Molecular phylogeny and population genetics of the *Grevillea* Thelemanniana group (Proteaceae).

Tanya Helena Hevroy (BSc Hons)

This thesis is presented for the degree of

Doctor of Philosophy of the University of Western Australia

School of Plant Biology 2016

THE UNIVERSITY OF WESTERN AUSTRALIA

Achieving International Excellence

Summary

Species are the basis for biological studies and ultimately the most important unit for conservation strategies. The definition and delimitation of species has been debated heavily in taxonomy, and no all encompassing definition has been defined yet. The recent expansion of molecular technology, allowing for larger sample sizes and multi-locus datasets, has shifted the field of systematics from morphological or single locus taxonomy towards integrative taxonomy applying multiple data sources. New concepts have been created which utilize population genetic theory to analyse multi locus data and discover patterns of evolutionary lineages to aid in discovering species limits and boundaries. By using molecular data from different regions of the genome, researchers can better understand evolutionary patterns. Additionally, new software employing Bayesian methods and molecular dating has increased our understanding of evolutionary patterns and speciation. Understanding the different patterns which make up species will ultimately lead to better conservation decision making.

In this study the *Grevillea* Thelemanniana group of Western Australia was investigated using chloroplast sequences from two regions and 12 microsatellites. Microsatellites were specifically designed for *Grevillea thelemanniana* ssp. *thelemanniana* and cross amplified across the group. The Thelemanniana group is made up of 16 species, four subspecies and two phrase named species that, based on morphology, likely have affiliations to the group. The resulting gene tree did not support all the original taxonomically defined species and species relationships and did not support the Thelemanniana group as monophyletic in relation to species from other taxonomically defined groups. However, two significantly different lineages were recovered from the chloroplast analyses, which corresponded to a northern and southern distribution. Using molecular methods and dating technology a significant lineage divergence within the Thelemanniana group was evident, dating back to the Miocene. The patterns of evolution within lineages in part support evolutionary hypothesis of the Southwest floristic region of Australia, but also illustrated alternate patterns.

Incongruence between resulting gene tree and species tree and also incongruence between sequence and microsatellite data, has often been attributed to incomplete lineage sorting (deep coalescence), historical introgression and/or hybridization.

Although systematical studies often discuss these issues, few have enough data to distinctly resolve them. In the second part of this thesis I aimed to resolve a subclade within the Thelemanniana group to address potential issues of incomplete lineage sorting, hybridization and introgression. I collected greater sample sizes from multiple populations for four taxon, applying chloroplast sequences and microsatellites, to address different levels of divergence between species and populations and delimit species boundaries. Results indicated molecular support for most species as unique and little evidence for recent introgression or hybridization.

Finally, I used a landscape genetic approach using microsatellites and geographic data to discover the spatial genetic structure of *Grevillea thelemanniana* ssp. Cooljarloo. Landscape genetics aims to provide information on how landscape features interact with evolutionary processes such as gene flow, migration and drift. *Grevillea thelemanniana* ssp. Cooljarloo is restricted to a 20 Km^2 region and is threatened by an active and expanding mine site within the region. The population sizes were large and continuous; however there appears to be habitat selection for creek lines and floodplains. Through riparian models and spatial Bayesian clustering analyses, regional clustering of creek line subpopulations and strong divergence between creek lines was revealed. Also, a surprising amount of genetic admixture from multiple sources was witnessed in flood plains which could be explained by regular flooding and pollinator movements.

The results from this study are important in explaining the evolutionary past of *Grevillea*, the third largest genus of flowering plants in Australia, and also adding to the greatly understudied southwest of Western Australia biodiversity hotspot. This work is one of the first to investigate the molecular phylogeny of *Grevillea*, and the first to do so on a Western Australian group. The results are important in informing conservation decisions as the majority of species within the Thelemanniana group are rare or threatened.

Table of Contents

Acknowledgments

I dedicate this PhD to my parents. They took me travelling around the world and showed me all the beauty it holds, which in turn sparked my love for science and nature. Their continued support and love throughout my life has allowed me to embark on all the adventures I could dream of and got me through this PhD. I would also like to thank my two brothers, who never in my 25 years have teased me for being a science nerd. And even though to this day they cannot explain what I am working on, they have voiced how proud they are on numerous occasions and I love them both very much.

Thank you to my supervisors Michael Moody and Siegy Krauss. Your knowledge and guidance throughout these 4 years has shaped me into the scientist I am today. I am truly proud of this work and I could not have achieved it without you both. Thank you for the financial support from The University of Western Australia and also Tronox Minerals, Kings Park Science and The Department of Parks and Wildlife for funding this work.

To the wonderful ladies (and Dave Coates) at the Department of Parks and Wildlife, you were a blessing to me. Your open door policy and 3'oclock chocolate breaks made working on this project fun. Even after 6 hours of DNA extractions. I cannot thank you enough for welcoming me with open arms and all that you have taught me over the past 4 years. A special thanks to Sarah (Battleship) Tapper, Heidi Nistleberger, Rachel Binks and Kym Ottewell, who were always there to talk me through a computer program, edit my figures, brainstorm my ideas and importantly, take me to the pub when necessary.

This project required me to travel the majority of the great state that is Western Australia. I have seen many places in this world and I still cannot fathom the beauty this place holds. I travelled from the limestone ranges of Cape Range National Park, to the southernmost forests of Mt. Frankland National Park, separated by 3000 kilometres. These journeys were made even more wonderful by my willing companions; Lindsey Spiller, Alison Ritchie, Jessica Stingemore and especially my mother. You were all much better at field work than me and I enjoyed these trips immensely.

A big thank you to my friends, who have been looking forward to calling me Dr. T. Thank you so much for supporting me, listening to my ideas and always encouraging me to reach further. A special thanks to Cameron Mounsy, for countless lunches and coffee breaks, discussing the world's problems with you every day was a joy. And lastly, thank you to Addie Bevilaqua, Tahnia Brumby and Tim Slate, for, well, just being my best friends.

DECLARATION OF CANDIDATE CONTRIBUTION

This thesis contains published and/or work prepared for publication, some of which has been co-authored. The bibliographical details of the works and where it appears in the thesis are outlined below.

One of the chapters of this thesis has been published. The text in this chapter has been modified slightly to suit the remainder of the thesis. The original manuscript is included as an Appendix.

Chapter 5:

Hevroy TH, Moody ML, Krauss SL, Gardner M (2013) Microsatellites for *Grevillea thelemanniana* ssp. *thelemanniana* and cross-species amplification in the *Grevillea* Thelemanniana group (Proteaceae). Conservation Genetics Resources **5**, 887-890.

Author contributions: TH 90% MM 2.5% SK 2.5% SG 5%

Chapter 1 – General Introduction

"We are here today to pay homage to our God, Proteus, and his wonderfully diverse family. If he made these plants in his own image and likeness, then either he has a personality crisis or else, after the initial creation, matters got out of hand." – Alex. S. George

Species Concepts and Delimitation

The taxonomic rank of species is the fundamental unit of biological evolutionary and conservation studies. Species delimitation and formal taxonomy of plants has relied greatly on morphological characters, as phenotypic similarity has been the criterion used historically by taxonomists to group individuals into species (Mallet 1995; Soltis & Gitzendanner 1999). Morphological taxonomy can however improperly subdivide species through inaccurate interpretation of natural phenotypic diversity over their distribution area (Duminil *et al.* 2012), or alternatively fail to discriminate between morphologically similar species, (i.e. cryptic species). Furthermore, limited morphological variation can mask considerable evolutionary diversity in many species complexes (Sinclair *et al.* 2004).

Although there are many papers that address the issue of species delimitation, there is little compatibility between different methods, which can lead to different conclusions for species boundaries (de Queiroz 2007). Sites JR and Marshall (2004) reviewed 12 of the the most commonly applied species delimitation methods, of which the majority applied molecular markers. With the advent of molecular sequencing in recent decades, the delimitation of species boundaries based on genetic information is becoming a routine and accepted practice within evolutionary biology. Molecular markers are advantageous as they provide insights into the genealogical descent of species and represent numerous alternative and powerful tools for species delimitation.

The shift from traditionally morphologically defined species to species defined on the basis of molecular data has also promoted the shift from traditional species concepts, which often use primarily one diagnostic tool. For example, the biological species

concept (Mayr 1942), is intrinsically based on defining reproductively isolated individuals, the ecological species concept (Van Valen 1976) is based on individuals in the same niche or adaptive zone, while the phylogenetic species concept (Cracraft 1983) emphasizes reciprocal monophyly and branch support, represent the diagnostic criteria for species delimitation. Since one of the main goals of systematics is to discover lineages at lower levels (Sites Jr & Marshall 2003), more recent species concepts define species as separately evolving lineages which need to satisfy a set of criteria (de Queiroz 2007; Shaffer & Thomson 2007; Miralles & Vences 2013). For example, the general lineage concept (GLC) was first introduced by de Queiroz (1998), based on the recognition that although species concepts are based on different assumptions, the overall similarity was that all concepts recognized species as variations of an evolutionary lineage. The aim of the general lineage concept is to move away from traditional definitions of species, and rather apply a set of criteria which species meet, such as reproductive isolation, phenetic differentiation, diagnosability (fixed differences) and reciprocal monophyly. When these criteria are met, species can be confidently delimited. Similarly, Padial *et al.* (2010) and Miralles and Vences (2013) argued for an integrative taxonomic approach to species delimitation, which implies that species are delimited from multiple criteria which can be supported by different lines of evidence. Padial *et al* (2010) outlined two frameworks: “Integration by congruence” and “Integration by cumulation”, both of which are based on integration which identifies a population-level phylogenetic lineage by several unlinked genetic loci that indicates genetic isolation from other such lineages, and thus qualifies as a species. Alternate species concepts can however also disagree on what criteria signify that speciation has occurred, or is occurring during the process of lineage divergence. Regardless, the primary goal of species delimitation, and what recent species concepts agree upon, is focusing on lineage discovery and delimitation (Shaffer & Thomson 2007) by considering all sources of evidence as potentially useful to support species (Padial *et al.* 2010).

Using an integrative approach in which species are delineated based on a combination of consistent genetic, morphological and geographic features may be a reasonable solution to the species delimitation debate and has prompted a coalescent based approach to species delimitation. Coalescent-based methods directly incorporate the

process of gene coalescence, which accounts for gene tree discordance often seen in phylogenetic studies and utilizes information from individual loci to estimate species membership, inter-relationships, population sizes, and divergence times (Carstens & Knowles 2007; Fujita *et al.* 2012). Another advantage of this approach is that it identifies quantitative differences in continuous variables, rather than distinct entities that do or do not conform to a prior species concept, and species taxa are recognized as evolving from population-level processes (Hart 2011). Numerous studies have proclaimed the benefits of a coalescence approach, combining theories of both population genetics and phylogenetic studies to delimitate species (Sites Jr & Marshall 2004; Petit & Excoffier 2009; Barrett & Freudenstein 2011; Hey & Pinho 2012; Pessoa *et al.* 2012). Analytical models that merge population genetic analyses with phylogenetics have resulted in an important shift in systematics that allows comparison of divergence levels. This shift has largely been driven by the increasing availability of large datasets consisting of greater sample sizes and multiple loci evolving at different rates.

Until recently it has been common practice for phylogenetic species delimitation to use a single gene sequence from a single individual organism as a proxy for an entire species (Heled & Drummond 2010). Technological advances mean it is now becoming more common to collect data sets containing multiple gene loci and multiple individuals per species (Goldstein *et al.* 2000; McCormack *et al.* 2009) especially since coalescence stochasticity leads to problems if focusing on a single locus (Omland *et al.* 2006). This enables a more objective assessment of species limits than delimiting species based solely on data from a single locus or morphology and is especially applicable to very recently diverged populations in which loci are subject to incomplete lineage sorting, making reciprocal monophyly in gene trees of independent lineages unlikely (Hudson & Coyne 2002).

Various processes may lead to incongruence between the species tree and gene trees (Maddison 1997; McCormack *et al.* 2009; Duminil *et al.* 2012), and the processes behind this are notoriously difficult to discriminate between (Spinks & Shaffer 2009). The most common causes for incongruence are the following processes;

1) Incomplete lineage sorting due to rapid species diversification and/or when species have originated recently can result in insufficient time for the fixation of gene lineages by genetic drift (Avise 2000; Carstens & Knowles 2007). Consequently, the retention and sorting of ancestral polymorphisms, where gene lineages that persist into deeper portions of the species trees may coalesce with gene lineages that are not from the most closely related species (i.e., deep coalescence), which causes a mismatch between the gene tree and the species tree (Maddison 1997; Knowles & Carstens 2007). Knowles (2009) showed that incomplete lineage sorting was four times as likely in shallower species phylogenies.

2) Hybridization and/or introgression can also cause incongruence between gene trees and species trees, as well as between different types of molecular markers. Hybridization may occur sporadically between sympatric species or be confined to hybrid zones. Introgression refers to the gene movement between species mediated by hybridization and back crossing (Avise 2000). Despite the fact that these processes can result in loss of diversity through the fusion of lineages, introgression can lead to the formation of novel species (Baack & Rieseberg 2007) and hybridization has been shown to be a driving force in speciation of numerous plant species (Andersson 1990; Arnold 1997; Rieseberg & Carney 1998). Although hybridization is more common in actively speciating groups (Andersson 1990), in many taxonomic groups, organisms separated for long periods of time may none the less retain anatomical and physiological capacity for hybrid production (Avise 2000).

3. Gene duplication with differential loss of loci producing comparisons of paralogous loci.

Knowles (2009) recommended a move away from the tradition of using gene trees as a literal interpretation of the species history, toward procedures that explicitly model the relationship between gene trees and species trees. The incongruence between gene trees and species trees have often been viewed as obstacles for inferring taxonomies, whereas recent authors recommend using that information as an opportunity to regard processes which have shaped species genomes (Degnan & Rosenberg 2009). Understanding the underlying causes for incongruence between gene trees and species trees is useful for inference of species phylogenies (Rosenberg

2002). The increasing use of different molecular markers and greater sample sizes can greatly expand on knowledge of the degree of incomplete lineage sorting, especially speciation times, hybridization and introgression (Arnold 1997; Wendel & Doyle 1998; Koch *et al.* 2003; Petit & Excoffier 2009).

Recent taxonomic studies have strongly recommended the use of different genomes in delimiting species rather than using just one line of evidence (Marshall *et al.* 2006; Wiens 2007; Hey & Pinho 2012). Studies incorporating molecular data from both chloroplast DNA sequences and nuclear DNA enable inferences on the hierarchies of time, analysed through the inherent properties of molecular makers and their different rates of evolution (Padial *et al.* 2010). By using both nuclear and chloroplast markers, and population and phylogenetic theory, incongruence can be explained and explored to a greater extent than before. Furthermore, the incorporation of geographic data into species delimitation has been encouraged (de Queiroz 2007; Wiens 2007), although this hasn't been seen much at interspecies levels, it has become increasingly popular at intraspecific levels.

Significant challenges remain in the empirical delimitation of species boundaries in recent radiations of closely related species (Hey *et al.* 2003), morphologically conservative species and species complexes with deep phylogenetic histories and limited gene flow (Sinclair *et al.* 2004). Recent studies applying methods for estimating species trees from multiple gene trees that combine Bayesian or maximum-likelihood models in a coalescent framework have been successful in resolving issues of incongruence in species delimitation. For example, multi-locus data has been successfully applied to resolving patterns of hybridization between species in *Helianthus* (Kane *et al.* 2009), levels of introgression between oaks (Lepais *et al.* 2009), closely related species (Franck *et al.* 2004; Baker *et al.* 2010; Rossetto *et al.* 2012) and species complexes in biodiversity hotspots (Edwards *et al.* 2008; Millar *et al.* 2011; Willis *et al.* 2012; Razafinarivo *et al.* 2013). Specifically, scaling up phylogenetic knowledge and use of genetic resources within species, and the evolutionary relationships of taxa to an integrative overview is particularly important in contributing to conservation in biodiversity hotspots (Hopper 2000).

Geographic and climatic history of the southwest of Western Australia

As Australia moved northwards from Antarctica during the Oligocene, it became progressively isolated from the remainder of Gondwanaland. During this time, seasonality became apparent in the southeast and southwest of Australia, and resulting climatic change is thought to have been responsible for fragmentation of the mesic zones (Martin 2006; Byrne *et al.* 2011). By the mid-Miocene, the first major step towards aridity became apparent (Martin 2006). The colder, drier conditions of the Miocene are reflected in contraction of the rainforest habitats and an expansion of sclerophyllous flora adapted to low nutrient soils (Beard *et al.* 2000; Byrne *et al.* 2008). Since the separation of Australia from Antarctica, Western Australia has not experienced glaciations, volcanism or significant mountain building leading to complex mosaics of nutrient deprived soils (Hopper & Gioia 2004). The marked increase in aridity during the mid-Miocene also marks the end of regular flows in palaeodrainage systems over much of Western Australia (Martin 2006). Furthermore, the flat terrain resulted in uncoordinated drainage, so that salts and rain are rarely flushed out into the ocean but rather have become extensive salt lake systems and ancient broad valley drainage lines (Hopper 2000).

In Western Australia, approximately 8400 native plant species are formally described, of which 62% are endemic (Beard *et al.* 2000; Hopper & Gioia 2004). The contemporary flora is dominated by sclerophyllus forests in the extreme southwest, woodlands and heathlands in the semi-arid zones, and shrublands and hammock grasslands in the arid zone (Byrne 2007). The greatest species richness and endemism in the state is in the southwest region, recognized as a global biodiversity hotspot (Myers *et al.* 2000) and referred to as the Southwest Australian Floristic Region (SWAFR, Fig.1.1, Hopper and Gioia 2004). The region is characterized by a Mediterranean type climate (Merwin *et al.* 2012) where there are approximately 5700 plant species, of which 79% are endemic (Beard 2000). Local endemism is prominent as there have been unprecedented periods of landscape stability and vegetative cover, which is suggested to have favoured selection for plants to stay put rather than disperse widely (Hopper 2000). Also, the formation of the Nullarbor Plain during the

mid Miocene is thought to have been one of the reasons the SWAFR has such high regional endemism (Crisp *et al.* 2004).

The development of the Nullarbor Plain has long been recognized as one of the most significant barriers between the southwest and southeast of Australia (Crisp *et al.* 2004; Crisp & Cook 2007). Molecular dating for several studies have suggested splits during the Miocene (20-5 Mya), corresponding closely to the formation of the Plain (Crisp & Cook 2007; Morgan *et al.* 2007; Rix & Harvey 2012). Cardillo and Pratt's (2013) phylogeny of *Banksia* revealed several SWAFR species as more closely related to east coast species, with a median estimate of divergence at only 8.1 Mya, post dating the formation of the Nullarbor Plain. They suggested that although the Plain formed an edaphic barrier, it is unlikely to have been a hard barrier to dispersal (Cardillo & Pratt 2013). Crisp and Cook (2007) suggested that populations of southwest and southeast Australian species could have remained linked via extensive siliceous soils of Central Australia, which during the late Miocene experienced wetter and warmer climate. Furthermore, the Late Miocene to early Pliocene as origin for multiple plant lineages have emerged in species-level molecular phylogenies of genera in the southwest, corresponding to these significant geographic events (Coates *et al.* 2003; Crisp & Cook 2007; Hopper 2009). This has also been evident in the fossil record which shows significant changes in vegetation, associated with periods of aridification, during the Pliocene in Western Australia (Byrne & Hopper 2008). These results indicate that a combination of the Nullarbor Plain and successive aridification is likely responsible for species diversity and endemism in the southwest of Western Australia. By the end of the Pliocene, arid conditions approaching those of today had been reached, but aridity has intensified especially in the last half million years (Martin 2006).

Hopper (1979) and Hopper and Gioia (2004) suggested that the recent climatic fluctuations during the Quaternary, which led cyclic contraction and expansion of the mesic zones in the SWAFR, brought about the extinction of many species and stimulated the explosion within other genera. McLoughlin and Hill (1996) suggested that the SWAFR has been the major centre for diversity of numerous families, especially the Myrtaceae and Proteaceae, since at least the Eocene. Recently, a molecular study of *Banksia* (Cardillo & Pratt 2013) supported a gradual accumulation of species, rather than an explosive speciation, as suggested by Hopper and Gioia

(2004). Ultimately, these studies highlight the need for further molecular phylogenetic studies of southwest flora to better understand patterns of speciation and divergence in the SWAFR.

Studies conducted within species and subspecies in the SWAFR have however revealed comparable patterns of divergence which have been attributed to fluctuations during the Quaternary (Byrne 2007; Byrne *et al.* 2008). Highly divergent lineages have been documented in numerous species in SWAFR, as populations likely contracted to geographically isolated refugia followed by subsequent expansion during favourable conditions in species phylogenies (Byrne *et al.* 1999; Byrne *et al.* 2002; Byrne *et al.* 2003; Coates *et al.* 2003; Byrne & Hines 2004; Wheeler & Byrne 2006; Yates *et al.* 2007; Nevill *et al.* 2014) and molecular dated species phylogenies (Edwards 2007; Cooper *et al.* 2011; Rix & Harvey 2012; Tapper *et al.* 2014).

The patterns of lineage divergence within many species reflect common responses of population contractions due to climatic change. Speciation studies in Mediterranean-climate regions usually invoke climate change as the driving force for geographic speciation in topographically heterogeneous environments (Andersson 1990; Cowling *et al.* 1996; Cardoso & Vogler 2005). However, species-specific patterns of lineage distribution within regions may reflect influences over different time scales or differing ecological tolerances (Byrne *et al.* 2011). Other factors in the SWAFR have been used to explain the high levels of species diversity, endemism and divergence within species. These include habitat specialization due to nutrient deprived soils (Cowling & Lamont 1998; Beard *et al.* 2000; Edwards *et al.* 2010) and high fire frequencies amounting to turn over of fire-killed species (Cowling & Lamont 1998; Byrne *et al.* 2011; Merwin *et al.* 2012). Mediterranean ecosystems provide many opportunities for the comparative study of the determinants of plant diversity at local and regional scales (Cowling *et al.* 1996). The ecology and unique geohistory of the southwest of Australia offers interesting opportunities is studying species diversity and landscape ecology.

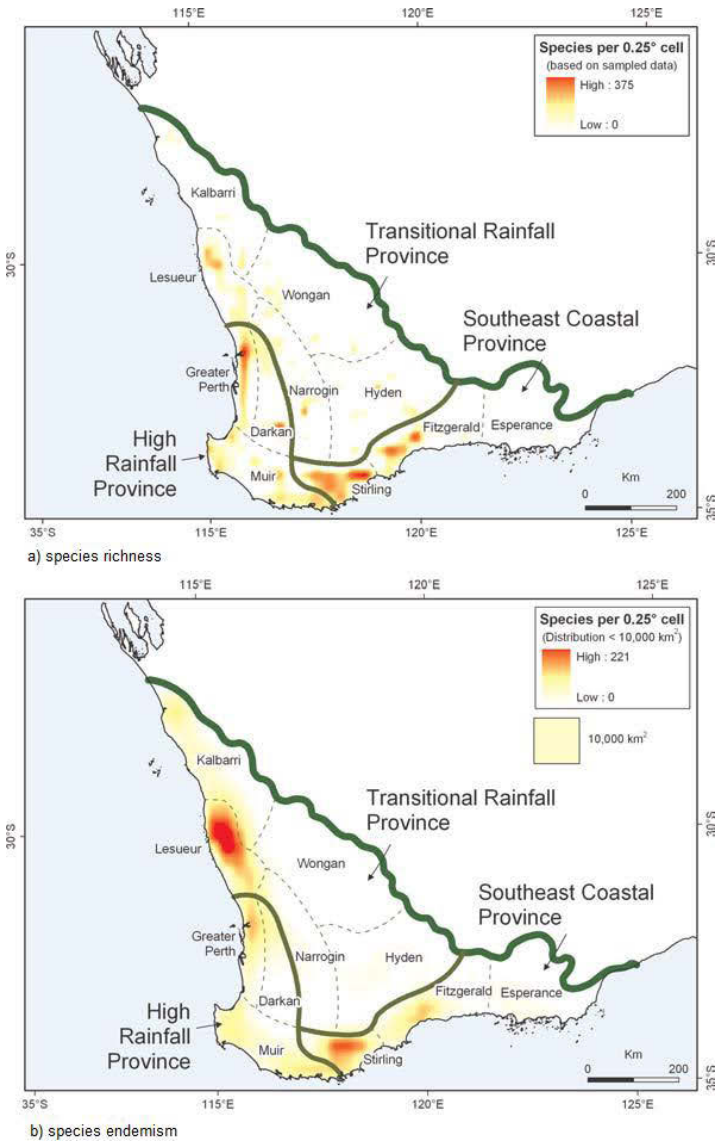

Fig.1.1. Phytogeographic maps of the SWAFR showing Botanical Provinces and Districts from Hopper and Gioia (2004). Shown are (*a*) the species richness in 0.25° latitude by 0.25° longitude grids and (*b*) the richness in 0.25° latitude by 0.25° longitude grids of local endemics.

Grevillea

The Proteaceae is one of the most ancient plant families in present day Australian flora (Dettmann & Jarzen 1998). The forbearers of extant Proteaceae have been found in the fossil pollen records 80-100 million years ago (Dettmann & Jarzen 1998), predating the breakup of Gondwanaland, and records through the entire tertiary showing affinities to numerous genera (Christophel 1989). There are currently 82 genera recognized(Mast *et al* 2008), with 1700 recorded species (Hoot & Douglas 1998),

making it one of the most species rich flowering plant families in the southern hemisphere

Grevillea is the largest genus in Proteaceae (Makinson 2000), and the third largest genus in the Australian flora behind *Acacia* and *Eucalyptus* (George 1998), with representatives in all states of the continent. It is classed in the tribe Grevilleeae, together with *Hakea* and *Finschia*(Weston & Barker 2006) . A phylogenetic analysis showed *Grevillea* to be sister to *Buckinghamia* (Hoot & Douglas 1998). Extensive phylogenetic studies of genera within Proteaceae have utilized fossil records to approximate divergence dates within *Macadamia* (Mast *et al.* 2008) and *Hakea* (Mast *et al.* 2012), and across the whole family (Barker *et al.* 2007). Within these studies, the authors included representatives from *Grevillea*, and dated the genus split with *Macadamia* at 52.7-68.0 Mya (Mast *et al.* 2008), the split from *Buckinghamia* at 45 Mya (Barker *et al.* 2007) and the split from *Finschia* 15 Mya (Sauquet *et al* 2009), which supports a long evolutionary history of *Grevillea* in Australia. Makinson (2000) highlighted that the monophyly of *Grevillea* with regards to *Hakea* is increasingly under question. Most recent results suggest divergence of *Hakea* and *Grevillea* at about 12 Mya (Mast *et al*2012), however, but questioned the reciprocal monophyly of the genera. Resolving this would require a genus wide phylogeny of *Grevillea,* which is currently underway (Mast *et al*, unpublished data). The long evolutionary past of *Grevillea* in Australia does provide unique opportunities to test whether phylogenetic patterns in *Grevillea* support current theories of speciation in southwest of Western Australia region.

The most recent revision of *Grevillea* listed 357 species, of which 45 are cited as vulnerable or endangered under the current federal Environmental Protection and Biodiversity Conservation Act, and 164 were on the most recent Rare or Threatened Australian Plants list (Makinson 2000). *Grevillea* is widely distributed and can be found in every state of Australia (Olde & Marriott 1994). The genus is known for extensive diversity in morphology, ecology and habitat, and is only entirely absent from the high altitude rainforest in northern Queensland and cool-temperate forest of Tasmania and Victoria (Makinson 2000). Species richness for the genus is highest in Western Australia (190 species), of which 117 listed as threatened, rare, or as Priority 1-4, under

the criteria set by the Department of Parks and Wildlife (DPaW, Smith 2013). Several of these species and subspecies are tentatively recognized as phrase named species but not yet published (DPaW).

