Spatial and temporal patterns of population genetic diversity in the fynbos plant, *Leucadendron salignum*, in the Cape Floral Region of South Africa

by

Erica Tassone

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved April 2013 by the
Graduate Supervisory Committee:

Brian Verrelli, Co-Chair
Thomas Dowling, Co-Chair
Reed Cartwright
Michael Rosenberg
Martin Wojciechowski

ARIZONA STATE UNIVERSITY

May 2013

ABSTRACT

The Cape Floral Region (CFR) in southwestern South Africa is one of the most diverse in the world, with >9,000 plant species, 70% of which are endemic, in an area of only ~90,000 km^2. Many have suggested that the CFR's heterogeneous environment, with respect to landscape gradients, vegetation, rainfall, elevation, and soil fertility, is responsible for the origin and maintenance of this biodiversity. While studies have struggled to link species diversity with these features, no study has attempted to associate patterns of gene flow with environmental data to determine how CFR biodiversity evolves on different scales. Here, a molecular population genetic data is presented for a widespread CFR plant, *Leucadendron salignum*, across 51 locations with 5-kb of chloroplast (cpDNA) and 6-kb of unlinked nuclear (nuDNA) DNA sequences in a dataset of 305 individuals. In the cpDNA dataset, significant genetic structure was found to vary on temporal and spatial scales, separating Western and Eastern Capes - the latter of which appears to be recently derived from the former - with the highest diversity in the heart of the CFR in a central region. A second study applied a statistical model using vegetation and soil composition and found fine-scale genetic divergence is better explained by this landscape resistance model than a geographic distance model. Finally, a third analysis contrasted cpDNA and nuDNA datasets, and revealed very little geographic structure in the latter, suggesting that seed and pollen dispersal can have different evolutionary genetic histories of gene flow on even small CFR scales. These three studies together caution that different genomic markers need to be considered when modeling the geographic and temporal origin of CFR groups. From a greater perspective, the results here are consistent with the hypothesis that landscape heterogeneity is one driving

influence in limiting gene flow across the CFR that can lead to species diversity on fine-scales. Nonetheless, while this pattern may be true of the widespread *L. salignum*, the extension of this approach is now warranted for other CFR species with varying ranges and dispersal mechanisms to determine how universal these patterns of landscape genetic diversity are.

DEDICATION

To my parents, Carol and Sam, who have always fostered my curiosity. Your love and continuous support has meant the world to me.

ACKNOWLEDGMENTS

There are number of people whom I have to thank for helping me get to where I am today. I would like to thank the members of my committee: Brian Verrelli, Thomas Dowling, Michael Rosenberg, Martin Wojciechowski, and Reed Cartwright. I especially thank my mentor and advisor, Brian Verrelli, for the time and energy he has dedicated to me throughout my graduate career and for helping me become an inquisitive scientist. He has always encouraged me to try new analyses or approaches and helped me whenever I struggled. A special thank you to Michael Rosenberg for developing numerous programs to help with my data analysis, his help and programming skills has saved me countless hours of data formatting and processing. I would also like to thank Thomas Dowling for his advice, help and mentoring, especially during my final semester. This work could not have been completed without the input and collaboration of our South African colleagues, Richard Cowling, Jan Vlok, and Karen Esler; they provided samples, knowledge, and expert advice throughout this project. Thank you to the past and present members of the Verrelli Lab who have not only helped me to collect data but often helped me to collect my thoughts. Thank you to my fellow graduate student friends and colleagues, you have provided me with ideas and insight throughout my graduate career. I would also like to thank my family, who has always been my cheerleaders, full of encouragement, and love throughout my life as well as throughout my graduate career, and a special thank you for Christian Snider, who has always been supportive and loving, and always quick to respond to my frustration with encouraging words. Lastly, thank you to all of my friends who been a constant source of much needed distractions.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Chapter 1

APPLYING POPULATION GENETICS TO EXPLAIN DIVERSITY WITHIN THE CAPE FLORAL REGION: A REVIEW OF PRIOR WORK

Introduction

A recent panel convened by the United Nations (January 2012) has recognized a biodiversity crisis, as mass extinctions and increased land use threatens species across the globe. Biodiversity, or the quantity or number of species found within an environment, can be a measurement of the health and productivity of an ecosystem (Hagan & Whitman 2006, Flombaum & Sala 2008). Losses of biodiversity within an ecosystem can have large impacts, reducing the sustainability of an environment through local extinctions and limiting available resources through habitat destruction (Costanza et al. 1997; Daily 1997; Balmford et al. 2002; Singh et al. 2012). Currently, several threats exist to endanger biodiversity across the globe including the human impact of exploiting the environment through land development, over fishing, pollution and the natural threats from invasive species and disease. Understanding what contributes to the generation and maintenance of biodiversity on both spatial and temporal scales is paramount to sustaining biodiversity in the long-term. However, in practice this is difficult given the technical challenges of measuring biodiversity and the theoretical challenges of how to account for all potential mechanisms that may contribute to biodiversity at various scales. Therefore, a theoretical framework to define and assess biodiversity would aid in the development of conservation plans by informing groups about all processes underlying the current biodiversity in a region. The information collected from these studies can then

be used as quantitative measures of biodiversity, providing data of factors that influence biodiversity, which can be accounted for in conservation plans.

One area facing this biodiversity crisis is the Cape Floral Region (CFR), a 'biodiversity hotspot' (Cowling et al. 2003) located in South Africa (Fig. 1.1). It is one of the world's richest botanical areas for its size and a regional area of endemism (Linder & Hardy 2004; Goldblatt 1978), often regarded as the smallest of the six Floral Kingdoms (Linder 2003). It covers less than 4% of the total land area of South Africa (Goldblatt 1997) yet contains over 20% of Africa's floral species. Nearly 9,000 species exist in the CFR, approximately 70% of which are endemic (Linder 2003), making it the most endemic and diverse geographic area excluding all island land masses and yet, all in a region covering only 90,000 km^2. Many studies have tried to understand what contributes to and drives diversity in this region, however, owing to the heterogeneity found in the region on climatic, geographical, topographical, and even temporal scales this has been difficult.

The climate of the CFR has a Mediterranean-like climate with warm, dry summers and mild, wet winters. Species diversity is typical of Mediterranean regions (Linder 2003); however, when comparing species density and endemism per square kilometer of land, the diversity of the CFR far outweighs its Mediterranean counterparts (Linder 2003, Fig. 1.2). The immense diversity found on such a small geographic scale may be in part due to the environmental heterogeneity within the CFR which includes complex vegetation, graduated rainfall regimes, multiple soil types, and changing topography on a spatially fine scale.

The complex vegetation of the CFR includes five distinctive vegetation types: fynbos, karoo steppe, succulent shrubland, forest thicket, and evergreen forest (Goldblatt 1997). Fynbos (meaning "fine bush" in Afrikaans) is endemic to the CFR, and itself is comprised of four fire-prone vegetation types: restiod reed-like bushes, ericoid fine leafed bushes, proteoid large leafed woody shrubs, and geophytes that contain large underground storage organs (Cowling et al. 1996). The high levels of plant endemism occur mostly in the fynbos vegetation that alone covers over half of the CFR and accounts for >80% of its plant species; several groups have hypothesized the unique composition of fynbos alone is what explains the species patterns found here (Goldblatt 1978; Campbell 1983; Linder 2003).

The CFR is naturally divided into two distinct coasts, the West and East coasts, which are referred to as the Western and Eastern Capes. This distinction is relevant because each cape is abutted by a different ocean; on the West is the Atlantic Ocean, and on the East is the Indian Ocean. The two oceans meet near the southern tip at Cape Agulhas, where the cooler Atlantic currents blend with the warmer Indian Ocean currents, creating unique oceanic environments and pressure systems. This blending of oceans also lends to substantial variation between the Capes with respect to climatic patterns. The Western Cape shows a Mediterranean climate pattern of summer rainfall and receives the highest concentration of rainfall near 60” per annum (Cowling et al. 2005). The amount of rainfall decreases in a step-wise fashion moving towards and into the Eastern Cape, where the climate changes to a bi-annual rainfall regime, resulting in unreliable amounts of rainfall, with an average of near 10” per annum (Cowling & Lombard 2005). These distinct climate patterns are thought to have emerged and

remained stable throughout the Pleistocene, allowing for unrestricted access to the Western Cape (Cowling et al. 2009; Valente et al. 2010). Several researchers have suggested that the flora and fauna inhabiting each coast are largely determined by the local conditions (Linder & Vlok 1991; Linder 2003), so these unique climatic patterns may be contributing to the local diversity of each Cape.

The terrain of the CFR changes considerably within its narrow borders; as mountain ranges separate the interior of the country from the coast, leading to a gradient of altitudes that can change from 3 m below sea level to 2,200 m above sea level and all within kilometers of each other (Latimer et al. 2006). Associated with rainfall and topographical instances is the variety of soils found across the CFR. These soils include rocky, nutrient-poor soil to nutritionally imbalanced dune and limestone sands of the coastal margin (Cowling & Proches 2005). Not only do these soils vary in type, but also quality, as the range of fertility and pH varies even on a fine scale. Altogether, this combination of topography, orography, and climatic diversity lends to the CFR areas rich with flora and desolate with sandy deserts; areas which can support the incredible plant diversity that has been well-documented. However, it is this exceptionally complex combination of factors on spatial and potentially temporal scales, which have been difficult to model both independently and collectively (Cowling et al. 2005).

In the development of a theoretical framework that both assesses the biodiversity of the CFR and also determines what factors may be creating or maintaining this biodiversity, a set of questions has been posited. These questions reflect current ongoing research aims in the CFR and include: understanding if overall patterns at regional scales affect local diversity; how these patterns of diversity vary across the geological and

topographical gradients found in the CFR; if the variations of the environment cause these patterns; when these patterns arose; and what evolutionary processes contributed to these patterns of diversity (Linder 2003; Cowling & Proches 2005; Barraclough 2006).

Numerous studies have been performed in the CFR in an effort to answer these questions; many studies have provided insight into how some of these processes have affected diversity of the CFR at some level. Despite these research efforts, many of these questions remain unanswered; therefore, several groups have suggested using population genetic studies as a means to understand population level genetic diversity. These studies can infer spatial and temporal information by generating data reflecting gene flow, population genetic structure and divergence, fine spatial scale variation, while also reflecting historic and recent expansions and bottlenecks in population size that relate information about temporal change on the scale of thousands of years (Perry et al. 2007; Zellmer & Knowles 2009).

Although this temporal and spatial scale analysis has been strongly encouraged for species in the CFR (Barraclough 2006; Hardy 2006; Linder 2006), this approach has not been largely adopted because of several obstacles, including organismal sampling, collection of molecular markers, and appropriate statistical models. In using a population genetic approach, three basic questions can be answered within population based models: (i) what is the genetic structure of gene flow, (ii) what is the relationship among these individuals in time, and (iii) what factors are influencing gene flow. However, in order to implement a population genetic approach to answer these questions, we must consider the organism of study, both the physical and genetic sampling of that organism, and the statistical tests used for data analysis.

Any CFR population can be used to understanding how genetic diversity is distributed in the CFR, but the biology and life history of that organism are necessary to consider when making extrapolating conclusions to the overall CFR. This includes the organisms range (i.e. is it a local endemic or is it widely distributed), reproductive strategy (i.e. can it self-reproduce?), and the ability for movement (i.e. is it a sessile plant or a mobile animal), all of which can influence the underlying genetic diversity within a population. For example, in studies that examine flora, the fact that plants have two mechanisms of dispersal and gene flow via pollen and seed dispersal may play a role in understanding levels of gene flow and spatial genetic structure and must be taken into consideration when implementing a population genetic study.

How an organism is sampled also determines how these questions can be answered as we can only answer questions regarding spatial genetic structure relative to the physical locations of the population sample. Population sampling can take on a number of strategies which include variations on the geographic range and number of individuals sampled. Some examples are: sampling which maximizes the number of locations but not the number of individuals across a broad range (e.g. single individuals sampled from many locations covering the entire CFR), focusing on maximizing individuals in a small range (e.g. sampling many individuals across a 10 x 10 m plot), or one that tries to both include a broad range while collecting a suitable number of individuals at each location to have the power to detect differences between each location using statistical analyses.

The genetic markers used can influence genetic diversity and may limit temporal resolution as each of these markers can reflect different historical or evolutionary

backgrounds given their biological and evolutionary differences. Genetic markers include haploid mitochondrial DNA (mtDNA) and chloroplast DNA (cpDNA), or diploid nuclear DNA (nuDNA), with an understanding that nuclear ploidy can vary depending on the organism examined. The inheritance of each of these markers varies; mtDNA and cpDNA are largely maternally inherited and nuDNA is bi-parentally inherited, therefore nuDNA undergoes recombination whereas mtDNA and cpDNA do not (Pharmawati 2004). Both mtDNA and cpDNA have a smaller effective population sizes because they are haploid and are more subject to genetic drift than nuDNA, they also can reflect different dispersal mechanisms, if female associated dispersal differs from males. The effects of drift in mtDNA and cpDNA may cause fixation of differences between sampled locations, which will appear as a signature of spatial genetic structure. Additionally, the mutation rates vary between these genomes, with nuclear DNA generally having a much slower mutation rate than both mtDNA and cpDNA, allowing for further resolution in time (Wolfe et al.1987; Anderson et al. 2010). These markers have their own advantages and disadvantages as they are each able to detect historical events at different temporal scales and through different lineages, therefore, when choosing which genetic markers to use for a population genetic study, it is important to consider these differences.

Lastly, an abundance of statistical tests are available to analyze data, the importance lies in whether these tests provide the power to answer population genetic questions given the collected data. Thus, in presenting the evidence existing or necessary to address questions regarding gene flow, temporal relatedness, and what may be

influencing gene flow in the CFR, these factors with respect to sampling, genetic markers, and statistical analyses will become underlying themes throughout this review.

Given these considerations for developing an overall framework to understand spatial genetic diversity in the CFR, as an overall hypothesis, a picture that compares and contrasts patterns of spatial genetic structure and gene flow within populations and across many species in the CFR could reveal where, geographically and regionally, specific focal points that are more important or relevant to generating species diversity. Here, previous genetic studies performed in the CFR are examined to see how each has contributed to the overall body of knowledge of understanding gene flow in the CFR and what factors may influence it.

To what extent does population genetic structure exist on different spatial scales in the CFR?

In determining levels of population structure and gene flow at different spatial scales, the heterogeneity of the CFR is an important component to consider and many studies have taken this into consideration. In continuing the theme, here studies are examined based on their sampling, genetic markers, and statistics in order to answer questions directly related to spatial genetic structure and gene flow for CFR populations.

Sampling

Locations

For any organism, sampling strategies for studies conducted in the CFR have taken on three broad strategies: (i) small-scale sampling that collects individuals from local regions located within the CFR, (ii) sampling across the entirety of the CFR, and (iii) sampling both within and outside the CFR, as illustrated in (Fig. 1.3). These

sampling strategies each provide data to determine the spatial genetic structure and gene flow for organisms in the CFR, although these strategies each have benefits and limitations.

Sampling across small geographic regions has provided a wealth of organismal-specific data (Cunningham et al. 2002; Suchentrunk et al. 2009; Prunier & Holsinger 2010; Segarra-Moragues & Ojeda 2010), while also increasing knowledge of within CFR species population structure and gene flow. Prunier & Holsinger (2010) examined four white protea species with local endemic ranges, for example, one species, *Protea mundii* has a small range distributed only along the slopes of the Potberg mountains (Manning 2008). This study measured the spatial genetic structure and patterns of gene flow among these species, identifying eight genetic clusters, almost equal to the number of species examined, and finding little to no gene flow between these species. It is apparent the strategy of sampling on small geographic scales is unavoidable when organisms have locally endemic ranges, and can be useful when asking organismal-specific questions. Here, although populations were sampled for each protea species, the study question focused on the gene flow and genetic structure between species and therefore provided little information about each of these protea populations. It is clear that the spatial genetic structure of organisms can be determined using locally endemic sampling strategies; however, it is difficult to extrapolate this population to what may be driving overall diversity in the CFR.

Several studies have sampled populations that span the heterogeneous environmental processes of the CFR in order to elucidate patterns of spatial genetic structure and gene flow (Lesia et al. 2002; Slingsby & Verboom 2006; Daniels et al.

2007; Price et al. 2007; Swart et al. 2009; Tolley et al. 2009). For example, Price et al. (2007) studied the CFR cicada species *Platypleura stridula* by collecting individuals from 66 sites encompassing the species' broad range, this included locations spanning the from the northwest of the Western Cape to East London in the Eastern Cape. They found evidence for spatial genetic structure occurring in eastern, central, and western lineages; with such low overlap between clades they suggested identifying these groups as separate species, indicating little gene flow between these clades. Interestingly, the boundary between the central and eastern clades was centered near the boundary between the Western and Eastern Capes, suggesting this boundary may influence spatial genetic structure in cicadas. The authors also point out that the boundary between the central and western clades occurs in close proximity (10 km apart in two pairs) to one another, indicating spatial genetic structure is occurring on a very fine scale in the CFR. In this case, this widespread sampling strategy elucidated two discreet areas associated with gene flow and spatial genetic structure within the broad CFR and it is evident that these effects from the landscape can be identified using a similar sampling strategy for other organisms.

Lastly, studies that sample both across the entirety of the CFR as well as outside its defined borders allow for extrinsic comparisons of genetic diversity. As Valente et al. (2010) indicates, without these comparisons it is difficult to attribute phenomena of genetic diversity specifically to the CFR; therefore, the authors used a sampling methodology to study the phylogenetic relationships among species in the genus *Protea,* by initiating a formal comparison between related clades both within and outside the CFR. They found that protea species outside of the CFR, sampled throughout Africa,

comprised only a single clade nested within Cape lineages. They also discovered that the increased diversification of the Cape lineages occurred at local scales across the landscape, concluding that understanding the diversity of the Cape requires knowledge for how Cape species diverge and persist at small spatial scales.

To correlate population level processes to overall patterns of genetic diversity in the CFR, a sampling scheme that reflects the entirety of the CFR range is essential. Although organisms with small endemic ranges will provide information regarding local regions, they are not ideal for understanding the questions proposed regarding overall diversity of the CFR; it is nearly impossible to associate the large scale heterogeneity of the CFR to a locally endemic organism. In sampling organisms that span the heterogeneous environmental processes, patterns of spatial genetic structure and gene flow can be attributed to potential barriers associated with the landscape as Price et al. (2007) found associating clade divergence to the Western and Eastern Capes. Additionally the density of sampling across these ranges can aid in detecting patterns of local diversity at smaller spatial scales. Ideally, sampling that allows for a formal comparison between regions found inside and outside the CFR will help in determining if patterns are reflective of the CFR or some other underlying factor.

Individuals

Aside from sample locations, the numbers of individuals sampled at each location can largely influence estimates of spatial genetic structure. When sampling individuals at a location, it is necessary to consider the overall objective and the question being asked. For example, if the objective is to evaluate the spatial genetic structure for a biogeographically widespread CFR organism, the entirety of the range is paramount to

answering the question; that is samples across the range will be a first priority. However, it may also be necessary to have enough statistical power to test fine-scale differences between sampled locations. Without an adequate sampling of both individuals and locations, it becomes difficult to determine spatial genetic structure and gene flow on fine scales. For example, in the Price et al. (2007) study noted above, the authors were able to determine the overall spatial genetic structure of *P. stridula*, however, across their 66 sampled locations only one individual per location was collected. So although they found patterns of spatial genetic structure, this sampling did not allow for testing of levels of genetic variation within any given location and testing where variation was partitioned between locations was severely limited. Similarly, Bergh et al. (2007) examined the spatial genetic variation in the renosterbos plant *Elytropappus rhinocerotis*, by comprehensively sampling across the entire known distribution range, which included locations inside and outside the CFR. In total, 26 locations were sampled, but the number of individuals collected at each location varied from one to ten. The authors note that this strategy met their goal of examining broad-scale patterns of spatial genetic structure; however, they recognize this was at a cost of dense within-population sampling. So although they found high levels of genetic diversity across broad scales (e.g. the "West" vs the "East, and "Northwest" vs "everywhere else"), their sampling strategy did not allow testing within each sampled region to determine how this genetic diversity might be generated at finer scales.

An example where sampling strategy did allow for testing between locations is that of Segarra-Moragues & Ojeda (2010) who examined two populations of the fynbos plant *Erica coccinea* that have different responses for post-fire survival, that of reseeding

and resprouting. They sampled ten locations for each plant type spanning from the Cape Peninsula to just east of the Agulhas Peninsula, for a total of 267 seeder and 235 resprouter individuals, with no less than 15 individuals collected per location. This allowed them not only to determine the spatial genetic structure and portioning of genetic variation between these populations, but also within populations. The authors found evidence for spatial genetic structure between these resprouter and reseeder populations, with evidence of four spatial genetic structure clusters. They were also able to calculate F_{ST} values within and between these life history strategies, finding values were twice as high in seeder populations than in resprouter populations. Additionally, levels of variation within each sampled location showed that seeder populations also had higher genetic diversity within each sampled location compared to the resprouter sampled locations.

Measures of genetic diversity can be made within populations that have a minimum of two sampled individuals because there are enough data to make a comparison, however as individuals are added, the power to detect differences that occur at lower frequencies is increased. With two individuals, this is limited to differences that are found in 50% of the population, with ten individuals, differences found at 10% can be determined. This is important when understanding genetic diversity as it is these differences that provide the data to quantify how and where variation is being partitioned among populations.

Ultimately, how many individuals are sampled reflects what question is being asked. Studies like Price et al. (2007) may not have looked at a large number of individuals within each location, but in determining population genetic structure across

cicadas in the CFR, their sample size was 66 individuals, and across their population they could detect differences. On the other hand, studies that collected more than 10 individuals at each sampled location were able to measure genetic variation both within each location and between locations, enabling an understanding of how genetic variation is partitioned on both broad and fine-scales. It is of note that although the number of individuals sampled dictates the ability to detect the frequency of variants the choice of genetic marker is important as well as the amount of information about genetic variation can differ between different genetic markers.

Genetic Markers

In performing a population genetic study to understand the spatial genetic structure of gene flow, molecular genetic markers provide a means to quantify levels of genetic variation, which allows us to measure levels of gene flow at the genetic level. There are no CFR organisms for which "whole" or even partial genome sequences are available, which makes neutral nuclear locus collection difficult for any population genetic study performed here. Because of this, the studies that have been performed for organisms in the CFR have included bi-parental nuclear intronic sequences, microsatellites, inter simple sequence repeats (ISSR), as well as uniparental data from chloroplast and mitochondrial regions. Each of these genetic markers provides unique insights into the history and relationships between individuals.

Commonly used loci include mitochondrial regions such as *cytb* and control regions (*CRI*) (Gaubert et al. 2004; Kryger et al. 2004; Herron et al. 2005; Daniels et al. 2007; Smit et al. 2007; Dubey et al. 2007; Tolley et al. 2009; Swart et al. 2009; Fritz et al. 2010), nuclear microsatellite or inter-simple sequence repeat (ISSR) markers

(Cunningham et al. 2002; Moodley et al. 2005; Prunier & Holsinger 2010) and intronic gene regions (Ingram et al. 2004; Willows-Munro & Matthee 2009), or a combination of two markers from different genomes, such as microsatellites and mtDNA (Ingram et al. 2004; Willows-Munro & Matthee 2009; Valente 2010) when conducting analyses.

Swart et al. (2009) used two mtDNA regions in a study of *Agama atra* lizards looking at spatial genetic structure between individuals across the CFR, finding congruency between the two mtDNA regions when building population networks. Since mtDNA is maternally inherited and does not undergo recombination, these two loci are presumably linked across the chromosome; therefore, this result was not surprising. The study did find geographical structuring of the *A. atra* lizard across the CFR, encompassed by four groups, that of the Central CFR, Cape Peninsula, Northern CFR and Limietberg. The authors associate the split between the Cape Peninsula clade to the presence of the Cape Fold Mountains and indicate that these clades were not previously identified in a broad scale study; therefore, further fine-scale sampling may reveal additional areas of genetic diversity. While this study provides resolution into the population genetic structure of *A. atra* lizards, it is limited in providing only information for the maternal lineage of these lizards. If female lizards are localized and male lizards disperse, this pattern will largely go unnoticed given this sampling.

To understand gene flow across the CFR it is beneficial to characterize a pattern of genetic diversity that reflects both lineages, this will provide an absolute picture of gene flow across the CFR, and in case there are genetic differences attributed to inheritance patterns of these haploid markers, comparisons between both sets would confirm if there are any discrepancies between marker sets.

In their study of the southern African shrew, Willows-Munro & Matthee (2011) found congruence between their mtDNA and nuDNA markers, with phylogenetic trees generated for both markers confirming relationships among clades, and measures of demographic history both indicated a recent expansion. Owing to their use of both mtDNA and nuDNA markers, the observations of four clades associated to the south, north, east, and west CFR are likely related to the organism and not the unique evolutionary history of the markers used.

Nuclear markers, such as AFLPs and microsatellites are bi-parental neutral markers which show patterns of spatial genetic structure and gene flow for both males and females. These markers have multiple alleles and so fewer loci may be needed to differentiate individuals in populations (Anderson et al. 2010). In a study of a renosterbos plant *Erica rhinocerotis*, Bergh et al. (2007) used three ISSR markers finding that *E. rhinocerotis* has a large amount of shared ISSR variation leading to a lack of population genetic structure. They attributed this observation partially to the high levels of recombination among ISSR loci, which was likely facilitated by high levels of gene-flow among populations, and high outbreeding rates. However, this nuclear marker could also reflect a singular dispersal mechanism (e.g. the dispersal of pollen); therefore, the addition of maternally inherited markers may help decipher if these levels of high gene flow are universal for all genetic markers.

Single nucleotide polymorphisms (SNPs) are bi-parental, recombining loci, which can be sampled from across the nuclear genome, they are largely slowly evolving, which provides historical temporal resolution, but can have difficulty to resolving recent phenomena. These markers are largely underused currently in CFR studies and are

beneficial for a number of reasons. Brumfield et al. (2003) shows that SNPs are often better for elucidating historical demographic patterns than microsatellites, and can be easily compared to patterns from mtDNA and cpDNA because they are on the same mutational scale (substitutions per site). Additionally, they suggest the variation using SNPs is easier to interpret than microsatellites or ISSR sequences, and when sampled blindly, can offer an unbiased description of genomic variation. SNPs have also been shown to outperform microsatellites in analyses of population genetic structure (Liu et al. 2005; Heylar et al. 2011). Previously, given the lack of genomic data available for CFR organisms, SNP data took considerable effort to obtain. In light of recent advances in DNA sequencing technology, these genetic markers can be readily sequenced in non-model organisms using Next Generation Sequencing technology (Davey et al. 2011), specifically through methods such as RAD-SEQ (Willing et al. 2011). In a recent study of a non-CFR organism, the globe artichoke, the use of RAD-SEQ generated a dataset of ~34,000 SNPS and nearly 800 indels (Scaglione et al. 2012) across the entire genome. Additionally, Baxter et al. (2011) used this technique to sequence 24 diamondback moth individuals, generating 3,177 Rad alleles assigned to 31 chromosomes, and Peterson et al. (2012) collected 10,000 SNPs for an emerging model system, the deer mouse.

Overall, each marker type will elicit information in determining the spatial genetic structure of CFR organisms. The extent of which each marker can reveal historical demography and patterns of spatial genetic structure is tied directly to modes of inheritance, effective population size, and mutation rates underlying each marker. Therefore in considering questions of gene flow and spatial genetic structure in the CFR, each of these factors must be considered relative to conclusions and the underlying

question being asked. Ideally, the use of combinations of these marker sets will provide information about overall levels of gene flow for an organism independent of the evolutionary histories associated with each marker type.

Statistical models

In determining patterns of spatial genetic structure and gene flow for CFR organisms, several statistical analyses have been used to measure levels of genetic diversity and how genetic diversity is partitioned within and among groups. Measurements of spatial genetic structure range from individual based models to network building across a range. Each model can provide information regarding relatedness across a sampled population; however, that information is not always consistent between methods. Tests used in the estimate spatial genetic structure and genetic diversity in the CFR largely include analyses of molecular variance (AMOVA) or spatial analyses of molecular variance (SAMOVA), F-tests (e.g. F_{ST}, F_{SC}) and computer programs such as STRUCTURE (Pritchard et al. 2000).

An AMOVA analysis is useful for partitioning genetic variation between groups and is often used as a first step in determining population genetic structure in many CFR studies (Kryger et al. 2004; Gaubert et al. 2005; Bergh et al. 2007; Smit et al. 2007, 2010; Swart et al. 2008; Willows-Munro & Matthee 2011) additionally, the analyses can add a spatial component (SAMOVA) which not only defines populations based on their genetic variation but can also identify genetic barriers between groups. Smit et al. (2007) used a SAMOVA in their study of two species of southern African elephant-shrews sampled from Namibia to the Eastern Cape to determine whether the spatial distribution of haplotypes found within and between the elephant-shrew species were significantly

different from random. They found for one species, *Macroscelides proboscideus*, there were two groups, one of the eastern most localities and the other containing the remainder of the population. However, one drawback when using AMOVA or SAMOVA analysis is the need of *a priori* information to bin individuals for testing.

F-statistics are used to determine the distribution of variance among groups, with the main statistic F_{ST} determining how variation is partitioned among sub-populations as compared to the total population. While this statistic provides quantification for the amount of genetic diversity present in a population, it does not provide information regarding the relatedness of samples. Computation of an F_{ST} requires some assumptions, including the designation of what constitutes a population, which can become problematic when working on a new system where relationships have not been characterized among individuals in the population. This statistical analysis uses pairwise comparisons and only takes into account data from two populations at a time. Additionally, the bounded nature of F_{ST} values (i.e., between 0 and 1) limits quantification. Studies using F_{ST}, or a similar fixation index, as a measurement of spatial genetic structure are common in the CFR (Lesia et al. 2003, Swart et al. 2008, Prunier & Holsinger 2010, Segarra-Moragues & Ojeda 2010, Willows-Munro & Matthee 2011). In a study of the southern African shrew *Myosorex varius*, Willows-Munro & Matthee (2011) calculated F-statistics between populations of *M. varius* and in comparing Southern versus Northern populations, the variation between these groups was calculated as $F_{ST} = 0.72$, with levels of genetic variation between populations within the groups calculated as $F_{SC} = 0.34$. While this measurement provides information about how and where genetic variation is partitioned among these *M. varius* groups, it can be difficult to

interpret these F_{ST} values, especially if little information is known about the underlying genetic diversity present in an organism. A value at or near "1" reflects a large portion of genetic variation is found within groups, and little genetic variation is shared among groups; however, because of the bounded nature of the statistic, if the population contains enough genetic variation then these values will quickly become saturated, especially if initially grouped without understanding the underlying population genetic structure.

These tests are useful in quantifying spatial genetic structure when population groups are known however one drawback for both of these tests is the *a priori* assignment of individuals to the groups being compared. The relationship of an individual with a specific population is generally unknown for CFR populations. Ideally an approach that uses an unbiased individual based assignment test will help in first determining which individuals are grouped together, so they may then be quantified using F-statistical tests.

Analyses have been used in the CFR that determine spatial genetic structure without the *a priori* assignment of individuals. The individual-based program STRUCTURE (Pritchard et al. 2000, Falush et al. 2003, Hubisz et al. 2009) uses a Bayesian algorithm to cluster individual sequences based on the probability of relatedness between samples. Although few studies use individual based clustering methods in the CFR, those that do, tend to use STRUCTURE (Prunier & Holsinger 2010, Segarra-Moragues & Ojeda 2010). Segarra-Moragues & Ojeda (2010) used the program to examine the spatial genetic structure among the seeder and resprouter *Erica coccinea* plant populations, finding evidence for four structure clusters; the seeder populations clustered almost exclusively to two clusters and the resprouter population grouping into

another, the last cluster contained resprouter populations and one seeder population. Without this analysis, the authors may have grouped each resprouter and seeder population into separate groups, however since this preliminary analysis indicated similarities between the two plant types as well, they were able to perform unbiased tests to determine how variation was partitioned within and between these clusters. Here is an example where using an unbiased example can elucidate information about how individuals are clustered regardless of any bias. Programs such as STRUCTURE provide a good first step in determining the spatial genetic structure of gene flow. Some limitations do exist with this analytical approach, including difficulty in detecting hierarchical STRUCTURE (Fogelqvist 2010; Kalinowski 2011), for example, in their study of white proteas, Prunier & Holsinger (2010) used STRUCTURE to determine the spatial genetic clustering among all sampled protea species. They were able to associate individuals to clusters based on species, but they were unable to detect any fine scale differences within protea species, likely due to the large differences between these species. Additionally, this program is unable to quantify levels of variation found within and between assigned clusters.

Understanding the magnitude of genetic differences between groups is important when trying to quantify the levels of gene flow necessary to create these population clusters. Therefore, in marrying this STRUCTURE analyses with those mentioned above an unbiased estimation of the partitioning of genetic variation can be made. Once this is accomplished, gene flow can be measure directly through programs such as MIGRATE (Beerli & Felsenstein 2001) or IMa (Hey & Neilson 2007), which use F_{ST} values in order to calculate rates of migration.

One estimate that may be of use in the CFR is the conditional genetic distance (cGD), a value derived from the topology of a 'population graph' that analyzes data within a graph theoretic framework requiring no *a priori* knowledge of population arrangements (Dyer et al. 2004). Together, Population Graphs and cGD values can show how samples are related to one another across space, identifying complex population genetic structure while also quantifying that variation. Although currently unused in the CFR, this method has been used to determine genetic structure and gene flow in numerous organisms (Dyer et al. 2010; Domingues et al. 2012; Lopez & Barreiro 2013; Kluetsch et al. 2012). Domingues et al. (2012) built a population graph while understanding the tobacco budworm, a pest species, finding evidence for two population groups while also determining the connectedness of locations within those groups. They posit that historical demographic patterns could explain these patterns, although since this was the first examination of population genetic structure in budworms, they could not relate this pattern to any specific ecological feature. The cGD analysis has also been shown to outperform the traditional genetic distance measure of F_{ST} (Dyer et al. 2010) and this would be a useful tool for determining spatial genetic structure in the CFR given the abundance of non-model organisms and dearth of *a priori* population level information.

Progress has been made to evaluate CFR populations using a population genetic approach since the original call for these studies (Barraclough 2006; Linder 2006; Hardy 2006). Through these studies, it is apparent that spatial genetic structure exists across the CFR landscape as many of these groups have identified large patterns of genetic structure attributed to the boundary between the Eastern and Western Capes or the presence of the

Cape Fold Mountains. However, while these studies have been able to determine patterns across broad scales, many lack the sampling resolution to identify patterns at local scales. Therefore understanding the spatial genetic structure of CFR populations and how that structure may have led to the overall diversity in the CFR requires population sampling with enough resolution to test if patterns of spatial genetic structure are occurring on both broad and fine scales. Also, owing to the lack of *a priori* knowledge for many CFR populations, initial individual-based analyses should be prioritized. Altogether, this sampling and statistical methodology will aid in an unbiased assessment of where spatial genetic structure is occurring across the landscape and the magnitude and partitioning of that diversity. This will ultimately describe patterns of gene flow at local scales, and allow tests to determine if these patterns of genetic diversity can be attributed to the heterogeneous environment of the CFR.

To what extent do demographic and temporal models of population genetics explain patterns of diversity within the CFR?

Understanding the temporal relationship among individuals in a population provides a backdrop to associate historical events that may have effected population level variation. These events may be historical population expansions, colonization events, or shifts in the climate that have taken place recently, or in the past. In the CFR, there are a number of known temporal and demographic events that have been postulated to drive diversity in the area. This includes climate changes leading to the establishment of the Benguela current around the Miocene-Pliocene boundary (Goldblatt & Manning 2000), a change in climate leading towards a full glacial period near the Plio-Pleistocene border (Hallam 2004), and the elevation of the Great Escarpment during the Pliocene (Partridge

& Maud 1990). The climate change that occurred near the Plio-Pleistocene border is thought to have established the current climatic conditions of the East and West Capes and is also associated with creating historic areas of refugia in the West and increasing extinction in the East (Cowling & Lombard 2002). Therefore, to associate any species or population events with these known spatial and temporal events, the divergence times within a population must first be estimated. Using the same population genetic approach here as was used for determining the genetic structure of gene flow, molecular genetic data can be analyzed to determine temporal relatedness, however, to test these temporal hypotheses, new considerations for sampling, molecular markers, and statistical models must be made.

