Spatial Structure and Population Genetic Variation in a Eucalypt Species Complex.

Rachel King B.Sc.(Hons)

Australian School of Environmental Studies
Faculty of Environmental Sciences
Griffith University

Submitted in fulfilment of the requirement of the degree of Doctor of Philosophy
May 2004
ABSTRACT

In this study, the relative influences of selection, gene flow, and other evolutionary forces on the spatial structure of genetic variation within a eucalypt species complex (the spotted gums: genus *Corymbia*, section Poltaria) were assessed. The study investigated the spatial genetic structure among four putative species of spotted gum (broad-scale), as well as within a single population (fine-scale) of one species, using both molecular and quantitative markers. The spotted gum complex occurs naturally across a range of 2500 km in eastern Australia.

Spatial genetic variation within and between the four putative spotted gum species was examined using both chloroplast and nuclear markers. No significant differentiation was found between the three northern species of the complex, *C. citriodora*, *C. variegata* and *C. henryi*. The southern species, *C. maculata*, shared no haplotypes with any of the three northern species. These results disagree in part with those reported in a previous allozyme based study in which *C. henryi* was found to be significantly divergent from *C. variegata* (with which it is sympatric) and more closely aligned with *C. maculata*. Re-analysis of the allozyme data provided evidence of selection acting at the *PGM2* locus within populations of *C. variegata* and *C. henryi*. The exclusion of this locus from the data set led to concordance between the cpDNA and nDNA analyses. Restricted gene flow and evidence of isolation by distance were identified as the dominant processes influencing the contemporary distribution of the cpDNA haplotypes. No geographic structure of haplotypes was found and complex genealogical relationships between haplotypes indicated the combined effects of past fragmentation, range expansion and possible long distance dispersal events.

The variation and spatial structure in both neutral molecular markers and quantitative genetic traits were compared to explore the relative influences of dispersal and selection within a single eucalypt population. Both mature trees (n=130) from a natural population of *C. variegata* and their progeny (n=127) were sampled. A very high outcrossing rate (98%) was estimated for the population based on data from seven microsatellite loci. This suggested regular pollen-mediated gene flow into the population, further supported by the observed high levels of genetic diversity and
polymorphism. Significant positive spatial structure was found between parent trees occurring up to 150 m apart in the natural forest, although genetic distance between these individuals suggested limited relatedness (i.e. less than half-sib relatedness). The effect of pollen-mediated gene flow appears, therefore, to swamp any effect of nearest neighbour inbreeding which has been reported in other studies of eucalypt populations and has been attributed to limited seed dispersal.

Resistance to the fungal disease *Sporothrix pitereka* (Ramularia Shoot Blight) was measured on progeny from each of the population study trees. Substantial resistance variability was found, along with a high estimate in heritability of resistance (0.44 ± 0.06), indicating significant additive genetic variation within the population. Spatial analysis showed no significant spatial structure with resistant and susceptible genotypes apparently distributed randomly throughout the population. The lack of concordance between the molecular and quantitative markers suggests that there may be a cost to resistance. Temporal variation in the severity of disease outbreaks may have then led to differential selection of seedlings across many generations, maintaining variability in disease resistance and facilitating the apparent random distribution of disease resistant and susceptible genotypes throughout the population.

*C. variegata* is an important commercial forestry species. The identification of strong genetic control in the disease resistance trait, as well as significant adverse genetic and phenotypic correlations between susceptibility and growth traits, will aid future breeding programs. Controlled crosses between resistant genotypes from this population should result in strong genetic gains in both resistance and growth, with little costs associated with inbreeding depression due to the highly outcrossed nature of the population.
Declaration

This work has not previously been submitted for a degree or diploma at this or any other university. To the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the thesis itself.

Signed:

Date:
**Table of Contents**

Abstract ............................................................................................................ i  
Declaration ...................................................................................................... iii  
Table of Contents ............................................................................................ v  
List of Figures ................................................................................................ vii  
List of Tables ................................................................................................ viii  
Acknowledgments ........................................................................................... ix  

**CHAPTER 1: General Introduction**

1.1 Genetic Variation and Structure ................................................................. 1  
  1.1.1 Processes of Population Genetic Variation .......................................... 1  
  1.1.2 Processes of Genetic Structure ............................................................ 2  
  1.1.3 Processes of Speciation .................................................................... 4  
  1.1.4 Spatial Genetic Structure Among Forest Tree Populations .............. 7  
1.2 Study Species ............................................................................................ 9  
  1.2.1 Eucalypts .......................................................................................... 9  
  1.2.2 Spotted Gums .................................................................................. 11  
  1.2.3 Ramularia Shoot Blight .................................................................... 15  
1.3 Thesis Objectives ...................................................................................... 17  

**CHAPTER 2: Broad-scale Genetic Variation**

2.1 Introduction ............................................................................................... 19  
2.2 Objectives ................................................................................................. 20  
2.3 Methods .................................................................................................... 20  
  2.3.1 Sampling Strategy ........................................................................... 20  
  2.3.2 Molecular Methods ......................................................................... 22  
  2.3.3 Data Analysis .................................................................................. 23  
2.4 Results ....................................................................................................... 26  
  2.4.1 Sequence Polymorphism .................................................................. 26  
  2.4.2 Geographic Distribution of Haplotypes ........................................... 30  
  2.4.3 Nested Clade Analysis .................................................................... 34  
  2.4.4 Reassessment of Allozyme Data ....................................................... 39  
2.5 Discussion ................................................................................................. 43  
  2.5.1 Species Relationships ...................................................................... 43  
  2.5.2 Historical Processes and Contemporary Population Structure ....... 46
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1</td>
<td>Schematic distribution map</td>
<td>12</td>
</tr>
<tr>
<td>Figure 1.2</td>
<td>Mature <em>C. variegata</em> trees in a natural forest</td>
<td>14</td>
</tr>
<tr>
<td>Figure 1.3</td>
<td>Infected <em>C. variegata</em> leaves by <em>Quambalaria pitereka</em></td>
<td>16</td>
</tr>
<tr>
<td>Figure 2.1</td>
<td>Distribution map of sampled populations</td>
<td>21</td>
</tr>
<tr>
<td>Figure 2.2</td>
<td>Correlation between population pairwise genetic distance (<em>F&lt;sub&gt;ST&lt;/sub&gt;</em>) and geographic distance</td>
<td>29</td>
</tr>
<tr>
<td>Figure 2.3</td>
<td>Hypothesised haplotype network</td>
<td>32</td>
</tr>
<tr>
<td>Figure 2.4</td>
<td>19 haplotypes mapped according to the populations in which they occur</td>
<td>33</td>
</tr>
<tr>
<td>Figure 2.5</td>
<td>Hypothesised haplotype network by morpho-species</td>
<td>36</td>
</tr>
<tr>
<td>Figure 2.6</td>
<td>Average pairwise location distances</td>
<td>38</td>
</tr>
<tr>
<td>Figure 2.7</td>
<td>Cluster analysis</td>
<td>40</td>
</tr>
<tr>
<td>Figure 2.8</td>
<td>Multi-dimensional Scaling</td>
<td>41</td>
</tr>
<tr>
<td>Figure 2.9</td>
<td>Allele frequencies at PGM2 locus</td>
<td>42</td>
</tr>
<tr>
<td>Figure 3.1</td>
<td>Location map of study population</td>
<td>52</td>
</tr>
<tr>
<td>Figure 3.2</td>
<td>Adult and progeny allele frequencies at seven microsatellite loci</td>
<td>57</td>
</tr>
<tr>
<td>Figure 3.3</td>
<td>The number of loci mismatches for the first and second most likely paternal parents</td>
<td>61</td>
</tr>
<tr>
<td>Figure 3.4</td>
<td>Delta scores, measuring the reliability that the most likely paternal parent assigned is in fact the true parent</td>
<td>62</td>
</tr>
<tr>
<td>Figure 3.5</td>
<td>Distogram of adult trees (A) and progeny (B)</td>
<td>63</td>
</tr>
<tr>
<td>Figure 4.1</td>
<td>Percentages of the total phenotypic variance attributable to each effect in the open-pollinated progeny</td>
<td>79</td>
</tr>
<tr>
<td>Figure 4.2</td>
<td>Plot of BLUP values from 127 families of <em>C. variegata</em></td>
<td>81</td>
</tr>
<tr>
<td>Figure 4.3</td>
<td>Map of family BLUP values representing the relative genetic merit of maternal parent trees in the natural forest</td>
<td>82</td>
</tr>
<tr>
<td>Figure 4.4</td>
<td>A: Distogram showing the relationship between similarity in family disease and the proximity of adult trees using 50m spatial classes. B: Distogram of adult trees using 50m spatial classes. Genetic distance calculated from 7 microsatellite loci</td>
<td>83</td>
</tr>
</tbody>
</table>
List of Tables

Table 2.1: List of haplotypes ................................................................. 28
Table 2.2: Species group pairwise differences based on cpDNA .................. 29
Table 2.3: The number of individuals of each haplotype within each population and morphological species group .................................. 31
Table 2.4: Results of nested clade geographic distance analysis ................ 37
Table 3.1: Levels of genetic variation at seven microsatellite loci ............... 59
Table 4.1: Estimated heritabilities ($h^2$) with standard errors for each trait measured in the *C. variegata* open-pollinated progeny trial .............. 80
Acknowledgments

Thanks to Professor Jane Hughes who, as the principle supervisor of this study, always provided valuable advice and critical assessment. Jane, I am very glad that you were a part of this project and appreciated you letting me work as independently as I wanted, while still being available to answer my questions and encourage new ideas. I was fortunate and am grateful that I had this chance to learn from you. Co-supervisors Brad Potts and David Lee also provided valuable information, particularly in the early stages of the development of this project and in relation to the quantitative analysis component of this thesis. Special thanks to John Kanowski for proof reading the final draft of this thesis.

A number of organisations and individuals helped in the collection of samples for this study. I appreciate the cooperation of CSIRO Forestry and Forest Products who provided seed samples from multiple populations of *Corymbia citriodora*, *C. variegata*, *C. henryi* and *C. maculata*, and Rhonda Stokoe who provided leaf samples from the outgroup species *C. torelliana*. David Lee and AFFS Forestry Research provided access to the Woondum State Forest and the related progeny field trial. I particularly thank Alan Ward, Damien Cotter and John Oostenbrink from AFFS Forestry Research whose assistance in the collection of leaf samples from the Woondum State Forest and in the assessment of quantitative traits in the progeny trial was invaluable. I also thank Annette Ritchie who faced the shotguns with me.

Thanks to all the genetics geeks, past and present, who have contributed to the accumulated knowledge and experience in Jane’s lab. For providing advise, encouragement and support at various stages throughout this study I particularly thank Michael Arthur, Andrew Baker, Naomi Doak, David Gopurenko, James Holman, Mia Hillyer, Jing Ma, Mark Ponniah, Alicia Schultheis, Steve Smith, Alicia Toon and Jacinta Zalucki. I am especially grateful for the advise and support of Michael Arthur, Naomi Doak and Cath Moran throughout the course of this project, it was great sharing an office with you.

Finally, I would like to thank family and friends for their encouragement.
CHAPTER 1: General Introduction

1.1 Genetic Variation and Structure

Understanding the nature, source and maintenance of heritable variation is one of the fundamental aims of evolutionary biology. The significance of populations as the units of evolution was recognised by Dobzhansky in 1951, and population variation forms the basis of most evolutionary theories and hypotheses of speciation. Variation occurs among individuals within populations and among populations themselves. However, it is not merely its occurrence but rather the structure and significance of this variation, which is of interest in an evolutionary context (Mayr 1970).

1.1.1 Processes of Population Genetic Variation

Mutation and gene flow are the main mechanisms responsible for producing and maintaining intra-population genetic variation; while genetic drift and natural selection can act in opposition to reduce it. The relative influence of each of these processes is also related to the demographic and life history characteristics of the species in question, such as effective population size (Ne), the dispersal / migration potential and breeding system. Dispersal potential is a product of both the physiology of the organism and geographic elements, which may limit or enhance migration between populations (Endler 1977; Page and Holmes 1998).

Genetic drift is the random fluctuation each generation in allele frequencies as a result of random sampling among gametes within a population. Over sufficient generations genetic drift within an isolated population can lead to the loss of variation and ultimately the fixation of an allele. The probability of a newly arisen allele going to fixation within a population is 1/(2N), where N is the number of individuals in the population. Smaller isolated populations with small Ne’s, therefore, are more likely to experience the effects of genetic drift (Endler 1977; Page and Holmes 1998). In the absence of gene flow the effects of genetic drift will result in high levels of genetic differentiation between populations (e.g. $F_{ST}$=1 where there are fixed differences between populations). Gene flow can counter this effect, homogenising allele frequencies and reducing levels of differentiation. Measures such as $F_{ST}$ are also influenced by levels of intra-population variation (Hedrick 1999). Mutation, therefore, will reduce estimates of differentiation between populations even though alleles arising
in an isolated population may be unique to that population. The time to reach genetic
equilibrium between genetic drift and gene flow is, therefore, dependant on effective
population size \((N_e)\), migration \((m)\) and mutation rates \((\mu)\).

\[
F_{ST} = \frac{1}{1 + 4N_e(m + \mu)}
\]  

(1.1)

Equilibrium in this sense can take many generations to occur (Wright 1969; Endler
1977; Crow and Aoki 1984; Slatkin 1987).

Natural selection can both decrease certain allele frequencies and/or maintain allele
frequencies at non-neutral loci. Selection acts on the variation in genes determining
phenotypes, which affect the fitness of individuals. The fitness of an individual is
decreased in evolutionary terms only when deleterious effects stop (or limit) the
individual from reaching reproductive maturity and passing its genes on to the next
generation (Endler 1986). Selection can influence allele frequencies in a number of
ways. For example, directional and disruptive selection remove alleles from the
population based on the relative fitness of the phenotype they express. Frequency
dependant selection occurs where the fitness of a trait is dependant on the frequency of
individuals with that trait in the population. Positive frequency dependant selection
favours alleles in high frequency and often leads to the fixation of that allele over few
generations. In contras, negative frequency dependant selection favours alleles in low
frequency, so that over time alternative alleles are favoured. As selection increases the
frequency of one allele another becomes less frequent and will then be favoured.
Alternation between favoured alleles over time will result in changes in the relative
frequencies of alleles, but does not tend to remove alleles from the population.
Balancing selection will also act to maintain allelic variation because in each
generation, extreme forms are selected against, yet in the following generation they
reappear due to the independent assortment of alleles (Endler 1977; Endler 1986).

1.1.2 Processes of Genetic Structure

Structuring of genetic variation can occur between groups of individuals within a
population, between populations and ultimately between groups of populations leading
to a level of genetic differentiation resulting in speciation. Intra- and inter-population
 genetic structures are a product of the strength of the mechanisms controlling genetic
variation (i.e. gene flow, genetic drift, selection etc.), the interactions between them and the biogeographic history of the study organism. The observable spatial patterns, therefore, capture the cumulative effects of these forces acting over many generations (Endler 1977; Epperson 2003).

At the finest spatial scale, the extent of within population genetic structure between individuals is dependant on the level of gene flow (dispersal) and/or selection pressures. For example, in species where inbreeding and asexual reproduction are common, the degree of spatial structure is generally high. Conversely, outbreeding species with high rates of dispersal may show limited genetic structure even over broad spatial scales. Populations of a species generally occur in one or more of the following ways: as a cline or a series of gradually changing contiguous populations, as geographic isolates or populations that are geographically separated, or as relatively narrow intergradation/hybrid zones (Endler 1977; Savolainen and Kuittinen 2000).

Climatic variation arises primarily due to differential selection along environmental gradients throughout the geographic distribution of the contiguous populations. For a cline to form, the homogenising effects of gene flow must be small relative to the effects of selection (Endler 1977). Higher levels of gene flow between neighbouring populations than between geographically distant populations can offset some of the differentiating effect of strong selection pressure, maintaining a cline of gradual change even in areas of abrupt environmental change. The occurrence and slope of a cline (the rate of phenotypic change) is, ultimately, the product of two conflicting forces: selection which would make every population uniquely adapted to its local environment, and gene flow, which would tend to make all populations of a species similar (Mayr 1970; Endler 1977; Page and Holmes 1998).

Geographic isolates are populations that are separated from the rest of the species by an extrinsic barrier, which prevents free gene flow with other populations. In an isolated population, the differentiating influence of genetic drift outweighs any cohesive effect of limited gene flow. The extent of differentiation between isolated populations depends on the degree of discontinuity, that is, the geographic and temporal extent of the barrier. The occurrence of isolated populations depends on the structure of the environment and
the dispersal ability of the species. Consequently they tend to occur more frequently at
the periphery of a species’ range (Mayr 1970; Endler 1977).

Zones of intergradation can occur as part of a cline between populations, or between
subspecies or species, the latter known as hybrid zones. These zones can arise through
two different evolutionary processes: primary differentiation and secondary contact
(Endler 1977). Primary differentiation occurs gradually over time while populations are
in continuous contact. In response to environmental change, a corresponding change in
selection pressure occurs leading to a steepening in the slope of the cline in the zone of
intergradation (Endler 1977). Secondary contact refers to cases in which two
populations have been isolated from each other for a sufficiently long time for genetic
divergence to have occurred through the accumulation of mutations, genetic drift and
localised selection pressures. When geographic isolation breaks down and contact is re-
established, gene flow between the two populations increases the genetic variation
within each and diminishes differentiation. The extent of genetic homogenisation
evident after secondary contact is dependent on the length of time populations were
isolated and the degree to which gene flow has been re-established (Endler 1977; Hewitt

1.1.3 Processes of Speciation
Speciation requires significant genetic divergence between populations and, therefore,
some level of segregation for population processes to achieve divergence (specifically
in diploid, outbreeding organisms). The two key hypothetical modes of speciation differ
with respect to the way in which it is believed that effective segregation can occur. The
process of allopatric speciation requires both spatial segregation and spatial
differentiation to initiate the process, leading to divergence and the development of
reproductive isolation during the time that differentiating populations are geographically
disjunct. Sympatric speciation, however, requires neither spatial segregation nor spatial
divergence. It is hypothesised that sympatric speciation could occur where there is
ecological or temporal segregation, assortative mating or host choice, resulting in
significant genetic divergence. However, this concept is controversial and there is no
general agreement on the occurrence of sympatric speciation in nature (Mayr 1970;
The potential for genetic divergence differs between animals and plants; in particular the potential for sympatric speciation in diploid, outbreeding plants is limited (Endler 1977). In animals, species-specific niche requirements often lead to a large degree of specialization. Consequently, hybridization between two closely related animal species results in increased variation, and reduced efficiency of the hybrids. Hybrid animals, therefore, are often less fit and selected against (Endler 1977; Harrison 1998). In plants, however, both genetic variability and phenotypic plasticity are favoured by selection (Endler 1977; Hamrick and Nason 2000). Plants are less mobile and cannot actively search for mates and suitable habitat. Seed must succeed where it falls and increased variability increases the potential of the plant to cope with a changing environment from which it cannot leave. Furthermore, the transfer of pollen is extrinsic to the plant (e.g. via wind or insects) and so transfer between species (or ‘mistakes’ in mate choice) are more likely, resulting in hybridization. For these reasons, both the ecological requirements and the reproductive mechanisms of plants facilitate hybridization and a lack of genetic divergence (Endler 1977; Templeton1998a; Hamrick and Nason 2000).

