little influence on the regeneration dynamics.

It is worth noting that the results for management scenario 2 with 5 remnant trees (with and without adjacent primary forest) and for 1 remnant tree with adjacent primary forest are similar in terms of average genetic diversity and level of inbreeding (although more variation was observed for 1 remnant tree). However, the level of diversity in these simulations is lower than for regeneration from adjacent primary forest only (management scenario 1). Thus, although a small number of remnant trees may speed up regeneration, they appear to reduce genetic diversity within the regenerating area. Possibly through

domination of the founding population by seeds from the remnant trees as opposed to gradual colonisation from a higher number of trees in adjacent forest.

In the regenerated forest, all populations showed a deficit of heterozygotes, which decreased with increasing numbers of remnant trees. However, there was less inbreeding found with regeneration from a single tree after the first and second harvest. Beyond this, inbreeding increased as in the other model runs (although with adjacent forest this pattern was not as evident). When there was a single tree left the initial selfed populations were within pollen dispersal distance of the single remnant tree and were then, once mature, a single, randomly mating population. After harvest the population can contain individuals increasingly isolated from each other and so inbreeding may increase. Inbreeding slowly declined in all runs and this decline may be related to colonisation time. When the array is completely colonised, patches of homozygotic individuals can experience greater gene flow.

7.4.4 Management study, planting V. ferruginea from collected seed

The levels of diversity were high in planted populations, even when the seed for planting was collected from only one individual (average diversity is 0.77). This contrasted with the low diversity experienced when regeneration was from one remnant tree. Collecting seed for planting from 5 primary forest trees led to high levels of diversity similar to those found when seed was collected from 50 or 200 mother trees. The increase in diversity observed when the population was regenerating from seed rather than remnant trees occurred because a single mother tree samples the diversity from a large number of fathers within the pollination distance to the mother tree. Therefore, although seeds are only taken from one mother tree, the genetic diversity of all trees within the pollen dispersal kernel was effectively sampled. This shows that the spatial distribution of individuals, and the genetic structure of the population, may have a significant impact on the levels of variation found in a large seed sample taken from a single individual. Consequently, when collecting seed

material the extent of the reproductive population contributing pollen to a seed tree will determine the extent of variation sampled. This ability to sample a large proportion of population diversity is therefore dependent on the number of pollen donors. Collecting seed from isolated trees will sample a significantly lower amount of diversity than seeds collected from within continuous forest. The extent of the reproductively effective population is largely determined by the extent of pollen flow, with restricted pollen dispersal expected to severely decrease the possibility of sampling a majority of the population variation by collecting seed from one individual.

Although diversity was high when seed was collected from a single tree it still increased when more mother trees were sampled. All 10 alleles were present in the seedling population when seeds were collected from a single tree, however, this may be a consequence of even allele frequencies in the initial primary forest population. If a population shows genetic structuring then sampling from a single individual is less likely to sample a population's genetic diversity. Degen *et al.* (1996) found that clustered populations decreased the effective population size as did limiting wind direction and decreasing the number of seed trees from 20 to 10. From these results they suggested that both the spatial distribution of a population and the number of seed trees were important in estimating the extent of the reproductively effective population and therefore the potential for increased genetic drift. The model of regeneration from *V. ferruginea* did not contain rare alleles and these are most likely to be lost in natural regeneration from a low number of individuals and would then constitute a decrease in variation not shown in these results. Rare alleles are also more likely to be unsampled when collecting seed from a low number of individuals.

When the population originated from planted seeds, there was a pronounced deficit of homozygotes in those runs where seeds were taken from 1 or 5 mother trees. This was because the allele frequencies in the planted populations were dominated by the maternal

alleles yet as there was little selfing these common alleles were only found in the homozygotic state when the paternal allele matches that of the maternal allele. This effect is most evident when seeds were collected from one maternal tree as the population allele frequencies were dominated by very high frequencies of the two maternal alleles.

7.5 Conclusion

In conclusion, to utilize a pioneer tree for timber the model results suggest that natural regeneration from primary forest into a cleared area of the magnitude of 1 km^2, may not be the most efficient use of the available land. In model populations, natural regeneration led to colonisation of only approximately 2.5% of the model array after 30 years. If the area to be used for secondary forest utilisation has remnant trees then these effectively increase the colonisation rate. However, natural regeneration does not adequately utilise the available space unless the number of remnant trees is large. As large numbers of remnant trees decrease the utility of the initial clearance, colonisation rates can best be maximised by supplementing natural regeneration with planting. When collecting seed for planting it is best to collect from trees within continuous primary forest and where there is pollination from a number of fathers so as to maximise the genetic variation sampled. When the number of pollen donors is unknown, the population is at a low density or the species shows restricted pollen dispersal it would be most prudent to collect seed from a high number of mother trees. If seedlings are to be generated by cloning then a much higher number of trees should be sampled as in this case the genetic diversity of only one individual and not the pollen cloud, is sampled.

This chapter illustrates how a simple model of forest dynamics with seed and pollen dispersal can provide useful insight into the colonisation rate and genetic structure of pioneer tree species. However, the results should be viewed as highlighting important trends and not as quantitative predictions. The quantitative results depend critically on the model

composition, the parameter values and functional relationships chosen. Thus, qualitative trends have been highlighted but it should be recognised that the quantitative behaviour will depend on the characteristics of individual species. Natural regeneration of secondary forest can lead to losses in genetic diversity and so a species' genetic response to management should be a consideration for sustainable forestry. Effective sustainable management of a timber species may best be achieved by the integration of standard forest inventories, molecular analysis of gene flow and spatial modelling results.

Chapter 8: Discussion

8.1 Principal findings

8.1.1 Molecular analysis of secondary forest regeneration by Vochysia ferruginea

Deforestation, timber extraction and rural-urban migration in neotropical forest habitats have caused a recent increase in the area of secondary forest, arising largely from colonisation of abandoned pasture by pioneer tree species, to the extent that regenerated forest is becoming important as an economic and environmental resource. The conservation of genetic diversity is an important component in maintaining the long term fitness and evolutionary potential of tropical forest tree species. This study investigates some of the genetic characteristics pertinent to the establishment of secondary forest through colonisation by a pioneer species. Pioneer trees need light to germinate and so persist in primary forest through the exploitation of gaps in the canopy. They can also form dense stands of secondary forest after either human mediated or natural major disturbance. Using two different sites this study assessed mating system, genetic diversity and structure in populations of the pioneer tree *Vochysia ferruginea* in primary forest gaps and dense secondary forest.

Populations of *V. ferruginea* in both primary and secondary forest were found to be predominantly outcrossing. The progeny from two isolated remnant trees were also predominantly outcrossed and it was estimated that these progeny were from a greater number of pollen donors. *V. ferruginea* is self-compatible and there appeared to be some true uniparental selfing occurring in one primary forest population and in progeny from the two remnant trees. As these were results from seed collections it is unknown whether these selfed seeds would have been less able to survive to maturity than non-selfed progeny. These results suggest *V. ferruginea* has a mixed mating system but selfed progeny are rare as gene

flow is effective in *V. ferruginea* populations, thus, primary forest and remnant trees experience extensive gene flow within a large effective population of spatially distant adults.

Primary forest populations of *V. ferruginea* populations had lower density, higher diversity and greater allelic richness than secondary forest populations. Primary forest fragments, of only a small number of trees, also had high diversity but had lower allelic richness. Although the secondary forest populations had lower diversity, the difference was small (range: 0.60 to 0.74 in secondary forest and 0.68 to 0.75 in primary forest). Seedling populations varied in levels of diversity compared to adult secondary forest, but consistently showed a loss of alleles. Progeny taken from seed trees within both primary and secondary forest showed the highest levels of diversity and the highest levels of allelic richness.

The loss of alleles in the seedling populations is an expected result of colonisation by a low number of seed donors. The differences in diversity between the two seedling populations are most likely due to the differences in their age structure. The low diversity population was composed of seedlings from a single year and the high diversity population was composed of seedlings from several years. Progeny from trees within mature forest adjacent to seedling populations were highly differentiated from the seedling populations and also showed high diversity and high allelic richness. Therefore, whilst the initial colonisation event led to reduced genetic diversity and allelic richness in the first seedling population, gene flow into regenerated populations from subsequent seed dispersal effectively mediated this loss, with increased diversity in seedling populations derived from several years of colonisation. Thus, mature secondary forest showed higher diversity and higher allelic richness than seedling populations. The high allelic richness in progeny from trees in secondary forest also shows that there is gene flow into these dense stands from outside the population and, provided there are high diversity populations within the pollen flow distance, the genetic diversity of secondary forests can increase.

Secondary forests tended to show an excess of heterozygosity, this suggests that these regenerated populations are derived from a small number of outcrossing seed trees that have experienced gene flow from a high number of pollen donors. Thus, the maternal alleles are very common in the population but will be in the homozygotic state less than expected under random mating.

Of the three secondary forest populations one was found to be unusual in that it had a significant deficit of heterozygotes and a higher diversity than the primary forest population it was assumed to derive from (in this case the primary forest population sampled comprised all of the primary forest at that site). The deficit of heterozygosity in this population may be due to population structuring giving rise to a Wahlund effect. The Wahlund effect causes reduced heterozygosity in populations due to subpopulation structure, if two or more subpopulations that are in Hardy-Weinberg equilibrium have different allele frequencies then overall the heterozygosity in the population is reduced (Crow and Kimura 1970). In combination, the heterozygote deficit and higher than expected diversity suggested that additional unsampled donor trees must have contributed to the establishment of this population. Potential candidate donor trees were present as remnant trees bordering rivers at this site. The genetic diversity of this regenerated population suggests that small fragments of differentiated primary forest may play an important role as sources of diversity for regenerating populations.

In secondary forest populations there were higher levels of inbreeding, however, as estimates of uniparental inbreeding were very low, this increase can be accounted for by increased biparental inbreeding. Whether biparental inbreeding has a measurable effect on fitness after recruitment has not been resolved by this study. However, there was some evidence that, over time, trees with high heterozygosity may outperform trees with lower heterozygosity.

Increased biparental inbreeding in these dense populations led to a clustering of interrelated groups of trees and all secondary forest populations showed significant positive spatial autocorrelation at small scales (4 m to 30 m). In primary forest there was little or no structure at short distances but there was some structuring at greater distances (30 m to 90 m). In contrast, all secondary forest populations showed negative spatial autocorrelation at larger distances (50 m to 90 m). One possible explanation for this effect is that the density of *V. ferruginea* populations significantly alters pollinator behaviour. In dense patchy populations, pollinators mainly move between nearby plants, mean flight distance decreases and structure is brought about on short scales. Low density has the opposite effect, facilitating pollen dispersal events at longer distances and decreasing structuring at shorter scales. Inadequacies in the level of discrimination possible with the dataset made it difficult to perform a direct analysis of gene flow. However, there was some evidence that, whilst gene flow may be extensive, mean pollen dispersal distances are significantly shorter in dense secondary forest. Lower gene flow in secondary forests may then also compound the small scale structuring that is brought about by colonisation.

The genetic characteristics of these secondary forest populations suggest that small differences in the colonisation process, such as distance from source population and presence of remnant trees, may have important consequences for levels of genetic diversity, allelic richness and heterozygosity in regenerated populations. Although levels of diversity may not be appreciably lower in regenerated populations the differences in density can have consequences for gene flow and thus, the levels of biparental inbreeding.