The *Grevillea* Thelemanniana group

The Thelemanniana group (previously the *thelemanniana* complex) has been revised three times by different authors (McGillivray & Makinson 1993; Olde & Marriott 1994; Makinson 2000) with a general trend of recognizing more taxonomic entities (subspecies or species). George (1998) highlighted that the complex needed major research into the current species delimitations. Molecular data have not previously been applied to the complex, but will be a valuable means to test species hypotheses and improve our understanding of the relationships and taxonomic delineations within this complex. The latest revision of *Grevillea* (Makinson 2000) delimited the 357 species into 44 subgroups, of which the "Thelemanniana Group" included 16 species and 4 subspecies (examples of some in Fig.1.2 & Fig.1.3). The complex is endemic to Western Australia, with the majority of taxa occurring in the southwest region, ranging from Busselton to east of Manjimup/Bridgetown in the south, to Cape Range National Park to the north (Fig.1.4). The overlapping distribution of species within the Thelemanniana group indicates the species geography alone is a poor indicator for species limits. The group is characterized by the floral characteristic of the glabrous ovary being basally abrupt and often rigid (Makinson 2000). They vary little in reproductive features with most differences relating to vegetative morphology. They are generally found in open shrub/woodland of mixed *Eucalyptus*, *Acacia*, *Banksia* and *Melaleuca* over limestone or ironstone with clay, loam or moist sand in wetlands, valleys or creek lines (examples of some habitats in Fig.1.5, Makinson 2000). Makinson (2000) highlighted affinities with the groups Oncogyne and Hakeoides.

The Thelemanniana group has many taxa of high conservation priority. Of the 16 species, 6 are Declared Rare Flora (DRF), 5 are listed as Priority 1 through to 4 (Smith 2013) and 5 are listed as 'Poorly Known' (Makinson 2000). Many of the DRF or priority species in the Thelemanniana group have populations located in or near areas already disturbed or potentially disturbed due to urban expansion. This poses difficulties with conservation plans as there is currently little sound data available for decisions making.

For example *G. evanescens* was proposed for critically endangered status. However, due to its gradation of characters with *G. obtusifolia,* the re-classification has been postponed (G. Keighery, pers. comm., 2010). *Grevillea mccutcheonii* is known from a single population near Busselton, on the Swan Coastal Plain, which is part of an ironstone community that has been recognized for high conservation values due to the high level of species diversity and restricted ranges of species, due mostly to the clearing of almost 95% of the Swan Coastal Plain being cleared natural vegetation (Gibson *et al.* 2000). Another example is the conservation Priority 1 taxon *G. thelemanniana* ssp. Cooljarloo (BJ Keighery 28B), which is known from only one general location at the intersection of the Swan Coastal Plain and Geraldton Sand Plains. These populations are in close vicinity of the Tronox Cooljarloo mine site. This taxon is currently "phrase named", meaning it is considered to be a unique subspecies, it is also under consideration for declaration as 'rare flora', but requires further assessment and clarification of taxonomic status. Another phrase named species is *G.* sp. Gillingarra (R.J. Cranfield 4087) which is currently only known from one isolated population and has been proposed to be part of the *thelemanniana* complex based on leaf morphology and unique habitat (G. Keighery, pers comm., 2010). As taxonomic uncertainty surrounds many of these taxa, a phylogenetic and population genetic approach with DNA-based tools is critical for resolving species limits, phylogenetic relationships and population genetic diversity, thus informing and underpinning more informed conservation decisions.

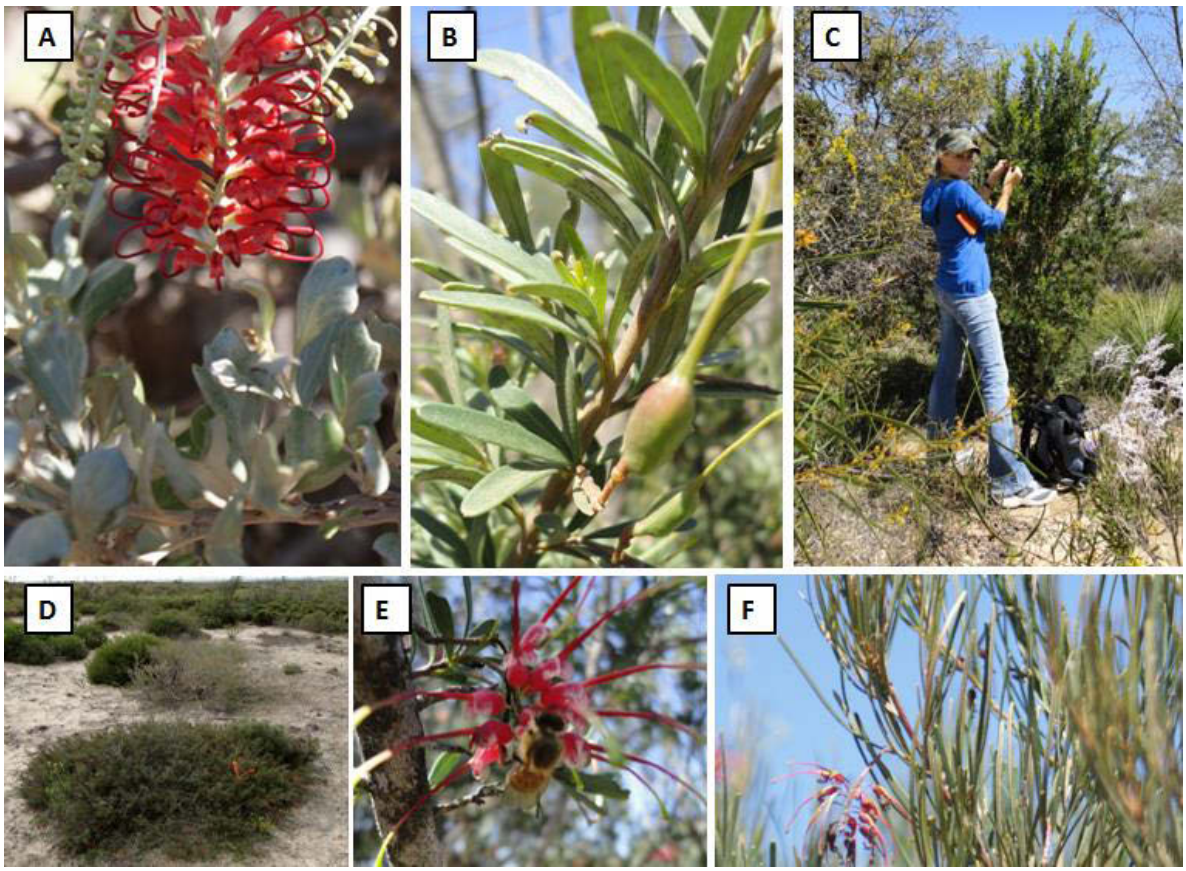

Fig.1.2. Images of select species from the *Grevillea* Thelemanniana group in Western Australia A) *G. variifolia* ssp. *variifolia* B) *G. obtusifolia* C) the author collecting *G. evanescens* D) *G. thelemanniana* ssp. Cooljarloo E) Bee visiting *G. evanescens* inflorescence and F) *G. pinaster*.

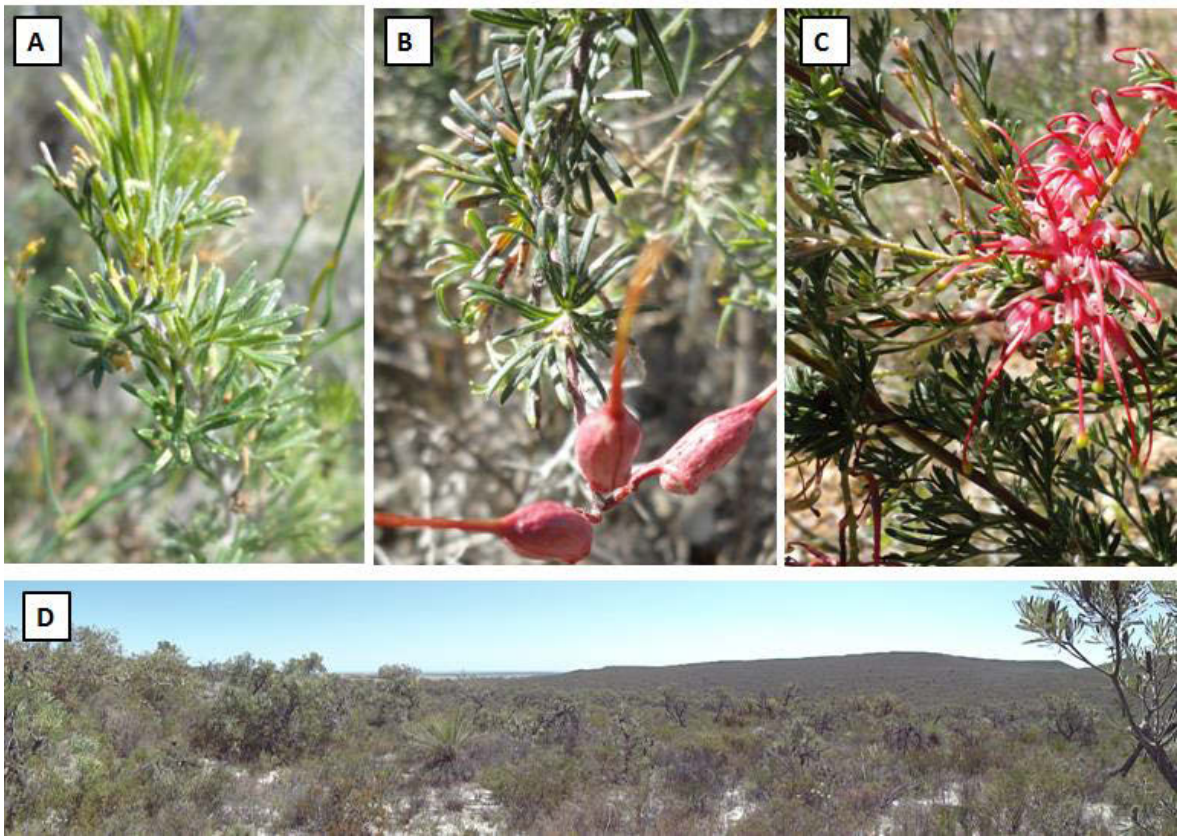

Fig.1.3. A) *Grevillea preissii* ssp. *preissii* B) *G. delta* with seeds C) *G.* sp. Gillingara and D) Overlooking Mt Lesueur National Park.

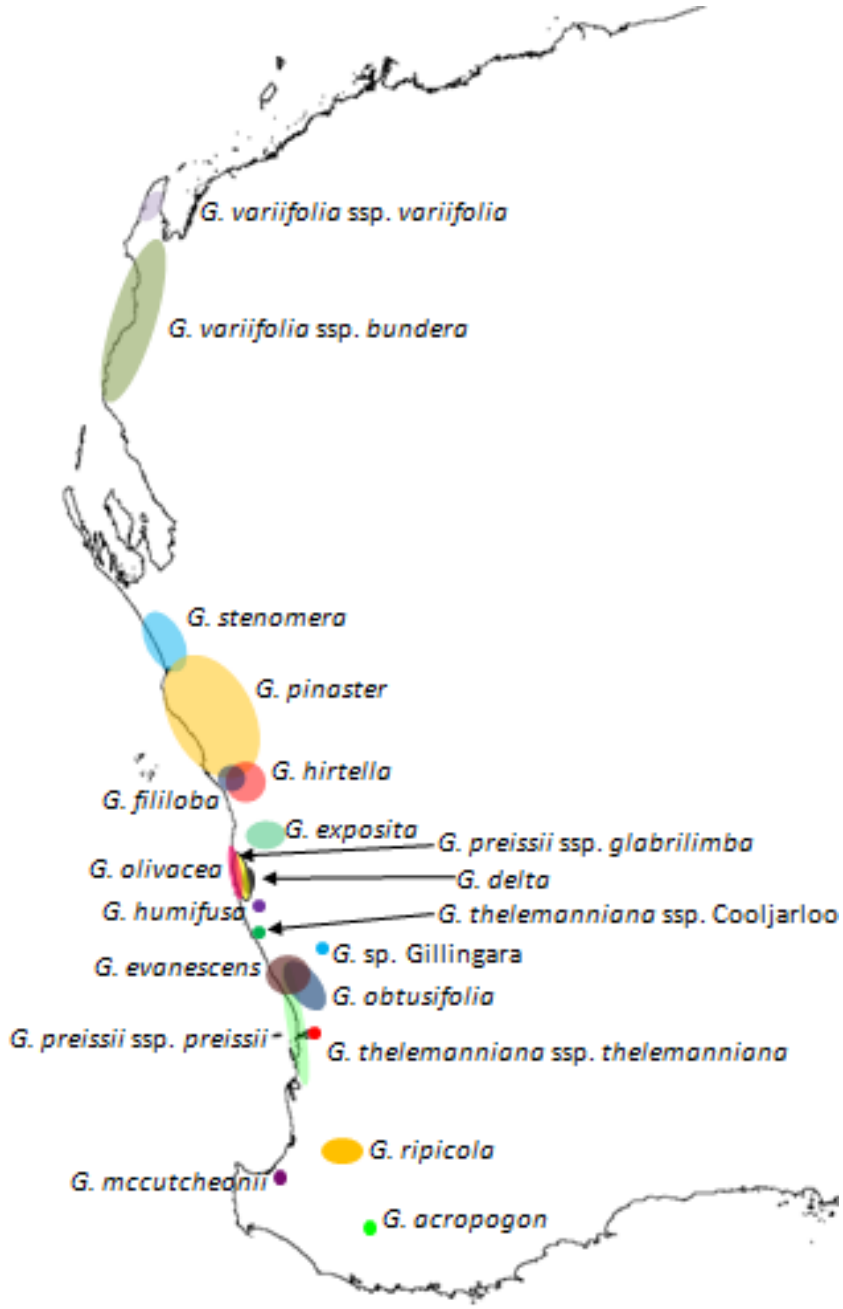

Fig.1.4. Distribution and known ranges of current taxa in the *Grevillea* Thelemanniana group in Western Australia.

Aims of research and structure of thesis

The first broad aim of this research is a systematic study of the *Grevillea* Thelemanniana group using multi-locus data from both chloroplast and nuclear DNA. A key objective within this chapter (Chapter 2) was to assess the taxonomic designation of the two phrase named taxa (*G. thelemanniana* spp. Cooljarloo and *G.* sp. Gillingara). In combination with a phylogenetic assessment of the species complex, I used molecular dating methods to determine the timing of diversification events and whether major patterns could be attributed to known geohistorical events which have occurred in the southwest of Western Australia. In Chapter 3, I combine phylogenetic and population genetic methods to better understand speciation and species boundaries in a subgroup within the species complex. This particular subgroup illustrated some incongruence in gene tree and species tree, and also between chloroplast and nuclear data, which could not be fully resolved in Chapter 2. Distinguishing among current incomplete lineage sorting in this subgroup, potentially due to hybridization, introgression or recent speciation, is particularly important as species and subspecies within the subgroup are of high conservation concern. In Chapter 4, I used a landscape genetics approach, using geographic data, genetic data and riparian models of dispersal to describe the spatial genetic structure and its drivers in the conservation priority taxon *Grevillea thelemanniana* ssp. Cooljarloo, which is currently facing threats from a local mine operation. Chapter 5 provides a description of the development of a microsatellite library for *Grevillea thelemanniana* ssp. *thelemanniana,* and the cross amplification to other species within the complex.

Grevillea in the southwest of Western Australia are largely understudied, especially using molecular data. Quantifying the level of speciation and diversity of *Grevillea* in this region provides information on the diverse processes which have been hypothesised to be responsible for the unique biodiversity of this region. To my knowledge this will be only the second phylogenetic study focusing on *Grevillea* species, and the first on Western Australian species.

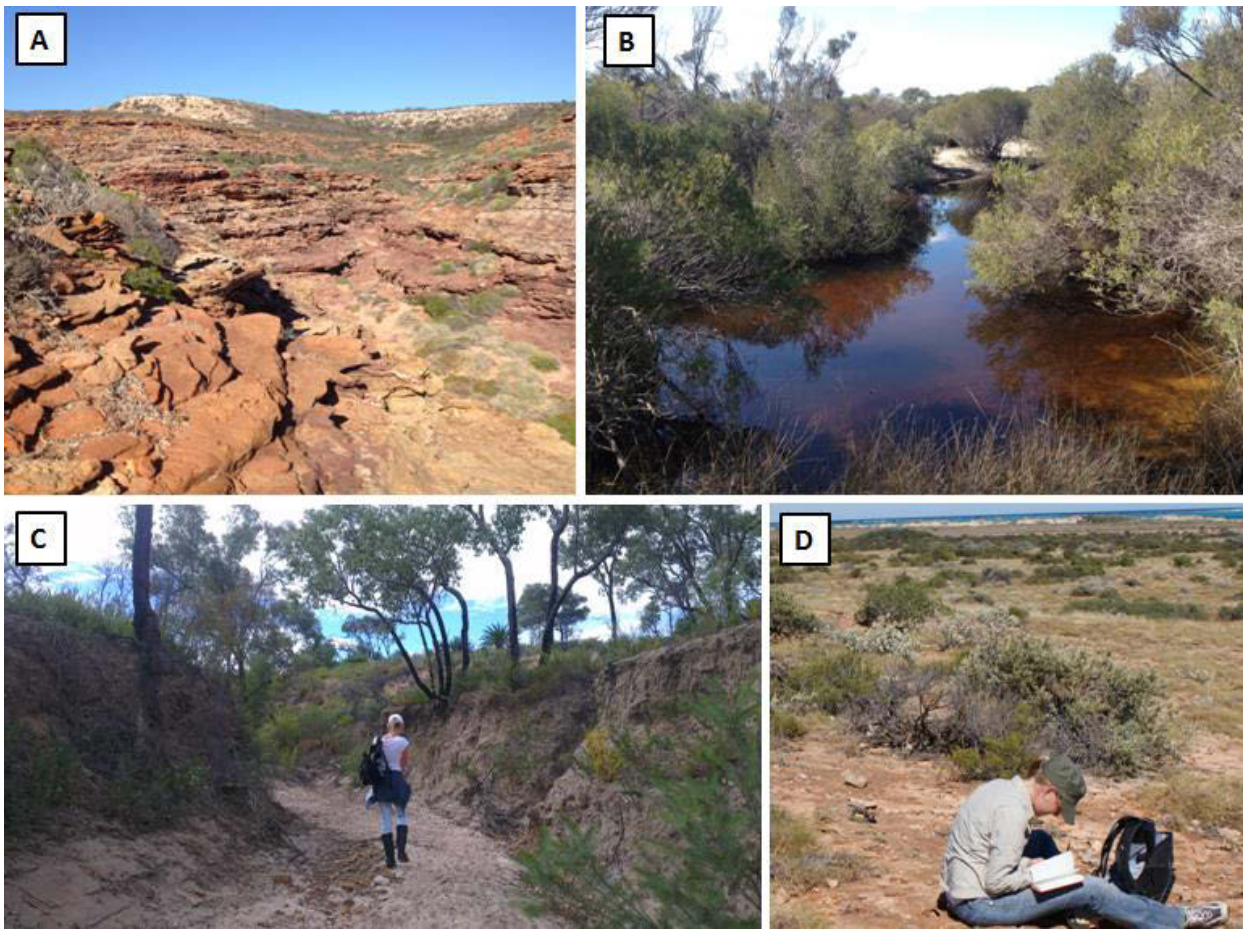

Fig.1.5. Different habitats of *Grevillea thelemanniana* species complex in Western Australia A) coastal cliffs and gorges in Kalbarri B) *G. thelemanniana* ssp. Cooljarloo distribution along creek lines inland from Cervantes C) *G. delta* distribution along creek beds in Mt. Lesueur National Park D) *G. variifolia* along the limestone ranges of Cape Range National Park.

Chapter 2 - Molecular systematics of the *Grevillea* Thelemanniana group (Proteaceae)

Introduction

Western Australia is one of the oldest landscapes on earth. It has remained mostly above sea level, been unglaciated and had no significant mountain building since the Permian, which has resulted in unique opportunities for continued evolution of terrestrial taxa (Hopper 1979; Hopper & Gioia 2004). The Southwest Australian Floristic Region (SWAFR) is particularity high in species richness and endemism, and is recognized as a global biodiversity hotspot (Myers *et al.* 2000), with at least 8,000 native plant species of which 50-75% are endemic (Hopper & Gioia 2004). The SWAFRs diversity has been largely attributed to geohistorical events which have promoted species evolution and diversification in the region. Two bursts of speciation are often referred to, the first occurring during early Tertiary with the onset of aridity and climatic fluctuations; and the second as a result of global cooling and progressive aridity during the late Tertiary and Quaternary (Hopper 1979; Hopper *et al.* 1996; Dettmann & Jarzen 1998; Coates & Hamley 1999; Crisp *et al.* 2004; Hopper & Gioia 2004; Hopper *et al.* 2009). In combination with geohistorical events, strong habitat affinities, complex soil mosaics and poor dispersal abilities have been suggested as reasons limiting migration and serving to isolate species into numerous narrowly distributed populations (Cowling *et al.* 1996; Hopper *et al.* 1996; Hopper & Gioia 2004; Cardillo & Pratt 2013). These isolating factors were further amplified by formation of the Nullarbor plain, a limestone plateau that developed during the Miocene creating an east to west barrier driving divergence between species (Hopper *et al.* 1996; Crisp *et al.* 2004; Hopper & Gioia 2004; Hopper *et al.* 2009).

The SWAFR is characterized by a transitional shift from a maritime coastal climate with high rainfall towards semi arid regions inland with lower rainfall (Hopper 1979). Hopper *et al* (1998) suggested that the climatic characteristics of these two zones act differently on speciation and evolution, with the high rainfall provenance (HRP) exhibiting older relict species, and the semi arid transitional rainfall provenance (TRP) represented by more recent speciation (Fig.2.1). This theory was in part supported in a

recent molecular dating study of *Banksia,* where the authors concluded that the TRP exhibited higher rates of speciation than in the HRP (Cardillo & Pratt 2013). However, Cardillo and Pratt (2013) suggested lower extinction rates in the TRP than in the HRP, and net migration from the TRP to the HRP, which does not support Hoppers' theory that the HRP is characterized by relictual species. These conflicting conclusions indicate that further molecular dating studies on species in this region is needed to test these hypotheses

The combination of climatic and geohistoric events are believed to have led to the particularly high species richness of several plant families in the SWAFR, of which one of the most diverse is the Proteaceae (Hopper *et al.* 1996; Crisp *et al.* 2004; Linder 2008). *Grevillea* is the most species rich genus within Proteaceae, currently comprising 357 species (Makinson 2000). Of these, most occur in Australia, with 5 species in New Caledonia, Papua New Guinea or Indonesia. The ancestors of extant Proteaceae have been identified in the fossil pollen record dating back 80 Mya, possibly even as far back as 100 Mya (Dettmann & Jarzen 1998), with *Grevillea* appearing between 65-80 Mya (Dettmann & Jarzen 1998; Mast *et al.* 2008). The presence of the genus in Australia is likely to have pre-dated the separation of Gondwana (Dettmann & Jarzen 1998), indicating a long evolutionary history in Australia. Today, *Grevillea* occurs across the entire Australian continent, and in most habitats throughout, with the greatest species richness found in Western Australia (ca. 190 species). The genus shows remarkable morphological diversity, especially in vegetative characters, although there is a great deal of plasticity within species.

There have been multiple revisions of *Grevillea* (Johnson & Briggs 1975; McGillivray & Makinson 1993; Makinson 2000), however taxonomic delimitation still remains contentious for several species complexes. Johnson and Briggs (1975) provided one of the most comprehensive early treatments of the genus. Their work was later revised by McGillivray and Makinson (1993), clarifying most of the uncertainties in the nomenclature. Olde and Marriot (1994) provided a descriptive treatment to the classification scheme developed by McGillivray and Makinson (1993), primarily intended as a horticulture guide. Makinson (2000) followed, recognizing 33 higher-order groups within the genus based on morphology. A detailed phylogenetic treatment of the genus using molecular data has not yet been completed, although

there is current progress (A. Mast, unpublished data). Phylogenetic analyses among and within *Grevillea* species complexes can help test hypothesised relationships depicted in the current treatment of the genus. Makinson (2000) points out several *Grevillea* species complexes in need of resolution using genetic data, including the *thelemanniana* complex. To date, only one study has addressed the phylogenetic relationships of *Grevillea* (Pharmawati *et al.* 2004), who investigated 11 species and 4 subspecies from several of the Makinson (2000) subgroups in New South Wales.

Within the *thelemanniana* complex (Table 2.1), the *Grevillea* treatments have recognized varying numbers of species and/or subspecies. The most recent treatment (Makinson 2000) recognized 16 species and 4 subspecies, all endemic to Western Australia, named the Thelemanniana group (referenced this way hereafter). The group ranges from Cape Range National Park in the northwest of Western Australian to the SWAFR, with most species occurring in the transistional rainfall province (TRP) in the SWAFR (Fig.2.1). The members of the group are characterized by: glabrous, truncate (subtriangular in side view) ovaries with basal ridges and a ridged fruit surface (at least near the base) (Makinson 2000). The species and subspecies of the group vary little in reproductive features with most variation occurring in vegetative morphology. The Thelemanniana group has many taxa of high conservation priority. Of the 16 species and subspecies, 3 species are listed as threatened flora and 9 species are listed as Priority 1 through 4 (Smith 2013). The remaining species are listed to designations of "Poorly Known" taxa (Briggs & Leigh 1996). Many of the threatened or priority taxa in the Thelemanniana group have populations located in or near areas already disturbed and/or threatened due to urban expansion. *Grevillea stenomera*, *G. delta*, *G. olivacea*, *G. humifusa* and *G. mccutcheonii* occur near some of Western Australia's most popular holiday spots; Kalbarri, Jurien Bay and the southwest coast. *Grevillea evanescens* and *G. obtusifolia* are located in fragmented populations in the wheatbelt region just north east of the Perth metropolitan area, while *G. thelemanniana* ssp. *thelemanniana* is located in a Perth suburban district and *G. ripicola* is located near the agricultural town of Collie.

Defining species limits has been particularly difficult as there is considerable geographic overlap in distribution of many of the taxa within the Thelemanniana group. Leaf morphology has been the primary source for species delimitation

(Makinson 2000), but there is considerable overlap and intermediacy between some species (eg. *G. evanescens* and *G. obtusifila*, *G. acropogon* and *G. ripicola*). This has posed difficulties for taxonomy and has conservation consequences. For example, the conservation priority 1 taxon *G. thelemanniana* ssp. Cooljarloo (BJ Keighery 28B), is known only from a restricted range (20km^2), north of the SWAFR (Fig.2.1), and occurs 2-8km from a heavy-mineral mine site (Tronox Cooljarloo). This taxon is currently a "phrase named" subspecies, meaning it is hypothesized to be a unique subspecies, but has not been formally assessed or described. It is also currently listed as a Priority 1 taxon, but is poorly known, and requires further assessment. The distribution of *G. thelemanniana* ssp. Cooljarloo is highly disjunct, approximately 150 km north of the nearest population of *G. thelemanniana* ssp. *thelemanniana*. Another phrase named species, *G.* sp. Gillingarra (R.J. Cranfield 4087), is only known from one isolated population (approximately 30 individuals) in the Wheatbelt region (Fig. 1) and has been proposed as a unique species based primarily on habitat and leaf morphology (G. Keighery, pers. comm., 2010). Given the historical taxonomic uncertainty of taxa in this group, molecular data will provide additional insight into resolving species limits and phylogenetic relationships, especially for the purpose of conservation assessment and for understanding the drivers of speciation and divergence in WA. Several studies addressing the origins of diversity of West Australian flora have supported geohistorical events as the trigger for speciation and divergence (Hopper 1979; Coates & Hamley 1999; Richardson *et al.* 2000; Byrne *et al.* 2003; Broadhurst *et al.* 2004; Hopper *et al.* 2009). However, many of these studies have been limited to few species, or restricted to the SWAFR only. The wide distribution and diversity of the Thelemanniana group makes it an ideal candidate for phylogenetic research into the drivers of diversification and speciation both inside and outside of the SWAFR.