Sampling

The considerations for sampling when asking questions of temporal relatedness depend on the underlying question. In the CFR there are a number of hypotheses about the origins of different groups across the landscape. When answering questions of temporal relatedness and attempting to try and estimate dates and origins of events, population sampling will be important. For instance when trying to estimate which region of the CFR may reflect the "oldest" or more ancestral region, a comparative sample across and outside the CFR is necessary. In this case, sampling in areas outside of the Western Cape where the CFR predominates and into the Eastern Cape will serve as an unbiased comparison. Of course, this gets back to arguments presented earlier in that if this question is of priority, then one must focus on an organism that is found within and across both Capes. On the other hand, and again most obvious, if studies are focused on aspects of the organism and the origin and the emergence of that particular organism,

then one will be limited to where the organism is and can be found as well as its natural range. This is simply a difference in priorities, where one of these two strategies can ask questions about age and origins of the CFR and different parts within it as a first objective, and not simply as a by-product of an analysis of which is actually research into a specific organism. Many studies in the CFR have examined the phylogenetic relationships both between species (Barker et al. 2004; Gaubert et al. 2005; Slingsby & Verboom 2006; Willows-Munro & Matthee 2009; Ingram et al. 2004; Fritz et al. 2010; Valente et al. 2010; Pirie et al. 2011; Schnitzler et al. 2011) and within a species (Daniels et al. 2007; Price et al. 2007; Smit et al. 2007; Swart et al. 2008; Willows-Munro & Matthee 2011) finding time estimates associated with their organism, but also those which coincide with known temporal events.

Studies that use species in temporal analyses are likely to detect divergence times between said species; however, these time estimates may reflect broad processes affecting all organisms within the CFR. For example, Gaubert et al. (2005) sampled 50 specimens representing 15 morphological genet species, and was able to estimate divergence times between species of genets in Africa as far back a 4.90 million years ago (MYA) between *Genetta victoriae* and *G.maculata*, with the most recent species split occurring around 0.31 MYA between *G. maculata* and *G. tigrina*. Within the species tree, several divergence points are associated with cyclical climate changes of the Quaternary period, including a climatic cooling around 2.80 MYA that corresponds to the emergence of *G. felina*; a time point that also coincides near the noted climatic shift during the Plio-Pleistocene border. So although this study says little about relationships between individuals within populations, or the micro-evolutionary processes that may have

eventually led to the speciation events described, it does provide additional information for a large scale temporal event that may have effected micro-evolutionary processes in the CFR.

Studies that examine populations can make inferences about divergence times for the overall population while also examining when any clustering groups began to diverge from one another. In Swart et al. (2008) the relationship between southern rock lizard populations was explored by collecting individuals from across the CFR, from locations ranging in the north near Elands Bay extending into the Eastern Cape near Port Elizabeth. These samples were used to produce a phylogenetic tree which dated the oldest of the four detected CFR clades to the northern CFR, which diverged around ~2.5 MYA and the most recent coalescent point between the Cape Peninsula and Central CFR clades, which diverged 0.64 MYA. Indicating these initial diversification events correspond to the climatic change of the Pleistocene. This sampling strategy provided data for temporal comparisons between different geographic regions indicating the spatial genetic structure of these lizards has occurred over time, with new populations differentiating as little as 0.64 MYA. This shows that, over time, these populations have changed and although some changes can be directly associated with known historical events, other recent factors may be driving diversity between these populations.

Although studies using separate species as samples can elicit temporal estimates for CFR events, these are likely to be ancient events, which are informative about processes driving macro-evolutionary events, such as the formation of those species, but do little to answer questions about understanding the temporal relationships among individuals in populations. These estimates of temporal relatedness largely complement

analyses of spatial genetic structure; therefore, to understand the temporal relatedness among individuals, it is best to mimic the ideal sampling scheme that initially determined the spatial genetic relationship among individuals while keeping in mind considerations for the questions being asked.

Genetic Markers

Questions addressing temporal relatedness can also be largely impacted by the molecular markers used in determining these relationships. As stated previously, CFR studies have utilized a number of molecular markers, including microsatellites, nuclear intronic regions, mitochondrial, and chloroplast DNA sequences. However, when choosing a marker to understand temporal relatedness among individuals, the choice of marker should reflect the time scale of interest. Each molecular marker has a unique biological and evolutionary background that can influence temporal results.

Both cpDNA and mtDNA DNA sequences evolve at comparatively slower molecular rates due to their low mutation rates than nuDNA sequences (Wolfe et al. 1987; Anderson et al. 2010) and are able to answer questions regarding historical change; however, these genomes are also more susceptible to drift over time. Temporal patterns discovered when using these markers may be a reflection of the uni-parental inheritance of these chromosomes; therefore, these temporal patterns may change if bi-parentally inherited markers are used to estimate the same relationship (Anderson et al. 2010). Both cpDNA and mtDNA markers have been effective in dating population level temporal relatedness in the CFR (Slingsby & Verboom et al. 2006; Tolley et al. 2006, 2009; Price et al. 2007; Swart et al. 2008, 2009; Fritz et al. 2010; Willows-Munro & Matthee 2011) and are the most widely used markers in most studies. Studies using these markers have

found ancient temporal relatedness, for example, Daniels et al. (2007) dated the divergence between clades of angulate tortoises between 10.4 and 8.4 MYA using mtDNA markers, which corresponds to climate change due to the development of the Benguela Current. In addition, Willows-Munro & Matthee (2011) were able to date the radiation of the southern African shrew around 2 MYA using mtDNA, near the Plio-Pleistocene border, as were Swart et al. (2008) in their examination of cape Agama lizards mentioned previously. Similarly, Price et al. (2007) placed the initial diversification of CFR cicadas to this time period as well, but also suggests that the neotectonic uplift ~2 MYA may also have driven the initial diversification within this group. It seems that from a number of studies, these markers place events in the distant past, with patterns occurring near the Plio-Pleistocene border around 1.5 – 2.5 MYA.

Alternatively, when asking questions that require a contemporary time scale, highly variable nuclear markers such as microsatellite loci, amplified fragment length polymorphism (AFLP), or inter simple sequence repeat (ISSR) sequences best reflect contemporary relationships between individuals (Anderson et al. 2010; Holderegger et al. 2008). However, the higher mutation rate of these markers can cause homoplasy, making associations over long periods of time difficult. Though not abundantly used in studies of the CFR, Prunier & Holsinger (2010) were able to use microsatellites in three species of white protea to estimate more recent divergence between species at 0.16-0.56MYA.

Optimally, for questions regarding ancient process on the order of millions of years, SNP data should be used. Nuclear genomic sequences have a biparental mode of inheritance and they have a slower mutation rate compared to mtDNA, cpDNA, and microsatellites. These SNP data are often able to provide the same or better genetic

resolution as microsatellites (Smouse 2010). In the cape, studies using SNPs generated from nuclear genomic intron sequences are used for building species and population phylogenies (Ingram et al. 2004; Pirie et al. 2011; Willows-Munro & Matthee 2011). Ingram et al. (2004) found in their study of African mole-rats located throughout Africa a divergence time around 38 MYA, and Willows-Munro & Matthee (2011) were able to date divergence within a South African shrew using both mtDNA and SNP data, finding divergence times were 1.75 MYA for mtDNA, and near 2.64 MYA using SNP data. To date no studies in the CFR have used genome wide SNP data in studies of spatial genetic structure or in determining temporal relatedness among populations.

Statistical Models

A number of methods are available to build phylogenetic trees to estimate divergence times between individuals, and each again has benefits and drawbacks, overall it is clear there is no singular method that is 'best'(Rutschmann 2006). Bayesian analyses are a coalescent-based estimation of demographic parameters from genetic data (Drummond et al. 2002; Wilson & Rannala 2003; Rannala & Yang 2003; Drummond & Rambaut 2007) and are commonly used for phylogenetic inferences in studies of the CFR (Ingram et al. 2004; Herron et al. 2005; Daniels et al. 2007; Tolley et al. 2009; Willows-Munro & Matthee 2011), though other methods such as neighbor-joining and maximum likelihood trees have been used (Van Der Walt et al. 2011).

However, one problem when asking questions about temporal relationships in the CFR is the lack of endemic fossil data for both flora and fauna needed to calibrate phylogenetic trees (Goldblatt & Manning 2002; Linder 2003; Cowling et al. 2005; Sauquet et al. 2009; Valente et al. 2009) and calculated mutation rates for endemic flora

and fauna. Some fossil data does exist, as Daniels et al. (2007) were able to use fossil data from *Chersina* tortoise species to date the relationship between clades of *Chersina angulata,* however, most studies lack the fossil data to appropriately calibrate rates of evolution and rely on estimates (Swart et al. 2008; Gaubert et al. 2004; Price et al. 2007). Time estimates for the emergence of *some* cape floral species have been calculated: Sauquet et al. (2009) used multiple calibration points for the Proteoideae to determine the emergence of some Protea genera; these data have successively been used to date the diversification of a number of species in genus *Protea* (Valente et al. 2009). Together, the Sauquet et al. (2009) and Valente et al. (2010) data were used by Prunier & Holsinger (2010), to estimate radiation times in white proteas, where they calculated that differentiation among the 6 species occurred between 0.16 and 0.56 MYA. These estimates have not been calculated for all endemic genera, so time point calibrations can still be difficult to ascertain for CFR species. For example, in the Schnitzler et al. (2011) study, two of the four CFR plant groups studied, *Moraea* and *Babiana,* lacked calibration data, so these points were estimated using recalibrated points from the Iridaceae family tree. As it is highly unlikely fossil data will appear in the CFR due to the poor soil found there, these estimates may be the best method to approximate dates in the CFR. The alternative method, which has been used in place of calibrating rates of evolution with fossil data, is to use relative estimates based on the molecular clock for genetic loci used. This method has drawbacks since many species do not have observed mutation rates, so proxy data are also used here. Some phylogenetic tree building programs do allow for hypothesis testing, such as BEAST (Drummond et al. 2007, 2012), which allows for either constant of variable rate molecular clock models, a choice of tree priors, and

flexibility of prior parameters and may be ideal for this area where little is known about fossil time points and rates of mutations specific to species.

Determining the temporal relationship between individuals associated with CFR populations can be difficult given the lack of fossil data to calibrate roots of trees, however, numerous questions regarding the timing of diversification events can still be answered from a population genetic perspective. Overall, when designing a population genetic study to answer temporal relatedness questions, it is important to consider sampling locations that encompass the question being asked. Molecular marker choice can potentially limit temporal resolution, so considerations must be made for the desired range of temporal resolution. Finally, it is useful to find a program for building phylogenetic trees that allows some flexibility in parameters when necessary, especially when *a priori* and fossil data are limited for many CFR species.

To what extent does the landscape influence patterns of gene flow in the CFR?

The heterogeneity of the CFR has been hypothesized to influence genetic diversity, and while a handful of studies have sought to understand what might be driving differences in the CFR (Cowling & Lombard 200; Daniels et al 2007; Smit et al. 2007; Swart et al 2009; Tolley et al. 2009; Prunier & Holsinger 2010), these forces have been difficult to detect at the population level. As noted above, many studies have attributed large historical climatic shifts in the Plio-Pleistocene to patterns of genetic variation or speciation (Tolley et al. 2006, 2008, 2009; Swart et al. 2009; Willows-Munro & Matthee 2011) while others have identified large geographic barriers, such as the Cape Fold Mountains (Daniels et al. 2007) as driving forces of genetic diversity, yet few have been able to quantify the effects of these barriers. It may be that the fine-scale heterogeneity of

the CFR may also be contributing to the diversity found here (Cowling et al. 1996; Linder 2005; Thuiller et al. 2006), yet this hypothesis has received little attention and empirical testing in the CFR. The cause of the immense floral species diversity in the CFR is still unknown, but by understanding what drives diversity within a population; it is possible to extrapolate this knowledge to learn about speciation and conservation on a broad scale across the CFR as boundaries identified as barriers to gene flow between locations within a population may eventually lead to speciation.

A newly emerging field deemed landscape genetics was coined by Manel et al. (2003) and merges ideas of landscape ecology and population genetics to understand 'the interaction between landscape features and micro-evolutionary processes'. This methodology may be helpful in testing hypotheses correlating landscape heterogeneity with population genetic spatial structure in the CFR. As with the questions of both temporal and spatial genetic relatedness, using a landscape genetic approach has similar requirements, although this approach also requires detailed landscape data. In landscape genetic studies, individuals become the operational unit, which avoids bias in *a priori* identification of populations, and allows for fine-scale sampling (Manel et al. 2003). The selection of appropriate genetic markers and statistical approaches to test correlations also need consideration.

Sampling

Landscape data

Detailed spatial data have been collected for the CFR by Latimer et al. (2006), which includes information for a number of factors such as elevation, temperature, soil fertility, rainfall, and urbanization. Additionally, the South African National Biodiversity

Institute (SANBI) has published a detailed record of the distribution of all vegetation in the CFR (Mucina et al. 2007) allowing correlations between vegetation type and population genetic variation. SANBI has also sponsored the Protea Atlas Project, where amateurs collected data for the locations of all protea species in the CFR, generating data that include GPS coordinates of all known protea species locations (http://protea.worldonline.co.za). These data are necessary and extremely useful when implementing a landscape genetic approach as they provide a number of testable variables to correlate with population genetic data.

Population sampling

To execute a landscape genetic approach the collection of population genetic data is necessary, however, in collecting these data, new considerations for population sampling must be considered. Anderson et al. (2010) suggest sampling should 'suit the ecological and evolutionary processes under consideration, which can include abiotic, biotic, and anthropogenic features. In the CFR, the landscape variables of interest may include vegetation, rainfall, elevation, or soil fertility. Also of consideration is the spatial scale when sampling, including the grain, extent, and resolution necessary for statistical analysis of landscape genetic data with the landscape features under question (Anderson et al. 2010). Given the heterogeneity of the CFR and the variable ranges between organisms, it is beneficial to consider first, the organism and its endemic range, and then the scale at which to sample the population. Therefore, in designing a landscape genetic study for the CFR, we must consider the organism, the range and distribution of the organism of interest, and how to sample within that range.

A review of landscape genetic studies by Storfer et al. (2007) indicated that the majority of landscape genetic studies examined vertebrates (62% of the total) compared to those which examined plants (14.5% of the total). Although these papers ask similar questions relating to the connectivity and barriers between populations sampled, the organismal sampling strategy must differ as vertebrates are mobile, whereas plants are sessile. Holderegger et al. (2010) reviews landscape genetic studies specifically studying plants, and notes one of the largest differences between studies of animals and plants is that plants have two methods of gene flow: first, the dispersal of diploid embryos in seeds, and second, the dispersal of the haploid pollen. Methods of dispersal for both of these structures can be influenced by abiotic factors such as wind or fire in the CFR, but also through insects and animals as well, and must be taken into consideration when analyzing data. Understanding the dispersal and range of any organism will direct sampling for a landscape genetic study, as these are both important when assessing gene flow and potential barriers.

Different strategies have been discussed for sampling individuals who occupy both continually distributed ranges and discontinuously distributed ranges. Organisms with a continuous range present a problem of determining population structure as differences in the population could be caused by genetic structure or the physical distance between individuals; therefore it is suggested to sample at regular intervals across a widespread range to avoid spurious results caused by unsampled individuals in a population (Anderson et al. 2010). An illustration of this sampling is seen in Smit et al. (2007), who collected a comprehensive spatial sampling of southern African shrew individuals across their broad, continuous range. Organisms with discontinuous ranges

have undergone ‘population’ based sampling, where a representative population sample from each occupied location is collected. In all cases, the biological range of species will dictate sampling strategy. It has been noted, for either sampling strategy, to achieve statistical significance sample sizes of 50-100 individuals are minimal for landscape genetic analysis, with an optimum number of individuals in the hundreds (Anderson et al. 2007).

Genetic Markers

Genetic markers provide the level of resolution for determining relatedness among individuals in a population. As mentioned previously, although microsatellites and both mtDNA and cpDNA are appropriate genetic markers, they can vary the temporal and spatial resolution detected, which must be taken into consideration when using these markers. Manel et al. (2003) suggests SNPs, however, Storfer et al. (2010) reported in their review microsatellites as the most commonly used molecular marker, so although SNPs are ideal, in the field they are not widely used. Neutral genome wide SNP data is posited to produce the most reliable data, with large numbers of SNPs preferred since any outlying loci that may have undergone selection can be removed (Manel et al. 2003). Given the Next Generation sequencing technologies available to generate these data mentioned previously, this method of marker collection should be implemented in future landscape genetic studies of CFR organisms.

Statistical Models

The traditional test to explain correlations between observed spatial patterns and spatial genetic structure used in landscape genetics is a simple isolation by distance (IBD) model, where linear physical distances are measured between sampling locations (Slatkin

1987). However, a number of spatial models have been used to characterize the connectivity between samples given their spatial distributions since they may reflect population networks better than a simple IBD model (Manel et al. 2003; Epperson et al. 2010; Fall et al. 2007). In the CFR, a number of studies have used spatial models such as the Delauney triangulation and Gabriel networks (Willows-Munro & Matthee 2011; Tolley et al. 2009) in tests of IBD. The physical or spatial measurement of distance is then correlated to a measurement quantifying the genetic variation found between locations and generally the statistical measurement for assessing genetic variation is a linearized F_{ST} value; however, other measurements such as genetic chord distance and Nei's distance can be used. This relationship between physical distance and genetic distance is then statistically tested using Mantel tests (Storfer et al. 2010).

Hypotheses of IBD have been tested in the CFR numerous times (Bergh et al. 2007; Smit et al. 2007; Prunier & Holsinger 2010), but the results within the cape have been mixed. Prunier & Holsinger (2010) found a significant correlation between physical distance and genetic distance as measured using a linearized F_{ST}, they found IBD in one population of Proteas out of seven independent populations tested. Failure to accept IBD as the mechanism driving diversity leads to testing other variables, such as the biotic and abiotic factors of the landscape. This hypothesis testing has been applied in the CFR when environmental variables are correlated to the patterns of isolation, but not tested empirically, for example, Tolley et al. (2009) generated spatial connectivity data for each of three lizard species using Delauney triangulation and together with estimates of genetic distances over the landscape, created genetic landscapes that indicated higher diversity in the western CFR clades than the eastern CFR clades. Additionally, Tolley et

al. (2009) used these data in conjunction with both present and future climatic scenarios to explain differences here, finding that climate models predict patchy distribution of these lizards, and in the future, these patches may shrink further due to climatic suitability.

While IBD provides an excellent null-model to test against, it has limited power to detect the effects of isolation on fine-scale levels as might be expected in the CFR, if the heterogeneous geography, topography, and climate of the CFR contribute to population isolation. Landscape genetics allows for the identification of landscape or environmental features as barriers to gene flow. To evaluate the impact of these variables, multiple analyses have been posited. One analysis is isolation by resistance (IBR), a term coined to explain how the landscape may resist or facilitate gene flow between regions. This method borrows from circuit theory, treating regions as nodes, gene flow as measured by a linearized F_{ST}, as current, and the landscape as the resistors or conductors that influence how gene flow occurs across the landscape (McRae 2006). The IBR method often generates a higher correlation with genetic distance than IBD (McRae and Beier 2007). The extent of use of IBR in the CFR is limited, but it was implemented by Smit et al. (2007) who reported similar results for both IBR and IBD analyses.

One limitation to both of these studies is the use of F_{ST} as a measure of gene flow between sampled populations. This pairwise test can prove difficult to use when extreme values are calculated for any pair of individuals or populations, even after linearization values approaching 0 or 1 quickly become significant outliers among the data, which effects correlation. As F_{ST} is not a measure of connectivity between groups, therefore, it is difficult to assess how populations contributing to these large values are associated

with the overall population and whether or not they should be removed from the analysis. This information is impossible to determine when using a strict IBD or IBR test.

To address the issue of both large F_{ST} values and connectivity of individuals and locations, applying the cGD statistic (Dyer and Nason 2004) would be useful, as it gives an unbounded assessment of the relationship between locations sampled and has been shown to outperform genetic distances measured by a linearized F_{ST} (Dyer and Nason 2004). Additionally, as 'population graphs' consider the topology of individuals, they can be directly correlated with measurements of resistance or distance. In the CFR implementing analyses of cGD as a measure of genetic distance with the potential resistance surfaces may provide insight into what landscape features are driving diversity.

Moving forward, to understand what is driving diversity in the CFR it is imperative to continue designing studies using a population genetic approach as called for by Barraclough (2006). Ideally, these studies could also be used for landscape genetic analyses, generating information across multiple organisms for the spatial genetic structure of each organism, how these relationships occur in time, and what may be driving the diversity found in each organism. Although study design is largely influenced by the organism, optimizing data collection by following the above recommendations will help in understanding the overall CFR diversity from a fine-scale perspective, which in turn, will allow us to understand how the fine-scale heterogeneity of the CFR influences these organisms. By consistently designing population genetic studies, these data can be combined to determine if a singular force or forces are contributing to the variation of organisms in the CFR.

Conclusion

In building a theoretical framework to define and assess biodiversity in the CFR, a population genetic approach will aid in understanding patterns of biodiversity at both local and broad scales across the CFR. Through both spatial and temporal modeling, the time points associated with diversification within populations can be addressed, and in landscape modeling, the heterogeneous features of the landscape can be examined as factors that contribute to resistance to gene flow. In the future, population genetic studies used to understand how diversity is generated in maintained in the CFR should include (i) a representative organismal sample with a widespread, continuous distribution across the CFR, (ii) a large sample of molecular markers from multiple genomes that reflect neutrality, (iii) a statistical model that identifies spatial genetic structure and temporal relatedness among samples without the input of *a priori* information, and finally (iv) a statistical model that correlates the landscape variables to estimates of genetic diversity.

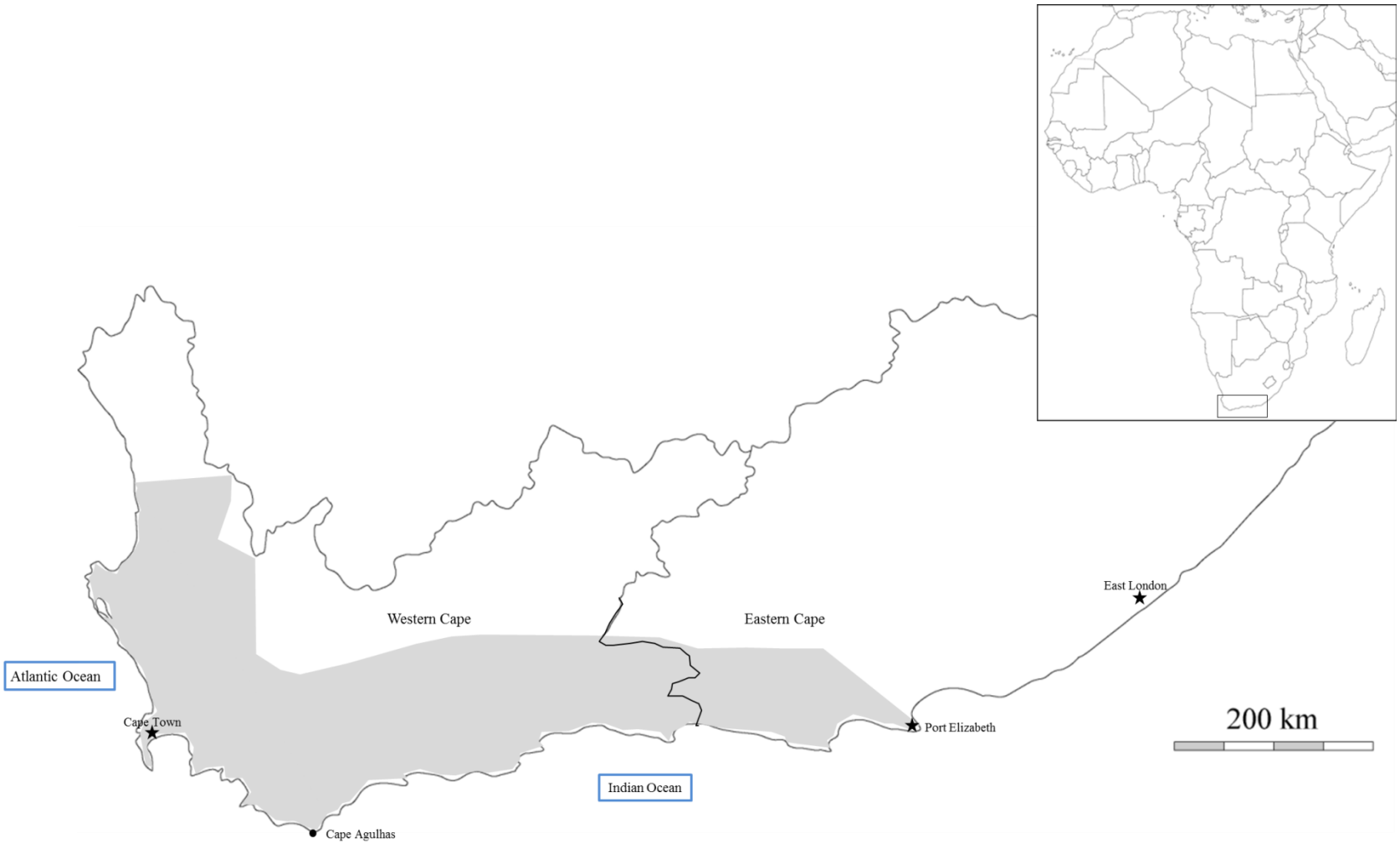

Fig. 1.1 Map of South Africa. The CFR (as defined by Turpie et al. 2003) is shaded in grey. Inset is a map of Africa with study area outlined by a box.

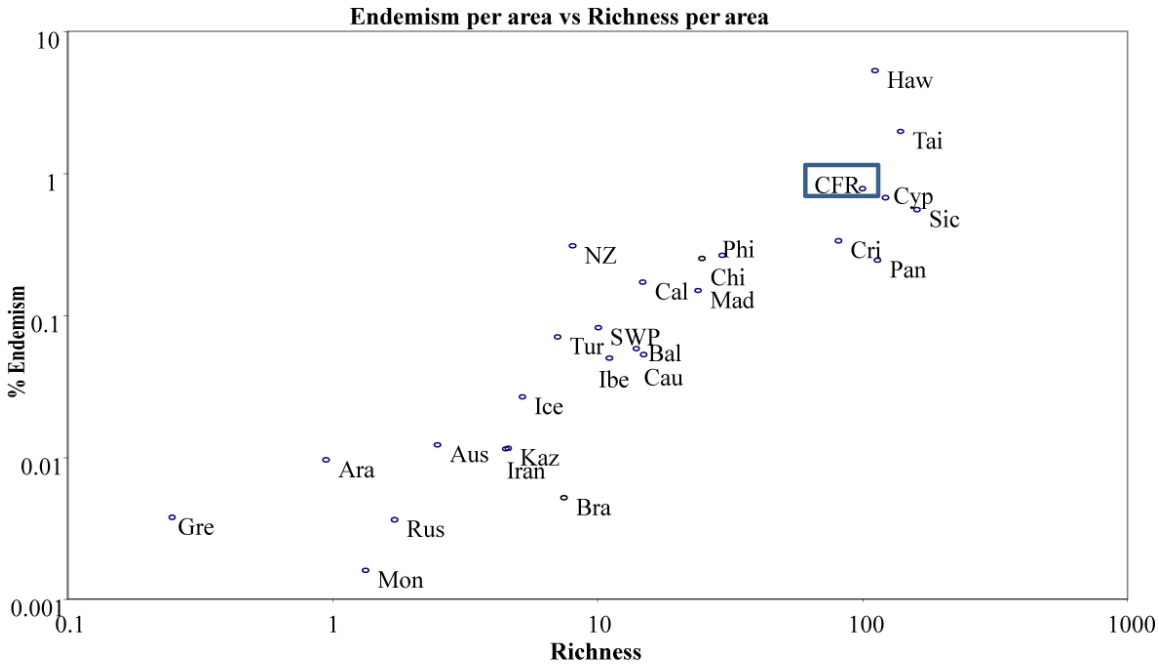

Fig. 1.2 Species percent endemism by area (km^2) compared to species richness per area (km^2) of the Cape Floristic Region and other areas. Axes are logarithmic (log10). Ara, Arabia; Aus, Australia; Bal, Balkans; Cal, California Floristic Province; Cau, Caucasus; Cape Floristic Region, CFR; Chi, Central Chile; Cri, Crimea; Cyp, Cyprus; Gre, Greenland; Haw, Hawaii; Ibe, Iberia; Ice, Iceland;Iran, Iran; Kaz, Kazakstan; Mad, Madagascar; Mon, Mongolia; NZ, New Zealand; Pan, Panama; Phi, Philippines; Rus, Soviet Union; Sic, Sicily; SWP, SW Australia; Tai, Taiwan; Tur, Turkey. Source of the data is Linder (2003).

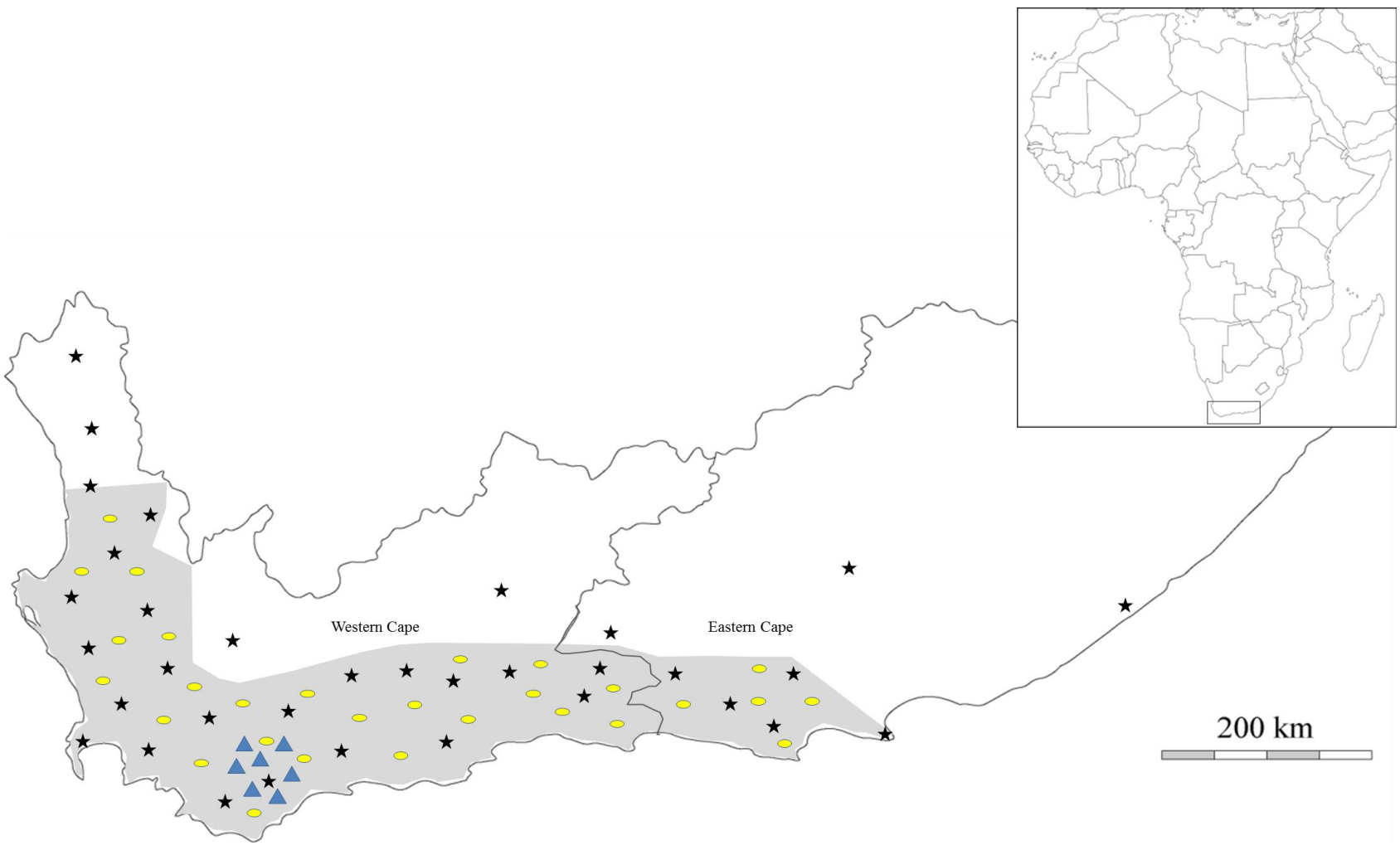

Fig. 1.3 Examples of sampling schemes. Sampling demarcations: a ▲ indicates small scale local sampling, ⬭ indicates sampling within the CFR, and ★ represents sampling both within and outside the CFR. The CFR (as defined by Turpie et al. 2003) is shaded in grey. Insert is map of Africa, with enlarged area delineated by the box.

Chapter 2

A POPULATION GENETIC STUDY OF THE FYNBOS PLANT *LEUCADENDRON SALIGNUM* FROM THE CAPE FLORAL REGION, SOUTH AFRICA

Introduction

Understanding what creates and maintains genetic diversity in geographic regions with high species diversity has been a paramount goal of evolutionary biology as biodiversity is one of the primary drivers of ecosystem health and productivity (Flombaum & Sala 2008). Studies to characterize diversity can take different approaches, including the use of molecular phylogenetics, which examine relationships and rates of diversification among extant species (Valente et al. 2009; Daniels et al. 2007), species accounts in abundance or density (Goldblatt 1997; Ojeda et al. 2001; Cowling et al. 2005, Proches et al. 2006; Manne et al. 2007) and fossil analyses, which look back through time, to measure ancient species diversity (Coetzee & Rogers 1982; Sauquet et al. 2009, Rector & Reed 2010). Although informative, these macro-evolutionary approaches to examine diversity lack insight into the underlying micro-evolutionary processes occurring within populations (Barraclough 2006), which are the starting points during which diversification occurs. Therefore, it is important to understand gene flow and genetic structure within populations as temporal and spatial variation in these factors can influence how species diversity originates.

One example of an area with such high diversity is the Cape Floral Region (CFR) that is located on the Western Cape in South Africa. Specifically, the CFR extends from the north-west near Namibia along the Cape Folded Mountain Belt and terminates near Port Elizabeth in the Eastern Cape (Fig. 2.1). The CFR is a Mediterranean region and

hotspot of biodiversity, containing about 9,000 vascular plant species, of which 70% are endemic, in an area of only ~90,000 km^2 (Linder 2003; Goldblatt & Manning 2002). Many groups have been working to maintain and preserve the diversity found in this region as recent urbanization has begun to threaten these species (Rouget et al. 2003). The predominant vegetation found here includes fynbos (meaning "fine bush" in Afrikaans), which accounts for >80% of the endemic CFR plant species (Cowling & Proches 2005). The fynbos alone is hypothesized to contribute to the unique speciation patterns found in the CFR (Goldblatt 1978; Campbell 1983; Linder 2003). Aside from the abundance of species, the CFR also has complex spatial heterogeneity, which includes vegetation, graduated rainfall regimes, and topography on a spatially fine scale (Latimer et al. 2006; Cowling & Proches 2005). This complex environment rich with species diversity has spawned numerous questions regarding the origins and maintenance of this diversity. One question in particular is to what extent patterns of gene flow and spatial structure contribute to the magnitude of endemic CFR species, and what factors influence these population processes.

Two competing theories have been posited to explain the diversity of the CFR, one of the stability of the environment over time, and conversely, the instability of the environment over time (Linder 2003; Cowling & Proches 2005). It has been hypothesized that a climatic shift in the Pleistocene that is associated with the development of the seasonal winter rainfall regime of the Western Cape and the non-seasonal rainfall in the Eastern Cape was a historical driving factor in creating the diversity found between the two Capes (Cowling & Lombard 2002). This transition zone exists today and is associated with vegetation and topographical differences between the Western and

Eastern Capes, as not only does rainfall vary near this border, but vegetation and topography do as well, as the density of fynbos decreases moving east as thicket and succulent Karoo become the predominant vegetation (Cowling, Proches & Vlok 2005). It is thought that the topography and climate in the West may have created refugia, whereas the East experienced increased extinction (Cowling & Lombard 2002). If the environment were stable, then that may have allowed unrestricted access to the Western Cape (Cowling et al. 2009; Valente et al. 2010); conversely if climatic patterns were unpredictable with regards to rainfall regimes and climatic fluctuations then this may have led to differential selection and an increase in CFR diversity (Linder 2003).

One under-utilized approach for answering these questions of diversity has been through population genetic studies, which has long been suggested for the CFR (Barraclough 2006; Valente et al. 2010), to provide insight into population level processes such as gene flow, population structure and divergence over spatial and temporal scales. Specifically, these studies can provide information about fine-scale genetic variation, as well as reflect historic and recent expansions and bottlenecks in population size, and key temporal changes on the scale of tens to hundreds of thousands of years (Perry et al. 2007; Zellmer and Knowles 2009). Whereas several previous studies have made attempts to examine diversity in the CFR, these have largely come as by-products of vertebrate and invertebrate organism-specific questions about phylogeography, in which case, species ranges, densities, and applicability to questions about the plant diversity on fine-scales have been serious limitations to understanding the dynamics of the CFR itself (Price et al. 2007; Swart et al. 2009; Tolley et al. 2009; Willows-Munro & Matthee 2009). Given the level of heterogeneity of the landscape in

such a small area of the CFR, several criteria should be met to conduct such a population genetic study: (i) a representative plant sample with a widespread biogeographic distribution, (ii) a large sample of molecular markers that reflect neutrality, and (iii) a statistical approach that can identify genetic structure on different scales in an unbiased way. Sampling diversity within and between locations, as well as inside and outside the CFR ranges, allows for formal comparisons between these groups and can answer temporal questions regarding when groups diverged from one another.