8.1.2 Modelling regeneration at a landscape and population scale

The landscape-scale, spatial model of gene flow via seeds showed that determining the seed dispersal kernel, particularly the long distance component, is vital in predicting the genetic response of a tree species to colonisation. Even very rare long distance dispersal significantly influenced colonisation speed and genetic structure.

It is predicted that the multiple founding events that characterise colonisation will lead to a decrease in allelic richness with distance from source that may not be mitigated by pollen dispersal in dense, fixed populations. In the seed dispersal model, population genetic structure was the product of multiple founder events from long distance dispersal with local dispersal confined to those cells surrounding the mother tree, yet within-population diversity increased as the proportion of long distance seed dispersal increased. This indicates that the creation of a patch structure from long distance dispersal of individual trees can maintain genetic diversity of populations because, although the founder events of long distance colonisation may result in low within-population diversity, the stability of these patches allows maintenance of migrant haplotypes within the total population.

Modelling colonisation at a population scale also illustrated an increase in the speed of colonisation by pioneer species with long distance dispersal. If there was long distance dispersal in shade tolerant species there was also much faster succession in the regeneration area, from a forest dominated by pioneer species to one dominated by shade tolerant species. However, if there was regular harvesting, with all non-target species being removed, there was long term, array-wide domination by the single target pioneer species.

Long distance dispersal may not be as important in established populations. In the population scale model with both seed and pollen dispersal, pioneer trees in established primary forest were able to colonise disturbance gaps equally well when seed and pollen

dispersal was limited. With long distance dispersal, diversity within the array was not affected, although limited dispersal led to lower diversity within gap populations, the whole forest population retained a high diversity. However, this increased dispersal led to a decrease in inbreeding over the long term. Thus, long distance dispersal may mitigate biparental inbreeding by allowing gene flow between spatially distant gap populations.

8.2 Suggestions for management and conservation

Fragments and primary forest do impact the genetic diversity of secondary populations. The effectiveness of gene flow via long distance seed dispersal, as seen in the simulations, means that primary forest populations form important sources of genetic diversity. The impact of this gene flow was evident in mature secondary populations, where one population had a higher diversity than the whole adjacent primary forest fragment, and could also be seen in seedling populations, where a multi-generation seedling population showed much higher diversity than a population from a single year. In the mature secondary forest population, gene flow was most likely to have come from both primary forest blocks and very small fragments of primary forest, e.g. along protected river courses. Hence, even very small fragments can be influential as they may be differentiated from primary forest blocks and can form a significant genetic resource.

A model of gene flow from a pioneer species was used to test a number of management strategies appropriate for the maintenance of genetic resources in secondary forest. The results suggest that, in terms of promoting the rate of colonisation, allowing natural dispersal alone is not the most efficient means to regenerate forest on abandoned land unless a large number of remnant trees are present within the area to be regenerated. To increase regeneration rates where there are few remnant trees, it is much more effective to plant trees either from collected seed or from nurseries. However, in this case, a careful selection of seed tree must be made to ensure maintenance of genetic diversity levels. Modelling showed

that when there is adequate gene flow from pollen, collecting a large number of seeds, even from a very small number of trees, can effectively sample the majority of the diversity within a population. However, this will only be the case when the seed trees are in continuous forest with high levels of gene flow from a number of fathers. It is current practice in some nurseries in Costa Rica to generate seedlings by cloning as this can be achieved outside of the fruiting season (FUNDECOR pers. comm.). Cloning will only sample the genetic diversity of one individual and not the population, therefore, if cloning is used, a much greater number of spatially distant trees will be required to sample a significant proportion of the diversity of a population.

8.3 Criticism of study and suggestions for future work

Initial assessment of possible study sites included a potential third site at Florencia, Costa Rica, although sampled, individuals from this site could not be analysed due to time constraints. This site consisted of a fragment of primary forest and a number of *V. ferruginea* dense mapped secondary plots (Finegan and Camacho 1999). If this site could have been included in the analysis then there would have been a replicated site of separate secondary forest populations at different distances from primary forest. Inclusion of a third site would also have given more statistical weight to all results.

The direct analysis of gene flow via pollen illustrated the difficulties in getting reliable estimates from paternity analysis when there is insufficient data. There were a high number of allele drop-outs within the dataset and whilst this in itself may not be a problem for some analyses, it rapidly decreases the power of statistical assignment of paternity. To successfully apply paternity assignment approaches, allelic dropout must be minimised, but it is also important to use as many loci as possible. Using maximum likelihood methods of paternity, increasing the number of loci in the dataset provides the best means to increase the power to identify real paternity. In this study the number of loci analysed was too low to

provide strong estimates of pollen flow, therefore if these are to be obtained for *V. ferruginea*, it is recommended that additional loci are examined.

The population model was developed with the aim to include dispersal curves derived from molecular analysis; however, the difficulties in obtaining a reliable direct analysis of gene flow meant that dispersal dynamics could only be estimated. Therefore, the pollen model was not as species specific as proposed, lessening its power as a management tool for sustainable forestry of *V. ferruginea* regeneration.

This study also attempted to find whether inbreeding had a notable effect on fitness by comparison of individual heterozygosity and long term growth rate (using data collected by CATIE). There was no statistically significant evidence for a fitness effect on growth; however, there was some evidence that, over time, trees with high heterozygosity may outperform trees with lower heterozygosity. The effect of inbreeding on fitness either as reproductive output or growth is very relevant to forestry so it would be valid to investigate this further with a greater extent of informative molecular data. This study assessed fitness effects in adults; however, the detrimental effects of inbreeding may be more important during recruitment and to the survival of seedlings. Therefore, the correlation of homozygosity with fitness in seedlings would also be a consideration for further study.

8.4 Conclusions

There are two main factors affecting genetic characteristics in secondary forest populations: the process of colonisation and the density of individuals. Regenerated populations are usually founded by a small number of individuals, leading to a loss of alleles. Therefore, the number of adults contributing to the regenerated forest will affect the levels of genetic diversity and the relatedness of individuals. In complex landscapes a regenerated population

may be founded by individuals from different populations and resultant populations can show high levels of heterozygosity and also diversity. The density of individuals in secondary forest can decrease the gene flow from pollen and so impacts on the fine scale structure of populations and the extent of biparental inbreeding.

Both the molecular and modelling studies show that while colonisation appears to lead to a loss of alleles and potentially a loss of genetic diversity, the high levels of gene flow found in pioneer species, particularly during colonisation, may counter that effect with time. The differences in diversity found in *V. ferruginea* seedling populations with different age structures show regeneration by successive generations may counter diversity loss of founding populations by a replacement effect through seed dispersal over multiple generations. Diversity can be further increased in mature secondary populations as they experience continued gene flow via post-colonisation pollen-mediated gene flow. The different genetic characteristics found in the three populations of *V. ferruginea* dominated secondary forest also show that the process of colonisation and differences in the diversity and differentiation of adults contributing to founding populations, markedly affect regenerated populations and can dominant the genetic characteristics of long lived tree species.

This study also highlights the potential use of modelling for the management of genetic resources and demonstrates that it could be an important tool for informing sustainable forest management.

REFERENCES

Adams W.T., Griffin A.R. and Moran G.F. 1992. *Using paternity analysis to measure effective pollen dispersal in plant populations*. The American Naturalist **140**, 762-780

Aide T.M. and Grau H.R. 2004. *Globalization, Migration, and Latin American Ecosystems.* Science, **305** (5692) 1915-1916

Aizen M.A. and Feinsinger P. 1994. *Forest fragmentation, pollination and plant reproduction in a chaco dry forest, Argentina*. Ecology, **75**, 330-351

Aldrich P.R., Chavarriaga H.J.L. and Kochert G. 1998, *Microsatellite analysis of demographic genetic structure in fragmented populations of the tropical tree Symphonia globulifera.* Molecular Ecology, **7**, 933-944

Alvarez-Buylla E.R. and Martinez-Ramos R. 1990. *Seed bank versus seed rain in the regeneration of a tropical pioneer tree*. Oecologia, **84**, 314-325

Alvarez-Buylla E.R. and Garcia-Barrios R. 1993. *Models of patch dynamics in tropical forests.* Trends in Ecology and Evolution, **8** (6), 201-204

Alvarez-Buylla E.R. and Garay A.A. 1994. *Population genetic structure of Cecropia obtusifolia, a tropical pioneer species*. Evolution, **48**, 437-453

Alvarez-Buylla E.R., García-Barrios R., Lara-Morena C. and Martinez-Ramos M.1996. *Demographic and genetic models in conservation biology: Applications and perspectives for tropical rain forest trees*. Annual Review of Ecological Systematics, **2**, 387-421

Alvarez-Buylla E.R., Chaos à, Piñero D. and Garay A.A. 1996. *Demographic genetics of a pioneer tropical tree species: patch dynamics, seed dispersal, and seed banks*. Evolution, **50**, 1155-1166

Anderson P.J. and Putz F.E. 2002. *Harvesting and conservation: are both possible for the palm, Iriartea deltoidea?* Forest Ecology and Management, **170** (1-3), 271-283

Andow D., Kareiva P., Levin S.A. and Okubo A.1990. *Spread of invading organisms*. Landscape Ecology, **4**, 177–188

Arnaez E. and Moreira I. 1995. *Fenología de Dipteryx panamensis (almendro) en la Región Huetar Norte, Costa Rica.* In Diferentes voces mencionando un problema: situación de la lapaverde (Ara ambigua) en Costa Rica. Experiencia de aprendizaje para la concertación en el manejo delos recursos naturales. San José, Costa Rica: MINAE-UICN (ed. V. Solis, P. Madrigal, I. Ayales, J.Calvo Domingo)

Ausubel F.M., Brent R., Kingston R.E., Moore D.D., Seidman J.G., Smith J.A. and Struhl K.1987. *Current Protocols in Molecular Biology*. Green Publishing Associates and Wiley-Interscience, New York

Austerlitz F. and Smouse P.E. 2001a. *Two-generation analysis of pollen flow across a landscape. II. Relation between Φ_{FT}, pollen dispersal and interfemale distance.* Genetics, **157**, 851-857

Austerlitz F. and Smouse P.E. 2001b. *Two-generation analysis of pollen flow across a landscape. III. Impact of within-population structure*. Genetical Research, **78**, 271-280

Austerlitz F. and Smouse P.E. 2002. *Two-generation analysis of pollen flow across a landscape. IV. Estimating the dispersal parameter*. Genetics, **161**, 355-363

Austerlitz F. and Garnier-Géré P.H. 2003. *Modelling the impact of colonisation on genetic diversity and differentiation of forest trees: interaction of life cycle, pollen flow and seed long-distance dispersal.* Heredity, **90**, 282-290

Austerlitz F., Dick C.W., Dutech C., Klein E.K., Oddou-Muratorio S., Smouse P.E. and Sork V.L. 2004. *Using genetic markers to estimate the pollen dispersal curve*. Molecular Ecology, **13**, 937-954

Avise J.C. 1989. A role for molecular genetics in the recognition and conservation of endangered species. Trends in Ecology and Evolution, **4**, 279-281

Baluarte-Vasquez J., Freitas Alvarado L., van de Kop P., de Jong W., Inga-Sanchez H., Melendez Torres W. 2000. *El bosque secundario de la llanura aluvial inundable en la zona de Jenaro Herrera, Amazonia Peruana: uso, manejo y contribucion a la economia del poblador ribereno*. Revista Forestal del Peru, **23** (1-2), 51-62

Barrett S.C.H. and Charlesworth D. 1991. *Effects of a change in the level of inbreeding on genetic load*. Nature, **352**, 522-524

Barrett. S.C.H., Eckert C.G. and Husband B.C. 1993. *Evolutionary processes in aquatic plant populations*. Aquatic Botany, **44**, 105-145

Bawa K.S. 1977. *Reproductive biology of Cupania guatemalensis Radlk (Sapindaceae)*. Ibid, **31**, 52-63

Barton A.M. 1993. *Factors controlling plant distributions: drought, competition, and fire in montane pines in Arizona.* Ecological Monographs, **63**, 367–397

Bawa K.S., Perry D.R. and Beach J.H. 1985a. *Reproductive biology of tropical lowland rain forest trees. I. Sexual systems and incompatibility mechanisms*. American Journal of Botany, **72**, 331-345

Bawa K.S., Bullock S.H., Perry D.R., Coville R.E. and Grayum M.H. 1985b. *Reproductive biology of tropical lowland rain forest trees. II. Pollination systems*. American Journal of Botany, **72**, 346-356

Bawa K.S. 1990 *Plant–pollinator interactions in tropical rainforest*. Annual Review of Ecology and Systematics, **21**, 399-422

Bawa K.S. 1998. *Conservation of genetic resources in the Dipterocarpaceae*. In: A review of Dipterocarps: Taxonomy, Ecology and Silviculture. Eds Appanah A. and Turnbull J.M. Center for International Forestry Research, Bogor, Indonesia.