DNA-based phylogenies provide morphologically independent results on species relationships and can be an integral tool for testing species limits and delimiting species (Tautz *et al.* 2003; Sites Jr & Marshall 2004; Wiens 2007). Taberlet et al 1991 used chloroplast DNA (cpDNA) to effectively in resolving lower level relationships (genus and species). They used universal primer pairs for cpDNA introns which have dominated the field under the assumption that these introns should be under less functional constraint (Shaw *et al.* 2005) and consequently display the highest

frequency of mutations which can be better applied to lower-level taxonomy (Taberlet *et al.* 1991). The *trnL-trnF* and/or *trnK-matK* regions have been applied successfully among and within a number of Proteaceae genera (Mast & Givnish 2002; Mast *et al.* 2005; Sauquet *et al* 2009; Mast *et al.* 2012; Cardillo & Pratt 2013).

The addition of nuclear DNA markers traditionally used for population level analyses (e.g. AFLP, microsatellites) have increasingly been used in phylogenetic research (Goldstein *et al.* 1995; Ritz *et al.* 2000; Richard & Thorpe 2001; Cook *et al.* 2008; Baker *et al.* 2010). They are particularly useful for delimiting closely related species due to their high polymorphism and distribution across multiple loci (Petit & Excoffier 2009; Duminil *et al.* 2012). Furthermore, microsatellites have been useful in resolving species complexes in other Australian plants (Millar *et al.* 2011; Janes *et al.* 2012; Jones *et al.* 2012). Recently, phylogenetic studies have advocated strongly for the use of multi-locus data originating from different genomes (Duminil & Di Michele 2009; Hausdorf & Hennig 2010; Miralles & Vences 2013) to investigate evolutionary processes such as reticulate evolution, introgression and linkage disequilibrium which often lead to disagreement between species trees and gene trees (Maddison & Knowles 2006). Due to their different modes of inheritance, the use and comparison of cpDNA sequence data and nDNA microsatellite markers are powerful tools for investigating evolutionary history, particularly in testing species limits (Carstens & Knowles 2007; Krak *et al.* 2013), and may provide valuable insights into the Thelemanniana group.

In this study, two cpDNA introns (*trnL-trnF, rps4R2-trnL*) and 12 microsatellite markers developed for *Grevillea thelemanniana* ssp. *thelemanniana* (Hevroy *et al.* 2013) were used for these major objectives: (1) evaluate the monophyly of the Thelemanniana group (2) determine the relationships among species within the complex; (3) assess species limits in the Thelemanniana group as proposed by Makinson (2000); (4) assess whether phrase named species *Grevillea thelemanniana* ssp. Cooljarloo and *Grevillea* ssp. Gillingara should be taxonomically recognised within the Thelemanniana group and (5) address the influence of geohistorical events in Western Australia on speciation and divergence of the Thelemanniana group, in the context of hypotheses proposed by Hopper *et al* (1979) and Hopper and Gioia (2004).

Table 2.1. Taxonomic treatments of the *Grevillea* Thelemanniana group listed alphabetically.

Johnson and Briggs (1974)	McGillivray (1993) - Group 14	Olde & Marriot (1994) - Group 14	Makinson 2000 - Thelemanniana group
			G. acropogon
	G. thelemanniana ssp. *delta*	*G. delta*	*G. delta*
		G. evanescens	*G. evanescens*
		G. exposita	*G. exposita*
	G. thelemanniana ssp. *fililoba*	*G. fililoba*	*G. fililoba*
	G. thelemanniana ssp. *hirtella*	*G. hirtella*	*G. hirtella*
		G. humifusa	*G. humifusa*
			G. mccutcheonii
G. obtusifolia	*G. thelemanniana* ssp. *obtusifolia*	*G. obtusifolia*	*G. obtusifolia*
G. olivacea	*G. olivacea* (Group 21)	*G. olivacea*	*G. olivacea*
G. pinaster	*G. thelemanniana* ssp. *pinaster*	*G. pinaster*	*G. pinaster*
G. preissii	*G. thelemanniana* ssp. *preissii*	*G. preissii*	*G. preissii*
		G. preissii ssp. *preissii*	*G. preissii* ssp. *preissii*
		G. preissii ssp. *glabrilimba*	*G. preissii* ssp. *glabrilimba*
G. ripicola	*G. ripicola*	*G. ripicola*	*G. ripicola*
G. stenomera	*G. stenomera*	*G. stenomera*	*G. stenomera*
G. thelemanniana	*G. thelemanniana* ssp. *thelemanniana*	*G. thelemanniana*	*G. thelemanniana*
G. variifolia	*G. variifolia*	*G. variifolia*	*G. variifolia*
			G. variifolia ssp. *variifolia*
			G. variifolia ssp. *bundera*

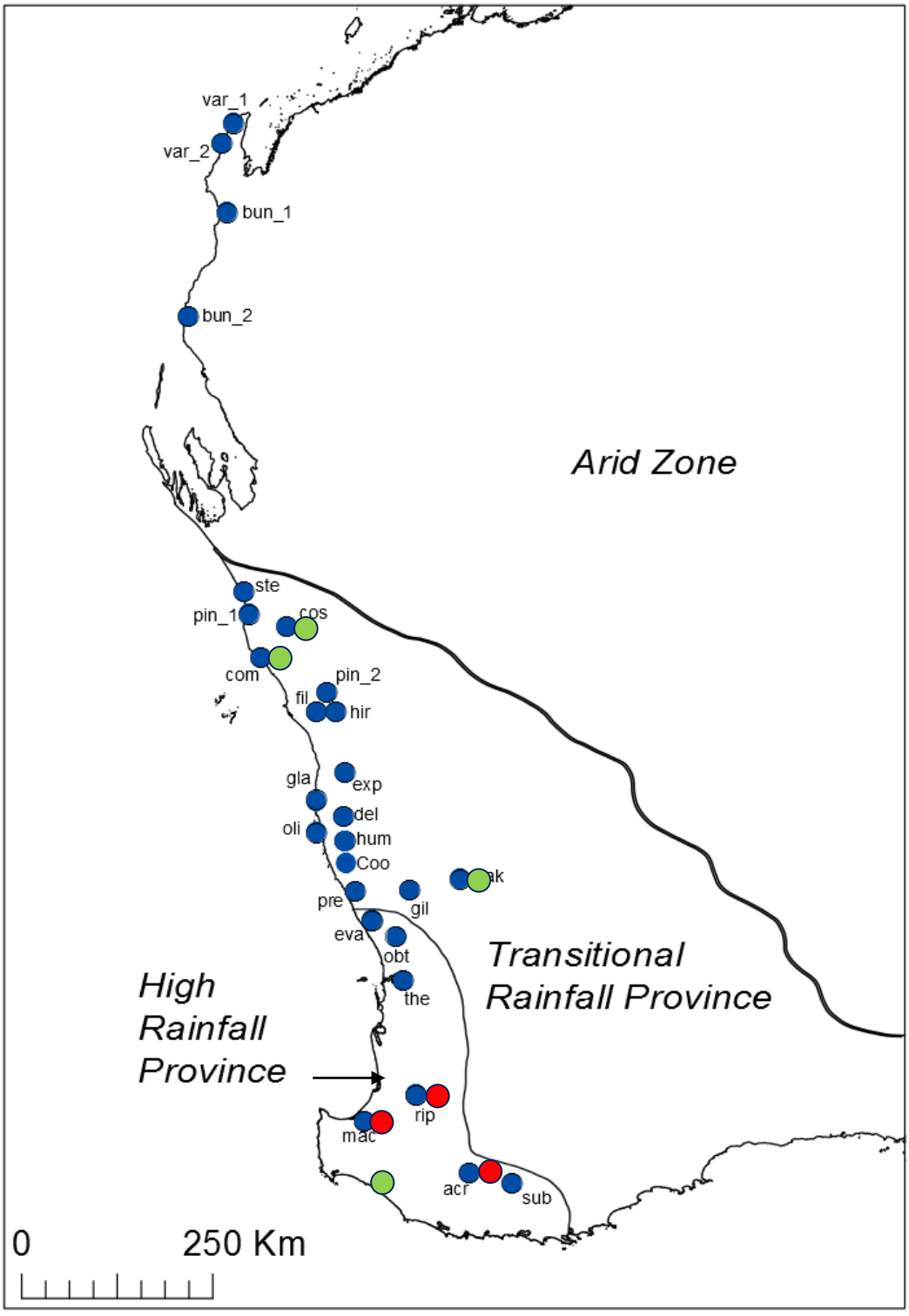

Fig.2.1. Sampling locations of *Grevillea* Thelemanniana group taxa and outgroup species (green circles) in Western Australia. Areas outlined in map are the high rainfall province (HRP) and transitional rainfall province (TRP) which together make up the Southwest Australian Floristic Region (SWAFR), and the Arid Zone. Species and populations sampled are identified by blue (northern lineage) and red (southern lineage) circles and abbreviated species names as in Table 2.2.

Materials and Methods

Taxon Sampling

Sampling included all 18 taxa (species, subspecies and phrased named taxa) from the Thelemanniana group recognized by Makinson (2000) and two phrase named species (Kieghery, pers comm., 2010). For each taxon, 8-10 accessions were collected from 1-3 populations. Multiple populations were collected where possible, however in some cases only one population was known or a second population could not be located where previously described. Geographic locations for collections were based on previously identified herbarium specimens (Department of Parks and Wildlife, WA) and their voucher numbers are listed in Table 2.2. *Grevillea hirtella* is known from a restricted natural range and could not be located (likely due to a recent fire), therefore DNA was extracted from recently collected herbarium specimens (Table 2.2). *Grevillea hakeoides* of the Hakeoids group was included as an outgroup based on Makinson (2000) classifications. Austin Mast (unpublished data) included two species as representative from the Thelemanniana group, *G. acropogon* and *G. preissii,* in his generic phylogeny of *Grevillea.* From his preliminary phylogeny, both Hakeoides and Linearifolia groups were found to be paraphyletic to members of the Thelemanniana group. Mast included *G. diversifolia* ssp. *subtersericata* of the Hakeoides group and *G. costata* of the Linearifolia group, which I also included for here comparison. In addition, *G. commutata* of the Hakeoides group was included due a suspected affiliation with *G. pinaster* of the Thelemanniana group (Digby Growns, pers comm., 2011).

DNA extraction and PCR amplification

Genomic DNA was extracted using the modified CTAB extraction method (Doyle 1990) for all species except *G. evanescens* and *G. stenomera* for which the method of Doyle and Doyle (Doyle 1987) was used, which is useful for species with resinous leaves. Initially I tested 12 cpDNA markers (Shaw *et al.* 2005; Shaw *et al.* 2007) for amplification success and variability on 2-3 accessions from each taxon in the *thelemanniana* complex. Two cpDNA regions were selected for their successful amplification across all taxa and because they exhibited the highest level of variation among species (*trnL-trnF, rps4R2-trnL*). Each region was amplified using the

Polymerase Chain Reaction (PCR). The *trnL-trnF* region was amplified with primers 'Tab-C' and 'Tab-F' and the *rpS4-trnL* region was amplified with primers 'Tab-B' (Taberlet *et al.* 1991) and 'rpS4R2' (Shaw *et al.* 2005). All PCR reactions contained 2-10 µg DNA, 3µl of 5xPCR buffer (Invitrogen, including dNTPs), 2mM $MgCl_2$, 0.1µM of each primer and 0.1µl *Taq* DNA polymerase (Invitrogen) and brought to a total volume of 15µl with sterile deionised water. PCR reactions were conducted as follows: One cycle at 97°C for 2min followed by 30 cycles at 94°C for 1 min, 52°C for 2 min, 72°C for 2 min and a final extension at 72°C for 7 min. PCR product was visualized for amplification success on a 1% agarose gel. Successful amplifications were sent to Macrogen Inc., South Korea, for purification and sequencing using the aforementioned primers. Both strands were sequenced to provide overlap and ensure accuracy in base calling.

Sequence alignment and the treatment of indels

Chloroplast sequences were edited and assembled into contigs using Sequencher v. 3.1.1 (Gene Codes Corp., Ann Arbor, MI). Sequences were trimmed to eliminate low quality data. The sequences were compared to the complete plastid sequence of Nicotiana tabacum L. (Shaw et al. 2007) to determine the beginning and end of each cpDNA intron. Sequence alignment was initially performed using the ClustalW (Larkin et al. 2007) function in the alignment package Mesquite 2.0 (Maddison & Maddison 2004), followed by manual adjustment. Insertion and deletion events (INDELS) were considered phylogenetically informative and were coded as presence/absence characters at the end of the data matrix. I used the JC (Jukes-Cantor) constant rates for analyses of INDEL characters. In total, 6 indels were coded in the trnL-trnF region and 14 in the rpS4-trnL region.

Phylogenetic reconstruction and Molecular Dating

Initial phylogeny of Thelemanniana group and its relationship with the outgroup species was evaluated using maximum likelihood (ML). Best fit nucleotide substitution models were determined for each region using jModelTest2 (Posada 2008) under the Akaike information criterion (AIC) (trnL-trnF: GTR+G+I and rps4-trnL: GTR+G). A ML tree was reconstructed using all 140 concatenated sequences in MEGA 6 (Tamura et al. 2013), using a heuristic Close-Neighbour interchange search under a GTR substitution model. Nodal support was assessed by 1000 bootstrap replicates. Although all 140

sequences were used in the analyses, the resulting tree topology was visualized in FigTree 1.3.1 (Rambaut 2009) and drawn with one representative from each species and haplotype using Inkscape 0.48.2-1 (http://inkscape.org/). Haplotypes (h), nucleotide diversities (π) and neutrality tests were calculated using Tajima's D and Fu's F statistics in DNASP 5.10.01 (Librado & Rozas 2009). A haplotype network was constructed using TCS 1.13 (Clement et al. 2000) with a 90% connection limit. TCS uses the statistical parsimony network approach. Each indel was considered as a single mutation event, and all indels were therefore coded as present/absent in the final alignment.

A Bayesian phylogenetic and dating analysis was performed using the software package BEAST 1.7.5 (Drummond & Rambaut 2007; Drummond *et al.* 2012). Initially, the analyses were performed on all 140 sequences to obtain posterior probability scores to assess the monophyly of individual species. However, for the dating analyses, only the unique haplotypes were used. Calibration points were inferred from a *Grevillea* phylogeny (provided by A. Mast, unpublished data), who used fossil calibration points and a broad sampling across the genus, but only *G. acropogon*, *G. preissii* of the *thelemanniana* complex and *G. diversifolia* and *G. costata* were included from these analyses. A crown node was constrained to include all Thelemanniana group species and all outgroup species except *G. hakeoides*. The common ancestor between *G. preissii* and *G. acropogon* was hypothesised at 24 Mya (± 4). The crown of the southern lineage was constrained based on the common ancestor of *G. diversifolia* and *G. acropogon* hypothesized at 8 Mya (±2). The two calibrated nodes were constrained as monophyletic. Divergence dates were estimated from the partitioned cpDNA data set, unlinking the substitution models, assuming a relaxed clock (uncorrelated lognormal), a Yule Tree prior and a randomly generated starting tree. Calibration points had standard deviations equal to degree of uncertainty from credibility intervals inferred from Mast (pers comm., 2013). Lognormal distribution priors were used for the two calibration nodes. Parameters and trees were estimated from two separate Markov Chain Monte Carlo (MCMC) runs for 50 million generations, sampling every 1000 generations. Tracer v 1.5 was used to check the performance of BEAST for adequate convergence and quality of parameter estimates by examining the likelihood ratio and whether effective sampling size (ESS) was greater than 200

(Drummond & Rambaut 2007; Drummond *et al.* 2012). Summary information and tree construction was implemented through Log Combiner and Tree Annotator 1.7.4 within BEAST package. A Maximum Clade Credibility consensus tree was reconstructed discarding the first 10% of trees as burn-in; the 95% highest posterior density (HPD) intervals for age estimate and posterior probability for each node were displayed using the graphical viewer Figtree 1.3.1.

Microsatellites

PCR amplification and optimization of microsatellite markers

The internal transcribed spacer (ITS) rDNA would not consistently amplify despite the use of Proteaceae-specific primers (Barker *et al.* 2002). Given this, and the low level of variation observed in the chloroplast sequences, a panel of 12 nuclear microsatellite primers, optimized for *Grevillea thelemanniana* ssp. *thelemanniana* (Hevroy et al. 2013) were employed. Since successful amplification was unlikely due to the level of phylogenetic divergence, microsatellites were not cross-amplified to the Hakeoides and Linnearifolia taxa. The microsatellites were amplified using PCR containing 2-10µg DNA, 2µl of 5xPCR buffer (Invitrogen, including dNTPs), 2mM $MgCl_2$, 0.1µM of the fluorescently labelled forward primer (6-FAM, NED &VIC) and unlabeled reverse primer, 0.1µl *Taq* DNA polymerase (Invitrogen) and brought to a total volume of 10µl with sterile water. PCR reactions were performed under the following conditions: One cycle at 94°C for 3 min followed by 30 cycles at 94°C for 40 secs, 54°C for 40 secs, 72°C for 40 secs and a final extension at 70°C for 15 min. PCR products were separated by capillary electrophoresis using a CEQ8800 (Beckman-Coulter). Alleles were sized using the CEQ 8800 Genetic Analysis System program, and Micro-Checker was used to identify any outliers and determine the presence of null alleles in the dataset (Van Oosterhout *et al.* 2004).

Genetic diversity and phylogenetic reconstruction

The genetic diversity of 170 individuals was analysed across the 12 microsatellite loci using GenAlEx 6.5 (Peakall & Smouse 2012). The number of alleles (N_A), effective alleles (N_E), observed and expected heterozygosity (H_O and H_E) and percentage of polymorphic loci (P_L) were calculated and averaged for each locus and population. A

hierarchical analysis of molecular variance (AMOVA) was also carried out to partition the total genetic variation into among species and within species levels.

A neighbour-joining (NJ) tree was generated to estimate phylogenetic relationships between all populations and taxa from microsatellite data. The NJ analysis was calculated from a matrix of Nei's genetic distance (D_A) among all populations, and executed in POPTREE 2.0 (Takezaki *et al.* 2010). D_A is based on the variance in allelic frequencies, which have been repeatedly shown to accurately calculate tree topology under simulation models (Takezaki & Nei 1996; Ritz *et al.* 2000; Richard & Thorpe 2001; Takezaki & Nei 2008), and is especially useful when sample sizes are small (Gaggiotti *et al.* 1999). Bootstrap analyses of 1,000 replicates were used to determine relative level of support for branches.

The model-based Bayesian clustering method implemented in STRUCTURE 2.2 (Pritchard *et al.* 2000) was used to test existing hypotheses of species delimitation based on the use of predefined taxonomic units (e.g. species). STRUCTURE has been primarily used for inferring population structure, where the program assigns individuals to genetic clusters that minimize Hardy Weinberg disequilibrium with no a priori information. However, STRUCTURE has also been applied as a supplementary tool to assess the degree of isolation of provisional species by visualizing the pattern of admixture (Edwards *et al.* 2008; Hausdorf & Hennig 2010; Duminil *et al.* 2012; Miralles & Vences 2013). Shaffer and Thomson (2007) suggested STRUCTURE is a useful tool in delimiting species of recent lineage diversification. The predefined populations were tested at two *K*-values, as either species (K = 17) or species and sub species (K = 20). Since admixture between clusters also represents co-ancestry and gene flow between individuals (Willis *et al.* 2012), I can test species hypotheses by identifying clusters of populations and comparing them to their taxonomic treatment. Divergence and subgroups recognised in chloroplast results were also investigated through Structure analyses by visualizing K at 2 – 5. For the analyses, the following parameters were used; admixture model with allele frequencies correlated among populations, ten runs of each *K*-value and 1 million Markov chain Monte Carlo generations with the first 100,000 discarded as burn-in.

Results

Chloroplast sequence

Once all 140 accessions were aligned, the two combined cpDNA regions totaled 1949 base pairs in length. Nucleotide diversity (π) was low; 0.0065 (0.008 when *G. hakeoides* was included). There were 41 parsimony-informative characters including 20 indels. Tajima's *D* and Fu & Li's *D* showed no significant deviation from neutrality (*P* > 0.01). The concatenated dataset resulted in 25 unique haplotypes (Table 2.2). Several species shared a cpDNA haplotype: *G. evanescens, G. exposita, G. fililoba, G. pinaster, G. preissii* ssp. *glabrilimba, G. variifolia* ssp. *variifolia* and *G. variifolia* ssp. *bundera* (Hap_3) and; *G. thelemanniana* ssp. *thelemanniana* and *G. obtusifolia* (Hap_15).

Phylogenetic reconstruction of chloroplast region

Posterior probability (PP) from Bayesian analyses and bootstrap support (BS) scores from maximum likelihood (ML) analyses are presented in the ML topology (Fig.2.2). Two lineages were supported in all analyses (1.0 PP, 100 BS): (1) a southern lineage including *G. acropogon, G. mccutcheonii, G. ripicola* and *G. diversifolia* ssp. *subterisercata* of the *Hakeoides* complex and (2) the northern lineage including the remainder of the *thelemanniana* group taxa and *G. commutata* of the Hakeoides group. There was low resolution within the northern lineage, however three subclades are recognized. Subclade 1 consisted of *G*. *thelemanniana* ssp. Cooljarloo *+ G. delta + G. humifusa* (.99PP, 50 BS); subclade 2 of *G. stenomera* + *G. evanescens* (.99PP, 50 BS) and subclade 3 is consisted of *G. preissii* + *G*. *olivacea* [Hap 17] (0.99 PP, 50 BS). *Grevillea costata* of the Linearifolia group resolved sister to the northern lineage with strong support (1.0 PP, 100 BS). There was strong support for monophyly of several species. Importantly, both phrase named species *G. thelemanniana* ssp. Cooljarloo and *G.* sp. Gillingara are supported as monophyletic (1.0 PP, 82 BS and 1.0 PP, 86 BS, respectively). There was also strong support (1.0 PP, >95 BS) for the monophyly of species; *G. ripicola G. diversifolia* ssp. *subterisercata* in the southern lineage and *G. commutata* and *G. hirtella* in the northern clade*. Grevillea stenomera* is paraphyletic in the northern clade.

The haplotype network showed 24 unique haplotypes, of which 22 were species specific (Fig. 2.3). The species that make up the southern clade did not connect with

the northern lineage at a 95% connection limit, but did at a 90% connection limit with a long branch of 22 mutational steps. *Grevillea costata* of the Hakeoides group connected to the northern lineage at 95%, albeit by 14 mutational steps. The network was dominated by one haplotype (Hap_3) shared by 5 species and 2 subspecies. Hap_3 ranges from the northern most part of the range (*G. variifolia*) to the northern Swan Coastal Plain (*G. evanescens*). There is little evidence of phylogeographic structuring in the northern lineage, where haplotypes were connected by one mutational step and few inferred haplotypes. *Grevillea evanescens, G. obtusifolia G. pinaster*, *G. preissii* ssp. *glabrilimba* and *G. variifolia* ssp. *bundera* were all collected from two populations and have two unique haplotypes. In all of these species, haplotypes were shared among populations (Table 2.2). On the other hand, *G. stenomera, G. olivacea* and *G. preissii* all had two haplotypes within a single population. Molecular variance was partitioned significantly by the AMOVA of cpDNA: among lineages (68%, $P < 0.0001$), among species (30.8%) and within species (1.2%, $P < 0.0001$; Table 2.3).

Molecular dating analyses

The mean of most age estimates of species within both the northern and southern lineages was <10Mya (Fig.2.4), diverging during the late Miocene/ early Plicocene (1.6 – 7 Mya) into the Pleistocene (1.6 – 0 Mya). *Grevillea costata* of the Linearifolia group is estimated to have emerged after the divergence of the northern and southern lineages 15.3 Mya (95% HPD = 12.9 – 24.4). *Grevillea commutata* of the Hakeoides group diverged from the northern lineage at 6.5 Mya (95% HPD = 4.2 – 11.3).

Microsatellites

Twelve microsatellite loci were cross-amplified across most species of the Thelemanniana group, with polymorphism within species ranging from six to all 12 loci (Table 2.4). Monomorphic genotypes were included in the analyses as they were unique among species. Three loci failed to amplify in *G. mccutcheonii,* two in *G. acropogon* and *G. ripicola,* and one in *G. stenomera.* The number of alleles, effective alleles, expected heterozygosity and observed heterozygosity were low across all species and subspecies, as is expected from small sample sizes. The AMOVA partitioned the majority of variation among species (62%), but also relatively high variation within species (38%; Table 2.3).

Analyses of microsatellite data

While the neighbour-joining tree was resolved, there was no support for the northern and southern lineages, as in the cpDNA data, nor was there support for most higher level relationships (Fig.2.5). The relationship between *G. thelemanniana* ssp. Cooljarloo and *G. delta* from subclade 1 was resolved. However, the inclusion of *G. humifusa* (as seen in the cpDNA tree) was not recovered in the NJ tree. The close relationship between *G. evanescens, G. obtusifolia* and *G. thelemanniana* ssp. *thelemanniana* from subclade 2 reflected the cpDNA tree, however *G. stenomera* showed no affinities in the NJ tree. Sublcade 3 was not recognised in the NJ tree. The neighbour-joining tree supported few relationships among species, however there was strong support for monophyly of several species sampled from multiple populations.

In most cases, STRUCTURE analyses assigned populations to clusters representing recognized species and subspecies at both *K* values (17 and 21) (Fig.2.6). The average assignment posterior probabilities were high for most clusters; *qi* = 0.88 for *K* = 17 and *qi* = 0.86 for *K* = 20. At *K* = 17, populations of *G. evanescens, G. obtusifolia* and *G. thelemanniana* ssp. *thelemanniana* were assigned to the same genetic cluster. *Grevillea acropogon* and *G. mccutcheonii* were also assigned to the same genetic cluster. The remaining species were recognized as distinct genetic clusters with limited admixture between them. Populations from the same species were assigned to the same cluster, except for *G. delta,* where one population admixed primarily with *G. thelemanniana* ssp. Cooljarloo. Genetic clusters at *K* = 20 were similar to *K* = 17, however *G. thelemanniana* ssp. *thelemanniana* and *G. mccutcheonii* were recognised as a unique genetic cluster, as was the population from *G. delta* which was recognised as having admixture with *G. thelemanniana* ssp. Cooljarloo at *K* = 17. A third unique genetic cluster was recognised as one population from *G. obtusifolia,* whilst the second population showed admixture or genetic identity with the *G. evanescens* cluster. *Grevillea acropogon* and *G. ripicola* assigned to the same genetic cluster. The analyses at low *K* values did not reveal support for the divergence between northern and southern lineages, nor were the subgroups revealed in the chloroplast data supported by the STRUCTURE analyses.

Table 2.2. List of the taxa and populations of the *Grevillea* Thelemanniana group and outgroups included in the analysis, current conservation status as listed by the Western Australia Department of Parks and Wildlife (Priority 1-4, T= Threatened and CR = Critically Endangered). Western Australian Herbarium (PERTH) specimen ID used as location source of collection, number of individuals used for cpDNA; H, haplotype number; ID, three letter code and GPS coordinates (Lat/Long) for each collection.