There is still much debate, however, regarding the level of divergence required between populations to constitute speciation, particularly in plants. Such debate has led to the development of several species concepts (only the major ones are considered in detail here) each attempting to define the criteria on which species can be differentiated. The biological species concept defines a species as a reproductive community whose boundaries are defined by a lack of potential gene flow with other such communities. This concept has worked relatively well for vertebrate species where morphological, ecological and geographic groups of individuals often coincide with the capacity to interbreed, and the testing of infertility between hybrids is relatively easy (Judd et al. 1999). However, mating success and gene flow vary greatly in and between plant populations. Subsequently, plant taxonomy has been historically dominated by the concept of morphological differentiation equating to species. This morphological species concept was often based on character similarities and differences across multiple morphological traits (Judd et al. 1999). The morphological species concept does not recognize lack of gene flow as central in the role of defining species, as is required by the biological species concept. Rather, the morphological species concept allows for some gene flow between species, acknowledges the occurrence of hybridization and embraces speciation in the presence of gene flow. Further, the
morphological species concept recognises the importance of evolutionary forces other than gene flow, such as selection on quantitative traits (Judd et al. 1999). For speciation to occur it is not necessary for selection to cause the cessation of gene flow, merely to override it, hence producing morphological divergence while evidence of gene flow may still exist (Mayr 1970; Endler 1977; Harrison 1998).

However, a classification system dominated by morphological similarities and differences does not recognise the importance of historical continuity within an evolutionary lineage. The use of phylogeny as an organising principle in systematics prompted the rise of the phylogenetic species concept, where the boundary between reticulate relationships is used to define species. However, this boundary is often quite ambiguous and at least three quite distinct criteria have been proposed as boundary criteria: monophyly (Donoghue 1985; Mishler 1985); unique combinations of character states (Nixon and Wheeler 1990); and basal exclusivity (Baum and Shaw 1995). The cohesion species concept extends the notion of basal exclusivity, as it requires the historical continuity of genetic cohesion within an evolutionary lineage. However, it still allows for limited inter-specific gene flow/hybridisation between lineages (Templeton 1989; Hull 1997; Hull 1999; Templeton 2001). Although there is no consensus about species concepts in plants, the cohesion species concept can apply equally to sexual and asexual taxa and acknowledges the opposing influences of asexual reproduction and hybridisation that challenge many other species concepts. This approach is therefore particularly applicable to plant populations where species boundaries are often confounded (and discordant with morphospecies boundaries) by the influences of hybridisation and selfing on population genetic structure (Hull 1997; Hull 1999; Judd et al. 1999; Schaal and Olsen 2000; Templeton 2001).

The cohesion species concept and coalescent theory also provide the means for testing plant evolutionary and population genetics hypotheses. Under coalescent theory, a cline, resulting from primary differentiation, for example, would represent a single evolutionary lineage and therefore a single species. However, a cline resulting from secondary contact between two formerly allopatric and independent evolutionary lineages would not (Hull 1997; Hull 1999).
Analytical techniques grounded in coalescent theory are not constrained by assumptions of equilibrium between gene flow and genetic drift which are considered unlikely to exist in most (particularly plant) species. Disequilibrium may be due to major climatic changes in recent evolutionary history dramatically influencing distribution patterns, or a tendency for substructuring to occur, with subpopulations cycling between colonisation and extinction events (i.e. metapopulation structure) (Templeton 1998a; Schaal and Olsen 2000) The development of nested clade analysis (NCA), in particular, has enhanced our ability to test hypotheses aimed at distinguishing between alternative historical processes leading to the contemporary distribution of haplotypes (Templeton 1998a). As speciation occurs at the interface of intra- and interspecific evolution, NCA extends our use of intraspecific phylogeographic techniques (based on coalescent theory) to infer the influence of historical and recurrent processes on the evolution of populations and closely related species (Posada and Crandall 2001; Templeton 2001; Masta et al. 2003).

1.1.4 Spatial Genetic Structure Among Forest Tree Populations

Forest trees are often thought to exist as large continuous populations. However there are many studies that have shown the occurrence of small isolated populations within a species distribution (eg. Pinus radiata, Moran and Bell 1987; Eucalyptus spp., Moran 1992; Quercus robur and other hardwoods, Mattilla et al. 1994). Historical episodes of range expansion and contraction in response to climate change are also likely in currently widespread species. During periods of severe range contraction resulting in population bottlenecks, the influence of small population processes such as genetic drift becomes more influential in determining the genetic structure between populations (Savolainen and Kuittinen 2000).

Plant dispersal can occur during two independent life cycle stages: (i) pollen dispersal before fertilisation, and (ii) seed dispersal after fertilisation and embryo development. Gene flow, therefore, occurs when a seedling develops within a population due to fertilisation from immigrant pollen, and/or where seed dispersal occurs from one population to another. Although this dual mode of gene flow is not unique to plants, it is an important difference between plants and most animals (Hamrick et al. 1993; Hamrick and Nason 2000). Long-range pollen dispersal and localised seed dispersal are characteristic of many forest trees (Moran 1992). This can result in broad scale
homogenisation of variation between geographically distant populations in conjunction with fine scale differentiation between family groups (Hamrick et al. 1993; Hamrick and Nason 1996). Studies of forest tree species in the northern hemisphere have shown similar levels of genetic differentiation among populations of temperate (mostly wind-pollinated) and tropical (mostly insect-pollinated) tree species (Hamrick et al. 1992; Chase et al. 1995; Rocha and Lobo 1996). This suggests that, among tree species in general, seed dispersal may be relatively more important than pollen dispersal in determining spatial genetic structure (Hamrick et al. 1992).

The mating system of tree species also influences spatial genetic structure. In the absence of self-incompatibility mechanisms, or where such mechanisms are incomplete, monoecious species are capable of self-fertilization (Boshier 2000). Inbreeding at this level severely reduces genetic diversity and, where seed dispersal is limited, further increases differentiation between family groups within a population. Differences in flowering phenology within and between populations can also substantially affect population structure. Within a population, flowering times may differ between groups of trees due to microenvironmental fluctuations. Asynchronous flowering times between populations or between groups within a population may lead to inbreeding between proximal individuals, which may already be closely related to each other due to limited seed dispersal. Individual trees may also disproportionately contribute genetic material to the next generation, reducing the effective population size and consequently reducing variation across the whole population (Boshier 2000; Epperson 2003).

Interactions between selection, dispersal and spatial structure can have important consequences on population inbreeding. Bi-parental inbreeding (breeding between related individuals) in small finite populations may reduce the average fitness of individuals within a population, decreasing the population size further. This could result in even more inbreeding, leading to extinction. Fenster (1991) showed that in Chamaecrista fasciculata, short-distance inbred pollinations produced less fit individuals that were themselves less likely to reproduce. In this case, selection favoured increased distances of effective pollen dispersal. Increased homozygosity as a result of inbreeding can lead to inbreeding depression when selection acts against homozygotes. This may occur where individuals homozygous for recessive deleterious alleles are less fit (dominance), or where all homozygotes are less fit than heterozygotes.
(over-dominance). Inbreeding depression can lead to substantial changes in the spatial structure of plants (Fenster and Galloway 2000; Epperson 2003).

The assessment of quantitative genetic variation can potentially provide a different view of spatial genetic variation and structure from that identified using molecular markers of neutral or unknown adaptive significance. Traits under quantitative genetic control are generally traits of ecological significance and the scale at which selection acts is often trait specific. Heritability of traits is the ratio of the additive genetic to total phenotypic variance measured (Falconer 1960). A measure of heritability therefore is based on the degree of similarity between relatives and is population specific. The spatial distribution of a highly heritable polymorphic trait can reflect the relative influences of dispersal (gene flow) and selection within the population.

Many plant species are susceptible to infection by pathogens, and variation in the physical effects of infection on growth and form among individuals is often used to estimate genetic variation in resistance (Kover and Schaal 2002). The potential for some pathogens to detrimentally affect the overall fitness of the plant host can result in selection acting against susceptible trees. Disease resistance, therefore, can be an important factor in the population dynamics of trees (Byrne 2000). Host-pathogen interactions are often very complex, however basic theoretical models describing the co-evolution and survival of both host and pathogen propose genes for resistance in the host and genes for virulence in the pathogen. Both genes exert a selective force in the population and both may have fitness costs associated with them (Byrne 2000). Host-pathogen relationships are discussed in more detail in Chapter 4.1.

1.2 Study Species

1.2.1 Eucalypts

Eucalypts species dominate extensive areas of the forested regions of Australia. Currently there are over 700 eucalypt species in three genera (Eucalyptus, Corymbia and Angophora) recognised in Australia (Williams and Brooker 1997). Classification of eucalypts has often been considered difficult due to their propensity for hybridisation and because fine morphological differences across multiple traits have often been used to distinguish between closely related species (Ladiges 1997).
Closely related eucalypt species tend to reflect a wide range of distribution patterns and varying levels of genetic diversity accompanied by variable yet consistent morphological differentiation between species. The group in general, therefore, lends itself well to the study of historic and contemporary patterns of gene flow; genetic sub-structuring and differences in genetic variability between and within populations; and the quantitative assessment of variation in the expression of traits of ecological significance (Potts and Wiltshire 1997).

As with most forest tree species, wide-ranging eucalypts have been found to have greater overall and within population genetic diversity than species of more restricted distribution, while regional or localised species have tended to show greater genetic diversity between populations (Potts and Wiltshire 1997). Widespread species have also tended to have a lower proportion of localised alleles and more alleles per polymorphic locus than regional species. Differentiation has also regularly been found to be greatest among populations of species that are both regional and disjunct (Moran and Hopper 1987; Potts and Wiltshire 1997). Eucalypts have small woody seeds and most are characterised by limited seed dispersal, possibly resulting in individual populations forming a ‘mosaic of family-group clusters’ due to reduced gene flow and subsequent neighbourhood inbreeding and disequilibrium (non-random association of alleles from different loci). However, there have, been few studies of genetic variation at the family level for eucalypt species, nor on the relative importance of gene flow and selection in maintaining genetic spatial sub-structuring within populations (Griffin 1980; Eldridge et al. 1993).

Within the eucalypt group, many species are of commercial interest and hence there are numerous studies of the distribution of genetic variation in quantitative traits (Eldridge et al. 1993; Dutkowski and Potts 1999). There are, however, few studies addressing whether the heritable variation of these traits is spatially or ecologically structured within a single natural population (Potts and Wiltshire 1997). Such studies may identify superior attributes of individual trees for future breeding improvement programs with commercial outcomes. They may also provide information regarding the relative influences of localised genetic processes such as selection, and any inbreeding effects resulting from limited seed dispersal, on genetic structure.
1.2.2 **Spotted Gums**

The genus *Corymbia* (one of the three genera encompassing all eucalypt species) was established by Hill and Johnson (1995). Within this genus, four morphological species were classified into the informal section Politaria, and are collectively known as the spotted gums: *Corymbia citriodora* (Hook) K.D Hill & L.A.S Johnson (syn. *Eucalyptus citriodora*), *C. variegata* (F. Muell.) K.D Hill & L.A.S Johnson, *C. henryi* (S.T. Blake) K.D Hill & L.A.S Johnson (syn. *E. henryi*) and *C. maculata* (Hook) K.D Hill & L.A.S Johnson (syn. *E. maculata*). The taxonomic standing of these species has been controversial and some taxonomists recognise *C. variegata* populations as a northern extension of *C. maculata* or *Eucalyptus maculata* (Brooker and Kleinig 1999). A study by McDonald *et al.* (2000) of the intra- and interspecific allozyme variation in the spotted gum complex identified generally low levels of divergence between the species and two distinct genetic alliances, one between *C. citriodora* and *C. variegata* and another between *C. henryi* and *C. maculata*. In this study *C. citriodora* and *C. variegata* could not be distinguished based on the allozyme profiles, prompting *C. variegata* to be formally subsumed under *C. citriodora* as a subspecies (McDonald and Bean 2000; McDonald *et al.* 2000). For ease of reference, the four morphological entities will be referred to as species for the remainder of this thesis.

The geographic distribution of the species complex extends throughout the eastern states of mainland Australia, from northern Queensland to eastern Victoria, a distance of approximately 2500 km (Figure 1.1). This range encompasses a variety of habitats including upland tropical regions and coastal and sub-coastal woodlands and forests. Three of the species, *C. citriodora*, *C. variegata* and *C. maculata*, replace each other latitudinally, while *C. henryi* has a relatively restricted distribution that is broadly sympatric with the middle member of the north-south series, *C. variegata*. The ranges of *C. citriodora* and *C. variegata* overlap to some extent. However *C. variegata* and *C. maculata* appear to have only limited contact. *C. citriodora* and *C. maculata* are wholly allopatric, as are *C. henryi* and *C. maculata*. *C. citriodora* and *C. maculata* have large disjunctions in their ranges. *C. variegata* also has a number of discontinuities in its distribution, although not as extensive as *C. citriodora* and *C. maculata* (Hill and Johnson 1995; McDonald *et al.* 2000).
**Figure 1.1:** Schematic distribution map of the four species of the spotted gum complex along the east coast of Australia. Only major distribution discontinuities are shown.
Morphological differentiation between the four study species is based on fine scale differences in the size and shape of adult leaves and capsules. The presence of citronella producing oil glands on the leaves of *C. citriodora* is regarded as the sole differentiating character between *C. citriodora* and *C. variegata*, and also distinguishes it from the other two species. *C. henryi* is generally considered to have thicker and more coarsely veined leaves, with fruits and buds on thicker pedicels, than both *C. variegata* (with which it is sympatric) and *C. maculata*. However, there is considerable overlap in observed size ranges. The adult and juvenile leaves of *C. variegata* are generally narrower than those of *C. maculata* (Hill and Johnson 1995; McDonald et al. 2000).

Flowering is often irregular and is influenced by environmental conditions. Several years between flowering events and multiple flowering events within a year have both been reported (Hill and Johnson 1995). Pollination is thought to occur via insects, birds, arboreal mammals and megachiropteran bats (Hill and Johnson 1995; Southerton et al. 2004). Several species of flying-foxes (megachiropterans) including *Pteropus scapulatus* and *P. poliocephalus* are known to feed on both *C. citriodora* and *C. maculata*, and may undertake large-scale migrations to feed on mass flowering eucalypts, including these species of spotted gums (Southerton et al. 2004). Flying foxes have also been observed with large amounts of pollen on their fur and may act as long-distance pollen vectors of eucalypts (Southerton et al. 2004). In contrast, seed release from capsules of spotted gums is passive, resulting in limited dispersal dependant on tree height (30-50 m) and wind conditions (Hill and Johnson 1995; McDonald et al. 2000).
Figure 1.2: Mature *C. variegata* trees in a natural forest.
1.2.3 Ramularia Shoot Blight

*Quambalaria pitereka* (Simpson) (previously known as *Ramularia pitereka*/*Sporothrix pitereka*) is a mitosporic fungus commonly known as Ramularia Shoot Blight (RSB). It is host-specific to species of *Corymbia*. However, debate over its taxonomic position has led to a number of name changes, and similar fungi have been found to affect other eucalypt species (Potts and Pederick 2000; Ivory *et al.* 2004). The disease affects new foliage causing pure white, shiny or powdery fungal lesions leading to distortion of leaves and curling and twisting of stems. The first symptoms usually appear approximately five days after infection, with pin-head-sized white pustules appearing on both sides of the leaves, particularly along leaf margins and leaf veins, and on young stems. The pustules become necrotic and the epidermis of the diseased shoots is ruptured with a waxy cuticle forming, giving the leaf surface a shiny white appearance (Potts and Pederick 2000; Ivory *et al.* 2004). Severe infections can result in leader damage and the proliferation of apical and lateral shoots. Severe and/or repeated infections can ultimately lead to restricted, stunted growth and even death (Ivory *et al.* 2004).

The RSB fungus has only been found in an asexual form with no known sexual phase. On leaf surfaces at least 50% of spores have been found to germinate within five hours, with large sporulating lesions evolving by day ten. The disease cycle can, therefore, be completed within ten days under favourable conditions. When grown in the laboratory, growth has been optimal between 20-25°C. In Queensland the disease appears to be active throughout the year. Outbreaks occur after short periods of rain, followed by dry conditions during which lesions develop. When these conditions coincide with flushes of new growth on the plants, most trees will be infected (Ivory *et al.* 2004).

RSB susceptibility is an ideal trait for quantitative genetic analysis as it potentially affects fitness and therefore may be under natural selection pressure. RSB was observed in provenance seedling seed orchard field trials of *C. variegata* planted in 1997 by the Department of Primary Industries and Fisheries’ (DPI&F) Agency for Food and Fibre Sciences Forestry Research (AFFS Forestry Research). Differences in susceptibility varied significantly between and within provenance regions (AFFS Forestry Research internal reports; Ivory *et al.* 2004). As *C. variegata* is an important timber species, any loss of form or growth potential is disadvantageous.
Figure 1.3: *C. variegata* leaves infected by the fungal disease *Quambalaria pitereka*. Top photo shows white pustules forming along the mid-vein of a leaf and the twisting and curling of an infected leaf stem (photo credit: Geoff Pegg). Bottom photo shows infection on most of the leaf surface (photo credit: David Lee).
1.3 Thesis Objectives

This thesis examines the structure of genetic variation within a eucalypt species complex at two spatial scales, using both molecular and quantitative genetic markers. It explores the relative influences of the key opposing mechanisms of genetic variation, genetic drift, gene flow and selection, in relation to spatial structure.

The key objectives of this thesis are to:

- To examine the broad-scale geographic variation and structure within and between the four putative species of the spotted gum complex, using a chloroplast DNA marker and comparing the results to existing allozyme data. In particular, to test the hypothesis that *C. henryi* and *C. maculata* are more closely related to each other than either species is to *C. citriodora* and *C. variegata*. (Chapter 2).

- To test the hypothesis that potentially limited seed dispersal may lead to nearest neighbour bi-parental inbreeding within a single population of *C. variegata*. The fine-scale spatial structure of genetic variation and relatedness between individuals within a single natural population of *C. variegata* will be assessed using nuclear microsatellite markers (Chapter 3).

- To compare the relative influences of dispersal and selection in determining fine-scale patterns in variation and spatial structure. The heritability, variation and spatial structure of a quantitative trait potentially under selection pressure will be compared to the spatial structure of molecular variation identified in Chapter 3 (Chapter 4).

- The results of these analyses are drawn together in Chapter 5, which discusses the implications of the study for our understanding of the spotted gum complex, for commercial development and conservation of the species, and more broadly, for our understanding of spatial genetic structure within forest tree species.
2.1 Introduction

The use of different molecular markers can often reveal different patterns of genetic variation between populations and species due, in large part, to their mode of inheritance. Nuclear DNA (nDNA) is always bi-parentally inherited and subject to intergenerational recombination (Judd et al. 1999). However, chloroplast DNA (cpDNA) is nearly always uniparentally inherited and therefore, with no recombination, it more easily allows genealogical relationships to be identified (Judd et al. 1999: Cornu and Dulieu 1998; McKinnon et al. 2001b). The uniparental inheritance of the chloroplast genome also means that it has a smaller effective population size than nDNA. CpDNA has also been found to have a slower mutation rate than nDNA. Consequently, the uniparental inheritance, smaller effective population size and slower mutation rate of cpDNA means that the influence of genetic drift on this marker is increased (Judd et al. 1999; Schaal et al. 1998). Therefore, the chloroplast genome provides a highly conserved marker useful for clarifying phylogenetic relationships, and is often the marker of choice in plant phylogenetic and phylogeographic studies (Avise 1994; Schaal et al. 1998; Judd et al. 1999; Murray et al. 2000).