Beaumont M.A. and Nichols R.A. 1996. *Evaluating loci for use in genetic analysis of population structure*. Proceedings of the Royal. Society London. Series B, **263**, 1619-1626

Bennett K.D. 1997. *Evolution and Ecology: The Pace of Life*. Cambridge: Cambridge University Press.

Bialozyt R, Zeigenhagen B. and Petit R. 2004. *Long distance dispersal and genetic diversity during range expansion*. In: Plant population ecology and genetics in the cultural landscape. Eds. Eckstein R.L. and Fischer M. Verhandlungen der Gesellschaft für Ökologie

Birks H.J.B. 1989. *Holocene isochrone maps and patterns of tree spreading in the British Isles*. Journal of Biogeography, **16**, 503-540

Boshier D.H., Chase M.R. and Bawa K.S. 1995. *Population genetics of Cordia alliodora (Boranginaceae), A neotropical tree, 2, Mating system*. American Journal of Botany, **82**, 476-483

Boucher D.H., Vandermeer J.H., Mallona M.A., Zamora N. and Perfecto I. 1994. *Resistance and resilience in a directly regenerating rainforest: Nicaraguan trees of the Vochysiaceae after Hurricane Joan*. Forest Ecology and Management, **68**, 127-136

Boucher D.H. and Mallona M.A.1997. *Recovery of the rain forest tree Vochysia ferruginea over 5 years following Hurricane Joan in Nicaragua: A preliminary population projection matrix*. Forest Ecology and Management, **91** 195 - 204

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

Brewer S., Cheddadi R., de Beauliu J.L. and Reille M. 2002. *The spread of deciduous Quercus throughout Europe since the last glacial period*. Forest Ecology and Management, **156**, 27-48

Brohede J. and Ellegren H. 1999. *Microsatellite evolution: polarity of substitutions within repeats and neutrality of flanking sequences*. Proceedings of the Royal Society of London, Series B, **266**, 825-833

Brown A.H.D. 1989. Core collections: *A practical approach to genetic resources management*. Genome, **31**, 818-824

Brown A.H.D. 1990. *Genetic characterization of plant mating systems*, pp. 145–162 in Plant Population Genetics, Breeding, and Genetic Resources, edited by A. H. D. Brown, M. T. Clegg, A. L.Kahler and B. S. Weir. Sinauer Associates, Sunderland, MA

Bullock J.M. and Clark R.T. 2000. *Long distance seed dispersal by wind: measuring and modelling the tail of the curve*. Oecologia, **124**, 506-521

Burczyk J., Adams W.T. and Shimizu J.Y. 1996. *Mating patterns and pollen dispersal in a natural knobcone pine (Pinus attenuata Lemmon.) stand*. Heredity, **77**, 251–260

Burczyk J., DiFazio S.P. and Adams W.T. 2004. *Gene flow in trees: How far do genes really travel?* Forest Genetics, **11**, 1-14

Burczyk J. and Koralewski T.E. 2005. Parentage versus two-generation analyses for estimating pollen-mediated gene flow in plant populations. Molecular Ecology, **14** (8), 2525-2537

Bush M. B. and Rivera R. S. 1998. *Pollen dispersal and representation in a neotropical forest.* Global Ecology and Biogeography Letters, **7**, 78–9

Butcher P.A., Decroocq S., Gray Y. and Moran G.F. 2000. *Development, inheritance and cross-species amplification of microsatellite markers from Acacia mangium*. Theoretical Applied Genetics, **101**, 1282-1290

Cain M.L., Damman H. and Muir A. 1998. *Seed dispersal and the Holocene migration of woodland herbs*. Ecological Monographs **68** (3), 325-327

Cain M.L., Milligan B.G. and Strand A.E. 2000. *Long distance seed dispersal in plant populations*. American Journal of Botany, **87**, 1217-1227

Callen D.F., Thompson A.D., Shen Y., Phillips H.A., Richards R.I., Mulley J.C. and Sutherland G.R. 1993. *Incidence and origin of "null" alleles in the (AC)n microsatellite markers*. American Journal of Human Genetics, **52**, 922–927

Cannon C.H., Peart D.R. and Leighton M. 1998. *Tree Species Diversity in Commercially Logged Bornean Rainforest* Science, **281**, 1366-1368

Cardoso M.A., Provan J.N., Powell W., Ferreira P.C.G. and De Oliveira D.E. 1998. *High genetic differentiation among remnant populations of the endangered Caesalpinia echinata Lam. (Leguminosae-caesalpinioideae*). Molecular Ecology, **7**, 601-608

Cavers S. 2002. *Population genetic structure and phylogeography of two commercially important neotropical tree species: Vochysia ferruginea Mart. and Cedrela odorata L.*. PhD, University of Edinburgh

Cavers S., Navarro C., Hopkins P. and Lowe A.J. 2005. *Genetic diversity and population structure of Vochysia ferruginea Mart. in Costa Rica, assessed using cpDNA and AFLP markers.* Silvae Genetica, **54** (6), 258-264

Cercueil A., Bellemain E. and Manel S. 2002. *Parente: a software package for parentage analysis.* Journal of Heredity, **93**, 458–459

Cespedes M., Gutierrez M.V., Holbrook N.M. and Rocha O.J. 2003. *Restoration of genetic diversity in the dry forest tree Swietenia macrophylla (Meliaceae) after pasture abandonment in Costa Rica.* Molecular Ecology, **12**, 3201-3212

Chai C.K. 1976. *Genetic evolution*. Chicago: University of Chicago Press

Chakraborty R. and Hedrick P.W. 1983. *Paternity exclusion and the paternity index for two linked loci*. Human Heredity, **33**, 13-23

Chakraborty R., Meagher T.R. and Smouse P.E. 1988. *Parentage analysis with genetic markers in natural populations. I The expected proportion of offspring with unambiguous paternity*. Genetics, **118**, 527-536

Chambers J.C. and MacMahon J.A. 1994, *A day in the life of a seed: movements and fates of seeds and their implications for natural and managed systems.* Annual Review of Ecology and Systematics **25**: 263–292

Charlesworth B. and Charlesworth D.B. 1987 *Inbreeding depression and its evolutionary consequences*. Annual review of Ecological Systematics, **18**, 237-268

Charlesworth B. Charlesworth D.B and Morgan M.T. 1990a. *Genetic loads and estimates of mutation rates in highly inbred plant populations*. Nature, **347**, 380-382

Charlesworth D.B. and Charlesworth B. 1990. *Inbreeding depression with heterozygote advantage and its effect on selection for modifiers changing the outcrossing rate*. Evolution, **44**, 870–888

Charlesworth D., Morgan M.T. and Charlesworth B. 1990b. *Inbreeding depression, genetic load and the evolution of outcrossing rates in a multi-locus system with no linkage*. Evolution, **44**, 1469-1489

Chambers J.Q., Higuchi N. and Schimel J.P. 1998. *Ancient trees in Amazonia*. Nature, **391**, 135-136

Chase M.R., Moller C., Kesseli R. and Bawa K.S. 1996. *Distant gene flow in tropical trees*. Nature, **383**, 398-399

Chinea J.D. and Helmer E.H. 2001. *Diversity and composition of tropical secondary forests recovering from large-scale clearing: results from the 1990 inventory in Puerto Rico.* Forest Ecology and Management, **180**, 227-240.

Chinea J.D. 2002. *Tropical forest succession on abandoned farms in the Humacao Municipality of eastern Puerto Rico.* Forest Ecology and Management, **167** 195-207

Chokkalingam U. and de Jong, W. 2001. *Secondary forest: a working definition and typology*. International Forestry Review **3** (1) 19-26

Chokkalingam U., Bhat D.M. and von Gemmingen G. 2001. *Secondary forests associated with the rehabilitation of degraded lands in tropical Asia: a synthesis*. Journal of Tropical Forest Science, **13** (4), 816-831

Chung M.Y., Epperson B.K. and Chung M.G. 2003. *Genetic structure of age classes in Camellia japonica (Theaceae).* Evolution Int J Org Evolution, **57** (1), 62-73

Church P., Mowbray R., Berwick N. and Martin R. 1994. *Forestry and the Environment: Costa Rica Case Study*, CDIE Working Paper No. 205

Clark J.S. 1998. *Why trees migrate so fast: Confronting theory with dispersal biology and the paleorecord*. American Naturalist, **152** 204-224

Clark J.S., Silman M., Kern R., Macklin E. and HilleRisLambers J. 1999. *Seed dispersal near and far: Patterns across temperate and tropical forests*. Ecology, **80** (5), 1475-1494

Clark J.S., Silman M., Kern R., Macklin E. and HilleRisLambers J. 2001. *Invasion by extremes: Population spread with variation in dispersal and reproduction*. American Naturalist, **157**, 537-554

Cliff A. and Ord J.K. 1981. *Spatial Processes, Models and Applications*. London: Pion, pp.17-19

Cockerham C.C. and Weir B.S. 1993. Estimation *of gene flow from F-statistics.* Evolution, **47**, 855-863

Collevatti R.G., Grattapaglia D. and Hay J.D. 2001. *Population genetic structure of the endangered tropical tree species Caryocar brasiliense, based on variability at microsatellite loci*. Molecular Ecology, **10**, 346-356

Colpaert N., Cavers S., Bandou E., Caron H., Gheysen G. and Lowe A.J. (2005). *Sampling tissue for DNA analysis of trees: Trunk cambium as an alternative to canopy leaves.* Silvae Genetica, **54** (6), 265-269

Comps B., Gömöry D., Letouzey J., Thiébaut B. and Petit R.J. 2001. *Diverging trends between heterozygosity and allelic richness during postglacial* colonisation *in the European beech*. Genetics, **157**, 389-397