Voucher	Taxon	Status	cpDNA	H	ID	Source
7249462	*G. acropogon*	T	8	1	acr	34°17'39.00"S 116°45'25.00"E
6340024	*G. delta*	2	5	5	del	30°08'43.00"S 115°11'23.30"E
6797873	*G. delta*		2	5	del2	30°03'46.60"S 115°10'44.00"E
6097553	*G. evanescens*	1	3	3,6	eva	31°19'05.09"S 115°37'09.80"E
7007205	*G. evanescens*		3	3,6	eva2	31°21'46.40"S 115°41'44.30"E
7213697	*G. exposita*		5	3	exp	29°37'57.40"S 115°13.53.80"E
4160096	*G. fililoba*		6	3	fil	28°53'07.71"S 114°56'10.90"E
	G. sp. Gillingarra	CR	5	7	gil	30°57'35.80"S 116°03'13.90"E
6330983	*G. hirtella*	3	1	9	hir	29°00'22.50"S 114°56'54.20"S
8157928	*G. hirtella*		1	9	hir	28°59'49.20"S 115°20'20.40"S
5942888	*G. hirtella*		1	9	hir	28°52'47.00"S 115°11'42.90"S
3213048	*G. humifusa*	T	6	10	hum	30°17'25.60"S 115°13'42.20"E
4168186	*G. mccutcheonii*	T	8	14	mac	33°39'57.20"S 115°34'32.50"E
8118272	*G. obtusifolia*		3	15	obt	31°37'30.50"S 116°00'49.10"E
7007167	*G. obtusifolia*		2	15,16	obt2	31°28'04.68"S 115°56'00.78"E
7908687	*G. olivacea*	4	5	17,18	oli	30°12'32.09"S 115°01'49.74"E
7440111	*G. pinaster*		3	3	pin	27°45'15.87"S 114°14'11.08"E
7275951	*G. pinaster*		3	3,19	pin2	28°48'07.50"S 115°04'36.20"E
4368711	*G. preissii* ssp. *preissii*		5	20,21	pre	30°58'45.97"S 115°25'02.16"E
6197167	*G. preissii* ssp. *glabrilimba*		5	3,8	gla	29°51'58.10"S 115°01'37.90"E
7909306	*G. preissii* ssp. *glabrilimba*			3	gla2	29°49'49.72"S 115°07'48.63"E
5909929	*G. ripicola*	4	7	22	rip	33°21'16.80"S 116°05'30.80"E
4157559	*G. stenomera*	2	5	23,24	ste	27°14'35.07"S 114°06'25.00"E
2161532	*G. thelemanniana* ssp. *thelemanniana*	4	3	15	the	32°01'15.00"S 115°59'12.00"E
2161516	*G. thelemanniana* ssp. *thelemanniana*		2	15	the2	32°01'30.00"S 115°58'40.00"E
2161583	*G. thelemanniana* ssp. *thelemanniana*		2	15	the3	32°00'50.00"S 115°59'10.00"E
6378498	*G. thelemanniana* ssp. Cooljarloo	1	2	4	coo	30°31'38.01"S 115°13'54.70"E
	G. thelemanniana ssp. Cooljarloo		2	4	coo2	30°34'00.77"S 115°21'17.05"E
	G. thelemanniana ssp. Cooljarloo		3	4	coo3	30°40'50.87"S 115°18'43.57"E
7431139	*G. variifolia* ssp. *variifolia*		2	3	var	21°58'00.09"S 113°57'03.24"E
7431155	*G. variifolia* ssp. *variifolia*		3	3	var2	22°02'04.14"S 113°55'02.34"E
4055217	*G. variifolia* ssp. *bundera*		6	3,2	bun	22°58'09.17"S 113°54'03.67"E
4997697	*G. variifolia* ssp. *bundera*			2	bun2	24°11'05.07"S 113°27'03.99"E
Outgroups						
	Hakeoides group					
7649826	*G. hakeoides*		6	25	hak	30°51'26.00"S 116°42'07.80"E
6452019	*G. commutata*		6	13	com	27°49'05.00"S 114°08'39.90"E
7061757	*G. diversifolia* ssp. *subtersericata*		5	11	sub	34°46'35.00"S 116°58'51.00"E
	Linearifolia group (Costata subgroup)					
4160207	*G. costata*	3	6	12	cos	27°53'01.50"S 114°34'04.80"E

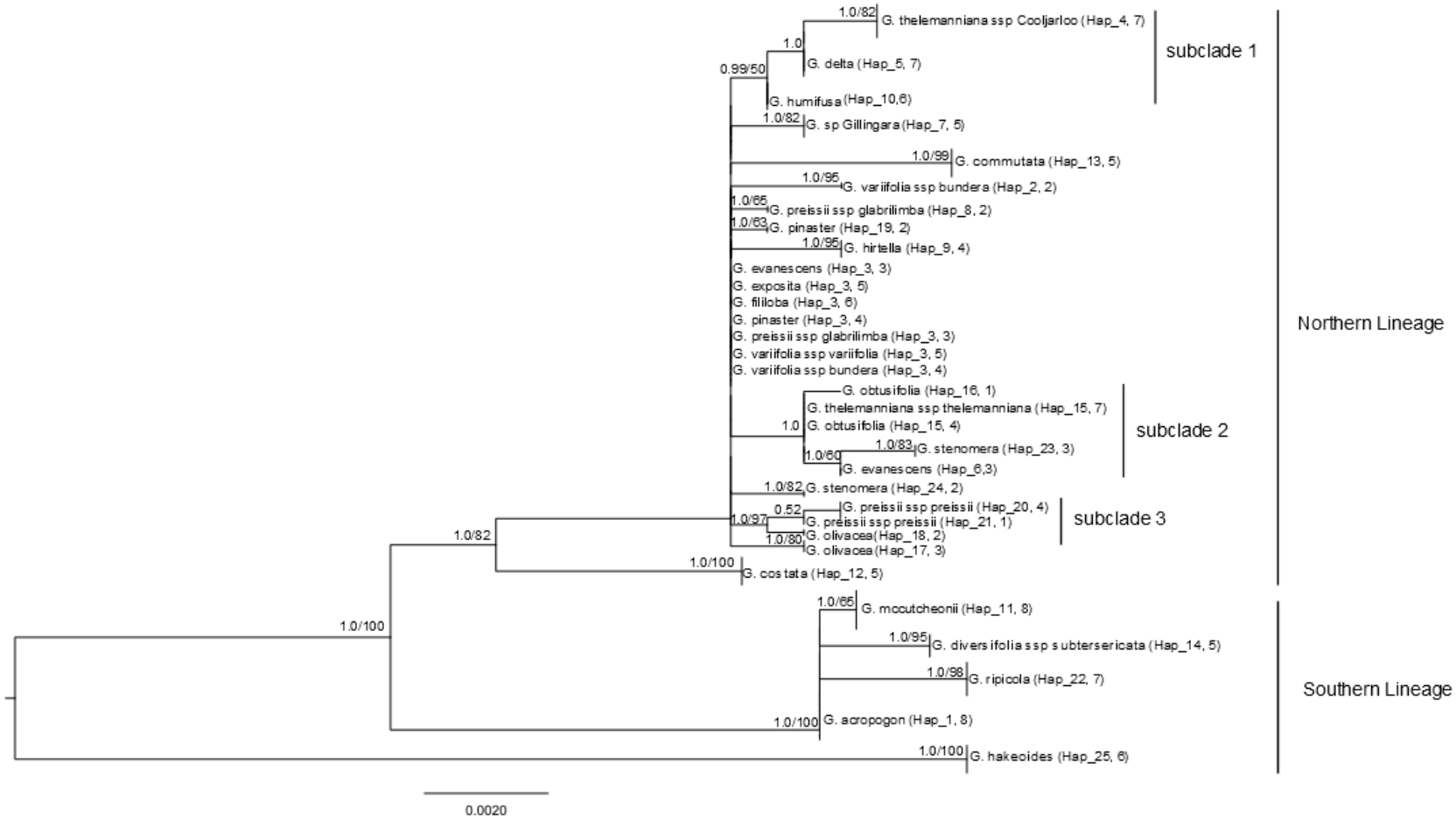

Fig.2.2. Maximum likelihood phylogram based on the concatenated cpDNA data set showing two major lineages within the *Grevillea* Thelemanniana group taxa and four outgroup taxa indicated in bold. The haplotype numbers as well as the number of accessions per haplotype are in brackets (see also Table 2.2). Bootstrap values >50% shown above branches, Bayesian posterior probabilities below branches and branch lengths correspond to the genetic distances (subtitutions per site).

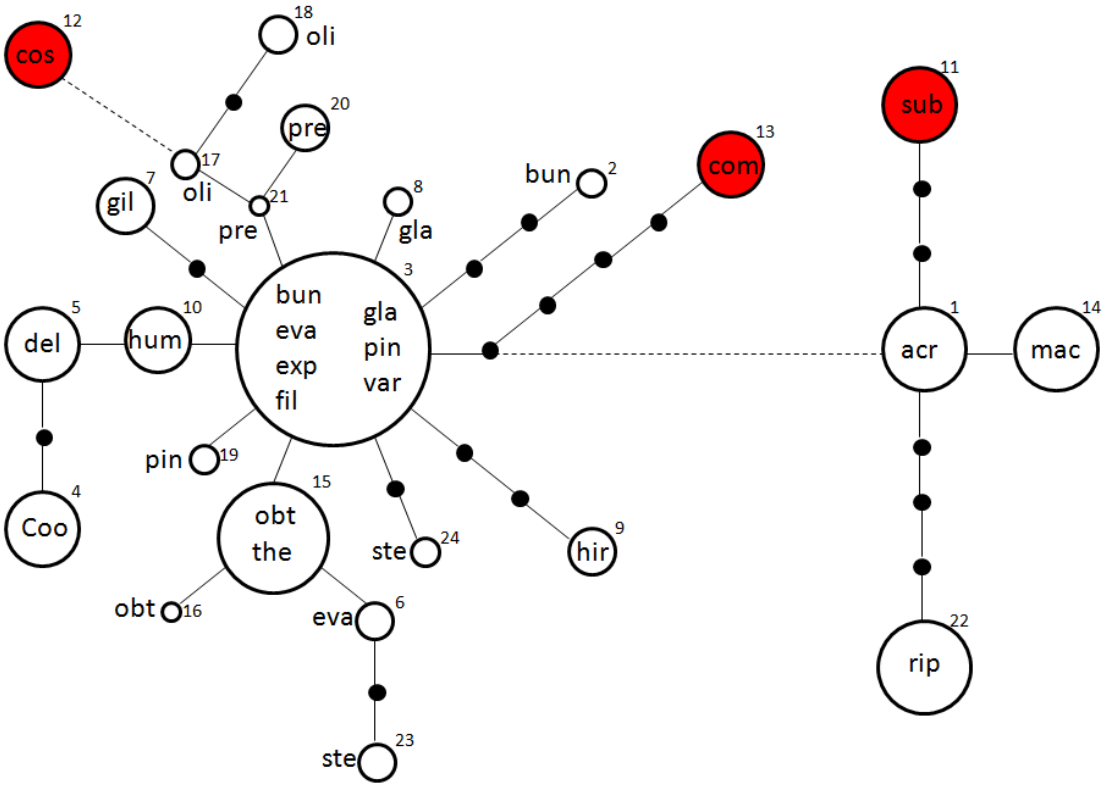

Fig.2.3. Statistical parsimony network of 24 cpDNA haplotypes found in the *Grevillea* Thelemanniana group including outgroup species *G. costata, G. diversifolia* ssp. *subterisercata* and *G. commutata* illustrated by red circles. Circle sizes correspond to the number of accessions possessing the haplotype. Species names are abbreviated by the first three letters of the species or subspecies epithet. Number represents unique haplotype number from Table 2.2. Black circles represent inferred extinct/missing haplotypes. Dashed black line represents 22 extinct/missing haplotypes (connection limit 90%) between the southern and northern lineage and 14 steps between *G. costata* and the northern lineage.

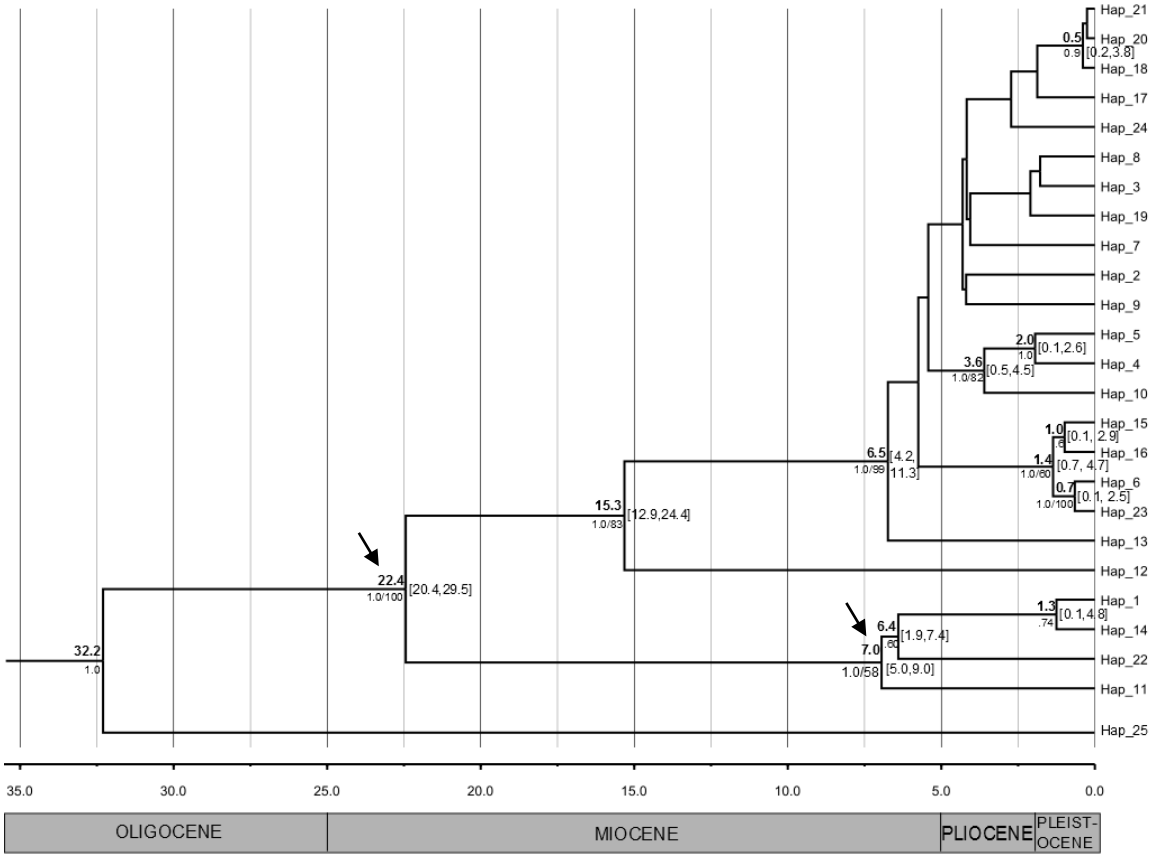

Fig.2.4. Maximum clade credibility chronogram of all *Grevillea* Thelemanniana group and outgroup species, illustrating haplotypes only (Table 2.2). The numbers in bold above branches is the mean inferred age (Millions of years). Numbers below axis represent posterior probabilities derived from Bayesian inference and bootstrap support derived from maximum likelihood analyses (PP: BS). Numbers in brackets are describing 95% HDP intervals for node ages. Arrows indicate nodes used for calibration points.

Table 2.3. Hierarchical analyses of molecular variance (AMOVA) for microsatellites and cpDNA sequence data from the *Grevillea* Thelemanniana group.

Microsatellites			
Source of variation	*df*	Percentage of variation	Significance
Among species	29	62	$P < 0.0001$
Within species	140	38	$P < 0.0001$
Total	169		
cpDNA			
Among lineages	1	68	$P < 0.0001$
Among species within lineages	22	30.8	$P < 0.0001$
Within species	117	1.2	$P < 0.0001$
Total	121		

Table 2.4. Genetic diversity measures for all sampled species of the *Grevillea* Thelemanniana group based on analysis of 12 microsatellite loci. *N*, sample size; N_A, average number of alleles per loci; N_E, average effective number of alleles per locus ; H_O and H_E, observed and expected heterozygosity; P_L, number of polymorphic loci (/12).

Collection	N	N_A	N_E	H_O	H_E	#P_L
G. acropogon	8	2.167	1.820	0.250	0.324	7
G. delta	4	2.417	1.752	0.250	0.362	11
G. delta 2	5	2.750	2.110	0.317	0.410	9
G. evanescens	3	1.833	1.663	0.222	0.269	6
G. evanescens 2	5	2.417	1.821	0.175	0.341	9
G. exposita	8	2.500	1.873	0.333	0.371	10
G. fililoba	9	3.667	2.220	0.434	0.495	12
G. sp. Gillingarra	10	1.667	1.154	0.075	0.108	6
G. hirtella	5	3.333	2.467	0.533	0.512	11
G. humifusa	8	2.167	1.654	0.231	0.256	7
G. mccutcheonii	8	1.417	1.011	0.100	0.146	6
G. obtusifolia	5	2.917	2.268	0.400	0.462	11
G. obtusifolia 2	4	2.583	2.048	0.292	0.393	10
G. olivacea	9	2.500	1.734	0.241	0.284	8
G. pinaster	5	3.500	2.698	0.383	0.537	11
G. pinaster 2	4	3.167	2.505	0.438	0.505	10
G. preissii ssp. *preissii*	8	2.833	1.939	0.308	0.338	8
G. preissii ssp. *glabrilimba*	5	2.583	1.993	0.267	0.385	9
G. preissii ssp. *glabrilimba 2*	4	2.417	2.015	0.313	0.357	8
G. ripicola	8	2.083	1.609	0.177	0.303	7
G. stenomera	8	2.250	2.222	0.438	0.427	7
G. thelemanniana ssp. *thelemanniana*	4	2.583	1.830	0.528	0.370	10
G. thelemanniana ssp. *thelemanniana* 2	3	2.083	1.596	0.417	0.282	9
G. thelemanniana ssp. *thelemanniana* 3	3	1.833	2.236	0.281	0.410	8
G. thelemanniana ssp. Cooljarloo	6	2.917	1.870	0.438	0.393	8
G. thelemanniana ssp. Cooljarloo 2	4	2.417	1.835	0.368	0.373	7
G. variifolia ssp. *variifolia*	4	2.250	1.404	0.313	0.255	9
G. variifolia ssp. *variifolia 2*	4	1.750	2.281	0.317	0.470	8
G. variifolia ssp. *bundera*	5	3.083	1.568	0.208	0.299	10
G. variifolia ssp. *bundera 2*	4	1.917	1.590	0.186	0.259	7

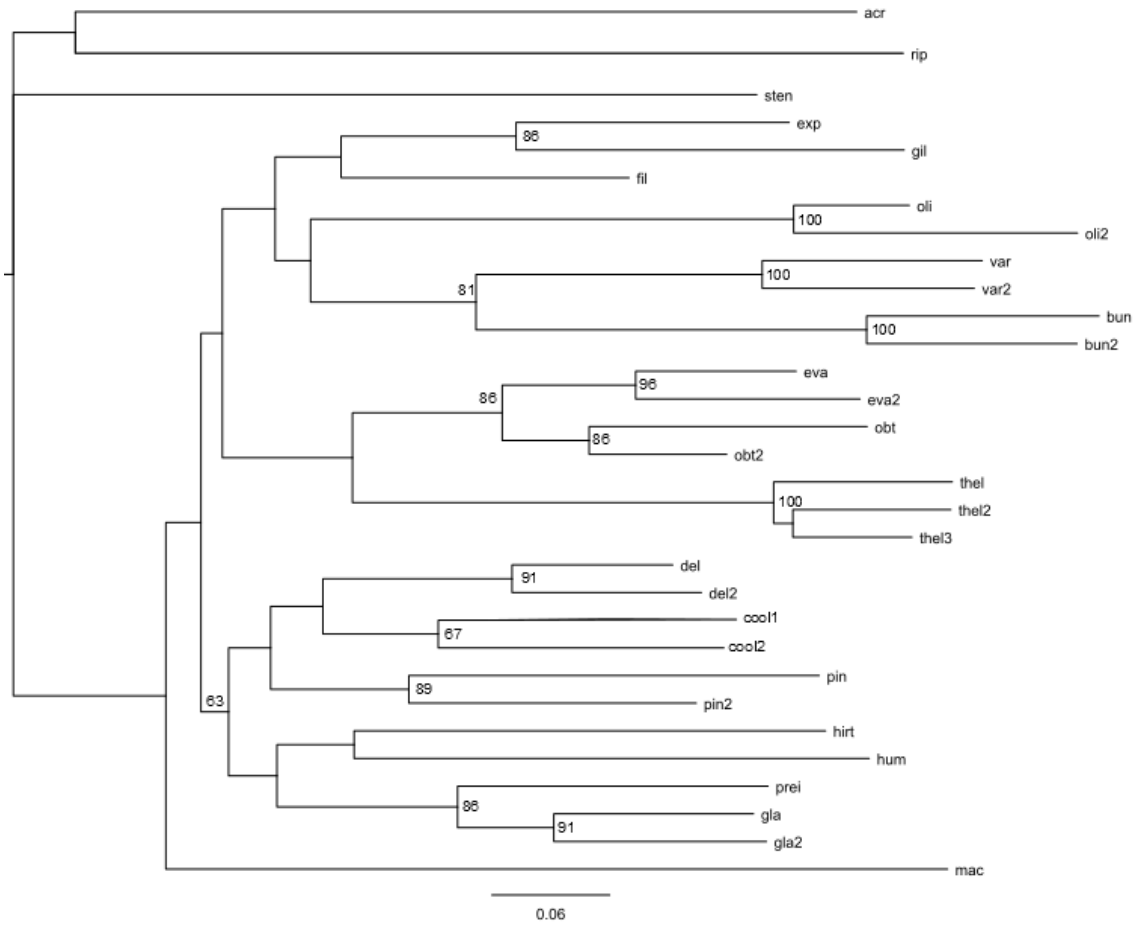

Fig.2.5. Neighbour-joining tree of *Grevillea* Thelemanniana taxa reconstructed from Nei's genetic distance (Da), derived from the allelic frequencies of 170 individuals from 12 microsatellite markers. Bootstrap support values over 50% are given and axis represents genetic distance. Species names are abbreviated by the first three letters of the species or subspecies epithet (Table2.2).

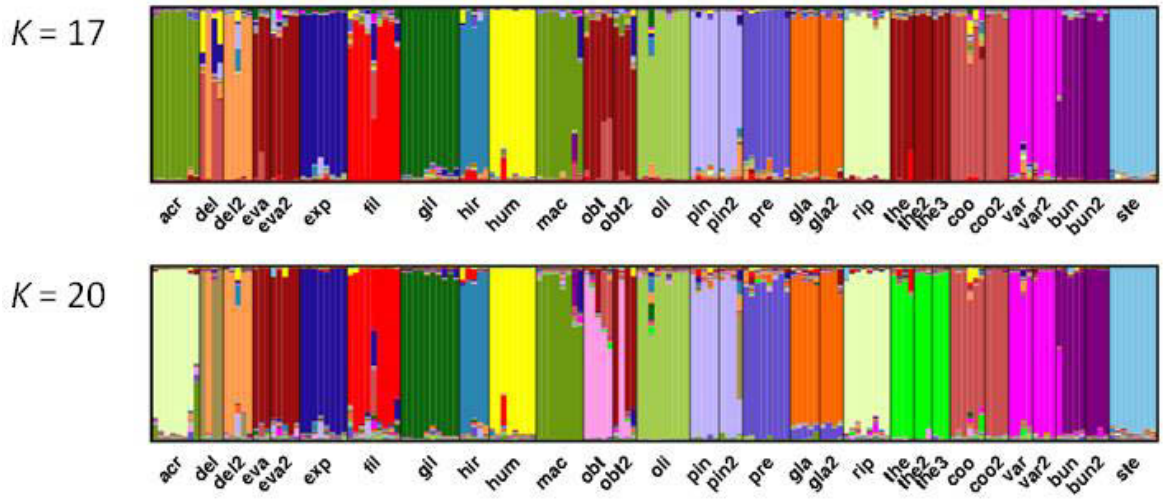

Fig.2.6. Bayesian analysis of microsatellite data for the *Grevillea* Thelemanniana group using STRUCTURE. Bar-plots represent the genotypes given in pre-defined species (*K*=17) and species and subspecies (*K*=20) and coloured following their membership to the K-dependent clustering. Individuals with multiple colours have admixed genotypes. Species and population number/code (as specified in Table 2.2) are given alongside the relevant bar-plots.

Discussion

Phylogeny of the Thelemanniana group

One of the most striking results in this study was the well-supported deep divergence between Northern and Southern clades of the Thelemanniana group. The southern lineage consisted of *G. acropogon, G. ripicola, G. mccutcheonii,* including *G. diversifolia* ssp. *subtersericata* of the Hakeoides group. The northern lineage consisted of the remaining Thelemanniana species, including *G. costata* of the Linearifolia group and *G. commutata* of the Hakeoides group. The two lineages correspond to a northern and southern geographic distribution in southeastern Australia. The presence of two strongly divergent lineages in the chloroplast data confirms preliminary results from Mast *et al* (unpublished data), whose generic phylogeny of *Grevillea* supported a Miocene lineage divergence between sampled species of the Thelemanniana group. The strong partitioning of total chloroplast variation by AMOVA between the northern and southern lineages supports the divergence between the northern and southern lineages (68%). Overall sequence variation was very low within the lineages, with only 1.2% of the variation within species, especially in the northern lineage. Low chloroplast

variation in the *trn*L - *trn*F sequence region has been reported in previous phylogenetic studies on plants, especially at lower taxonomic levels (Krak *et al.* 2013).

Monophyly of the Thelemanniana group

Phylogenetic analysis of the cpDNA data did not support the monophyly of the Thelemanniana group as proposed by Makinson (2000), as of the group was paraphyletic in relation to species from the Hakeoides and Linearifolia groups. Makinson (2000) proposed a close relationship of the Thelemanniana group with species from the Hakeoides group, a hypothesis which has been supported by this study. Paraphyly of the Thelemanniana group with species of the Hakeoides group also supports initial findings from Mast *et al.* (unpublished data). His results showed *G. diversifolia* ssp. *subtersericata* as sister to *G. acropogon* of the southern lineage and *G. costata* as diverging from the same lineage as *G. preissii* ssp. *preissii,* from the northern lineage. Furthermore, this study found *G. commutata,* also of the Hakeoides group, to be nested within the Thelemanniana group rather than being closely related to *G. hakeoides* of the Hakeoides group. Pharmawati et al. (2004) also uncovered incongruence between the morphological classifications of Makinson (2000) and molecular data, as their phylogeny did not support the morphological classification regarding four species in the Linearifolia group.

Genetic relationships among species of the Thelemanniana group

There was little resolution of relationships among species outside the split between the Northern and Southern clades. In fact many species shared common haplotypes. The sharing of the central haplotype (Hap_3) among several morphologically well defined species in the northern lineage, may have occurred through two processes; (1) hybridization and/or introgression; or (2) incomplete lineage sorting through the retention of an ancestral haplotype. The distribution of species with shared haplotypes can give insight into the evolutionary history that leads to the occurrence (Meudt & Bayly 2008; Krak *et al.* 2013). The species which shared a haplotype are distributed across a wide geographic range with no known overlap of populations, which provides a strong argument against recent hybridization and introgression, since this would require close geographic proximity or overlapping populations (Rieseberg 1998). Furthermore, structure analyses of microsatellite data showed little to no

genetic admixture between the species which shared haplotypes, as would be expected if hybridization were the cause (Kane *et al.* 2009). The one exception being an intermediate between *G. variifolia* subspecies, given their close ranges and similar flowering times, hybridization and/or introgression could potentially be occurring between *G. variifolia* subspecies. Nevertheless, the large portion of species sharing one haplotype could be explained by recent speciation, and substitution rates at these cpDNA loci being low in comparison to the the more variable microsatellite loci. This hypothesis is supported by the molecular dating analyses which revealed a late Pliocene/ early Pleistocene divergence for the shared haplotype.