In setting out to collect a population genetic dataset to address these questions of how, when, and where diversity has originated in the CFR, here, we focused on the endemic fynbos plant *Leucadendron salignum*, a shrub of the family Proteaceae. *Leucadendron. salignum* is the most biogeographically widespread fynbos plant (Barker et al. 2004), found not only in the CFR, but inhabits almost every area of fynbos in South Africa from the Northern Province across both the Eastern and Western Capes, covering mountain and coastal areas, as well as all rainfall regimes. It is a diploid, dioecious plant whose chloroplast DNA is maternally inherited (Pharmawati et al. 2004). *Leucadendron salignum* is insect pollinated, and has evolved a re-sprouting strategy as a fire-survival mechanism; however it also uses serotiny as a mechanism for seed dispersal, likely triggered by fire (Williams 1972; Hattingh & Giliomee 1989; Barker et al. 2004). In this respect, this fynbos plant is an ideal model for a much-needed population genetic approach in the CFR to generate hypotheses about how genetic diversity is found across temporal and spatial scales. Specifically, our analysis includes sampled locations from across the entire species range of Western and Eastern Capes, develops multiple DNA

sequence markers de novo, and applies several statistical and evolutionary tests of geographic and genetic structure.

Methods

Samples

Given our understanding of genetic diversity at any level with respect to geography and genomic regions is poorly understood for *L. salignum*, our sampling rationale intended to maximize information over the entire range while still estimating some level of within-location diversity. Specifically, we do not know on what spatial scale diversity is partitioned, and thus, our intention was to not oversample from any one locale/spot only to find that diversity is partitioned more between locales. This approach was especially important since our primary goal was to determine, initially, whether we can identify large-scale structure and geographic regions that may then require further sampling and resource focus on fine-scale. In this respect, we sampled leaves from a total of 305 *Leucadendron salignum* individuals from 51 locations spread over the entire geographic range of the species (Fig. 2.1, Table 2.1). This sampling reflects a number of different aspects of the landscape including western and eastern rainfall regimes of the different Capes, mountain and coastal fynbos regions, as well as urban and undisturbed sites. Leaves were preserved immediately in silica gel and stored at room temperature. DNA was extracted using the QIAGEN DNeasy Plant Mini Kit (QIAGEN, Inc., Valencia, CA) according to the manufacturer's instructions.

Data generation

As no genome sequences are available for *L. salignum*, or any fynbos plant for that matter, we chose to develop markers for DNA sequence analyses. While studies have

previously used regions such as ITS for resolving phylogenetic relationships among genera and species, these markers by definition are best for examining more distant evolutionary events due to their more slowly-evolving nature, and thus, we needed markers that best capture population-level processes. On the other hand, these conserved regions can be useful for generating primers. Specifically, we took advantage of one full genome sequence published, that of *Arabidopsis thaliana,* and designed many degenerate primers in chloroplast (cpDNA) regions typically conserved across taxa (Chaw et al. 2005; Hu et al. 2000; Raspe et al. 2000; Birky 1995) to amplify across intergenic sequences via PCR for an initial dataset of markers. A set of individuals were first used to test many primer pairs and evaluate DNA sequences to omit those with complex repeats and otherwise unreadable data.

In total, ~5000 bp of cpDNA sequence across 6 different intergenic markers were collected from each of the 306 individuals. PCR products ranged from sizes of ~500 to ~1500 bp and were prepared for DNA sequencing using shrimp alkaline phosphate and exonuclease I (Us Biochemicals, Cleveland, OH). Nucleotide sequences were generated on an Applied Biosystems 3720 capillary sequencer, and trace files were edited and aligned in Sequencher v. 4.6 (Gene Codes). All primer pairs and PCR conditions are available upon request.

Population Structure and Gene flow analyses

Unless otherwise noted, all summary statistics and parameters were calculated in DNAsp v 5.5 (Rozas et al. 2003). Estimates of nucleotide diversity (π) and the number of polymorphic sites (S) following Watterson's (1979) θ were estimated across all individuals as well as within geographic regions or genetic clusters as noted below. As

there is no *a priori* information on what dictates "neutrality" in the *L. salignum* genome, we generated a single nucleotide polymorphism (SNP) frequency spectrum and evaluated its departure from that expected under a standard coalescent model using Tajima's (1989) D. Values significantly greater and less than one may be considered indicative of certain demographic events such as population structure (excess of common alleles) and expansion (excess of rare alleles), respectively.

The program STRUCTURE v 2.3.3 (Pritchard et al. 2000; Falush et al. 2003; Falush et al. 2007; Hubisz et al. 2009) was used to determine the underlying genetic relationship among *L. salignum* individuals. The analyses implemented in STRUCTURE present an unbiased picture of how diversity is distributed across the sampled localities, using an individual-based clustering approach. All individuals were assigned as separate entities and run using the assumption of K = 2-20 clusters; with 5 replicates of 10^6 generations and a burn-in of 10^4 generations. The program Structure Harvester (Earl 2009) was used to interpret the data using the Evanno method (Evanno et al. 2005) and produced the files necessary for the program CLUMPP 1.1.2 (Jakobsson & Rosenberg 2007), which combined results from the 5 runs at the estimated K value. Clusters were visualized using the program DISTRUCT 1.1 (Rosenberg 2004). In addition, to assess how variance was partitioned among different geographic groups or genetic clusters as defined by STRUCTURE, F_{ST} statistics and their associated p-values (testing standard hypothesis of $F_{ST}=0$) were calculated using ARLEQUIN (Excoffier et al. 2005).

A coalescent approach was also used to calculate the effective population size (N_e) and asymmetrical migration rates between different geographic regions and genetic cluster groups. The program MIGRATE-N 3.2.16 (Beerli and Felsenstein 2001; Beerli

2002) employs a Markov Chain Monte Carlo (MCMC) sampling of gene trees to estimate N_e and migration rates among groups. MIGRATE-N was run under the maximum likelihood framework, with the datatype set to nucleotide polymorphism; each run consisted of 10 short chains and 3 long chains with a burn-in period of 10,000 trees.

Divergence Time Estimates

Due to the difficulty of obtaining divergence times for South African plants in the fossil record, we employed the approach of Valente et al. (2009), which used estimated divergence times from fossil outgroup data in an iterative process to constrain the molecular clocks in a series of phylogenetic tree analyses. This analysis first required estimating the divergence times of species within the *Leucadendron* genus, which we did using a relaxed Bayesian MCMC approach implemented in BEAST v.1.7.5 (Drummond et al. 2007, 2012). ITS nucleotide sequence data from Barker et al. (2004) for 14 *Leucadendron* species including *L. salignum* and one outgroup, *Serruria adscendens* (Table 2.2), for which phylogenetic relationships were estimated, but no divergence times were generated at the time, were downloaded from Genbank and analyzed in BEAUTi v.1.6.2 (Drummond et al. 2007, 2012). In BEAUTi a speciation model following a Yule process was chosen as the tree prior with the starting tree randomly generated, branch rates were set as uncorrelated lognormal, with rates estimated among branches. This tree reflected that first published in Barker et al. (2004), and thus, the species relationships were unchanged. Now, using the stem clade of *Leucadendron* estimated at 29.3-46.2 MYA in Sauquet et al. (2009), we constrained the root node age in this tree to a normal distribution for three separate values across this time range. Five independent runs of 5 million generations, sampling every 2000 generations were performed in BEAST for

each root value and the effective sample size (ESS) was assessed in Tracer v. 1.5 (Rambaut & Drummond 2007, 2012). All runs were combined using LogCombiner and trees were annotated using TreeAnnotator v1.7.5 (Drummond & Raumbaut 2007) and visualized using FigTree v.1.3.1. The divergence time of *L. salignum* from closely-related *Leucadendron* species in these runs was conservatively estimated at 8 to 22 MYA.

Based on the Barker et al. (2004) analysis, *Leucadendron* species that are estimated to be closely related to *L. salignum* are very rare or endangered, and those potentially available to calibrate rates of evolution would appear to be highly diverged based on our new estimates here. Thus, in order to calibrate rates of evolution on our population tree, we collected samples of a *Leucadendron* species from the Cape Peninsula, which while unidentified, closely resemble *L. salignum* but are sufficiently genetically diverged. Specifically, although our cpDNA markers amplified in all individuals from multiple *Leucadendron spp.*, they exhibit no shared variants with *L. salignum* but many fixed differences, with an estimated between species nucleotide divergence of 2.88% compared to that of 0.245% estimated between any two *L. salignum* individuals. In addition, our examination of genome-wide nuclear markers in *L. salignum* (see Chapter 4) could not be amplified in any of the *Leucadendron spp.* individuals, indicating that while the nuclear DNA has become sufficiently diverged, the cpDNA is still relatively conserved and appropriate for phylogenetic rooting here. Thus, as explained below, we calibrated rates of evolution for our cpDNA *L. salignum* population tree with these outgroup sequences and constrained the root time with the estimated divergence time from above.

To evaluate the *L. salignum* population genetic data in BEAST, first, MODELTEST (Posada & Crandall, 1998) was used to determine the appropriate model of nucleotide evolution for the *L. salignum* SNP dataset. We used standard published estimates of the chloroplast sequence divergence range of 1.0-3.0 X 10^{-9} substitutions per site per year (Wolfe et al. 1987) in multiple runs. Independent runs for trees with roots constrained to a normal distribution across the range of 8-20 MYA were conducted. Markov Chain Monte Carlo (MCMC) analyses were performed using both strict and relaxed (uncorrelated lognormal) clocks as well as using the constant size, exponential growth, and Bayesian skyline population models. Rate variation was allowed under all models by setting a normal prior distribution for the molecular clock rate (mean = 0.001, SD = 0.00001). Each BEAST profile ran five times for 60 million generations with the first 6,000,000 discarded as burn-in, logging every 6000. All initial runs were viewed in the program TRACER 1.5 to analyze the parameter distributions estimated from BEAST and to check for convergence of the chains. Log files from each set of runs were then combined using LogCombiner and TreeAnnotator v1.6.1 which yielded a consensus tree for each set of analyses. Finally, as a comparison to these results regarding the relationships among individuals, a simple neighbor-joining tree, which employs fewer assumptions, was created in MEGA v. 5 (Tamura et al. 2011).

Results

Nucleotide Diversity

A dataset of ~5000 bp of cpDNA sequence totaling 104 variable sites that included 94 SNPs and 10 indels was collected for each of the *L. salignum* individuals. Overall nucleotide diversity was measured as $\theta_\pi = 0.0025$ and $\theta_s = 0.0030$, with Tajima's

(1989) D = -0.908 (Table 2.3). This latter analysis indicates no significant departures from that expected under neutrality, and thus, the dataset may be expected to fit a drift-mutation model. Analyses performed between geographical groups show some differences from the overall statistics; for example, the Eastern Cape had much lower levels of overall nucleotide diversity (θ_π = 0.0011, θ_s = 0.0013, D = -0.378). On the other hand, the Western Cape had levels of diversity similar to that found in the overall sample (θ_π = 0.0019, θ_s = 0.0029, D = -1.090).

Population Genetic Structure

The ΔK method (Evanno et al. 2005) implemented in Harvester showed a single peak at K=4 (average LnPD= -3116.3) with no secondary peak at larger K values, thus, it would be appropriately conservative to conclude that, based on our data and sample locations, we can detect four structure clusters. Thus, results are presented here for all K runs between 2 and 6 to illustrate the structure in these data (Fig. 2.2). Of note is that some clustering patterns emerged early in the runs that did not change significantly at higher K values; for example, samples collected from the Eastern cape clustered together beginning at K=2, and remained clustered throughout all K values. The four STRUCTURE clusters correlate well with distinct geographic areas of the CFR (north, southwest, central, and east), thus, signaling a geographic distance-based clustering. In addition, in cases where individuals could not be "binned" entirely into one structure cluster, they appear to be located on the edge of cluster "breaks" (Fig. 2.3).

Interestingly, the results for K=5-6 indicate further structure in the north and central regions (Fig. 2.2). These clusters are not fully resolved; however, the rate of increase in the relationship between K and ln(K) (Fig. 2.4) continues through K = 7, at

which point the variance of ln(K) becomes much larger between runs, suggesting there is already some evidence for significant structure even within these cluster groups.

To test this hypothesis of hierarchical structure, additional STRUCTURE analyses were run on each of the original clusters using the same parameters. Within the eastern cluster, which originally remained undifferentiated between all K values, STRUCTURE identified K=3 in the secondary analysis. In the subsequent *post hoc* analyses of the original clusters, STRUCTURE identified K=3 for the central, east and southwest clusters, as well. In order to explore this hierarchical pattern further and objectively, we examined the amount of variance at several levels from within and between locations to within and between the geographic cluster groups. The global F_{ST} among all individuals was equal to 0.78 ($p < 0.001$), with pairwise F_{ST} analyses spanning the entire range (0.1-1.0, Table 2.4). Fst analyses of the genetic cluster groups indicate there is less variation present between the East and West ($F_{ST} = 0.43$, $p<0.001$), and within the East ($F_{ST} = 0.50$, $p<0.001$), than in the overall sample (Table 2.5). However, within the central cluster ($F_{ST} = 0.73$, $p<0.001$), variation is similar to that found in the overall sample. Migration rates estimated using MIGRATE-N indicate patterns of asymmetrical migration across large geographical regions, such as the between the Western and Eastern Cape, with <1 migrant moving from West to East, and >1 migrant travelling from East to West, with possibly limited migration among STRUCTURE groups (e.g., <1 migrant per generation, Table 2.6). Finally, estimates of N_e from MIGRATE-N for all STRUCTURE clusters show lower values for the East than all other regions (Table 2.6).

Temporal Genetic Structure

The results of the BEAST analysis (Fig. 2.5) corresponds to the four STRUCTURE clusters originally detected, which is not unusual given that cpDNA may be expected to reflect lineage sorting due to its non-recombining nature. The topology of the BEAST tree shows three key findings (Fig. 2.5). First, the coalescent point for all CFR locations ("A", Fig. 2.5) indicates a deep divergence between groups with high support (posterior probability = 1.0) approximately ~1.17 MYA (with an estimated 95% CI of 0.73-1.63 MYA). Second, the tree shows the Eastern Cape locations form a monophyletic clade ("B") and shares a common ancestor with locations only sampled in the south west, indicating more recent emergence from this geographic area. The age estimate at ~ 0.321 MYA (0.17-0.47 MYA) of the eastern clade is also significantly more recent than the other clades. Finally, there appears to be a much deeper ancestry for the remaining central and south west locations ("C") ~ 0.885 MYA (0.518-1.26 MYA), indicating that although there are deep roots for many of the locations, the population structure here is not static as new groups have emerged recently. Finally, the neighbor joining tree analysis resulted in tree topologies similar to that found with our BEAST analysis, further indicating the robustness of these results when using a number of different parameters.

Discussion

This analysis of 306 *L. salignum* individuals across 51 locations is one of the most comprehensive studies to examine population genetic estimates of diversity and structure within the CFR. By comparing locations both inside and outside the CFR, we were able to uncover significant amounts of spatial genetic structure that appears to be driven by the partitioning of genetic diversity at fine scales. Additionally, we were able to identify both

spatial and temporal differences for samples collected in the Eastern Cape compared to those in the Western Cape. Interestingly, we also found evidence that the creation of this spatial genetic structure, both inside and outside the CFR, has occurred across different time periods, indicating that the landscape of the CFR has been dynamic over time.

Temporal differences within L. salignum

Our phylogenetic analysis confirmed the initial clustering detected in STRUCTURE and suggests these broad groups have not been static; as the emergence of each clade varies across both time and space, with the oldest clade located in the central region and the newest clade in the East. Previous studies have indicated that the Eastern Cape is influenced by different climate controls than the West (Cowling & Lombard 2002; Linder 2003; Daniels et al. 2007; Tolley et al. 2009), and these differences may contribute to diversity on both spatial and temporal scales. It has been hypothesized that either a historic climate shift or the current climatic regime near the border between the Western and Eastern Capes influences species near the East/West Cape boundary. The estimated relative divergence time between all *L. salignum* groups falls around ~1.17 MYA, well within the estimated Plio-Pleistocene climate range, indicating that this historic climatic shift may have influenced the initial divergence of *L. salignum*. The oldest divergence point within *L. salignum* between groups is the coalescent point of the Southwest and Central groups, who shared a common ancestor ~0.885 MYA. This time period falls near the climatic upheaval associated with the Plio-Pleistocene, which was thought to create areas of refugia across the Western Cape. It may be, during this time, these two groups became isolated from one another, ultimately generating this diversity. However, one of the most surprising results of this study is the origin of the monophyletic

clade representing the East, which appears to have originated from the West more recently than any other clades ~0.321 MYA, a time period that is not associated with historical Plio-Pleistocene climatic shifts, but may be associated with the current climate conditions.

This observation is consistent with our observations of genetic diversity; when comparing the values for Tajima's D for all locations in the Eastern Cape (D = -0.38) to the value calculated for all locations found in the Western Cape (D = -1.09), the value is much larger in the East than the West. It is likely, given the low cpDNA mutation rate (Wolfe et al. 1987), the pattern observed for the East is a result of the recent colonization from a subset of 'founding' individuals, and these newly established groups have yet to acquire enough new rare variation to produce a signature of a recent expansion, in the form of a highly negative D value, and are still largely reflecting the colonizing individuals genotypes. Analyses from MIGRATE-N also corroborate the expansion hypothesis, as the effective population sizes of the East are much smaller than any other group, which is expected after a recent expansion. Additionally, gene flow can be detected at broad spatial scales outside defined structure clusters between locations in the north and east, with >1 effective migrants detected moving from the north to the east. This may be indicative of a signature of historical gene flow that is still detectable given the recent colonization and shared genetic variation.

Interestingly, this recent colonization of the East suggests that the initial historic climatic shifts from the Plio-Pleistocene epoch that are attributed to the differences in vegetation across the East and West Capes did not facilitate the move of fynbos into the Eastern Cape. This observed pattern is perhaps in part due to the stability of the

environment over time since the climatic upheavals, which has maintained the seasonal winter rainfall and bi-annual rainfall patterns found between the Western and Eastern Capes. Several groups have hypothesized these current climatic differences may influence diversity between the Capes (Cowling & Lombard 2002; Linder 2003), and these data of a recent expansion seem to support this hypothesis; however, the exact underlying mechanism that facilitated this recent expansion into the East is still unknown.

Spatial genetic structure in L. salignum

Considering the unique plant endemism found in the CFR, it is surprising that little is known about the genetic variation and micro-evolutionary processes occurring within plant species found in this region. The results presented here show that within *L. salignum* there are high levels of spatial genetic structure occurring on a fine-scale; however, many of these patterns were not detected until *post hoc* analyses were applied. Initially, these *L. salignum* data group into four broad STRUCTRUE clusters, which correspond to geographical regions of the CFR: the north, southwest, central, and east regions (Fig. 2.3). However, *post hoc* STRUCTURE analyses also indicate hierarchical structure occurring on spatially fine-scales; therefore, although four cluster groups largely represents the overall structure, it is not a definitive value, but rather a guide for understanding the broad spatial genetic structure of *L. salignum*.

One limitation of STRUCTURE is that it can be less effective at identifying fine scale variation when there are large differences between groups (Fogelqvist 2010; Kalinowski 2011). In one of the few CFR studies that have used STRUCTURE, Prunier & Holsinger (2010) were able to confirm the genetic distinctiveness of each species; however, they were unable to detect any fine-scale structure within their individual

population samples. While STRUCTURE is a useful tool in determining population structure when no *a priori* information is available, it does not provide the magnitude of differences that separate these groups. For example, in this study, the observed levels of genetic variation, among the initial four STRUCTURE clusters are high, as measured by F_{ST} (Table 2.5), but within some of these clusters, F_{ST} values exceed those values found between the furthest groups of the East and West (Table 2.4). This evidence suggested hierarchical structure within the initial four clusters, confirmed by *post hoc* analyses. The clustering of smaller, localized areas detected in *post hoc* analyses indicated that genetic diversity in *L. salignum* is driven at the fine-scale. One hypothesis to explain these patterns is, if the complex heterogeneity of the CFR drives genetic diversity, we would expect that genetic variation will become partitioned between groups at small spatial scales.

The magnitude of F_{ST} values found in this study suggests that the majority of genetic variation results from differences among groups found within sub-populations and is not shared. This was a surprising result as other studies of fynbos plants have reported low levels of genetic variation even between fynbos species (Bergh et al. 2007; Prunier & Holsinger 2010). However, these studies used bi-parentally inherited nuclear markers, which are difficult to compare to this study using cpDNA, given the differences in dispersal, effective population sizes, and inheritance between these marker sets. While it is difficult to know if these large F_{ST} values are specific to *L. salignum* or indicative of CFR plant genetic variation, we can compare these values within the overall sample to see that even within this population, the partitioning of genetic variation differs across groups. In the geographic region encompassed by the 'central' STRUCTURE cluster, F_{ST}

values are much higher than those found within any other cluster, indicating that, here, more variation is being partitioned between groups than in any other defined CFR clustered region. These large values may be attributed to the age of this STRUCTURE cluster, as it is one of the oldest clades identified by BEAST, more variation may have accumulated simply due to the temporal differences between this group and all others. Alternatively, this area may have been fragmented during periods of climatic shifts, creating many regions of refugia which were isolated over time, generating higher amounts of diversity in this region compared to others. The locations represented by the east STRUCTURE cluster have the lowest amount of genetic variation which may be characteristic of its recent origins compared to all other groups. When comparing F_{ST} values between structure clusters, values are highest when compared to the East, indicating the variation between the East and all other clusters is not shared. This may be expected because the genetic variation found in all other groups is significantly older and more differentiated than in the Eastern cluster, and therefore, variation is portioned largely within each of these groups simply due to the temporal differences between them.

Lastly, sampling locations of *L. salignum* found both inside and outside the CFR allowed for comparisons in determining where genetic diversity is being generated. Given the range of *L. salignum*, we were able to collect four locations outside of the CFR for preliminary comparisons. Although this is a small proportion of the total number of locations collected, it is of note that in comparing these 'outside' samples to those found within the CFR, these outside samples all showed spatial genetic relatedness to CFR locations in STRUCTURE and common ancestry with CFR locations in BEAST. It does not appear that these locations are generating CFR diversity, but rather they show

origination from within the CFR but are then able to move outside of the defined CFR boundary. The influence of genetic diversity within the CFR is able to also contribute to genetic diversity outside of this region, a fact which may have gone unnoticed without this comparison.

Conclusion

In this molecular population genetic study of *L. salignum*, we find evidence for significant amounts of spatial genetic structure through at least four STRUCTURE clusters, with evidence of hierarchical structure within each of these groups. It appears this spatial variation occurs on a fine-scale, as estimates of genetic diversity show that variation is largely found within sampled locations and is not shared between locations. In dating the temporal origins of these groups we find the emergence of these groups occurred across different times, with the east being the most recently colonized. These temporal differences indicate that the overall population of *L. salignum* is not yet at equilibrium, and that the genetic structure may not be spatially generated. In understanding what drives these patterns of diversity across spatial and temporal scales, we must consider alternate approaches to the traditional measures of isolation by distance generally evoked to explain patterns of diversity. It is apparent that the landscape of the CFR is dynamic on both spatial and temporal scales and without this preliminary understanding of the spatial genetic structure and temporal relatedness among individuals in the *L. salignum* population these patterns may not have been discovered. Moving forward, in determining what drives the diversity present in *L. salignum* it will be important to consider the fine-scale genetic variation and temporal differences between these identified groups, as well as compare the patterns found in this study to those from

bi-parentally inherited genetic markers to determine if these patterns are specifically associated with maternal markers or both lineages.

Table 2.1 Names and geographic coordinates for all sampled *L. salignum* locations

	Abbreviation	Location Name	latitude	longitude
1	VRP	Vanrhyns Pass	-31.37087	19.01587
2	GIP	Gifberg Pass	-31.76927	18.76999
3	PAP	Pakhuis Pass	-32.14442	19.02492
4	MIP	Middelberg Pass	-32.63051	19.15221
5	VEP	Versfeld Pass	-32.84253	18.73164
6	GYP	Gydo Pass	-33.23595	19.33648
7	WOR	Worcester	-33.60255	19.33463
8	GBY	Grotto Bay	-33.52580	18.35388
9	PHI	Philadelphia	-33.71750	18.54403
10	CPT	Cape Town	-33.91698	18.40488
11	SVR	Silvermine	-34.08278	18.41490
12	SIT	Simons Town	-34.20063	18.41121
13	SWB	Smitswinkel Bay	-34.26193	18.46104
14	PRB	Pringle Bay	-34.31126	18.83138
15	STL	Stellenbosch	-33.93238	18.87687
16	FRP	Franschhoek Pass	-33.91495	19.15701
17	GRY	Greyton	-34.03345	19.60752
18	STA	Stanford	-34.41032	19.58887
19	AGU	Agulhas	-34.66967	19.77530
20	DHP	DeHoop	-34.37937	20.52973
21	BTK	Bontebok	-34.04368	20.46947
22	KOP	Kogmanskloof Pass	-33.80377	20.10572
23	TRP	Tradouws Pass	-33.93706	20.71161
24	HID	Heidelberg	-34.01497	20.96603
25	GAP	Garcia Pass	-33.94164	21.20187
26	STB	Still Bay	-34.26279	21.37145
27	ALB	Albertinia	-34.22925	21.59360
28	VLB	Vlees Bay	-34.34412	21.86730
29	MOB	Mossel Bay	-34.16592	22.00785
30	GRB	Grootbrak	-34.06319	22.20193
31	CLP	Cloete Pass	-33.93095	21.76138
32	ROP	Robinsons Pass	-33.86390	22.02835
33	OUP	Outeniqua Pass	-33.88690	22.39971
34	DAS	Daskop	-33.76938	22.65600
35	RBP	Rooiberg Pass	-33.65101	21.63852
36	SEW	Seweweekspoort	-33.38633	21.40805
37	ANB	Anysberg	-33.47264	20.58203
38	SWP	Swartberg Pass	-33.35097	22.04737
39	BLB	Blesberg	-33.40775	22.73231
40	POP	Potjiesberg Pass	-33.70272	23.04364
41	PLB	Plettenberg Bay	-34.01325	23.38842
42	MIS	Misgund	-33.76043	23.48274
43	KOU	Kougaberg	-33.67525	23.50330
44	JOB	Joubertina	-33.82058	23.85400
45	NOT	Nooitgedacht	-33.82692	24.25445
46	AGS	Assegaaibos	-33.93455	24.30351
47	BAV	East Baviaanskloof	-33.63444	24.46917
48	HUM	Humansdorp	-33.95992	24.76670
49	SHB	Stinkhoutberg	-33.81534	24.95069
50	SUP	Suurberg Pass	-33.28263	25.72010
51	GRT	Grahamstown	-33.34028	26.51658

Table 2.2 Summary of species used in generating divergence times

Species	GenBank Accession Numbers
Leucadendron nervosum	AY692171.1
Leucadendron album	AY692167.1
Leucadendron ericifolium	AF508855.1
Leucadendron flexuosum	AY692169.1
Leucadendron salignum	AY692172.1
Leucadendron lanigerum	AY692170.1
Leucadendron discolor	AY692202.1
Leucadendron modestum	AY692221.1
Leucadendron dregei	AY692166.1
Leucadendron singulare	AY692209.1
Leucadendron platyspermum	AY692205.1
Leucadendron rubrum	AY692186.1
Leucadendron argenteum	AY692184.1
Leucadendron osbornei	AY692168.1
Serruria adscendens	AF508823.1

Table 2.3 Summary statistics of nucleotide variation

	N	Length (bp)	S^a	$\theta_S{}^b$	$\theta_\pi{}^c$	D^d
All samples	305	4649	88	0.0030	0.0025	-0.908
East Samples	59	4649	28	0.0013	0.0011	-0.378
West Samples	246	4649	83	0.0029	0.0019	-1.090

[a]Total number of SNPs

[b]Watterson's θ

[c]Average number of pairwise diffrences between sequences

[d]Tajima's D statistic

Table 2.4 Pairwise F_{ST} values for all sampled locations

	VRP	GIP	PAP	MIP	VEP	GYP	CPT	SIT	PRB	STA	FRP	GRY	KOP	TRP	DHP	GAP	STB	ALB	VLB	MOB	GRB	ROP	RBP	OUP	SWP	POP	PLB	MIS	AGS	HUM	BAV	SHB	SUP	GRT	AGU	ANB	BTK	CLP	BLB	DAS	GBY	HID	JOB	KOU	PHI	SEW	SVR	SWB	STL	WOR	NOT
VRP	-																																																					
GIP	0.39	-																																																			
PAP	0.77	0.76	-																																																		
MIP	0.74	0.74	0.15	-																																																
VEP	0.86	0.80	0.59	0.58	-																																														
GYP	0.84	0.77	0.00	0.40	0.64	-																																													
CPT	0.86	0.80	0.57	0.58	0.62	0.61	-																																												
SIT	0.78	0.72	0.79	0.83	0.84	0.82	0.85	-																																										
PRB	0.74	0.75	0.33	0.40	0.68	0.44	0.61	0.81	-																																									
STA	0.85	0.80	0.68	0.74	0.86	0.81	0.87	0.93	0.13	-																																								
FRP	0.85	0.77	0.59	0.74	0.79	0.73	0.81	0.93	0.17	0.00	-																																						
GRY	0.78	0.69	0.27	0.46	0.59	0.34	0.58	0.82	0.07	0.20	0.20	-																																					
KOP	0.87	0.86	0.70	0.80	0.84	0.80	0.85	0.94	0.47	1.00	1.00	0.54	-																																				
TRP	0.90	0.86	0.82	0.80	0.91	0.86	0.89	0.94	0.48	1.00	1.00	0.54	1.00	-																																			
DHP	0.84	0.84	0.54	0.74	0.76	0.70	0.79	0.93	0.20	0.00	0.00	0.20	1.00	1.00	-																																			
GAP	0.85	0.84	0.64	0.75	0.80	0.75	0.82	0.92	0.38	0.80	0.80	0.43	0.91	0.00	0.80	-																																		
STB	0.60	0.61	0.49	0.51	0.58	0.53	0.70	0.51	0.41	0.47	0.47	0.35	0.57	0.57	0.47	0.52	-																																	
ALB	0.86	0.83	0.70	0.71	0.81	0.75	0.97	0.91	0.13	0.00	0.00	0.18	0.86	0.86	0.00	0.67	0.45	-																																
VLB	0.74	0.71	0.45	0.50	0.63	0.54	0.78	0.72	0.03	0.00	0.00	0.06	0.43	0.43	0.00	0.31	0.22	0.00	-																															
MOB	0.89	0.84	0.75	0.74	1.00	0.78	1.00	0.93	0.15	0.00	0.00	0.20	1.00	1.00	0.00	0.80	0.47	0.00	0.00	-																														
GRB	0.89	0.87	0.86	0.84	1.00	1.00	1.00	0.95	0.66	1.00	1.00	0.68	1.00	1.00	1.00	0.84	0.64	0.92	0.60	1.00	-																													
ROP	0.88	0.86	0.82	0.80	1.00	1.00	1.00	0.57	0.51	1.00	1.00	0.54	1.00	1.00	1.00	0.91	0.57	0.86	0.43	1.00	1.00	-																													
RBP	0.85	0.84	0.77	0.76	0.96	0.94	0.96	0.91	0.48	0.80	0.80	0.50	0.89	0.89	0.80	0.80	0.55	0.71	0.40	0.80	0.92	0.89	-																												
OUP	0.85	0.82	0.68	0.70	0.95	0.92	0.95	0.90	0.16	0.20	0.20	0.20	0.80	0.80	0.20	0.63	0.45	0.13	0.04	0.20	0.84	0.60	0.66	-																											
SWP	0.89	0.85	0.81	0.81	0.94	0.92	0.94	0.92	0.67	0.89	0.87	0.68	0.90	0.90	0.87	0.86	0.66	0.83	0.65	0.87	0.91	0.90	0.71	0.76	-																										
POP	0.88	0.86	0.82	0.80	1.00	1.00	0.94	0.57	0.35	1.00	1.00	0.54	1.00	1.00	1.00	0.91	0.57	0.86	0.43	1.00	1.00	0.00	0.73	0.47	0.79	-																									
PLB	0.93	0.86	0.82	0.80	1.00	1.00	1.00	0.57	0.45	1.00	1.00	0.54	1.00	1.00	1.00	0.91	0.57	0.86	0.43	1.00	1.00	0.00	0.89	0.60	0.87	0.20	-																								
MIS	0.69	0.63	0.55	0.53	0.74	0.66	0.75	0.53	0.39	0.51	0.55	0.44	0.63	0.63	0.55	0.59	0.27	0.52	0.35	0.55	0.68	0.63	0.57	0.50	0.65	0.54	0.54	-																						
AGS	0.93	0.70	0.69	0.63	0.81	0.78	0.84	0.51	0.59	0.72	0.76	0.64	0.79	0.79	0.76	0.76	0.43	0.73	0.57	0.76	0.81	0.79	0.79	0.75	0.80	0.75	0.72	0.06	-																				
HUM	0.79	0.84	0.92	0.88	1.00	1.00	1.00	0.83	0.76	1.00	1.00	0.86	1.00	1.00	1.00	0.98	0.62	0.98	0.80	1.00	1.00	1.00	0.97	0.93	0.94	0.94	0.94	0.35	0.36	-																			
BAV	0.84	0.83	0.90	0.86	0.98	0.98	0.97	0.78	0.74	0.97	0.98	0.84	0.98	0.98	0.98	0.96	0.61	0.95	0.78	0.98	0.98	0.98	0.94	0.91	0.92	0.91	0.91	0.26	0.12	0.51	-																		
SHB	0.84	0.81	0.85	0.82	0.94	0.93	0.95	0.73	0.73	0.91	0.92	0.80	0.93	0.93	0.92	0.91	0.60	0.90	0.74	0.92	0.94	0.93	0.91	0.88	0.90	0.94	0.88	0.34	0.27	0.61	0.37	-																	
SUP	0.83	0.81	0.87	0.84	0.96	0.95	0.94	0.75	0.72	0.93	0.94	0.82	0.95	0.95	0.94	0.87	0.61	0.91	0.77	0.94	0.95	0.95	0.91	0.88	0.89	0.92	0.88	0.28	0.07	0.61	0.41	0.44	-															
GRT	0.88	0.86	0.94	0.91	1.00	1.00	1.00	0.68	0.82	1.00	1.00	0.89	1.00	1.00	1.00	0.98	0.72	0.98	0.85	1.00	1.00	1.00	0.97	0.95	0.95	0.95	0.94	0.57	0.59	0.80	0.79	0.77	0.70	-														
AGU	0.95	0.83	0.89	0.87	0.98	0.99	0.96	0.88	0.78	0.94	0.99	0.79	0.99	0.94	0.99	0.97	0.72	0.99	0.86	0.97	0.99	0.97	0.99	0.92	0.94	0.92	0.96	0.72	0.82	0.45	0.07	0.10	-0.08	0.64	-												
ANB	0.95	0.85	0.86	0.82	0.97	0.97	0.94	0.92	0.51	0.74	0.90	0.55	0.86	0.76	0.90	0.85	0.68	0.90	0.61	0.86	0.95	0.88	0.95	0.77	0.85	0.79	0.87	0.68	0.84	0.98	0.93	0.92	0.87	0.95	0.97	-										
BTK	0.84	0.74	0.73	0.69	0.84	0.83	0.81	0.76	0.29	0.48	0.54	0.38	0.63	0.49	0.54	0.56	0.46	0.54	0.33	0.54	0.73	0.65	0.73	0.59	0.69	0.59	0.65	0.49	0.67	0.81	0.78	0.76	0.71	0.83	0.84	0.59	-								
CLP	0.86	0.76	0.74	0.70	0.86	0.85	0.83	0.82	0.34	0.59	0.63	0.42	0.69	0.58	0.63	0.64	0.55	0.63	0.47	0.61	0.68	0.56	0.57	0.40	0.61	0.48	0.56	0.52	0.73	0.87	0.83	0.82	0.77	0.87	0.88	0.66	0.47	-							
BLB	0.94	0.84	0.85	0.81	0.96	0.97	0.94	0.91	0.50	0.79	0.92	0.57	0.94	0.80	0.92	0.88	0.66	0.92	0.65	0.87	0.94	0.86	0.75	0.65	0.71	0.75	0.84	0.62	0.82	0.97	0.92	0.91	0.86	0.95	0.97	0.88	0.65	0.38	-						
DAS	0.89	0.80	0.76	0.68	0.87	0.88	0.84	0.86	0.39	0.67	0.76	0.48	0.79	0.69	0.76	0.75	0.62	0.76	0.57	0.73	0.80	0.72	0.80	0.52	0.75	0.63	0.71	0.59	0.77	0.91	0.87	0.86	0.82	0.90	0.91	0.75	0.49	0.49	0.72	-					
GBY	0.94	0.82	0.79	0.73	0.92	0.94	0.89	0.86	0.67	0.90	0.97	0.69	0.97	0.91	0.97	0.94	0.66	0.97	0.81	0.94	0.97	0.94	0.97	0.88	0.91	0.89	0.93	0.67	0.77	0.96	0.89	0.88	0.82	0.94	0.95	0.94	0.76	0.83	0.94	0.88	-				
HID	0.88	0.78	0.80	0.76	0.90	0.90	0.88	0.80	0.41	0.57	0.67	0.47	0.73	0.54	0.67	0.63	0.52	0.67	0.44	0.67	0.81	0.73	0.81	0.66	0.75	0.66	0.73	0.55	0.73	0.88	0.84	0.82	0.76	0.88	0.89	0.69	0.09	0.54	0.74	0.64	0.83	-			
JOB	0.90	0.78	0.81	0.78	0.91	0.91	0.89	0.84	0.68	0.85	0.92	0.69	0.92	0.86	0.92	0.90	0.65	0.92	0.74	0.88	0.92	0.89	0.92	0.84	0.88	0.84	0.89	0.30	0.35	0.81	0.65	0.67	0.51	0.83	0.91	0.90	0.76	0.78	0.89	0.82	0.89	0.82	-		
KOU	0.91	0.78	0.85	0.82	0.94	0.94	0.92	0.85	0.72	0.87	0.94	0.73	0.94	0.88	0.95	0.92	0.64	0.94	0.76	0.91	0.94	0.91	0.94	0.86	0.90	0.87	0.91	0.25	0.30	0.82	0.59	0.62	0.39	0.84	0.92	0.92	0.77	0.81	0.91	0.86	0.89	0.83	0.27	-		
PHI	0.77	0.66	0.54	0.49	0.65	0.59	0.59	0.70	0.47	0.70	0.78	0.49	0.79	0.73	0.78	0.77	0.56	0.78	0.62	0.74	0.79	0.75	0.79	0.71	0.76	0.69	0.75	0.49	0.60	0.78	0.71	0.71	0.65	0.80	0.78	0.77	0.65	0.65	0.75	0.70	0.54	0.70	0.62	0.65	-		
SEW	0.93	0.82	0.83	0.79	0.95	0.95	0.92	0.90	0.48	0.73	0.87	0.53	0.89	0.75	0.87	0.83	0.64	0.87	0.60	0.81	0.85	0.71	0.77	0.62	0.71	0.57	0.70	0.62	0.81	0.96	0.91	0.90	0.85	0.94	0.95	0.84	0.60	0.42	0.48	0.66	0.92	0.69	0.87	0.90	0.74	-		
SVR	0.80	0.69	0.44	0.36	0.54	0.50	0.42	0.71	0.44	0.75	0.79	0.48	0.81	0.75	0.79	0.79	0.58	0.79	0.67	0.78	0.81	0.78	0.81	0.73	0.77	0.71	0.77	0.56	0.66	0.83	0.77	0.76	0.72	0.83	0.80	0.79	0.63	0.67	0.77	0.67	0.43	0.71	0.72	0.75	0.19	0.75	-		
SWB	0.92	0.77	0.81	0.78	0.96	0.96	0.92	0.79	0.63	0.88	0.96	0.63	0.97	0.88	0.96	0.93	0.66	0.96	0.77	0.93	0.97	0.93	0.97	0.85	0.90	0.85	0.92	0.64	0.79	0.97	0.91	0.90	0.84	0.95	0.95	0.94	0.79	0.79	0.93	0.85	0.93	0.85	0.86	0.89	0.64	0.90	0.68	-		
STL	0.77	0.67	0.45	0.38	0.54	0.50	0.52	0.72	0.29	0.62	0.69	0.35	0.72	0.65	0.69	0.69	0.53	0.69	0.55	0.68	0.71	0.66	0.70	0.60	0.67	0.58	0.65	0.50	0.64	0.80	0.75	0.74	0.70	0.81	0.79	0.70	0.55	0.50	0.63	0.60	0.48	0.63	0.68	0.71	0.21	0.62	0.15	0.63	-		
WOR	0.96	0.84	0.82	0.77	0.96	0.98	0.92	0.89	0.71	0.93	0.99	0.72	0.99	0.93	0.99	0.97	0.72	0.99	0.85	0.97	0.99	0.96	0.99	0.91	0.93	0.92	0.96	0.72	0.82	0.99	0.92	0.91	0.86	0.96	0.97	0.97	0.81	0.87	0.96	0.91	0.64	0.87	0.91	0.93	0.58	0.95	0.48	0.95	0.58	-	
NOT	0.86	0.75	0.76	0.72	0.85	0.86	0.82	0.81	0.63	0.80	0.86	0.66	0.87	0.81	0.86	0.85	0.63	0.86	0.70	0.83	0.87	0.84	0.87	0.80	0.84	0.80	0.84	0.38	0.44	0.77	0.66	0.66	0.57	0.80	0.87	0.85	0.71	0.74	0.84	0.74	0.85	0.79	0.25	0.48	0.58	0.82	0.66	0.80	0.64	0.88	-

Table 2.5 Pairwise F_{ST} values within and between identified Structure Clusters

	North	South West	Central	East	*within Structure cluster*	
North	-	0.001	0.001	0.001	0.726	p<0.001
South West	0.38	-	0.001	0.001	0.626	p<0.000
Central	0.34	0.48	-	0.001	0.731	p<0.001
East	0.41	0.55	0.54	-	0.501	p<0.001

F_{ST} values are listed below the diagonal and corresponding p-values are listed above the diagonal.