Condit R., Hubbell S.P. and Foster R.B. 1992. *Short-term dynamics of a neotropical forest: change within limits*. BioScience, **42**, 822-888

Condit R., Ashton P.S., Manokaron N., LaFrankie J.V., Hubbell S.P. and Foster R.B. 1999. *Dynamics of the forest communities at Pasoh and Barro Colorado: comparing two 50-ha plots*. Philosophical Trans Royal Society London B, **354** (1391), 1739-48

Connell J.H. 1971. On the role of natural enemies in preventing competitive exclusion in some marine mammals and rain forest trees. Pages 298-312 in P. J. den Boer and G. R. Gradwell, editors. Dynamics of populations. Center for Agricultural Publishing and Documentation, Wageningen, The Netherlands

Conner J.K. and Hartl D.L. 2004. *A primer of ecological genetics*. Sunderland MA Sinauer

Conte R. Nodari R.O., Vencovsky R., and Sedrez dos Reis M.S. 2003. *Genetic diversity and recruitment of the tropical palm, Euterpe edulis Mart., in a natural population from the Brazilian Atlantic Forest.* Heredity, **91**, 401–406

Cottrell J.E., Munro R.C., Tabbener H.E., Gillies A.C.M., Forrest G.I., Deans J.D., and Lowe A.J. 2002. *Distribution of chloroplast DNA variation in British oaks (Quercus robur and Q-petraea): The influence of postglacial* colonisation *and human management*. Forest Ecology and Management, **156**, 181-195

Cottrell J.E., Munro R.C., Tabbener H.E., Milner A.D., Forrest G.I. and Lowe A.J. 2003. *Comparison of fine-scale genetic structure using nuclear microsatellites within two British oakwoods differing in population history*. Forest Ecology and Management, **176**, 287–303

Crnokrak P. and Barrett S.C.H. 2002. *Perspective: purging the genetic load: a review of the experimental evidence*. Evolution, **56**, 2347–2358

Cronquist A. 1981. *An Integrated System of Classification of Flowering Plants*. Columbia University Press., New York.

Crow J.F and Kimura M. 1970. *An Introduction to Population Genetics Theory*. Harper and Row, New York

Crow J.F. 1988. *Eighty years ago: the beginnings of population genetics.* Genetics, **119**, 473–476

Dakin E.E. and Avise J.C. 2004. *Microsatellite null alleles in parentage analysis*. Heredity, **93** (5), 504-9

Danzmann R.G. 1997. *PROBMAX: a computer program for assigning unknown parentage in pedigree analysis from genotypic pools of parents and progeny*. Journal of Heredity, **88**, 333

Davies N., Villablanca F.X. and Roderick G.K. 1999. *Determining the source of individuals: multilocus genotyping in nonequilibrium population genetics*. Trends in Ecology and Evolution, **14**, 17-2

Dayanandan S., Attygalla D.N.C., Abeygunasekara A.W.W.L., Gunatilleke I.A.U.N. and Gunatilleke C.V.S. 1990. *Phenology and floral morphology in relation to pollination of some Sri Lankan dipterocarps.* In: Bawa, K.S. and Hadley, M. (eds) Reproductive Ecology of Tropical Forest Plants, pp. 103-133. UNESCO, Paris/ Parthenon Publishing Group, Carnforth, UK

Dayanandan S., Dole J., Bawa K. and Kesseli R. 1999. *Population structure delineated with microsatellite markers in fragmented populations of a tropical tree, Carapa guianensis (Meliaceae)*. Molecular Ecology, **8**, 1585-1592

Dawson I.K., Waugh R., Simons A.J. and Powell W. 1997. *Simple sequence repeats provide a direct estimate of pollen-mediated gene dispersal in the tropical tree Gliricidia sepium.* Molecular Ecology, **6**, 179-183

Degen B., Gregorius H.R. and Scholz F. 1996. *ECO-GENE, a model for simulation studies on the spatial and temporal dynamics of genetic structures of tree populations* Silvae Genetica, **45** (5-6), 323-329

Degen B., Petit R. and Kremer A. 2001. *SGS - Spatial genetic software: A computer program for analysis of spatial genetic and phenotypic structures of individuals and populations.* Journal of Heredity, **92**, 447-449

Degen B., Loveless M.D. and Kremer A. (Eds) 2002. *Modelling and experimental research on genetic processes in tropical and temperate forests*. Embrapa, Kourou, September 18-22 2000

Degen B. 2004. *Effects of animal pollination on pollen dispersal, selfing and effective population size of tropical trees: A simulation study*. Biotropica, **36**, 165–179

Degen B., Bandou E. and Caron H. 2004. *Limited pollen dispersal and biparental inbreeding in Symphonia globulifera in French Guiana.* Heredity, **93** (6), 585-91

Demesure B., Comps B. and Petit R.J. 1996. *Chloroplast DNA phylogeography of the common beech (Fagus sylvatica L.) in Europe*. Evolution, **50**, 2515 2520.

Devlin B., Roeder K. and Ellstrand N.C. 1988. *Fractional paternity assignment: theoretical development and comparison to other methods*. Theoretical Applied Genetics, **76**, 369–380

Devlin B. and Ellstrand N.C. 1990. *The development and application of a refined method for estimating gene flow from angiosperm paternity analysis.* Evolution, **44**, 248-259

Dick C.W. 2001. *Genetic rescue of remnant tropical trees by an alien pollinator.* Proceedings of the Royal Society of London Series B–Biological Sciences, **268**, 2391-2396

Dick C.W., Etchelecu G. and Austerlitz F. 2003. *Pollen dispersal of tropical trees (Dinizia excelsa: Fabaceae) by native insects and African honeybees in pristine and fragmented Amazoian rainforest*. Molecular Ecology, **12**, 753-764

DiFazio S.P., Slavov G.T., Burczyk J., Leonardi S. and Strauss S.H. 2004. *Gene Flow from tree plantations and implications for transgenic risk* In: Plantation Forest Biotechnology for the 21st Century, Eds: Christian Walter and Mike Carson. Rotorua, New Zealand, Forest Research

DiRienzo A., Peterson A.C., Garza J.C., Valdes A.M. and Slatkin M. *et al.* 1994. *Mutational processes of simple-sequence repeat loci in human populations*. Proceedings of the National Academy of Sciences, USA, **91**, 3166-3170

Doligez A. and Joly H.I. 1997. *Mating system of Carapa procera (Meliaceae) in the French Guiana tropical forest.* American Journal of Botany, **84**, 461-470

Doligez, A., Baril C. and Joly H.I. 1998. *Fine-scale spatial genetic structure with nonuniform distribution of individuals.* Genetics, **148**, 905-919

Dounias E. 1999. *Ecotone forêt-savane et système agraire des Tikar du haut Mbam*. In M Servant, S Servant-Vildary, eds, Dynamique à long terme des écosystèmes forestiers intertropicaux, Paris : Unesco, pp. 85-102

Dow B.D. and Ashley M.V. 1996. *Microsatellite analysis of seed dispersal and parentage of saplings of bur oak, Quercus macrocarpa.* Molecular Ecology, **5**, 615-627

Doyle J.J. and Doyle J.L. 1987. *Isolation of plant DNA from fresh tissue*. Focus, **12**, 13-15

Duchesne P., Godbout M.H. and Bernatchez L. 2002. *PAPA (Package for the analysis of parental allocation): a computer program for simulated and real parental allocation*. Molecular Ecology Notes, **2** 191-193

Dumolin S., Demesure B. and Petit R.J. 1995. *Inheritance of chloroplast and mitochondrial genomes in pedunculate oak investigated with an efficient PCR method*. Theoretical Applied Genetics, **91**, 1253-1256

Dutech C., Seiter J., Petronelli P., Joly H.I. and Jarne P. 2002. *Evidence of low gene flow in a neotropical clustered tree species in two rainforest stands of French Guiana*. Molecular Ecology, **11**, 725-738

Edwards A., Civitello A., Hammond H.A. and Caskey C.T. 1991. *DNA typing and genetic mapping with trimeric and tetrameric tandem repeats*. American Journal of Human Genetics, **49**, 746–756.

Edwards A., Hammond H.A., Jin L., Caskey C.T. and Chakraborty R. 1992. *Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups*. Proceedings of the National Academy of Sciences, **91**, 3166-3170

Eiten G. 1972. *The cerrado vegetation of Brazil.* Botanical Review, 38 201-341

El Mousadik A. and Petit R.J. 1996. *Chloroplast DNA phylogeography of the argan tree of Morocco*. Molecular Ecology, **5**, 547-555

Ellegren H. 2004. *Microsatellites: simple sequences with complex evolution*. Nature Reviews Genetics, **5**, 435-445

Ellstrand N.C. 1984. *Multiple paternity within the fruits of the wild radish, Raphanus sativus.* American Naturalist, **123**, 819-82

Ellstrand N.C. and Marshall D.L. 1985. *Interpopulation gene flow by pollen in wild radish, Raphanus sativus*. American Naturalist, **126**, 606-616

Ellstrand N.C., Devlin B. and Marshall D.L. 1989. *Gene flow by pollen into small populations: data from experimental and natural stands of wild radish*. Proceedings of the National Academy of Sciences, **86**, 9044-9047

Ellstrand N.C. 1992. *Gene flow among seed plant populations.* New Forests, **6**, 241–256

Endler J.A. 1977. *Geographic variation, speciation, and clines*. Princeton University Press, Princeton, New Jersey.

Ennos R.A. and Clegg M.T. 1982. *Effect of population substructuring on estimates of outcrossing rate in plant population.* Heredity, **48**, 283-292

Ennos R.A. 2001. *Inferences about spatial processes in plant populations from the analysis of molecular markers*. In: Integrating Ecology and Evolution in a Spatial Context, eds Silvertown, J. and Antonovics, J., pp.45-71. Blackwell Science, Oxford.