While there was generally little phylogenetic resolution for or cpDNA data, three clades in the northern lineage were supported in these analyses. Subclade 1 included *G. delta, G. humifusa* and *G. thelemanniana* ssp. Cooljarloo. The placement of *G. delta* and *G. humifusa* in subclade 1 supports the close morphological similarity outlined in the taxonomic treatment of both Olde and Marriott (1994) and Makinson (2000). *Grevillea thelemanniana* ssp. Cooljarloo has only recently been recognized, and these analyses support its position in subclade 1 rather than having a close relationship with *G. thelemanniana* ssp. *thelemanniana* (see below). This subclade is in part recognised in the microsatellite data also, however there is incongruence regarding the placement of *G. humifusa*, which is not resolved as closely related to the other two species using microsatellite data and a relationship between *G. delta* and *G. thelemanniana* ssp. Cooljarloo with *G. pinaster* is inferred from the NJ tree. This subclade will be examined in more detail in chapter 3.

Grevillea evanescens, G. obtusifolia, G. thelemanniana ssp. *thelemanniana* and *G. stenomera* make up subclade 2. *Grevillea obtusifolia* was treated as a subspecies of *G. thelemanniana* by McGillivray and Makinson (1993). However, Olde and Marriott (1994) and Makinson (2000) recognised them as unique species with close morphological affiliations, which is supported herein by the haplotype sharing between these two species. The microsatellite analyses supported close affiliations between *G. evanescens, G. obtusifolia* and *G. thelemanniana* ssp. *thelemanniana.* The NJ tree illustrated a subgrouping of these species and STRUCTURE analyses placed them in the same genetic cluster at $K = 17$. *Grevillea evanescens,* and one population of *G. obtusifolia,* was recognised as unique at $K = 20$, illustrating that current or recent

gene flow between *G. evanescens* and *G. obtusifolia* is possible. Given their overlapping ranges, similar flowering times and propensity to hybridise in glasshouse experiments (Growns, G. pers.,comm.), hybridisation and/or introgression could potentially be occurring between these two species in nature. The position of one haplotype of *G. stenomera* in subclade 2 was unexpected, given the species is found in isolated populations more than 300 kilometres north of *G. evanescens, G. obtusifolia* and *G. thelemanniana* ssp. *thelemanniana*. Furthermore, *G. stenomera* was described by both McGillivray and Makinson (1993) and Makinson (2000) as being morphologically similar to *G. pinaster, w*hich is illustrated by a second haplotype from *G. stenomera* which branches off directly from the central haplotype, shared by *G. pinaster*. The position of one *G. stenomera* haplotype within subclade 2 is likely due to incomplete lineage sorting which is not uncommon among recent rapidly diversifying lineages (Knowles & Carstens 2007).

Subclade 3 includes only *G. preissii* ssp. *preissii* and one haplotype of *G. olivacea.* The cpDNA would suggest that *G. preissii* subspecies are not closely related. However the nDNA microsatellite NJ tree is not congruent with the cpDNA phylogeny. The NJ tree illustrates a close relationship between the *G. preissii* subspecies which is one of the few well supported results in the NJ topology. The placement of *G. olivacea* with *G. variifolia* subspecies in the NJ tree is confounding given the large geographic distance separating them (800 Km) and lack of any relationship in chloroplast data, but as for most relationships among species the relationship has weak support. However, this result is likely an indication that the resolving power of microsatellite markers are better suited for delimiting species.

Species delimitation within the Thelemanniana group

The lack of resolution regarding many species limits in the chloroplast data demonstrates how microsatellite markers are particularly relevant for testing species hypotheses in this group. Utilising multi-locus data like microsatellites to resolve issues of incomplete lineage sorting due to recent speciation has been encouraged in numerous recent studies (Petit & Excoffier 2009; Carstens & Dewey 2010; Hausdorf & Hennig 2010; Miralles & Vences 2013). The poor support of the resolution in the NJ topology, where species and subspecies were supported but northern and southern

lineages, as well as subclades 1-3, were not supported, is likely caused by the rapid evolutionary rate of microsatellites (Bacon *et al.* 2012). Bayesian analysis of the microsatellite data using STRUCTURE, also assigned most individuals into species and subspecies, which strongly supports the current taxonomy, providing a higher level of resolution than the cpDNA sequencing.

Genetic admixture between species was rarely found in the STRUCTURE analysis, suggesting that cpDNA results of widespread haplotype sharing was not due to hybridization and/or introgression. The microsatellite data successfully assigned individuals correctly to the previously described species and subspecies in most cases. Furthermore, the microsatellite data provided support for species which were not recognised as unique or monophyletic when using the chloroplast data alone. These included; *G. exposita, G. fililoba, G. variifolia* ssp. *variifolia* and *G. thelemanniana* ssp. *thelemanniana.* On the other hand, some species which were supported as monophyletic in the phylogenetic analyses were not clearly defined using the microsatellite data. For example *Grevillea acropogon* + *G. mccutcheonii* (*K* = 17) and *G. acropopon* + *G. ripicola* (*K* = 20) were grouped together in respective analyses. This may in part be due to the reduced number of loci available for these species. Not all microsatellite loci cross amplified to the southern clade species, likely because the microsatellites were developed from distantly related Northern clade specie *G. thelemanniana* ssp. *thelemanniana*. Bacon *et al.* (2012) highlighted that older lineages and missing data can be detrimental to species delimitation using multi-locus microsatellite data, which can explain the difficulty in resolving the southern lineage species. This is likely true for the clustering of *G. mccutcheonii* and *G. ripicola* which were separated by many mutational steps using the cpDNA. However*, G. acropogon* and *G. mccutcheonii* were only differentiated by a single mutational step in analyses of the chloroplast data and the STRUCTURE analysis result of microsatellite data may well reflect poor differentiation between these species. Makinson (2000) tentatively placed *G. acropogon* and *G. mccutcheonii* in the Thelemanniana group due to morphological similarities with *G. ripicola*, and recognised the need for more thorough research into these species relationships.

The different rates of evolution of chloroplast sequences and microsatellite data can lead to different levels of resolution in these analyses. Given the recent diversification

of this species complex the assignment of individuals to their respective species is probabilistically more likely owing to the inherent properties of nuclear microsatellites (Petit & Excoffier 2009). The integration of microsatellite and chloroplast data in this study revealed that the chloroplast sequences used here are generally more adept at resolving higher level relationships, whereas microsatellites are more powerful for recognising species boundaries, supporting similar species delimitation studies (Shaffer & Thomson 2007; Duminil *et al.* 2012).

Phrase named taxa

The position of *G. thelemanniana* ssp. Cooljarloo in subclade 1 and *G. thelemanniana* ssp. *thelemanniana* in subclade 2 illustrates a significant level of divergence between these two subspecies. The initial classification of *G. thelemanniana* ssp. Cooljarloo was based on morphological and ecological similarity to *G. thelemanniana* ssp. *thelemanniana* (Kieghery, pers comm., 2010). However, given that the subspecies are separated by approximately 170 kilometres, and the continued pattern arising between *G. thelemanniana* ssp. Cooljarloo as a sister to *G. delta*, the recognition of these taxa as different species is warranted. Furthermore, *G. delta* and *G. thelemanniana* ssp. Cooljarloo, while closely related, each have unique cpDNA haplotypes, show no evidence of admixture in the nuclear data, and are geographically separated by approximately 60 kilometres, thus supporting their unique status. These results all lead to supporting *G. thelemanniana* ssp. Cooljarloo as a unique species.

The phrase named species *G.* sp. Gillingara was also strongly supported as monophyletic and and its recognition is recommended. Although, it was part of a large polytomy in the chloroplast phylogeny it was separated from all other haplotypes by two mutational steps. Similarly, the neighbour joining tree of the microsatellite data showed moderately high divergence, and the STRUCTURE analyses detected a unique genetic cluster with no admixture from other species. Given the highly restricted range of its single known population of approximately 50 individuals, a detailed assessment of current threats needs to be conducted and a conservation plan proposed.

Phylogeography

The inferred crown age of 22.4 Mya, suggests a split between southern and northern lineages of the Thelemanniana group in the Early Miocene. During this time seasonality had become apparent in southwest of Western Australia and the first steps toward aridity began (Martin 2006). Specifically the rainforest that had dominated these regions disappeared by the Miocene, suggesting paleoclimate change from mesic to arid (Byrne *et al.* 2011), which could have resulted in regional divergences in taxa. Major species diversification for the lineages began during the late Miocene (11.6 – 5.3 Mya), extending into the Pliocene (5 – 1.6 Mya), with the many species appearing during the Pleistocene (1.6Mya – present). These estimated divergence times in the Thelemanniana group correspond to major global cooling commencing 14 Mya, which intensified in the late Pliocene (Hopper 1979; Hopper & Gioia 2004). The cooling temperatures and greater aridity during the Pliocene and Pleistocene has generally been thought as the main drivers of speciation in the SWAFR (Crisp *et al.* 2004). Significant radiations of flora from the SWAFR during the Pliocene and/or Pleistocene have been recorded for *Banksia* (Merwin *et al.* 2012) and *Hakea* (Mast *et al.* 2012), and inferred for species complexes of *Acacia* (Byrne *et al.* 2002) and *Stylidium* (Coates *et al.* 2003).

Hopper (1979) and Hopper and Gioia (2004) suggested that the erosion and climatic stresses of the late Tertiary and Quaternary was greater in the TRP than the HRP, leading to more recently evolved taxa in the TRP compared with the HRP. In addition this may have resulted in the persistence of relictual species due to the relative climatic stability of the HRP compared to the TRP. The older lineages within the Thelemanniana group however did not indicate a relationship with geographic distribution in the SWAFR. Indeed, *G. ripicola* and *G. mccutcheonii* of the southern lineage and *G. costata* and *G. commutata* of the northern lineage both diverged during the late Miocene/early Pliocene (Figure 2.1). This result corroborates the results of Cardillo and Pratt (2013), which also did not find this pattern of a bias towards older relict species in the HRP in their molecular study on *Banksia*. On the other hand, Cardillo and Pratt (2013) did find higher rates of speciation in the TRP. These results are consistent with their finding as there was evidence for recent speciation of the

numerous species in Thelemanniana group, which for the most part, were located in the TRP (Figure 2.1).

Lastly, Hopper (1979) and Hopper and Gioia (2004) suggested that the climatic stresses of the late tertiary and Quaternary also resulted in explosive recent radiation of many SWAFR genera. This hypothesis was not supported by Cardillo and Pratt (2013), in which they showed that diversity within *Banksia* fit a pattern of steady accumulation rather than rapid recent speciation. The molecular dating of the Thelemanniana group illustrates a similar pattern of accumulation, with several lineages diverging in the late Miocene through the Pliocene. The results from this study in combination with those from Cardillo and Pratt (2013) show do not support these hypotheses of Hopper (1979) and Hopper and Gioia (2004) regarding diversification of the SWAFR flora. Cardillo and Pratt (2013) and Merwin *et al* (2012) highlighted that mating systems, life history and fire frequencies also have an important effect on extinction and speciation rates, which can explain different responses to climatic fluctuations of species in the SWAFR. These processes are likely to have had a significant impact on *Grevillea* species also, and can explain why hypotheses based on the impact of geohistory and climatic fluctuations alone are not supported in this study.

The calibration points used to estimate molecular time in this study are not directly from fossil records, but are based on a larger study which used fossil calibration (Mast, unpublished data). The use of secondary calibration points has been discouraged due to a potentially large degree of error (Renner 2005). However, direct fossil evidence is often not possible, as the fossil record is largely incomplete and biased (Linder *et al.* 2005). Therefore, several studies argue for the use of secondary calibration points as suitable for time estimation (Hedges & Kumar 2004; Linder *et al.* 2005), and even rough estimates of evolutionary time can greatly expand the information conveyed on a tree (Avise 2000). Thus, secondary calibration points are used to estimate divergence times in this study.

Conclusions

I have shown that the Thelemanniana group is not monophyletic, and current classifications based on morphology are incongruent with the molecular data.

Importantly, the recognition of both *G. thelemanniana* ssp. Cooljarloo and *G.* sp. Gillingara as unique lineages using both cpDNA sequence and nDNA microsatellite data warrants their recognition as unique species. Incomplete lineage is hypothesised as the cause of widespread haplotype sharing because of recent speciation within the Thelemanniana group, as microsatellite data do not support hybridization. This relatively recent radiation also explains the difficulty in obtaining reciprocal monophyly in the chloroplast data. Alternatively, the vast sharing of one haplotype among species makes the molecular dating somewehat unreliable since this is likely an ancestral haplotype that contrasts with the theory of recent speciation of the northern clade. The species by Bayesian structure analyses identified most species recognised in Makinson's (2000) treatment of *Grevillea.* These results highlight the importance of using molecular markers representing different evolutionary rates when resolving relationships among groups with both deep divergence and recent diversification. Lastly, the molecular dating of the Thelemanniana in part supports Hopper's theory that climatic oscillations during the Quaternary have lead to recent speciation in the transitional rainfall provenance in the SWAFR. However, similar to Cardillo and Pratt's (2013) study of *Banksia,* there appears to be steady accumulation of species, in that new species are branching of at multiple time points rather than singular rapid radiations, which does not support the view that the SWAFR flora is the result of rapid radiations.

Chapter 3 - Evolutionary patterns within and among species of a *Grevillea* Thelemanniana group subclade.

Introduction

Species are one of the fundamental units of biology (Coates 2000; de Queiroz 2007; Fujita *et al.* 2012). The identification of boundaries among closely related species and the inference of the number of species are essential for effective conservation strategies (Elliott & Byrne 2004; Duminil & Di Michele 2009; Petit & Excoffier 2009). Most commonly used approaches to species delimitation have been through morphologically defining characters specific to a species. Morphologically defined species in regions with low biodiversity and well defined geographic or ecological boundaries are often congruent with biological or evolutionary entities (Broadhurst *et al.* 2004). However, in regions like the southwest of Western Australia, where a complex evolutionary history has lead to high species diversity and endemism, species delimitation from phylogenies that incorporate an evolutionary perspective is often necessary (Byrne 2007). Phylogenies of flora in the south west of Western Australia have revealed a broad spectrum of intraspecific and interspecific evolutionary patterns (Coates 2000; Byrne 2007). For example, numerous molecular studies on flora in the region have discovered highly disjunct lineages within the same species (Coates & Hamley 1999; Byrne *et al.* 2003; Byrne & Hines 2004), and extensive speciation and genetic divergence between closely related species and subspecies (Broadhurst *et al.* 2001; Coates *et al.* 2003; Elliott & Byrne 2004; Butcher *et al.* 2007). The evolutionary patterns resolved in these studies have resulted in a strong congruence with geohistorical events and climatic oscillations, illustrating how the recognition of taxonomic entities through phylogenetics can reflect evolutionary histories. Byrne (2007) highlighted the importance of comparative studies across numerous species in contributing to the understanding of how historical factors have influenced flora in the region. However, for many species in the southwest of Western Australia knowledge on taxonomy, population genetics, ecology and evolution is lacking, and much work is needed in describing species and defining conservation units (Elliott *et al.* 2002; Coates *et al.* 2007; Byrne *et al.* 2011).

Many phylogenetic studies utilise the chloroplast introns to resolve species relationships in plants because they have lower mutation rates due to the genome elvoving at a slower rate than the nuclear genome. Intron sequences are valuable sources of data for taxonomic studies above species and but below family level (Barker *et al.* 2000). Although intron sequences are highly variable compared to other sequence data, microsatellites are more variable due to their high polymorphism and distribution of multiple loci. The combination of intron sequences and microsatellites can therefore be used to resolve species complexes, which often have cryptic genetic relations within and between species. relationships (REF). In the flora of the southwest of Western Australia the chloroplast has revealed relatively high levels of divergence in studies within and among species (Byrne *et al.* 1999; Byrne *et al.* 2001; Tapper *et al.* 2014). However, as seen in Chapter 2, the levels ofchloroplast variation was not comparable to previous studies on Western Australian species, and can result in poorly supported phylogenies in closely related species complexes. In this case nuclear loci such as microsatellites combined with chloroplast sequence data can be useful in identifying genetic structure and delimiting species (Coates 2000). The use of microsatellites for species delimitation are becoming more common in cases where there is low genetic divergence due to recent speciation (Petit & Excoffier 2009; Prunier & Holsinger 2010). Studies integrating different sources of data have improved species delimitation (Carstens & Knowles 2007; Oliver *et al.* 2009; Pessoa *et al.* 2012). Additionally, many recent taxonomic studies have begun to rely on population genetics approaches, typically with microsatellites, to resolve species relationships and discovering potential cryptic species (Millar *et al.* 2011; Jones *et al.* 2012; Kim *et al.* 2012). Hey and Pinho (2012) argued that using information on both population level divergence and species level divergence means that scientists can better compare the level of variation between two populations compared to two species. Species are expected to show greater evolutionary independence from one another than populations within species. Therefore, in theory, if this difference could be quatified researchers could use it to engage species level diagnosis.

Phylogenetic analyses of species and populations enable taxonomy to reflect evolutionary processes (Broadhurst *et al.* 2004; Elliott & Byrne 2004), and using different types of molecular markers in phylogenetic studies, and their comparison,

can provide valuable insights into the evolutionary history of closely related species (Duminil & Di Michele 2009; Barrett & Freudenstein 2011). Delimiting species that have low levels of genetic divergence, either because speciation is recent or because the species continue to exchange genes is a major challenge in taxonomy (Petit & Excoffier 2009). Phylogenies of derived species will often display incongruence between gene trees due to incomplete sorting of ancestral polymorphisms causing lack of reciprocal monophyly at some loci (Carstens & Knowles 2007), or due to hybridization and/or introgression (Arnold 1992; Baack & Rieseberg 2007). However, a multi-locus approach increases the likelihood of resolving incongruences through assessment of different evolutionary lineages (Duminil & Di Michele 2009). Microsatellites can be informative regarding recent gene exchange and can be used to compare patterns of divergence within and between species. For example, Rossetto *et al.* (2012) used chloroplast and nuclear microsatellites to resolve evolutionary patterns in *Teleopa*, which showed recent allopatric speciation through localised hybridization. Using multi-locus molecular data that represents lineages evolving under different rates and at different levels has led to an integrative perspective in taxonomic practice which should increase the power comparative studies (Duminil *et al.* 2012; Miralles & Vences 2013).

Low genetic diversity and recent speciation was revealed in the Western Australia *Grevillea* Thelemanniana group (Chapter 2). Several subclades were recognised in the phylogeny based on chloroplast sequence data, albeit with low support. One of these was subclade 1, which includes *G. delta, G. humifusa* and *G. thelemanniana* ssp. Cooljarloo. A NJ tree based on the nDNA microsatellite data, illustrated a different relationship between *G. delta, G. pinaster* and *G. thelemanniana* ssp. Cooljarloo. *Grevillea delta, G. humifusa and G. pinaster* have been variously treated as species and subspecies in past classifications of the larger Thelemanniana group, but most recently as species (Makinson 2000). The conservation status of *G. delta* and *G. humifusa* is priority 2 and threatened, respectively, while *G. pinaster,* is not listed (Smith 2013). *Grevillea humifusa* is currently recognised from a single roadside population and attempts are being made to translocate new populations (DPaW, pers., comm.). *Grevillea thelemanniana* ssp. Cooljarloo is recognized as a phrase-named subspecies and is listed as Priority 1 (Smith 2013) and is facing immediate threats from a nearby

minerals mining operation. *Grevillea pinaster* is widespread, whereas *G. delta, G. thelemanniana* ssp. Cooljarloo and *G. humifusa* all have restricted ranges within the Lesueur region, which has been recognised for extremely high levels of species diversity (Griffin *et al.* 1983; Hopper & Gioia 2004).

To examine species limits proposed on the basis of morphology, I used both cpDNA sequence data and nDNA microsatellite markers among taxa of subclade 1 (see chapter 2) of the *Grevillea* Thelemanniana group to 1) compare population genetic patterns among populations and species to determine if a significantly different level of divergence exists within species and between species and; 2) assess whether population genetic based measures of evolutionary independence can be used to delimit species and; 3) determine evolutionary patterns between taxa and to provide a better understanding of the complex evolutionary patterns in the flora of the southwest of Western Australia; and 4) address the conservation implications of these findings.

Methods and Materials

Sampling and species distributions

A total of 388 accessions were sampled from multiple populations for each of the targeted taxa: *G. delta*, *G. pinaster* and *G. thelemanniana* ssp. Cooljarloo, and one population of *G. humifusa* (Table 3.1). *Grevillea pinaster* ranges from Enneaba north to Murchison River. Populations are widespread in the northern range from Geraldton to Kalbarri, and inland to Mullewa (Fig 3.1), however southern populations around Enneaba are sparse. *Grevillea delta* and *G. thelemanniana* ssp. Cooljarloo are local endemics distributed along creek beds in their respective ranges, and *G. humifusa* in known from a single roadside population. Collections for *G. pinaster, G. delta* and *G. humifusa* were based on herbarium specimens GPS records (Herbarium 1998). Populations of *G. thelemanniana* ssp. Cooljarloo were identified from herbarium specimens and records from the environmental team from Tronox Ltd. Apart from distance and fragmented landscapes, geographic barriers such as Mt. Lesueur may be acting as a barrier separating *G. delta* from *G. humifusa* and *G. thelemanniana* ssp. Cooljarloo populations.

Species biology

Morphologically *G. thelemanniana* ssp. Cooljarloo differs from *G. delta* and *G. humifusa* by the revolute leaf margins which enclose the lower surface of the leaf blade, forming two grooves with the midvein. *Grevillea humifusa* is a prostrate shrub, shorter and less robust than *G. delta* with longer pistils (22-24mm as opposed to 10-12mm). *Grevillea thelemanniana* ssp. Cooljarloo has only primary leaf division (G. Keighery, 2013, *pers comm*.), whereas *G. delta* and *G. humifusa* have secondary leaf division on some or all leaves (Makinson 2000). *Grevillea pinaster* differs primarily by having leaves entire and is usually a robust shrub, growing up to 3 meters, but also has forms which are more prostrate like *G. humifusa,* growing along coastal granite cliffs. Makinson (2000) distinguished three forms of *G. pinaster*. Firstly, the 'typical form' which has mid-green leaves and upright habitat occurring from Geraldton East to beyond the Murchison River (RR, GLN, SPR, HUTT and GER, Table 3.1). Secondly, the 'secund form' which occurs on outcrops and in coastal gorges in Kalbarri (GRG and KAL,

Table 3.1). It has grey-green leaves often crowded to one side of branchlets and a lower more spreading habitat. This form might also occur near Enneaba (Makinson 2000). A third 'pinnate-leaved form' has been recognised occurring sporadically around Enneaba and Mullewa, however it's status is uncertain and no official records are recorded at the Western Australian Herbarium, DPaW. Therefore, this form was not sampled in this study. Lastly, a robust erect tree, 2-3 meters was discovered south of Kalbarri (D. Growns 2013, *pers comm*.), and included in this study (VIV, Table 3.1). Examples of forms are illustrated in Figure 3.2.

The species have several similar life history traits. Flowering times for all four taxa are long and overlap, beginning in May/June through to October/November (Makinson 2000). They exhibit bright red flowers, indicating that they are pollinated by birds (most likely honeyeaters, Meliphagidae) and insects (England et al 2003). *Grevillea pinaster* and *G. humifusa* are thought to regenerate from lignotuber and seeds, whereas regeneration mode for *G. delta* and *G. thelemanniana* ssp. Cooljarloo are unknown.

Table 3.1 Species name, location and geographical coordinates (UTM), population codes, sample sizes for microsatellite (simple sequence repeats or SSR's) and chloroplast (cpDNA) datasets and nuclear microsatellite diversity parameters for each population sampled from three *Grevillea* species and one subspecies. Mean values for: number of alleles (Na), observed heterozygosity (HO), expected heterozygosity (HE) and inbreeding coefficient (F_{IS}). P_A = number of private alleles across all loci and A_R = allelic richness standardized across minimum sample size.