Table 2.6 Estimates of migration rates ($N_e m$) and effective population sizes (N_e) between STRUCTURE clusters. In the matrix of migration migration rates, populations in columns are donating migrants and populations in rows are receiving migrants.

	$N_e m$				
	North	East	Central	SouthWest	N_e
North	-	0	0	1.41	1.1×10^8
East		-	0.25	0	5.9×10^7
Central			-	0.70	8.6×10^7
SouthWest				-	9.4×10^7

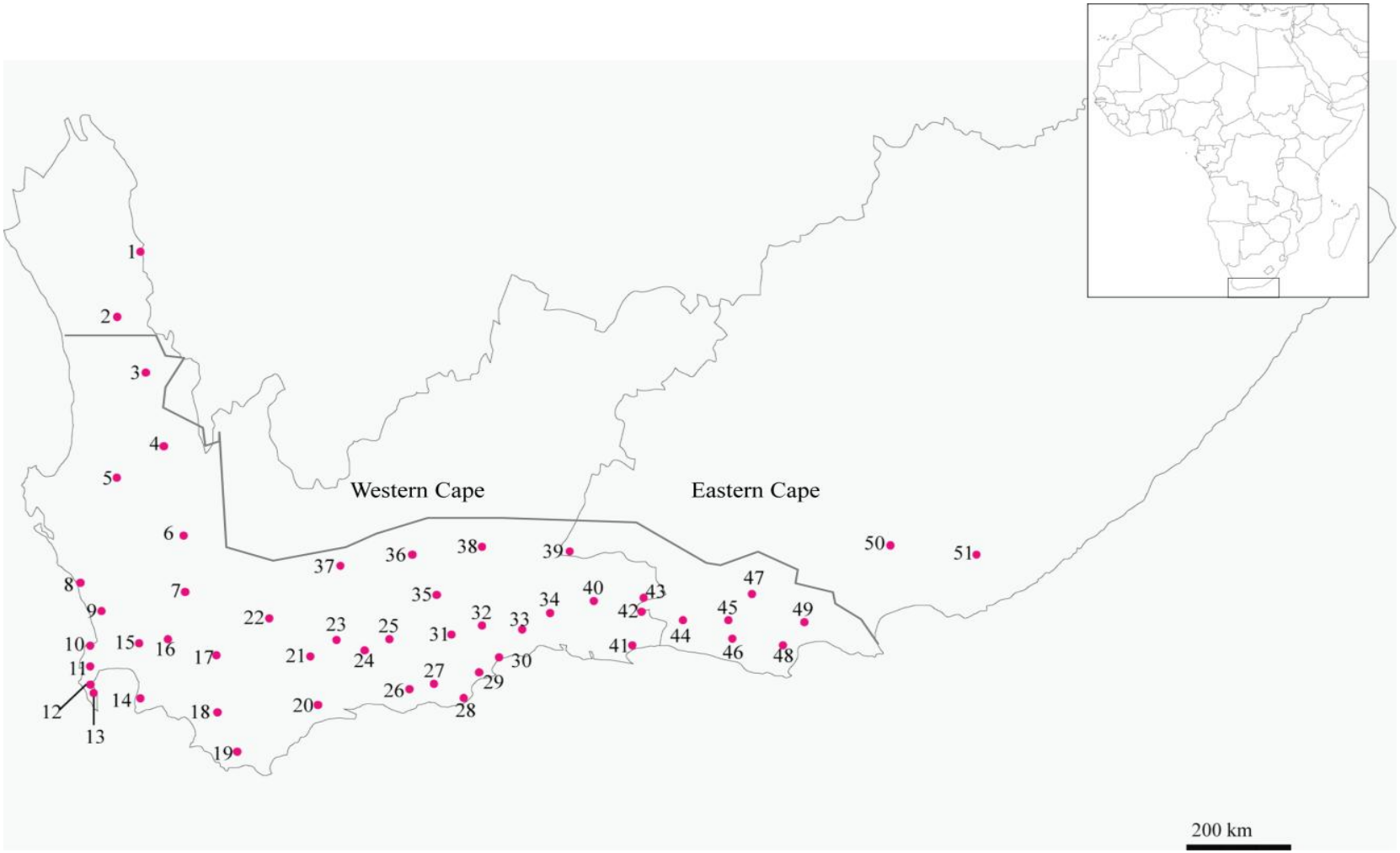

Fig. 2.1 All 51 locations of *L. salignum* sampled throughout the CFR and South Africa. Numbers correspond to locations listed in Table 2.1. The approximate boundary of the CFR (as in Turpie et al. 2003) is outlined in grey.

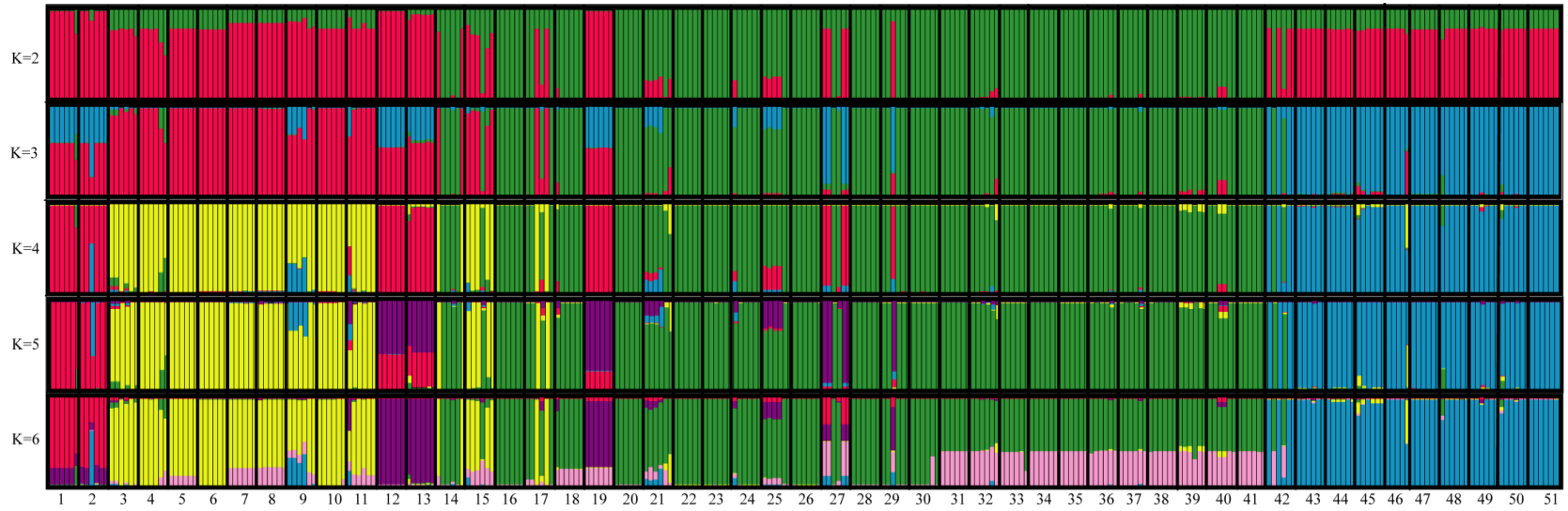

Fig. 2.2 Results of the STRUCTURE analysis with K ranging from 2 to 6. Each individual is represented by a vertical bar. Locations are separated by black lines. Numbers correspond to the following locations: 1. VRP 2. GIP 3. PAP 4. MIP 5. VEP 6. GYP 7. WOR 8. GBY 9. PHI 10. CPT 11. SVR 12. SIT 13. SWB 14. PRB 15. STL 16. FRP 17. GRY 18. STA 19. AGU 20. DHP 21. BTK 22. KOP 23. TRP 24. HID 25. GAP 26. STB 27. ALB 28. VLB 29. MOB 30. GRB 31. CLP 32. ROP 33. OUP 34. DAS 35. RBP 36. SEW 37. ANB 38. SWP 39. BLB 40. POP 41. PLB 42. MIS 43. KOU 44. JOB 45. NOT 46. AGS 47. BAV 48. HUM 49. SHB 50. SUP 51. GRT

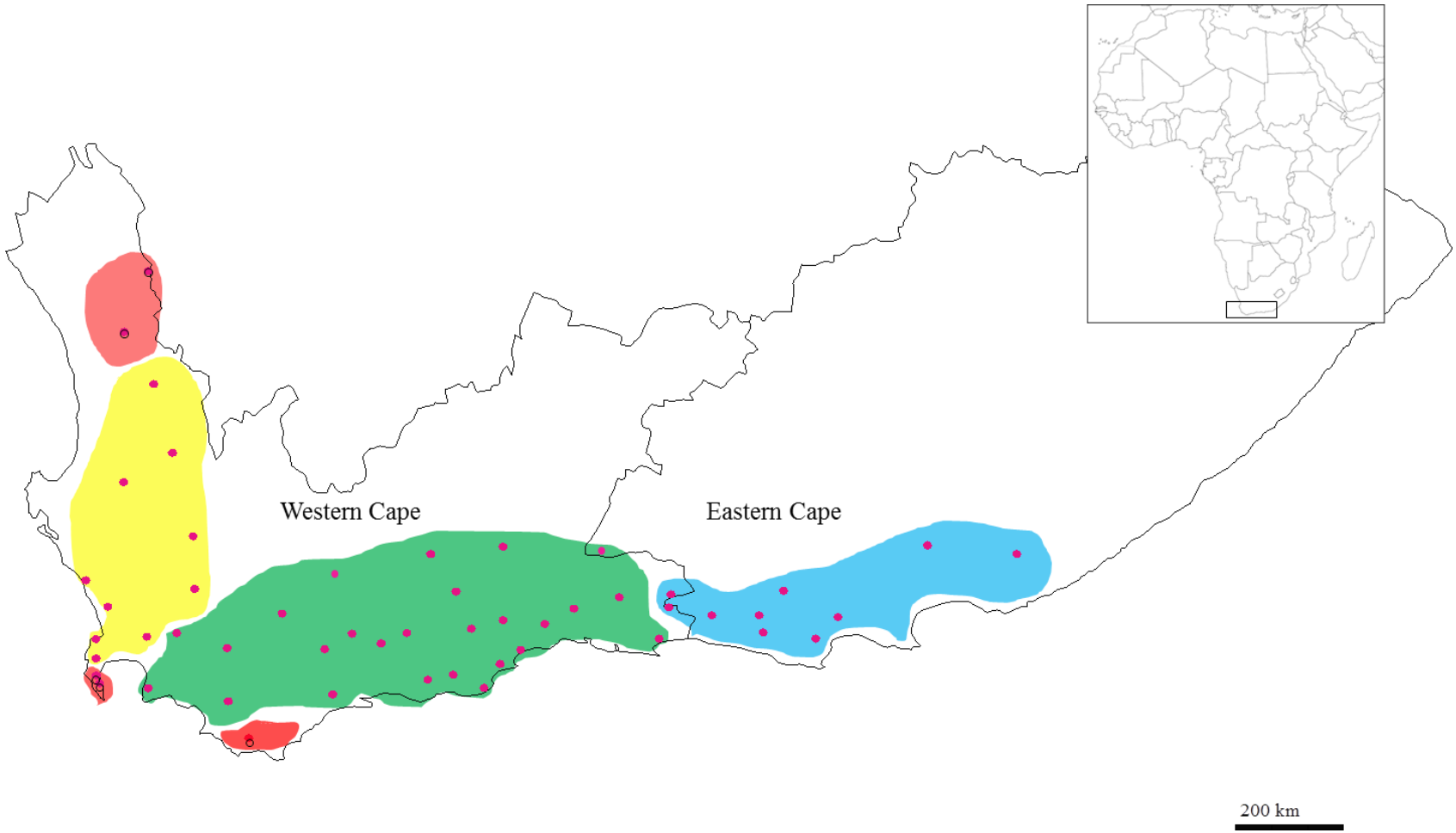

Fig. 2.3 Structure clusters are shaded as follows: North is in red , South West is in yellow , Central is in green , and East is in blue . Inset is a map of Africa, with the enlarged area demarcated by the box. The CFR boundary is outlined in blue.

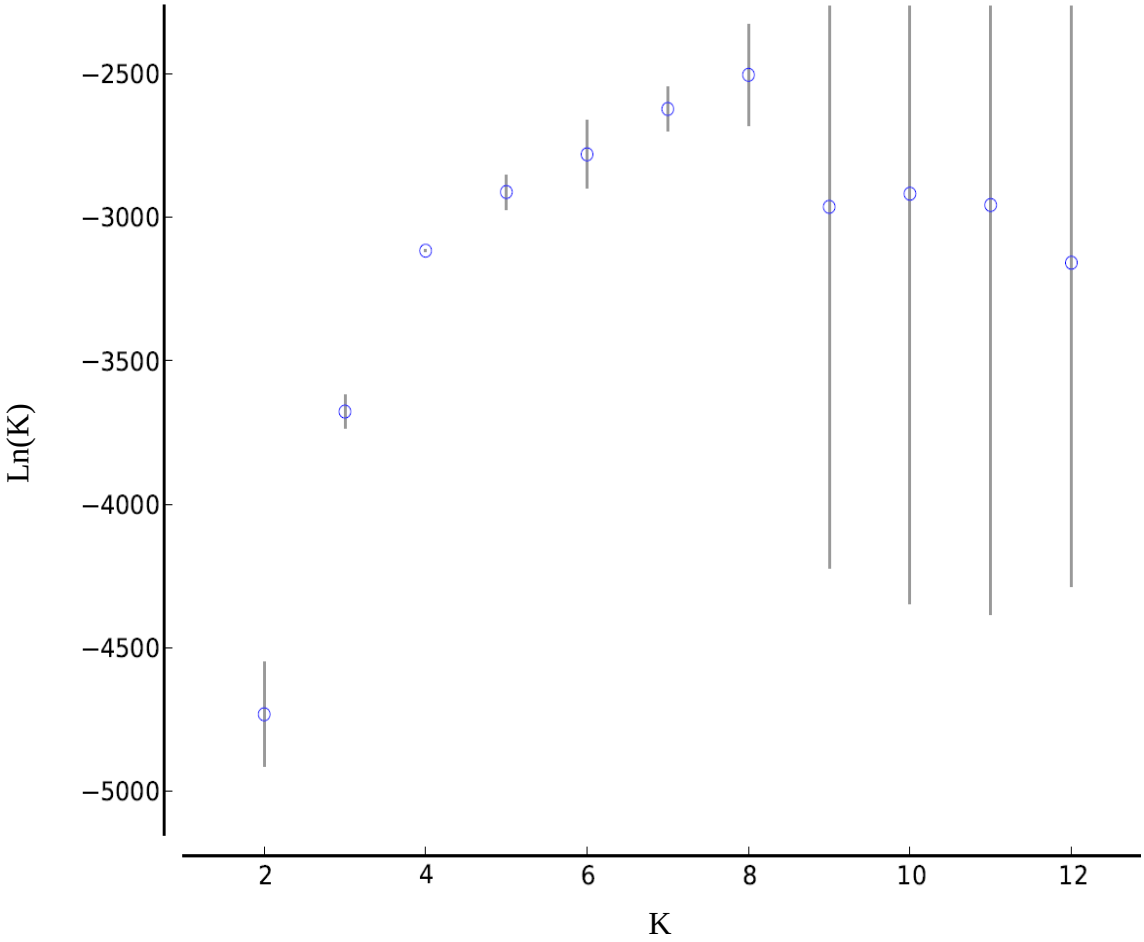

Fig. 2.4 The average log likelihood of the 20 STRUCTURE runs at each K. Error bars are standard deviations. Values for K 2 to 12 are shown.

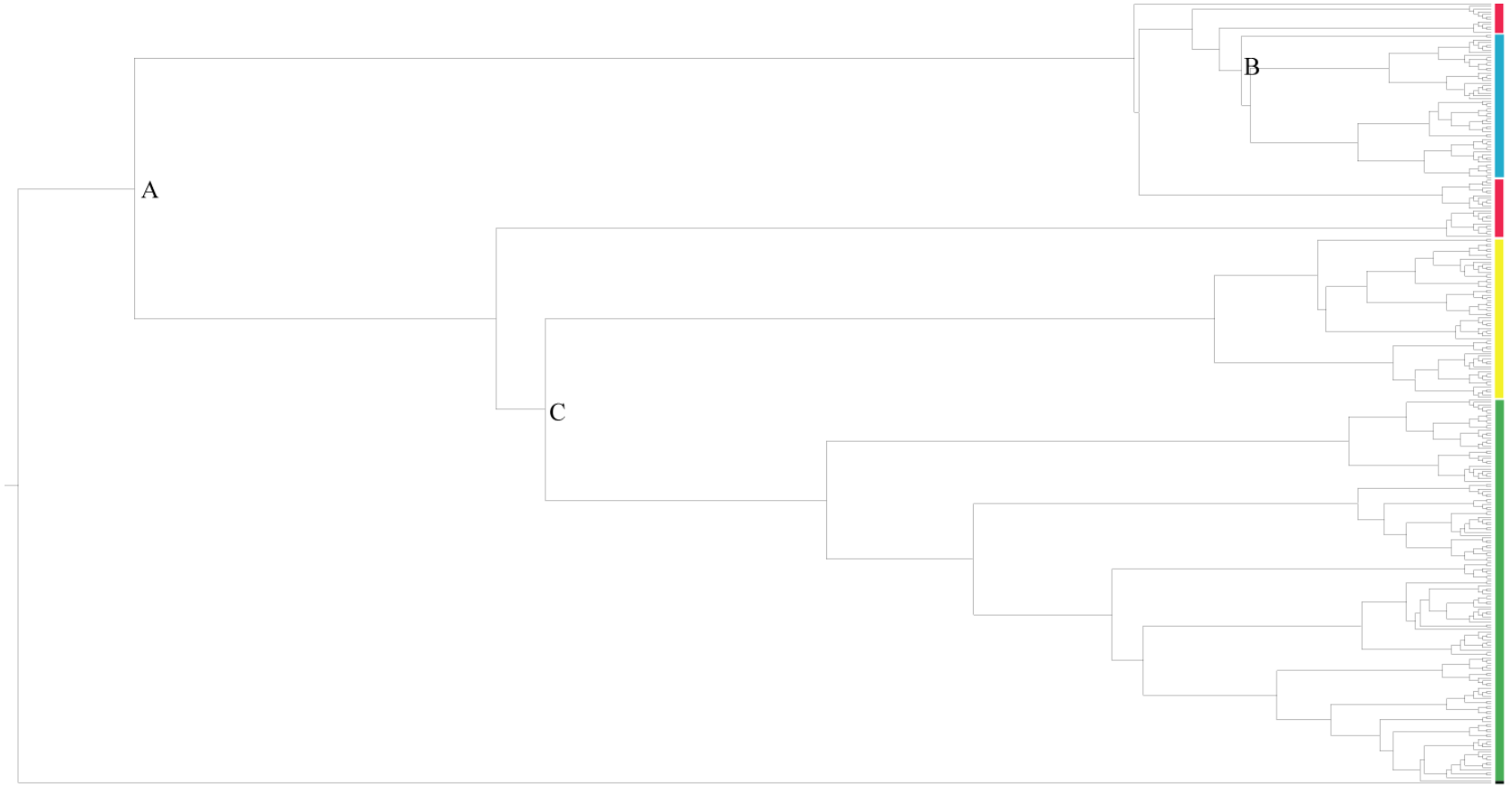

Fig. 2.5 Evolutionary estimated relationships of 306 individuals based on BEAST analysis. The four genetic "color" clusters shown in Fig. 2-3 are mapped to the right. Age estimates at nodes (noted in the text) in millions of years are (A) 1.17, (B) 0.321, and (C) 0.885

Chapter 3

A LANDSCAPE GENETIC STUDY OF THE FYNBOS PLANT *LEUCADENDRON SALIGNUM* FROM THE CAPE FLORAL REGION, SOUTH AFRICA

Introduction

The Cape Floral Region (CFR) is a biodiversity 'hot-spot' and area of global significance (Meyers et al. 2000) that has long been a priority for conservation (Cowling et al. 2003). The CFR is extremely diverse, with over 9,000 vascular plant species of which approximately 70% are endemic, all within an area encompassing only 90,000 km^2. This includes five endemic plant types: fynbos, renosterveld, subtropical thicket and forest, and succulent karoo (Goldblatt 1997; Cowling & Proches 2004). Of these, fynbos and renosterveld are the predominant vegetation in the CFR, with the thicket, forest and succulent karoo occupying much smaller areas. The fynbos, which in Afrikaans means 'fine-leaved bush' is characterized by restiod reed-like bushes, ericoid fine leafed bushes, proteoid large leafed woody shrubs, and geophytes that contain large underground storage organs (Cowling et al. 1996). The fynbos overall makes up 80% of the endemic plants found in the CFR; thus, many have hypothesized that the composition and unique characteristics of the fynbos alone may explain the high species diversity and patterns in the CFR (Goldblatt 1978; Campbell 1983; Linder 2003).

One hypothesis to explain the species richness of the CFR has been the heterogeneity of the CFR environment. For example, the coasts of South Africa that encompass the CFR create the distinct Western and Eastern Capes; substantial geological, topographical, and environmental variation exists between these two regions. The Western Cape receives the highest concentration of rainfall, receiving almost 69" per

annum. This volume decreases in an almost stepwise fashion across to the Eastern Cape which shifts to a bimodal rainfall pattern, with approximately 10" per annum, mostly falling in the spring and autumn (Cowling 1983, 2005a; Latimer et al. 2006). The Eastern Cape has been defined as a region of transition for climate, topography and geological processes (Cowling 1983). It has been hypothesized that the local conditions of each Cape determine the presence of flora and fauna in each region (Linder & Vlok 1991; Linder 2003). Fynbos becomes less common in the Eastern Cape and is replaced by thicket, which preferentially grows in semi-arid to subhumid and subtropical to warm-temperate environments (Acocks 1953; Low & Rebelo 1996; Cowling et al. 1999, 2005b). Soil composition and fertility ranges on a fine scale, from rocky, nutrient-poor soil to nutritionally imbalanced dune and limestone sands in the West while soils in the east contain more nitrogen and phosphorus (Campbell 1983). It has been suggested by Linder (2003) that 'soil nutrient status may form an important barrier' between different types of vegetation. Mountain ranges separate the interior of the country from the coast, acting as barriers to moisture and having a profound effect on the climate. These ranges create elevation gradients, where elevation can change from below sea-level upwards to over 2,200 meters all within a few kilometers. These folded mountains also form a set of ridges, with a range of eastern mountains falling parallel to the Eastern coast and a set which are parallel to the Atlantic, trending from the southern part of the Western Cape and moving north (Linder 2003). The mountain ranges, specifically the Cape Fold Mountains, which originate in the southwestern part of the country near Cape Town, have been identified as potential areas of refugia (Verboom et al. 2009) and plant species richness has been positively correlated with altitude (Cowling & Lombard 2002).

The flora of the CFR are heterogeneous as well, with high levels of turnover between plant habitats at local scales, and high turnover between analogous habitats along geographical gradients (Cowling et al. 1992; Linder 2003), which is evident by the vegetation composition of the CFR (Fig. 3.1). Several studies have identified geographic areas of interest for generating diversity by correlating patterns of genetic diversity with large-scale features (Daniels et al. 2007; Price et al. 2007; Smit et al. 2007; Swart et al. 2008; Willows-Munro & Matthee 2011). For example, by first observing the occurrence of three species clades in a study of angulate tortoises, Daniels et al. (2007) correlated the occurrence of these three distinct cape clades to the presence of the Cape Fold Mountains. Similar observations were made by Willows-Munro & Matthee (2011) in a study of southern African shrews, which found a distinct lineage that broadly followed the topology of the Cape Fold Mountains. In addition, Tolley et al. (2009) found in a study of Agama lizards that little gene flow occurred across the rainfall boundary which delineates the Western and Eastern Capes, suggesting this region presents a barrier to gene flow. However, the inability to localize these regions generating diversity beyond this large scale (i.e., "West vs East") and to identify the cause of these genetic patterns specifically, brings into question whether these studies lack the sampling resolution, especially since the growing hypotheses have all targeted fine-scale heterogeneity as an important factor in generating diversity. That is, genetic sampling of organisms with fine-scaled spatial distributions across the CFR would enable the empirical testing of landscape variables in correlation with spatial genetic structure. Using a landscape genetics approach, we could model the 'interaction between landscape features and

microevolutionary processes, such as gene flow' that may facilitate or isolate individuals and populations over short- and long-term periods (Manel et al. 2003).

Traditionally, landscape genetic approaches have used an 'isolation by distance' (IBD) model as a null hypothesis, where it is assumed that populations in close proximity to one another will share more genetic variation than those that are further away (Wright 1942). This has been tested by measuring the straight line distance between points; however, other spatial have been used. Spatial graphs connect points based on different underlying assumptions and may be helpful because they approximate connectivity between locations given their spatial distributions (Manel et al. 2003; Epperson et al. 2010; Fall et al. 2007). The Delauney triangulation and Gabriel networks are most frequently used in spatial analyses (Dale and Fortin 2010; Diniz-Filho & Bini 2012) and have been used in the CFR (Willows-Munro & Matthee 2011; Tolley et al. 2009). As an alternative to a model of IBD, landscape genetic studies have also used 'isolation by resistance' (IBR) models (i.e., McRae 2006), which allow any landscape feature (including geographic distance) to be parameterized as a 'resistance surface' that weighs these different features in their ability to impede or facilitate gene flow. This method has been shown to have an increased correlation with genetic diversity over models of IBD (McRae &Beier 2007). While this IBR model could help in understanding landscape heterogeneity in the CFR, it has only been used rarely; a single study by Smit et al. (2007) tested patterns of IBR in the cape rock elephant-shrew using resistance calculated as the reciprocal of migration levels between sampled populations. They indicated patterns of clade divergence within the fynbos were significant under both models of IBD and IBR, however, although significant correlations were found, the IBR model did not

consider the landscape of the CFR directly as a factor in shaping the patterns of genetic diversity.

One consideration for a landscape genetic analysis in the CFR is to focus on a plant model species with a broad widespread distribution, and yet one that can also be sampled at the population level across an array of fine-scale landscape features. *Leucadendron salignum* is a widespread fynbos plant, found across both the Western and Eastern Capes, at all levels of elevation and across an array of different soil types. Our previous population genetic analysis of a *L. salignum* data set, using ~5kb across multiple chloroplast DNA markers in 306 individuals from across 51 locations, revealed significant geographic clustering across the entire CFR range (Chapter 2; Fig. 3.2). Surprisingly, when using a pairwise F_{ST} analysis to partition the variation within and between these geographic clusters, it appeared that significant genetic diversity exists between locations even within these cluster groups (Table 3.1), suggesting fine-scale hierarchical structure. Some of this structure may be explained by temporal variation; for example, our phylogenetic analyses suggest that Eastern Cape locations are due to more recent colonization as an expanding monophyletic group out of the Western Cape. Thus, genetic heterogeneity in the Western Cape may be attributed to an older evolutionary and environmental history, whereas, the Easter Cape pattern reflects emergence into a different climatic and vegetation region.

While our previous work characterized patterns of spatial and temporal geographic structure for a plant model across the CFR, in this present study, we test hypotheses about how landscape features may generate this genetic diversity. Specifically, we use our *L. salignum* population genetic dataset combined with spatial

network IBD and IBR models, which incorporate vegetation and other environmental features of the CFR, to determine to what extent these models explain patterns of gene flow on broad- and fine-scales that may lead to increased genetic and phenotypic diversity. We propose that with this higher level of sampling in locations, individuals, and genetic markers, that resolution will be afforded to model not only previously noted broad-scale features (i.e., "West vs East"), but also detect fine-scale variation within the CFR. From a larger perspective, these models can be used to identify geographic regions and environmental features that are consistent, or dissimilar, across taxa in presenting a strategy for developing conservation and management plans in the CFR.

Methods

Samples and Data

Our previous dataset of cpDNA markers totaling ~5kb of DNA sequence data collected from 306 *Leucadendron salignum* individuals from 51 locations inside and outside of the CFR (Fig. 3.1) was used here in all spatial and landscape analyses. These data include sampling across the defined geographic range of *L. salignum*, with GPS coordinates taken for each location. The sampling scheme also reflects the heterogeneity of the CFR, including samples from coastal and mountainous regions, found at all levels of elevation, across a broad range of vegetation, and located on both the Western and Eastern Capes of South Africa. Details on the molecular data collection protocols and explanations of population structure analyses are found in Chapter 2.

Spatial models

Although there are a number of ways (and methods) designed for examining spatial connectivity, we explored only a few that generated simple networks among our

locations in accordance with standard IBD models. These networks provide us with a set of hypothetical scenarios and relaxed assumptions with which to compare and interpret our IBR models below. Spatial relationships between sampled locations were modeled, unless otherwise specified, using *PASSaGE* (Rosenberg & Anderson 2011). Raw geographic distances were estimated from latitude and longitude coordinates as great-circle distances for direct connections among populations. Hypotheses related to more limited spatial connectivity were also tested among locations by constructing four geometric graphs describing hypothetical connection schemes; each represents a fully connected graph. In order from most- to least-densely connected, these included the (1) Delaunay triangulation (Delaunay 1928, 1934), (2) Gabriel graph (Gabriel & Sokal 1969), (3) relative neighborhood network (Toussaint 1980) and (4) minimum spanning tree (Gallager et al. 1983). As an alternative to direct geographic distance, the shortest path distance was determined among locations for each graph as a measure of network distance; the connectivity represented by each of these models is shown in Fig. 3.3

Landscape resistance models

The program Circuitscape 3.5.4 (McRae 2006) was used to model resistance to gene flow among sampled locations as isolation by resistance (IBR). The Circuitscape algorithm borrows from circuit theory (McRae & Beier 2007) and measures the resistance to gene flow between sampled locations by treating the system like a circuit board, where each location is a node connected by a series of resistors or conductors and gene flow is the current. The program allows flexibility in determining the resistance surface and weighting schemes applied for each environmental or landscape feature.

As many have previously suggested that landscape features can be highly correlated (i.e., rainfall and elevation), we recognize that with the significant heterogeneity in the CFR, a variable that takes into consideration as many of these features as possible would be most appropriate from a statistical perspective (i.e., relaxes the problem of independent contrasts). Thus, as a first attempt in generating landscape cost surfaces that can be modeled with our genetic data, we used the variable "vegetation type", which reflects Floral and ecological characteristics of the vegetation strata including soil composition and fertility with fine-scale resolution (Mucina & Rutherford 2007), as categorized by the National Vegetation Map study of the South African National Biodiversity Institute (SANBI) (Fig. 3.1). First, GIS data files of the SANBI database containing >404 vegetation types (Mucina et al. 2007) were overlapped with a file containing density data on *L. salignum* (Protea Atlas Project, http://www.proteaatlas.org.za) in the program ArcGIS. Next, the density of *L. salignum* was calculated for each vegetation type as the number of plants estimated per km^2 (again using ArcGIS); this resulted in a return of 32 unique vegetation types (Table 3.2).

The input file required for Circuitscape is a raster data set, where each cell represents the resistance value corresponding to the probability of the organism moving through that cell to adjacent cells. To create this input, the vegetation type layer from the SANBI vegetation map was exported to a raster map with 10m x 10m grid cells using SAGA GIS v2.0.8 (Conrad 2006). Vegetation types were then reclassified based on the density of *L. salignum* present in each vegetation type, with higher densities given a weight of lower resistance and lower densities given a weight of higher resistance. Given the unknown "true" cost of these different resistance surfaces, the scaled classifications

were modeled with a number of parameters following similar previous simulations (Koen et al. 2012). The models generated here use three classes of curves to represent the cost relationship, linear, bi-modal, and logistic, and each model was scaled with variable "steepness" controlled by the transition between high and low resistance costs. This was done using a logarithmic model that varied both the width of the curve (w), which controlled the steepness of the transition between high and low coasts, and the maximum cost (m) of the surface. After simulations of extremes using these two parameters presented a range of cost surfaces, we created nine resistance surfaces that captured the variation within these extremes by varying w = 1, 5, and 20, and m = 15, 50, 100.