Epperson B.K. 1992. *Spatial structure of genetic variation within populations of forest trees*. New Forests, **6**, 257-278

Epperson B.K. and Alvarez-Buylla E. 1997. *Limited seed dispersal and genetic structure in life stages of Cecropia obtusifolia.* Evolution, **51**, 275-282

Epperson B.K. 2000. *Spatial genetic structure and non-equilibrium demographics within plant populations*. Plant Species Biology, **15**, 269-279

Estoup A., Tailliez C., Cornuet J.-M. and Solignac M. 1995. *Size homoplasy and mutational processes of interrupted microsatellites in two bee species, Apis mellifera and Bombus terrestris (Apidae)*. Molecular Biology and Evolution, **112**, 1074-1084

Evett I.W. and Weir B.S. 1998. *Interpreting DNA Evidence*: Statistical Genetics for Forensic Scientists. Sinauer Associates,Sunderland, MA, USA

FAO, FLD, IPGRI 2004. *Forest genetic resources conservation and management. Vol. 1: Overview, concepts and some systematic approaches*. International Plant Genetic Resources Institute, Rome, Italy

Fenster C.B., Armbruster W.S., Wilson P., Dudash M.R. and Thomson J.D. 2004. *Pollination syndromes and floral specialization.* Annual Review of Ecology, Evolution, and Systematics, **35**, 375-403

Ferris C., Oliver R.P., Davy A.J. and Hewitt G.M. 1993. *Native oak chloroplasts reveal an ancient divide across Europ*e. Molecular Ecology, **2**, 337-344

Ferris C., Oliver R.P., Davy A.J. and Hewitt G.M. 1995. *Using chloroplast DNA to trace postglacial migration routes of oaks into Britain*. Molecular Ecology, **4**, 731-738

Fichtler E., Clark D.E. and Worbes M. 2003. *Age and Long-term Growth of Trees in an Old-growth Tropical Rain Forest, Based on Analyses of Tree Rings and* $^{14}C^{1}$. Biotropica **35** (3), 306–317

Fimbel R.A. and Fimbel C.C. 1996. *The role of exotic conifer plantations in rehabilitating degraded forest lands: a case study from the Kibale Forest in Uganda*. Forest Ecology and Management, **81**, 215–226

Finegan B. 1992. *The management potential of neotropical secondary lowland rainforest.* Forest Ecology and Management, **47**, 295-321

Finegan B. and Camacho M. 1999. *Stand Dynamics in a logged and silviculturally treated Costa Rican rain forest 1988-1996.* Forest Ecology and Management, **121**, 177-189

Finegan B. and Delgado D. 2000. *Structural and Floristic Heterogeneity in a 30-year-old Costa Rican Rain Forest restored on pasture Through Natural Secondary Succession.* Restoration Ecology, **8**, 380-393

Flores E.M. 1993. *Trees and Seeds from the Neotropics.* Volume 2 (2). Museo National de Costa Rica

Fowler N.L. 1995. *Density-dependent demography in two grasses: a five-year study*. Ecology, **76**, 2145–2164

Franceschinelli E.V. and Bawa K.S. 2000. *The effect of ecological factors on the mating system of a South American shrub species (Helicteres brevispira).* Heredity, **84**, 116-123

Fuhr M., Delegue M. A. and Nasi R. 2001. *Vegetation structure, floristic composition and growth characteristics of Aucoumea klaineana Pierre stands as influenced by stand age and thinning*. Forest Ecology and Management, **140** (2-3), 117-132

Gerber S., Mariette S., Streiff R., Boénès C. and Kremer A. 2000. *Comparison of microsatellites and amplified fragment length polymorphism markers for parentage analysis.* Molecular Ecology, **9**, 1037-1048

Gerber S., Chabrier P. and Kremer A. 2003. *FaMoz: a software for parentage analysis using dominant, codominant and uniparentally inherited markers*, Molecular Ecology Notes, **3** (3), 479-481

GFC. 2002. *Code of practice for timber harvesting - second edition*. Guyana Forestry Commission, Georgetown, Guyana.

Ghazoul J., Liston K.A. and Boyle T.J.B. 1998. *Disturbance-induced density-dependent seed set in Shorea siamensis (Dipterocarpaceae), a tropical forest tree.* Journal of Ecology, **86**, 462-473

Ghazoul J. and McLeish M. 2001. *Reproductive ecology of tropical forest trees in logged and fragmented habitats in Thailand and Costa Rica*, Plant Ecology, **153** (2), 335–345

Gjertson D.W. and Morris J.W. 1995. *Assessing probability of paternity and the product rule in DNA systems.* Genetica, **96**, 89-98

Goldstein D.B., Roemer G.W., Smith D.A., Reich D.E. and Bergman A. *et al.* 1999 *The use of microsatellite variation to infer patterns of population structure and demographic history in a natural model system*. Genetics, **151**, 797-801

Goodman S.J. 1998. *Patterns of extensive genetic differentiation and variation among european harbor seals (Phoca vitulina vitulina) revealed using microsatellite DNA polymorphisms*. Molecular Biology and Evolution, **15**, 104–118

Goodnight K.F. and Queller D.C. 1999. *Computer software for performing likelihood tests of pedigree relationship using genetic markers.* Molecular Ecology, **8**, 1231-1234

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. Updated from Goudet, J. 1995. *FSTAT v-1.2. A computer program to calculate F-statistics.* Journal of Heredity, **86**, 485-486

Greene D.F. and Johnson E.A. 2000. *Tree recruitment from burn edges*. Canadian Journal of Forest Research, **30**, 1264-1274

Gribel R., Gibbs P. and Queir L. 1999. *Flowering Phenology and Pollination Biology of Ceiba pentandra (Bombacaceae) in Central Amazon.* Journal of Tropical Ecology, **15** (3), 247-263

Grivet D., Smouse P.E. and Sork, V.L. 2005. *A novel approach to an old problem: tracking dispersed seeds*. Molecular Ecology, **14**, 3585-3595

Guariguata M.R., Chazdon R.L., Denslow J.S., Dupuy J.M. and Anderson L. 1997. *Structure and floristic of secondary and old growth stands in lowland Costa Rica.* Plant Ecology, **132**, 107-120

Guariguata M.R. 1999. *Early response of selected tree species to liberation thinning in a young secondary forest in North-eastern Costa Rica.* Forest Ecology and Management, **124** (2-3), 255-261

Guariguata M.R. 2000. *Seed and seedling ecology of tree species in neotropical secondary forests: Management implications*. Ecological Applications, **10** (1), 145-154

Hall P., Orrell L.C. and Bawa K.S. 1994. *Genetic diversity and mating system in a tropical tree, Carapa guianensis (Meliaceae)*. American Journal of Botany, **81**, 1104-1111

Hamilton M.B. 1999. *Tropical tree gene flow and seed dispersal*. Nature, **40**, 129

Hamilton M.B. and Miller J. M. 2002. *Comparing relative rates of pollen and seed gene flow in the island model using nuclear and organelle measures of population structure.* Genetics, 162, 1897-1909

Hammond D.S. and Brown V.K. 1998. *Disturbance, phenology and life-history characteristics: factors influencing distance/ density-dependent attack on tropical seeds and seedling*s. In: Dynamics of Tropical Communities (eds Newbery, D.M., Prins, H.H.T. and Brown, N.D.). Blackwell Science, Oxford, pp 51–78

Hamrick J.L. 1982. *Plant population genetics and evolution*. American Journal of Botany, **69**, 1685-1693

Hamrick J.L. and Godt M.J.W. 1990. *Allozyme diversity in plant species.* In: Brown, A.H. D.; Clegg, M.T.; Kahler, A.L. and Weir, B.S. (Eds) Plant population genetics, breeding and genetic resources. Sinauer, Sunderland, Massachusetts, pp 43-63

Hamrick J.L., Godt M.J.W. and Sherman-Broyles S.L. 1992. *Factors influencing levels of genetic diversity in woody plant species*, New Forests, **6** (1), 95-124

Hamrick J.L., Murawski D.A. and Nason J.D. 1993. *The influence of seed dispersal mechanisms on the genetic structure of tropical tree populations*. Plant Ecology, **107**, 281-297

Hamrick J.L. and Gray M.J.W. 1996. *Conservation genetics of endemic plant species.* In H. J. L. Avise J. C., editor. Conservation Genetics: Case histories from Nature. Chapman and Hall

Hamrick J.L. and Nason J.D. 1996. *Consequences of dispersal in plants*. Pp. 203–236 in O. E. J. Rhodes, R. K. Chesser, and M. H. Smith, eds., Population dynamics in ecological space and time. The Univ. of Chicago Press, Chicago, IL

Hardesty B.D., Hubbell S.P. and Bermingham E. 2006. *Genetic evidence of frequent long-distance recruitment in a vertebrate-dispersed tree*. Ecology Letters, **9** (5), 516

Hardy O.J. and Vekemans X. 2002. *SpaGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population level*. Molecular Ecology Notes, **2**, 618-620

Hardy O.J., González-Martínez S.C., Fréville H., Boquien H., Mignot A., Colas B. and Olivieri I. 2004a. *Fine-scale genetic structure and gene dispersal in Centaurea corymbosa (Asteraceae) I. Pattern of pollen dispersal.* Journal of Evolutionary Biology, **17** (4), 795-806

Hardy O.J., González-Martínez S.C., Colas B., Fréville H., Mignot A. and Olivieri I. 2004b. *Fine scale genetic structure and gene dispersal in Centaurea corymbosa (Asteraceae). II. Correlated paternity within and among sibships*. Genetics, **168**, 1601-1614

Hardy O.J., Maggia L., Bandou E., Breyne P., Caron H., Chevallier M.H., Doligez A., Dutech C., Kremer A., Latouche-Halle C., Troispoux V., Veron V. and Degen B. 2006. *Fine-scale genetic structure and gene dispersal inferences in 10 neotropical tree species.* Molecular Ecology, **15** (2), 559-71

Harper J.L. 1977. *Population biology of plants*. Academic Press.

Harr B. and Schlötterer C. 2000. *Long microsatellite alleles in Drosophila melanogaster have a downward mutation bias and short persistence times, which cause their genome-wide under representation.* Genetics, **155**, 1213-1220

Harris S.A. and Ingram R. 1991. *Chloroplast DNA and biosystematics: the effect of intraspecific diversity and plastid transmission*. Taxon, **40**, 393-412

Hedrick P.W. 1999. *Perspective: Highly variable loci and their interpretation in evolution and conservation*. Evolution, **53**, 313-318

Hedrick P.W. 2000. *Genetics of Populations,* 2nd Edition. Sudbury MA, Jones and Bartlett Publishers Inc

Herrera B. and Finegan B. 1997. *Substrate conditions and foliar nutrients and the distributions of two canopy tree species in a Costa Rican secondary rain forest.* Plant and Soil 19**1**, 259–267

Herrera B., Campos J.J., Finegan B. and Alvarado A. 1999. *Factors affecting site productivity of a Costa Rican secondary rain forest in relation to Vochysia ferruginea, a commercially valuable canopy tree species.* Forest Ecology and Management, **118**, 73–81

Hewitt G.M. 1993. *Postglacial distribution and species substructure: lessons from pollen, insects and hybrid zones.* In: Lees DR, Edwards D (eds). Evolutionary Patterns and Processes. Linnean Society Symposium Series, **14**, 97-123 London Academic Press

Hewitt G.M. 1996. *Some genetic consequences of ice ages, and their role in divergence and speciation.* Biological Journal of the Linnaean Society, **58**, 247-276

Hewitt G.M. 1999. *Post-glacial re*-colonisation *of European biota*. Biological Journal of the Linnaean Society, **68**, 87-112

Heywood J.S. 1991. *Spatial analysis of genetic variation in plant populations.* Annual Review of Ecology and Systematics, **22**, 335-355

Higgins S.I. and Richardson D.M. 1996. *A review of models of alien plant spread*. Ecological Modelling, **87**, 249-265

Higgins S.L. and Richardson D.M. 1999. *Predicting plant migration rates in a changing world: the role of long distance dispersal.* American Naturalist, **153** (5), 464-475

Howe H.F. and Smallwood J. 1982. *Ecology of seed dispersal.* Annual Review of Ecological Systematics, **13** 201-228

Hubbell S.P. 2005. *Neutral theory in community ecology and the hypothesis of functional equivalence.* Functional Ecology 19, 166-172

Huntley B. and Birks H.J.B. 1983. *An atlas of past and present pollen maps for Europe.* Cambridge University press, Cambridge

Hurlbert S.H. 1971. *The non-concept of species diversity: A critique and alternative parameters*. Ecology, **52**, 577-586

Husband B.C. and Schemske D.W. 1995. *Magnitude and timing of inbreeding depression in a diploid population of Epilobium angustifolium (Onagraceae)*. Heredity, **75** 206-215

Husband B.C. and Schemske D.W. 1996. *Evolution of the magnitude and timing of inbreeding depression in plants.* Evolution, **50** (1), 54-70