							Genetic diversity					
Species	Location	GPS coordinate		Code	SSR	cpDNA	Na	Ho	He	A_R	P_A	F_{IS}
G. pinaster	Rob Rd	253627	6875118	RR	22	6	6.67	0.61	0.69	5.63		0.13
	Glenn Point	245919	6878041	GLN	24	7	7.33	0.68	0.70	6.21	1	0.02
	Spring Rd	242070	6880722	SPR	20	7	6.83	0.63	0.70	6.09		0.11
	Pots Alley Gorge	217488	6925764	GRG	23	6	6.33	0.70	0.70	5.40	3	0.01
	Hutt Provinance	250367	6892006	HUTT	12	8	4.67	0.55	0.60	4.67		0.20
	Viv Property	253657	6878849	VIV	8	7	5.33	0.58	0.65			0.08
	Eneba Rd	315729	6820142	GER	5	4	3.33	0.45	0.48			0.02
	Outcrop	222284	6930732	KAL	4	3	3.50	0.58	0.55		1	-0.08
G. delta	Coorow Gully	324325	6672807	GUL	24	12	6.17	0.42	0.58	3.38	1	0.32
	Cockshell Gully	321764	6664316	CG	35	9	3.83	0.39	0.59	3.67	1	0.30
	Mt Leu Base	325651	6663733	GS	4	3	2.50	0.39	0.41		1	0.04
	East C. Gully	323561	6664959	GE	0	2						
G. humifusa	Cantabilling Rd	329615	6647234	HUM	32	12	4.00	0.41	0.47	2.50		0.18
G. thelemanniana ssp. Cooljarloo	North Creekline	339361	6617412	NC	17	7	5.67	0.56	0.67	5.22	4	0.20
	Wongonderrah Rd	338417	6618446	WON	20	10	5.83	0.58	0.67	5.20		0.14
	North floodplain	334933	6615989	F34	18	6	5.50	0.66	0.71	5.18	1	0.08
	West creek line	331808	6611324	WC	28	6	6.50	0.69	0.72	5.50		0.07
	West floodplain	333550	6610495	F10	12	2	5.50	0.67	0.70	5.50	1	0.07
	South creek line	338718	6606276	SC1	21	5	4.50	0.61	0.64	4.08		0.07
	South creek line	332275	6610516	SC2	15	5	4.33	0.62	0.63	4.17		0.02
	South floodplain	338074	6604528	F15	20	3	6.00	0.71	0.71	5.42		0.04

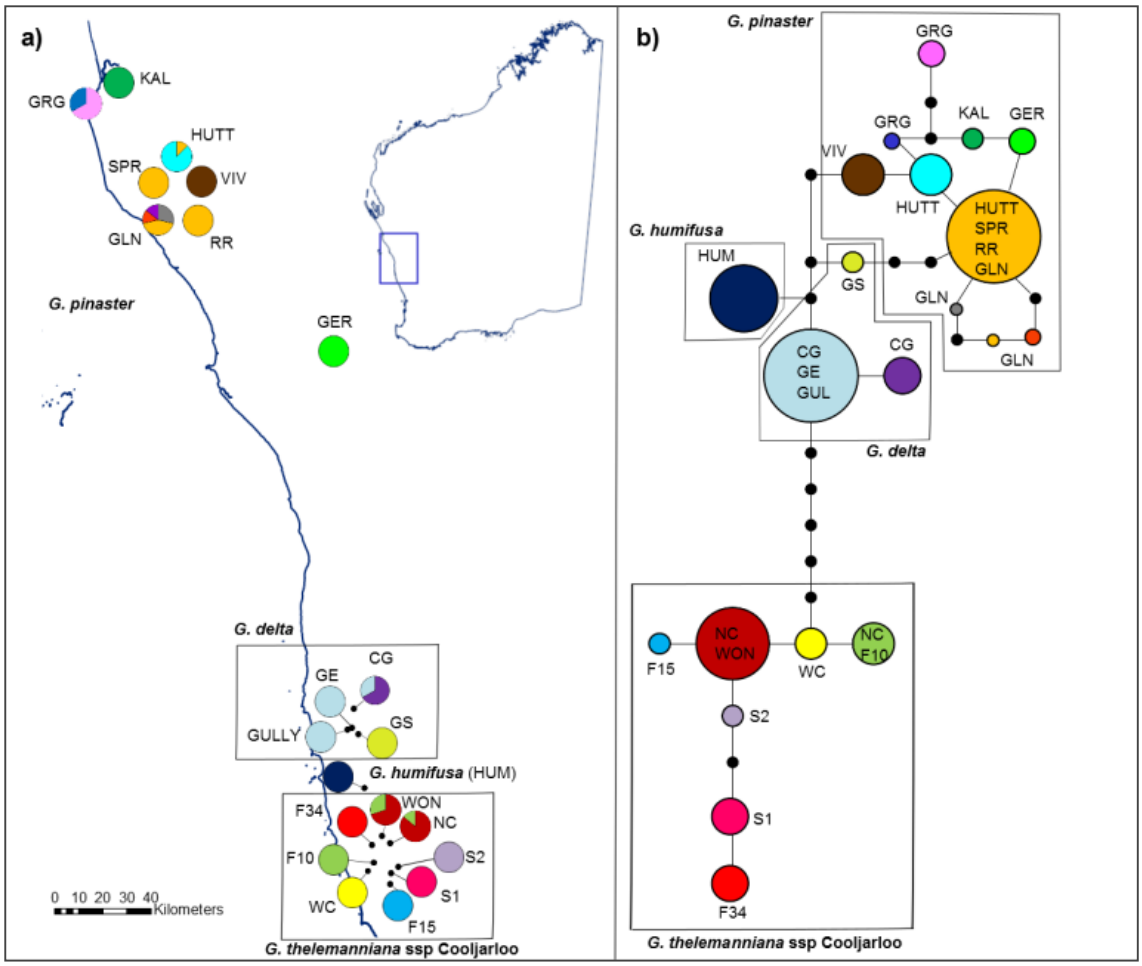

Fig.3.1 a) Location in Western Australia of sampled populations for *G. pinaster, G. delta, G. humifusa* and *G. thelemanniana* ssp. Cooljarloo and the distribution of chloroplast DNA haplotypes found in populations. The size of proportions of pie charts represents the number of individuals of that haplotype. Colour coding of haplotypes correspond to those in the parsimony haplotype network and population names are abbreviated by the population code (Table 3.1). b) Statistical parsimony network of 130 concatenated cpDNA sequences from populations of *Grevillea delta, G. humifusa, G. pinaster* and *G. thelemanniana* ssp. Cooljarloo. Circle sizes correspond to the number of individuals possessing the haplotype, and colours correspond to unique haplotype. Inferred missing or extinct haplotypes (not present in the data set) are depicted as black dots.

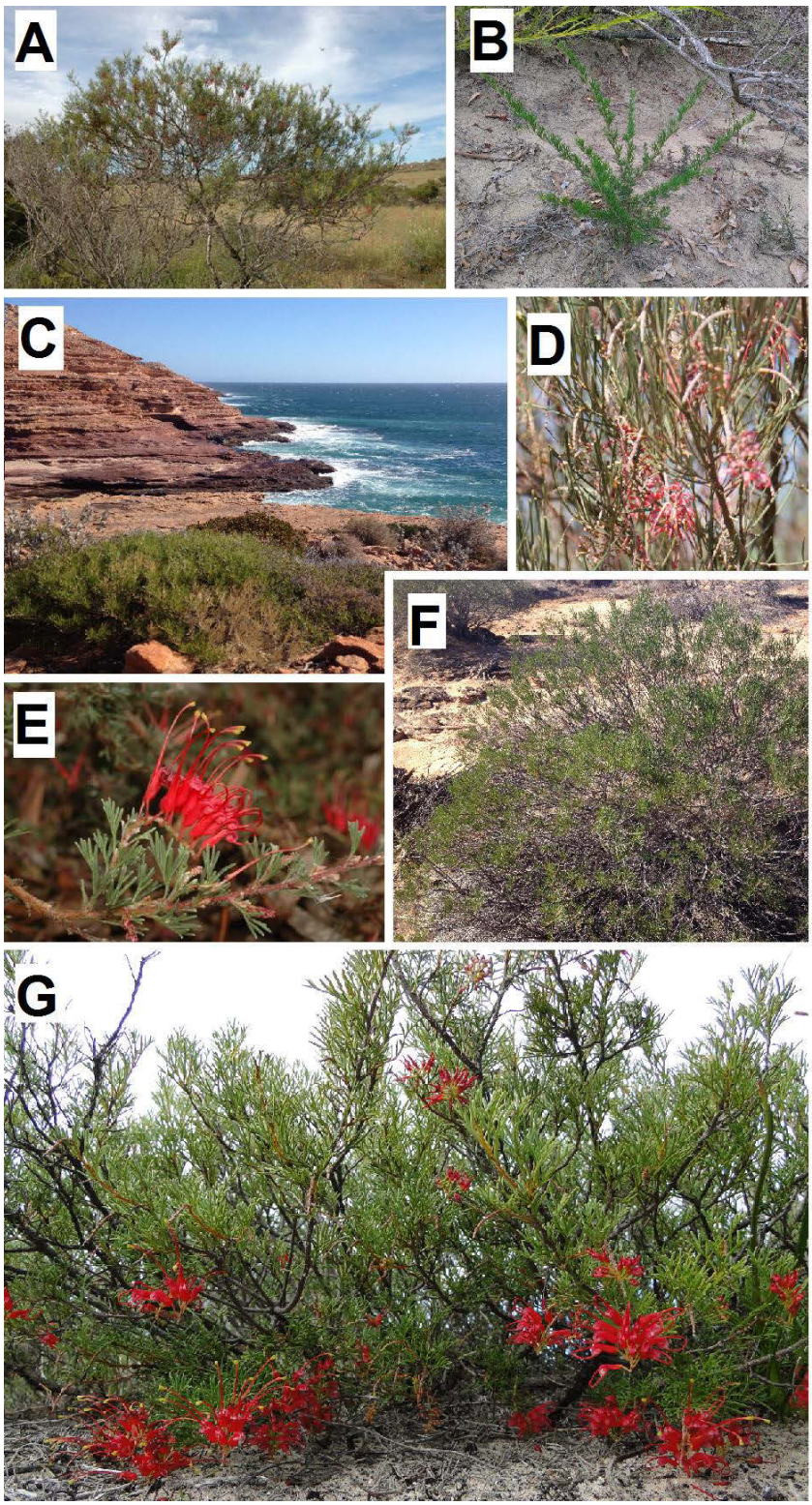

Fig.3.2 *Grevillea* sampled in this study (A) erect tree of *G. pinaster* (VIV); (B) *G. delta* (CG); (C) 'secund form' of *G. pinaster* along Kalbarri coast (GORGE); (D) Entire leaves of *G. pinaster* (GER); (E) *G. humifusa* (F) 'typical form' of *G. pinaster* (RR) and; (G) *G. thelemanniana* ssp. Cooljarloo. Images A-D, F &G by Tanya Hevroy and E by Fagg, M Australian Plant Index, Atlas of Living Australia.

Amplification of molecular markers

Genomic DNA was extracted from 20-30mg freeze dried leaf material using the modified CTAB extraction method (Doyle 1990). Quantity and quality of DNA was assessed by visualization on a 1.5% agarose gel and a Nanodrop (Thermo Fisher, Wilmington, DE). Two chloroplast regions were amplified using the Polymerase Chain Reaction (PCR) for all accessions. The trnL-trnF region was amplified with primers 'Tab-C' and 'Tab-F' of Taberlet *et al.* (1991) and the rpS4-trnL region was amplified with primers 'Tab-B' (Taberlet *et al.* 1991) and 'rpS4R2' (Shaw *et al.* 2005). All PCR reactions contained 2-10µg DNA, 3µl of 5xPCR buffer (Invitrogen, including dNTPs), 2mM $MgCl_2$, 0.1µM of each primer and 0.1µl Taq DNA polymerase (Invitrogen) and brought to a total volume of 15µl with sterile water. PCR reactions were performed in Eppendorf thermocyclers under the following conditions: One cycle at 97°C for 2 min followed by 30 cycles at 94°C for 1 min, 52°C for 2min, 72°C for 2min and a final extension at 72°C for 7min. PCR product was visualized for amplification success on a 1% agarose gel. Successful amplifications were sent to Macrogen Inc., South Korea, for purification and sequencing. Both strands were sequenced for overlap and to assure accuracy in base calling. The DNA sequences were edited for miscalls in Sequencher 3.1.1 (Gene Codes Corp.) and manually aligned by eye and concatenated in Mesquite (Maddison & Maddison 2004). Insertion and deletion events (INDELS) were considered phylogenetically informative and were coded as single base transversions characters at the end of the data matrix. In total 6 indels were coded for the concatenated data set.

Twelve microsatellites from *Grevillea thelemanniana* ssp. *thelemanniana* (Hevroy *et al.* 2013) and four from *G. macleayana* (England *et al.* 1999) were cross amplified and tested for polymorphism. Of these, five from Hevroy *et al.* (2013) and one from England *et al.* (1999) showed illustrated polymorphism across all four species. The microsatellites were amplified using PCR containing 2-10 µg DNA, 2µl of 5xPCR buffer (Invitrogen, including dNTPs), 2mM $MgCl_2$, 0.1 µM of the fluorescently labelled forward primer (6-FAM, NED &VIC) and unlabeled reverse primer, 0.1µl Taq DNA polymerase (Invitrogen) and brought to a total volume of 10 µl with sterile water. PCR reactions were performed in Eppendorf thermocyclers under the following conditions: One cycle at 94°C for 3 min followed by 30 cycles at 94°C for 40 secs, 54°C for 40 secs, 72°C for 40 secs and a final extension at 70°C for 15 min. PCR products were separated by capillary electrophoresis using a CEQ8800 (Beckman-Coulter). Micro-Checker was

used to identify any outliers and determine the presence of null alleles in the dataset (Van Oosterhout *et al.* 2004).

Chloroplast data analyses

Identification of haplotypes, nucleotide and haplotype diversity was calculated in DNASP 5.10.01 (Librado & Rozas 2009). Analyses hereafter were performed with one representative of each haplotype. Chloroplast genetic variance was also partitioned into three hierarchical levels; among species, among populations within species and within populations using analysis of molecular variance (AMOVA) in the program Arlequin v. 3.11 (Excoffier & Lischer 2010), with 1000 permutations based on distance matrices.

A maximum likelihood tree (ML) was constructed in MEGA version 6 (Tamura *et al.* 2013), using uncorrected genetic distances as the model of sequence evolution because of the low average sequence divergence among samples, as recommended by the authors (Tamura *et al.* 2013). For both analyses, 1000 bootstrap replicates were used to assess nodal support.

The hierarchical Bayesian model implemented in *BEAST v. 1.7.5 (Drummond *et al.* 2012) was used to estimate species trees from a partitioned data file with both trnL-trnF and rpS4-trnL sequences. I used the regular BEAST analyses model since I had sinle locus and concatened loci. BEAST incorporates uncertainty associated with gene trees and nucleotide substitution model parameters (Drummond *et al.* 2012). jModelTest 2.0 (Posada 2008) was used to determine the substitution model for each cpDNA region (HKY+G+I for trnL-trnF and GTR+G for rpS4-trnL). The analysis was repeated twice with MCMC run for a total of 50 million generations (sampling 1000 steps and excluding the first 10% as burn-in. Convergence was assessed by examining the likelihood plots through time using TRACER v. 1.4.1 (Rambaut & Drummond 2003) and ensuring that all ESS (effective sample size) were above 200, as recommended by the authors (Drummond & Rambaut 2007; Drummond *et al.* 2012).

I constructed a haplotype network using the statistical parsimony method as implemented in the program TCS version 1.21 (Clement *et al.* 2000), using a 95% connection limit. The program uses maximum parsimony to join haplotypes into an

unrooted network until the chosen limit of confidence is exceeded. The coded indels were included in the network analyses.

Microsatellite data analyses

A variety of genetic diversity statistics were used to assess and compare genetic variation within populations, averaged for each locus: number of alleles per locus (N_a), observed and expected heterozygosity (H_o & H_e), fixation index (F_{IS}), and private alleles (P_R) were calculated using GenAlEx 6.05 (Peakall & Smouse 2012). Allelic richness was calculated using HP-Rare 1.0 (Kalinowski 2005) and standardized for minimal population size, populations with less than 10 individuals were removed to avoid bias caused by uneven sampling.

To determine the genetic variance between and within populations of the four taxa, pairwise F_{ST} was calculated between all sampled populations. An analysis of molecular variance (AMOVA) was used to partition the total genetic variance, from pairwise F_{ST}, into three hierarchical levels; between species, among populations within species and within populations. Both F_{ST} and AMOVA were also calculated using the program ARLEQUIN, and significance was assessed at 95% and 99% confidence intervals by using 10000 bootstrap permutations. Mantel tests were performed to test the influence of geographic distances on observed patterns of genetic differentiation (isolation by distance; IBD). A plot of matrices of F_{ST} (F_{ST} / (1- F_{ST})) against the log of geographic distances was performed using 10,000 random permutations across all populations and taxa, as well as separate tests for populations of *G. thelemanniana* ssp. Cooljarloo and *G. pinaster*. Due to low population sampling in *G. delta* and only one population of *G. humifusa* separate tests could not be performed for these taxa. A principal coordinate analysis (PCA) based on Nei's genetic distance was used to produce a comparative graphical representation of genetic distances among all populations. Mantel tests and PCA were calculated using GenAlEx 6.5.

The Bayesian clustering program STRUCTURE 2.1 (Pritchard *et al.* 2000) was used to determine the affinity of individuals to genetically homogenous clusters (*K*) using both an admixture and no admixture model with correlated allele frequencies. Analysis parameters were set as 1,000,000 Markov chain Monte Carlo (MCMC) with burn-in length of 100,000 and repetitions with 10 simulations for each proposed *K* (*K* = 1-17). A

value of *K* = 17 was chosen as the upper limit of the number of possible clusters based on the number of populations sampled across all species and subspecies. Results were uploaded to STRUCTURE HARVESTER (Earl & von Holdt 2012) to identify the most probable value of K based on a plot of likelihood values against delta K (Evanno *et al.* 2005). The resulting matrices were run through CLUMPP 1.1.2 (Jakobsson & Rosenberg 2007) to determine similarity coefficient (*h*') and clustering patterns were visualized in DISTRUCT 1.1 (Rosenberg 2004).

Results

Chloroplast sequences

Allignments were 953 base pairs for the *trnL-trnF* region and 1117 base pairs for the *rpS4-trnL* region. The concatenated sequences had 19 parsimony informative sites (6 indels), haplotype (gene) diversity of 0.97 and nucleotide diversity of 0.002.

There were 21 unique haplotypes in the chloroplast sequence dataset (Fig 3.1). No haplotypes were shared among species, however there was haplotype sharing within species among populations. *Grevillea pinaster* had 10 unique haplotypes in total, of which nine were population specific. The one shared haplotype was among populations; GLN, SPR, RR and HUTT. Four haplotypes were found in GLN, two in HUTT and GRG and one in GER, KAL and VIV. *Grevillea pinaster* haplotypes were closely related to each other, most are connected by no more than two mutational steps. There are three loops in the network, indicating homoplasy. Two loops are evident in the *G. pinaster* network and a third loop is between *G. pinaster* and *G. delta.* The first two loops could not be resolved unambiguously, however the loop concerning the connection of *G. delta* (GS) to *G. pinaster was* caused by homoplasy involving one of the indels. *Grevillea delta* has 3 unique haplotypes, of which one is shared between three populations (CG, GE and GUI). CG and GS have unique haplotypes, however GS unique haplotype is based on one indel (a single base pair), and when removed it expresses the same haplotype as the central *G. pinaster* haplotype. There are five mutational steps, or missing intermediate haplotypes, separating *G. thelemanniana* ssp. Cooljarloo from *G. delta*. *Grevillea thelemanniana* ssp. Cooljarloo has seven unique

haplotypes. Most of the unique haplotypes were between different sampling localities; south (F15) and north floodplains (F34), west creek line (WC) and both populations from south creek lines (S1 & S2). Shared haplotypes were between northern populations (WON & NC) and between northern creek line and western floodplain (NC & F10).

The AMOVA analyses supported the existence of strong and significant segregation between the four named taxa (73.2%, $P < 0.0001$), with 23.1% ($P < 0.0001$) among populations within species and 3.7% ($P < 0.0001$) within populations (Table 3.2).

The Bayesian 50% majority consensus tree had little resolution for the backbone but support for most names taxa (Fig.3.3). *Grevillea humifusa* was supported as monophyletic (1.0 PP, 72 BS) as was *G. pinaster* (0.98 PP, 56 BS), and *G. thelemanniana* ssp. Cooljarloo (0.99 PP, 80 BS). There is weak support for most *G. delta* accessions and *G. thelemanniana* ssp. Cooljarloo as sister species (0.79 PP, 62 BS). However, *G. delta* is not monophyletic as all accession from one population (GS) are part of a clade including *G. pinaster* and *G. humifusa* (1.0PP). Some *G. thelemanniana* ssp. Cooljarloo populations were supported. In *G. pinaster*, only the VIV population (0.98 PP, 56 BS) was supported as monophyletic.

Microsatellite markers

All loci were polymorphic in all populations, except for THP4 in GER and KAL, THP13 in HUM and GM13 in GS. There was no evidence for large allele drop out, stuttering or null alleles for any of the 6 loci. Tests for Hardy-Weinberg equilibrium (HWE) following sequential Bonferroni correction revealed significant ($P<0.05$) departure from HWE in only 16 out of 120 locus-population combinations. Mean expected heterozygosity was 0.47 for the one *G. humifusa* population, and ranged from 0.41 – 0.53 in *G. delta* populations, 0.48 – 0.70 in *G. pinaster* populations and 0.63 – 0.72 in *G. thelemanniana* ssp. Cooljarloo populations (Table 3.1). The mean number of alleles ranged from 4.0 in *G. humifusa,* 4.17 in *G. delta,* 5.5 in *G. pinaster* and 5.48 in *G. thelemanniana* ssp. Cooljarloo. *Grevillea pinaster* (5), *G. delta* (4) and *G. thelemanniana* ssp. Cooljarloo (6) had private alleles, whereas *G. humifusa* did not. Number of alleles, allelic richness and heterozygosity was significantly lower in *G. humifusa* compared with *G. pinaster, G. delta* and *G. thelemanniana* ssp. Cooljarloo populations ($P < 0.01$, two-tailed *t*-test).

There was significant genetic structuring between all pair wise population F_{ST} comparisons at P = 0.05, however comparisons between some *G. pinaster* populations, and between KAL and GS, were not significant at P = 0.01 (Table 3.3). The global mean F_{ST} across all comparisons was 0.124, with a large range from 0.038 to 0.375. The mean pairwise population F_{ST} comparisons between species were high (F_{ST} = 0.216). In comparisons, F_{ST} within species was significantly lower than between species; *G. delta* (0.133), *G. thelemanniana* ssp. Cooljarloo (0.122) and *G. pinaster* (0.076). The PCA showed polarity of populations from the same species with no overlap (Fig.3.4). The relative position of the *G. pinaster* outlier populations of GER and KAL most likely reflects low sample numbers.

The AMOVA partitioned 77% of the total genetic variation within populations, with a significant proportion (P <0.0001, Table 3.2) of the variation partitioned among populations (8.3%) and among species (14.5%). There was significant isolation by distance among all taxa (r^2=0.18, P = 0.001, Fig.3.5). There was no evidence for isolation by distance among populations of *G. thelemanniana* ssp. Cooljarloo (r^2=0.08, P = 0.072, Fig 3.6a), whereas a significant pattern of IBD was evident in *G. pinaster* populations (r^2=0.28, P = 0.004, Fig 3.6b).

The ΔK method of Evanno *et al* (2004) from Bayesian assignment results obtained from STRUCTURE indicated K = 2 as the best estimate of genetic clusters under an admixture model and K = 4 under a noadmixture model (Fig.3.7). At K = 2 *G. pinaster, G. delta* and *G. humifusa* made up one genetic cluster and *G. thelemanniana* ssp Cooljarloo another. Assignment probabilities (*qi*) were greater than 0.95 for almost all the individuals. At K = 4 taxa were individuals were assigned correctly to their defined taxa, with assignment probabilities (*qi*) higher than 0.9 for almost all the individuals, indicating clear genetic structuring and low levels of gene flow between species. Individuals of *G. delta* population GS showed evidence of admixture (*qi* = 0.72) with *G. pinaster*. *Grevillea pinaster* population GLN illustrated some admixture from *G. delta* (*qi* > 0.8). Additionally five individuals in *G. pinaster* populations RR, GLN and SPR showed signs of admixture with *G. humifusa* and *G. delta* (*qi* ≈ 0.5). *Grevillea delta* and *G. thelemanniana* ssp. Cooljarloo illustrated one individual each with a genotype from *G. pinaster (qi* > 0.9). Attempts to recognize if there was further clustering within *G.*

pinaster did not support the existence of within species structure. At *K* = 5 and *K* = 6 there was no evidence for substantial genetic structure within *G. pinaster*. At *K* = 6 a split emerged in the *Grevillea thelemanniana* ssp. Cooljarloo populations, however most of this structure illustrates a high level of admixture which indicates no clear structure in these populations.

Table 3.2 Hierarchical analyses of molecular variance (AMOVA) among and within taxa and populations of three *Grevillea* species and one subspecies based on both cpDNA sequence and nDNA microsatellite data.

	cpDNA			Microsatellites		
	df	% Variation	*P*	*df*	% Variation	*P*
Among species	3	73.2	<0.0001	3	14.5	<0.0001
Among populations within species	17	23.1	<0.0001	16	8.3	<0.0001
Within populations	129	3.7	<0.0001	367	77.2	<0.0001

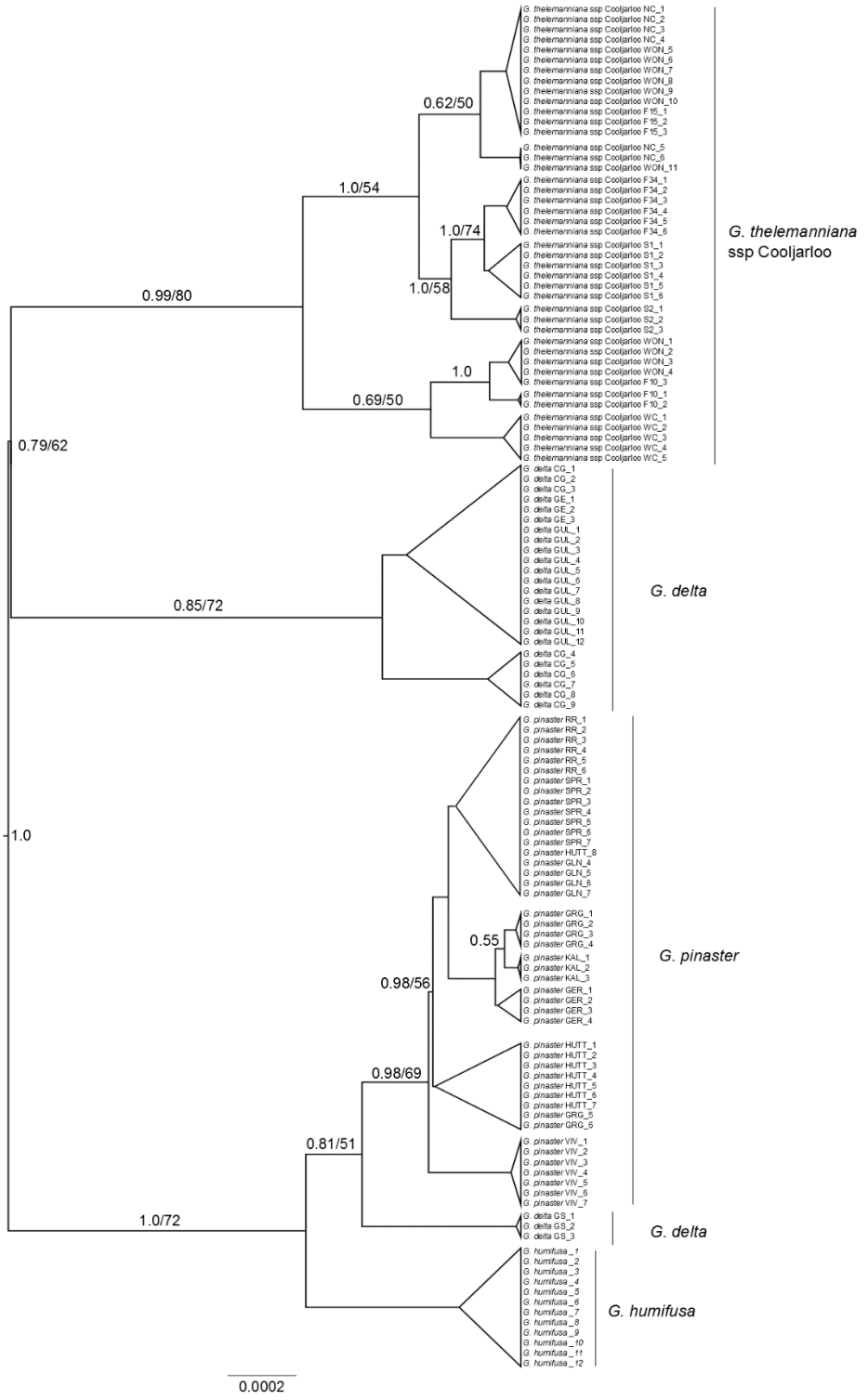

Fig.3.3 Bayesian 50% majority consensus tree of 130 chloroplast sequences from the *trnL-trnF* and *rpS4-trnL* concatenated data. Numbers above branches refer to posterior probabilities/maximum likelihood bootstrap support (PP/BS). If many individuals occurred in the same clade and the relationship within the clade remained unresolved, the branches of the original tree were replaced by triangles. Sequences are labelled by species name followed by population sampled (Table 3.1).

Table 3.3 Pairwise F_{ST} between populations collected from the *Grevillea* subclade 1 taxa. All values are significant at $P < 0.05$, values indicated in bold are not significant at $P < 0.01$.