When vegetation types within the geographic region spanning our sample locations were found to have *L. salignum* densities ~0 they were assigned "infinite" resistance values to be conservative with respect to cost of gene flow, as suggested within the Circuitscape protocol. However, vegetation areas assigned infinite resistance such as the karoo that borders the entire northern range of the CFR, as well as areas such as the surrounding ocean, and bays and water bodies within the CFR, all accounted for <1% of the total geographic region encompassing our sample locations across the Western and Eastern Capes. In fact, the rarity of these areas resulted in very little impact on the cost surfaces at any scale (data not shown). Each cost surface, which now reflects different predictions of gene flow as a function of *L. salignum* density across the terrain, was analyzed using Circuitscape to generate a matrix of landscape resistance values. Runs were conducted between all pairs of locations based on average resistance and four-neighbor connections settings (McRae 2006; McRae et al. 2008).

Statistical Analyses

Mantel tests were used to evaluate the amount of genetic variation explained by the various spatial networks and our vegetation resistance models. Previously generated F_{ST} values calculated for this data set were used as measures of spatial genetic diversity (Chapter 2), that is, pairwise genetic distances, as calculated in ARLEQUIN v. 3.5 (Excoffier & Lisher 2010). Statistical significance of the Mantel correlation between the F_{ST} matrix and each spatial and landscape resistance distance matrix was determined using standard permutation tests (999 iterations) in *PASSaGE*. An underappreciated aspect of resistance studies in landscape genetics is that there is often a particularly high correlation between spatial distance and resistance, and this was particularly true here for our data, as well (data not shown). Thus, to account for this correlation and examine the effect of the resistance models in the absence of the spatial component, partial Mantel tests (Smouse et al. 1986) were conducted between landscape resistance and F_{ST} while holding spatial distance constant. Comparisons of Mantel R values (and corresponding p) allowed for an assessment of the performance of each model relative to another. This statistical model and rationale was applied to the overall sampled locations of *L. salignum* as well as to local geographic areas as dictated by predictions based on previous analyses, as well as by our own genetic clustering data (Chapter 2, Fig 3.2).

Results

Pairwise genetic distances between sampling localities based on the 5KB of cpDNA loci collected in Chapter 2 ranged from 0.00-1.0 (Table 3.3, same Results as in Chapter 2). When testing patterns of IBD among all sampled locations, a statistically significant pattern was observed for all spatial models (Table 3.4, R=0.296, all p-values <0.05). In comparison, our model using vegetation type as a measure of resistance for the

entire geographic scale that incorporates all locations was non-significant (R=0.090) after accounting for geographic distance in the model.

We tested distinct geographic areas identified by STRUCTURE analyses in Chapter 2 (Fig. 3.2), to examine the extent that models of IBD and IBR explain spatial genetic diversity. In the three regions tested, represented by the South West, East, and Central clusters, all models were non-significant for tests of IBD and IBR. Interestingly, although the Northwest cluster appears to be a significantly different cluster group in our data set, this clustering of only a few locations was not examined here due to low sample size (but see Discussion). Although models of IBD and IBR were non-significant for the Central cluster, it does represent a large geographic region of ~45,000 km^2, and the degree of genetic diversity among sampled locations within this cluster is still unexpectedly high (F_{ST} = 0.731, p< 0.001). Therefore, like the overall sample, this region may be too large to detect effects of fine-scale heterogeneity on the landscape. To test this hypothesis, several *post hoc* tests were run by breaking this region arbitrarily into two smaller groups, Group A and Group B, each including 10 locations (Fig. 3.4). Group B, located in the easternmost part of the Central cluster, showed no significant correlations for IBD; however, our vegetation resistance model significantly explained patterns of genetic distance (R=0.556, all p<0.02) after controlling for geographic distance. In Group A, located in the west of the Central cluster, IBD was not significant and the vegetation type resistance model suggested a negative correlation with genetic diversity (R = -0.050, all p<0.010) after controlling for geographic distance.

Lastly, we tested geographic areas previously described as correlated to genetic diversity in CFR studies, these being the regions in the southwest CFR near the Cape

Fold Mountains and the border between the Western and Eastern Capes. Interestingly these areas also fall on "breaks" between our identified structure clusters (Fig. 3.2). Near the Cape Fold Mountains, along the border between the South West and Central clusters, the IBD model was non-significant, however, our model of vegetation resistance significantly explained genetic diversity ($R=0.343$, all $p<0.03$) after controlling for geographic distance. At the border between the Western and Eastern capes and the Central and East structure clusters (Fig. 3.2), models of IBD were significant ($R=0.549$, all $p<0.001$) as were models of IBR using vegetation type ($R=0.340$, all $p<0.05$) after controlling for geographic distance.

Discussion

Our previous study showed significant spatial genetic structure and genetic diversity on both large and small scales for the most biogeographically distributed CFR plant *L. salignum*. To understand how the landscape may be contributing to this genetic diversity, we tested hypotheses associating both spatial models of IBD and models of landscape IBR using the variable vegetation type across the broad population sampling of *L. salignum*. Overall, isolation by distance is a significant predictor of patterns of genetic diversity on a broad scale for *L. salignum.* However, when these models were applied to smaller spatial areas where previous studies had identified genetic patterns of divergence, including our own, an IBR model using vegetation type appears to predict fine-scale genetic patterns better than IBD models.

Vegetation type as a predictor of genetic variation in the CFR

The CFR contains a complex landscape filled with heterogeneity across numerous biotic and abiotic features, which implies that resistance from this landscape may explain

genetic diversity patterns and eventual high species diversity. In testing this hypothesis of IBR, we chose a variable that reflects a number of these different features while also being heterogeneous itself at small spatial scales. Vegetation type generated hypotheses regarding the distribution of fynbos owing to the patchy nature of these vegetation types, and the idea being the heterogeneity of vegetation itself may be driving diversity in *L. salignum*. This variable also reflects many features of the terrestrial biodiversity (Mucina et al. 2007) while allowing hypothesis testing of landscape resistance on a fine-scale. However, because the heterogeneity of vegetation type varies so considerably across fine scales (Fig. 3.1), further sampling resolution may be necessary in some regions. For example, the distribution of fynbos is highly concentrated in the west decreases in the east as thicket and forest begin to enter the region along the coast, and succulent and dwarf shrubland extend down from the Karoo (Werger 1978; Cowling 1983). Our variable of vegetation type does take these factors into consideration and was found to correlate significantly with measures of genetic diversity for our samples; however, it may be that the change in fynbos vegetation density moving from west to east is important in explaining patterns of genetic diversity between these two regions.

Vegetation type not only captures where fynbos occurs on the landscape, but we can identify the density of *L. salignum* found in each vegetation type. It is apparent that the density at which *L. salignum* occurs across each vegetation type is highly variable, for example, sandstone fynbos occupies almost 60% of the region, yet the density of *L. salignum* is much lower here than in granite fynbos, which only occupies about 4% of the region but has a density twice that of sandstone. This observation alone suggests that although granite fynbos may appear to be the preferable vegetation type for *L. salignum*,

sandstone fynbos appears to provide connectivity across the CFR for *L. salignum*. Following this result, it may be interesting to examine how sandstone alone provides this connectivity on the landscape by modeling it directly with samples found only on its stratum, especially since it is more prevalent in the east than the west, and may have facilitated the colonization of the Eastern Cape by *L. salignum*. In addition, it will be interesting to study how this vegetation type, given its prevalence may or may not facilitate the dispersal and distribution of other taxa, noting the fact that the patterns here may be a result of our organism sample alone.

While vegetation type significantly predicts genetic diversity at fine scales, it is worth noting that it is only a single feature, albeit a complex one that encapsulates and is correlated with a number of variables, including the presence of fynbos, soil type, and climate (Mucina et al. 2007). Adding to this feature, it will be interesting to examine how other landscape variables may further explain patterns of gene flow in the CFR at fine-spatial scales, possibly as a step-wise multi-regression analysis (Epperson et al. 2010; Legendre &Fortin 2010). For example, rainfall is one climatic variable that might help determine or drive genetic diversity given its graduated regime in the CFR. However, this variable may be correlated to vegetation type, and it does not vary at a fine scale, but rather at the extremes in the northern part of the Western Cape and in the Eastern Cape (Latimer et al. 2006) and so, we may expect that it predicts only broad scale changes at these extremes edges. One heterogeneous variable with a limited correlation to vegetation type might be topographic features of the CFR terrain, which includes the presence of the Cape Fold Mountains.

Landscape resistance in the CFR

In determining how spatial and temporal patterns affect genetic diversity in different regions across the CFR landscape, several hypotheses were tested. First, we tested spatial models of IBD across all sampled locations, finding the overall genetic pattern fit a model of IBD. However, when looking at the residual plot it appears that IBD explains only a small percentage (~8%) of the overall diversity (R=0.336, Fig. 3.5). Other than there being a large number of observations that lead to this "significance", a number of points saturate the area where genetic diversity among locations is high and geographic distance is low, which is not expected by an IBD model. For example, Grootbrak and Mossel Bay are only 21 km apart from one another along the southern coast in the Central cluster region, yet, they appear very diverged ($F_{ST} = 1$). Therefore, although spatial models can explain some population genetic diversity, it is clear that they do not explain a majority of the genetic variation in our sample, especially on fine scales.

We then tested how landscape resistance as measured by vegetation type contributes to the overall genetic diversity of our sampled *L. salignum* population and found no significant correlations. This indicates that, on a large scale, vegetation type is not a good predictor of genetic variation, likely because it reflects heterogeneity on fine-scales, as well as it not being evenly distributed across the CFR. Additionally, models of IBR assigned high costs to regions (e.g. the karoo, bodies of water) through which *L. salignum* likely does not disperse, however, models of IBD do not account for these features. On the other hand, when looking at regional scales, we were able to shed light on how spatial and temporal variation is important. In the East, neither models of IBD nor IBR significantly explained genetic patterns, however; our previous study indicates

that this area reflects a recent colonization and appears much "younger" than Western clades. In this region levels of genetic diversity and structure are much lower than the west, which may simply be an artifact of the recent colonization of this region. Enough time may not have passed since the initial colonization for locations within this region to accumulate genetic variation on the same scale as locations from the West. Additionally, rainfall patterns switch from a winter regime to one of a seasonal rainfall in the East (Cowling et al. 2005; Latimer et al. 2006; Tolley et al. 2009), and the vegetation transitions to include the presence of thicket, forest, and succulent shrubland (Cowling 1983). In the East, "vegetation type" as the resistance factor defined for this analysis, may not be a good predictor of genetic diversity because of the patchy presence of fynbos and the increase in different alternate vegetation types as noted above.

Similar to some previous studies, our genetic structure data also showed a break between the East and Central clusters in this area, and it appears that our IBR model fits the patterns of genetic diversity. Here, the transition zone from largely fynbos to other vegetation types may act as a barrier to gene flow across this border, and thus, this transition zone may explain the recent emergence into the Eastern Cape by this species. It is worth noting again that *L. salignum* is the most widely distributed plant in the fynbos, and thus, it is possible that the pattern here may be unique with respect to CFR plants and specifically fynbos ones.

In the South West structure cluster, measures of IBD and IBR were both non-significant, and it is likely that scale might be important for determining spatial genetic structure here. In this region, there are no temporal differences as in the East, but the presence of vegetation types is very heterogeneous within this area so it may be that

sampling resolution must occur at even finer scales to detect patterns of IBR using vegetation type here. This geographic area has been previously identified as a barrier to gene flow, often attributed to the Cape Fold Mountains (Daniels et al. 2007; Price et al. 2007; Verboom et al. 2009). In our analyses that examined locations along the genetic cluster border we previously identified (Fig. 3.2), we found that IBD models do not predict genetic diversity, but that vegetation type did. This result corroborates with previous studies; however, it does not definitively identify what about this region is ultimately contributing to genetic diversity as vegetation type could be correlated with the Cape Fold Mountain terrain. To improve resolution here, additional sampling would help to fill in areas, allowing for more fine-scale testing.

In the Central region, neither our IBD nor IBR models significantly explain genetic diversity on large scale. However, this region harbors the largest amount of genetic diversity and it may be that the scale of this area (~45,000 km^2) may be too large to detect resolution on fine-scales. Our *post hoc* analyses that broke up the region into smaller areas suggest that scale size may be the explanation; that is, until we examine correlations between genetic and landscape features on an even finer scale in this region, we are unlikely to have the resolution. Nevertheless, it is clear that this region exhibits the highest amount of genetic diversity on fine-scale and that landscape resistance plays a greater role than simply geographic distance.

Finally, it is of note that we were unable to test the North cluster due to insufficient sample size. Previous studies have found similar patterns of spatial genetic structure in this region, with clusters appearing in the northwest (Daniels et al. 2007; Matthee & Robinson 2007; Smit et al. 2010; Willows-Munro & Matthee 2011), so

although we were unable to apply our models of IBD and IBR to this area at this point, further sampling across this area would be justified to try and elucidate what may be driving the diversity in this region as well. Nonetheless, we should also note that this area does not reflect "recent" lineages such as the Eastern Cape, but instead, this area that is isolated looks quite the opposite in that it actually appears to harbor "older" diverged lineages (see Chapter 2), possibly even ancestral refugia as compared to those locations in the Western Cape.

Conclusion

Using a molecular population genetic model of *L. salignum*, the most biogeographically widespread fynbos plant, we find that incorporating fine-scale sampling and variables of landscape resistance are more likely to elucidate how the CFR landscape contributes to spatial genetic variation among locations across the Western and Eastern Capes than does distance alone. Given the small area of the CFR and its fine-scale heterogeneity, identifying patterns of gene flow on these small scales could have a large impact on identifying potential conservation areas within the CFR. Specifically, certain patterns, such as those in the Eastern Cape, indicate temporal factors have been influential in driving patterns of diversity, while patterns of variation in the Southwest may reflect that barriers, such as the Cape Fold Mountains, play a role, and even more in contrast is the Central region, where locations separated by close geographic range appear to be genetically very different. In all, temporal and spatial landscape variation contributes to genetic diversity and gene flow across the CFR, but clearly on different scales in these different regions. Our findings may not be universal across all organisms, suggesting that diversity may not be impacted the same way across the CFR for all taxa.

Specifically, while there may be certain "environmental triggers" (Linder and Hardy 2004; Sauquet et al. 2009) to explain radiation and diversity in the CFR, there may be examples of certain taxa and certain geographic regions where this is not the case. Moving forward, studies seeking to explain genetic diversity using a similar model need to be wary of not only the number of genetic markers and individuals, but specifically the number and location of their samples inside and outside of the CFR.

Table 3.1 Pairwise F_{ST} values within and between Structure Clusters using cpDNA

	North	South West	Central	East	*within Structure cluster*	
North	-	0.001	0.001	0.001	0.726	p<0.001
South West	0.38	-	0.001	0.001	0.626	p<0.000
Central	0.34	0.48	-	0.001	0.731	p<0.001
East	0.41	0.55	0.54	-	0.501	p<0.001

F_{ST} values are listed below the diagonal and corresponding p-values are listed above the diagonal.

Table 3.2 Vegetation types found within the *L. salignum* range

Vegetation type	total sq km[a]	% of total sq km	*L. salignum* density/sq km[b]
Granite Fynbos	1073	3.979	0.959
Ferricrete Fynbos	706	2.197	0.8045
Shale Band Vegetation	753	2.038	0.6999
Alluvium Fynbos	1216	2.862	0.6086
Shale Fynbos	2830	6.431	0.5876
Silcrete Fynbos	869	1.876	0.5581
Sandstone Fynbos	30106	59.908	0.5145
Limestone Fynbos	2104	3.562	0.4377
Conglomerate Fynbos	624	0.781	0.3237
Limestone Renosterveld	502	0.333	0.1713
Sand Fynbos	8189	4.927	0.1556
Estuarine Vegetation	163	0.058	0.092
Silcrete Renosterveld	310	0.108	0.0903
Alluvium Renosterveld	561	0.193	0.0891
Quartzite Fynbos	4512	1.257	0.072
Shale Renosterveld	23909	6.126	0.0663
Western Strandveld	3001	0.603	0.052
Zonal & Intrazonal Forests	4485	0.677	0.039
Waterbodies	673	0.093	0.0357
Freshwater Wetlands	1326	0.112	0.0219
Alluvial Vegetation	8817	0.267	0.0078
Albany Thicket	29128	0.746	0.0066
Seashore Vegetation	349	0.008	0.0057
Rainshadow Valley Karoo	27893	0.611	0.0057
Knersvlakte Bioregion	5061	0.054	0.0028
Eastern Strandveld	472	0.004	0.0021
Inland Saline Vegetation	17855	0.124	0.0018
Sub-Escarpment Savanna Bioregion	35523	0.031	0.0002
Namaqualand Hardeveld Bioregion	19352	0.012	0.0002
Lower Karoo Bioregion	31863	0.015	0.0001
Namaqualand Sandveld Bioregion	9097	0.004	0.0001
Trans-Escarpment Succulent Karoo Bioregion	15416	0.004	0.0001

[a]**Mucina & Rutherford** (2007)

[b]Protea Atlas Project

Table 3.3 Pairwise F_{ST} values for all sampled locations using cpDNA

	VRP	GIP	PAP	MIP	VEP	GYP	CPT	SIT	PRB	STA	FRP	GRY	KOP	TRP	DHP	GAP	STB	ALB	VLB	MOB	GRB	ROP	RBP	OUP	SWP	POP	PLB	MIS	AGS	HUM	BAV	SHB	SUP	GRT	AGU	ANB	BTK	CLP	BLB	DAS	GBY	HID	JOB	KOU	PHI	SEW	SVR	SWB	STL	WOR	NOT
VRP	-																																																				
GIP	0.39	-																																																		
PAP	0.77	0.76	-																																																	
MIP	0.74	0.74	0.15	-																																																
VEP	0.86	0.80	0.59	0.58	-																																														
GYP	0.84	0.77	0.00	0.40	0.64	-																																													
CPT	0.86	0.80	0.57	0.58	0.62	0.61	-																																													
SIT	0.78	0.72	0.79	0.83	0.84	0.82	0.85	-																																											
PRB	0.74	0.75	0.33	0.40	0.68	0.44	0.61	0.81	-																																										
STA	0.85	0.80	0.68	0.74	0.86	0.81	0.87	0.93	0.13	-																																									
FRP	0.85	0.77	0.59	0.74	0.79	0.73	0.81	0.93	0.17	0.00	-																																								
GRY	0.78	0.69	0.27	0.46	0.59	0.34	0.58	0.82	0.07	0.20	0.20	-																																							
KOP	0.87	0.86	0.70	0.80	0.84	0.80	0.85	0.94	0.47	1.00	1.00	0.54	-																																						
TRP	0.90	0.86	0.82	0.80	0.91	0.86	0.89	0.94	0.48	1.00	1.00	0.54	1.00	-																																					
DHP	0.84	0.84	0.54	0.74	0.76	0.70	0.79	0.93	0.20	0.00	0.00	0.20	1.00	1.00	-																																					
GAP	0.85	0.84	0.64	0.75	0.80	0.75	0.82	0.92	0.38	0.80	0.80	0.43	0.91	0.00	0.80	-																																				
STB	0.60	0.61	0.49	0.51	0.58	0.53	0.70	0.51	0.41	0.47	0.47	0.35	0.57	0.57	0.47	0.52	-																																			
ALB	0.86	0.83	0.70	0.71	0.81	0.75	0.97	0.91	0.13	0.00	0.00	0.18	0.86	0.86	0.00	0.67	0.45	-																																		
VLB	0.74	0.71	0.45	0.50	0.63	0.54	0.78	0.72	0.03	0.00	0.00	0.06	0.43	0.43	0.00	0.31	0.22	0.00	-																																	
MOB	0.89	0.84	0.75	0.74	1.00	0.78	1.00	0.93	0.15	0.00	0.00	0.20	1.00	1.00	0.00	0.80	0.47	0.00	0.00	-																																
GRB	0.89	0.87	0.86	0.84	1.00	1.00	1.00	0.95	0.66	1.00	1.00	0.68	1.00	1.00	1.00	0.84	0.64	0.92	0.60	1.00	-																															
ROP	0.88	0.86	0.82	0.80	1.00	1.00	1.00	0.57	0.51	1.00	1.00	0.54	1.00	1.00	1.00	0.91	0.57	0.86	0.43	1.00	1.00	-																														
RBP	0.85	0.84	0.77	0.76	0.96	0.94	0.96	0.91	0.48	0.80	0.80	0.50	0.89	0.89	0.80	0.80	0.55	0.71	0.40	0.80	0.92	0.89	-																													
OUP	0.85	0.82	0.68	0.70	0.95	0.92	0.95	0.90	0.16	0.20	0.20	0.20	0.80	0.80	0.20	0.63	0.45	0.13	0.04	0.20	0.84	0.60	0.66	-																													
SWP	0.89	0.85	0.81	0.81	0.94	0.92	0.94	0.92	0.67	0.89	0.87	0.68	0.90	0.90	0.87	0.86	0.66	0.83	0.65	0.87	0.91	0.90	0.71	0.76	-																												
POP	0.88	0.86	0.82	0.80	1.00	1.00	0.94	0.57	0.35	1.00	1.00	0.54	1.00	1.00	1.00	0.91	0.57	0.86	0.43	1.00	1.00	0.00	0.73	0.47	0.79	-																											
PLB	0.93	0.86	0.82	0.80	1.00	1.00	1.00	0.57	0.45	1.00	1.00	0.54	1.00	1.00	1.00	0.91	0.57	0.86	0.43	1.00	1.00	0.00	0.89	0.60	0.87	0.20	-																										
MIS	0.69	0.63	0.55	0.53	0.74	0.66	0.75	0.53	0.39	0.51	0.55	0.44	0.63	0.63	0.55	0.59	0.27	0.52	0.35	0.55	0.68	0.63	0.57	0.50	0.65	0.54	0.54	-																								
AGS	0.93	0.70	0.69	0.63	0.81	0.78	0.84	0.51	0.59	0.72	0.76	0.64	0.79	0.79	0.76	0.76	0.43	0.73	0.57	0.76	0.81	0.79	0.79	0.75	0.80	0.75	0.72	0.06	-																							
HUM	0.79	0.84	0.92	0.88	1.00	1.00	1.00	0.83	0.76	1.00	1.00	0.86	1.00	1.00	1.00	0.98	0.62	0.98	0.80	1.00	1.00	1.00	0.97	0.93	0.94	0.94	0.94	0.35	0.36	-																							
BAV	0.84	0.83	0.90	0.86	0.98	0.98	0.97	0.78	0.74	0.97	0.98	0.84	0.98	0.98	0.98	0.96	0.61	0.95	0.78	0.98	0.98	0.98	0.94	0.91	0.92	0.91	0.91	0.26	0.12	0.51	-																						
SHB	0.84	0.81	0.85	0.82	0.94	0.93	0.95	0.73	0.73	0.91	0.92	0.80	0.93	0.93	0.92	0.91	0.60	0.90	0.74	0.92	0.94	0.93	0.91	0.88	0.90	0.94	0.88	0.34	0.27	0.61	0.37	-																					
SUP	0.83	0.81	0.87	0.84	0.96	0.95	0.94	0.75	0.72	0.93	0.94	0.82	0.95	0.95	0.94	0.87	0.61	0.91	0.77	0.94	0.95	0.95	0.91	0.88	0.89	0.92	0.88	0.28	0.07	0.61	0.41	0.44	-																				
GRT	0.88	0.86	0.94	0.91	1.00	1.00	1.00	0.68	0.82	1.00	1.00	0.89	1.00	1.00	1.00	0.98	0.72	0.98	0.85	1.00	1.00	1.00	0.97	0.95	0.95	0.95	0.94	0.57	0.59	0.80	0.79	0.77	0.70	-																			
AGU	0.95	0.83	0.89	0.87	0.98	0.99	0.96	0.88	0.78	0.94	0.99	0.79	0.99	0.94	0.99	0.97	0.72	0.99	0.86	0.97	0.99	0.97	0.99	0.92	0.94	0.92	0.96	0.72	0.82	0.45	0.07	0.10	-0.08	0.64	-																		
ANB	0.95	0.85	0.86	0.82	0.97	0.97	0.94	0.92	0.51	0.74	0.90	0.55	0.86	0.76	0.90	0.85	0.68	0.90	0.61	0.86	0.95	0.88	0.95	0.77	0.85	0.79	0.87	0.68	0.84	0.98	0.93	0.92	0.87	0.95	0.97	-																	
BTK	0.84	0.74	0.73	0.69	0.84	0.83	0.81	0.76	0.29	0.48	0.54	0.38	0.63	0.49	0.54	0.56	0.46	0.54	0.33	0.54	0.73	0.65	0.73	0.59	0.69	0.59	0.65	0.49	0.67	0.81	0.78	0.76	0.71	0.83	0.84	0.59	-															
CLP	0.86	0.76	0.74	0.70	0.86	0.85	0.83	0.82	0.34	0.59	0.63	0.42	0.69	0.58	0.63	0.64	0.55	0.63	0.47	0.61	0.68	0.56	0.57	0.40	0.61	0.48	0.56	0.52	0.73	0.87	0.83	0.82	0.77	0.87	0.88	0.66	0.47	-													
BLB	0.94	0.84	0.85	0.81	0.96	0.97	0.94	0.91	0.50	0.79	0.92	0.57	0.94	0.80	0.92	0.88	0.66	0.92	0.65	0.87	0.94	0.86	0.75	0.65	0.71	0.75	0.84	0.62	0.82	0.97	0.92	0.91	0.86	0.95	0.97	0.88	0.65	0.38	-									
DAS	0.89	0.80	0.76	0.68	0.87	0.88	0.84	0.86	0.39	0.67	0.76	0.48	0.79	0.69	0.76	0.75	0.62	0.76	0.57	0.73	0.80	0.72	0.80	0.52	0.75	0.63	0.71	0.59	0.77	0.91	0.87	0.86	0.82	0.90	0.91	0.75	0.49	0.49	0.72	-							
GBY	0.94	0.82	0.79	0.73	0.92	0.94	0.89	0.86	0.67	0.90	0.97	0.69	0.97	0.91	0.97	0.94	0.66	0.97	0.81	0.94	0.97	0.94	0.97	0.88	0.91	0.89	0.93	0.67	0.77	0.96	0.89	0.88	0.82	0.94	0.95	0.94	0.76	0.83	0.94	0.88	-						
HID	0.88	0.78	0.80	0.76	0.90	0.90	0.88	0.80	0.41	0.57	0.67	0.47	0.73	0.54	0.67	0.63	0.52	0.67	0.44	0.67	0.81	0.73	0.81	0.66	0.75	0.66	0.73	0.55	0.73	0.88	0.84	0.82	0.76	0.88	0.89	0.69	0.09	0.54	0.74	0.64	0.83	-					
JOB	0.90	0.78	0.81	0.78	0.91	0.91	0.89	0.84	0.68	0.85	0.92	0.69	0.92	0.86	0.92	0.90	0.65	0.92	0.74	0.88	0.92	0.89	0.92	0.84	0.88	0.84	0.89	0.30	0.35	0.81	0.65	0.67	0.51	0.83	0.91	0.90	0.76	0.78	0.89	0.82	0.89	0.82	-				
KOU	0.91	0.78	0.85	0.82	0.94	0.94	0.92	0.85	0.72	0.87	0.94	0.73	0.94	0.88	0.95	0.92	0.64	0.94	0.76	0.91	0.94	0.91	0.94	0.86	0.90	0.87	0.91	0.25	0.30	0.82	0.59	0.62	0.39	0.84	0.92	0.92	0.77	0.81	0.91	0.86	0.89	0.83	0.27	-			
PHI	0.77	0.66	0.54	0.49	0.65	0.59	0.59	0.70	0.47	0.70	0.78	0.49	0.79	0.73	0.78	0.77	0.56	0.78	0.62	0.74	0.79	0.75	0.79	0.71	0.76	0.69	0.75	0.49	0.60	0.78	0.71	0.71	0.65	0.80	0.78	0.77	0.65	0.65	0.75	0.70	0.54	0.70	0.62	0.65	-			
SEW	0.93	0.82	0.83	0.79	0.95	0.95	0.92	0.90	0.48	0.73	0.87	0.53	0.89	0.75	0.87	0.83	0.64	0.87	0.60	0.81	0.85	0.71	0.77	0.62	0.71	0.57	0.70	0.62	0.81	0.96	0.91	0.90	0.85	0.94	0.95	0.84	0.60	0.42	0.48	0.66	0.92	0.69	0.87	0.90	0.74	-		
SVR	0.80	0.69	0.44	0.36	0.54	0.50	0.42	0.71	0.44	0.75	0.79	0.48	0.81	0.75	0.79	0.79	0.58	0.79	0.67	0.78	0.81	0.78	0.81	0.73	0.77	0.71	0.77	0.56	0.66	0.83	0.77	0.76	0.72	0.83	0.80	0.79	0.63	0.67	0.77	0.67	0.43	0.71	0.72	0.75	0.19	0.75	-		
SWB	0.92	0.77	0.81	0.78	0.96	0.96	0.92	0.79	0.63	0.88	0.96	0.63	0.97	0.88	0.96	0.93	0.66	0.96	0.77	0.93	0.97	0.93	0.97	0.85	0.90	0.85	0.92	0.64	0.79	0.97	0.91	0.90	0.84	0.95	0.95	0.94	0.79	0.79	0.93	0.85	0.93	0.85	0.86	0.89	0.64	0.90	0.68	-	
STL	0.77	0.67	0.45	0.38	0.54	0.50	0.52	0.72	0.29	0.62	0.69	0.35	0.72	0.65	0.69	0.69	0.53	0.69	0.55	0.68	0.71	0.66	0.70	0.60	0.67	0.58	0.65	0.50	0.64	0.80	0.75	0.74	0.70	0.81	0.79	0.70	0.55	0.50	0.63	0.60	0.48	0.63	0.68	0.71	0.21	0.62	0.15	0.63	-	
WOR	0.96	0.84	0.82	0.77	0.96	0.98	0.92	0.89	0.71	0.93	0.99	0.72	0.99	0.93	0.99	0.97	0.72	0.99	0.85	0.97	0.99	0.96	0.99	0.91	0.93	0.92	0.96	0.72	0.82	0.99	0.92	0.91	0.86	0.96	0.97	0.97	0.81	0.87	0.96	0.91	0.64	0.87	0.91	0.93	0.58	0.95	0.48	0.95	0.58	-	
NOT	0.86	0.75	0.76	0.72	0.85	0.86	0.82	0.81	0.63	0.80	0.86	0.66	0.87	0.81	0.86	0.85	0.63	0.86	0.70	0.83	0.87	0.84	0.87	0.80	0.84	0.80	0.84	0.38	0.44	0.77	0.66	0.66	0.57	0.80	0.87	0.85	0.71	0.74	0.84	0.74	0.85	0.79	0.25	0.48	0.58	0.82	0.66	0.80	0.64	0.88	-

Table 3.4 Summary Mantel R values for all spatial and resistance models tested.

		Geographic Genetic Clusters			**Geographic Regions**		***Ad hoc* Analyses**	
Spatial Models	All locations	South West	Central	East	Cape Fold Mountains	Capes Border	Central Group A	Central Group B
Geographic Distance	**0.297**	0.047	0.059	0.384	0.243	**0.535**	0.055	-0.036
Delaunay Triangulation	**0.290**	0.055	0.051	0.385	0.248*	**0.532**	0.066	-0.025
Gabriel Network	**0.290**	0.08	0.047	0.383	0.232	**0.553**	0.003	0.002
Relative Neighbor	**0.301**	0.100	0.069	0.383	0.227	**0.564**	-0.046	0.018
Minimum Spanning Distance	**0.303**	0.197	0.11	0.383	0.0665	**0.565**	-0.017	0.083
Resistance Models								
W1M15	0.098	0.200	0.091	0.435	0.352*	0.338*	**-0.533**	**0.475**
W1M50	0.099	0.217	0.106	0.432	0.349*	0.357*	**-0.52**	**0.501**
W1M100	0.100	0.222	0.109	0.432	0.347*	0361*	**-0.517**	**0.505**
W5M15	0.087	0.186	0.117	0.444	0.348*	0.385*	**-0.522**	**0.524**
W5M50	0.088	0.201	0.137	0.443	0.345*	0.414*	**-0.51**	**0.556**
W5M100	0.088	0.205	0.142	0.442	0.343*	0.420*	**-0.507**	**0.564**
W20M15	0.079	0.161	0.182	0.458	0.341*	0.319	**-0.443**	**0.595**
W20M50	0.080	0.173	0.203	0.463	0.339*	0.237	**-0.424**	**0.626**
W20M100	0.079	0.179	0.209	0.463	0.335*	0.217	**-0.417**	0.635*

p-values corresponding to each Mantel R value: **p<0.05**, *p<0.001

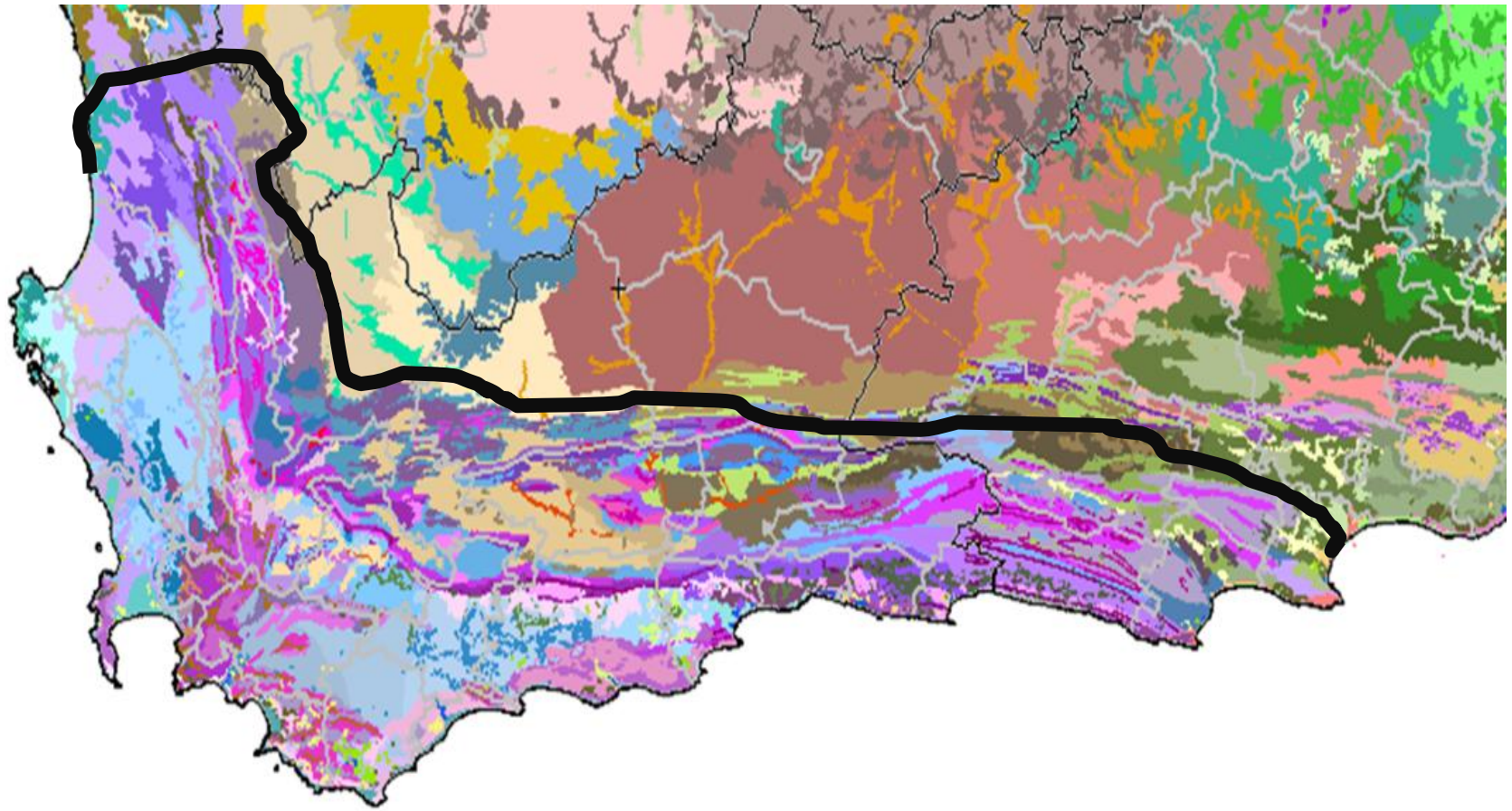

Fig. 3.1 Vegetation of the CFR and South Africa. Each color represents a unique vegetation type as measured by SANBI. (Mucina et al. 2007).The approximate boundary of the CFR (as in Turpie et al. 2003) is outlined in black.