In eucalypts, both high levels of intraspecific polymorphism and interspecific sharing of cpDNA haplotypes have been found (Steane et al. 1998; Jackson et al. 1999; McKinnon et al. 1999; McKinnon et al. 2001a). The chloroplast genome is maternally inherited in eucalypts (as in most angiosperms) and may therefore be useful in comparing the extent of seed and pollen mediated gene flow when compared to bi-parentally inherited markers (Judd et al. 1999; Murray et al. 2000). In eucalypt species with limited seed dispersal, cpDNA variation has also been more closely associated with geographic distribution than with species boundaries based on morphological divergence (Steane et al. 1998; Jackson et al. 1999; McKinnon et al. 1999; McKinnon et al. 2001a; McKinnon et al. 2001b).

In contrast, the intergenerational recombination of the nuclear genome increases the level of variation observed and limits its usefulness when attempting to assess historical patterns of gene flow and lineage relationships (Judd et al. 1999; Murray et al. 2000).
The propensity for hybridisation between closely related species of plants, particularly eucalypts, confounds this problem (Avise 1994; Judd et al. 1999; Murray et al. 2000). Due to ongoing debate over the taxonomic standing of the spotted gums and the generally low levels of nDNA divergence found between morphospecies (McDonald et al. 2000), the species complex lends itself well to a comparative study using cpDNA sequence data and phylogeographic analysis based on coalescent theory and the cohesion species concept. Analysis of the distribution of cpDNA haplotypes throughout the geographic range of the complex and the genealogical relationships between haplotypes may clarify the lineage status of the current morphological species designations.

2.2 Objectives
The objectives of this study were to extend the work of McDonald et al. (2000), explore further the status of the species within the group, and the relationships among populations of the four species. Specific aims were to test the hypothesis that C. henryi and C. maculata are more closely related to each other than to either C. citriodora or C. variegata by: 1) determining any correspondence, or lack of, between the nuclear and chloroplast markers; 2) inferring the historical microevolutionary processes that may have led to the contemporary inter- and intra-specific population structure; 3) assessing the degree of correlation between the geographic distributions and cpDNA genetic variation of populations.

2.3 Methods
2.3.1 Sampling Strategy
Seed from multiple, open-pollinated families from each of 23 populations across the species’ range-wide distributions (Figure 2.1) were obtained from CSIRO Forestry and Forest Products. The seed had been collected during the study of the complex by McDonald et al. (2000) and its use in this study was employed to allow a direct comparison between the nuclear based results of that study and the chloroplast analysis of this study. The population identity numbering used by McDonald et al. (2000) was retained for this study. Populations were numbered sequentially (1-28) from north (1) to south (28). Populations 2, 11, 13, 20 and 24 were not used in the chloroplast analysis, as
the seed did not germinate. No bulked seedlots were used, therefore the identities of individual mother trees was maintained. Each seedlot was germinated in an individual germination tray using a commercial peat:sand mixture (80:20).

Figure 2.1: Distribution map of the sampled populations of spotted gums: *C. citriodora* (populations 1-8), *C. variegata* (populations 9-16), *C. henryi* (populations 17-20) and *C. maculata* (populations 21-28), and the outgroup species *C. torelliana* (population 29). Location coordinates from McDonald et al. (2000).
Leaf samples were collected for chloroplast analysis from seedlings between 12–16 weeks of age. As the key aim of this study was to explore the relationships between species, the priority was to sample as many populations as possible rather than a large number of individuals from a few locations. In total, 68 families (each representing an individual maternal parent) were sampled across the four species; three individuals from each of seven populations for both *C. citriodora* and *C. maculata* (42 in total), three individuals from each of six populations for *C. variegata* (18 in total), and three individuals from each of two populations and two individuals from one population of *C. henryi* (8 in total). One individual (population 29) from the section Cadagaria, *Corymbia torelliana* was also sampled from Kuranda, which is located between populations 1 and 2 (Figure 2.1). *C. torelliana* is considered the closest sister species to the spotted gum complex and was included as an outgroup. Leaf tissue was stored in a freezer at -70°C until used for DNA extraction.

2.3.2 Molecular Methods

Total genomic DNA was extracted using the QIAGEN DNeasy Plant Mini-kit (QIAGEN Industries) after first grinding the leaf tissue (100 mg) using a mortar and pestle in the presence of liquid nitrogen and Poly-vinyl pyrolidone (PVP) powder (5.0 mg). PVP powder was used to assist in the removal of DNA-degrading phenolic compounds (Loomis *et al.* 1979).

The hypervariable J_{LA}+ region of the chloroplast genome was amplified using the published primers *rpl2* (GATAATTTGATTCTTCGCC; (Goulding *et al.* 1996)) and *eucpsbA* (GGAGCAATAACCAACTCTTG; (Freeman *et al.* 2001)). This region surrounds a junction between one of the inverted repeats and the large single copy region in the chloroplast genome. It is comprised of the *rpl2-trnH* intergenic spacer, the *trnH* gene and the *trnH-psbA* intergenic spacer (Freeman *et al.* 2001). Polymerase chain reaction (PCR) amplification conditions were the same as those described by Vaillancourt and Jackson (2000) for the J_{LA} region. The resultant PCR product was purified using QIAquick columns (QIAGEN Industries) following the manufacturers product protocol. Purified PCR product was sequenced in both directions using the *rpl2* and *eucpsbA* primers with PE terminator dye chemistry and sequenced on an ABI-377 automated DNA sequencer. Sequence alignments were performed manually using Bioedit (Hall 1999).
2.3.3 Data Analysis

Sequence alignment

Due to large poly-A region(s) (approximately 20 base pairs (bp) long) starting 124bp downstream of the rpl2 primer, much of the subsequent sequence was unresolved due to slippage during the PCR process. This occurred in the sequences from both directions and therefore no overlap was available during alignment to ensure strand homology. Therefore, to clarify any unresolved base positions in the first 124 bp from the rpl2 primer and the first 317 bp from the eucpsbA primer, a consensus sequence was obtained for each sample in each direction by performing the PCR amplification and sequencing process a number of times on each individual. The forward and reverse consensus sequences were then joined to provide a single aligned sequence for each individual of 441 bp.

cpDNA analysis

To test for genetic structure within and between the four morphological species, analysis of molecular variance (AMOVA) (Excoffier et al. 1992) was performed in the program Arlequin (Schneider et al. 2000) using the Felsenstein 1981 (F81) nucleotide substitution model (gamma=0.0058). This substitution model was determined to be the most suitable for the dataset using the program ModelTest (Posada and Crandall 1998) which compares 64 models with different combinations of parameters for DNA site substitution (Posada and Crandall 2001). The F81 model of evolution is a simple model that assumes that all base substitutions are equally likely, with variable base frequencies (gamma=site to site rate variation) (Felsenstein 1981). An overall pattern of isolation by distance due to limited seed dispersal was assessed using a Mantel test which identifies significant correlation between population pairwise genetic and geographic distances. The program PATN (Belbin 1995) was used to perform the analysis and 1000 permutations were conducted.

To assess the cohesion between populations without the restriction of taxonomic morphospecies preconceptions, a haplotype network was generated using the program TCS version 1.13 (Clement et al. 2000) which estimates gene genealogies from DNA sequences. The network method is a coalescent based approach and is a more appropriate method of phylogenetic reconstruction at the population level than traditional phylogenetic methods (such as trees which are used for studies of higher
taxonomic levels) (Posada & Crandall 2001). Within populations, ancestral haplotypes, for example, can give rise to a number of derived haplotypes and therefore produce a tree with multifurcations, violating the assumption of a strictly bifurcating tree topology. It is also unlikely that the rise of a new haplotype would lead to the extinction of the ancestral type. In fact, coalescent theory predicts that ancestral types will not only persist but also be the most frequent and geographically widespread. However as the oldest (ancestral) haplotypes they would be interior (extinct) nodes of a haplotype tree (Posada & Crandall 2001). Therefore, temporal relationships between haplotypes are often more clearly represented within a network.

The program TCS uses the statistical parsimony method described by Templeton et al. (1992) to estimate gene genealogies from DNA sequences. It estimates the cladogram with the 95% most parsimonious branch connections and displays the number of missing haplotypes along branches between extant haplotypes (Clement et al. 2000). In the present study, sequence gaps were treated as a fifth character state so that indel information was included in the analysis.

Nested clade analysis (NCA) (Crandall 1996; Templeton 1998a; Templeton 1998b) was used to assess further the geographic structure of haplotype distribution. Using the number of mutational differences between haplotypes and the geographic distances between populations, NCA tests whether individuals from the same or neighbouring populations are closer to each other in the haplotype network than would be expected by chance (Hey & Machado 2003). The hierarchical nature of the nesting procedure allows for differentiation between the historical and contemporary evolutionary processes affecting the distribution of haplotypes. The hierarchical nesting of the haplotypes (clades) was determined using the nesting algorithm of Crandall (1996). Degenerate clades (those at any nesting level containing less than two known haplotypes) were uninformative and excluded from analysis. Straight-line geographic distances between populations were calculated from the latitude/longitude coordinates given in McDonald et al. (2000).

The program GeoDis (Posada et al. 2000) was used to first calculate a simple test of geographic association, where geographic locations are treated as categorical variables (i.e. not considering distances between locations) using chi-squared exact tests for each
clade at each nesting level (Posada et al. 2000). GeoDis was also used to calculate the clade distances (Dc) (the geographic range of a given clade) and the nested-clade distances (Dn) (the average geographic distance of a clade from the geographic centre of the clade within which it is nested). Average distances between interior and tip clades within a nested group (I-Tc) and between a clade and the higher clade within which it is nested (I-Tn) were also calculated. The significance of these statistics at the 5% level of significance was determined using a Monte Carlo permutation procedure (Posada et al. 2000).

The permutation method used in GeoDis for NCA permutes individuals or clades across the geographical area considered in the clade within which they are nested, while haplotype and population numbers are held constant. Petit and Grivet (2002) argued that, particularly where few populations (<20) have been sampled, this permutation procedure might lead to insufficiently conservative spatial inferences. However, the most problematic scenario identified by Petit and Grivet (2002) was one with high levels of fixation, a low number of populations and a high number of individuals sampled per population, none of which apply in this case.

Interpretation of results followed an updated, downloaded version of the inference key (Posada et al. 2003). The basis of this inference key is that historical processes such as restricted gene flow, range expansion and allopatric fragmentation are expected to result in different patterns of geographic association (Hey & Machado 2003).

To test for evidence of secondary contact, the average pairwise distances between the geographical centres of haplotypes and clades was plotted in sequential order of population identity (from northern to southern populations) (Templeton 2001; Byrne et al. 2002). This method identifies how different the haplotypes or clades are within each population. Under an isolation by distance model, haplotypes found within a single location tend to be more closely related evolutionarily. The Dc values (the geographic range) for each haplotype and then each clade are used to calculate pairwise distances at each clade-step level. As similar haplotypes clade together at lower nesting levels (i.e. 1-step or 2-step clades), the pairwise distances between Dc values will fall until all the haplotypes within a population are nested in the same clade, at which point the pairwise location distance will equal zero. If secondary contact has occurred the haplotypes...
found in that population will not clade together until one of the higher nesting levels, and therefore the pairwise differences in the Dc values for that population will remain high. Haplotypes not included in the analysis of 1-step clades (i.e. they initially nest with a missing haplotypes) have the same Dc values as the 2-step clade that they nest within. This is because missing haplotypes add nothing to the Dc values. If a haplotype occurs in only one population the Dc value for that haplotype will be zero (Templeton 2001).

Reassessment of allozyme data

Initial analysis of the cpDNA generated from this study suggested different genealogical relationships between the four species from those proposed by McDonald et al. (2000) based on allozyme data. The possibility that one or more of the allozyme loci had not reflected true gene flow patterns between the species was then considered. To explore the relative discriminating influence of each allozyme locus in determining the taxonomic relationships supported by McDonald et al. (2000) the GD program of Ritland (1990) was used to produce 19 UPGMA trees based on Nei’s (1978) unbiased genetic distance. One tree was based on allele frequencies of all loci assayed by McDonald et al. (2003), while for each of the other 18 trees only 17 of the 18 loci were used (i.e. each locus was alternatively removed from the dataset). Multidimensional scaling (MDS) ordinations based on the pairwise population genetic distance matrices calculated in GD were then produced in PATN (Belbin 1995) (and plotted in SAS) to further explore shifts in the topologies of the initial tree (including all loci) and the tree with individual loci omitted. Plots of allele frequencies for discriminating loci were also produced.

2.4 Results

2.4.1 Sequence Polymorphism

The total aligned sequence length for the cpDNA J_{LA+} region was 441bp, with 23 variable characters (5.2%) defining 19 haplotypes (Table 2.1). Many previous studies have reported significant variation due to indels rather than single base mutations in this region (Vaillancourt & Jackson 2000; Freeman et al. 2001). However, with the necessary exclusion of the large (and likely variable) poly-A section(s) (described previously) from the sequence data, 21 of the 23 variable characters scored across the
four species in this section were simple base pair substitutions. Of the three indels scored, two involved single base positions and one a 6 bp duplication. In all but two cases, single base differences between haplotypes were observed in more than one individual. In total, 19 haplotypes were identified (Appendix 1). The sequence for the outgroup species, *C. torelliana*, showed an additional nine base substitutions and one additional 2bp insertion (Table 2.1) not found in any of the ingroup species.

AMOVA identified significant genetic differentiation among the four morphological species groups (*p* < 0.001), with 24% of the molecular variation attributed to among species variation and 76% to within species variation. A comparison of species group pairwise $\Phi_{ST}$’s and the pairwise percent sequence differences (Table 2.2) showed that statistically significant (*p* <0.05) differentiation occurs between *C. citriodora* and both *C. variegata* and *C. maculata*. Contrary to the species relationships proposed by McDonald *et al.* (2000), *C. henryi* was not found to be significantly different from either *C. citriodora* or *C. variegata*, but was significantly different from *C. maculata*. *C. citriodora* populations had by far the highest within group percent sequence variation (0.13), almost eight times that found in *C. maculata* (0.017) (Table 2.2).
Table 2.1: List of haplotypes found across the four species in *Corymbia* section Politaria, aligned against the study outgroup, and close relative of the section, *Corymbia torelliana* (section Cadagaria). ~ indicates the loss or insertion of base pair(s). The unsequenced poly-A region began at position 120 of the aligned sequence using *rpl2* as the forward primer.

| base position | 22 26 32 | 37 48 62 72 130 141 151 168 175 176 202 210 218 238 239 242 248 264 274 283 284 289 301 342 359 369 371 380 437 |
|---------------|----------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|
Table 2.2: Species group pairwise differences based on cpDNA.
Above Diagonal: % average sequence difference between species (* p<0.05).
Diagonal Elements: % average sequence difference within species.
Below Diagonal: pairwise $\Phi_{ST}$ s (* p<0.05).

<table>
<thead>
<tr>
<th></th>
<th>C. citriodora</th>
<th>C. variegata</th>
<th>C. henryi</th>
<th>C. maculata</th>
</tr>
</thead>
<tbody>
<tr>
<td>C. citriodora</td>
<td>0.132</td>
<td>0.032*</td>
<td>0.023</td>
<td>0.040*</td>
</tr>
<tr>
<td>C. variegata</td>
<td>0.250*</td>
<td>0.053</td>
<td>0.000</td>
<td>0.003</td>
</tr>
<tr>
<td>C. henryi</td>
<td>0.150</td>
<td>0.000</td>
<td>0.044</td>
<td>0.003</td>
</tr>
<tr>
<td>C. maculata</td>
<td>0.350*</td>
<td>0.080</td>
<td>0.160*</td>
<td>0.017</td>
</tr>
</tbody>
</table>

The Mantel test identified an overall significant effect of isolation by distance (p<0.001). However, plotting of all population pairwise comparisons (Figure 2.2) showed that the correlation between genetic and geographic distance was driven mostly by the greater genetic divergence between geographically distant populations (i.e. those greater than 1500km apart). Pairwise comparisons between geographically close populations (up to 1500 km apart) reflected both large and small genetic differences. Under an isolation by distance model, it is expected large genetic distance would not be found among geographically close populations, where individuals should be more closely related than random.

![Figure 2.2: Correlation between population pairwise genetic distance ($F_{ST}$) and geographic distance.](image-url)
2.4.2 Geographic Distribution of Haplotypes

The number of individuals of each haplotype within each population is shown in Table 2.3. The haplotype network (Figure 2.3) shows no obvious pattern of haplotype distribution across the populations: northern and southern populations share haplotypes (e.g. H-5), geographically distant populations share genetically similar haplotypes (e.g. H-13 and H-17) and genetically divergent haplotypes are found in geographically close populations (e.g. H-17 and H-18). This lack of geographic structure, specifically in the three northern species, is shown in Figure 2.4 where haplotypes have been mapped according to the populations in which they occur. The network also illustrates that the largest number of haplotypes was found in populations of *C. citriodora* (populations 1-8), encompassing up to 12 mutational changes, and that sharing of haplotypes (H-5, H-11 and H-13) occurred between *C. citriodora*, *C. variegata* (populations 9-16) and *C. henryi* (populations 17-19). The tip haplotype H-13 occurred in five populations (10, 14, 15, 18, 19): three *C. variegata* and two *C. henryi* populations. Only two haplotypes were found in *C. maculata* (H-17 and H–18). Both haplotypes were unique and each was associated with a different region of *C. maculata*’s distribution (Table 2.3)
Table 2.3: The number of individuals of each haplotype within each population and morphological species group. Populations numbered 1-28 sequentially from the northern most population of *C. citriodora* to the southern most population of *C. maculata*. A maximum of three individuals per population were sequenced (no individuals from populations 2, 11, 13, 20 and 24 were sequenced).

<table>
<thead>
<tr>
<th>Population</th>
<th><em>C. citriodora</em></th>
<th><em>C. variegata</em></th>
<th><em>C. henryi</em></th>
<th><em>C. maculata</em></th>
<th>Southern</th>
<th>Northern</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>3 3 3 3 3 3 3 3</td>
<td>3 3 3 3 3 3 3 3</td>
<td>3 3 3 3 3 3 3 3</td>
<td>3 3 3 3 3 3 3 3</td>
<td>3 3 3 3</td>
<td>3 3 3 3</td>
</tr>
</tbody>
</table>
Figure 2.3: Hypothesised haplotype network. Haplotypes (H1-19) are represented by large coloured circles, each colour signifying a different haplotype. Each connection between haplotypes is representative of a single base change (or insertion/deletion) in the chloroplast sequence (see Table 2.1). Each small open circle along clade connections represents a putative ancestral haplotype that was not sampled.
Figure 2.4: 19 haplotypes of the spotted gum complex mapped according to the populations in which they occur. Each haplotype is represented by a different colour and shows the lack of overall clear geographic structure in haplotype distribution. C = *C. citriodora* populations, V= *C. variegata* populations, H= *C. henryi* populations, M= *C. maculata* populations.
One ambiguous network loop occurred between H-3, H-4, H-5 and H-11 and was resolved by removing a 2-step connection between H-3 and H-11. The removal of the alternate connection between H-5 and H-4 would have resulted in H-5 becoming a tip clade. H-5 is more geographically widespread, more frequent and has a higher outgroup probability than H-3 (0.088 and 0.039 respectively). Under the predictions of coalescent theory, this haplotype is more likely to be an internal haplotype. To meet the 95% probability of parsimony criterion required for the network, thirteen missing haplotypes were generated by TCS. The most likely ancestral haplotype was H-2 with an outgroup probability of 0.204 and was the haplotype with the least number of base differences (11) from the outgroup species (Table 2.1). The sequence from an individual from the closest sister taxon, Corymbia torelliana, was included in the network construction but was rejected by TCS as too genealogically divergent from the Politaria group to establish a branch connection which would meet the 95% plausibility criterion.