		G. pinaster								*G. humifusa*	*G. delta*			*G. thelemanniana* ssp. Cooljarloo							
		RR	GLN	SPR	GRG	HUTT	VIV	GER	KAL	HUM	GUL	CG	GS	F15	F34	NC	WON	S1	S2	WC	F10
G. pinaster	RR	-																			
	GLN	0.04	-																		
	SPR	0.05	0.06	-																	
	GRG	0.07	0.08	0.05	-																
	HUTT	**0.04**	0.05	0.06	0.08	-															
	VIV	0.07	0.05	0.05	0.08	**0.06**	-														
	GER	0.11	0.1	0.09	0.1	0.13	0.15	-													
	KAL	**0.06**	0.09	**0.07**	**0.05**	**0.12**	0.12	**0.05**	-												
G. humifusa	HUM	0.3	0.28	0.26	0.23	0.38	0.36	0.31	0.26	-											
G. delta	GUL	0.21	0.16	0.17	0.2	0.23	0.22	0.23	0.22	0.34	-										
	CG	0.2	0.21	0.19	0.24	0.25	0.24	0.26	0.2	0.37	0.19	-									
	GS	0.17	0.12	0.15	0.17	0.25	0.17	0.29	**0.21**	0.37	0.17	0.2	-								
G. thelemanniana ssp. Cooljarloo	F15	0.17	0.19	0.16	0.18	0.19	0.2	0.14	0.09	0.34	0.26	0.28	0.25	-							
	F34	0.17	0.19	0.16	0.17	0.18	0.19	0.16	0.09	0.31	0.26	0.25	0.22	0.08	-						
	NC	0.2	0.2	0.18	0.17	0.21	0.21	0.18	0.15	0.33	0.26	0.25	0.23	0.14	0.07	-					
	WON	0.17	0.2	0.14	0.17	0.21	0.19	0.17	0.1	0.33	0.27	0.25	0.22	0.11	0.09	0.09	-				
	S1	0.23	0.25	0.2	0.21	0.26	0.27	0.19	0.13	0.34	0.32	0.33	0.29	0.08	0.15	0.19	0.14	-			
	S2	0.2	0.22	0.19	0.2	0.21	0.25	0.17	0.09	0.34	0.32	0.32	0.3	0.08	0.09	0.16	0.13	0.04	-		
	WC	0.18	0.2	0.15	0.15	0.22	0.2	0.15	0.1	0.3	0.25	0.25	0.23	0.05	0.06	0.08	0.06	0.13	0.14	-	
	F10	0.14	0.15	0.1	0.11	0.15	0.15	**0.08**	**0.04**	0.24	0.23	0.25	0.23	0.06	0.04	0.1	0.09	0.1	0.07	0.06	-

Fig.3.4. Plots of the first two principal coordinates obtained from the principal coordinate analyses of Nei's genetic distance from microsatellite data. Points are labelled by population name (Table 3.1) and ovals correspond to species or subspecies (Light blue = *G. thelemanniana* ssp. Cooljarloo, dark blue = *G. delta* and purple = *G. pinaster*), *G. humifusa* is represented by only one population (HUM).

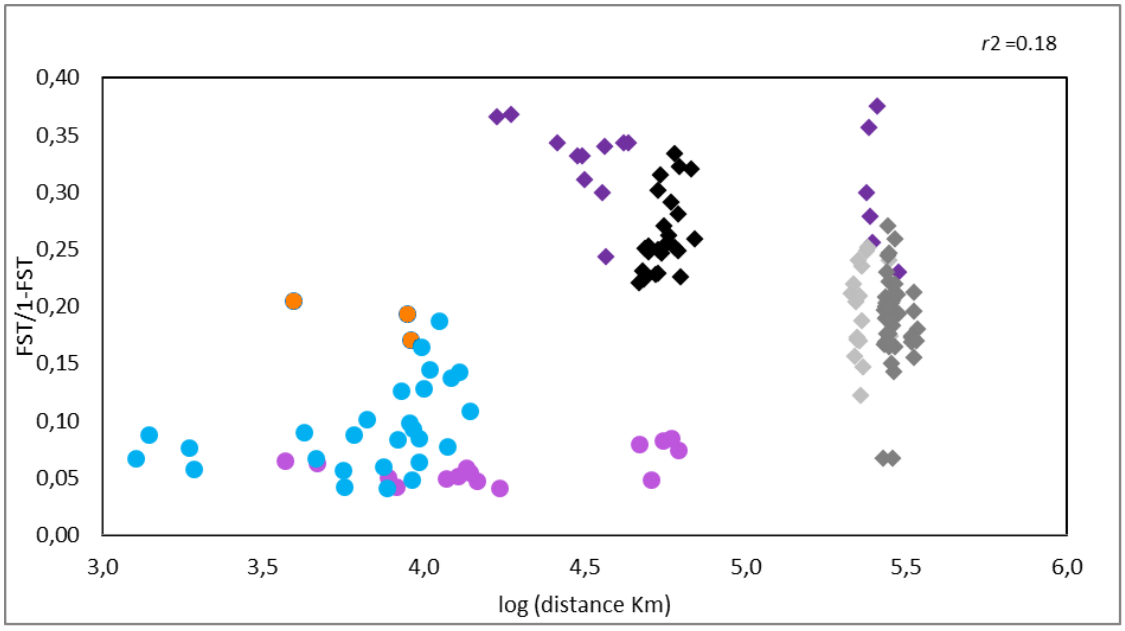

Fig.3.5. Association between genetic distance and log geographic distance across all taxa. Circles indicate comparisons within taxa; Light blue = *G. thelemanniana* ssp. Cooljarloo, light purple = *G. pinaster* and orange = *G. delta*. Diamonds indicate comparisons between taxa; Dark purple = *G. humifusa,* light grey = *G. pinaster* vs *G. thelemanniana* ssp. Cooljarloo, grey = *G. pinaster* vs *G. delta* and dark grey = *G. delta* vs *G. thelemanniana* ssp. Cooljarloo.

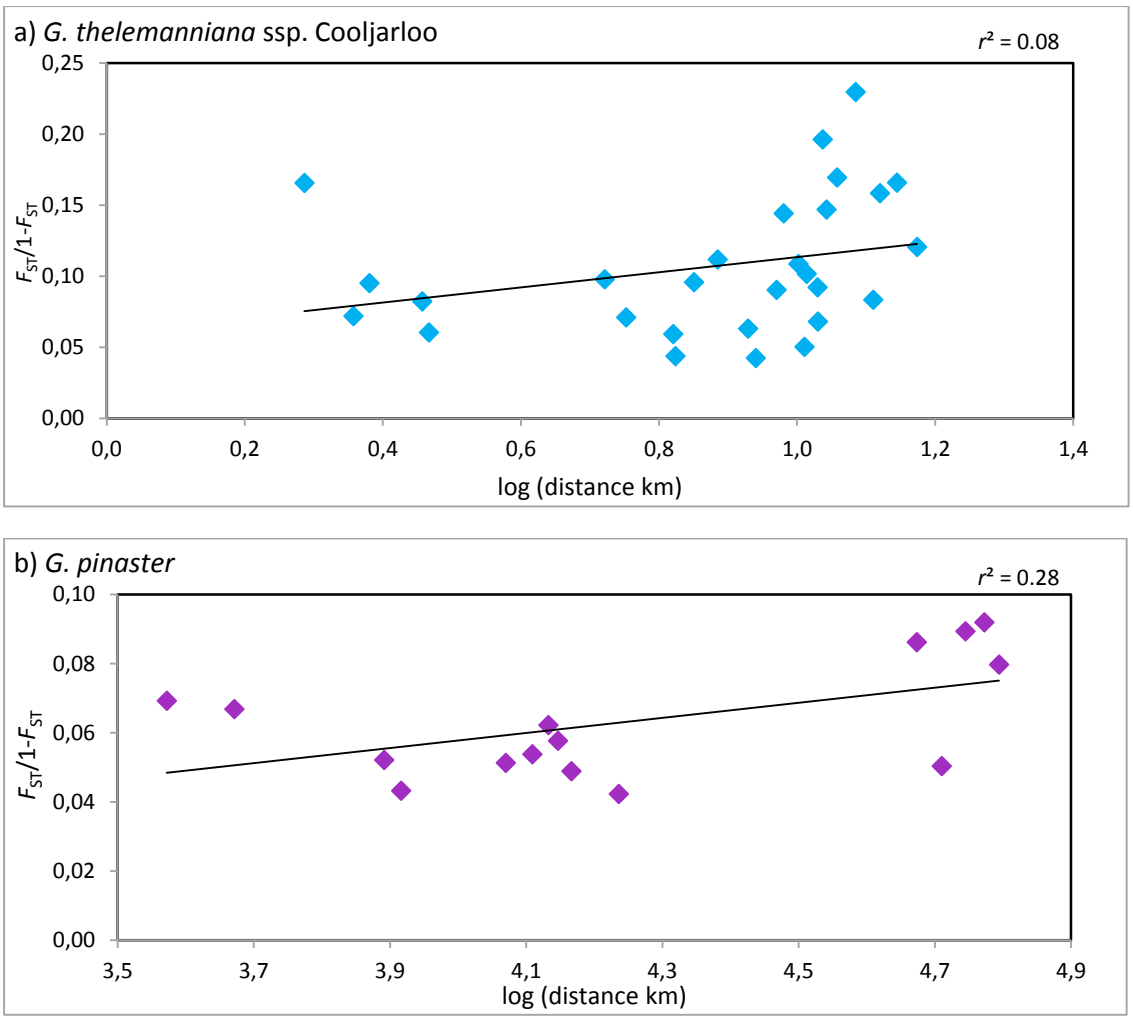

Fig.3.6. Association between genetic distance and log geographic distance across all a) within populations of *G. thelemanniana* ssp. Cooljarloo (blue) and; b) within populations of *G. pinaster* (purple).

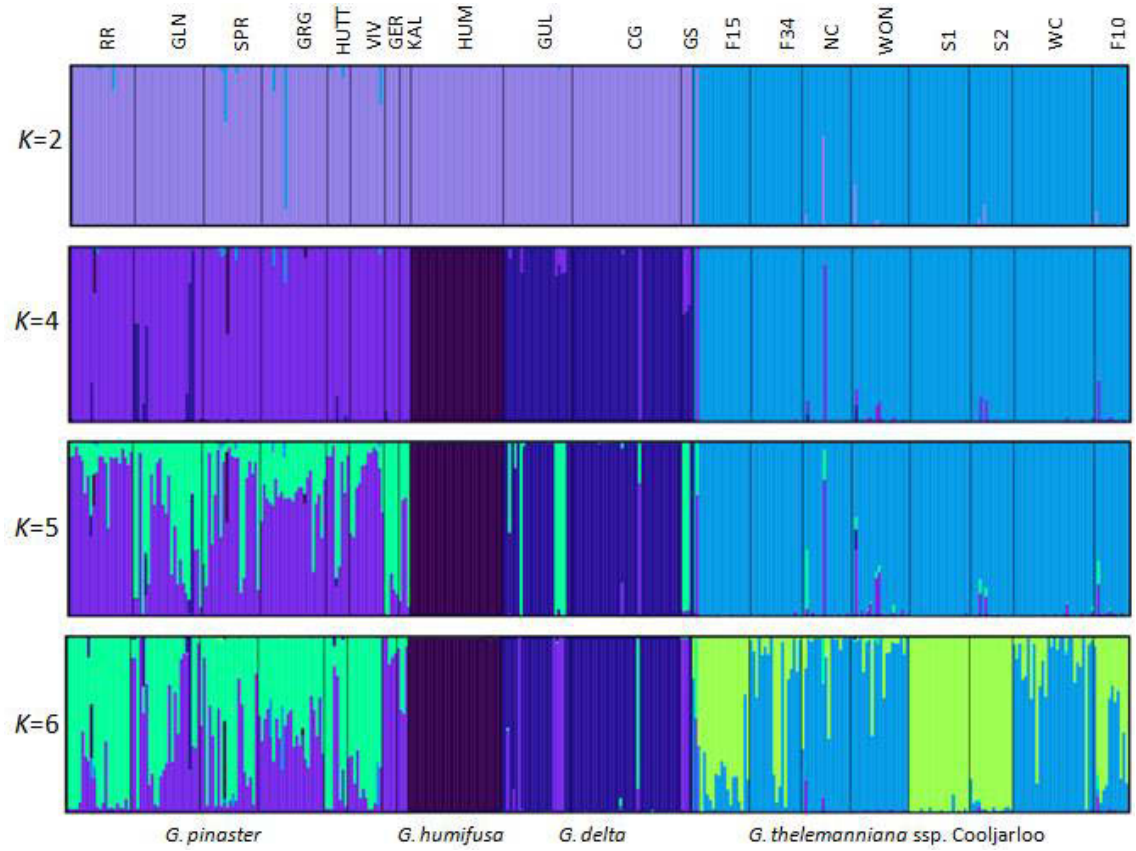

Fig.3.7. Results from STRUCTURE analyses of microsatellite data, optimal *K* under the admixture model was 2, and using no admixture; *K* = 4, illustrating proportion of membership for 367 individuals representing *G. pinaster*, *G. humifusa, G. delta* and *G. thelemanniana* ssp. Cooljarloo species. Each individual is represented by a thin vertical line, partitioned into coloured segments representing the individuals estimated membership in *K* genetic clusters. Individuals are grouped by population (indicated above the figure as for Table 3.1).

Discussion

Phylogenetic analyses of the cpDNA sequence data results supported the monophyly of *G. thelemanniana* ssp. Cooljarloo *G. pinaster* and *G. humifusa*, however the support is not strong. *Grevillea delta* on the other hand was found to be paraphyletic (Fig.3.3). Although chloroplast markers did not show strong support for monophyly of all taxa, this is likely explained by low sequence divergence which is expected for closely related species. STRUCTURE analyses of nDNA microsatellite data showed strong support for species delimitation, assigning individuals to putative species and subspecies with little apparent admixture. The different levels of resolution recovered by chloroplast and nuclear markers is commonly attributed to differential rates of gene

flow between the two genomes (Pessoa *et al.* 2012). Overall, cohesion is clear for all taxa, and results are mostly congruent between nuclear and chloroplast analyses.

The paraphyly of *G. delta* could be explained by incomplete lineage sorting or hybridisation and/or introgression with *G. pinaster*. Recent hybridisation however is not likely between *G. delta* and *G. pinaster,* since their current distributions are separated by approximately 200 kilometres. Although pollinators for these specific species are unknown, honeyeaters are commonly seen to pollinate *Greviilea* species (England *et al* 2003), however the greatest ranges they have been recoreded between plants in Western Australia is 12.5 kilometers (Byrne *et al* 2008). STRUCTURE analyses of the microsatellite data detected some evidence of genetic admixture in *G. delta* GS population from *G. pinaster*, however this is unlikely due to recent introgression since individuals bearing introgressed alleles should be geographically close to individuals or species from which the allele is derived (Rieseberg 1998). The results are therefore indicative of either incomplete lineage sorting or ancient hybridization, the latter of which could explain the apparent admixture based on analyses of microsatellite data. Under incomplete lineage sorting, species would be expected to be nonmonophyletic in phylogenetic analyses, but would be genetically differentiated at microsatellite loci because they evolve at a much more rapid rate than DNA sequences (Jarne & Lagoda 1996; Petit & Excoffier 2009). Alternatively under the hypothesis of ancient hybridization, which may have occurred during the process of speciation when species where not fully genetically differentiated resulting in little or no genetic footprint left (Edwards *et al.* 2008). Both hypotheses match aspects of the results and therefore it is currently inconclusive as to whether incomplete lineage sorting or historic hybridisation, or both, are responsible for the paraphyly of *G. delta.*

Numerous studies which have utilised population based approaches using microsatellite data to investigation incongruence between molecular datasets or phylogenies with low chloroplast diversity have been successful at delimiting species and informing on underlying reasons for incongruence (Edwards *et al.* 2008; Millar *et al.* 2011; Rossetto *et al.* 2012; Razafinarivo *et al.* 2013). Using rapidly evolving microsatellites and fine-scale population-level sampling did provide a much finer resolution of patterns of genetic structure of the subclade 1 revealed in chloroplast data and the relationship with *G. pinaster* illustrated in microsatellite data from

Chapter 2. Willis et al (2012) argued that the distinctiveness of two sets of individuals will often be the inverse of the number of specimens examined. In effect, morphological or genetic discontinuities between smaller sample sets may disappear as more specimens are examined. Interestingly, the opposite was illustrated in this study. Where originally a close relationship between *G. delta* and *G. humifusa* with *G. thelemanniana* ssp. Cooljarloo was suggested (Chapter 2), results herein illustrate a significant divergence in both chloroplast and microsatellite data. This implies that to adequately test species hypothesis it is necessary to sample densely enough to observed discontinuities between described species. Furthermore, the level of divergence between species is apparently greater than any divergence between populations using both cpDNA sequence and nDNA microsatellite data. For example, there is support from the sequence data both *G. pinaster* and *G. thelemanniana* ssp. Cooljarloo as species, but the populations within were not well resolved, with less sequence variation among species. Furthermore, Structure analyses of microsatellite data supported species level divergences but the populations were not recognised. The difference in the proportion of divergence between species as compared to the divergence between populations can be objectively used to discriminate between inter- and intra-specific differences for species diagnosis.

Grevillea thelemanniana ssp. Cooljarloo

The phylogenetic and network analyses of cpDNA data illustrated a noteworthy divergence of *G. thelemanniana* ssp. Cooljarloo from the *G. pinaster, G. delta* and *G. humifusa* lineages. Given the strong divergence of *G. thelemanniana* ssp. Cooljarloo from *G. thelemanniana* ssp. *thelemanniana* (Chapter 3) and clear differentiation from other species in this clade, formal species recognition is warranted. Additionally, there was significant genetic differentiation among populations of *G. thelemanniana* ssp. Cooljarloo, relatively high chloroplast diversity and limited sharing of haplotypes between populations are indicative of isolation among populations. This pattern would explain the significant and large divergence from *G. humifusa, G. delta* and *G. pinaster* taxa, which are likely more recently diverged. Lack of significant isolation by distance across the species range suggests that there is still ongoing gene flow among populations.

Similar levels of haplotype diversity and genetic divergence have been reported in several southwest Western Australian species, albeit across much larger geographic ranges (Byrne *et al.* 2003; Shepherd *et al.* 2013) or have had major geographic barriers between populations (Millar *et al.* 2013; Tapper *et al.* 2014). A similar pattern of significant population differentiation was found among narrowly distributed populations of the southwest Western Australia *Calothamnus quadrifidus* ssp. *teretifolius,* where the genetic system was characterised by inbreeding and historically low gene flow and poor dispersal mechanisms (Sampson *et al.* 2013). This suggests that the mating system of *G. thelemanniana* ssp. Cooljarloo could be responsible for current genetic structure, especially since no significant isolation by distance indicates that geographic distance between populations alone is unlikely to explain the high levels of differentiation. However, its distribution along creek lines and flood plain suggests a relationship with the landscape features which may also be responsible for the current genetic structure (Chapter 4).

Grevillea pinaster

There was no significant support for divergence between different forms of *G. pinaster,* and the lack of genetic clustering in the microsatellite data could indicate that current levels of gene flow are high. Although there was no strong support for divergence in the chloroplast data, there was high haplotype diversity. Since cpDNA is maternally inherited, the greater chloroplast divergence between populations indicates seed dispersal is limited, whereas paternally inherited genes are being exchanged regularly between populations resulting in low nuclear divergence. Genetic variation was expected between different recognized forms of *G. pinaster* populations, given their illustrated morphological variation, however this was not the case. Morphological variation unsupported by genetic data is not rare, considering that neutral molecular markers are in principle independent of environment (Duminil & Di Michele 2009). Given the diverse landscape around the Kalbarri region, with coastal cliffs, deep valley gorges and numerous spreading creek beds, these results could be indicative of adaptations to differing environmental conditions which is not mirrored in the molecular data. However the vast morphological differences could also be due to plasticity. Low genetic differentiation and lack of significant isolation by distance

among populations of *G. pinaster* is indicative of recent expansion (Wheeler *et al.* 2003).

Interestingly, the pairwise F_{ST} differentiation between *G. pinaster* populations was significantly lower than between *G. thelemanniana* ssp. Cooljarloo populations, even though *G. pinaster* populations occupy a much greater range. This result indicates that *G. thelemanniana* ssp. Cooljarloo may be an older lineage where populations have been isolated from one another for a longer period of time. On the other hand, gene flow in *G. pinaster* populations may be higher due to different landscape and pollinators. It also highlights that morphological variation in the flora of southwest of Western Australia can differ vastly across even closely related species and populations, irrespective of evolutionary time.

Grevillea humifusa

Lower chloroplast and nuclear genetic diversity in *G. humifusa* suggests they may have undergone substantial range contraction with a degree of associated genetic erosion. Although only one population could be sampled, the number of individuals sampled was higher than in all other population collections. Current low diversity and rarity could be a direct consequence of a recent bottleneck due to anthropogenic disturbance, given that the remaining roadside population is bounded by cleared agricultural land on both sides of the road. Alternatively narrow endemism could be due to gradual long environmental change in the region (Rossetto *et al.* 2008).

Grevillea delta

Grevillea delta also exhibited lower levels of genetic variation and significantly higher F_{ST} values between populations than *G. pinaster* and *G. thelemanniana* ssp. Cooljarloo. The populations of *G. delta* are situated inbetween the Mt. Lesueur ranges, indicating that the elevated landscape may be acting as a barrier to gene flow between populations. High divergence among populations distributed across outcrops or ridges, regardless of close proximity, has been reported in species in the south west of Western Australia (Byrne & Hopper 2008; Tapper *et al.* 2014) and *Grevillea* on the east coast of Australia (Llorens *et al.* 2004). Both large populations smpled from *G. delta* (CG & GUL) do have moderate levels of inbreeding as evident from high F_{IS} values

(>0.30). Inbreeding may influence estimates of population structure by overestimating subdivision or by producing spurious signals of admixture in STRUCTURE analyses (Edwards *et al.* 2008).

Recent contraction of both *G. humifusa* and *G. delta* ranges due to climate oscillations could explain both lower diversity in these species and close affinities with the widespread *G. pinaster*, especially if *G. pinaster* populations do in fact exist in the Enneaba region which is adjacent to the Lesueur region. Yates et al (2007) suggested that the divergence between subspecies of south west Western Australian *Verticordia staminosa* indicated a previously widespread progenitor given that suitable habitat existed between subspecies. Results were attributed to climatic fluctuations of the Pleistocene and recurrent fires which has resulted in extinction of species populations throughout their range, effectively causing populations to evolve in isolation (Yates et al 2007).

Evolutionary implications

Strong genetic differentiation between closely related species is consistent with the status of the SWAFR as a known diversification hotspot reflected in high endemism and phylogeographic breaks (Hopper 1979; Hopper & Gioia 2004; Yates *et al.* 2007; Byrne *et al.* 2008; Cooper *et al.* 2011). In particular, (Hopper & Gioia 2004) hypothesised greater rates of recent speciation in the transitional rainfall province (Figure 2.1; Chapter 2), in which all four taxa reside. The climatic oscillations during the Pleistocene, resulted in cyclic expansion and contraction of mesic habitat in the SWAFR (Hopper 1979; Hopper & Gioia 2004), which have the potential to isolate populations and initiate allopatric speciation, even in the absence of strong intrinsic barriers (Rossetto *et al.* 2012).

The high level of genetic diversity in *G. pinaster,* that dominates the more arid transitional rainfall province, is consistent with a historically widespread distribution. A similar pattern of diversity and distribution was found in *Acacia lindleyi* of the *Acacia saligna* species complex, which is found in the mesic northern catchments of the TRP (Millar et al. 2011), suggesting that the TRP may have been the centre for diversity, or a refugium, during the contraction of flora during arid phases of the Pleistocene. Given that southern populations of *G. pinaster* have been recorded near the distributions of

G. delta and *G. humifusa*, and the close phylogenetic relationship, could indicate that these species reflect divergence through allopatric speciation. Basic allopatric differentiation implies that geographic barriers to gene flow are, or have been, important enough to cause reproductive isolation and subspecies differentiation (Rossetto *et al.* 2007). Under this evolutionary model, differentiation is purely driven by IBD and genetic drift and can potentially lead to the accumulation of mutations, resulting in reproductive incompatibility among populations. These conclusions on evolutionary patterns and speciation add to a growing list of studies that illustrate a complex relationship with climatic fluctuations during the Pleistocene, which has resulted in southwest Western Australia's high endemism and species diversity.

Conservation implications

Grevillea humifusa is currently listed as threatened (Smith 2013), and given that only one population is known and genetic diversity of this population is low, reclassification of this species to a more critical status is needed. The lower genetic diversity in populations of *G. delta*, indicates that conservation efforts should focus on protecting its remaining habitat, perhaps treating populations as separate conservation units and managing accordingly (Coates 2000). There was no significant support for genetically different forms of *G. pinaster.* Consequently focusing conservation efforts towards potential cryptic species in this group is not necessary. The strong support for *G. thelemanniana* ssp. Cooljarloo to be recognised as unique species has conservation repercussions for the conservation and restoration management of this species, as it is distributed near an operating mine. The results from this study illustrate how genetic data can greatly contribute to and inform conservation outcomes.

Chapter 4 - Landscape genetic analyses reveal barriers and corridors for gene flow within and among riparian populations of a rare *Grevillea* (Proteaceae).

Introduction

One of the key ecological processes affecting spatial genetic structure is gene flow (Slatkin 1985; Manel *et al.* 2003; Leblois *et al.* 2004; Holderegger *et al.* 2010). Gene flow in and among plant populations occurs mainly through two processes, dispersal of diploid embryos in seeds and haploid gametes in pollen (Holderegger *et al.* 2010). These processes therefore determine the genetic structure of plant populations, and are in turn influenced by the composition of landscapes they inhabit, as well as life history traits such as pollination mechanisms and mating systems (Storfer *et al.* 2007; Holderegger *et al.* 2010). For example, riparian landscapes have been shown to influence the directional dispersal of seeds in a wetland hibiscus resulting in an accumulation of genetic variation at the base of a river system Kudoh and Wigham (1997). On the other hand, Kramer *et al.* (2011) found that differences in primary pollinators drove the different levels of genetic structure seen in three species of the perennial *Penstemon,* which all had similar seed dispersal and habitats. This combination of ecology and population genetics has led to the emergence of landscape genetics as a scientific discipline which aims to assess how landscape features might influence gene flow and spatial genetic structure (Manel *et al.* 2003; Manel & Holderegger 2013). Landscape genetic studies on plants are largely under-explored (Holderegger *et al.* 2010), and are especially underrepresented in biodiversity hotspots (Storfer *et al.* 2010).

Identifying potential barriers to gene flow or how the landscape provides functional connectivity or corridors and how these influence spatial genetic structure is a key focus of landscape genetic studies (Storfer *et al.* 2007). In some habitats, geographic barriers impacting gene flow may be obvious, for example mountain ranges (Hu *et al.* 2008; Trénel *et al.* 2008), intervening habitats between riparian environments (Kitamoto *et al.* 2005; Mitsui *et al.* 2010) or anthropogenic induced fragmentation (Cegelski *et al.* 2003; Epps *et al.* 2007). However, in many landscapes, obstacles to

gene flow may be more cryptic, for instance, historical vegetation changes (Jacquemyn *et al.* 2004), plant density (Torres *et al.* 2003) and/or a combination of both biotic (e.g. pollinator response to habitat features) and abiotic (e.g. wind as a dispersal vector) factors (Holderegger *et al.* 2010). Alternatively, landscape features such as rivers, creek lines and flood plains can provide corridors facilitating dispersal (Storfer *et al.* 2007; Hu *et al.* 2008).

More than a decade of landscape genetic studies (Manel & Holderegger 2013) coincides with the recent expansion of statistical approaches combining analyses of both genetic (e.g. microsatellites) and spatial data available for describing population genetic structure. Through spatial analyses of individuals and the spatial arrangement of genetic clusters, landscape features affecting dispersal among and within populations can be better understood. Clusters can be viewed as genetically divergent groups of individuals where gene flow is inferred to be impeded by a physical barrier (Guillot *et al.* 2005; Chen *et al.* 2007; Corander *et al.* 2008). Clustering models applying Bayesian assignment methods have become very popular in landscape genetics (Storfer *et al.* 2010), promoting the shift from population based to individual based analyses (Segelbacher *et al.* 2010). These methods are especially useful when individuals are not obviously clustered in discrete populations but rather arranged more continuously across habitats, such as in many riparian habitats, and when barriers to gene flow may be more cryptic. In these situations, the landscape genetic approach, employing geographic information, is most valuable (Manel *et al.* 2003).