Ibrahim K.M., Nichols R.A. and Hewitt G.M. 1996. *Spatial patterns of genetic variation generated by different forms of dispersal during range expansion*. Heredity, **77**, 282-291

Imbert E. and Lefèvre F. 2003. *Dispersal and gene flow of Populus nigra (Salicaceae) along a dynamic river system.* Journal of Ecology, **91**, 447-456

Janzen D.H. 1970. *Herbivores and the number of tree species in tropical forests*. American Naturalist, **104**, 501-528

Janzen D.H. 1971. *Seed predation by animals*. Annual Review of Ecology and Systematics, **2**, 465-492

Janzen D.H. 1984. *Dispersal of small seeds by big herbivores: foliage is the fruit.* American Naturalist, **123**, 338–353

Jennersten O. 1988. Pollination in Dianthus deltoides (Caryophyllaceae): *Effects of habitat fragmentation on visitation and seed set.* Conservation Biology, **2**, 359-366

Jones A.G., Kvarnemo C., Moore G.I., Simmons L.W. and Avise J.C. 1998. *Microsatellite evidence for monogamy and sex-biased recombination in the Western Australian seahorse, Hippocampus angustus*. Molecular Ecology, **7**, 1497-1505

Jones A.G. and Arden W.R. 2003. *Methods of parentage analysis in natural populations*. Molecular Ecology, **12**, 2511-2523

Kadmon R. and Shmida A. 1990. *Spatiotemporal demographic processes in plant populations: an approach and case study.* American Naturalist, **135**, 382-397

Kartawinata K., Riswan S., Gintings A.N. and Puspitojati T. 2001. *An overview of post-extraction secondary forests in Indonesia*. Journal of Tropical Forest Science, **13** (4), 621-638

Kaufman S.R., Smouse P.E. and Alvarez-Buylla E.R 1998. *Pollen mediated gene flow and differential male reproductive success in a tropical pioneer tree, Cecropia obtusifolia Bertol. (Moraceae): a paternity analysis.* Heredity, **81**, 164-173

Keller L.F. and Waller D.M. 2002 *Inbreeding effects in wild populations*. Trends in Ecology and Evolution, **17** (5), 230-241

Kenta T., Isagi Y., Nakagawa M., Yamashita M. and Nakashizuka T. 2004. *Variation in pollen dispersal between years with different pollination conditions in a tropical emergent tree.* Molecular Ecology, **13**, 3575-3584

Kevin K.S.Ng., Lee S.L. and Koh C. L. 2004. *Spatial structure and genetic diversity of two tropical tree species with contrasting breeding systems and different ploidy levels.* Molecular Ecology, **13** (3), 657-669.

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

King R.A. and Ferris C. 1998. *Chloroplast DNA phylogeography of Alnus glutinosa (L.) Gaertn*. Molecular Ecology, **4**, 95-103

Kitamura K., Abdul Rahman M.Y., Ochiai Y. and Yoshimaru H. 1994. *Estimation of the outcrossing rate on Dryobalanops aromatica Gaertn. f. in primary and secondary forest in Brunei, Borneo, Southeast Asia*. Plant Species Biology, **9**, 37-41

Knowles P., Perry D.J. and Foster H.A. 1992. *Spatial genetic structure in two tamarack [Larix laricina (Du Roi) K. Koch] populations with differing establishment histories.* Evolution, **46**, 572-576

König A.O., Ziegenhagen B., van Dam B., Csaikl U.M., Coart E., Degen B. *et al.* 2002. *Chloroplast DNA variation in western Central Europe and genetic consequences of human influences*. Forest Ecology and Management, **156**, 147-166

Konuma A., Tsumura Y., Lee C.T., Lee S.L. and Okuda T. 2000. *Estimation of gene flow in the tropical rainforest tree Neobalancarpus heimii (Dipterocarpaceae), inferred from paternity analysis.* Molecular Ecology, **9**, 1843-1852

Kreitman M. and Akashi H. 1995. *Molecular evidence for natural selection*. Annual Review of Ecological Systematics, **26**, 403-422

Lande R. and Schemske D.W. 1985. *The evolution of self-fertilization and inbreeding depression in plants. I. Genetic models.* Evolution, **39**, 24-40

Latouche-Halle C., Ramboer A., Bandou E., Caron H. and Kremer A. 2004. *Long-distance pollen flow and tolerance to selfing in a neotropical tree species.* Molecular Ecology, **13**, 1055-1064

Lazzeroni L.C. and Lange K. 1997. *Markov chains for Monte Carlo tests of genetic equilibrium in multi-dimensional contingency tables.* Annals of Statistics, **25**, 138-168

Leblois R., Rousset F. and Estoup A. 2004. *Influence of Spatial and Temporal Heterogeneities on the Estimation of Demographic Parameters in a Continuous Population Using Individual Microsatellite Data*. Genetics, **166**, 1081-1092

Le Corre V., Machon N., Petit R.J. and Kremer A. 1997. *Colonisation with long-distance seed dispersal and genetic structure of maternally inherited genes in forest trees: a simulation study.* Genetical Research Cambridge, **68**, 117-125

Le Corre V. and Kremer A. 1998. *Cumulative effects of founding events during* colonisation *on genetic diversity and differentiation in an island and stepping-stone model.* Journal of Evolutionary Biology, **11**, 495-512

Le Corre V. and Kremer A. 2003. *Genetic variability at neutral markers, quantitative trait loci and trait in a subdivided population under selection*. Genetics, **164** 2005-2019

Lemes M., Gribel R., Proctor J. and Grattapaglia D. 2003. *Population genetic structure of mahogany (Swietenia macrophylla King, Meliaceae) across the Brazilian Amazon: implications for conservation.* Molecular Ecology, **12**, 2875-2883

Lewontin R.C. and Krakauer J. 1973. *Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms*. Genetics, **74**, 175-195

Levin D.A., Kerster H.W. and Niedzlek M. 1971. *Pollinator flight directionality and its effect on pollen flow*. Evolution, **25**, 113-118

Levin D.A. and Kerster H.W. 1974. *Gene flow in seed plants*. Evolutionary Biology, **7**, 139-220

Levinson G. and Gutman G.A. 1987. *Slipped-Strand Mispairing: A Major Mechanism for DNA Sequence Evolution*. Molecular Biology and Evolution, **4**, 203-221

Lim L.S., Wickneswari R., Lee S.L. and Latiff A. 2002. *Genetic variation of Dryobalanops aromatica Gaertn. F. (Dipterocarpaceae) in Peninsular Malaysia using microsatellite DNA markers*. Forest Genetics, **9** (2), 127-138

Lira C.F., Cardoso R.S., Ferreira C.G. Cardoso M.A. and Provan J. 2003. *Long-term population isolation in the endangered tropical tree species Caesalpinia echinata Lam. Revealed by chloroplast microsatellites.* Molecular Ecology, **12**, 3219-3225

Litrico I., Ronfort J., Verlaque R. and Thompson J. D. 2005. *Spatial structure of genetic variation and primary succession in the pioneer tree species Antirhea borbonica on La Réunion.* Molecular Ecology, **14** (5), 1575-1584

Lobo J.A., Quesada M. and Stoner K.E. 2005. *Effects of pollination by bats on the mating system of Ceiba pentandra (Bombacaceae) populations in two tropical life zones in Costa Rica.* American Journal of Botany, **92**, 370–376

Loveless M.D. and Hamrick J.L. 1984. *Ecological determinants of genetic structure in plant populations*. Annual Review of Ecological Systematics, **15**, 65-95

Loveless M.D. 2002. *Genetic diversity and differentiation in tropical trees In: Modelling and experimental research on genetic processes in tropical and temperate forests* Eds Degen B, Loveless M.D. and Kremer A. Embrapa

Lowe A.J., Goodall-Copestake W.P., Caron H., Kremer A. and Decroocq S. 2002. *A set of polymorphic microsatellites for Vochysia ferruginea, a promising tree for land reclamation in the neotropics*. Molecular Ecology Notes, **2**, 153-155

Lowe A.J., Breyne P., Colpaert N., Navarro C., Wilson J. and Cavers S. 2003. *Fine-scale genetic structure and gene flow within Costa Rican populations of mahogany (Swietenia macrophylla)*. Heredity, **90**, 268-275

Lowe A., Harris S. and Ashton P. 2004. *Ecological Genetics: Design, Analysis and Application*. Blackwell Publishing, Malden, MA

Lowe A.J. *et al.* 2005. *Population genetics of neotropical trees, focus issue*. Heredity, **95**

Luikart G. and England. P.R. 1999. *Statistical analysis of microsatellite data*. Trends in Ecology and Evolution, **14**, 253-256

Lynch M. and Walsh B. 1998. *Genetics and analysis of quantitative traits*. Sunderland MA Sinauer

Lynch M. and Ritland K. 1999. *Estimation of pairwise relatedness with molecular markers*. Genetics, **152**, 1753-1766

Marshall T.C., Slate J., Kruuk L.E.B. and Pemberton J.M. 1998. *Statistical confidence for likelihood-based paternity inference in natural populations.* Molecular Ecology, **7**, 639-655

McCauley D.E. 1991. *Genetic consequences of local population extinction and re*colonisation. Trends in Ecology and Evolution, **6**, 5-8

Meagher T.R. 1986. *Analysis of paternity within a natural population of Chamaelirium luteum. I. Identification of most likely male parents*. American Naturalist, **128** 199-215

Meagher T.R. and Thompson E.A. 1987. *Analysis of parentage for naturally established seedlings within a population of Chamaelirium luteum*. Ecology, **68**, 803-812

Miyazaki Y. and Isagi Y. 2000. *Pollen flow and the intrapopulation genetic structure of Heloniopsis orientalis on the forest floor as determined using microsatellite markers.* Theoretical Applied Genetics, **101**, 718-723

Montagnini F. and Jordan C.F. 2005. *Tropical Forest Ecology: The Basis for Conservation and Management* Springer

Moran-Palma P. and Snow A. 1997. *The Effect of interplant distance on mating success in federally threatened self-incompatible Hymenoxys herbacea = H. Acaulis var. glabra (Asteraceae).* American Journal of Botany, **84** (2), 233-238

Morgan M.T. 1998. *Properties of maximum likelihood male fertility estimation in plant populations.* Genetics, **149**, 1099-1103

Muona O. 1989. *Population genetics in forest tree improvement*. pp. 282-298. In: A. H.D. Brown *et al.* (eds.), Plant Population Genetics, Breeding, and Genetic Resources. Sinauer Association, Sunderland, MA.