2.4.3 Nested Clade Analysis

The nineteen identified haplotypes were partitioned into fifteen 1-step clades, seven 2-step clades and two 3-step clades (Figure 2.5). Due to the large number of missing haplotypes in the network, twelve of the 1-step clades and one of the 2-step clades were degenerate and could not be included in the analysis. Two additional 2-step clades were also degenerate due to no geographic distance between haplotypes: i.e., the haplotypes were endemic to the same population. This may limit the effectiveness of NCA as a method of inferring patterns of geographic association between the most genetically similar haplotypes (i.e. most of the 1-step clades), however, associations at the higher nesting levels were still considered potentially informative.

The chi-squared contingency table analysis identified significant geographic structure of haplotypes (p<0.05) at seven of the ten nested clade levels, including the total cladogram (Table 2.4). Inclusion of geographic distance into the analysis and use of the inference key suggested that patterns of geographic variation at the 1-step clade level were due to the combined effects of past range expansion (RE) and fragmentation (F) events, with the possible influence of long distance movements (LDC). Restricted gene flow/isolation by distance (RGF/IBD) was inferred at all nested clade levels. The total cladogram suggests that the predominant historical evolutionary process influencing the observed geographical association between all haplotypes has been RGF/IBD (Table
2.4). Plotting of the pairwise location distances (Figure 2.6) showed distances tending towards zero at nearly all locations: i.e. for nearly all populations the average pairwise location distance shown in Figure 2.6A (haplotypes) has reached zero in Figure 2.6D (3-step clade level). At the 1-step clade level (the level at which possible fragmentation was detected), large pairwise location distances were found at many of the northern and some of the centrally located populations. This indicates that at least one of the haplotypes found in each of these populations also occurs in other geographically distant populations. At the 3-step level, a non-zero distance was found at one location, population 10.
Figure 2.5: Hypothesised haplotype network of the spotted gum complex by morphospecies. Haplotypes (H1-19) are represented by large coloured circles, each colour signifying a different morphologic species (see key). Each connection between haplotypes is representative of a single base change in the chloroplast sequence. Each small open circle along clade connections represents a putative ancestral haplotype that was not sampled.
Table 2.4: Results of nested clade geographic distance analysis of the JLA+ haplotypes of the spotted gum complex. The nested design is shown in Figure 2.5. The populations represented by each haplotype are shown in Table 2.3. Significant $\chi^2$ values indicated by *. The Dc and Dn values are the greater circle distances. Significantly small (S) and significantly large (L) values are shaded (p<0.05). For the total cladogram, clade 3-1 was identified as the interior clade as it included the haplotypes (zero-step clades) with the largest out-group probabilities (haplotypes 2 and 17). The inference key couplet sequence and the biological inference for each significant nested clade are listed (PF=past fragmentation, RE= range expansion, LDC=long distance dispersal; RGF=restricted gene flow; IBD=isolation by distance).

<table>
<thead>
<tr>
<th>Nesting</th>
<th>Clade Location</th>
<th>$\chi^2$ Statistic</th>
<th>Haplotype or Clade Number</th>
<th>Dc</th>
<th>Dn</th>
<th>Inference Key</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>One-step Clades</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-4</td>
<td>Interior</td>
<td>17.00 *</td>
<td>2</td>
<td>360 S</td>
<td>920 L</td>
<td>1, 19, 20, 2, 11, 12, 13, 21, No Tip</td>
<td>15</td>
</tr>
<tr>
<td>Tip</td>
<td>19</td>
<td>0</td>
<td>645</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-T</td>
<td>360</td>
<td>227</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-5</td>
<td>Tip</td>
<td>22.00 *</td>
<td>13</td>
<td>215 S</td>
<td>613 L</td>
<td>1, 19, 20, 2, 11, 12, 13, 21, No Interior</td>
<td>17</td>
</tr>
<tr>
<td>1-T</td>
<td>-72</td>
<td>-97 S</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-15</td>
<td>Tip</td>
<td>7.77</td>
<td>3</td>
<td>0</td>
<td>218</td>
<td>1, 2, 3, 4, No Interior</td>
<td>4</td>
</tr>
<tr>
<td>Tip</td>
<td>9</td>
<td>0</td>
<td>272</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-T</td>
<td>237 L</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Two-step Clades</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-2</td>
<td>Tip</td>
<td>49.11 *</td>
<td>1-1</td>
<td>0 S</td>
<td>905</td>
<td>1, 2, 11, 17, 4, No Interior</td>
<td>1-4</td>
</tr>
<tr>
<td>Tip</td>
<td>1-5</td>
<td>560 S</td>
<td>695 S</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-T</td>
<td>280</td>
<td>157 L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-4</td>
<td>Interior</td>
<td>5.00</td>
<td>1-10</td>
<td>0</td>
<td>985</td>
<td>No significant values Interior</td>
<td>1-13</td>
</tr>
<tr>
<td>1-T</td>
<td>no interior/tip clades exist in this group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-5</td>
<td>Tip</td>
<td>4.00</td>
<td>1-8</td>
<td>0</td>
<td>222</td>
<td>No significant values Interior</td>
<td>1-9</td>
</tr>
<tr>
<td>1-T</td>
<td>0</td>
<td>222</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-7</td>
<td>Interior</td>
<td>11.00 *</td>
<td>1-14</td>
<td>893 L</td>
<td>694 L</td>
<td>1, 2, 3, 4, No Tip</td>
<td>1-15</td>
</tr>
<tr>
<td>1-T</td>
<td>616 L</td>
<td>262 L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Three-step Clades</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-1</td>
<td>Tip</td>
<td>90.00 *</td>
<td>2-1</td>
<td>0 S</td>
<td>1745 L</td>
<td>1, 2, 3, 5, 6, 7, Yes Interior</td>
<td>2-2</td>
</tr>
<tr>
<td>Tip</td>
<td>2-3</td>
<td>157 S</td>
<td>604 S</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-T</td>
<td>640 L</td>
<td>-100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-2</td>
<td>Interior</td>
<td>41.40 *</td>
<td>2-4</td>
<td>603</td>
<td>480</td>
<td>No significant values Tip</td>
<td>2-5</td>
</tr>
<tr>
<td>Tip</td>
<td>2-6</td>
<td>0</td>
<td>340</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tip</td>
<td>2-7</td>
<td>544</td>
<td>497</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-T</td>
<td>175</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total Cladogram</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interior</td>
<td>65.02 *</td>
<td>3-1</td>
<td>817</td>
<td>845 L</td>
<td>1, 2, 3, 4, No Tip</td>
<td>3-2</td>
<td>475 S</td>
</tr>
<tr>
<td>1-T</td>
<td>341 L</td>
<td>82</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 2.6: Average pairwise location distances between the geographical centres of A – haplotypes, B- 1-step clades, C- 2-step clades and D- 3-step clades of the spotted gum complex. Populations ordered from northern to southern most.
2.4.4 Reassessment of Allozyme Data

The UPGMA tree of all allozyme loci (k=18) shows the topology and species associations published by McDonald et al. (2000) (Figure 2.7A). In this analysis there are two main groups comprised of the *C. citriodora* and *C. variegata* populations and the *C. maculata* and *C. henryi* populations respectively. The southern *C. maculata* populations also showed greater affinity to each other than to the northern *C. maculata* populations. This topology remained consistent in all other trees based on k-1 loci except in the case where the *PGM2* locus was removed. The removal of *PGM2* resulted in a substantial shift in the clustering of the *C. variegata* populations 14, 15 and 16 from one main clade to the other, indicating a closer association between these populations and those of *C. henryi* and *C. maculata* (Figure 2.7B). Populations 14, 15 and 16 occur in the same geographic area as the populations of *C. henryi* with which they cluster in this case.

The MDS ordination of all loci (k=18) showed a similar clustering of populations to that found using UPGMA (Figure 2.8A). The two main groupings reflected the same relationships identified in the two main UPGMA clades (Figure 2.7A). The removal of *PGM2* positioned the *C. henryi* and the southern *C. variegata* populations more closely together in ordination space. However the shift occurred in the populations of both species (Figure 2.8B), not simply those of *C. variegata* as suggested by the UPGMA tree (Figure 2.7B). The position of the *C. maculata* populations remained relatively unchanged.

Four alleles were scored at the *PGM2* locus, although two occurred at only low frequencies in random populations. For the two informative alleles (A and B), there is an abrupt switch in the dominant allele between population 16 (*C. variegata*) and population 17 (*C. henryi*) (Fig 2.9). Populations 1-16 are all *C. citriodora* and *C. variegata* populations and are fixed or have very high frequencies of allele A, whereas populations 17-28 are all *C. henryi* and *C. maculata* populations and are fixed or have very high frequencies of allele B. An abrupt shift at one locus, with no such shift at any of the other 17 loci, suggests the effects of selection acting at the *PGM2* locus.
Figure 2.7: Cluster analysis based on Nei’s (1978) unbiased genetic distance and the UPGMA algorithm for *C. citriodora*, *C. variegata*, *C. henryi* and *C. maculata* populations. Tree A is based on 18 allozyme loci and tree B excludes the *PGM2* locus. Green and blue coloured branches highlight the change in clustering between the two data sets. Removal of any other locus apart from *PGM2* produced trees with similar topology to that shown in A (results not shown).
Figure 2.8: Multi-dimensional Scaling (MDS) 2-dimensional ordination plots for populations 1-28 using Nei’s (1978) unbiased genetic distance. Species identifiers c, v, h and m (representing *C. citriodora*, *C. variegata*, *C. henryi* and *C. maculata* respectively) are used along side population numbers. Plot A is based on 18 allozyme loci and plot B excludes the *PGM2* locus. Green and blue coloured circles highlight the change in relative positions of *C. variegata* and *C. henryi* populations. Removal of any other locus apart from *PGM2* gave a similar display of populations as plot A (results not shown).
Figure 2.9: Allele frequencies at $PGM2$ locus for populations 1-28. An abrupt shift in dominance of alleles A and B occurs between populations 16 and 17.
2.5 Discussion

2.5.1 Species Relationships

In the present study, *C. henryi* was found to be significantly different to *C. maculata*, but not to *C. variegata* and *C. citriodora* (AMOVA). The distribution of haplotypes among species in the network showed shared haplotypes between all species except *C. maculata*. This suggests a closer affinity between the three northern species and a divergence of *C. maculata* from this group. These results contrast with those of the previous study of this species complex based on allozyme data (McDonald et al. 2000), which found *C. henryi* to be more closely aligned with *C. maculata*, than with *C. variegata* and (possibly) *C. citriodora* as suggested by the cpDNA data.

Data from nuclear and chloroplast markers often indicate different potential genealogies, due to the larger effective population size of nuclear genes. Larger effective population sizes result in longer coalescent times and increases the likelihood of shared ancestral polymorphisms (Schaal and Olsen 2000). However, reanalysis of the allozyme data with the exclusion of the *PGM2* locus, supported the assertion that *C. henryi* is more closely associated with *C. variegata* and *C. citriodora* than with *C. maculata*. This suggested the influence of selection acting at this locus and that differences in effective population size may not, therefore be the sole cause of the initial difference in results between these markers.

The genetic alliance of *C. henryi* with *C. maculata* reported by McDonald et al. (2000) was supported to some extent by the morphological and habitat similarities between them. However, the complex as a whole is differentiated by only subtle morphological differences (Hill and Johnson 1995; Brooker and Kleinig 1999; Larmour et al. 2000). McDonald et al. (2000) also found that although *C. henryi* has the most restricted distribution, and therefore potentially the smallest effective population size, it showed the greatest genetic diversity. Alternatively, given that *C. henryi* is broadly sympatric with *C. variegata* and allopatric with both *C. maculata* and *C. citriodora*, then hybridisation between *C. henryi* and *C. variegata* (as suggested by the cpDNA results), would also explain the high level of nuclear diversity maintained within *C. henryi*.

A comparative study of the chemical composition in seedling leaf oils of the four putative spotted gum species by Asante et al. (2001) reported species affinities more in
line with the results from the cpDNA analysis than those of McDonald et al. (2000). Based on chemical composition of leaf oils, Asante et al. (2001) reported a clear separation of *C. maculata* from the other three species in this group, as well as affinities between neighbouring populations of *C. variegata* and *C. henryi*. *C. citriodora* was also distinct in this study due to the presence of citronellal and citronellol in all samples. However, the abundance of these oils in southern populations of *C. citriodora* (i.e. those close to *C. variegata*) was often low (Asante et al. 2001).

The plots of *PGM2* allele frequencies by population highlighted a distinct shift in the most common allele found in neighbouring populations of *C. variegata* and *C. henryi*, although they shared cpDNA haplotypes. For example, populations 14 (*C. variegata*) and 19 (*C. henryi*) are neighbouring populations and have high frequencies for different *PGM2* alleles while sharing H-13. As H-13 is a tip haplotype, sharing of this haplotype suggests either contemporary gene flow or homoplasy among the populations of the two species. However, it is unlikely that the sharing of this haplotype between species is the result of homoplasy because the most closely related haplotype in either species is at least three characters different from H-13. This haplotype is also shared with *C. citriodora* further supporting the earlier assertion in this study of a closer contemporary affinity of *C. henryi* to *C. variegata* and *C. citriodora* than to *C. maculata*.

The removal of *PGM2* and the subsequent reanalysis of the data suggested a broad isolation by distance effect occurring between the *C. citriodora*, *C. variegata* and *C. henryi* populations. The distinct clustering of the *C. maculata* populations into two subgroups was supported by the cpDNA analysis in which one of two unique haplotypes (H-17 and H-18) was found in each of the two sub-groups. The shift of population positions in the UPGMA tree and MDS plot, and the change in most common alleles suggests that *PGM2* is not selectively neutral. The homogenising effect of gene flow evident at all other loci was not strong enough to mask the effect of selection on *PGM2*. Therefore, the inclusion of *PGM2* in analysis may have misrepresented the true relationship between *C. henryi* and *C. maculata*.

Endler (1977) showed that a sharp differentiation in allele frequencies separating zones of otherwise relatively constant gene frequencies, or step clines, (as seen in *PGM2*) are unlikely to occur simply due to genetic drift, especially where gene flow is known to be
occurring (as suggested by all other allozyme loci in these species). The development of step clines due to primary differentiation requires selection acting along an environmental gradient, although step clines are not necessarily associated with spatially abrupt environmental changes (Endler 1977). Under coalescent theory and the cohesion species concept, a cline derived from primary differentiation should represent a single evolutionary lineage and be identified as a single species. Where populations have been historically isolated and divergence has occurred in allopatry, secondary contact between population groups can also result in stepped clines (Endler 1977). Under this model, population groups would represent different evolutionary lineages (Endler 1977). However, there was no evidence of the latter interpretation from the cpDNA analysis, and therefore secondary contact is unlikely to be the cause of the cline observed in this study.

Little detailed information is available on the environmental preferences of the species, particularly where they occur sympatrically. Therefore, the nature of the environmental gradient driving the selection pattern in these species is not known and was beyond the scope of this study. However, PGM is known to be associated with metabolism, metal–binding domains, and indirectly in the formation of cell wall polysaccharide components in plants (Sterky et al. 1998). Selection has also been demonstrated to affect other allozyme loci involved in the metabolism of animals and plants (Watt 1977; Hilbish and Koehn 1985; Carter and Watt 1988; Travers and Mazer 2001).

AMOVA also indicated significant genetic differences between C. citriodora and C. variegata. However, these two taxa, along with C. henryi, share haplotypes (H-4, H-5 and H-11). C. citriodora shows a greater number of haplotypes and divergence between haplotypes than C. variegata. It is likely that this feature is driving the significant difference between the two species, and indicates a larger overall effective population size for C. citriodora.

Sharing of haplotypes can indicate either incomplete lineage sorting or ongoing gene flow/hybridisation between taxa. Incomplete lineage sorting is simply the retention of ancestral haplotypes in some populations by chance despite nuclear divergence (as a result of population isolation). Shared haplotypes due to incomplete lineage sorting are generally internal haplotypes with participant taxa or populations also expressing
additional unique haplotypes derived from both shared and other unique (non-shared) haplotypes (Wendel and Doyle 1998) such as H-5 and H-11. As a chance event, the retention of ancestral haplotypes is also expected to occur randomly among populations. However, H-5 and H-11 occur in populations that are geographically relatively close to each other. Both of these shared haplotypes are also found in two of the same populations (9 and 17). The probability of two ancestral haplotypes retained simply through drift in the same populations seems unlikely and supports the occurrence of some hybridisation between these three species. H-4 is also shared between C. variegata and C. citriodora and is derived from one of the more internal, shared haplotypes (H-5). This more derived haplotype is found within the known area of range overlap between the two taxa, lending support to the inference of gene flow/hybridisation between populations of C. citriodora and C. variegata and the proposed sub-species status of C. variegata by McDonald and Bean (2000).

2.5.2 Historical Processes and Contemporary Population Structure
Patterns of RGF/IBD were found at all NCA clade levels, including the total cladogram. The tendency for most pairwise location distances to fall with increasing clade level, with most reaching zero by the 3-step clade level, also supports a model of IBD (Templeton 2001). Additional support for IBD from the Mantel test indicates that limited seed dispersal between proximal populations has been the dominant process influencing population structure throughout the species complex, at all temporal scales. However, with the exception of C. maculata, the expected geographical pattern in haplotype distribution, associated with an ongoing process of IBD, was not obvious from the network. Given the divergence of the closest sister taxon, C. torelliana, and its exclusion from the network, it is unlikely that the lack of geographic structure in haplotype distribution is the result of introgression.

The lack of geographic association between haplotypes suggests a complex history of population processes shaping the contemporary distribution. For example, H-13 is derived from one of the two unique C. maculata haplotypes (H-17) found only in the southern populations of C. maculata range (i.e. southern NSW and eastern Victoria). The three intermediate geographical populations are all C. maculata populations and are fixed for the H-18 haplotype. The two C. maculata haplotypes both appear to have diverged from the most ancestral haplotype in the network, H-2. This suggests that these
haplotypes may once have been more widespread and, over evolutionary time, they have been lost from populations of the three northern species through genetic drift during times of range contraction. The *C. maculata* populations must also have been historically geographically isolated from the other species and themselves gone through periods of range contraction that isolated the northern subset of populations (H-18) from the southern subset (H-17). Range contractions reduce the effective population size \( N_e \), which in turn amplifies the effect of genetic drift. Severe reductions in Ne over long periods may have led to the chance fixation of divergent haplotypes (H-17 and H-18) in the different areas of *C. maculata*’s distribution. Therefore, the observed pattern of haplotype distribution may be the result of several contributing factors. These may include the localised loss of the *C. maculata* haplotypes, specifically H-17, after H-13 arose in the region in which it is now found. In addition, the subsequent geographic isolation of *C. maculata* from *C. variegata* and *C. henryi*, and further range contractions and loss of haplotypes within *C. maculata*, may have occurred.

Since H-13 is a recently derived haplotype, alternatively it may have arisen in the southern populations of *C. maculata*. Its dispersal to northern populations, by-passing several intermediate populations, may be due to artificial dispersal via intentional or accidental movement of seed by Aboriginal people. Due to the distance (approximately 800 km) involved this may have required capsules to be passed between several groups of people as part of trading or ceremonial meetings. However, it would have only required one successful event for the given distribution pattern to arise and may therefore reflect a true LDC evident.