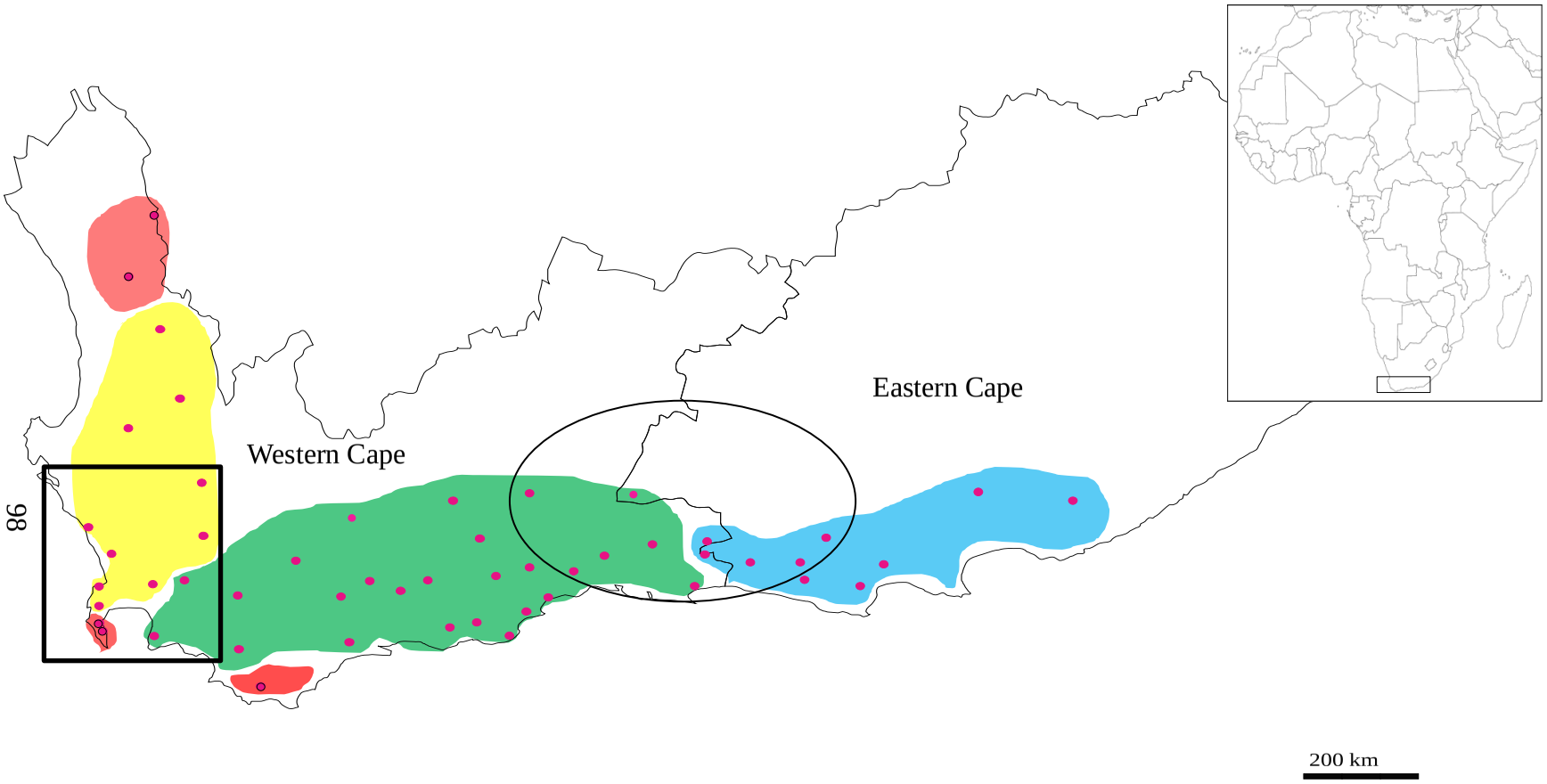

Fig. 3.2 Sampling locations (indicated by dots) for *L. salignum*. Structure clusters are shaded as follows: North is in red, South West is in yellow, Central is in green, and East is in blue. Locations in the square and ovals represent geographical testing for the Cape Fold Mountains and boundary between the Western and Eastern Capes, respectively. Inset is a map of Africa, with the enlarged area demarcated by the box. The CFR boundary is outlined in blue.

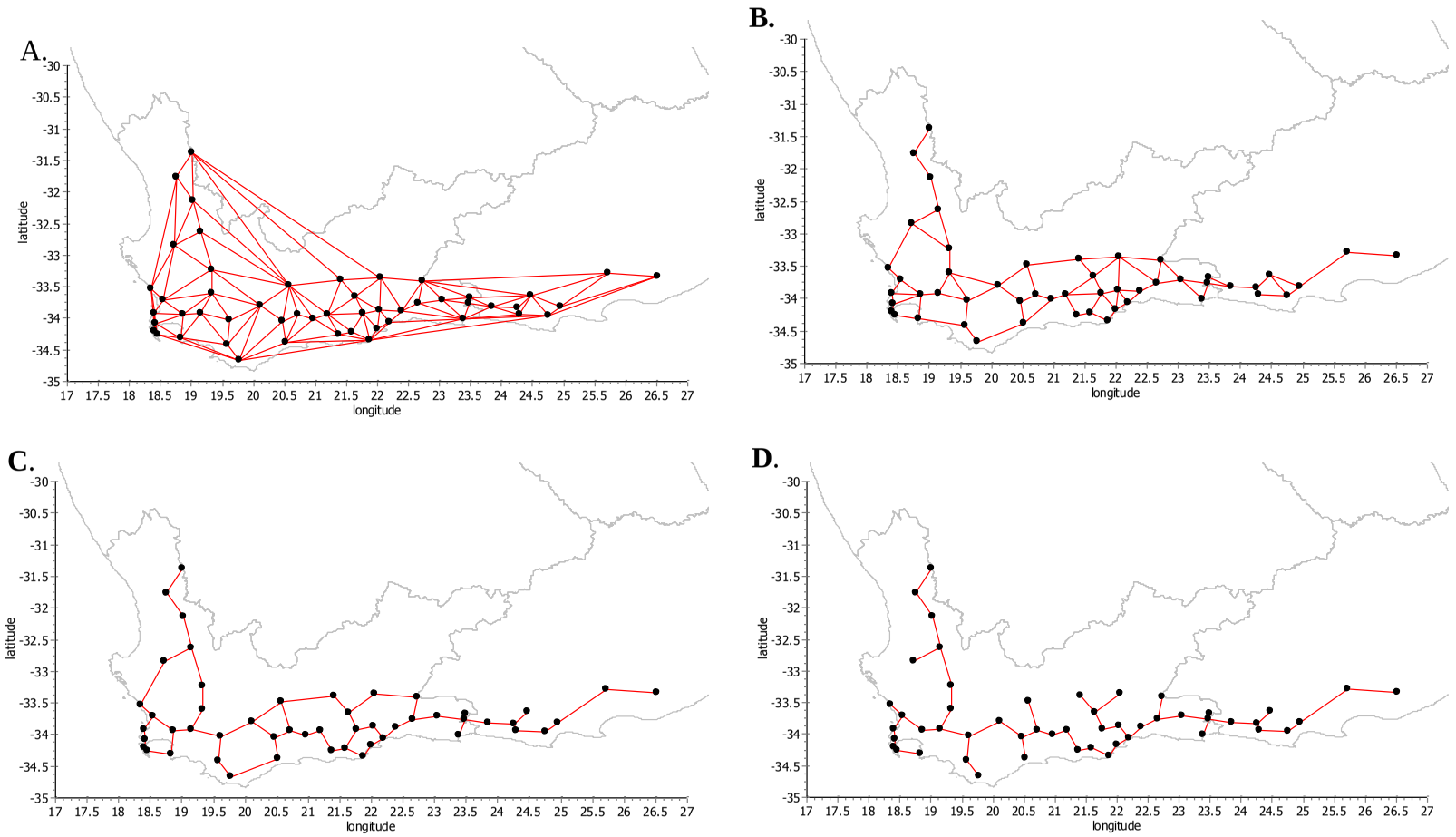

Fig. 3.3 Spatial networks connecting *L. salignum* decreasing from most to least connected: A. The Delaunay Triangulation B. Gabriel Network C. Relative Neighborhood Network D. Minimum Spanning Tree

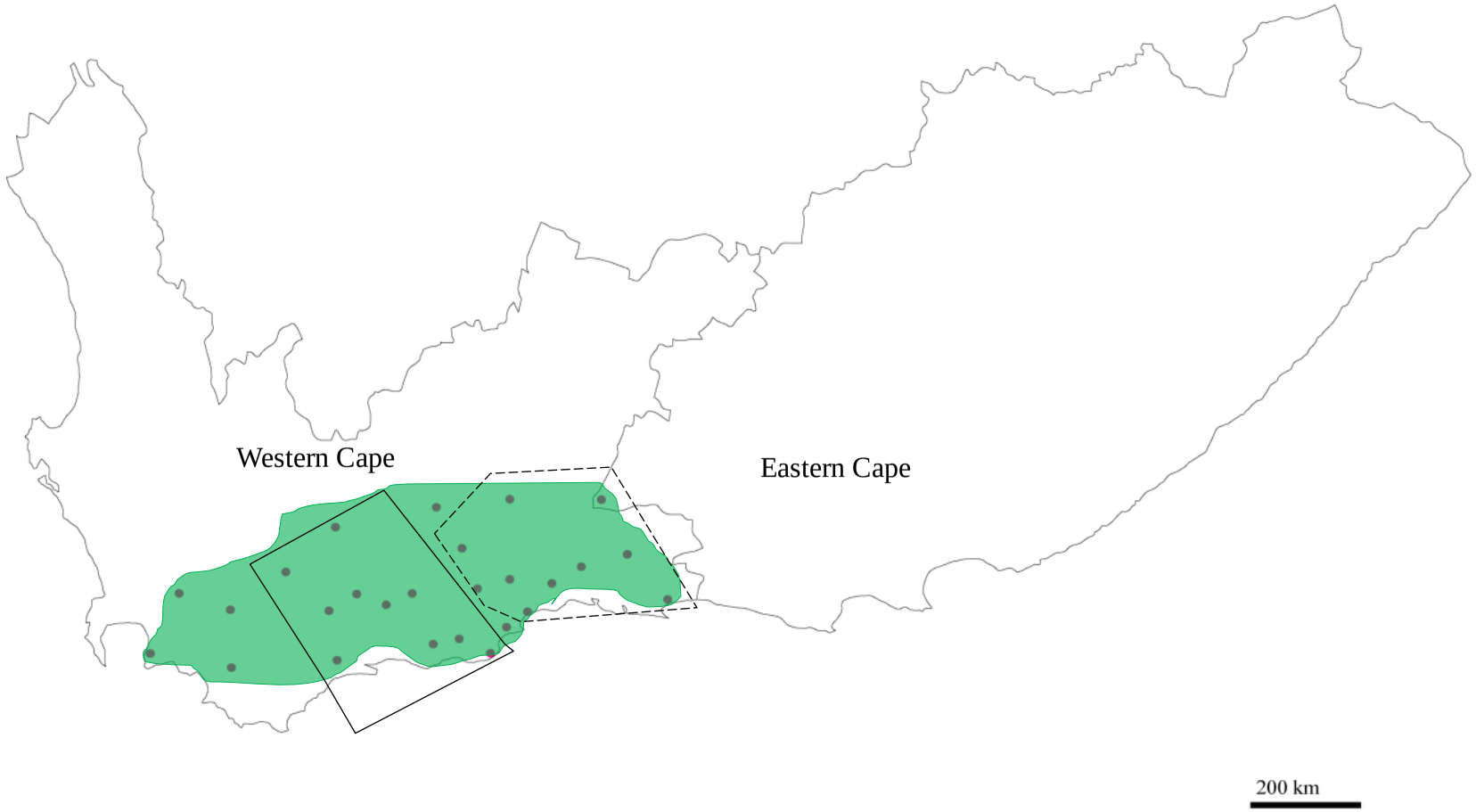

Fig. 3.4 *Post hoc* grouping of Central STRUCTURE cluster. Group A is outlined by a solid black line, Group B is outlined by a dashed line.

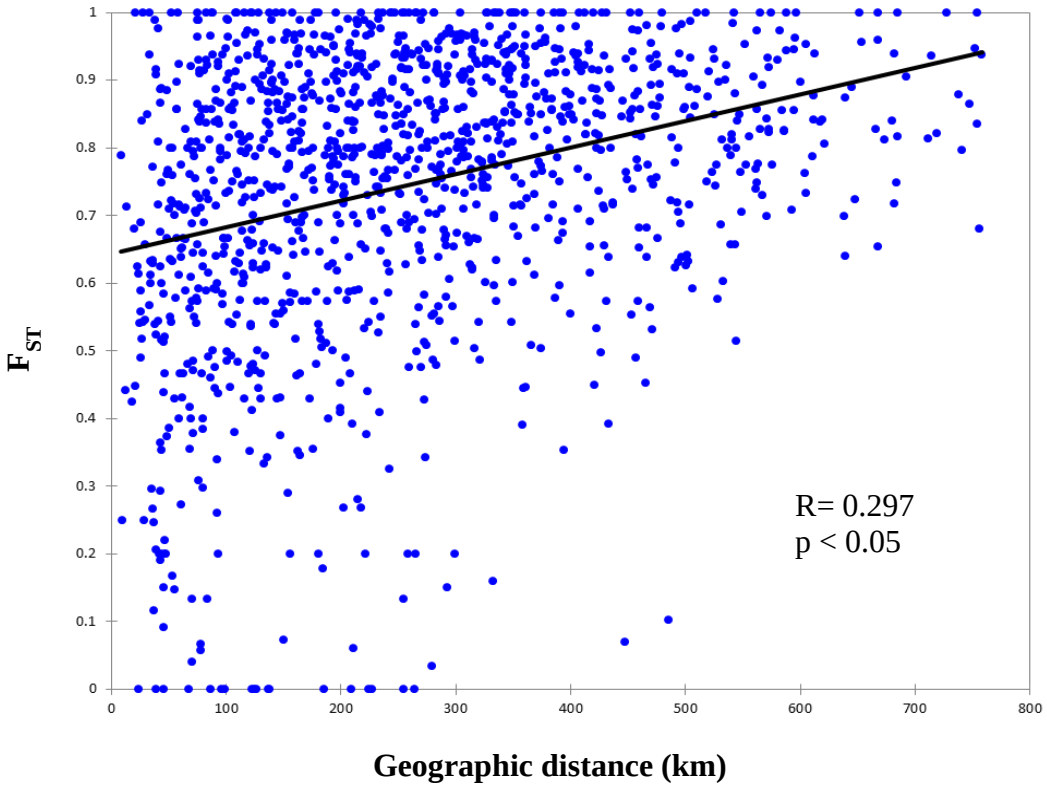

Fig. 3.5 Mantel residuals for the test of IBD for all sampled locations (N=51) using geographic distance.

Chapter 4

COMPARISONS OF SPATIAL AND TEMPORAL PATTERNS BETWEEN GENOMIC MARKERS FOR *LEUCADENDRON SALIGNUM*

Introduction

The Cape Floral Region (CFR) located in South Africa is considered the smallest of the six Floral kingdoms (Goldblatt 1978), with 9,000 vascular plant species, of which approximately 70% are endemic, all found in a region occupying only 90,000 km^2 (Goldblatt & Manning 2002; Linder 2003; Cowling & Proches 2005). The diversity of this biogeographical region is similar to that found only on islands (Linder 2003), yet the vegetation found in the CFR is largely encapsulated by five plant types: fynbos, renosterveld, thicket, forest, and succulent karoo. Of these vegetation types the fynbos and renosterveld are predominant in the CFR, with fynbos accounting for 80% of the vegetation in the region (Cowling 1983). The fynbos, meaning 'fine-leaved bush' in Afrikaans, is comprised of four vegetation types: restiod reed-like bushes, proteoid large leafed woody shrubs, ericoid fine leafed bushes, and geophytes that contain large underground storage organs (Cowling et al. 1996), and alone makes up 80% of the endemic plants found in the CFR.

Numerous hypotheses have been suggested to explain the richness and endemism of floral species in the CFR, with many studies focusing on patterns contributing to lineage and species diversification (Linder & Hardy 2004; Goldblatt & Manning 2002, Sauquet et al. 2009), however, an underlying factor that may lead to the speciation of the CFR is gene flow and the factors that may limit it between CFR plant populations (Linder 2003). It is thought that the diverse limitations to plant specific gene flow present in the

CFR (e.g. heterogeneous landscapes) predict the floral speciation in the Cape (Goldblatt 1996; Linder 2003). Gene flow maintains connectivity among individuals in populations, allowing genetic variation to be shared in a population by connecting organisms across the landscape; a reduction in gene flow will increase the probability of differentiation between populations (Slatkin 1987), and a permanent loss of gene flow can lead to speciation (Mayr 1947); therefore, understanding levels of gene flow within populations is paramount to understanding how diversity is generated at a population level.

Levels of gene flow between plant populations can differ due to a number of underlying factors, and these differences in gene flow can create or maintain diversity in CFR plant populations. Within plants, there are specific biological mechanisms directly related to gene flow, including, the dispersal of seeds and the dispersal of pollen (Ennos 1994). The levels of gene flow can vary between these two mechanisms depending on the distance of dispersal and can be measured directly through observations of pollen and seed movement or indirectly through estimates of population structure for the genomes associated with each mechanism (Slatkin 1985, 1989; Ennos 1994). Patterns of seed dispersal be apparent in maternally inherited organelles (Birky et al. 1989), whereas patterns of pollen dispersal will influence the nuclear genome. Within the CFR, the mechanism of seed dispersal is largely represented by two methods, passive dispersal and ant-mediated dispersal. Passive dispersal is a mechanism that is often associated with plants existing in nutrient poor soils (Goldblat 1996), and given that these soils are predominant throughout the CFR (Cowling & Proches 2005), it is unsurprising that passive dispersal is the common dispersal mechanism for plants here, distributing seeds around 5 m from the original plant. Another mechanism, ant mediated seed dispersal, is

unusually high in the CFR, as plants existing on such nutrient-poor soils will generally not produce seeds with the fatty bodies that attract ants (Bond & Slingsby 1983; Goldblatt 1996). Although this is an alternative to passive dispersal, this mechanism still only distributes seeds up to 6 m away from the maternal plant (Goldblatt 1996). Given the majority of seed dispersal mechanisms in the CFR are short ranged, this will likely impact levels of gene flow in plant species. Although seed dispersal may be limited, pollen dispersal can aid in connecting individuals or populations (Austerlitz et al. 2004; Ennos 1994). With respect to pollination and pollen dispersal for CFR plant species, wind pollination has been identified in two of the largest fynbos species, *Erica* and *Leucadendron* (Linder 2003; Barker et al. 2004), while anemophily is also very common in the Cape flora (Koutnik 1987). However, the apparent ability of pollen to disperse further than seeds in the CFR and thereby increasing gene flow across a larger range may likely maintain and introduce variation into plant populations. As a whole, pollination in the CFR is not significantly different from any other region in the world (Linder 1985) and therefore may not be a main contributor to the unique CFR species diversity; however, it is influential in mediating gene flow among populations.

The physical dispersal of plant gametes is one factor that contributes to gene flow in plant species, but in addition to pollen and seed dispersal, another factor we must also consider is the complex genetic makeup of plants that is associated with these dispersal patterns. Plants have three separate genomes: chloroplast (cpDNA), mitochondrial (mtDNA), and nuclear (nuDNA) and the mode of inheritance of these genomes varies: cpDNA is often maternally inherited (Conde et al. 1979; Ennos 1994) as is mtDNA (Neale et al. 1991; Palmer 1992) whereas nuDNA is bi-parentally inherited. Additionally,

the nuclear genome of many plants can be polyploidy (e.g. have multiple sets of chromosomes) due to duplication events. It has been noted, that a consequence of these different modes of inheritance is that the extent of gene flow can differ for each of these genomes, which may be reflected in patterns found in genetic markers associated with each mode of inheritance (e.g., Birky et al. 1983, 1989; Takahata & Palumbi 1985). For example, we may expect that nuclear markers reflect higher levels of gene flow since they are carried across populations via pollen, whereas chloroplast or mitochondrial markers are represented by seed dispersal and may show lower levels of gene flow due to more limited dispersal patterns (Ennos 1994). This may lead to different patterns of genetic variation and spatial genetic structure among a singular population depending on which genetic marker is examined (Birky et al. 1989). Additionally, the effective population sizes differ between these genomes, with cpDNA and mtDNA having an effective population size one quarter that of nuDNA (Wright 1931). Therefore, the maternally inherited cpDNA and mtDNA are more likely to experience the effects of genetic drift, which may quickly produce differences in these markers between populations. Lastly, mutation rates vary between these genomes with rates almost an order of magnitude higher in nuDNA than in cpDNA or mtDNA (Wolfe et al. 1987), allowing for more variation to accumulate within the nuDNA over time.

Studies of spatial genetic structure and gene flow for plant populations are infrequent in the CFR. Studies that measure gene flow indirectly speculate about the role of gene flow in speciation and population diversification (Goldblatt 1978; Bergh et al, 2010; Prunier & Holsinger 2010; Rymer et al. 2010; Segarra-Moragues et al. 2010). In a study using nuclear ISSR markers, Bergh et al. (2007) found that within a species of the

widely distributed renosterbos, *Elytropappus rhinocerotis*, 80% of the total variation was found among individuals in the CFR. They suspected that this result reflected high levels of recombination among ISSR loci, a bi-product of high levels of gene flow and outbreeding rates. In a study of locally endemic white protea species, Prunier & Holsinger (2010) used 10 microsatellite nuclear loci to determine the spatial genetic structure between species and levels of gene flow within species. They found evidence for little gene flow between sampled population, with eight genetic clusters identified, nearly concordant with the number of species tested. The only exception was a single species, *Protea mundii*, which was divided into two groups from the western and eastern capes. This result is in contrast to the Bergh et al. (2007) study, which found high levels of gene flow among a widespread plant, which suggests that gene flow is potentially correlated to plant ranges. Although these studies both used nuclear markers, the patterns are markedly different which may be a result of different dispersal mechanisms between the two plants, however, by only examining a single marker, these studies have not captured all potential avenues of gene flow in plants and any conclusion regarding dispersal cannot be made. It may be these patterns reflect the genetic markers used, the underlying dispersal associated with these markers, or potential influences from the CFR landscape. Investigating patterns correlated to both maternal and paternal markers may allow for tests for differences in dispersal between these plants, a perspective that cannot be obtained without such a study.

In a previous study, the spatial genetic diversity of the most widespread fynbos plant *Leucadendron salignum* was characterized using cpDNA across its entire range, both inside and outside the CFR (Chapter 2). *Leucadendron salignum* is a dioecious plant

that uses insect-mediated pollination and seed dispersal by ants (Barker et al. 2004; Williams 1972; Hattingh and Giliomee 1989). We were able to detect at least four major genetic cluster groups, with further evidence for hierarchical structure even between locations within close proximity within these major geographically clustered groups. In a second study (Chapter 3), we noted that this genetic diversity can be largely explained by resistance on the landscape due to vegetation type, which considers different aspects of the flora with respect to distribution of strata and landscape factors such as soil composition and fertility, and that this isolation can occur on a very fine-scale. However, this study only included cpDNA markers. Given arguments above, this one picture may limit our understanding of overall population structure and gene flow on temporal and spatial scales due its single mode of dispersal as well as a possible bias towards more recent events due to increased drift.

In *L. salignum,* if seed dispersal is local and if the cpDNA is experiencing drift within these local regions, we would expect to see patterns of spatial genetic structure within geographic regions as in our previous study. By comparing patterns of spatial genetic structure found in the nuclear genome to those found in cpDNA we can determine if the same pattern is occurring across both markers. If patterns of spatial genetic structure are the same, regardless of the marker used, then it is clear markers are not sensitive to evolutionary history, dispersal, or inheritance, and conclusions made from any marker would reflect the influence of factors on the landscape. If, however, patterns are different across markers, then these differences in evolutionary history, dispersal, and inheritance must be taken into consideration when making conclusions. Therefore, to understand to what extent the use of different genomes alters our perspective of spatial

and temporal genetic and evolutionary histories and patterns of gene flow in *L. salignum*, we present a genome-wide assay of nuclear diversity in these same *L. salignum* population samples and compare these patterns with those seen in the haploid cpDNA dataset.

Methods

Samples

To test the spatial genetic structure of *L. salignum*, leaves were collected from four individuals from 51 locations across South Africa, for a total of 204 samples representing the species' entire known distribution range, including samples on the Western and Eastern Cape (Fig. 4.1, Table 4.1). Leaves were immediately placed in powdered silica for preservation and stored at room temperature. Details on the molecular data collection protocols are found in Chapter 2.

Nuclear genetic loci generation

As no previous genomic data are available for *L. salignum*, we developed a method to identify and collect random DNA sequences from homologous genome-wide nuclear regions. A total of 8 ISSR primers (Table 4.2) designed for use in plants were used in pairwise PCR reactions as in Bergh et al. (2007). All ISSR PCR reactions were carried out in 25-µl reactions each containing 100 ng DNA from one *L. salignum* individual at first, 10X GoTaq Flexi buffer (supplied with Promega GoTaq Flexi DNA polymerase, Promega), 3.0 mM $MgCl_2$, 2.0 mM each dNTP, 10 µM of two of the ISSR primers, and 1 U Promega GoTaq flexi DNA polymerase (Promega). PCR amplifications were performed under the following conditions: 95 ˚C for 1 minute, followed by 39 cycles of 95˚C for 30 seconds; 39˚C for 1 minute, and finally 72˚C for 2 minutes and 30

seconds. PCR products were run on 1.2% Agarose gels at 120 volts for 60 minutes, and lanes with multiple bands ranging from ~500 to 1500 bp were chosen for cloning. Each PCR product was first cleaned using shrimp alkaline phosphate and exonuclease 1 (US Biochemicals, Cleveland, OH); products were then cloned using the TOPO® XL PCR Cloning kit (Invitrogen, Carlsbad, Ca) according to the manufacturer's instructions. To determine the sequence of each cloned fragment, single colonies were picked from agar media plates and separately placed in a PCR tube with 100 µL of ddH_20, mixed thoroughly and incubated at 95 ˚C for 5 minutes; 4 µL of this solution was then used as the template for a PCR reaction using the following conditions: 10 X buffer, 3.0 mM MgCl2, 2.0 mM each dNTP, 10 µM M13 Forward primer (Invitrogen) and 10 µM M13 Reverse primer (Invitrogen), and 1 U Promega GoTaq flexi DNA polymerase. PCR cycle conditions were run same as above. PCR products were cleaned using the same protocol as above, and nucleotide sequences were collected on an Applied Biosystems 3720 capillary sequencer.

Resulting sequences were examined for length variants, repeat elements and microsatellites, then run through the NCBI BLAST-n algorithm. One attempt to prune non-nuclear sequences from the dataset was made by removing DNA sequences matching previously identified mitochondrial or chloroplast regions. A second step was made to determine if regions had open reading frames (whether they could be identified in GenBank or not), and these were discarded as well in order to build a dataset composed of variation that most closely resembled "neutrally-evolving" markers of the genome. *Leucadendron salignum* specific PCR primers were designed from these initial fragments to amplify 20 independent fragments adhering to these criteria, resulting in 9

random regions that could be readily replicated in multiple individuals of *L. salignum*. Finally, all regions, when eventually amplified and sequenced in multiple individuals, were also checked for "heterozygous" sites - as well as Hardy-Weinberg and linkage equilibrium - for their inclusion in our putatively nuclear region dataset. In addition to these markers, conserved primers for the nuclear ITS region 18S-26S were used to generate a single fragment as in Baldwin and Markos (1998). The DNA fragments were amplified for each individual using the following conditions. All PCR amplifications were carried out in 25-µl reactions with 100 ng DNA, 1X PCR buffer (supplied with KAPA2G Robust *Taq* polymerase, KAPA Biosystems, Woburn, MA) 2.5 mM $MgCl_2$, 2.0 mM each dNTP, 10 µM forward and reverse primers, and 0.3 U KAPA2G Robust HotStart DNA Polymerase (KAPA). PCR amplification was performed under the following conditions for all loci: 95 ^{o}C for 1 minute, then 39 cycles of 95 ^{o}C for 30 seconds; 52 ^{o}C annealing and 78 ^{o}C for extension.

Illumina MiSeq sequencing preparation

After an initial dataset was compiled on individuals to check for the above criteria using the Applied Biosystems 3720 capillary sequencer, we collected the remaining nucleotide sequences via the Illumina MiSeq. For each individual, the concentration of each of the 10 amplicons was determined using a Nano Drop 1000 (Thermo Scientific). Amplicons were then pooled in equimolar concentrations for each individual, resulting in 192 pooled samples. The concentration of each pool was then measured using Agilent Technologies Bioanaylzer and then diluted to 0.2 ng/µl for barcoding and preparation. Using the Nextera XT DNA sample preparation kit, according to the manufacturer's instructions, 96 individuals were uniquely barcoded per run. Upon completion of sample

preparation, all 96 barcoded samples were pooled and sent to Ambry Genetics (Aliso Viejo, Ca) to be run on their Illumina MiSeq via standard protocol.

Data processing

The bioinformatics process of SNP calling of MiSeq data were performed using CLC Genomics Workbench v. 5.5 (CLC Bio, Cambridge, Ma) implemented by Ambry Genetics. Data sets were combined by aligning the FASTA files of Illumina data with the trace files from capillary sequences in Sequencher. All files were aligned and trimmed to equal lengths. FASTA files for the combined data sets were used to create two haplotypes for each individual, where nucleotides from heterozygous sites were assigned randomly to one of two haplotypes using the script PhaseSeqs (pers comm, M. S. Rosenberg) to be used in all downstream analyses.

Population Differentiation

Unless otherwise noted, all population genetic statistics were computed using DnaSP v 5.5 (Rozas et al. 2003). Estimates of diversity based on the number of segregating sites (S), corrected by sample size, were calculated using Wattersons's θ (1975), as were estimates of π, which calculates diversity based on the average pairwise differences among sequences. Under neutrality, these two estimates are expected to be equal, and this hypothesis was tested using Tajima's (1989) test to calculate per marker single nucleotide frequency spectra, as well as to generate a distribution of values across markers (as we have no a priori view of what this "neutral" distribution looks like across the genome). Values of Tajima's D can indicate both natural selection and demographic history, with negative D values associated with directional selection or population expansions and positive D values with balancing selection or population structure.

Comparing D values across all loci can help identify unusual markers (i.e., outliers) in an attempt to reflect the genome-wide demographic history.

The spatial genetic structure across *L. salignum* locations was determined previously for the chloroplast (cpDNA) genome (Chapter 2). In order to compare patterns of spatial genetic structure between genomes, genetic structure was estimated using the program Structure 2.3.3 (Pritchard et al. 2000, Falush et al. 2003, Falush et al. 2007, Hubisz et al. 2009), which uses an unbiased individual based clustering approach to partition variation across sampled locations. All individuals were assigned to separate "population" samples and run using the assumption K = 2-20; with 5 replicates of 10^6 generations and a burn-in of 10^4 generations. Results were analyzed using the program Structure Harvester (Earl 2012) which both uses the Evanno method (Evanno et al. 2005) to evaluate the appropriate number of clusters and produces the files necessary to visualize the data using the programs CLUMP 1.1.2 (Jacobsson & Rosenberg 2007) and DISTRUCT 1.1 (Rosenberg 2004).

Following the use of STRUCTURE, a secondary analysis was performed to estimate potential inbreeding given the sessile nature of plants using InStruct (Gao et al. 2007). This program uses an approach similar to STRUCTURE to determine population clusters; however, InStruct also takes into consideration inbreeding among samples. InStruct was run to infer population structure and individual inbreeding coefficients, with 100,000 Markov chain Monte Carlo (MCMC) iterations, a burn-in of 10000, thinning of 10, and testing for k groups of 1–5 with five separate chains for each K grouping. The individual runs were combined using the program CLUMP 1.1.2, and a final output was created using DISTRUCT 1.1.

To determine how variance was partitioned within and among groups, an F_{ST} analysis (e.g., Wright 1951) was conducted with 10,000 permutations calculated to assess significant differences (from a model of F_{ST} =0) using ARLEQUIN (Excoffier et al. 2005). Three different groupings were assessed, first samples were grouped by sampling location to assess the average variation between locations and second, samples were grouped geographically by their presence in either the Western or Eastern Cape. Lastly samples were grouped to compare with the cpDNA patterns, specifically, they were based on previous STRUCTURE clusters (Chapter 2). These statistics can also serve as an estimation of migration between tested groups.

Divergence time estimates

Given the limited information available for *L. salignum*, the flexible program BEAST 1.5.6 (Drummond and Rambaut 2007; Drummond et al. 2012) was used to determine the temporal relationships among samples. DNA sequence data were run through MODELTEST (Posada and Crandall, 1998) to determine the appropriate nucleotide substitution model for the *L. salignum* SNP dataset. The sequence divergence range was estimated using the conservative value of 30×10^{-8} substitutions per site per year (Wolfe et al. 1987). The Markov Chain Monte Carlo (MCMC) analysis was performed using a relaxed uncorrelated lognormal clock, using the Bayesian skyline population models. The BEAST profile was run five times for 30 million generations with the first 3,000,000 discarded as burnin, logging ever 3000. All runs were viewed in the program TRACER 1.5 (Drummond et al. 2012) to check for convergence of the chains and estimate parameter distributions. All log files and tree files were combined using LogCombiner and tree files were annotated using TreeAnnotator v.1.6.1

(Drummond et al. 2012), which produced a consensus tree for these analyses. As a comparison to these BEAST results, a neighbor-joining tree was also created in MEGA v. 5 (Tamura et al. 2011).

Results

Nucleotide Diversity

These data were obtained primarily to estimate neutral levels of variation and patterns of geographic and temporal structure for the nuclear genome for comparison to data obtained for cpDNA in *L. salignum.* This nuclear data set represents the first large-scale nuclear data set for any South African plant species. A total of 5600 KB of nuclear DNA sequence was collected for each of the *L. salignum* individuals resulting in a total of 616 SNPs. Population genetic analyses detected some variance in SNP diversity (Table 4.3), with the average values measuring as $\theta_\pi = 0.005$ and $\theta_s = 0.01$ across all loci. Using Tajima's (1989) test, similar negative values were observed across all loci, with an average value of D = -1.79, which is non-significant (compared to 0) and overall these unlinked random markers reflect neutrality. Analyses of θ_π and θ_s performed between cluster groups identified previously by STRUCTURE analyses on cpDNA (Chapter 2) showed similar D values as compared to the overall population sample (Table 4.3). After adjusting for effective population size in the cpDNA dataset (i.e., nuclear sequences expected to have an effective population size four times that of the haploid genome), values of of θ_π and θ_s were similar for both marker types (Table 4.3). In comparing these data to the previously generated cpDNA dataset we can associate the different spatial and temporal scales between markers (Anderson et al. 2010, Wolfe et al. 1987) and how those

inherent differences may influence analyses of nucleotide diversity and historical demographics for the sampled *L. salignum* population across the CFR landscape.

Population Genetic Structure

Structure results are presented here for K= 2 to 4 to illustrate the lack of diversity across samples (Fig. 4.2). These results suggest that we cannot reject K > 1, as it appears that we have little evidence to suggest that the nuclear diversity is consistent with anything but a single population sample. This is in contrast to the cpDNA dataset where K is at least four, and evidence for hierarchical structure indicates possibly more. Results from InStruct were similar, with evidence for no more than one cluster. Levels of individual inbreeding coefficients (F) within the sample were detected between 0.025-0.880, with an average of F= 0.307. The variance of this estimator is likely the result of small samples within locations; thus, the overall estimate is not unusually high and is consistent with the lack of isolation seen from the STRUCTURE analysis.

Although statistically significant, levels of genetic variation among all locations as measured by F_{ST} were low in the *L. salignum* nuclear dataset (F_{ST} =0.14, p <0.0001; Table 4.4), compared to values obtained for the overall cpDNA data set (F_{ST} = 0.78, p<0.001). This analysis was also repeated among groups in the nuclear dataset that were binned based on the group analyses in the cpDNA as a comparison. Levels of genetic variation measured between samples located on the Western and Eastern capes were very low, although again, significant (F_{ST} = 0.03, p<0.0001), compared to levels of genetic variation found in our previous cpDNA study (F_{ST}=0.43, p<0.001). Additional analyses measuring genetic variation for the STRUCTURE clusters all showed lower levels of variation ranging from F_{ST} = 0.08 to 0.10, p<0.00001, than those detected in our cpDNA

study (Table 4.5 which found values between F_{ST} = 0.50 to 0.73, $p<0.001$. Altogether, these analyses show much lower levels of genetic variation among locations, both at far and near proximities in the nuclear dataset than in the cpDNA dataset.