Murawski D.A. and Hamrick J.L. 1991. *The effect of the density of flowering individuals on the mating systems of nine tropical tree species.* Heredity, **67**, 167-174

Murawski D.A. and Hamrick J.L. 1992. *Mating system of Cavanellesia platanifolia (Bombacaceae) in relation to flowering tree density: a test of predictions*. Biotropica, **24**, 99-101

Murawski D.A. and Bawa K.S. 1994. *Genetic structure and mating system on Stemonoporus oblongifolius (Dipterocarapaceae) in Sri Lanka*. American Journal of Botany, **81**, 155-160

Murawski D.A., Nimal Gunatilleke I.A.U. and Bawa K.S. 1994. *The effect of selective logging on inbreeding in Shorea megistophylla (Diptocarpaceae) from Sri Lanka.* Conservation Biology, **8**, 997-1002

Murawski D.A. 1995. *Reproductive biology and genetics of tropical trees from a canopy perspective.* In: Loman M and Nadkarni N (eds) Forest Canopy. Academic Press, New York, pp 457-493

Nathan R. and Muller-Landau H.C. 2000. *Spatial patterns of seed dispersal, their determinants and consequences for recruitment.* Trends in Ecology and Evolution, **15**, 278-285

Nathan R. and Casagrandi R. 2004. *A simple mechanistic model of seed dispersal, predation and plant establishment: Janzen-Connell and beyond*. Journal of Ecology, **92**, 733-746

Nasi R. 1997. *Les peuplements d'Okoumé au Gabon. Leur dynamique et croissance en zone côtière*. Bois et Forêts des Tropiques, **251**, 5-27

Nason J.D. and Hamrick J.L. 1997. *Reproductive and genetic consequences of forest fragmentation: Two case studies of neotropical canopy trees.* Journal of Heredity, **88**, 264-276

Nason J.D., Aldrich P.R. and Hamrick J.L. 1997. *Dispersal and the dynamics of genetic structure in fragmented tropical tree populations*. In: Laurance, W. F., and R. O.

Bierregaard, Jr. (eds.), Tropical Forest Remnants: Ecology, Management, and Conservation of Fragmented Communities. University of Chicago Press, Chicago

Nason J.D. Herre E.A. and Hamrick J.L. 1998. *The breeding structure of a tropical keystone plant resource.* Nature, **391**, 685-687

Neff B.D., Repka J. and Gross M.R. 2001. *A Bayesian framework for parentage analysis: the value of genetic and other biological data*. Theoretical Population Biology, **59**, 315–331

Neigel J.E. 1997. *A comparison of alternative strategies for estimating gene flow from genetic markers*. Annual Review of Ecology and Systematics, **28**, 105-128

Nei M. 1973. *Analysis of gene diversity in subdivided populations*. Proceedings of the National Academy of Sciences, USA,**70**, 3321-3323

Nei M. 1975. *Molecular population genetics and evolution*. Edited by A. Neuberger and E. Tatum. Vol. 40, Frontiers of Biology. Amsterdam: North-Holland

Nei M., Maruyama T. and Chakraborty R. 1975. *The bottleneck effect and genetic variability in populations.* Evolution, **29**, 1-10

Nei M. 1987. *Molecular evolutionary genetics* / Masatoshi Nei. New York, Columbia

Newport M.E. 1989. *A test for proximity-dependent outcrossing in the alpine skypilot, Polemonium viscosum.* Evolution, **43**, 1110-1113

Nielsen R., Mattila D.K., Clapham P.J. and Palsboll P.J. 2001. *Statistical approaches to paternity analysis in natural populations and applications to the North Atlantic humpback whale.* Genetics, **157**, 1673-1682

Novick R., Dick C.W., Lemes M.R., Navarro C., Caccone A. and Bermingham E. 2003. *Genetic structure of Mesoamerican populations of Big-leaf mahogany (Swietenia macrophylla) inferred from microsatellite analysis*. Molecular Ecology, **12**, 2885-2893

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 Camb, **22** 201-204

Oliveira P. and Gibbs P. 1994. *Pollination biology and breeding systems of six Vochysia species (Vochysiaceae) in Central Brazil*. Journal of Tropical Ecology, **10**, 509–522

Oosterhout C.V., Hutchinson W.F., Wills D.P.M. and Shipley P. 2004. *Program Note Micro-Checker: software for identifying and correcting genotyping errors in microsatellite data.* Molecular Ecology Notes, **4**, 535-553

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

Pakeman R.J. 2001. *Plant migration rates and seed dispersal mechanisms*. Journal of Biogeography, **28**, 795-800

Packer A. and Clay K. 2003. *Soil pathogens and Prunus serotina seedling and sapling growth near conspecific trees.* Ecology, **84**, 108-119

Parker K.C., Hamrick J.L., Parker A.J. and Nason J.D. 2003. *Fine-scale genetic structure in Pinus clausa (Pinaceae) populations: effects of disturbance history*. Heredity, **87**, (1), 99-113

Parrota J.A., Turnbull J.W. and Jones N. 1997. *Catalyzing native forest regeneration on degraded tropical lands*. Forest Ecology and Management, **99**, 1-7

Pearce D. 2001 How valuable are the tropical forests? Demonstrating and capturing economic value as a means of addressing the causes of deforestation. Seminar paper for Conseil d'Analyse Économique, Seminaire Economie de L'Environnement, Et du Developpement Durable, Paris, December 2001

Pena S.D.J. and Chakraborty R. 1994. *Paternity testing in the DNA era*. Trends in Genetics, **10** 204-209

Petit R.J., Pineau E., Demesure B., Bacilieri R. and Ducousso A. 1997. *Chloroplast DNA footprints of postglacial recolonisation by oaks.* Proceedings of the National Academy of Sciences, USA, **94**, 9996-10001

Petit R.J., Bialozyt R., Brewer S., Cheddadi R. and Comps B. 2000. *From spatial patterns of genetic diversity to postglacial migration processes in forest trees. Integrating Ecology and Evolution in a Spatial Context* (eds Silvertown J. and Antonovics J.), Blackwell Science pp. 295-318

Petit R.J., Csaikl U.M., Bordacs S., Burg K., Coart E., Cottrell J.E., van Dam B. *et al.* 2002a. *Chloroplast DNA variation in European white oaks Phlogeography and patterns of diversity based on data from 2600 populations*. Forest Ecology and Management, **156**, 5-26

Petit R.J., Brewer S., Bordacs S., Burg, K., Cheddadi R.. Coart E., *et al.* (2002b). *Identification of refugia and post-glacial* colonisation *routes of European white oaks based on chloroplast DNA and fossil pollen evidence*. Forest Ecology and Management, **156**, 49-74

Piotto D., Montagnini F., Ugalde L. and Kanninen, M. 2003. *Growth and effects of thinning of mixed and pure plantations with native trees in humid tropical Costa Rica.* Forest Ecology and Management, **177** (1-3), 427-439.

Piotto D., Montagnini F., Ugalde L. and Kanninen, M. 2003. *Performance of forest plantations in small and medium-sized farms in the Atlantic lowlands of Costa Rica.* Forest Ecology and Management, **175** (1-3) 195-204

Piotto D., Montagnini F., Kanninen M., Ugalde L. and Viquez E. 2004. *Forest plantations in Costa Rica and Nicaragua: performance of species and preferences of farmers*. Journal of Sustainable Forestry, **18** (4), 59-77

Pons O. and Petit R.J. 1995. *Estimation, variance and optimal sampling of gene diversity I. Haploid locus.* Theoretical Applied Genetics, **90**, 462-470

Putz F.E., Blate G.M., Redford K.H., Fimbel R. and Robinson J. 2001. *Tropical Forest Management and Conservation of Biodiversity: an Overview.* Conservation Biology, **15** (1), 7-20

Ralls K.J., Ballou D. and Templeton A. 1988. *Estimates of lethal equivalents and the cost of inbreeding in mammals*. Conservation Biology, **2**, 185-193

Raymond M. and Rousset F. 1995. *GENEPOP, version 1.2: a population genetics software for exacts tests and ecumenicism*. Journal of Heredity, **86**, 248-249

Ribeiro R..A and Lovato M.B. 2004. *Mating system in a neotropical tree species, Senna multijuga (Fabaceae).* Genetics and Molecular Biology, **27**, 418–424

Richards C.M. 2000. *Genetic and demographic influences on population persistence: gene flow and genetic rescue in Silena alba*. Pp. 271–291 in Genetics, Demography and Viability of Fragmented Populations (A.G. Young and G.M. Clarke, eds.). Cambridge University Press, Cambridge, UK

Ritland K. and Jain S. 1981. *A model for the estimation of outcrossing rate and gene frequency using independent loci.* Heredity, **47**, 35-52

Ritland K. 1984. *The effective proportion of self-fertilization with consanguineous matings in inbred populations.* Genetics, **106**, 139-152

Ritland K. 1988. *The genetic mating structure of subdivided populations. II. Correlated mating models*. Theoretical Population Biology, **34**, 320-346

Ritland K. 1989. *Correlated matings in the partial selfer, Mimulus guttatus*. Evolution, **43**, 848-859

Ritland K. 1990. *A series of FORTRAN computer programs for estimating plant mating systems.* Journal of Heredity, **81**, 235-237

Ritland K. 1996. *MLTR: Multilocus mating system program. Version 1.1*. Available from the author

Ritland K. 2002. *Extensions of models for the estimation of mating systems using n independent loci.* Heredity, **88**, 221-228

Rocha O.J. and Aguilar G. 2001. *Reproductive biology of the dry forest tree Enterolobium cyclocarpum (guanacaste) in Costa Rica: a comparison between trees left in pastures and trees in continuous forest.* American. Journal of Botany, **88**, 1607-1614

Roeder K., Devlin B. and Lindsay B.G. 1989. *Application of maximum likelihood methods to population genetic data for the estimation of individual fertilities.* Biometrics, **45**, 363-379

Rossetto M., Jones R. and Hunter J. 2004. *Genetic effects of fragmentation in an early successional tree (Elaeocarpus grandis).* Heredity, **93,** 610-618

Rousset F. 2004. Genetic Structure and Selection in Subdivided Populations. Princeton University Press

Sambrook J., Fritsch E.F. and Maniatis T. 1989. *Molecular cloning: a laboratory manual*. Second edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, USA

Sanchez J.G. and Jacquez Tournon. 2000. *Evaluacion fitosociologica y etnobotanica de un bosque secundario cerca de Pucalipa- Ucayali, Amazonz peruana.* Revista Forestal del Peru, **23** (1-2), 79-90

Sancristobal M. and Chevalet C. 1997. *Error tolerant parent identification from a finite set of individuals.* Genetical Research, **70**, 53-62

Sauer J.D. 1988. *Plant migration: the dynamics of geographic patterning in seed plant species*. University of California Press, Berkeley, California, USA

Schierenbeck K.A., Skupski M., Lieberman D. and Lieberman M. 1997. *Population structure and genetic diversity in four tropical tree species in Costa Rica.* Molecular Ecology, **6**, 137-144

Schoen D.J. and Stewart S.C. 1986. *Variation in male reproductive investment and male reproductive success in white spruce.* Evolution, **40**, 1109-1120

Schlotterer C. and Tautz D. 1992. *Slippage synthesis of simple sequence DNA.* Nucleic Acids Research 20, 211-215

Schlötterer C. and Pemberton J. 1994. *Molecular ecology and evolution: approaches and applications: the use of microsatellites for genetic analysis of natural populations.* Basel: Birkhäuser

Schlötterer C. 2002. *A microsatellite-based multilocus screen for the identification of local selective sweeps*. Genetics, 160, 753–763

Schnabel A., Nason J.D. and Hamrick J.L. 1998. *Understanding the population genetic structure of Gleditsia trianthos L.: seed dispersal and variation in female reproductive success.* Molecular Ecology, **7**, 819-822

Sezen U.U., Chazdon R.L. and Holsinger K.E. 2005. *Genetic consequences of tropical second-growth forest regeneration.* Science, **307**, 891-892

Sherwin W.B. and Moritz C.M. 2000. *Managing and monitoring genetic erosion in genetics, demography and viability of fragmented populations.* Eds Young A.G. and Clarke G. M. Cambridge University Press