The occurrence of population range contraction (fragmentation) and expansion events in conjunction with an overall tendency for RGF between populations is likely to have led to the loss of haplotypes from intermediate locations. This may account for the lack of geographic structure found among haplotypes. If neighbouring populations have been historically isolated then the effect of genetic drift, acting independently within each population, may have led to the localized loss of some haplotypes. Historical isolation in response to climate change among many plant populations capable of contemporary RGF has been hypothesised (Schaal and Olsen 2000). Due to the limited seed dispersal potential of these species, the degree of range contraction may not necessarily have been drastic in order to result in effective isolation. The large variance in genetic distance
between proximate locations (identified in Figure 2.2) also suggests that the effect of genetic drift between neighbouring populations has been substantial on a historical timescale, although genetic isolation has not been maintained. Large average pairwise location distances (at the 1-step level) within many of the more northern populations where fragmentation/expansion was inferred, support the occurrence of secondary contact between previously isolated populations (Templeton 2001). The single large pairwise location distance found at the 3-step clade level was within a population located in the recognised hybrid zone between *C. citriodora* and *C. variegata*. Although no inference of fragmentation was found at this clade level, evidence of secondary contact in this population may lend further support to a history of range contraction and expansion in this geographic region.

The combined effects of these processes may also have led to the extinction of haplotypes from all populations throughout the species complex (Schaal and Olsen 2000). Therefore, the large number of missing haplotypes identified within the haplotype network may not be simply the failure of sampling to identify extant haplotypes, but may also reflect extinction of these haplotypes.

An inference of fragmentation leading to temporally sustained evolutionary lineages is required in NCA as an indication of genetic cohesion in a species (Templeton 2001). However, all fragmentation events in this study were inferred at the 1-step level, with subsequent gene flow through range expansion or long distance movement. Therefore, no support of independent lineage status for any of the four species was found from the analysis of this cpDNA data.

The exclusion of *PGM2* from the nDNA analysis resulted in correspondence between the cpDNA and allozyme markers, and identified *PGM2* as a non-selectively neutral locus. Both data sets showed limited divergence between populations of *C. citriodora* and *C. variegata*, and a greater affinity of *C. henryi* to these two northern species than to populations of *C. maculata*. In conjunction with our cpDNA results, the reanalysis of the allozyme data suggests that *C. citriodora*, *C. variegata* and *C. henryi* are part of one morphologically variable species influenced by the variable environment within which they occur across their ranges.
CHAPTER 3: Fine-scale Genetic Variation

3.1 Introduction

3.1.1 Spatial Structure

The potential for significant spatial structuring of genotypes due to restricted dispersal and environmental heterogeneity has been shown in both plants and animals (Epperson 2003). However, localised genetic structure within populations is thought more likely in plants due to the inability of adults to move and the moderate to strong limits in the dispersal potential of both pollen and seed (Heywood 1991; Vekemans and Hardy 2004). The internal structuring of plant populations in a relatively homogenous environment may occur through the formation of local pedigree structures as a consequence of limited gene dispersal. Genetic similarity among neighbouring individuals is consequently higher than between distant ones (Epperson 2000; Vekemans and Hardy 2004). This follows the pattern predicted under the theory of isolation by distance where genetic drift and dispersal are at equilibrium. The fine-scale structure of plant populations is therefore believed to be dependant on the degree to which limitations to seed dispersal produce and maintain groups of kin-related individuals, and the degree to which this structure is swamped by long-distance pollen dispersal (Epperson 2000).

3.1.2 Molecular marker

Microsatellites have increasingly become the molecular marker of choice for population genetic studies (Hancock 1999). They occur throughout the genome, are usually highly polymorphic, and when occurring in non-coding regions are generally considered selectively neutral and so reflect unbiased patterns of mutation, gene flow and genetic drift (Hancock 1999). Mutational changes in microsatellites occur as the addition or loss of repeat units, with differences in alleles based on the length of the sequence (i.e. the number of repeat units) (Jarne and Lagoda 1996; Estoup and Cornuet 1999). Whether microsatellite mutations occur one repeat unit at a time, in a step-wise pattern, or whether mutations involving multiple repeat units can occur in one event, is still largely unknown. However, as mutations are simply the loss or addition of motif repeats, and both are considered equally likely, it is clear that microsatellite mutations can result in alleles being identical in state (IIS) rather than identical by descent (IBD) i.e. homoplasy (Jarne and Lagoda 1996; Estoup and Cornuet 1999).
Fundamental measures of population genetic differentiation such as $F_{ST}$ are based on the infinite allele model (IAM) (Kimura and Crow 1964) of evolution, which assumes that each point mutation produces a new unique allele. This is considered unlikely in microsatellites due to both homoplasy and potential upper limits in repeat length. $R_{ST}$, an analogue of $F_{ST}$, is based on the stepwise mutation model (SMM) (Kimura and Otha 1978). Under the SMM, new alleles are created via the loss or addition of one microsatellite repeat unit at a time and differences in allele size reflect relatedness between individuals. Although neither the IAM nor the SMM deal with the occurrence of homoplasy in microsatellites, the SMM is generally considered as a more appropriate model of evolution because similarities and/or differences in allele length variations are considered, rather than simple allelic frequencies as in the IAM (Balloux and Lugon 2002). As there is still debate surrounding this issue many studies include calculations of both $F_{ST}$ and $R_{ST}$ measures to estimate population differentiation.

In studies where microsatellites are used to determine deviations from genetic equilibrium, null alleles, or point mutations in the primer site that prevent amplification of the associated allele, can be problematic (Chakraborty 1992). Null alleles can inflate estimates of homozygosity, as some heterozygotes for the null allele will be scored as homozygotes of the single amplified allele (Chakraborty 1992). Null alleles are of less concern, however, when microsatellites are used to measure population structure based on relative allelic frequencies, rather than estimates of outcrossing rates, and relatedness.

### 3.2 Objectives

The aim of this study was to assess the genetic diversity and localised spatial structure within a single natural population of *C. variegata*. Specific aims were to:

1) to test the hypothesis that neighbouring trees are more closely related than would be expected by chance

2) to determine if differentiation is occurring across spatially disjunct potential sub-populations

3) to identify the proportion of known maternal trees also acting as pollen/paternal parents, and

4) to identify any differences in allele frequencies and heterozygosity between two groups of trees: the maternal parents and their progeny.
3.3 Materials and Methods

3.3.1 Study Species

*Corymbia* variegata is the central species of the three spotted gum species forming a replacement series along the east coast of Australia (Hill & Johnson 1995). Flowering in the species is often irregular and pollination is predominantly by insects and birds (Hill & Johnson 1995). However, Southerton *et al.* (2004) recently reported long-distance migration of several species of flying-fox including *Pteropus scapulatus* and *P. poliocephalus*, both of which are known to feed on *C. citriodora* and *C. maculata* and may act as long-distance pollen dispersal vectors. Seed release from capsules is passive, resulting in limited dispersal, dependant on tree height and wind conditions (Hill & Johnson 1995; McDonald *et al.* 2000). A study of the mating system of *C. citriodora* (the closest sister taxa of *C. variegata*) by Yeh *et al.* (1983), estimated an average inbreeding co-efficient of 0.205 with 14.7% of open-pollinated seed resulting from self-fertilization (based on allozyme data). This study however was based on open-pollinated seed collected from a seed production area and not from a natural stand. As *C. variegata* is an important commercial source of hardwood, levels of population inbreeding and the extent of localised spatial structure in genetic variation are of commercial as well as evolutionary interest.

3.3.2 Study Site and Field Methods

A natural stand of *Corymbia variegata* in the Woondum State Forest was chosen by the Department of Primary Industries & Fisheries Agency for Food and Fibre Sciences, Forestry Research (AFFS FR) as the base population for a progeny field trial, part of a long-term breeding improvement program for the species. The population is located 15 km south of Gympie in southeast Queensland (-26.1833°S, 152.6667°E). This population was included in the cpDNA analysis of the previous chapter (population 12, Figure 2.1). An individual tree seed collection of 130 trees was undertaken by AFFS FR in January 2000. An attempt was made by AFFS FR at that time to sample trees at least 50 m apart and all sampled trees were permanently marked. Seed from 127 of the trees sampled by AFFS FR were planted in a progeny field trial at Tiaro, 92 km north of Gympie, southeast Queensland. Trees sampled covered an area of approximately 7500 m². A disjunction of approximately 1 km occurs in the middle of the sample area, potentially separating the otherwise continuous population distribution into northern (n=103) and southern sub-populations (n=27). There are no *C. variegata* occurring in
this area of discontinuity, which is lower in elevation than the surrounding sampled areas (Figure 3.1).

**Figure 3.1:** Location map of the study population (within Woondum State Forest) and the locations of each of 130 adult trees of *C. variegata* sampled.
Both the natural population and the field trial were used in this project to assess the
localised spatial structure of both molecular and quantitative spatial variation. The
design of the trial and quantitative trait assessment are discussed in Chapter 4 of this
thesis. With the assistance of AFFS FR staff, leaf samples were collected from all 130
mature trees for molecular analysis and their locations recorded with a GPS. The
position of each tree was then mapped using ArcView software (Figure 3.1). Leaf
samples were also collected from one progeny of each of 127 individual tree seed lots
(maternal parents) planted in the progeny trial. Progeny were 18 months old at the time
of sampling.

3.3.3 Molecular Methods
Total genomic DNA was extracted using the QIAGEN DNeasy Plant Mini-kit
(QIAGEN Industries) after first grinding the leaf tissue (100 mg) using a mortar and
pestle in the presence of liquid nitrogen and Poly-vinyl pyrolidone (PVP) powder (5.0
mg). PVP powder was used to assist in the removal of DNA-degrading phenolic
compounds (Loomis et al. 1979).

Fourteen sets of *C. variegata* microsatellite primers developed by Jones et al. (2001)
were screened across 32 individuals each. Seven loci were selected (EMCRC26,
EMCRC27, EMCRC34, EMCRC35, EMCRC37, EMCRC38, EMCRC40) which
showed substantial polymorphism and no signs of excess homozygosity (suggesting the
presence of null alleles) or indications that more than one locus was being amplified.
PCR was performed using 12.5µL reaction mix containing: 1.25µL of 10xTaq
polymerase buffer, 2mM magnesium chloride, 0.2mM of dNTP’s, 0.4µM of forward
and reverse primers, 0.25 Units of Biotech Taq polymerase and 10-15ng of DNA
template. Amplification was carried out using an Eppendorf Thermocycler under the
following conditions: 3 mins denaturation at 95°C, then 30 cycles of 30 secs at 94°C, 30
secs at 54°C, and 30 secs at 72°C; followed by a final extension phase of 7 mins at
72°C.

PCR product mixed with formamide loading dye (1:1) was denatured and immediately
cooled on ice before loading in 5% acrylamide denaturing gels on a Gelscan 2000 DNA
analyser (Corbett Pty. Ltd.). To ensure consistency in scoring, a commercial standard
(TAMRA 350) was loaded into 4 lanes on each gel. A ladder of pooled individuals, identified from the screening phases to represent alleles across the range at each locus, was also used. Gels were scored using the One-Dscan program (Corbett Pty. Ltd.) and the information known about repeat unit length (Jones et al. 2001).

3.3.4 Statistical Analysis
Genetic differentiation among potential sub-populations was tested using weighted analysis of variance $F_{ST}$ (Cockerham 1973; Weir and Cockerham 1984) and $R_{ST}$ (Rousset 1996; Michalakis and Excoffier 1996) genetic measures in the program GENEPOP (web version) (Raymond and Rousset 1995). Each locus (for both adults and progeny) was first assessed for the likely occurrence of null alleles using Micro-Checker (van Oosterhout et al. 2003). The program estimates null allele frequencies using both methods by Chakraborty et al. (1992) and Brookfield (1996). No loci showed significant estimates of null alleles. Allele frequencies and expected ($H_e$) and observed ($H_o$) heterozygosity for each locus, exact tests for deviations from Hardy-Weinberg equilibrium for each locus and across all loci (p< 0.05; 1000 iterations), and linkage disequilibrium for all locus pairs (p< 0.05; 1000 iterations); were also calculated using GENEPOP (web version) (Raymond & Rousset 1995). Weir and Cockerham’s (1984) $F_{IS}$ values (p< 0.05; 1000 iterations) calculated for each locus and across all loci, were based on the infinite allele (IAM) model. Genetic differentiation in allele frequencies between adults and progeny was also tested using both $F_{ST}$ and $R_{ST}$ measures.

The program CERVUS (Marshall et al. 1998) was used to identify possible paternal parents for each of the progeny and estimate the degree of selfing in the population. When given the multi-locus genotypes of both the progeny and the known maternal parent, Cervus calculates the progeny’s likelihood of decent (LOD) for all nominated candidate paternal parents. An error rate of 1% was specified to allow for possible mistypings, null alleles and mutations, as recommended by Marshall et al. (1998). Due to the inclusion of this error rate, the highest LOD score may not coincide with the candidate parent expressing simply the lowest number of mismatches, but rather the candidate parent that is statistically most likely to be the true parent, given the allele frequencies of all sampled individuals. Delta scores (the difference in LOD scores between the first and second most likely candidate parents) are also calculated by
CERVUS and are a measure of the reliability of assigning parentage to the most likely candidate. Based on the simulation of a large number of parentage tests, strict (95%) and relaxed (80%) confidence levels are applied to the delta values to indicate the probability that true parentage has been assigned (Marshall et al. 1998). All other sampled trees (130) were included as possible candidate paternal parents for all progeny sampled (127). In addition the known maternal parents were also included as possible paternal parents, to allow for possible selfing. Therefore, for each of the 127 progeny sampled, there were 131 possible paternal parents screened. The significance values of the delta values were calculated twice based on estimates that 20% and 40% of the trees in the natural population were sampled.

3.3.5 Spatial Analysis

Distograms showing spatial autocorrelation were produced using Spatial Genetic Software (SGS) (Degen et al. 2001). Distograms were calculated for both the parent trees from the natural population and the sampled progeny. Progeny were assigned the spatial coordinates of their maternal parents for the purpose of this analysis. The program calculates the mean genetic distance across all loci and between all individuals (Gregarious 1978) within a user specified spatial distance class. The genetic distance measure was calculated as:

\[ D_G = \frac{1}{L} \sum_{l=1}^{L} \left( \frac{1}{2} \sum_{k=1}^{n_l} \left| p_{kl} - p'_{kl} \right| \right), \]  

(3.1)

where L is the number of loci, nl is the number of alleles at a locus, and \( p_k - p'_{kl} \) is the difference in allele frequencies between two individuals. The frequencies of alleles at a locus are either 1 (homozygote for that allele), 0.5 (heterozygote with one copy of the allele) or 0 if the allele is absent.
A minimum distance class of 50 m was used as the average distance from one tree to the next was 52.95 m (calculated in SGS). Six pairs of data points fell into the 0-50 m size class (trees were 40 – 50 m apart), well below the minimum of 30 pairs of data points recommended for statistical validity (Degen et al. 2000). These observations were therefore pooled with the 50 –100 m data points for spatial analysis. A Monte-Carlo permutation procedure is used by SGS to test for significant deviation from a random spatial distribution of genetic (or phenotypic) data over the spatial co-ordinates of the sampled points (Manly 1997). A mean value is calculated over all pairs of data points (regardless of distance class), which acts as a reference value against which spatial genetic structure is measured. For each spatial class, the observed genetic distance between individuals is compared with a distribution obtained from 1000 permutations and 95% confidence intervals are calculated. Values outside these intervals are assumed to indicate significant positive (individuals more closely related than random) or negative spatial structure.

Correlation between individual pairwise linear genetic and log transformed geographic distances was assessed using a Mantel test (Smouse et al. 1986; Smouse and Long 1992), to determine any significant isolation by distance (IBD) across the population (999 random permutations, p<0.05) using GenAIEx (Peakall and Smouse 2001).

3.4 Results

Most of the seven microsatellite loci assayed were highly polymorphic (Figure 3.2), and no significant difference (p>0.05) was found in allele frequencies between adults and progeny at any loci based on either $F_{ST}$ or $R_{ST}$ measures. The average number of alleles across the seven loci was 22 for the adult trees and 24 for their progeny (Table 3.1). The observed and expected heterozygosities of the parent trees sampled ranged from 0.84 to 0.90 and 0.88 to 0.94, with averages of 0.86 and 0.91 respectively. Among the progeny sampled the observed and expected heterozygosities were very similar to the adults and ranged from 0.80 to 0.87 and 0.86 to 0.93, with averages of 0.82 and 0.90 respectively (Table 3.1).

Significant deviations from Hardy Weinberg Equilibrium proportions were found at three loci in the parent trees and at five loci among the progeny (Table 3.1). All $F_{IS}$
values for parent trees were positive but they were small (<0.08) and non-significant (p>0.05). No evidence of linkage disequilibrium was found among any locus pairs, suggesting independence between all loci in the parent trees. Progeny $F_{IS}$ values were slightly larger than the parent values at four of the seven loci (Table 3.1). However all were non-significant (p>0.05). Significant linkage disequilibrium (p<0.05), was found only between loci EMCRC26 and EMCRC27, and EMCRC38 and EMCRC40 for the progeny.

**Figure 3.2:** Adult and progeny allele frequencies at seven microsatellite loci for *C. variegata* (continued over the page)
Figure 3.2: Adult and progeny allele frequencies at seven microsatellite loci for *C. variegata.*
Table 3.1: Levels of genetic variation at seven microsatellite loci for adults sampled from a natural population and their progeny planted out in a seed orchard field trial.

<table>
<thead>
<tr>
<th>Locus</th>
<th>Adults (n=130)</th>
<th>Progeny (n=127)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. of alleles</td>
<td>Allele size</td>
<td>H₂O</td>
</tr>
<tr>
<td>EMCRC26</td>
<td>26</td>
<td>92-148</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>92-148</td>
<td>0.85</td>
</tr>
<tr>
<td>EMCRC27</td>
<td>13</td>
<td>141-186</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>141-183</td>
<td>0.80</td>
</tr>
<tr>
<td>EMCRC34</td>
<td>30</td>
<td>60-142</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>66-146</td>
<td>0.83</td>
</tr>
<tr>
<td>EMCRC35</td>
<td>32</td>
<td>136-202</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>138-206</td>
<td>0.81</td>
</tr>
<tr>
<td>EMCRC37</td>
<td>19</td>
<td>244-280</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>244-284</td>
<td>0.83</td>
</tr>
<tr>
<td>EMCRC38</td>
<td>20</td>
<td>176-216</td>
<td>0.87*</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>168-216</td>
<td>0.87</td>
</tr>
<tr>
<td>EMCRC40</td>
<td>22</td>
<td>120-168</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>122-170</td>
<td>0.81</td>
</tr>
</tbody>
</table>

A (H₀) is the average observed heterozygosity
B (Hₑ) is the average expected heterozygosity
No mismatches were found between any progeny and their known maternal parent. One allele in each of the progeny could, therefore, be confidently assigned to the known maternal parent, with no mutational changes having occurred between the generations. Evidence of selfing was found on only two occasions, where the maternal parent was also identified as the most likely possible paternal parent. In both of these instances there were no allele mismatches at any locus (Figure 3.3). This indicates a minimum outcrossing rate for this population of 98%. Limited evidence of selfing in the population further supports lack of population inbreeding inferred from the non-significant $F_{IS}$ values.