Temporal Genetic Structure

Given the low levels of genetic diversity within these samples, there was little differentiation and phylogenetic signal in the nuDNA dataset compared to the cpDNA dataset. Both trees in MEGA and BEAST failed to resolve clusters seen in the cpDNA dataset. In fact, where we had previously identified a monophyletic branch for the East, here all individuals previously associated with that cluster were found distributed throughout the tree. Similarly, the branches associated with the North, South West, and Central groups disappeared and individuals found from these locations were placed in random places throughout the tree (Fig. 4.3). Although branch patterns were inconsistent from our cpDNA tree, the age of the nuDNA tree was consistent, with an estimated root around 1.7 MYA.

Discussion

Our previous study of spatial genetic structure using the cpDNA markers showed significant spatial genetic structure and genetic diversity for *L. salignum*, including at least four structure clusters, with evidence for even more in the detection of hierarchical structure within these groups. In contrast, this study using nuDNA showed patterns of panmixia among populations, with no evidence for spatial genetic structure. The different patterns observed between these two data sets indicate that these markers do not provide concordant patterns of spatial genetic structure and genetic diversity for *L. salignum*; therefore, they may be influenced by underlying factors affecting the markers themselves

such as dispersal, inheritance, or evolutionary history which are unique to each of these markers.

Tests of marker neutrality indicated that both marker sets approximate neutrality and reflect similar evolutionary patterns. The levels of nucleotide diversity measured by θ_S were similar between the two data sets, after adjusting for the difference in effective population sizes, indicating that genetic diversity reflected by sequence data are comparable. Measures of Tajima's (1989) D showed similar negative values between all amplicons, as well as between the cpDNA and nuDNA data sets, indicating these markers all reflect similar demographic histories. The underlying measures of genetic mutations are consistent and comparable across both data sets, however, given that patterns of spatial genetic structure are not consistent between data sets, additional forces including the evolutionary history associated with each marker genome or gamete dispersal may be contributing to patterns of spatial genetic structure.

Previous studies indicate that cpDNA is maternally inherited and does not recombine in *L. salignum* (Pharmawati et al. 2004); it has an effective population size that is smaller than the nuclear genome (Wright 1931) and reflects seed dispersal (Ennos 1994). Therefore, this marker likely reflects the effects of drift more so than a nuclear marker. As mentioned previously, drift is more likely to create spatial genetic structure on local levels for our markers; therefore the pattern of spatial genetic structure found in the cpDNA dataset may be due to this driving force alone. Ennos (1994) suggested that gene flow will likely be higher in pollen than maternally inherited markers and population differentiation is expected to differ between these markers (Birky et al. 1989). The bi-parental nature of nuDNA reflects recombination and a much larger effective

population size, and it also represents pollen dispersal in our organism of study here. Given the data, it appears that patterns from evolutionary history are different between these two markers and must be taken into consideration when drawing conclusions about how the landscape affects the *overall* population genetic structure in *L. salignum.*

An alternative to the evolutionary history driving these patterns is that the dispersal associated with each of the two genome types is contributing to the patterns of spatial genetic structure found in each data set. The lack of a spatial genetic pattern observed for the nuDNA data set suggest that, in *L. salignum* pollen is widely dispersed, whereas the highly structured cpDNA data suggest seed dispersal is local. If pollen dispersal were local, we would expect to see similar spatial genetic structure patterns as in cpDNA, however, these results show conflicting patterns, so it is likely that pollen is dispersing at a level that is maintaining connectivity across the entire population over time, and therefore maintaining the high levels of shared variation evident in the nuDNA data. Interestingly, in our previous study, there was a positive association between vegetation type in the fynbos and patterns of cpDNA spatial genetic structure; it may be that vegetation acts to facilitate seed dispersal in *L. salignum* but has little influence on pollen dispersal. Specifically, on local scales, the heterogeneity of the landscape as measured by vegetation type predicts seed dispersal patterns in *L. salignum,* while pollen is maintaining connectivity amongst locations on possibly broad and fine-scales.

The interaction between seed and pollen dispersal is noteworthy, as previous CFR studies have indicated limited seed dispersal for fynbos plants (Goldblatt 1996). The limited dispersal of seeds is likely generating local diversity in associated maternal genomes and without the influx of variation from pollen, these local regions could

become genetically different from one another very quickly. Since it seems as though pollen maintains variation across the landscape, any change resulting in a reduction of pollen dispersal could likely result in eventual speciation. Although we did not directly measure levels of pollen dispersal, understanding how both pollen and seed dispersal contribute to genetic diversity on the landscape is essential when studying plant species.

Our study suggests that *L. salignum* is a genetically diverse fynbos plant, with levels of spatial genetic structure more apparent in cpDNA than nuDNA. Some limitations of this study may include our sample collection of eight chromosomes (i.e. marker sequences) per location, which suggests that variants at less than ~13% (i.e., 1 of 8) within each of these locations would not be identified. However, while additional samples at each location would help detect these rare variants, it is unlikely to uncover many common variants to come even close to the patterns seen among the cpDNA markers. Thus, the most one could say is that subtle differences in nuDNA among our locations, if they exist, are generating some population structure. It is worth noting that differences are not even found at large scales as no IBD model fits diversity between locations separated at the most extremes of our sampling. On the other hand, with fewer individuals per location and fewer variants, we find non-subtle cpDNA variation among locations at not only broad scales across the Western and Eastern Capes, but also within locations in close proximity. Thus, it is clear that cpDNA simply exhibits much more population structure than nuDNA and overall reflects a very different evolutionary history of gene flow across the Western and Eastern Capes. That said, although *L. salignum* is the most widespread fynbos plant in the CFR and sheds insight into how landscape features and genetic markers can influence patterns of diversity, it still is an

"N" of 1. Therefore, it is essential that population genetic data be collected from other CFR plants with varying dispersal and geographic ranges. This will allow for comparisons of patterns of structure and gene flow to see if and how similar forces are driving and maintaining genetic diversity across the landscape.

Conclusion

Using only haploid markers will result in patterns that are likely associated with the evolutionary history of that marker and dispersal of seedd and not necessarily the contributions of the landscape. Likewise, using only nuclear markers may result in an overall picture, but the influences of the maternal lineage through seed dispersal will not be found. Using both haploid and nuclear markers will help determine if pollen and seed dispersal are similar and if the landscape is affecting the patterns of gene flow in both markers. It is clear from this study each genome marker can give decidedly different pictures of gene flow and consequently different conclusions regarding the importance of landscape in the CFR. Conservation efforts need to consider perspectives from both genomes when designing conservation plans because pollen dispersal is maintaining connectivity among populations on broad scales, however, it appears the landscape is creating diversity on local scales by affecting seed dispersal. Therefore, the loss of vegetation associated with these patterns in cpDNA may result in the loss of diversity in *L. salignum* and potentially in the CFR as a whole. This may make it necessary to focus on these smaller areas when designing conservation plans, even though nuDNA is still maintaining diversity at broad scales.

Table 4.1 Names and geographic coordinates for all sampled *L. salignum* locations

	Abbreviation	Location Name	latitude	longitude
1	VRP	Vanrhyns Pass	-31.37087	19.01587
2	GIP	Gifberg Pass	-31.76927	18.76999
3	PAP	Pakhuis Pass	-32.14442	19.02492
4	MIP	Middelberg Pass	-32.63051	19.15221
5	VEP	Versfeld Pass	-32.84253	18.73164
6	GYP	Gydo Pass	-33.23595	19.33648
7	WOR	Worcester	-33.60255	19.33463
8	GBY	Grotto Bay	-33.52580	18.35388
9	PHI	Philadelphia	-33.71750	18.54403
10	CPT	Cape Town	-33.91698	18.40488
11	SVR	Silvermine	-34.08278	18.41490
12	SIT	Simons Town	-34.20063	18.41121
13	SWB	Smitswinkel Bay	-34.26193	18.46104
14	PRB	Pringle Bay	-34.31126	18.83138
15	STL	Stellenbosch	-33.93238	18.87687
16	FRP	Franschhoek Pass	-33.91495	19.15701
17	GRY	Greyton	-34.03345	19.60752
18	STA	Stanford	-34.41032	19.58887
19	AGU	Agulhas	-34.66967	19.77530
20	DHP	DeHoop	-34.37937	20.52973
21	BTK	Bontebok	-34.04368	20.46947
22	KOP	Kogmanskloof Pass	-33.80377	20.10572
23	TRP	Tradouws Pass	-33.93706	20.71161
24	HID	Heidelberg	-34.01497	20.96603
25	GAP	Garcia Pass	-33.94164	21.20187
26	STB	Still Bay	-34.26279	21.37145
27	ALB	Albertinia	-34.22925	21.59360
28	VLB	Vlees Bay	-34.34412	21.86730
29	MOB	Mossel Bay	-34.16592	22.00785
30	GRB	Grootbrak	-34.06319	22.20193
31	CLP	Cloete Pass	-33.93095	21.76138
32	ROP	Robinsons Pass	-33.86390	22.02835
33	OUP	Outeniqua Pass	-33.88690	22.39971
34	DAS	Daskop	-33.76938	22.65600
35	RBP	Rooiberg Pass	-33.65101	21.63852
36	SEW	Seweweekspoort	-33.38633	21.40805
37	ANB	Anysberg	-33.47264	20.58203
38	SWP	Swartberg Pass	-33.35097	22.04737
39	BLB	Blesberg	-33.40775	22.73231
40	POP	Potjiesberg Pass	-33.70272	23.04364
41	PLB	Plettenberg Bay	-34.01325	23.38842
42	MIS	Misgund	-33.76043	23.48274
43	KOU	Kougaberg	-33.67525	23.50330
44	JOB	Joubertina	-33.82058	23.85400
45	NOT	Nooitgedacht	-33.82692	24.25445
46	AGS	Assegaaibos	-33.93455	24.30351
47	BAV	East Baviaanskloof	-33.63444	24.46917
48	HUM	Humansdorp	-33.95992	24.76670
49	SHB	Stinkhoutberg	-33.81534	24.95069
50	SUP	Suurberg Pass	-33.28263	25.72010
51	GRT	Grahamstown	-33.34028	26.51658

Table 4.2 ISSR primers used in PCR reactions

ISSR 812	$(GA)_8A$
ISSR 825	$(AC)_8T$
ISSR 830	$(TG)_8G$
ISSR 835	$(AG)_8YC$
ISSR 841	$(GA)_8YC$
ISSR 846	$(CA)_8RT$
ISSR 855	$(AC)_8YT$
ISSR 856	$(AC)_8YA$

Table 4.3 Summary statistics of Nucleotide Variation

Marker	Length (bp)	S [a]	θ_S [b]	θ_π [c]	D [d]
18s26s	570	73	0.019	0.005	-2.193
Amplicon 1	660	78	0.018	0.004	-2.199
Amplicon 2	600	113	0.029	0.009	-2.019
Amplicon 3	521	47	0.014	0.005	-1.717
Amplicon 4	465	50	0.016	0.007	-1.571
Amplicon 5	393	42	0.016	0.005	-1.859
Amplicon 6	600	83	0.021	0.004	-2.235
Amplicon 7	540	81	0.023	0.011	-1.452
Amplicon 8	576	44	0.012	0.005	-1.608
Amplicon 9	697	5	0.001	0.0004	-1.12
All Amplicons	5622	616	0.017	0.005	-2.06
cpDNA	4649	88	*0.012	0.002	-0.9077

[a] Total number of SNPs

[b] Watterson's θ (1975)

[c] Average number of pairwise differences between sequences

[d] Tajimia's D statistic

*value adjusted to represent 4Ne

Table 4.4 Pairwise F_{ST} values for all sample locations using nuDNA

	VRP	GIP	PAP	MIP	VEP	GYP	WOR	BOK	PHI	CAT	SVR	SIT	TAB	PRB	STL	FRP	GRY	STA	AGU	DHP	BTK	KOP	TRP	HID	GAP	STB	ALB	VLB	MOB	GRT	CLO	ROP	OUP	DAS	RBP	SEW	ANB	SWP	BLB	POP	PLB	MIS	KOU	JOB	NOT	AGS	BAV	HDP	SHB	SUP	GRW
VRP	-																																																				
GIP	0.07	-																																																		
PAP	0.17	0.17	-																																																	
MIP	0.12	0.14	0.07	-																																																
VEP	0.14	0.13	0.11	0.05	-																																														
GYP	0.08	0.07	0.15	0.06	0.03	-																																													
WOR	0.10	0.13	0.08	0.09	0.12	0.09	-																																													
BOK	0.09	0.11	0.07	0.05	0.08	0.05	0.02	-																																											
PHI	0.09	0.11	0.09	0.08	0.09	0.09	0.02	0.01	-																																										
CAT	0.07	0.02	0.09	0.11	0.14	0.09	0.05	0.07	0.05	-																																									
SVR	0.11	0.16	0.06	0.04	0.07	0.10	0.03	0.02	0.00	0.06	-																																							
SIT	0.10	0.06	0.05	0.09	0.11	0.10	0.03	0.04	0.05	0.00	0.05	-																																						
TAB	0.15	0.16	0.09	0.08	0.07	0.13	0.07	0.07	0.04	0.10	0.03	0.09	-																																					
PRB	0.11	0.13	0.08	0.08	0.15	0.08	0.01	0.01	0.02	0.06	0.05	0.04	0.13	-																																			
STL	0.14	0.16	0.06	0.10	0.13	0.15	0.03	0.06	0.05	0.08	0.05	0.06	0.09	0.06	-																																			
FRP	0.09	0.07	0.05	0.06	0.08	0.03	0.01	-0.01	0.04	0.00	0.02	0.00	0.06	0.00	0.09	-																																		
GRY	0.08	0.11	0.04	0.03	0.04	0.04	0.02	-0.01	0.02	0.05	0.02	0.03	0.05	0.02	0.05	-0.01	-																																	
STA	0.14	0.14	0.08	0.12	0.17	0.17	0.07	0.09	0.07	0.06	0.08	0.05	0.11	0.09	0.08	0.06	0.05	-																																
AGU	0.11	0.13	0.07	0.08	0.11	0.11	0.08	0.02	0.05	0.07	0.07	0.05	0.09	0.05	0.09	0.01	-0.01	0.08	-																															
DHP	0.10	0.07	0.05	0.05	0.07	0.05	0.03	0.00	0.05	0.01	0.03	0.02	0.08	0.02	0.08	-0.04	0.01	0.04	0.03	-																														
BTK	0.12	0.11	0.10	0.11	0.16	0.13	0.05	0.02	0.03	0.04	0.08	0.04	0.08	0.05	0.07	0.01	0.05	0.11	0.02	0.03	-																													
KOP	0.13	0.20	0.11	0.10	0.18	0.15	0.03	0.04	0.04	0.10	0.05	0.06	0.16	0.02	0.09	0.03	0.03	0.07	0.05	0.03	0.08	-																												
TRP	0.07	0.07	0.11	0.10	0.11	0.07	0.03	0.05	0.06	0.03	0.08	0.01	0.09	0.04	0.10	0.00	0.03	0.05	0.06	0.02	0.05	0.08	-																											
HID	0.05	0.05	0.11	0.09	0.10	0.06	0.04	0.00	0.04	0.03	0.07	0.03	0.10	0.03	0.09	-0.01	0.01	0.08	0.03	0.00	0.03	0.08	0.03	-																										
GAP	0.13	0.11	0.12	0.11	0.11	0.11	0.07	0.06	0.08	0.08	0.10	0.05	0.13	0.05	0.11	0.03	0.04	0.13	0.06	0.05	0.04	0.10	0.04	0.06	-																									
STB	0.13	0.16	0.05	0.09	0.11	0.14	0.04	0.02	0.02	0.07	0.05	0.03	0.07	0.04	0.05	0.03	0.03	0.04	0.04	0.02	0.04	0.05	0.04	0.04	0.06	-																								
ALB	0.19	0.21	0.16	0.16	0.24	0.20	0.06	0.11	0.10	0.12	0.15	0.12	0.17	0.05	0.10	0.10	0.09	0.13	0.12	0.10	0.09	0.12	0.10	0.10	0.11	0.10	-																							
VLB	0.11	0.14	0.07	0.08	0.10	0.10	0.04	0.06	0.04	0.06	0.05	0.04	0.07	0.07	0.07	0.03	0.03	0.02	0.09	0.02	0.10	0.08	0.05	0.05	0.13	0.03	0.12	-																						
MOB	0.14	0.14	0.10	0.14	0.18	0.16	0.07	0.06	0.07	0.09	0.10	0.02	0.13	0.05	0.09	0.05	0.07	0.10	0.08	0.07	0.06	0.08	0.06	0.06	0.06	0.02	0.08	0.10	-																				
GRT	0.19	0.18	0.11	0.16	0.22	0.21	0.09	0.10	0.11	0.11	0.13	0.08	0.14	0.10	0.09	0.08	0.09	0.12	0.09	0.10	0.06	0.14	0.10	0.09	0.10	0.06	0.06	0.13	0.04	-																		
CLO	0.14	0.13	0.05	0.08	0.15	0.12	0.02	0.01	0.05	0.05	0.06	0.00	0.09	0.03	0.07	-0.01	0.01	0.05	0.02	0.01	0.01	0.05	0.02	0.02	0.04	0.01	0.07	0.05	0.03	0.05	-																	
ROP	0.17	0.18	0.09	0.14	0.17	0.19	0.06	0.07	0.05	0.10	0.08	0.06	0.09	0.10	0.07	0.09	0.07	0.07	0.10	0.08	0.08	0.07	0.09	0.12	0.11	0.03	0.12	0.06	0.07	0.09	0.08	-																
OUP	0.17	0.20	0.09	0.12	0.17	0.17	0.04	0.05	0.06	0.11	0.08	0.06	0.12	0.05	0.09	0.05	0.06	0.09	0.05	0.04	0.06	0.06	0.07	0.07	0.09	-0.01	0.10	0.09	0.05	0.07	0.01	0.07	-														
DAS	0.18	0.16	0.10	0.15	0.19	0.14	0.02	0.08	0.09	0.08	0.13	0.04	0.13	0.06	0.12	0.01	0.08	0.09	0.09	0.04	0.07	0.10	0.03	0.05	0.12	0.06	0.13	0.06	0.09	0.11	0.02	0.11	0.04	-												
RBP	0.17	0.17	0.10	0.14	0.11	0.16	0.04	0.08	0.07	0.10	0.10	0.07	0.09	0.09	0.11	0.06	0.06	0.07	0.10	0.06	0.10	0.11	0.03	0.08	0.07	0.00	0.14	0.07	0.08	0.10	0.05	0.06	0.03	0.05	-											
SEW	0.18	0.21	0.15	0.14	0.18	0.19	0.11	0.11	0.07	0.13	0.11	0.13	0.12	0.10	0.00	0.13	0.10	0.14	0.11	0.12	0.09	0.15	0.12	0.12	0.15	0.07	0.11	0.11	0.13	0.09	0.10	0.12	0.12	0.16	0.14	-										
ANB	0.15	0.21	0.13	0.13	0.16	0.17	0.06	0.08	0.04	0.14	0.09	0.10	0.10	0.09	0.12	0.10	0.09	0.13	0.11	0.08	0.10	0.12	0.10	0.07	0.14	0.05	0.16	0.09	0.12	0.16	0.08	0.11	0.10	0.11	0.08	0.13	-									
SWP	0.11	0.14	0.13	0.10	0.15	0.14	0.04	0.06	0.05	0.08	0.08	0.07	0.11	0.06	0.09	0.05	0.07	0.11	0.09	0.05	0.07	0.10	0.05	0.03	0.10	0.03	0.11	0.08	0.07	0.10	0.08	0.09	0.06	0.07	0.07	0.12	0.07	-								
BLB	0.23	0.20	0.22	0.20	0.22	0.20	0.15	0.16	0.14	0.19	0.22	0.18	0.19	0.14	0.20	0.16	0.16	0.22	0.17	0.17	0.12	0.22	0.16	0.14	0.14	0.17	0.17	0.21	0.18	0.19	0.15	0.21	0.16	0.20	0.18	0.21	0.21	0.18	-							
POP	0.15	0.16	0.09	0.12	0.14	0.16	0.04	0.05	0.00	0.08	0.04	0.04	0.03	0.07	0.06	0.04	0.04	0.07	0.08	0.06	0.04	0.07	0.07	0.08	0.09	0.03	0.12	0.05	0.08	0.09	0.04	0.03	0.04	0.06	0.05	0.09	0.08	0.06	0.17	-						
PLB	0.19	0.22	0.07	0.11	0.14	0.19	0.03	0.06	0.06	0.12	0.05	0.07	0.09	0.09	0.07	0.07	0.04	0.09	0.09	0.06	0.10	0.08	0.09	0.10	0.10	0.00	0.10	0.07	0.07	0.08	0.03	0.03	-0.01	0.08	0.03	0.09	0.08	0.08	0.21	0.04	-					
MIS	0.18	0.20	0.14	0.15	0.20	0.19	0.06	0.07	0.08	0.11	0.10	0.09	0.13	0.08	0.12	0.08	0.05	0.12	0.10	0.07	0.08	0.09	0.08	0.08	0.08	0.05	0.09	0.13	0.07	0.07	0.03	0.09	0.03	0.11	0.07	0.12	0.11	0.09	0.18	0.05	0.01	-				
KOU	0.15	0.15	0.13	0.14	0.15	0.15	0.09	0.08	0.06	0.13	0.13	0.09	0.14	0.07	0.10	0.10	0.05	0.14	0.04	0.09	0.06	0.12	0.08	0.05	0.10	0.07	0.12	0.12	0.07	0.09	0.06	0.12	0.08	0.11	0.11	0.11	0.09	0.09	0.16	0.09	0.10	0.04	-			
JOB	0.11	0.13	0.14	0.11	0.14	0.11	0.11	0.04	0.08	0.12	0.10	0.12	0.17	0.07	0.12	0.08	0.06	0.12	0.08	0.08	0.12	0.08	0.10	0.07	0.12	0.11	0.16	0.15	0.13	0.18	0.12	0.17	0.16	0.19	0.18	0.15	0.15	0.13	0.18	0.15	0.17	0.14	0.09	-		
NOT	0.17	0.19	0.13	0.13	0.14	0.17	0.08	0.07	0.05	0.13	0.10	0.09	0.10	0.07	0.10	0.06	0.05	0.14	0.04	0.07	0.04	0.12	0.09	0.05	0.09	0.04	0.14	0.10	0.09	0.10	0.04	0.11	0.04	0.10	0.08	0.09	0.08	0.09	0.18	0.04	0.06	0.06	0.02	0.13	-		
AGS	0.16	0.16	0.08	0.14	0.20	0.18	0.08	0.06	0.09	0.06	0.11	0.05	0.13	0.08	0.09	0.07	0.06	0.11	0.05	0.09	0.04	0.10	0.09	0.08	0.09	0.04	0.10	0.13	0.05	0.06	0.03	0.09	0.05	0.11	0.10	0.13	0.13	0.10	0.17	0.08	0.06	0.04	0.06	0.14	0.09	-		
BAV	0.16	0.18	0.14	0.16	0.18	0.18	0.07	0.05	0.09	0.12	0.10	0.08	0.11	0.12	0.12	0.09	0.06	0.15	0.10	0.08	0.08	0.15	0.07	0.06	0.09	0.05	0.14	0.14	0.07	0.06	0.07	0.11	0.06	0.13	0.09	0.14	0.11	0.07	0.21	0.09	0.06	0.00	0.06	0.15	0.08	0.05	-		
HDP	0.21	0.23	0.25	0.20	0.25	0.23	0.19	0.18	0.18	0.22	0.21	0.20	0.22	0.20	0.22	0.19	0.17	0.24	0.19	0.18	0.19	0.23	0.18	0.16	0.20	0.18	0.20	0.22	0.20	0.20	0.20	0.21	0.21	0.23	0.22	0.21	0.22	0.18	0.26	0.21	0.21	0.19	0.18	0.21	0.21	0.21	0.18	-		
SHB	0.19	0.20	0.17	0.17	0.20	0.21	0.09	0.09	0.08	0.14	0.12	0.11	0.16	0.10	0.12	0.11	0.12	0.15	0.12	0.10	0.10	0.14	0.12	0.09	0.12	0.06	0.15	0.14	0.10	0.13	0.10	0.08	0.07	0.14	0.10	0.14	0.10	0.08	0.22	0.09	0.06	0.05	0.05	0.16	0.08	0.09	0.08	0.20	-	
SUP	0.18	0.20	0.16	0.17	0.19	0.22	0.09	0.09	0.07	0.13	0.13	0.09	0.15	0.13	0.12	0.12	0.09	0.13	0.12	0.09	0.09	0.13	0.09	0.09	0.12	0.03	0.17	0.13	0.11	0.11	0.10	0.07	0.08	0.14	0.07	0.13	0.08	0.06	0.23	0.09	0.06	0.06	0.09	0.16	0.08	0.08	0.06	0.20	0.05	-	
GRW	0.24	0.27	0.25	0.23	0.30	0.28	0.22	0.19	0.17	0.22	0.23	0.18	0.23	0.22	0.22	0.19	0.19	0.25	0.17	0.19	0.14	0.25	0.21	0.16	0.21	0.19	0.25	0.22	0.21	0.16	0.18	0.22	0.21	0.24	0.23	0.22	0.22	0.22	0.31	0.19	0.22	0.21	0.11	0.25	0.15	0.19	0.18	0.29	0.19	0.23	-

Table 4.5 Values for all F_{ST} calculations within and between STRUCTURE clusters

	F_{ST} between STRUCTURE clusters				F_{ST} within Structure clusters	
Structure Clusters:	North	SouthWest	Central	East		
North	-	0.001	0.001	0.001	North	0.102
SouthWest	0.0131	-	0.001	0.001	SouthWest	0.070
Central	0.0223	0.024	-	0.001	Central	0.077
East	0.0488	0.055	0.029	-	East	0.107

p-values shown above diagonal

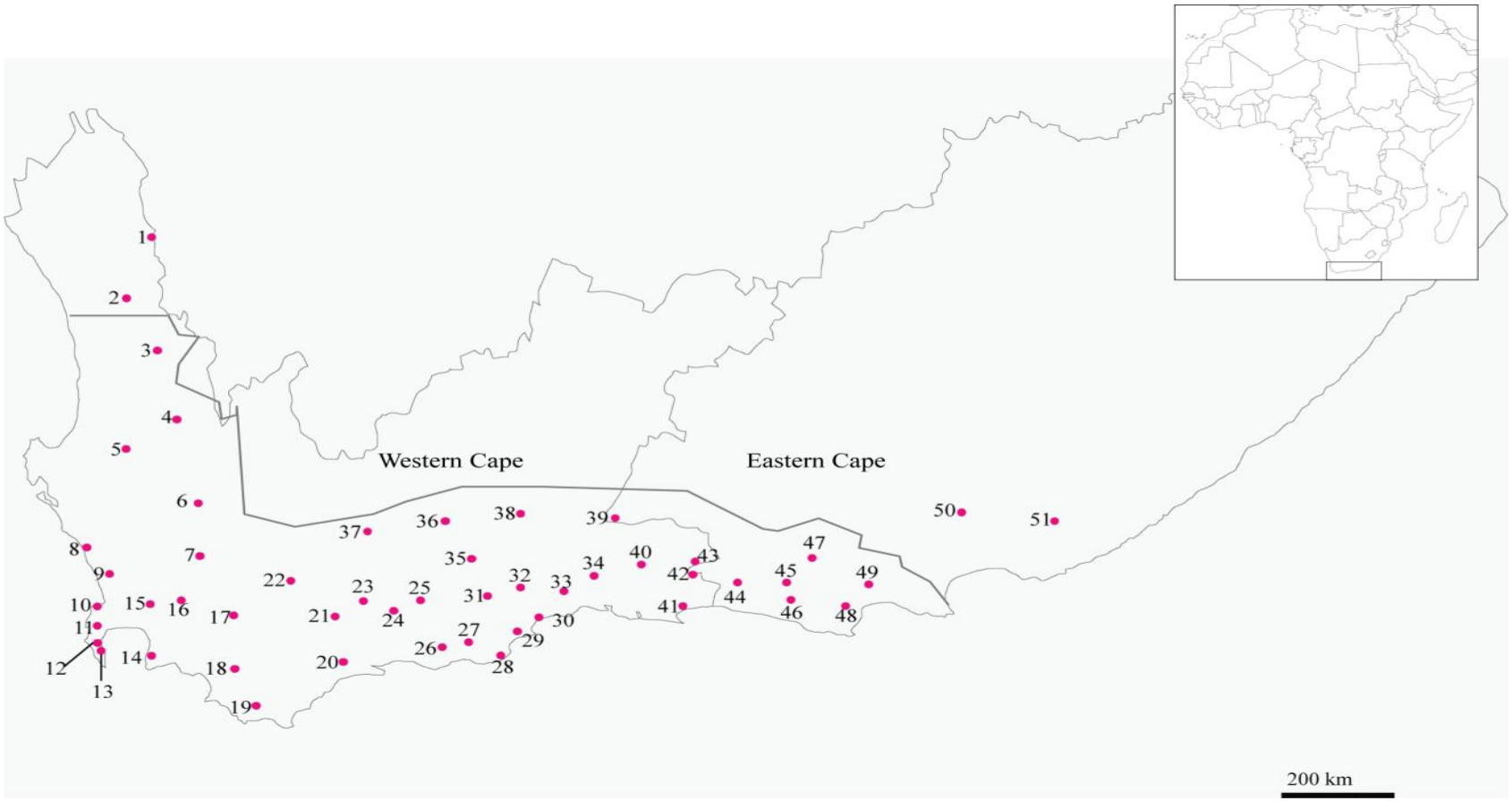

Fig. 4.1 All 51 locations of *L. salignum* sampled throughout the CFR and South Africa. Numbers correspond to locations listed in Table 4.1. The approximate boundary of the CFR (as in Turpie et al. 2003) is outlined in grey.

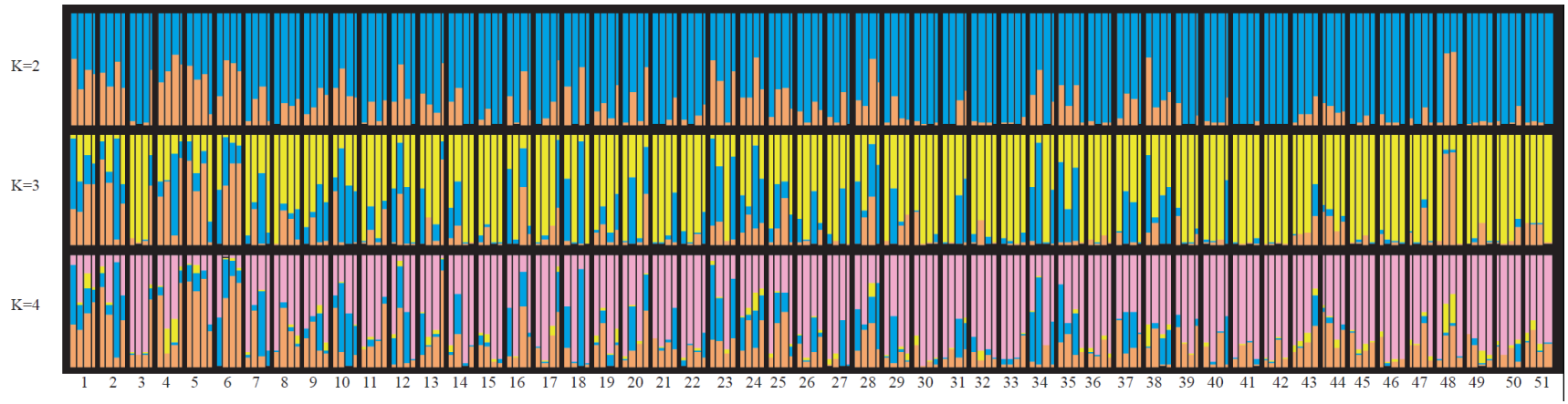

Fig. 4.2 Results of the Structure analyses with K ranging from 2 to 9. Each individual is represented by a vertical bar and populations are separated by black lines. 1. VRP 2. GIP 3. PAP 4. MIP 5. VEP 6. GYP 7. WOR 8. GBY 9. PHI 10. CPT 11. SVR 12. SIT 13. SWB 14. PRB 15. STL 16. FRP 17. GRY 18. STA 19. AGU 20. DHP 21. BTK 22. KOP 23. TRP 24. HID 25. GAP 26. STB 27. ALB 28. VLB 29. MOB 30. GRB 31. CLP 32. ROP 33. OUP 34. DAS 35. RBP 36. SEW 37. ANB 38. SWP 39. BLB 40. POP 41. PLB 42. MIS 43. KOU 44. JOB 45. NOT 46. AGS 47. BAV 48. HUM 49. SHB 50. SUP 51. GRT

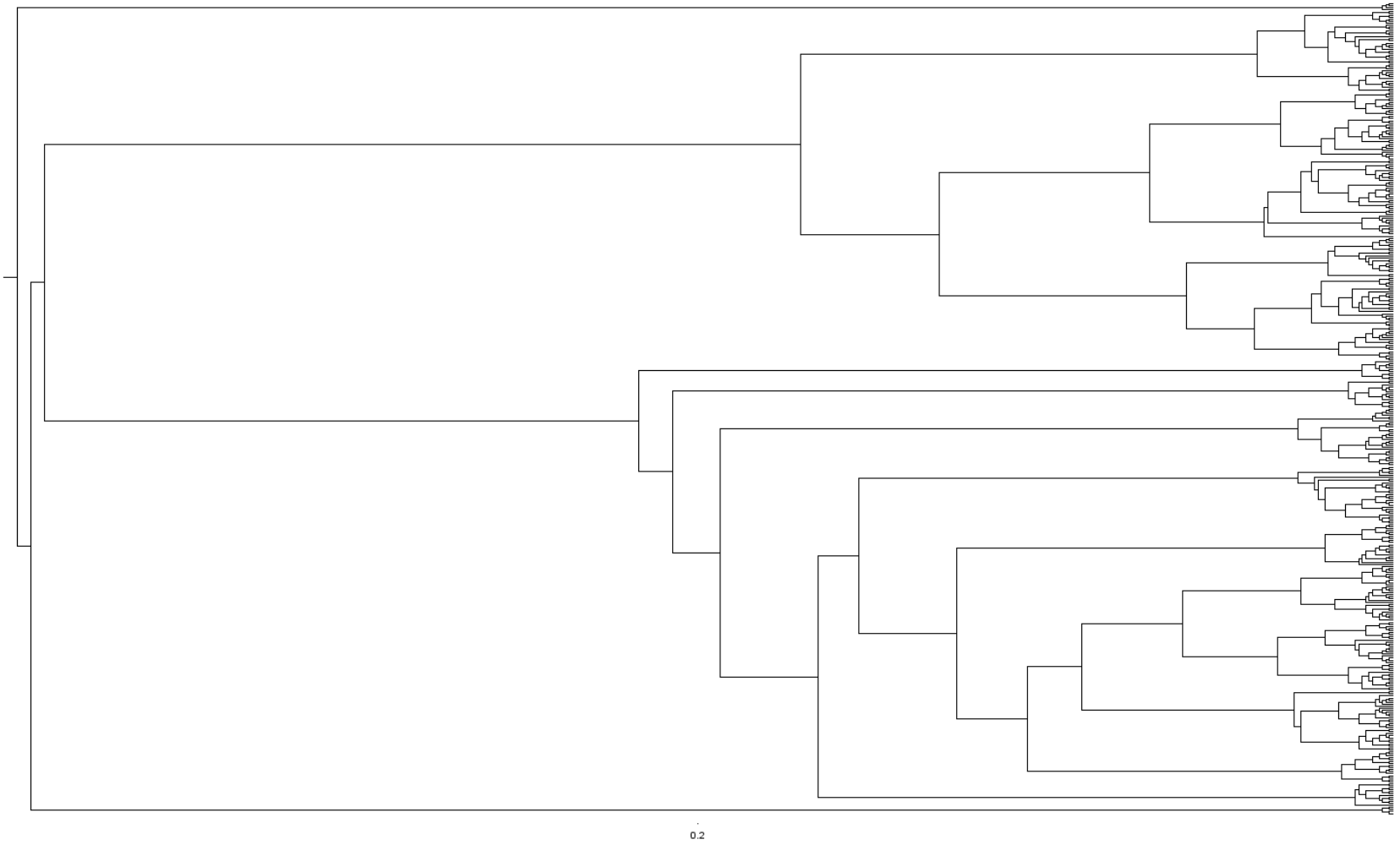

Fig. 4.3 Evolutionary estimated relationships of 408 individual-sequences based on BEAST analysis.