Shimatani K. and Takahashi M. 2003. *On methods of spatial analysis for genotyped individuals*. Heredity, **91**, 173–180

Skellam J.G. 1951. *Random dispersal in theoretical populations*. Biometrika, **38** 196-218

Slatkin M. 1985. Gene flow in natural populations. Annual Review of Ecological Systematics, **16**, 393-430

Slatkin M. 1989. *Population structure and evolutionary progress*, Genome, **31**,196-202

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

Smith J., Ferreira S., van de Kop P., Ferreira C.A.P. and Sabogal C. 2003. *The persistence of secondary forests on colonist farms in the Brazilian Amazon.* Agroforestry Systems, **58**, 125-135

Sloane M.A., Sunnucks P., Wilson A.C.C. and Hales D.F. 2001. *Microsatellite isolation, linkage group identification and determination of recombination frequency in the peach potato aphid, Myzus persicae (Sulzer) (Hemiptera: Aphididae)*, Genetical Research Cambridge, **77**, 251-260

Smouse P.E. and Meagher T.R. 1994 *Genetic analysis of reproductive contributions in Chamaelirium luteum (L.) Gray (Liliaceae).* Genetics, **136**, 313–322

Smouse P.E., Dyer R.J., Westfall R.D. and Sork V.L. 2001. *Two-generation analysis of pollen flow across a landscape. I. Male gamete heterogeneity among females*. Evolution, **55**, 260-271

Smouse P.E. and Sork V.L. 2004. *Measuring pollen flow in forest trees: an exposition of alternative approaches.* Forest Ecology and Management 19**7**, 21–38

Sokal R.R. and Oden N.L. 1978a Spatial autocorrelation in biology, 1. Methology. Biological Journal of the Linnean Society, **10** 199-228

Sokal R.R. and Oden N.L. 1978b Spatial autocorrelation in biology, 2. Some biological implications and four applications of evolutionary and ecological interest. Biological Journal of the Linnean Society, **10**, 229-249

Sokal R.S. and Wartenberg D.E. 1983. *A test of spatial autocorrelation analysis using an isolation-by-distance model.* Genetics, **105**, 219-237

Sokal R.R. and Jacquez G.M. 1991. *Testing inferences about microevolutionary processes by means of spatial autocorrelation analysis.* Evolution, **45**, 152-168

Sorensen A.E. 1986 *Seed dispersal by adhesion*. Annual Review of Ecology and Systematics, **17**, 443–463

Sork V.L., Nason J., Campbell D.R. and Fernandez J.F. 1999. *Landscape approaches to historical and contemporary gene flow in plants.* Trends in Ecology and Evolution, **14** (6), 219-224

Sork V.L., Davis F., Smouse P.E. *et al.* 2002. *Pollen movement in declining populations of California valley oak, Quercus lobata: where have all the fathers gone?* Molecular Ecology, **11**, 1657-1668

Sousa V.A. and Hattemer H.H. 2003. *Pollen dispersal and gene flow by pollen in Araucaria angustifolia*. Australian Journal of Botany, **51**, 309-317

Stacy E.A., Hamrick J.L., Nason J.D., Hubell S.P., Foster R.B. and Condit R. 1996. *Pollen dispersal in low density populations of three neotropical tree species*. American Naturalist, **148**, 275-298

Streiff R., Labbe T., Bacilieri R., Steinkellner H., Glössl J. and Kremer A. 1998. *Within-population genetic structure in Quercus robur L. and Quercus petraea (Matt.) Liebl. assessed with isozymes and microsatellites*. Molecular Ecology, **7**, 317-328

Takenaka A. 2001. *Individual-based model of a forest with spatial structure and gene flow*. Pages 415-420 in Matsuno T. and Kida H., editors. Present and future of modelling global environmental change: toward integrated modelling. Terrapub

Templeton A.R. 1986. *Co-adaptation and outbreeding depression*. In: M. Soule (ed.) Conservation Biology: The Science of Scarcity and Diversity, Sinauer, Sunderland MA.

Turner T.F., Salter L.A. and Gold J.R. 2001. *Temporal-method estimates of Ne from highly polymorphic loci*. Conservation Genetics, **2**, 297–308

Vekemans X. and Hardy O.J. 2004. *New insights from fine-scale spatial genetic structure analyses in plant populations*. Molecular Ecology, **13**, 921-935

Vandermeer J., Brenner A. and Granzow De La Cerda I. 1998. *Growth rates of tree height six years after hurricane damage at four localities in eastern Nicaragua*. Biotropica, **30**, 502–509

Van de Wal H. 1998. Chinantec shifting cultivation and secondary vegetation. BOS document 19, BOS foundation, Wageningen, Netherlands.

Waines J.G. and Hegde S.G. 2003. *Intraespecific gene flow in bread wheat as affected by reproductive biology and pollination ecology of wheat flowers*. Crop Science, **43**, 451-463

Wakeley J. 2005. *The limits of theoretical population genetics.* Genetics, **169** (1), 1-7

Wang J., Hill W.G., Charlesworth D. and Charlesworth B. 1999. *Dynamics of inbreeding depression due to deleterious mutations in small populations: mutation parameters and inbreeding rate*. Genetical Research, **74**, 165–178

Wang J. 2002. *An estimator for pairwise relatedness using molecular markers.* Genetics, **160**, 1203-1215

Ward M., Dick C.W., Gribel R., Lemes M., Caron H. and Lowe A.J. 2005. *To self, or not to self... A review of outcrossing and pollen-mediated gene flow in neotropical trees*. Heredity, **95**, 246-254

Waser N.M. and Price M.V. 1994. *Crossing-distance effects in Delphinium nelsonii: Outbreeding and inbreeding depression in progeny fitness.* Evolution, **48**, 842–852

Watson J.D., Gilman M., Witkowlsi J. and Zoller M. 1992. *Recombinant DNA.* Second edition. W. H. Freeman, New York, New York, USA

Wattier R., Engel C.R., Saumitou-Laprade P. and Valero M. 1998. *Short allele dominance as a source of heterozygote deficiency at microsatellite loci: experimental evidence at the dinucleotidelocus Gv1CT in Gracilaria gracilis (Rhodophyta).* Molecular Ecology, **7**, 1569-1573

Webb C.J. and Bawa K.S. 1983. *Pollen dispersal by hummingbirds and butterflies: a comparative study of two lowland tropical plants*. Evolution, **37**, 1258-1270

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 G.M., Boshier D.H. and Powell W. 1999. *Genetic variation within a fragmented population of Swietenia humilis Zucc.* Molecular Ecology, **8**, 1899-1909

White G.M., Boshier D.H. and Powell W. 2002. *Increased pollen flow counteracts fragmentation in a tropical dry forest: An example from Swietenia humilis Zuccarini.* Proceedings of the National Academy of Sciences, USA, **99** (4), 2038-2042

Whitlock M.C. and McCauley D.E. 1999 *Short Review: Indirect measures of gene flow - F_{ST} does not equal 1/(4Nm+1)*. Heredity, **82**, 117-125

Whittaker J.C., Harbord R.M., Boxall N., Mackay I., Dawson G. and Sibly R.M 2003. *Likelihood-based estimation of microsatellite mutation rates*. Genetics, **164**, 781-787

Willson M.F. 1988. *Spatial heterogeneity of post-dispersal survivorship of Queensland rainforest seeds*. Australian Journal of Ecology, **13**, 137-145

Willson M.F. 1993. *Dispersal mode, seed shadows, and colonisation patterns*. Vegetatio **107** (108), 261–280

Winkler E. and Stocklin J. 2002. *Sexual and vegetative reproduction of Hieracium pilosella L. under competition and disturbance: a grid based simulation model*. Annals of Botany, **89**, 525-526

Wolff R.K., Plaetke R., Jeffreys A.J. and White R. 1989. *Unequal crossing over between homologous chromosomes is not the major mechanism involved in the generation of new alleles at VNTR loci.* Genomics, **5**, 382-384

Worthington Wilmer J., Allen P.J., Pomeroy P.P., Twiss S.D. and Amos W. 1999. *Where have all the fathers gone? An extensive microsatellite analysis of paternity in the grey seal (Halichoerus grypus).* Molecular Ecology, **8**, 1417-1429

Wright S. 1931. *Evolution in Mendelian populations*. Genetics, **16**, 97-159

Wright S. 1940. *Breeding structure of populations in relation to spceiation*. American Naturalist, **74**, 232-248

Wright S. 1943. *Isolation by distance*. Genetics, **6**, 114-138

Wright S. 1951. *The genetical structure of populations*. Annals of Eugenics, **15**, 323-354

Wright S. 1965. *The interpretation of population structure by F-statistics with special regard to systems of mating*. Evolution 19, 395-420

Wright S. 1969 1978 *Evolution and Genetics of Populations, vols 1-4*. Chicago: University Chicago Press

Wright S.J. 2005. *Tropical forests in a changing environment*. Trends in Ecology and Evolution 20, 553-560

Zagt R.J. 1997. *Tree demography in the tropical rain forest of Guyana*. Ph.D. thesis, Utrecht University, Utrecht, The Netherlands. Tropenbos-Guyana Series 3, Tropenbos-Guyana Programme, Georgetown, Guyana.

Zagt R.J., van der Hout P., Phillips P.D. and van Gardingen P.R. 2005. *Logging scenarios for sustainable timber harvesting in the tropical rain forests of central Guyana.* In. Long-term responses of populations and communities of trees to selective logging in tropical rain forests in Guyana Eric J.M.M. Arets Tropenbos International

Zucchi M.I., Brondani R.P.V. and Pinheiro J.B. 2003. *Genetic structure and gene flow in Eugenia dysenterica DC in the Brazilian Cerrado utilizing SSR markers*. Genetics and Molecular Biology, **26**, 449-457

APPENDIX

Data validation

Linkage disequilibrium estimate

Each locus was tested for linkage disequilibrium. The P-value for genotypic disequilibrium was based on 200 permutations. Adjusted P-value for 5% nominal level is: 0.005000. Tests using FSTAT

Loci combination	P-value
A1-5 X A1-10	0.025
A1-5 X A1-15	0.005
A1-5 X A1-20	0.02
A1-5 X A1-35	0.17
A1-10 X A1-15	0.055
A1-10 X A1-20	0.07
A1-10 X A1-35	0.07
A1-15 X A1-20	0.01
A1-15 X A1-35	0.25
A1-20 X A1-35	0.005
A1-5 X A1-10	0.025

The tests estimated a significant linkage disequilibrium for some loci combinations, notably A1-5 x A1-35 (P = 0.17) and A1-15 x A1-35 (P = 0.25). However, this is less likely to be because of linked loci and more a consequence of a similar pattern of allelic drop out.

Genotyping error estimate

Each locus was tested for the identification of genotyping errors due to non-amplified alleles (null alleles), short allele dominance (large allele dropout) and the scoring of stutter peaks. Tests performed using MICROCHEKER (Oosterhout *et al.* 2004). All loci showed no evidence of genotyping errors.

Locus	Presence of null alleles	Large allele drop-out	Mis-scoring of stutters
A1-5	No evidence	No evidence	No evidence
A1-10	No evidence	No evidence	No evidence
A1-15	No evidence	No evidence	No evidence
A1-20	No evidence	No evidence	No evidence
A1-35	No evidence	No evidence	No evidence

1