For all other progeny, the majority of the first and second most likely paternal parents had mismatches at between two and five loci (Figure 3.3). The delta scores, which indicate the difference between the LOD scores calculated for the first and second most likely paternal parent, ranged between 0.004 and 3.564. When it was assumed that 20% of the population had been sampled an 80% confidence level was calculated for only five of the candidate parents assigned as the most likely paternal parent and a 95% confidence level was calculated for only one candidate paternal parent (Figure 3.4). Therefore, for 121 of the progeny there was a < 80% chance that any one individual adult tree was more likely to be the paternal parent than any other adult tree sampled. Increasing the estimated percent of the population sampled to 40% simply increased the significance of individuals with a large number of mismatches and smaller delta scores. Of the four progeny with significant LOD scores (>80%) (i.e. not including the two selfs), the nominated paternal parents were at least 400 m from the maternal tree, and in one case nominated parents were separated by the spatial disjunction in the population distribution. Both the lack of likely paternal parents, and the distance of the few potential paternal parents from the maternal parents, suggest very low levels of paternity within the population.

The Mantel test based on linear genetic distances and log transformed geographic distances was significant ($p=0.04$) (no significant relationship was found using linear geographic distances). This indicates an IBD effect across the spatial distribution of the population.
No significant genetic differentiation was found between the two potential subpopulations (Figure 3.1) based on either $F_{ST}$ (0.005, $p>0.05$) or $R_{ST}$ (0.01, $p>0.05$) measures. This suggests that the spatial disjunction is not acting as a barrier to gene flow and that the significant IBD effect is not driven by genetic differentiation at this scale (i.e. between sub-populations due to the effects of genetic drift within each). Therefore, further assessment of genetic variation and spatial structure was made on the population as a whole.

**Figure 3.3:** The number of loci mismatches for the first and second most likely paternal parents identified in CERVUS. *C. variegata* trees identified as the 1st most likely paternal parent had allele mismatches with their potential progeny at between 0 and 5 loci. Trees identified as the 2nd most likely paternal parent had allele mismatches with their potential progeny at between 2 and 5 loci.
Figure 3.4: Delta scores, measuring the reliability that the most likely paternal parent assigned is in fact the true parent, for each of the *C. variegata* progeny sampled (assuming 20% of the population was sampled). Open circles represent progeny for whom parentage was assigned with 80% confidence and the open square represents the progeny for whom parentage was assigned with 95% confidence.

Spatial autocorrelation analysis for the adult trees showed a high overall mean genetic distance between all individuals across all spatial classes (\(D_G=0.8506\)) (Figure 3.5). Observed genetic distances between all pairs of trees ranged from 0.82730 to 0.8736. Within the first two spatial classes (i.e. 50-150 m) significant positive spatial structure was evident, with genetic distance between proximal (<150 m) individuals significantly less than the reference (mean) value (Figure 3.5). Within these significant spatial classes, however the mean genetic distance was still quite high (0.8273-0.8356). The average genetic distance between known maternal parents and progeny would be \(\leq 0.5\), equivalent to at least a full-sib relationship as they share a minimum of half their alleles at all loci (selfs would contribute a genetic distance of zero). Negative spatial structure occurred consistently above 1600 m. This distance is greater than that separating the disjunct groups (approximately 1000 m) and supports the lack of genetic differentiation found across the disjunction (i.e. non-significant \(F_{ST}\) and \(R_{ST}\) values).

Autocorrelation of progeny showed a slightly higher overall mean genetic distance across all individuals (\(D_G=0.8587\)) (Figure 3.6). Observed genetic distances were
between 0.8395 and 0.8970 with significant positive spatial structure occurring up to 150 m and then again at 200-250 m. As each of the single progeny individuals sampled for each parent was randomly selected they represent an independent sample of potential trees occurring in the forest. As the autocorrelation results of the progeny are comparable to that of the adult trees, it is likely that the significant spatial positive structure between adult trees up to 150 m apart is real and not an anomaly of the sampling design.

**Figure 3.5:** Distogram of adult trees (A) and progeny (B) of *C. variegata* using 50m spatial classes. Each progeny assigned the spatial co-ordinates of their maternal parent. The straight horizontal line represents the mean genetic distance between all individuals sampled. Grey lines represent the upper and lower 95% confidence levels calculated for each spatial class. The black line represents the observed genetic distance between individuals within each spatial distance class. Significant positive spatial structure occurs where the observed genetic distance between individual’s falls below the lower 95% confidence limit, and significant negative spatial structure occurs where it extends above the upper 95% confidence limit.
3.5 Discussion

3.5.1 Population Genetic Diversity and Outcrossing

All microsatellite loci in adults and progeny showed high levels of allelic polymorphism and heterozygosity ($H_E$) that was equal to, or in excess of, that found in previous studies of eucalypt species (Jones et al. 2002; Holman et al. 2003; Smith et al. 2003) and other forest tree species (Rossetto et al. 1999; Ueno et al. 2000). Although there have been few studies to date of eucalypt species using microsatellite markers, those published have sampled across greater geographical areas than this study (Jones et al. 2002; Holman et al. 2003; Smith et al. 2003). Given that the sampling was among individuals within only 3000 m of each other, the extent of polymorphism identified was surprising.

Deviations from HWE at three loci in the adult trees were accompanied by small and non-significant $F_{IS}$ values, indicating that deviations were due to a lack of certain genotypes rather than homozygote excess as a result of inbreeding. The effect of long-term population inbreeding is a loss of independence between loci resulting in linkage disequilibrium (Ayala 1982). No evidence of linkage disequilibrium was found in the adult population, nor was any evidence of high levels of inbreeding through selfing, as the estimated outcrossing rate was at least 98%. There was also no evidence of high levels of bi-parental inbreeding (breeding between related individuals), as all putative paternal parents identified in this study were at least 400 m from the maternal parents and significant relatedness between individuals was found up to only 150 m. Therefore, likely paternal parents were not significantly related to the maternal parent. These results support the idea that population processes other than inbreeding are influencing the non-random assortment of genotype frequencies at some loci.

Previous studies in eucalypts using allozyme markers have generally shown a homozygote deficit in adult populations, which has been attributed to natural selection acting against homozygosity within natural populations (Potts & Wiltshire 1997). The slightly larger (although still non-significant) progeny $F_{IS}$ values at all loci may indirectly support the occurrence of selection acting against homozygotes in the natural population. As the progeny sampled were the result of open pollinated seed collected, germinated under nursery conditions and then planted into a progeny trial once established, selection acting against homozygotes in the natural environment had
effectively been removed and may have lead to the slightly larger observed homozygosity in the progeny through the retention of selfs. However, the $F_{is}$ values in both parents and progeny were very small. Along with the high outcrossing estimate (98%), it appears that protandrous development of the flower and post-mating mechanisms of self-incompatibility, such as reduced seed production found in other eucalypt species (Potts and Savva 1988; Eldridge et al. 1993; Hardner and Potts 1995), may be effective in limiting the influence of selfs on the genetic diversity of this *C. variegata* population.

Reported single population multi-locus estimates of outcrossing rates for other widespread eucalypt species include: 54% in *Eucalyptus pellita* (House and Bell 1996), 74% in *E. regnans* (Moran et al. 1989), 75% *E. camaldulensis* (Moncur et al. 1995) and 83% in *E. leucoxylon* (Ellis and Sedgley 1993). Mean heritabilities estimated from multiple populations of eucalypts include: 81% in *E. marginata* from the jarrah forest in south-west Western Australia (Millar et al. 2000) and 95% in *E. camaldulensis* from provenance/progeny trials in Thailand (Butcher and Williams 2002). Each of these estimates was lower than found in the current study population (i.e. 98%) and, given that the two progeny identified as maternal selfs could still be outcrossed individuals whose paternal alleles happen to match the maternal parents, this estimate may well underestimate true outcrossing in the Woondum population of *C. variegata*. However, the previous studies were based on the analysis of glasshouse germinates, while sampling of progeny in this study occurred at 18 months of age. It is possible in this study, therefore, that selfed seedlings either died in the glasshouse or that there was some degree of artificial selection disproportionately favouring more vigorous outcrossed seedlings prior to planting. In either case selfed progeny that would normally be included in outcrossing estimates based on germinates could have been erroneously excluded and may have artificially inflated the outcrossing estimate in this study. In several studies of other eucalypt species, however, little evidence has been found to suggest the occurrence of inbreeding depression for the survival or growth of selfs (or open-pollinated progenies) in the nursery (Potts et al. 1992; Hardner and Potts 1995). It is likely, therefore, that the current estimate accurately indicates that selfing is rare in this population.
3.5.2 *Source of Pollen/Paternal Parents*

The complete allelic match found between all 127 progeny and their known maternal parents suggests no mutational change between generations via the female germ-line. Although there has recently been some evidence supporting a male-biased mutation rate at neutral sites in gymnosperms (Whittle & Johnston 2003), there has been no known work done in this area on angiosperms, particularly with regard to microsatellite mutation rates. However, studies of sex-biased microsatellite mutation rates in mammals have shown that any influence of sex on mutation rate seems to vary among loci (Estoup & Cornuet 1999). Therefore, given that we found no mutational changes between maternal parents and progeny, and allowing for some male-biased mutation at some loci, we expected that if the true paternal parent was among the adults sampled, then mismatches should reasonably occur at none or only one or two loci. The majority of all the first and second most likely paternal parents identified for each of the non-selfed progeny, however, showed mismatches at between two and five of the seven loci screened. Also, for 121 of the progeny, the most likely paternal parent was not significantly more likely to be the ‘true’ paternal parent than any of the other adult trees. Therefore, it is unlikely that the true paternal parents were among the sampled adult trees from the population.

The most likely paternal parents for the four non-selfed progeny with significant LOD scores (>80%) were also found to occur at least 400 m from the maternal parent, and in one case across the 1 km spatial disjunction. Therefore, both pollen contribution to neighbouring trees within the population, and the overall contribution of the sampled trees as pollinators in the population, appears to occur infrequently. Although there were unsampled trees in the forest (i.e. within less than 50 m of each other) it seems unlikely that they would have disproportionately contributed any more pollen to the population than the trees sampled. Ongoing gene flow via pollen into the sample area would account for the high levels of genetic diversity and polymorphism found.

3.5.3 *Localised Spatial Genetic Structure*

The significant IBD effect found across the population indicates increased genetic differentiation between individuals as geographic distance between them increases. Genetic isolation by distance is a product of limited dispersal of progeny from the
maternal parent. Each generation there is an increased chance of related individuals mating, and even when this does not occur, progeny still disperse limited distances and face the same situation when they reach reproductive maturity (Epperson 2003). However, no significant genetic differentiation was found between the two potential sub-populations, and parentage analysis provided some evidence of potential pollen parents occurring on opposite sides of the disjunction from the maternal parents. This suggests that pollen-mediated gene flow is strong enough to overcome the spatial discontinuity in the distribution of *C. variegata* in this population. The significant IBD effect may therefore be driven by limited seed mediated gene flow, with subtle spatial substructuring occurring at a fine-scale ‘family group’ level.

Studies of other eucalypt species have shown significant population genetic differences over relatively short distances (Moran and Griffin 1983; Sampson *et al.* 1989; Sale *et al.* 1996). These differences have been attributed to the generally poor seed dispersal potential of eucalypts and have lead to the hypothesis that natural forests may, therefore, consist of a mosaic of family-group clusters with proximal individuals sharing at least half-sib relationships (Griffin 1980; Eldridge *et al.* 1993). There has been little work addressing this hypothesis directly, although a study of *E. globulus* by Skabo *et al.* (1998) (using RAPD’s) did find strong evidence for spatial sub-structuring of individuals within 25 m of each other. Also, Moran and Griffin (1983) suggested that the higher outcrossing rates found in a seed orchard of *E. regnans* compared to natural populations was indirect evidence of family group structure.

Autocorrelation results showed significant positive spatial structure up to 150 m in the adult trees of this study, far beyond that found in *E. globulus* by Skabo *et al.* (1998). However, the mean genetic distance between individuals in the 50-100 m and 100-150 m classes was quite high (0.8273 and 0.8356 respectively). These values suggest only limited relatedness between proximal individuals at the scale sampled i.e. less than half-sib relatedness. Given the high outcrossing rate estimated from the parentage assignment of the progeny, and the lack of significant candidate paternal parents identified among the sampled adults, this relatively high genetic distance between neighbouring trees is not surprising. As virtually all individuals sampled were at least approximately 50 m apart, it is possible that sampling at a finer scale may have
identified higher levels of relatedness (smaller genetic distances) within smaller spatial classes, more in-line with the results of Skabo et al. (1998).

The high levels of genetic diversity and limited spatial genetic differentiation found in the population appear to be the result of a high population outcrossing rate and adequate gene flow via both pollen and seed to balance any effect of genetic drift. The study population forms the basis of a breeding improvement program by AFFS FR. It is clear from this study that there is limited relatedness amongst the maternal trees sampled and that the open-pollinated progeny, which form the basis of the breeding program, are likely to be highly outcrossed. *C. variegata* is known to be affected by the parasitic fungus *Quambalaria pitereka* (Ramularia Shoot Blight). The relationship between the distribution of resistance to this disease and the distribution of genotypes in the population is assessed in the following chapter.
4.1 Introduction

Knowledge of the magnitude of heritable variation and genetic correlations in natural populations provides information relevant to the rates and directions of short-term evolution and historical patterns of natural selection (Ritland 1996). Comparison of genetic differentiation at neutral marker loci with that in quantitative traits can help determine the relative importance of genetic drift and natural selection as causes of among population genetic differentiation in quantitative traits, although few studies have taken this approach (Merila and Crnokrak 2001). In many cases, the degree of genetic differentiation in quantitative traits has been found to exceed that of nuclear molecular markers (allozymes) (Merila and Crnokrak 2001). This is consistent with quantitative traits being under directional natural selection and variation in the direction and magnitude of selection causing differentiation among local populations. Where gene flow may be evident based on neutral molecular markers, differentiation among populations in quantitative traits may still occur. This may be due to environmental differences and varying selection intensity favouring specific phenotypes (Merila and Crnokrak 2001).

4.1.1 Background to host/pathogen relationships

The study of variation in, and the heritability of, disease resistance differs from the study of many other quantitative traits in that the evolution of effective resistance and the maintenance of resistance and susceptible genotypes (resistance polymorphism) in the host may be directly linked to these same characteristics in the pathogen. Many studies have focused on the development of theoretical models as well as the assessment of natural systems in order to understand the genetic systems that govern the co-evolution of hosts and pathogens (Thrall and Burdon 2000; Thrall and Burdon 2002; Brown 2003). The application of mathematical models to investigate the relationship between ecological and evolutionary processes in the coevolution of host-parasite systems most often invoke a fitness cost associated with resistance in the host and virulence in the pathogen to account for the occurrence of susceptible host genotypes (Bowers et al. 1994; Boots and Haraguchi 1999; Thrall and Burdon 2000).
The genetic interactions in host-parasite relationships are still under much debate. However, most theoretical models are based on fairly simple genetic systems (Neema et al. 2001). Empirical studies have most commonly supported the gene-for-gene model (reviewed by Thompson and Burdon 1992), first described by Flor (1955) and later by Burdon (1987 and 1994) and it is regularly used in plant breeding. In the gene-for-gene relationship, plants are resistant to some genotypes of a parasite species but not others. Effective resistance is elicited only if the presence of an avirulence (Avr) gene in the parasite is recognised by a matching resistance (R) gene in the plant. Most R and Avr genes are dominant or semi dominant and so where resistance is controlled by a number of genes, it only takes one R-Avr pair for resistance to be effective. However, many plant-pathogen interactions in natural systems have not followed this model, and alternative models such as the matching allele model, which assumes one locus with many alleles in the host, have also been proposed. Under either of these systems, however, plants with the resistance allele should always have an advantage over those with a susceptibility allele, ultimately diminishing polymorphism in resistance and susceptibility to the point of fixation of the resistant allele (R) (Neema et al. 2001; Brown 2003).

To explain the maintenance of polymorphism in resistance and susceptibility, a cost to the host of resistance has often been considered important in gene-for-gene systems. The potential costs of resistance, or selection for it, include direct energetic costs incurred by the host, through being able to resist the pathogen (Loehle and Namkoong 1987) and other adverse pleiotropic effects (where a single gene affects numerous traits). Probably the most crucial factor, though, is that disease resistance is almost always one of many components of fitness and any selection for resistance is likely to incur a cost in the selection differential for other fitness components (Burdon 2001).

Under the cost of resistance model, in the absence of the disease resistance is costly, reducing the fitness of host plants. This places them at a selective disadvantage, while at times of exposure to the disease these individuals have a selective advantage over susceptible plants. In this way, balancing selection maintains resistance polymorphism in the population. A recent study in Arabidopsis thaliana provided the first evidence of sufficiently high costs balanced by benefits of the R-gene in plant defence, maintaining
polymorphism in susceptibility and tolerance to the disease (*Pseudomonas syringae*) (Tian *et al.* 2003).

The extent and nature of the costs required in the host for polymorphism to be maintained have been questioned in many natural populations. Thrall and Burdon (2002) argued that it was not always necessary to assume resistance or virulence costs to maintain high levels of polymorphism when host and pathogen dispersal is very local. In such cases, virulence polymorphism in the pathogen was maintained, indicating that the coevolution of host-pathogen relationships among small isolated populations subject to extinction may be quite different from that occurring in large populations. This study supported previous assertions that frequency dependant selection among populations that incur changes in population size may have the potential to promote genetic polymorphism in both the host and pathogen (May and Anderson 1990, Thrall and Burdon 2000).

It is also possible that the response to selection for resistance may be limited (even though response to susceptibility may be high) especially at higher levels of resistance. This would particularly be the case if there were important non-additive (especially epistatic) components of inheritance, in which case the contribution of resistance to fitness may offer diminishing returns (Burdon 2001).

In a study of the response of *Arabidopsis thaliana* to the bacteria *Pseudomonas syringae*, Kover and Schaal (2002) showed that plants may vary not only in their resistance to pathogens but also in their tolerance. Resistance traits can be measured in several ways and are broadly defined as host traits that reduce the extent of pathogen infection. In crop varieties, yield with and without the pathogen infection is often used to measure resistance. In contrast, genetic variation for disease resistance in natural populations is usually measured by quantitative variation in visual symptoms. Kover and Schaal (2002) proposed that although some traits may confer resistance by preventing or reducing pathogen growth on the host, completely different host traits might increase host tolerance by diminishing the effect of infection on overall fitness.
4.1.2 Spatial structure and disease resistance

A number of studies have shown considerable variation for disease resistance in plant species in natural environments (Parker 1985; Burdon 1987; Burdon and Jarosz 1991; Bevan et al. 1993; Burdon et al. 1999). Most of these studies have taken a metapopulation approach aimed at understanding not only why polymorphism in disease resistance exists and how it is maintained, but also the role that spatial genetic structure (in both the host and pathogen) may have in determining the persistence of variation in resistance and susceptibility within the host species (reviewed by Burdon and Thrall 1999; Burdon et al. 1990).

A combined study by Carlsson-Graner and Thrall (2002) using both empirical (host plant Lychnis alpina and pathogen smut fungus Microbotryum violaceum) and modelling approaches concluded that differences in disease incidence and prevalence seen in natural populations is most likely because host resistance varies as a function of the degree of connectedness between populations. However, even in the most highly connected systems, with substantial gene flow between populations, the susceptible genotype persisted (Carlsson-Graner and Thrall 2002).