Chapter 5

CONCLUSION

The main objective of this research was to examine spatial genetic structure and gene flow over both spatial and temporal scales for a widespread fynbos plant, *L. salignum,* to understand levels of genetic diversity within endemic populations of the Cape Floral Region (CFR) and ultimately discover what factors might contribute to the diversity found within and among populations and to the region as a whole. The genetic diversity of *L. salignum* was characterized using both chloroplast (cpDNA) and nuclear DNA (nuDNA) markers to estimate temporal and spatial diversity using each of these markers. This method captured the two modes of gene flow available for plants through seed and pollen dispersal, allowing for comparisons between levels of gene flow which contribute to the overall genetic diversity of *L. salignum.*

As discussed in Chapter 2, the spatial genetic structure of *L. salignum* found using cpDNA varies on even a fine scale, with at least four broad STRUCTURE clusters and evidence for more through *post hoc* analyses. Interestingly, it was revealed that while locations within the *L. salignum* population vary spatially, they vary temporally as well, as it was shown that there has been a recent expansion of *L. salignum* into the Eastern Cape. These spatial and temporal differences found between groups suggest the landscape of the CFR is rather dynamic. These results support the idea that the current climatic differences between the Western and Eastern capes may be contributing to genetic variation currently. Given these patterns of fine scale genetic structure, it was next of interest to determine what may be driving these patterns of fine scale differences.

As discussed in Chapter 3 a model of isolation by distance explained the overall diversity in *L. salignum*; however, the fine-scale heterogeneity of the landscape was likely contributing to diversity as well. It was found that in measuring how the heterogeneity of the landscape contributes to genetic diversity through isolation by resistance analyses, vegetation type was an appropriate resistance variable that could explain some of the patterns of fine scale spatial genetic structure. However, in some regions, it appeared that other factors may better explain diversity. It was also clear, given these results, that further sampling in certain areas would help resolve how landscape resistance is generating diversity, for near the Cape Fold Mountains. In this region patterns of resistance were complex, and other groups (Daniels et al. 2007; Tolley et al. 2009) have hypothesized the mountains themselves are barriers to gene flow. Sampling in this area could disentangle the effects of the mountain topology from other factors on the landscape, such as vegetation type. Additionally, further sampling in the northern area of the CFR would allow for tests of isolation by resistance, as the current sampling scheme included too few individuals from that region. This region is especially of interest because it contained samples 'outside' the CFR, so understanding how *L. salignum* was able to move outside this area would add to our understanding of how plants are able to colonize outside their endemic region. Lastly, further sampling within the region that encompasses both the Western and Eastern Capes could elucidate where the definitive break is between the identified genetic groups. This area has been hypothesized to generate diversity over both contemporary and historical time scales; therefore, increased sampling here will improve our understanding of how *L. salignum* was able to recently expand into the area, as well as identify if the potential for expansion

is available for other fynbos plants. Overall, while vegetation resistance was able to predict genetic diversity in many areas, it is not the only landscape variable contributing to the patterns of genetic diversity found here. Further tests that capture additional geographic factors, such as topography, and their interaction with vegetation type may better explain genetic diversity than our single variable alone. In the future, it would be beneficial to incorporate more complex variables to better reflect the complex heterogeneity of the landscape in understanding patterns of gene flow and spatial genetic structure.

Lastly, as discussed in Chapter 4, patterns of gene flow and spatial genetic diversity changed significantly when using a nuDNA model. Here, we failed to detect spatial genetic structure with nuDNA markers. This contrasting pattern of genetic admixture to the highly structured cpDNA indicated that dispersal may be essential when studying gene flow, especially in plants which have two methods of dispersal and gene flow. It was clear from these data that these dispersal mechanisms are not equivalent in *L. salignum*, and therefore both methods were generating and maintaining diversity on different scales. The seed dispersal appeared to generate fine-scale spatial genetic structure, whereas pollen was maintaining diversity across the broad distribution of *L. salignum*. In moving forward with any study of gene flow and genetic diversity in CFR plants, it will be important to examine both maternally and bi-parentally inherited markers since levels of gene flow can vary between the two markers.

In summary, this research is the first to generate a population genetic data set for a widespread fynbos plant. This study provides insight into the micro-evolutionary processes occurring within a CFR species for two genomic data sets. For conservation

efforts, this study identified areas of importance which should be the foci of management efforts. Important areas identified were the region near the Cape Fold Mountains or the border between the Western and Eastern Capes, where diversity is apparently generated. This study also identified landscape factors that may be generating genetic diversity on a fine scale for one of the plant genomes, which can now also be incorporated into conservation plans. These fine-scale areas are important because although pollen maintains connectivity throughout the *L. salignum* population, it is the local dispersal of seeds that allows for the initial colonization of an area. Therefore, conserving vegetation that aids in seed dispersal will help maintain genetic diversity in *L. salignum.* This work has provided a model on which further studies can build upon, through examining other fynbos plants. In generating data for a number of fynbos plants, this methodology can contribute to understanding gene flow and spatial genetic structure across populations, allowing for the identification of patterns and processes that are shared among multiple plant species, which may be contributing to the diversity of the species rich CFR.

REFERENCES

Anderson, C. D., B.K. Epperson, M.J. Fortin, R. Holderegger, P.M.A. James, M. S. Rosenberg, K. T. Scribner, and S. Spear. 2010. Considering spatial and temporal scale in landscape-genetic studies of gene flow. *Molecular Ecology* 19 (17) (SEP): 3565-75.

Austerlitz, F., C.W Dick, C. Dutech, E.K. Klein, S. Oddou-Muratorio, P.E. Smouse, and V.L. Sork. 2004. Using genetic markers to estimate the pollen dispersal curve. *Molecular Ecology* 13 (4) (APR): 937-54.

Balmford, A., A. Bruner, P. Cooper, R. Costanza, S. Farber, RE Green, M. Jenkins, et al. 2002. Ecology - economic reasons for conserving wild nature. *Science* 297 (5583) (AUG 9): 950-3.

Barker, N. P., A. Vanderpoorten, C. M. Morton, and J. P. Rourke. 2004. Phylogeny, biogeography, and the evolution of life-history traits in *Leucadendron* (Proteaceae). *Molecular Phylogenetics and Evolution* 33 (3) (DEC): 845-60.

Barraclough, T. G. 2006. What can phylogenetics tell us about speciation in the cape flora? *Diversity and Distributions* 12 (1) (JAN): 21-6.

Baxter, S. W., J.W. Davey, J.S. Johnston, A.M. Shelton, D.G. Heckel, C.D. Jiggins, and M.L. Blaxter. 2011. Linkage mapping and comparative genomics using next-generation RAD sequencing of a non-model organism. *Plos One* 6 (4) (APR 26): e19315.

Beerli, P., and J. Felsenstein. 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences of the United States of America* 98 (8) (APR 10): 4563-8.

Bergh, N.G., T.A. Hedderson, H.P. Linder, and W.J. Bond. 2007. Palaeoclimate-induced range shifts may explain current patterns of spatial genetic variation in Renosterbos (*Elytropappus rhinocerotis*, asteraceae). *Taxon* 56 (2) (MAY): 393-408.

Birky, C. W. 1995. Uniparental inheritance of mitochondrial and chloroplast genes - mechanisms and evolution. *Proceedings of the National Academy of Sciences of the United States of America* 92 (25) (DEC 5): 11331-8.

Birky, C.W., P. Fuerst, and T. Maruyama. 1989. Organelle gene diversity under migration, mutation, and drift - equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. *Genetics* 121 (3) (MAR): 613-27.

Birky, C. W. 1988. Evolution and variation in plant chloroplast and mitochondrial genomes. In: Gottlieb, L. D. and Jam, S. K. (eds) *Plant Evolutionary Biology*, pp. 23—55. Chapman and Hall, London.

Birky, C. W., Maruyama, T. and Fuerst, I. 1983. An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results. *Genetics*, 103,513—527.

Bond, W.J., and P. Slingsby. 1983. Seed dispersal by ants in shrublands of the cape province and its evolutionary implications. *South African Journal of Science* 79 (8): 231-3.

Brumfield, R.T., P. Beerli, D.A. Nickerson, and S.V. Edwards. 2003. The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology & Evolution* 18 (5) (MAY): 249-56.

Chaw, S. M., T. W. Walters, C. C. Chang, S. H. Hu, and S. H. Chen. 2005. A phylogeny of cycads (Cycadales) inferred from chloroplast matK gene, trnK intron, and nuclear rDNA ITS region. *Molecular Phylogenetics and Evolution* 37 (1) (OCT): 214-34.

Coetzee, J.A., and J. Rogers. 1982. Palynological and lithological evidence for the miocene palaeoenvironment in the saldanha region (South-Africa). *Palaeogeography Palaeoclimatology Palaeoecology* 39 (1-2): 71-85.

Conde, M. F., Pring, D. R. and Levings, C. S. 1979. Maternal inheritance of organelle DNAs in Zea mays—Zea perennis reciprocal crosses. *J. Hered*., 70, 2—4.

Conrad, O. 2006. SAGA – Program Structure and Current State of Implementation. in: Böhner, J., McCloy, K.R. & Strobl, J. [Eds.]: SAGA – Analysis and Modelling Applications. *Göttinger Geographische Abhandlungen*, 115: 39-52.

Costanza, R., R. dArge, R. deGroot, S. Farber, M. Grasso, B. Hannon, K. Limburg, et al. 1997. The value of the world's ecosystem services and natural capital. *Nature* 387 (6630) (MAY 15): 253-60.

Cowling, R.M., S. Proches, and T.C. Partridge. 2009. Explaining the uniqueness of the cape flora: Incorporating geomorphic evolution as a factor for explaining its diversification. *Molecular Phylogenetics and Evolution* 51 (1) (APR): 64-74.

Cowling R.M. & Proches, S. 2005. Patterns and evolution of plant diversity in the Cape Floral Region. *Biol. skr.* 55: 273-288

Cowling, R., S. Proches, and J. H. J. Vlok. 2005. On the origin of Southern African subtropical thicket vegetation. *South African Journal of Botany* 71 (1) (MAR): 1-23.

Cowling, R.M., F. Ojeda, B.B. Lamont, P.W. Rundel, and R. Lechmere-Oertel. 2005. Rainfall reliability, a neglected factor in explaining convergence and divergence of plant traits in fire-prone Mediterranean-climate ecosystems. *Global Ecology and Biogeography* 14 (6) (NOV): 509-519.

Cowling, R.M., and RL Pressey. 2003. Introduction to systematic conservation planning in the cape Floral region. *Biological Conservation* 112 (1-2) (JUL-AUG): 1-13.

Cowling, R. M., and A. T. Lombard. 2002. Heterogeneity, speciation/extinction history and climate: Explaining regional plant diversity patterns in the cape Floral region. *Diversity and Distributions* 8 (3) (MAY): 163-79.

Cowling, RM. 1999. Classics in physical geography revisited - acocks,J.P.H. 1953: Veld types of South Africa. *Progress in Physical Geography* 23 (2) (JUN): 251-5.

Cowling, R. M., P. W. Rundel, B. B. Lamont, M. K. Arroyo, and M. Arianoutsou. 1996. Plant diversity in mediterranean-climate regions. *Trends in Ecology & Evolution* 11 (9) (SEP): 362-6.

Cowling, R.M. 1983. Phytochorology and vegetation history in the Southeastern cape, South-Africa. *Journal of Biogeography* 10 (5): 393-419.

Cowling, R.M., and B.M. Campbell. 1983. A comparison of fynbos and non-fynbos coenoclines in the lower gamtoos river valley, Southeastern cape, South-Africa. *Vegetatio* 53 (3): 161-178.

Cunningham, J., E.H.W. Baard, E.H. Harley, and C. O'Ryan. 2002. Investigation of genetic diversity in fragmented geometric tortoise (psammobates geometricus) populations. *Conservation Genetics* 3 (3): 215-23.

Dale, M. R. T., and M-J Fortin. 2010. From graphs to spatial graphs. *Annual Review of Ecology, Evolution, and Systematics, Vol 41* 41 : 21-38.

Daniels, S.R., M. D. Hofmeyr, B.T. Henen, and K.A. Crandall. 2007. Living with the genetic signature of miocene induced change: Evidence from the phylogeographic structure of the endemic angulate tortoise *Chersina angulata*. *Molecular Phylogenetics and Evolution* 45 (3) (DEC): 915-26.

Davey, J.W., P.A. Hohenlohe, P.D. Etter, J.Q. Boone, J.M. Catchen, and M.L. Blaxter. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics* 12 (7) (JUL): 499-510.

Delaunay, H., and J. Berthoumeyun. 1934. Microdosage of glucose in sea water. *Comptes Rendus Des Seances De La Societe De Biologie Et De Ses Filiales* 116 : 230-1.

Delaunay, B. 1928. Complete solution to the indefinite equation X(2)q+Y-2=1 in whole numbers. *Mathematische Zeitschrift* 28 : 1-9.

Diniz-Filho, J., F. Alexandre, T. Siqueira, A.A. Padial, T.F. Rangel, V.L. Landeiro, and L.M. Bini. 2012. Spatial autocorrelation analysis allows disentangling the balance between neutral and niche processes in metacommunities. *Oikos* 121 (2) (FEB): 201-10.

Domingues, F.A., K.L. Silva-Brandao, A.G. Abreu, O.P. Perera, C.A. Blanco, F.L. Consoli, and C. Omoto. 2012. Genetic structure and gene flow among brazilian populations of *Heliothis virescens* (Lepidoptera: Noctuidae). *Journal of Economic Entomology* 105 (6) (DEC): 2136-46.

Drummond, A.J., M.A. Suchard, D.Xie, and A. Rambaut. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* 29 (8) (AUG): 1969-73.

Drummond, A.J., and A. Rambaut. 2007. BEAST: Bayesian evolutionary analysis by sampling trees RID A-3209-2010 RID B-2481-2009. *Bmc Evolutionary Biology* 7 (NOV 8): 214.

Drummond, A.J., G.K. Nicholls, A.G. Rodrigo, and W. Solomon. 2002. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics* 161 (3) (JUL): 1307-20.

Dubey, S., M. Antonin, C. Denys, and P. Vogel. 2007. Use of phylogeny to resolve the taxonomy of the widespread and highly polymorphic African giant shrews (*Crocidura olivieri* group, Crocidurinae, Mammalia). *Zoology* 110 (1): 48-57.

Dyer, R. J., J.D. Nason, and R.C. Garrick. 2010. Landscape modeling of gene flow: Improved power using conditional genetic distance derived from the topology of population networks. *Molecular Ecology* 19 (17) (SEP): 3746-59.

Dyer, R.J., and J.D. Nason. 2004. Population graphs: The graph theoretic shape of genetic structure. *Molecular Ecology* 13 (7) (JUL): 1713-27.

Earl, D.A., and B.M. vonHoldt. 2012. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 4 (2) (JUN): 359-61.

Ennos, R.A. 1994. Estimating the relative rates of pollen and seed migration among plant-populations. *Heredity* 72 (MAR): 250-9.

Epperson, B.K., B.H. McRae, K. Scribner, S.A. Cushman, M.S. Rosenberg, M.J. Fortin, Patrick M. A. James, et al. 2010. Utility of computer simulations in landscape genetics. *Molecular Ecology* 19 (17) (SEP): 3549-64.

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology* 14 (8) (JUL): 2611-2620.

Excoffier, L., and H.E.L. Lischer. 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under linux and windows. *Molecular Ecology Resources* 10 (3) (MAY): 564-7.

Fall, A., M.J. Fortin, M. Manseau, and D. O'Brien. 2007. Spatial graphs: Principles and applications for habitat connectivity. *Ecosystems* 10 (3) (APR): 448-61.

Falush, D., M.Stephens, and J.K. Pritchard. 2007. Inference of population structure using multilocus genotype data: Dominant markers and null alleles. *Molecular Ecology Notes* 7 (4) (JUL): 574-8.

Falush, D., M. Stephens, and J. K. Pritchard. 2003. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* 164 (4) (AUG): 1567-87.

Flombaum, Pedro, and Osvaldo E. Sala. 2008. Higher effect of plant species diversity on productivity in natural than artificial ecosystems. *Proceedings of the National Academy of Sciences of the United States of America* 105 (16) (APR 22): 6087-90.

Fogelqvist, J., Anne Niittyvuopio, Jon Agren, Outi Savolainen, and Martin Lascoux. 2010. Cryptic population genetic structure: The number of inferred clusters depends on sample size. *Molecular Ecology Resources* 10 (2) (MAR): 314-23.

Fritz, U., S.R. Daniels, M.D. Hofmeyr, J. Gonzalez, C.L. Barrio-Amoros, P. Siroky, A.K. Hundsdoerfer, and H.Stuckas. 2010. Mitochondrial phylogeography and subspecies of the wide-ranging sub-saharan leopard tortoise stigmochelys pardalis (testudines: Testudinidae) - a case study for the pitfalls of pseudogenes and GenBank sequences. *Journal of Zoological Systematics and Evolutionary Research* 48 (4) (NOV): 348-359.

Gabriel, K.R., and R.R. Sokal. 1969. A new statistical approach to geographic variation analysis. *Systematic Zoology* 18 (3): 259.

Gao, H., Williamson,S., and Bustamante, C.D. 2007.An MCMC Approach for Joint Inference of Population Structure and Inbreeding Rates from Multi-Locus Genotype Data. *Genetics* (online)

Gallager, R.G., P.A. Humblet, and P.M. Spira. 1983. A distributed algorithm for minimum-weight spanning-trees. *Acm Transactions on Programming Languages and Systems* 5 (1): 66-77.

Gaubert, P., P. J. Taylor, C. A. Fernandes, M. W. Bruford, and G. Veron. 2005. Patterns of cryptic hybridization revealed using an integrative approach: A case study on genets (carnivora, viverridae, genetta spp.) from the Southern African subregion. *Biological Journal of the Linnean Society* 86 (1) (SEP): 11-33.

Goldblatt, P. 1997. Floral diversity in the cape flora of South Africa. *Biodiversity and Conservation* 6 (3) (MAR): 359-77.

Goldblatt, P. 1978. Analysis of the flora of Southern-Africa - its characteristics, relationships, and origins. *Annals of the Missouri Botanical Garden* 65 (2): 369-436.

Goldblatt, P., and J. C. Manning. 2002. Plant diversity of the cape region of Southern Africa. *Annals of the Missouri Botanical Garden* 89 (2): 281-302.

Hagan, J.M., and A.A. Whitman. 2006. Biodiversity indicators for sustainable forestry: Simplifying complexity. *Journal of Forestry* 104 (4) (JUN): 203-10.

Hardy, C. R. 2006. Reconstructing ancestral ecologies: Challenges and possible solutions. *Diversity and Distributions* 12 (1) (JAN): 7-19.

Hattingh, V., and J.H. Giliomee. 1989. Pollination of certain Leucadendron species (proteaceae). *South African Journal of Botany* 55 (4) (AUG): 387-93.

Herron, M. D., J. M. Waterman, and C. L. Parkinson. 2005. Phylogeny and historical biogeography of African ground squirrels: The role of climate change in the evolution of Xerus. *Molecular Ecology* 14 (9) (AUG): 2773-88.

Hey, J., and R. Nielsen. 2007. Integration within the Felsenstein equation for improved markov chain monte carlo methods in population genetics. *Proceedings of the National Academy of Sciences of the United States of America* 104 (8) (FEB 20): 2785-90.

Helyar, S. J., J. Hemmer-Hansen, D. Bekkevold, M. I. Taylor, R. Ogden, M. T. Limborg, A. Cariani, et al. 2011. Application of SNPs for population genetics of nonmodel organisms: New opportunities and challenges. *Molecular Ecology Resources* 11 (MAR): 123-36.

Holderegger, R., D. Buehler, F. Gugerli, and S. Manel. 2010. Landscape genetics of plants. *Trends in Plant Science* 15 (12) (DEC): 675-83.

Holderegger, R., and H.H. Wagner. 2008. Landscape genetics. *Bioscience* 58 (3) (MAR): 199-207.

Hu, J.M., M. Lavin, M.F. Wojciechowski, and M.J. Sanderson. 2000. Phylogenetic systematics of the tribe *Millettieae* (Leguminosae) based on chloroplast trnK/matK sequences and its implications for evolutionary patterns in Papilionoideae. *American Journal of Botany* 87 (3) (MAR): 418-30.

Hubisz, M.J., D. Falush, M. Stephens, and J.K. Pritchard. 2009. Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources* 9 (5) (SEP): 1322-32.

Ingram, C. M., H. Burda, and R. L. Honeycutt. 2004. Molecular phylogenetics and taxonomy of the African mole-rats, genus Cryptomys and the new genus Coetomys gray, 1864. *Molecular Phylogenetics and Evolution* 31 (3) (JUN): 997-1014.

Jakobsson, M., and N.A. Rosenberg. 2007. CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23 (14) (JUL 15): 1801-6.

Kalinowski, S. T. 2011. The computer program STRUCTURE does not reliably identify the main genetic clusters within species: Simulations and implications for human population structure. *Heredity* 106 (4) (APR): 625-32.

Kluetsch, C. F. C., R. J. Dyer, and B. Misof. 2012. Combining multiple analytical approaches for the identification of population structure and genetic delineation of two subspecies of the endemic arabian burnet moth *Reissita simonyi* (Zygaenidae; Lepidoptera). *Conservation Genetics* 13 (1) (FEB): 21-37.

Koen, E.L., J. Bowman, and A.A. Walpole. 2012. The effect of cost surface parameterization on landscape resistance estimates. *Molecular Ecology Resources* 12 (4) (JUL): 686-96.

Koutnik, D. (1987). Wind pollination in the Cape Flora. In A Preliminary Synthesis of Pollination Biology in the Cape Flora, *South African National Scientific Programmes Report No. 141* (ed. A. G. Rebelo), pp. 126–133. CSIR, Pretoria.

Kryger, U., T. J. Robinson, and P. Bloomer. 2004. Population structure and history of Southern African scrub hares, *Lepus saxatilis*. *Journal of Zoology* 263 (JUN): 121-33.

Kumar, S., M. Nei, J. Dudley, and K. Tamura. 2008. MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. *Briefings in Bioinformatics* 9 (4) (JUL): 299-306.

Latimer, A. M., S. S. Wu, A. E. Gelfand, and J. A. Silander. 2006. Building statistical models to analyze species distributions. *Ecological Applications* 16 (1) (FEB): 33-50.

Legendre, P., and M.J. Fortin. 2010. Comparison of the mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Molecular Ecology Resources* 10 (5) (SEP): 831-44.

Lesia, M.G.A., M.D. Hofmeyr, and M.E. D'Amato. 2003. Genetic variation in three *Chersina angulata* (angulate tortoise) populations along the west coast of South Africa. *African Zoology* 38 (1) (APR): 109-17.

Librado, P., and J. Rozas. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25 (11) (JUN 1): 1451-1452.

Linder, H. P. 2006. Investigating the evolution of floras: Problems and progress - an introduction. *Diversity and Distributions* 12 (1) (JAN): 3-5.

Linder, H. P. 2005. Evolution of diversity: The cape flora. *Trends in Plant Science* 10 (11) (NOV): 536-41.

Linder, H. P., and C. R. Hardy. 2004. Evolution of the species-rich cape flora. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* 359 (1450) (OCT 29): 1623-32.

Linder, H.P. 2003. The radiation of the cape flora, Southern Africa. *Biological Reviews* 78 (4) (NOV): 597-638.

Linder, H.P., and J.H. Vlok. 1991. The morphology, taxonomy and evolution of rhodocoma (restionaceae). *Plant Systematics and Evolution* 175 (3-4): 139-160.

Liu, N.J., L. Chen, S. Wang, C.G. Oh, and H.Y. Zhao. 2005. Comparison of single-nucleotide polymorphisms and microsatellites in inference of population structure. *Bmc Genetics* 6 (DEC 30): S26.

Lopez, L., and R. Barreiro. 2013. Genetic guidelines for the conservation of the endangered polyploid centaurea borjae (asteraceae). *Journal of Plant Research* 126 (1) (JAN): 81-93.

Manel, S., M. K. Schwartz, G. Luikart, and P. Taberlet. 2003. Landscape genetics: Combining landscape ecology and population genetics. *Trends in Ecology & Evolution* 18 (4) (APR): 189-97.

Manne, L.L., P.H. Williams, G.F. Midgley, W. Thuiller, T. Rebelo, and L.Hannah. 2007. Spatial and temporal variation in species-area relationships in the fynbos biological hotspot. *Ecography* 30 (6) (DEC): 852-61.

Matthee, C. A., and T. J. Robinson. 1996. Mitochondrial DNA differentiation among geographical populations of *Pronolagus rupestris*, Smith's red rock rabbit (mammalia: Lagomorpha). *Heredity* 76 (MAY): 514-23.

McRae, B.H. 2006. Isolation by resistance. *Evolution* 60 (8) (AUG): 1551-61.

McRae, B.H., B.G. Dickson, T.H. Keitt, and V.B. Shah. 2008. Using circuit theory to model connectivity in ecology, evolution, and conservation. *Ecology* 89 (10) (OCT): 2712-24.

McRae, B.H., and P. Beier. 2007. Circuit theory predicts gene flow in plant and animal populations. *Proceedings of the National Academy of Sciences of the United States of America* 104 (50) (DEC 11): 19885-90.

Moodley, Y., and E.H. Harley. 2005. Population structuring in mountain zebras (*Equus zebra*): The molecular consequences of divergent demographic histories. *Conservation Genetics* 6 (6) (DEC): 953-68.

Mucina, L., Rutherford, M.C. & Powrie, L.W. (eds) 2007. Vegetation Map of South Africa, Lesotho and Swaziland, edn 2, 1:1 000 000 scale sheet maps. South African National Biodiversity Institute, Pretoria. ISBN 978-1-919976-42-6.

Neale, D. B., Marshall, K. A. and Harry, D. E. 1991. Inheritance of chloroplast and mitochondrical DNA in incense cedar (*Calocedrus decurrens*). Can. J. For. Res., 21, 717—720.

Ojeda, F., M. T. Simmons, J. Arroyo, T. Maranon, and R. M. Cowling. 2001. Biodiversity in South African fynbos and mediterranean heathland. *Journal of Vegetation Science* 12 (6) (DEC): 867-74.

Palmer, J. P. 1992. Mitochondrial DNA in plant systematics: applications and limitations. In: Soltis, P. S., Soltis, D. E. and Doyle, J. J. (eds) *Molecular Systematics of Plants*, pp. 36-49. Chapman and Hall, New York.

Partridge, T.C., D.M. Avery, G.A. Botha, J.S. Brink, J. Deacon, R.S. Herbert, R.R. Maud, et al. 1990. Late pleistocene and holocene climatic-change in Southern Africa. *South African Journal of Science* 86 (7-10) (JUL-OCT): 302-6.

Pearce, D. 1997. Nature's services. societal dependence on natural ecosystems - daily,GC. *Science* 277 (5333) (SEP 19): 1783.

Perry, G.H., R.D. Martin, and B.C. Verrelli. 2007. Signatures of functional constraint at aye-aye opsin genes: The potential of adaptive color vision in a nocturnal primate. *Molecular Biology and Evolution* 24 (9) (SEP): 1963-70.

Peterson, B. K., J.N. Weber, E.H. Kay, H.S. Fisher, and H. E. Hoekstra. 2012. Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *Plos One* 7 (5) (MAY 31): e37135.

Pharmawati, M., G. Yan, R. Sedgley, and PM Finnegan. 2004. Chloroplast DNA inheritance and variation in Leucadendron species (proteaceae) as revealed by PCR-RFLP. *Theoretical and Applied Genetics* 109 (8) (NOV): 1694-701.

Pirie, Michael D., E. G. H. Oliver, and D.U. Bellstedt. 2011. A densely sampled ITS phylogeny of the cape flagship genus Erica L. suggests numerous shifts in floral macro-morphology. *Molecular Phylogenetics and Evolution* 61 (2) (NOV): 593-601.

Posada, D., and K.A. Crandall. 1998. MODELTEST: Testing the model of DNA substitution. *Bioinformatics* 14 (9): 817-8.

Price, B. W., N. P. Barker, and M. H. Villet. 2007. Patterns and processes underlying evolutionary significant units in the *Platypleura stridula* L. species complex (hemiptera : Cicadidae) in the cape Floral region, South Africa. *Molecular Ecology* 16 (12) (JUN): 2574-2588.

Pritchard, JK, M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155 (2) (JUN): 945-959.

Proches, S., R. M. Cowling, P. Goldblatt, J. C. Manning, and D. A. Snijman. 2006. An overview of the cape geophytes. *Biological Journal of the Linnean Society* 87 (1) (JAN): 27-43.

Prunier, R., and K.E. Holsinger. 2010. Was it an explosion? Using population genetics to explore the dynamics of a recent radiation within Protea (Proteaceae L.). *Molecular Ecology* 19 (18) (SEP): 3968-3980.

Rannala, B., and Z.H. Yang. 2003. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics* 164 (4) (AUG): 1645-56.

Raspe, O., P. Saumitou-Laprade, J. Cuguen, and A. L. Jacquemart. 2000. Chloroplast DNA haplotype variation and population differentiation in *Sorbus aucuparia L.* (Rosaceae : Maloideae). *Molecular Ecology* 9 (8) (AUG): 1113-22.

Rector, A.L., and K.E. Reed. 2010. Middle and late pleistocene faunas of pinnacle point and their paleoecological implications. *Journal of Human Evolution* 59 (3-4) (SEP-OCT): 340-57.

Rosenberg, M.S., and C.D. Anderson. 2011. PASSaGE: Pattern analysis, spatial statistics and geographic exegesis. version 2. *Methods in Ecology and Evolution* 2 (3) (JUN): 229-32.

Rosenberg, N. A. 2004. DISTRUCT: A program for the graphical display of population structure. *Molecular Ecology Notes* 4 (1) (MAR): 137-8.

Rouget, M., R.M. Cowling, R.L. Pressey, and D.M. Richardson. 2003. Identifying spatial components of ecological and evolutionary processes for regional conservation planning in the Cape Floral Region, South Africa. *Diversity and Distributions* 9 (3) (MAY): 191-210.

Rouget, M., D.M. Richardson, and R.M. Cowling. 2003. The current configuration of protected areas in the Cape Floral Region, South Africa - reservation bias and representation of biodiversity patterns and processes RID B-7249-2008 RID A-1495-2008. *Biological Conservation* 112 (1-2) (JUL-AUG): 129-45.

Rutschmann, F. 2006. Molecular dating of phylogenetic trees: A brief review of current methods that estimate divergence times. *Diversity and Distributions* 12 (1) (JAN): 35-48.

Rymer, P.D., J.C. Manning, P. Goldblatt, M. P. Powell, and V. Savolainen. 2010. Evidence of recent and continuous speciation in a biodiversity hotspot: A population genetic approach in Southern African Gladioli (gladiolus; iridaceae). *Molecular Ecology* 19 (21) (NOV): 4765-82.

Sauquet, H., P.H. Weston, N.P. Barker, C.L. Anderson, D. J. Cantrill, and V. Savolainen. 2009. Using fossils and molecular data to reveal the origins of the cape proteas (subfamily proteoideae). *Molecular Phylogenetics and Evolution* 51 (1) (APR): 31-43.

Scaglione, D., A. Acquadro, E. Portis, M. Tirone, S.J. Knapp, and S. Lanteri. 2012. RAD tag sequencing as a source of SNP markers in *Cynara cardunculus* L. *Bmc Genomics* 13 (JAN 3): 3.

Schnitzler, J., T.G. Barraclough, J.S. Boatwright, P. Goldblatt, J.C. Manning, M.P. Powell, T.Rebelo, and V. Savolainen. 2011. Causes of plant diversification in the cape biodiversity hotspot of South Africa. *Systematic Biology* 60 (3) (MAY): 343-57.

Segarra-Moragues, G.J., and F.Ojeda. 2010. Postfire response and genetic diversity in erica coccinea: Connecting population dynamics and diversification in a biodiversity hotspot. *Evolution* 64 (12) (DEC): 3511-24.

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

Slatkin, M. 1987. Gene flow and the geographic structure of natural-populations. *Science* 236 (4803) (MAY 15): 787-92.

Slingsby, J.A., and G. Anthony Verboom. 2006. Phylogenetic relatedness limits co-occurrence at fine spatial scales: Evidence from the schoenoid sedges (Cyperaceae : Schoeneae) of the cape Floral region, South Africa. *American Naturalist* 168 (1) (JUL): 14-27.

Smit, H.A., J. Watson, and B. J. van Vuuren. 2010. Relative importance of habitat connectivity in shaping the genetic profiles of two Southern African elephant-shrews. *Journal of Biogeography* 37 (5) (MAY): 857-64.

Smit, H. A., T. J. Robinson, and B. J. Van Vuuren. 2007. Coalescence methods reveal the impact of vicariance on the spatial genetic structure of *Elephantulus edwardii* (afrotheria, macroscelidea). *Molecular Ecology* 16 (13) (JUL): 2680-2692.

Smouse, P.E. 2010. How many SNPs are enough? *Molecular Ecology* 19 (7) (APR): 1265-6.

Smouse, P.E., J.C. Long, and R.R. Sokal. 1986. Multiple-regression and correlation extensions of the mantel test of matrix correspondence. *Systematic Zoology* 35 (4) (DEC): 627-32.

Storfer, A., M. A. Murphy, J. S. Evans, C. S. Goldberg, S. Robinson, S. F. Spear, R. Dezzani, E. Delmelle, L. Vierling, and L. P. Waits. 2007. Putting the 'landscape' in landscape genetics. *Heredity* 98 (3) (MAR): 128-42.

Swart, B. L., K.A. Tolley, and C.A. Matthee. 2009. Climate change drives speciation in the Southern rock agama (*Agama atra*) in the cape Floral region, South Africa. *Journal of Biogeography* 36 (1) (JAN): 78-87.

Takahata, N. and Palumbi, S. R. 1985. Extranuclear differentiation and gene flow in the finite island model. *Genetics*, 109,441—457.

Tajima, F. 1989. Statistical-method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123 (3) (NOV): 585-95.

Thuiller, W., G.F. Midgley, M. Rouget, and R.M. Cowling. 2006. Predicting patterns of plant species richness in megadiverse South Africa. *Ecography* 29 (5) (OCT): 733-44.

Tolley, K. A., J.S. Makokha, D.T. Houniet, B. L. Swart, and C. A. Matthee. 2009. The potential for predicted climate shifts to impact genetic landscapes of lizards in the South African cape Floral region. *Molecular Phylogenetics and Evolution* 51 (1) (APR): 120-30.

Tolley, K.A., B.M. Chase, and F. Forest. 2008. Speciation and radiations track climate transitions since the miocene climatic optimum: A case study of Southern African chameleons. *Journal of Biogeography* 35 (8) (AUG): 1402-14.

Tolley, K.A, M. Burger, A.A. Turner, and C.A. Matthee. 2006. Biogeographic patterns and phylogeography of dwarf chameleons (Bradypodion) in an African biodiversity hotspot. *Molecular Ecology* 15 (3) (MAR): 781-93.

Toussaint, G.T. 1980. The relative neighborhood graph of a finite planar set. *Pattern Recognition* 12 (4): 261-8.

Turpie, J.K., Frazee, S.R., R.M. Cowling, R.L. Pressey, and N. Lindenberg. 2003. Estimating the costs of conserving a biodiversity hotspot: A case-study of the cape Floral region, South Africa. *Biological Conservation* 112 (1-2) (JUL-AUG): 275-90.

Valente, L.M., G. Reeves, J. Schnitzler, I.P. Mason, M.F. Fay, T.G. Rebelo, M.W. Chase, and T.G. Barraclough. 2010. Diversification of the African genus Protea (Proteaceae) in the cape biodiversity hotspot and beyond: Equal rates in different biomes. *Evolution* 64 (3) (MAR): 745-59.

Verboom, G.A, J.K. Archibald, F.T. Bakker, D.U. Bellstedt, F. Conrad, L.L. Dreyer, F. Forest, et al. 2009. Origin and diversification of the greater cape flora: Ancient species repository, hot-bed of recent radiation, or both? *Molecular Phylogenetics and Evolution* 51 (1) (APR): 44-53.

Watterson, G. A. 1975. Number of segregating sites in genetic models without recombination. *Theoretical Population Biology* 7 (2): 256-76.

Werger, M.J.A. (1978a) The Karoo-Namib Region. Biogeography and ecology of southern Africa (ed. by M. J. A. Werger), pp. 233-299. Junk, The Hague.

Williams, I.J.M., 1972. A revision of the genus Leucadendron (Proteaceae). *Contrib. Bol.Herb.*
3, 1–425.

Willing, E.M., M. Hoffmann, J.D. Klein, D. Weigel, and C. Dreyer. 2011. Paired-end RAD-seq for de novo assembly and marker design without available reference. *Bioinformatics* 27 (16) (AUG 15): 2187-93.

Willows-Munro, S., and C.A. Matthee. 2009. The evolution of the Southern African members of the shrew genus Myosorex: Understanding the origin and diversification of a morphologically cryptic group. *Molecular Phylogenetics and Evolution* 51 (2) (MAY): 394-398.

Wilson, G.A., and B. Rannala. 2003. Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* 163 (3) (MAR): 1177-91.

Wolfe, K.H., WH Li, and P.M. Sharp. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences of the United States of America* 84 (24) (DEC): 9054-8.

Wright, S. 1931. Evolution in Mendelian populations. *Genetics* 16 (2) (MAR): 0097-159.

Zellmer, A. J., and L. L. Knowles. 2009. Disentangling the effects of historic vs. contemporary landscape structure on population genetic divergence. *Molecular Ecology* 18 (17) (SEP): 3593-602.