Within population variability in resistance has been found in forestry breeding programs, where breeding for resistance requires identification of individuals carrying the disease resistance genes. For example, blight resistant European Chestnut trees (Castanea sativa) have been developed for reforestation in Switzerland (Devey et al. 1995), Cronartium fusiforme resistant seed orchards for Pinus taeda and Pinus elliottii have been established in the southern USA (Powers 1984), and Phytophthora cinnamomi rootrot resistant lines have been developed in Eucalyptus marginata (Stukely and Crane 1994). Within population variability in the resistance and susceptibility of C. variegata to the parasitic fungus Quambalaria pitereka has also been observed (AFFS Forestry Research internal reports). However, few studies have compared the spatial relationship between localised molecular genetic structure and the distribution of disease resistance and susceptibility within a single natural population (Merila and Crnokrak 2001).

Pathogens can potentially be distributed uniformly across a single population when that population exists within a relatively uniform environment. Within such an environment,
microenvironmental changes are less likely to lead to significantly different responses of host trees to a disease. Also, climatic conditions, which may influence the dispersal and virulence of the disease, are less likely to fluctuate across the distribution of the population. The degree of resistance polymorphism and its distribution within a population as a result of selection pressure may, therefore, be regulated primarily with respect to the relative resistance and virulence expressed by individual host plants and the disease, rather than by dispersal barriers or spatial structure in the distribution of the pathogen itself. However, spatial genetic structure of the plant host can still occur at the local scale (as shown in Chapter 3) due to limited seed dispersal and the resulting formation of family group clusters.

4.2 Objectives
As outlined in Chapter 1, populations of *C. variegata* are affected by the fungal disease *Quambalaria piterea* (Ramularia Shoot Blight or RSB). As such, disease resistance in *C. variegata* may provide a model study system to explore within population variation of a quantitative trait potentially influenced by selection. Furthermore, *C. variegata* a commercial timber species and is the focus of breeding programs managed by the AFFS Forestry Research. Identification of genetic control in the disease resistance trait may aid future breeding programs.

The specific aims of this study were: 1) to estimate the heritability of disease (RSB) resistance and its association with other fitness traits (growth and form) within the Woondum population of *C. variegata*; 2) to estimate the genetic resistance of each of the maternal parent trees in the forest based on the mean resistance of their offspring, for the purpose of aiding future tree improvement programs; and 3) to determine any localized patterns of spatial structure in resistance polymorphism and any correlation with genetic spatial structure found in neutral molecular markers (refer to Chapter 3).

4.3 Methods
4.3.1 Field trial and traits measured
The progeny trial studied was established by AFFS Forestry Research in July 2000 to advance the long-term genetic improvement of *C. variegata* by testing the resistance of both families and provenances to the fungal disease *Quambalaria piterea* (RSB). The
trial focused on screening open-pollinated families derived from several provenances, including the Woondum provenance. This provenance had shown evidence of variable tolerance to RSB in previous provenance trials established by AFFS Forestry Research (AFFS Forestry Research internal reports). The families from this provenance included in the trial were derived from open-pollinated seed collected by AFFS Forestry Research in January 2000 from 127 of the 130 trees that formed the basis of the molecular and spatial analysis of the previous chapter.

The trial was planted at Tiaro, approximately 60 km north of the Woondum population sampled for inclusion in the trial. The trial design was generated using Alpha + version 2.3 (CSIRO Australia) and was a randomised incomplete block with seven replicates, each containing 21 incomplete plots. Each plot contained 12 separate families, each represented by a line of four trees (see Appendix 2). Each family was randomly assigned to a plot within each of the seven treatments and each family was equally represented throughout the trial at the time of planting (i.e. 4 trees in each of seven plots = 28 open-pollinated seedlings per family). The trial also contained seedlots from 12 other *C. variegata* provenances (both individual tree seedlots and provenance bulks). However, only the Woondum families were considered in this study. As only a subset of families planted in the field trial were analysed, a randomised complete block design was used in analyses.

Seedlings were planted in November 2000, when they were 5 months old. The height of seedlings at planting (Height0) was recorded by AFFS Forestry Research. RSB occurs naturally in the area surrounding the trial location and a severe outbreak of the RSB disease was first noted in March 2001, when the plants were 9 months old. Individual assessment of physical crown damage due to RSB susceptibility (RSB) and the presence or absence of leader damage (Leader) was made in May 2001. Susceptibility was categorised using a 6 point scale reflecting the visible percentage of crown cover damage on each tree: 1 = 0%, 2 = 1-10%, 3 = 11-25%, 4 = 26-50%, 5 = 51-75%, 6 = 76-100%. The height (m) at 1 year (Height1) was also measured in August 2001.
4.3.2 Statistical analysis

The general analytical approach followed Jordon et al. (1999). The program ASReml was used to calculate variance and covariance components and estimate fixed and random effects (Gilmour et al. 1997a,b). This program uses a restricted maximum likelihood approach with an average information algorithm and sparse matrix technology (Gilmour et al. 1995; Gilmour et al. 1997b). The model fitted to the data was:

\[ y = \text{mean} + \text{replicate} + \text{family} + \text{plot} + \text{residual} \quad (4.1) \]

where mean was the trait mean, replicate was the fixed effect of replicate, family was the random family effect, plot was the random variation between the seven plots of 4 trees within a family, and residual was the unexplained random variation between trees within a family within a plot. The variances and covariances relating to family are purely genetic and reflect genetic differences between the female parents in the Woondum population. The incomplete block and replicate effects reflect quantitative trait variation within a family due to small-scale environmental differences among blocks and replicates. Error effects include both residual additive genetic effects and non-additive genetic effects (due to possible dominance or epistasis) and microenvironment variation. This model was fitted as a univariate analysis for the traits disease susceptibility (RSB), height of seedlings at time of planting into the trial (Height0), height after 1 year and after exposure to the disease (Height1) and leader damage (Leader). As Leader was a presence/absence trait, a binary model was fitted for the analysis using a probit link function.

The model was then fitted for the bivariate case to obtain direct estimates of the correlations between traits for each random term using the CORR option in ASReml. Univariate variance component estimates were used as starting values in the bivariate model. The additive genetic correlation between traits was estimated from the variance and covariance components of each family (Jordan et al. 1999). The binary trait Leader was treated as a normal trait in these bivariate analyses, as the genetic correlations between a binary trait and a quantitative trait are not affected by the frequency of the
binary trait and are therefore unbiased (Olausson and Rönningen 1975; Chambers et al. 1996). Genetic correlations of disease susceptibility with height at the time of planting (Height0), height at one year (Height1) and leader damage (Leader) were also calculated as:

\[ r_{\text{family}(1,2)} = \frac{\sigma_{\text{family}(1,2)}}{\sqrt{\sigma_{\text{family}(1)}^2 \sigma_{\text{family}(2)}^2}}, \quad (4.2) \]

where \( r_{\text{family}(1,2)} \) is the correlation between traits 1 and 2 at the family level, which is the level of additive genetic variation, \( \sigma_{\text{family}(1,2)} \) is the family covariance between the traits, and \( \sigma_{\text{family}(1)}^2 \) and \( \sigma_{\text{family}(2)}^2 \) are the family variance components for each trait. The phenotypic correlation between traits was simply derived from the Pearson’s correlation between individual traits.

The significance of variance components and correlations was tested using likelihood ratio tests. To test that the variance components were significantly greater than zero a one-tailed likelihood ratio test, which uses an approximate \( \chi^2_{n,s} \) distribution for –2log L was used (Stram and Lee 1994). Tests in the univariate analysis compared models including and excluding the relevant effects. Tests of correlations compared the full model with one in which the correlations were constrained to zero and a likelihood ratio test using \( \chi^2_1 \) was used to test significance (Stram and Lee 1994; Gilmour et al. 1997b).

Given the high outcrossing rate (98%) estimated for this population in the previous chapter, narrow-sense heritabilities \( h^2 \) for each trait were calculated in ASReml using:

\[ h^2 = \frac{\sigma_{\text{family}}^2}{r(\sigma_{\text{family}}^2 + \sigma_{\text{plot}}^2 + \sigma_{\text{residual}}^2)}, \quad (4.3) \]
where $\sigma^2_{\text{family}}$ is the among family variance, $\sigma^2_{\text{plot}}$ is the plot variance component and $\sigma^2_{\text{residual}}$ is the residual unexplained random variance from the univariate analysis for each trait. These heritability ($h^2$) estimates assume no self-pollination and that all open-pollinated progeny from the same parents have a half-sib relationship, i.e. complete outcrossing ($r=0.25$). Heritability ($h^2_{OP}$) estimates were also calculated where the coefficient of relatedness ($r$) was assumed to be 0.4, which is equivalent to an outcrossing rate of 70%. This rate is commonly assumed in open-pollinated eucalypt families (Griffin and Cotterill 1988; Potts and Jordan 1994) and was calculated in this case to allow comparison with other studies. Standard errors of heritabilities are estimated using an expanded Taylor series in ASReml (Gilmour et al. 1997b).

Best Linear Unbiased Predictions (BLUP’s) of family effects were calculated from RSB and Height1 using the variance components estimated in the univariate model. These family BLUP’s represent the family deviation around the grand mean of a given trait and are an estimate of the General Combining Ability (GCA) of the female parent trees in the forest. The GCA is the influence a parent has on its progeny when randomly crossed with other individuals, while breeding values represent the value of an individual as a parent. Therefore, both measures are indicators of additive gene action and the GCA is equivalent to half the female breeding value (Wilcox and Shelbourne 1975; Nicholas 1987; Falconer and Mackay 1996). GCA values may be poor estimates of the breeding values if there is significant selfing occurring under open-pollination (in which case the breeding value would be less than twice the GCA), particularly if the trait is subject to inbreeding depression (Wilcox and Shelbourne 1975; Nicholas 1987; Falconer and Mackay 1996). However, a strong correlation between open-pollinated and control pollinated breeding value estimates for a disease trait (Mycosphearella ssp.) has been reported in Eucalyptus (Dungey et al. 1997) and extremely low levels of selfing were observed in the Woondum population (Chapter 3).

Using the known location of each tree in the forest, these family BLUP’s were mapped to show the distribution of susceptible (positive values) and resistant (negative values) trees in the Woondum forest. The RSB and Height1 BLUP’s for each family were also graphed to show the relationship between the two traits.
Spatial analysis of the distribution of disease resistance in the natural forest was undertaken using least square means for RSB, calculated for each family in SAS (SAS Institute Inc 1990). The model used was the same as that shown in equation 4.1, however, family was treated as a fixed effect. These values were then assigned to the relevant parent trees in the forest and, using each tree’s spatial position, distograms were produced using Spatial Genetic Software (SGS) (Degen et al. 2001). The program calculates the mean trait difference between all pairs of individuals within the user specified spatial distance classes. A minimum distance class of 50m was used which included greater than 30 pairs of data points per class as recommended for statistical validity (Degen et al. 2001). The distance measure used in the program is a simple city-block distance measure (Deichsel and Trampisch 1985):

\[ t_{im} = \frac{a_{im} - \bar{a}_m}{st_m} \]  
(4.4)

\[ D_{ij} = \frac{\sum_{m=1}^{n} |t_{im} - t_{jm}|}{n} \]  
(4.5)

where \( a_{im} \) is the value of individual \( i \) for trait \( m \), \( \bar{a}_m \) represents the mean value for trait \( m \), \( st_m \) is the standard deviation of trait \( m \), and therefore \( t_{im} \) gives the value of individual \( i \) for trait \( m \) after z-transformation. Within each size class, the distance between all pairs of individuals, \( D_{ij} \), is the sum of the absolute distances divided by the \( n \) number of traits, which is one in this case.
4.4 Results
Substantial phenotypic variation in RSB disease resistance and susceptibility, tree height (Height1) and leader damage (Leader) was measured from the field trial. Trees were recorded in all categories of RSB disease susceptibility, from virtually unaffected to highly affected. The partitioning of variance components for each of the traits RSB, Height1 and Leader are shown in Figure 4.1. The largest proportion of explained phenotypic variation in all three traits was attributable to genetic variation between families (18-23 %). There was very little variation among replicates (0.2-3.0 %), indicating that any microenvironmental differences between replicates had no significant effect on either family growth rates or the dispersal of the pathogen across the trial area. This suggests that all trees were equally exposed to the disease. The plot effect was highly significant for all three traits (4-12 %), indicating variation within a family between plots. All three traits also showed moderate to high heritabilities, between 0.30–0.58 when an outcrossing rate of 70% (r=0.4) was assumed, and between 0.48-0.93 when it was assumed that no self-pollination was occurring in the population (r=0.25) (Table 4.1). In particular, substantial family variance was found for RSB (17.5 %), along with high heritabilities of 0.44 assuming 70% outcrossing and 0.71 assuming complete outcrossing. This indicates significant genetic variation in disease resistance among the maternal parents within the natural population.

Figure 4.1: Percentages of the total phenotypic variance attributable to each effect in the open-pollinated progeny trial of C. variegata. Significance of the variances are based on likelihood ratio tests with *** = p<0.001 and n.s=non-significant.
The physical effect of RSB is a reduction in growth and loss of form in infected trees. Therefore, a large proportion of the genetic variation expressed in the heritabilities of Height1 and Leader may be due to the effect of RSB. To test this both the phenotypic and genetic correlations between these traits (Height1 and Leader) and RSB were calculated to determine the strength of any linear relationship between them (Table 4.1).

There were substantial negative phenotypic ($r = -0.53 \pm 0.076$) and genetic ($r_g = -0.84 \pm 0.04$) correlations between disease susceptibility (RSB) and the height of trees after exposure to the disease (Height1). This indicates that genetic variation in susceptibility is associated with genetic differences in vigour. A very strong, positive correlation between leader damage (Leader) and disease susceptibility (RSB) was found ($r_g = 0.99 \pm 0.01$), suggesting that these traits were measuring the same genetic effect. No significant phenotypic or genetic correlation was found between RSB and Height0. This indicates that the initial height of seedlings did not influence their susceptibility to the disease and did not contribute to the correlation found between RSB and Height1.

**Table 4.1:** Estimated heritabilities ($h^2$) with standard errors for each trait measured in the *C. variegata* open-pollinated progeny trial. Heritabilities were calculated assuming half-sib relationships ($h^2$, $r=0.25$) and 70% outcrossing ($h_{OP}^2$, $r=0.40$). Phenotypic Pearson’s correlations ($r$) between disease susceptibility (RSB); and the genetic correlations ($r_g$) between disease susceptibility (RSB) and Height1 and Leader.

<table>
<thead>
<tr>
<th>Trait</th>
<th>$h^2$</th>
<th>$h_{OP}^2$</th>
<th>$r$</th>
<th>$r_g$</th>
</tr>
</thead>
<tbody>
<tr>
<td>RSB</td>
<td>0.71±0.092</td>
<td>0.44±0.058</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Height1</td>
<td>0.48±0.048</td>
<td>0.30±0.048</td>
<td>-0.53±0.076</td>
<td>-0.84±0.004</td>
</tr>
<tr>
<td>Leader</td>
<td>0.93±0.115</td>
<td>0.58±0.072</td>
<td>0.89±0.041</td>
<td>0.99±0.004</td>
</tr>
</tbody>
</table>

A negative relationship was evident between the family BLUP values calculated for disease susceptibility and the height of progeny after exposure to the disease in the trial (Height1) (Figure 4.2). Highly resistant families consistently showed substantially better growth than highly susceptible families whose growth was considerably reduced.
Mapping of BLUP values according to the location of each maternal parent tree in the forest illustrated the distribution of polymorphism in disease susceptibility throughout the population and the relative locations of highly resistant and highly susceptible trees (Figure 4.3). There appeared to be no obvious spatial structure or grouping of disease susceptibility or resistance between neighbouring trees, suggesting that disease resistant genotypes are distributed randomly throughout the forest.

Spatial analysis confirmed the lack of spatial structure in the distribution of disease resistance among trees sampled within the Woondum population (Figure 4.4A). Given the results from molecular analysis of this population (Figure 4.4B, result of molecular analysis in Chapter 3), this suggests that disease resistance is dispersed independently of the positive spatial structure in neutral genetic markers.
Figure 4.3: Map of family BLUP values representing the relative genetic merit of maternal parent trees in the natural forest calculated from up to 28 progeny per family established in a progeny field trial. Positive BLUP values (red circles) indicate highly disease susceptible trees, while negative BLUP values (green circles) represent highly disease resistant trees.
Figure 4.4: A: Distogram showing the relationship between similarity in family disease resistance and the proximity of adult trees using 50m spatial classes. B: Distogram of showing the relationship between genetic distance and the geographic proximity of adult trees using 50m spatial classes. Genetic distance was calculated from 7 microsatellite loci. The straight grey horizontal line represents the mean distance between all individuals sampled. Light grey lines represent the upper and lower 95% confidence levels across all spatial comparisons. The black line represents the average pairwise difference in RSB (A) or the average pairwise genetic distance (B) between all individuals within each spatial distance class. Significant positive spatial structure occurs where the actual distance (black line) falls below the lower 95% confidence limit, and significant negative spatial structure occurs where it extends above the upper 95% confidence limit. (Figure 4.4B is reproduced from Chapter 3, Figure 3.5A).
4.5 Discussion

Substantial resistance variability, moderate to high heritability and strong adverse genetic and phenotypic correlations between a tree’s response to the disease and its growth and form were found in this population. Pathogens can be a selective force on plant evolution only when the pathogen infection affects host fitness and when there is heritable variation in resistance traits among individuals over several generations (Kover and Schaal 2002). The inability of plants to move means that any loss in growth potential and/or performance may reduce their fitness in comparison to other trees competing for limited space and resources in the natural forest (Byrne 2000; Namkoong et al. 2000). From a forestry perspective, loss of form and reduced growth in susceptible trees also represents a loss of fitness with respect to future tree improvement programs.

Estimated heritability of disease resistance increased substantially when complete outcrossing among all individuals was assumed \((h^2=0.71)\) compared to that calculated assuming an outcrossing rate of 70\% \((h^2=0.44)\). The accuracy of heritability estimates based on open-pollinated progeny has been questioned because of the potential for unpredictable and possibly differential inbreeding within eucalypt populations (Hardner and Potts 1995; Potts et al. 1995; Hodge et al. 1996). It has also been suggested that accurate estimates of heritabilities and other genetic parameters require fully pedigreed controlled crosses (Potts et al. 1995; Hodge et al. 1996). Although open-pollinated, rather than controlled cross progeny were used in the current study, given the limited number of potentially selfed individuals identified based on microsatellite analysis (Chapter 3), it is likely that heritability of RSB resistance is greater than 0.5 and therefore highly heritable within the Woondum population of \textit{C. variegata}.

No other studies have estimated the heritability of RSB resistance within \textit{C. variegata} or other potentially affected species, and therefore no direct comparison can be made. However, compared to other studies of disease resistance in eucalypts, the level of heritability in resistance found in this study is quite high. For example, \textit{Eucalyptus globulus} and \textit{E. nitens} are both susceptible to \textit{Mycrosphaerella} leaf disease, which causes leaf necrosis and defoliation that can be highly detrimental to growth, similar to the effects of RSB. Several studies of this host-parasite relationship have assessed heritability of resistance to the disease. Reinoso (1992) estimated heritability of 0.31
based on open pollinated progeny of *E. globulus*, while Dungey *et al.* (1997) reported estimates between 0.12 and 0.34, based on both open pollinated families and controlled intra-specific crosses of *E. globulus*. In *E. nitens*, Dungey *et al.* (1997) also estimated heritabilities of resistance to *Mycrosphaerella* leaf disease between 0.01 and 0.20, and between 0.17 and 0.51 for controlled inter-specific crosses between *E. globulus* and *E. nitens*.

The high heritability estimated for disease resistance in *C. variegata*, as well as the strong genetic and phenotypic correlations between the degree of infection and growth traits, suggests that there are significant benefits associated with resistance. It also indicates directional selection pressure may be acting against highly susceptible genotypes. Therefore, the maintenance of variation in disease resistance would likely require a cost of resistance at times when the disease is not prevalent.

Given the high heritability of disease resistance estimated for this population the apparent random dispersal of resistant and susceptible genotypes in the Woondum population was unexpected. No spatial structure was evident from either spatial autocorrelation or the mapping of maternal BLUP values in the natural forest. High heritability suggests that related trees should express similar, inherited levels of disease tolerance or resistance. However, although significant positive spatial genetic structure was found between trees within 150 m of each other (based on microsatellite markers) no spatial structure in disease resistance and susceptibility was found.

It is unlikely that the disparity between the molecular and quantitative results is due to inadequate spatial sampling within the population. Further sampling to include trees occurring between those sampled may have clarified the degree of relatedness between trees occurring directly next to each other. However, it is unlikely that this would have helped to clarify the lack of spatial structure in resistance and susceptibility. If the level of sampling was inadequate, it is more likely that the results would have identified clustering of disease resistant and susceptible phenotypes in the population with no positive spatial structure in the molecular markers. As the level of sampling and the sensitivity of the selectively neutral microsatellite markers was adequate enough to identify positive spatial genetic structure, and therefore significant relatedness, between
trees up to 150 m apart, it is likely that it also reflects the true spatial extent of relatedness within the population.

Although RSB is thought to exist consistently within the forest area, the severity of outbreaks may vary (Ivory et al. 2004). Optimal weather conditions to facilitate severe outbreaks include temperatures between 20-25° C after short periods of heavy rain, coinciding with new flushes of immature foliage. Further development of the lesions caused by the fungus requires dry conditions (Ivory et al. 2004). Periods of drought are common in the region where the Woondum population occurs and therefore intense outbreaks, during which high resistance would be of most advantage, may occur only intermittently. As the sampled parent trees in the Woondum population are unlikely to represent a single aged stand, they may reflect several generations that established under varying environmental conditions. Therefore, these trees would have been subjected to different levels of disease intensity as seedlings. If a tree’s exposure to the disease was low (due to unfavourable conditions for the disease) any advantage bestowed on a tree through the retention of resistance genes for RSB may have been lost or selected against. With no advantage of resistance in this situation, susceptible genotypes could successfully establish.

Given the substantial damage to growth and form observed in the progeny, it seems likely that repeated exposure to extreme outbreaks of the disease would remove highly susceptible trees from the forest through natural selection. However, Ivory et al. (2004) suggest that where plantations have remained free of the disease in the first year, they have been less likely to be affected as badly in subsequent years, suggesting that trees become more able to recover from infection as they get older. This also suggests that although disease infection has growth and form costs, these costs may not be significant in terms of a plant’s overall fitness. Although differences in the recovery potential of different aged stands was not assessed in this study, it may explain how highly susceptible trees, which have probably been exposed to a severe outbreak of RSB at some time, have reached maturity within the population. If plants can withstand the effects of infection more readily as they get older, and they are not exposed to severe outbreaks of the disease while young, they could still reach reproductive maturity. Susceptible alleles may then be passed on to the next generation and, along with costs
associated with resistance, this may be maintaining resistance variability within the population and the apparent random dispersal of resistant and susceptible genotypes.

Another hypothesis that does not require fitness costs for the maintenance of resistance polymorphism has also been proposed by a number of authors (reviewed by Neema et al. 2001). Genes for resistance to pathogens may be clustered in complex loci in the host genome. Selection by local strains would maintain polymorphism for resistance at these clusters through selection for genes that also confer resistance on other strains or other pathogens, or through hitchhiking effects on the rest of the cluster. As no other pathogens are known to affect the fitness potential or incur alternative selection costs in *C. variegata* as severely as RSB, this hypothesis seems less likely than costs directly associated with RSB resistance.

The identification of highly distinct resistant and susceptible trees within the natural population will benefit AFFS Forestry Research spotted gum improvement programs. As *C. variegata* is a plantation hardwood species, obtaining improved seed from trees expressing optimal growth rates and form, despite exposure to RSB, is important for hardwood production programs. As all trees within the forest were permanently marked by AFFS Forestry Research at the time of seed collection, the results of this study will enable direct sampling of seed from trees identified in this study as highly resistant to RSB. This study indicates that any future sampling of the natural Woondum population for seed from RSB disease resistant trees should not be concentrated in one particular area of the forest as disease resistant genotypes do not appear to be spatially grouped. More importantly, this study found that RSB disease resistance is under very strong genetic control and highly correlated with growth and form. Therefore, resampling of resistant genotypes from either the natural forest or the progeny trial will result in large genetic gains in resistance and growth traits in future breeding programs.
CHAPTER 5: General Discussion

5.1 Implications of this Study for the Spotted Gum Complex

No evidence of historical lineage divergence was found among the four putative spotted gum species. Contemporary gene flow was evident between populations of *C. citriodora*, *C. variegata* and *C. henryi*. This species alliance did not support the existing taxonomic relationships in the complex. The overall pattern of spatial structure was one of isolation by distance, with no indication that contemporary phenotypic differences, including differences in morphology and leaf oil composition, arose while populations were in isolation. This indicated that the three northern species of spotted gums represent a single species cline, with morphological differences and variability the product of primary differentiation, maintained by natural selection across a heterogeneous environment.

The influence of natural selection on neighbouring populations of *C. variegata* and *C. henryi* was also inferred from this study. The effect of selection in this case masked the true extent of both historic and contemporary gene flow between the two putative species. Based on both cpDNA sequences and 17 allozyme loci, there was no indication of genetic differentiation between these populations. However, the intensity of directional selection reflected in one allozyme locus (*PGM2*) was strong enough to mask this evidence of ongoing gene flow. The population divergence suggested by this locus reflected morphological differences between the populations of the two putative species. The influence of environmental selection in this way highlights many of the issues surrounding species concepts in plants, in particular eucalypts. Sharp spatial heterogeneity in environmental conditions, when coupled with the immobility of plants, means that environmental selection can be a strong force in creating morphological differences between geographically close populations (Bradshaw *et al.* 1965). As seen in the present study, natural selection may then mask the true genealogical relationships between putative species.

The single population of *C. variegata*, which was the basis of the analysis of fine scale spatial structure in this study, showed phenotypic differences between individual trees in response to exposure to a fungal disease. In this case, it appears that directional selection acting on the plants due to the adverse effects of the disease was not strong
enough, or the fitness costs to the plant were not consistently high enough, to either limit variation in disease resistance or contribute to the spatial structuring of variability over many generations. Some spatial structure in molecular data was found in this population, but the interaction between effective gene flow via both pollen and seed dispersal and possible temporal variability in selection pressure due to changes in the intensity of exposure to the disease, has potentially led to the apparent random distribution of resistant and susceptible genotypes.

The fine scale analysis of molecular and quantitative genetic variation and structure within a single population of *C. variegata* provided information on the evolutionary dynamics of the population as well as significant outcomes relevant to future tree improvement programs by AFFS Forestry Research. The outcrossing rate in this population was estimated to be 98%. This is considerably higher than estimates for many other eucalypt species (Potts and Wiltshire 1997). High levels of genetic diversity were found across multiple microsatellite loci and there was little evidence of either significant selfing or bi-parental inbreeding. Disease resistance was found to be under strong genetic control, which has several important implications for future breeding programs. Due to the high heritability estimated for disease resistance and susceptibility, large genetic gains in disease resistance could be made by collecting seed from highly resistant trees identified in this study (or their progeny from the field trial). Strong adverse genetic and phenotypic correlations were identified between susceptibility and tree height and form after exposure to the disease. This suggests that if disease resistant trees are selected and grown in environments where the disease occurs, strong gains in growth traits and timber production may also be made. Most importantly, however, if resistant maternal and paternal parents occurring greater than 150 m apart are selected from the Woondum population for controlled crossing, they are likely to be unrelated. This means that the genetic and growth gains made through controlled crossing of resistant genotypes from the Woondum population will not be diminished by any possible effects of inbreeding depression. This will also maintain genetic diversity in source stock.
5.2 Relevance of the Study in the Context of Conservation and Forest Management

Genetic diversity has important implications for forest resource management issues, however, many authors have debated the importance and relevance of genetics to conservation issues. Some argue that current ecological, demographic and stochastic events are of greater importance than genetic ones, particularly in the case of endangered species (Caro and Laurenson 1994). However, this view does not recognise the inherent influence of ecological, demographic and stochastic events on the genetic structure of natural systems. Molecular genetic data can provide information not only in relation to extant levels of genetic diversity and contemporary demographic parameters, but also in relation to historical patterns or trends of interest to conservation biologists. Although traditional long-term ecological studies provide valuable information on population demographic trends, molecular genetic markers can potentially provide similar knowledge in a shorter time frame (Boshier and Young 2000).

Demographic parameters such as population size and mating systems can vary from generation to generation. Consequently, direct demographic observations over the short term may not provide a representative view of the long-term status of a population. For example, small population size may be unimportant to a population that has always been small, at least in terms of potential loss of genetic diversity. A small population size may be critical, however, for a population that has recently become small or whose size is increasingly under threat (Boshier and Young 2000). Although the species studied in this project are not endangered, their distributions encompass areas subject to some of the highest levels of human disturbance and land clearing in Australia, resulting in increasing levels of forest fragmentation.

Forest fragmentation alters population connectivity and consequently has direct implications for potential gene flow between populations. Large or highly connected populations generally have large effective population sizes that are a source of genetic diversity through continuous mutation. Potential gene flow between a large number of individuals can promote and maintain high levels of genetic diversity, which enables individuals to respond to shifts or fluctuations in environmental conditions or selection pressures. Small isolated populations have a limited source of variation and reduced ability to maintain diversity. This subsequently reduces their ability to respond to
environmental change. The effect of genetic drift within small populations can exacerbate the loss of variability through inbreeding and inbreeding depression (Young and Boyle 2000).

This study of the spotted gum species complex exemplifies how the contemporary spatial structure of populations can be the result of a complex history of the varying influences of population processes, and the long-term effects that may result from historical demographic changes in forest tree species. For example, in the seven *C. maculata* populations sampled only two cpDNA haplotypes were found. This level of diversity was much lower than that found in each of the three other putative species. Each of the two *C. maculata* haplotypes was unique and fixed in a geographically separate subgroup of the *C. maculata* populations. Genealogical relationships between the observed haplotypes, suggested historical isolation of these two groups of populations from all other sampled populations in the species complex, as well as from each other. Loss of genetic diversity in most populations is a symptom rather than a cause of reduced population size. Therefore, the contemporary lack of genetic diversity in these populations suggests an historic reduction in effective population size and distribution (i.e. a bottleneck), resulting in genetic drift exerting a dominant effect on spatial genetic structure and variability. The lack of evidence of any subsequent gene flow into or between these areas suggests the continued isolation of these populations. Nevertheless, *C. maculata* currently has a widespread distribution and potentially large effective population size. Therefore, this species is an example of how the use of genetic markers can contribute to our understanding of the demographic properties of populations that may not be evident via traditional methods, as well as historical trends in these properties.

In contrast, the three northern species of spotted gums display higher levels of genetic diversity, indicative of a highly connected system of populations. Even though individual populations have undergone their own complex histories, the homogenising effect of effective gene flow is evident through lack of divergence. From a conservation perspective, the patterns shown in these species highlight the importance of gene flow in the maintenance of genetic variation and patterns of spatial genetic structure in forest tree species. In the past, some authors suggested that gene flow contributed little to the determination of spatial genetic structure of a species (Ehrlich and Raven 1969; Levin
and Kerster 1974), while more recently others have argued that gene flow has a major influence on the spatial structure of genetic variation (Ellstrand and Marshall 1985; Hamrick 1987; Hamrick and Nason 2000).

The genealogical relationships identified among the three northern species of spotted gums also highlighted limitations associated with the use of morphological traits as the basis for plant taxonomy. In contrast, species classification based on the cohesion species concept accommodates the propensity for hybridisation, selfing and bi-parental inbreeding in many plant species. It provides a framework for testing specific hypotheses to determine if spatial patterns of genetic diversity are the product of primary differentiation or secondary contact and can help identify historical events such as range fragmentation and range expansion (Templeton 1989; Hull 1997).

Adoption of the cohesion species concept in eucalypt taxonomy in particular could also help clarify the nature and evolutionary significance of hybrid zones, which are often considered evidence of secondary contact. For example, Holman et al. (2003) proposed that two eucalypt clines previously thought to be the result of secondary contact might actually be the result of continuous morphological divergence promoted by a directional selection gradient. The application of the cohesion species concept and associated analytical procedures by Holman et al. (2003) enabled interspecific populations to be distinguished from independent evolutionary lineages. The evolutionary significance of selection is not minimised through this approach, but it does reduce the propensity of selection to mask the effect of other processes such as gene flow.

In many cases, effective conservation requires some component of ‘conservation through use’ of trees (Boshier and Young 2000). Commercial native tree species have little history of domestication in Australia. When grown for timber, they are managed using long rotation times and are often planted in natural settings, where genetic diversity is important for adaptability. Knowledge of the genetic diversity in populations used as a seed source in these situations is critical in order to maintain genetic variation and, therefore, adaptability to a range of environmental conditions over both space and time. Levels of inbreeding may be unimportant from an evolutionarily sustainable context if selfed or highly bi-parentally inbred individuals are selected against at various stages of regeneration. However, levels of inbreeding may be
critical in terms of the levels of diversity sampled for tree breeding or plantation programs or for ex situ conservation, such as farm forestry (Boshier and Young 2000). Revegetation of farm production areas through the selective planting of forest trees and other plant species aims to provide habitat, while also providing future income to landholders through timber production. Farm forestry projects potentially re-establish or maintain connectivity between populations. As these trees will then contribute to local genetic diversity, the selection of source seed can have substantial impacts on the genetic diversity and structure of neighbouring populations. In this context the results from the analysis of localised genetic structure and outcrossing in the single population of *C. variegata* are important. They identify the Woondum population as a viable source of seed with high levels of both genetic and disease resistance variation.

### 5.3 Future Directions

While species designations based on morphological similarities and differences have legitimate benefits for field identification, they often do not reflect true genealogical relationships. There is increasing evidence, including this study, that much of the morphological variation and differentiation among eucalypts in particular does not necessarily reflect genetic diversity or divergence and may be environmentally driven (Holman *et al* 2003). Consensus among plant ecologists is needed with regards to an appropriate species concept that tolerates the influences of selfing and hybridisation, while still providing species classifications from which a relatively consistent level of relatedness can be assumed. Although the results of this study support the application of the cohesion species concept, the application of this concept in future inter-specific molecular studies of plants would provide further evidence of its appropriateness.

Further sampling of the Woondum population of *C. variegata* assessed in this study would clarify the level of relatedness between directly adjacent trees. Whether or not this revealed higher levels of bi-parental inbreeding and/or selfing than identified in the current study, it would provide more specific information on the extent of seed dispersal in this species. Further work on the biology of the fungal disease Ramularia Shoot Blight may also clarify aspects of the host-parasite relationship, as well as the maintenance of resistance variability within this population.
References


References


Moran GF, Bell JC, Griffin AR. (1989) Reduction in levels of inbreeding in a seed orchard of *Eucalyptus regnans* F. Muell. compared with natural populations. *Silvae Genetica* 38: 32-36.


References


Posada D, Crandall KA, Templeton AR (2003) Nested Clade Analysis Inference Key http://inbio.byu.edu/Faculty/kac/crandall_lab/geodis.htm


Ritland K (1990) Genetic distance, diversity and clustering programs: GD and GDD.


Stukely MJC, Crane CE (1994) Genetically based resistance of *Eucalyptus marginata* to *Phytophthora cinnamomi*. *Phytopathology* 84: 650-656.


Appendices

Appendix 1: cpDNA haplotype sequences (JLA+ region) of *C. citriodora, C. variegata, C. henryi, C. maculata* (H-1–H-19) and *C. torelliana* (H-20)
|---|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>H-1</td>
<td>AATTCATAAT GGAATAAT TCAATTTCAT TGTTATTTT TAACATTTTT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

114
| H-1       | TTATCTTAAT TATGAGATAG AAGAAGCAGA AAATTCTAAC CTTCCTATTTT |
| H-2       | ................. ................. ................. ................. G.. |
| H-3       | ................. ................. ................. ................. G.. |
| H-4       | ................. ................. ................. ................. G.. |
| H-5       | ................. ................. ................. ................. G.. |
| H-6       | ................. ................. ................. ................. G.. |
| H-7       | ................. ................. ................. ................. G.. |
| H-8       | ................. ................. ................. ................. G.. |
| H-9       | ................. ................. ................. ................. G.. |
| H-10      | ................. ................. ................. ................. G.. |
| H-11      | ................. ................. ................. ................. G.. |
| H-12      | ................. ................. ................. ................. G.. |
| H-13      | ................. ................. ................. ................. G.. |
| H-14      | ................. ................. ................. ................. G.. |
| H-15      | ................. ................. ................. ................. G.. |
| H-16      | ................. ................. ................. ................. G.. |
| H-17      | ................. ................. ................. ................. G.. |
| H-18      | ................. ................. ................. ................. G.. |
| H-19      | ................. ................. ................. ................. G.. |
| H-20      | ................. ................. ................. ................. G.. |

| H-1       | TATTTGAAAA AAA--CTAGA AGATAATAAT CTCCCAAACC CTTACAAAGG |
| H-2       | ................. ................. ................. ................. G. ................. |
| H-3       | ................. ................. ................. ................. G. ................. |
| H-4       | ................. ................. ................. ................. G. ................. |
| H-5       | ................. ................. ................. ................. G. ................. |
| H-6       | ................. ................. ................. ................. G. ................. |
| H-7       | ................. ................. ................. ................. G. ................. |
| H-8       | ................. ................. ................. ................. G. ................. |
| H-9       | ................. ................. ................. ................. G. ................. |
| H-10      | ................. ................. ................. ................. G. ................. |
| H-11      | ................. ................. ................. ................. G. ................. |
| H-12      | ................. ................. ................. ................. G. ................. |
| H-13      | ................. ................. ................. ................. G. ................. |
| H-14      | ................. ................. ................. ................. G. ................. |
| H-15      | ................. ................. ................. ................. G. ................. |
| H-16      | ................. ................. ................. ................. G. ................. |
| H-17      | ................. ................. ................. ................. G. ................. |
| H-18      | ................. ................. ................. ................. G. ................. |
| H-19      | ................. ................. ................. ................. G. ................. |
| H-20      | ................. ................. ................. ................. G. ................. |

| H-1       | TATTTGAAAA AAA--CTAGA AGATAATAAT CTCCCAAACC CTTACAAAGG |
H-1       GTTAAAGTAA AGAAAAACTT ATGTAAGAA AAGAGCACTC
H-3       .......... .......... .......... ..........
H-4       .......... .......... .......... ..........
H-5       .......... .......... .......... ..........
H-6       .......... .......... .......... ..........
H-7       .......... .......... .......... .......C...
H-8       .......... .......... .......... .......C...
H-10      .......... .......... .......... ..........
H-12      .......... .......... .......... ..........
H-16      .......... .......... .......... ..........
H-17      .......... .......... .......... ..........

Appendix 2: Progeny field trial design (source: AFFS Forestry Research internal report)

4 trees per family
12 families per plot