Riparian distributions are typically characterized by largely linear patterns in which gene flow is likely to occur more readily within each habitat than between (Wei *et al.* 2013). For example, seeds that can be dispersed by water (hydrochory) will promote gene flow within but not among riparian corridors. Hydrocohory as a force of dispersal can be responsible for influencing spatial genetic structuring of populations along riparian environments at both small and large scales (Kudoh *et al.* 2006; Nilsson *et al.* 2010). In *Myricaria laxiflora* hydrochory influences directional dispersal of seeds and propagules from upstream toward downstream populations (Liu *et al.* 2006). However, pollinator behaviour, zoochory or wind dispersal have been shown to be efficient mechanisms at counteracting downstream accumulation of genetic diversity, by creating bidirectional gene flow (Tero *et al.* 2005; Markwith & Scanlon 2007).

Additionally, flooding of riparian environments can significantly increase the range of seed dispersal (Nilsson *et al.* 2010). For example, in the perennial *Primula sieboldii,* flooding within streams led to greater seed dispersal than by gravity alone (Kitamoto *et al.* 2005), resulting in an erosion of spatial genetic structure. In the tree *Fraxinus mandshurica*, weak population genetic structure was found among rivers due to large scale flooding occasionally connecting rivers (Hu et al. 2008). However, populations of the stream lily *Helmtholtzia glaberrima* were strongly genetically structured within a creek (Prentis *et al.* 2004), indicating that gene flow is not always enhanced in riparian habitats by hydrochory. These studies show that patterns of genetic variation within and among riparian plant populations are not necessarily predictable.

To better understand the impact of a riparian landscape on the spatial distribution of genetic variation and gene flow, Tero *et al* (2003) derived five different theoretical models of dispersal (Fig 4.1). The first three of these assume that populations or patches are principally permanent, with predicted gene flow ranging from freely among a more or less uniform "spatially extended population", to a "regional ensemble" or "classic stepping stone" model with limited migration between populations. The "source-sink" model predicts that dispersal is unidirectional, resulting in downstream populations with greater genetic diversity relative to upstream source populations. When dispersal is not limited to a linear habitat and occurs across the landscape, patterns of genetic differentiation may result in alternative models of gene flow showing isolation by distance (IBD) or a "classic metapopulation" model. The "classic metapopulation" model assumes that populations are more or less ephemeral, where the patterns of migration may be more diverse and populations can be founded by different number of individuals from single or multiple source populations (Tero *et al.* 2003). These models provide a context for the analysis of the spatial distribution of neutral genetic variation within and among populations, which can be inferred to characterise the impact of landscape features on gene flow.

The rare and threatened *Grevillea thelemanniana* ssp. Cooljarloo (Proteaceae) is currently a phrase named subspecies listed as Priority 1 by the Department of Parks and Wildlife (Smith 2013). This taxon is naturally restricted to an area of 20km^2 just south of Cervantes in the Southwest Australian Floristic Region (SWAFR) in Western Australia. Although the taxon is abundant within this range, its distribution is restricted

along creek lines and within floodplains. Within its range is an active mineral sand mining operation, for which expansion is being considered. Expansion of the mine would likely result in further land clearing and impact of vital landscape features such as drainage lines, as well as disruption to pollinator abundance and behaviour. The quantification of genetic diversity, its spatial structuring and the influence of landscape features on these patterns underpins better conservation management of *G. thelemanniana* ssp. Cooljarloo.

Through combining landscape genetics and riparian modelling, I aim to uncover patterns of gene flow, genetic structure and diversity within and between populations of *G. thelemanniana* ssp. Cooljarloo. Specifically I aim to answer the following questions: (1) Does the spatial genetic structure reflect landscape characteristics? (2) Do creek lines and floodplains impact differently on dispersal and genetic structure of *G. thelemanniana* ssp. Cooljarloo populations? (3) Do patterns of gene flow support any of the riparian models of Tero et al (2003)? To answer these questions, I used landscape genetic statistical approaches to analyse data from nine nuclear microsatellites.

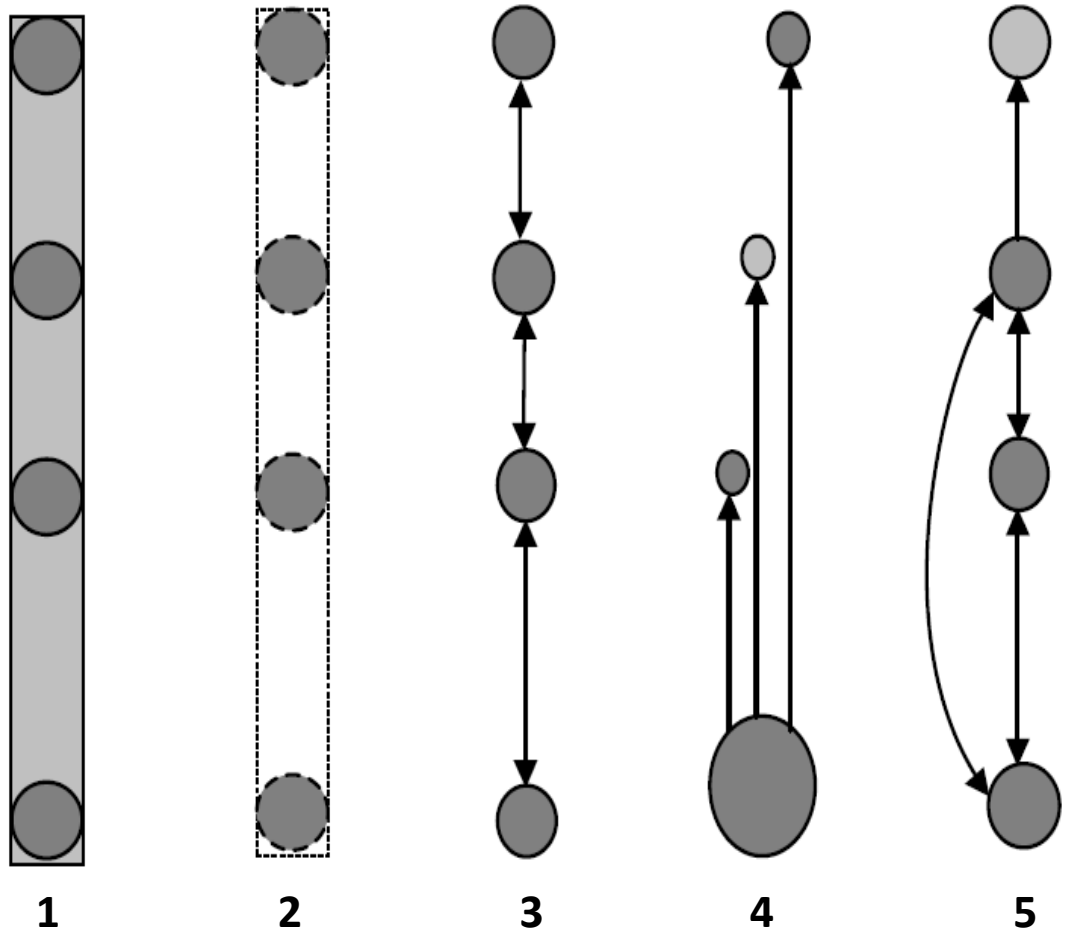

Fig.4.1. Diagram representing different models of dispersal between populations in riparian habitats (1) spatially extended or genetically uniform population with gene flow freely is moving across all populations (2) regional ensemble with limited recurrent gene flow between populations (3) classic stepping stone (4) source-sink and (5) classic metapopulation (adapted from Tero *et al* (2003)).

Materials and Methods

Distribution and study site

Grevillea thelemanniana ssp. Cooljarloo populations are located on the Swan Coastal Plain (SCP), 30 km southeast of Cervantes, Western Australia. The SCP is a narrow, low relief coastal strip of alluvial and aeolian sands confined by the Indian Ocean to the west and the Darling Scarp to the east (Seddon 2004). Plants are restricted to, but occur somewhat continuously along, creek lines and within large floodplains. Although the floodplains frequently flood during the winter months, the creek lines often have stagnant unconnected pools of water which only flow together during heavy rainfall

(personal observation). The region is situated between Nambung National Park (S 30°35,12' E 115°09'45') and Tronox sand mine (S 30°39'00', E115°27'14') (Fig.4.2). The mine began in 1989 and currently occupies 9800 hectares of land. The north-western expansion had the greatest impact on *G. thelemanniana* ssp. Cooljarloo populations, where clearing of the land on both sides of the creek line containing individuals, has effectively isolated them from the remainder of the subspecies range. The land cleared between the northern and southern populations

Subspecies biology

Grevillea thelemanniana ssp. Cooljarloo has no obvious adaptations for seed dispersal. Pollination experiments in the glasshouse have shown that plants are self-compatible (pers comm., Digby Growns). Based on its numerous inflorescences, style length, nectar laden and bright red flowers, the main pollinators are likely to be a variety of insects and native birds, specifically honeyeaters (Meliphagidae) (England *et al.* 2003; Phillips *et al.* 2010). Ants and small mammals have been shown to effectively disperse seed and pollen in *Grevillea* (Auld & Denham 1999; Auld & Denham 2001). The introduced honeybee *Apis mellifera* has been recorded to visit flowers of multiple species (Celebrezze & Paton 2004). Studies on *Grevillea* have revealed highly variable mating strategies, ranging from self compatible (Ayre *et al.* 1994; Smith & Gross 2002) to self-incompatible (Hermanutz *et al.* 1998). However, *Grevillea* species often lack seed dispersal adaptations (Makinson 2000; England *et al.* 2003), and generally have low fruit to flower ratio (Harriss & Whelan 1993; Hermanutz *et al.* 1998; Vaughton 1998; Caddy & Gross 2006). Low fruit to flower ratio can result in strongly genetically structured populations, which have been reported in numerous studies (England *et al.* 2001; Hoebee & Young 2001; Llorens *et al.* 2004; Holmes *et al.* 2009). Furthermore limited pollen dispersal by predominantly nearest neighbour mating has resulted in strongly genetically structured populations (Hoebee & Young 2001; Smith & Gross 2002; England *et al.* 2003).

Sampling strategy

A 30m digital elevation model (DEM) was constructed from 974 spot heights provided by Tronox Inc. (CAD resources) using ArcGIS 10.1 (ESRI, Redlands, CA, USA) and visualized as contour lines to illustrate elevation changes within the study region

(Fig.4.2). Creek lines were drawn onto the map based satellite images and previous landscape mapping provided by Tronox Inc. Leaf samples were collected from three regions within the study area based on GPS points provided by the mine site environmental surveyors, who had previously recorded all sites of *G. thelemanniana* ssp. Cooljarloo populations in the area. All individuals were confined to either creek beds or floodplain populations. Three large creeks occur within the region. These and three adjoining floodplains were sampled for this study (Fig.4.2). The creek line in the west starts with a *Eucalyptus* dominated woodland. *Grevillea thelemanniana* ssp. Cooljarloo individuals begin to occur where habitat changes to *Banksia* dominated woodland. This creek opens up into a floodplain before continuing west as a creek into Nambung National Park, (Fig.4.2). The creek line in the north is cut off at both ends by a road with cleared agricultural land at one end and the Tronox mining operations at the other end. It is difficult to determine where the southern creek begins as it is cut off by farm land and the Tronox mine operations site in the east, but it comes out into the Cooljarloo swamp in the southwest. Two sites where *G. thelemanniana* ssp. Cooljarloo individuals have been recorded could not be sampled (Fig.4.2; indicated by stars). The first site was recently burnt and contained no live plants. The second site in the north was inaccessible due to current expansions of the Tronox Mine.

Due to the largely continuous distribution of *G. thelemanniana* ssp. Cooljarloo along the creek lines and flood plains I sampled in three ways:

1) Within the northern, southern and western creek lines I collected samples based on a constant sampling fraction (Lowe *et al.* 2009) by defining a 20mx20m perimeter on both sides of the creek every 0.5-1 kilometre along the creeks. I sampled every individual within the perimeter, and called these populations (Table 4.1).

2) In order to assess fine scale spatial genetic change along the three creek lines I sampled 1-3 individuals every 100 meters between populations and called them spatial collections (Table 4.1).

3) As with the creek line populations, I sampled every individual from a randomly selected 20x20 perimeter within three isolated floodplains adjacent to north, south and west creek lines and called these populations (Table 4.1).

Each individual sampled was recorded and mapped using GPS (Garmin) and plotted in ArcGIS 10.1 (ESRI, Redlands, CA, USA).

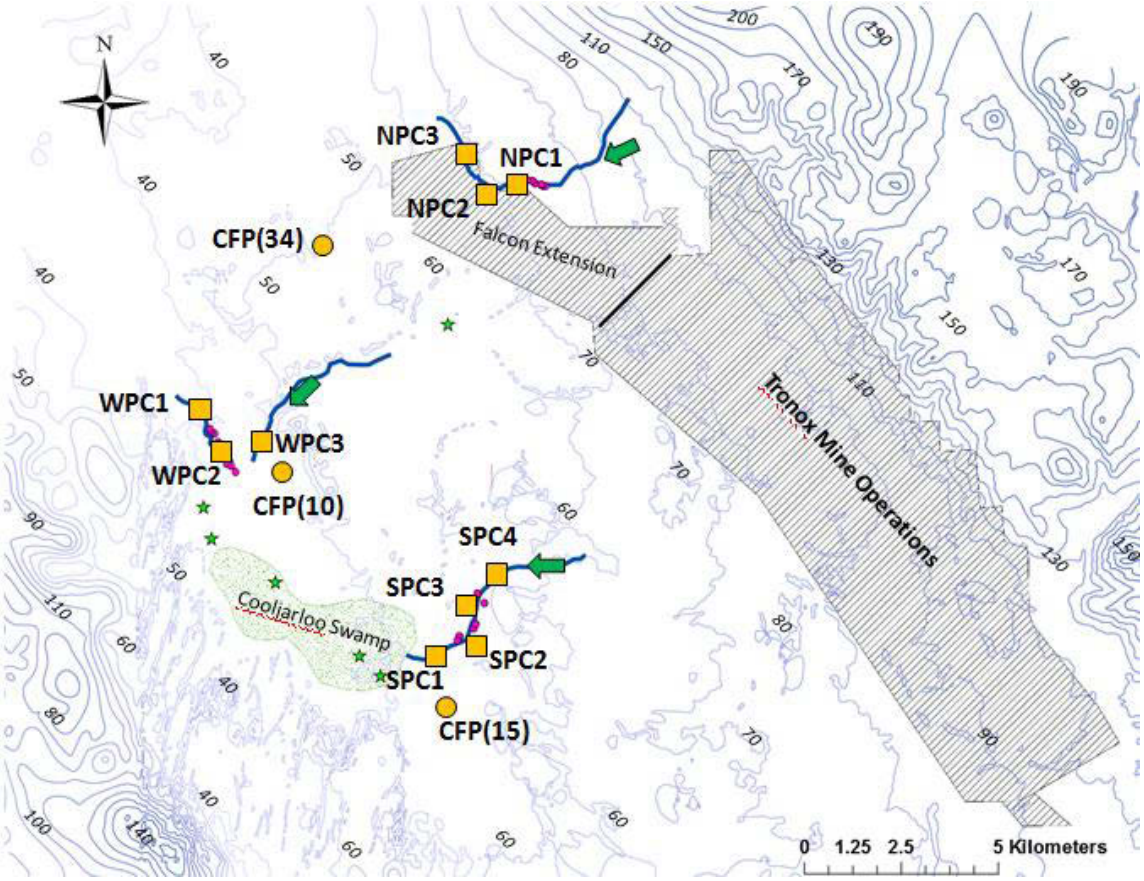

Fig. 4.2 Map of the locations of sampled populations of *G. thelemanniana* ssp. Cooljarloo from creeks (orange squares) and floodplains (orange circles). Sample site names correspond to those in Table 4.1. Creek lines are drawn onto the map and arrows indicate direction of flow. Numbered contour lines represent elevation, ranging from low (light blue) to high (dark blue) elevation (meters above sea level). The green stippled area illustrates Cooljarloo Swamp. Stars indicate recorded locations which could not be sampled due to fire and mining operations.

DNA extraction and genotyping

Genomic DNA was extracted using the modified CTAB extraction method (Doyle 1990). Quantity and quality of DNA was assessed by visualization on a 1.5% agarose gel and a Nanodrop (Thermo Fisher, Wilmington, DE). Five microsatellite loci from *Grevillea thelemanniana* ssp. *thelemanniana* (Hevroy *et al.* 2013) were successfully cross amplified in *G. thelemanniana* ssp. Cooljarloo and a further 2 primers were developed from the 454 shotgun sequencing results for *G. thelemanniana* ssp. *thelemanniana*. I also cross amplified two primers from *Grevillea macleayana* (England *et al.* 1999) which resulted in higher variation than the *G. thelemanniana* ssp. *thelemanniana* primers. The microsatellites were amplified using PCR containing 2-10ng DNA template, 2µl of 5xPCR buffer (Invitrogen, including dNTPs), 2mM $MgCl_2$, 0.1µM of the fluorescently labeled primer (6-FAM, NED &VIC) and unlabeled reverse primer and 0.1ul *Taq* DNA polymerase (Invitrogen) and brought to a total volume of 10µl with sterile water. PCR amplification was conducted in Eppendorf thermocyclers under the following amplification conditions; One cycle at 94°C for 3min followed by 30 cycles at 94°C for 40 secs, 54°C -60°C (Hevroy et al 2013) for 40 secs, 72°C for 40 secs and a final extension at 70°C for 15 min. A Beckman DNA Analyzer was used to visualize PCR fragments and genotypes were scored manually using Beckman Fragment Analyzer software.

Genetic diversity and relative tests

Standard genetic diversity parameters, total number of alleles (N_A), observed and expected heterozygosity (H_O and H_E), inbreeding coefficient (F_{IS}), were estimated for each of the floodplain and creek line population collections using GenAlEx 6.5 software (Peakall & Smouse 2012). Allelic richness (A_R) and private allelic richness (P_R), were estimated for each population at every locus with HP-rare (Kalinowski 2005) which uses rarefaction to account for differences in sample size. *S*ignificance of differences between allelic and private allele richness within populations along creek line compared to floodplain populations were assessed by t-tests. Deviations from Hardy-Weinberg equilibrium (HWE) and linkage disequilibrium were tested using an adjusted sequential Bonferroni correction (Rice 1989) and 10,000 randomizations at the

nominal significance level (5%) in GenAlEx 6.5. MicroChecker 2.2.3 (Van Oosterhout *et al.* 2004) was used to detect presence of null alleles and potential scoring errors.

Population based structure analyses

Pairwise population genetic differentiation was estimated by F_{ST} with ARLEQUIN V.3.5 (Excoffier & Lischer 2010). Since F_{ST} can falsely reflect differentiation among populations when heterozygosity within populations is high (Jost 2008; Heller & Siegismund 2009). I also calculated pair wise differentiation using Jost's D_{EST} with SMOGD (Crawford 2010). Jost's D_{EST} is a measure of pure differentiation, which is based on effective number of alleles and is useful in conservation genetics of rare species, especially when decisions have to be made on their contribution to overall population differentiation (Jost 2008; Meirmans & Hedrick 2011). Genetic distances among populations and spatial collections were also estimated using Nei's genetic distance (Nei's *D*). Ordinations were conducted with all three distance matrices using Principal Coordinates Analysis (PCoA), and the first two components were plotted. The total genetic diversity was partitioned into within and among population components by hierarchical Analysis of Molecular Variance (AMOVA), using ARLEQUIN, with 10,000 permutations and significance levels set with nominal values of 0.05 following Bonferroni correction. Isolation by distance among populations of all three creek lines was evaluated using Mantel tests between log transformed geographic distance and genetic distance [standardized F_{ST}/ (1-F_{ST})] (Rousset 2000). Paired Mantel tests of genetic distance against geographic distance between individuals sampled within each creek line were conducted to test for isolation by distance. Significance of the correlation between population differentiation and spatial distance was tested using *P*-values from 10,000 permutations in GenAlEx 6.5.

Individual based genetic structure analyses

Genetic structure in *G. thelemanniana* ssp. Cooljarloo was further analysed using STRUCTURE 2.3.3 (Pritchard *et al.* 2000), which uses a Bayesian algorithm to estimate probability of membership of an individual (*q*) to an assumed genetic cluster, using a Monte Carlo Markov Chain algorithm. Given the restricted range of populations, I implemented the admixture ancestry model with correlated allele frequencies. In admixture models, the allele frequencies are less contained than in no admixture

models and as a consequence can detect geographic clines in allele frequencies and ancestry coefficients (François & Durand 2010). The number of potential genetic clusters (*K*) in the dataset was tested from 1 - 10, replicated 10 times for 1 million generations and a burn-in of 100,000. Optimum *K* was identified using the Δ*K* function of (Evanno *et al.* 2005), which is based on second rate change of likelihood function with respect to *K,* and was calculated using STRUCTUREHARVESTER (Earl & von Holdt 2012). STRUCTURE makes the implicit assumption that individual cluster membership does not display any particular spatial pattern (Guillot *et al.* 2009). Complete spatial randomness however is unlikely in any population. I therefore also implemented clustering analyses that take into account a spatial component.

Since admixture may result from processes that are intrinsically spatial, recent improvements of Bayesian assignment methods have incorporated geographically explicit prior distributions (Storfer *et al.* 2007). Chen et al (2007) proposed an alternate clustering program that uses both genetic data as well as spatial data to estimate optimal number of clusters (*K*), based on spatial autocorrelation within an individual model of admixture. TESS 1.3 (Chen *et al.* 2007) delineates the maximum number of clusters in the dataset by implementing a Bayesian clustering algorithm that uses a hidden Markov random fields (HMRFs) model and allows for a detailed admixture analysis. The Markov fields method used in TESS is based on the spatial autocorrelation assumptions that allele frequencies at a geographic site are more likely to be similar to allele frequencies at neighbouring sites than at distant sites (François & Durand 2010). Spatial autocorrelation is more appropriate for populations which have a higher degree of assortative mating and are only weakly spatially related (Chen *et al.* 2007). It therefore also accounts for some degree of isolation by distance, and has been shown to outperform other spatial clustering programs when IBD is present in the dataset (Safner *et al.* 2011). The degree of spatial dependence incorporated in the analyses is determined by a spatial parameter (ψ). The spatial parameter represents the strength of spatial autocorrelation, ranging from non-informative spatial parameter at zero to highly spatially informative at one. In this analysis, the spatial parameter was set ψ = 0.6, as recommended by the authors (Chen *et al.* 2007). Each run was performed under the admixture model (used the non-admixture model for comparison), using 50,000 cycles with a burn-in of 10,000 for *K*=2-8 with 100 iterations for each *K*. Based

on the 10 highest deviance information criterion (DIC) values for each *K*, I determined the most likely number of clusters (*K*) from where the DIC values plateau (Chen *et al.* 2007). I obtained values of the proportion of membership of each individual to given clusters at the optimal *K* (q_i), to illustrate these values. Similarity among different runs was assessed using the similarity coefficient (*H*) in CLUMPP 1.1.2 (Jakobsson & Rosenberg 2007) for both replicates of STRUCTURE and TESS runs at optimal *K*.

Contemporary migration by seed among creek lines, measured as first generation migrants, was inferred through assignment tests in GeneClass2 (Piry *et al.* 2004). The likelihood of an individual belonging to the source from which it was sampled, or having come from another source was determined using Rannala and Mountain's (1997) likelihood Bayesian-based method. The Bayesian method performs well for any value of F_{ST} and because it does not assume that the true source population is included in the analysis (Cornuet *et al.* 1999), which is appropriate for this analysis. I ran the software using the L_home test statistic, as it could not be established with certainty that all potential sources had been sampled, with 10,000 simulated individuals at a threshold of 0.01. Furthermore, I applied the stringency criterion of He *et al.* (2004), where an assignment of an individual is considered unambiguous when the difference (δ) between the largest and second largest log-likelihood was above a predetermined threshold. I applied a stringency level of δ = 1.0, i.e. a multi-locus genotype had to be at least 10 times more likely to be assigned to a source population than the population from which it was sampled. Under this stringency level He *et al.* (2004, 2009) determined that individuals assigned to populations other than which they were sampled as a confident indication of a seed dispersal event.

Results

GIS data

Distribution of *G. thelemanniana* ssp. Cooljarloo was restricted to lower elevation habitats. Elevation within the study site (including mine operations site) ranged from 30 to 205 meters. Areas where *Grevillea thelemanniana* ssp. Cooljarloo populations were mapped and sampled had an average elevation of 51±0.48 meters (37-79m), which was significantly lower than the average elevation in habitats which *G. thelemanniana* ssp. Cooljarloo wasn't found, including the mine operations site, which was 91±2.3 (52-91m) (P<0.001).

Genetic diversity and relative tests

Only locus GM13 exhibited potential null alleles. Genetic analyses without this marker did not generate significant differences in results compared to the full data set. It was therefore included in analyses to increase robustness. The nine polymorphic loci amplified a total of 102 alleles across 13 populations. Mean estimates (and range) of genetic parameters across 9 loci were mean number of alleles 5.2 (range = 3.6 – 6.2), mean allelic richness 5.1 (3.9 – 5.9), observed heterozygosity 0.64 (0.58 – 0.77) and expected heterozygosity 0.69 (0.63 – 0.76). Private allele frequency was significantly lower in the south and west populations than in the floodplain and north populations (P < 0.001). F_{IS} deviated significantly from zero within five populations (Table 4.1), with fewer heterozygotes than expected, indicating possible inbreeding in these populations. There were no significant differences in any genetic parameters (A_R, N_A, P_R, H_E, F_{IS}) between upstream and downstream populations (P < 0.001).

Table 4.1 Sampling information and genetic diversity statistics averaged over 9 microsatellite loci for 13 natural populations of *Grevillea thelemanniana* ssp. Cooljarloo.

Site	Collection ID	Collection type	N	N_A	A_R	pA_R	H_O	H_E	F_{IS}
North	NPC1	Population	22	5.3	5.1	0.12	0.608	0.691 *	0.125
	NPC2	Population	17	5.2	5.0	0.27	0.606	0.684 *	0.114
	NPC3	Population	23	5.7	5.0	0.14	0.629	0.677	0.067
	NSC	Spatial collection	28	5.9	4.8	0.05			
	CFP(34)	Floodplain	27	6.7	5.6	0.2	0.622	0.695 *	0.108
South	SPC1	Population	23	5.7	4.9	0.01	0.619	0.663	0.116
	SPC2	Population	15	5.1	4.8	0	0.670	0.727	0.065
	SPC3	Population	28	5.2	3.9	0	0.586	0.644	0.005
	SPC4	Population	15	4.0	4.3	0	0.633	0.628	0.079
	SSC	Spatial collection	16	4.6	4.2	0.14			
	CFP(15)	Floodplain	30	6.1	5.2	0.25	0.672	0.696	0.037
West	WPC1	Population	22	6.6	6.0	0.28	0.777	0.739	-0.053
	WPC2	Population	26	6.8	5.9	0.09	0.566	0.723 *	0.221
	WPC3	Population	29	6.7	5.6	0.05	0.696	0.729	0.043
	WSC	Spatial collection	16	5.9	5.4	0.13			
	CFP(10)	Floodplain	28	6.7	5.9	0.15	0.667	0.760	0.118

N number of individuals genotyped, N_A total number of alleles detected, A_R and pA_R allelic and private allele richness based on the minimum population size of 15 diploid individuals respectively. H_O observed heterozygosity, H_E expected heterozygosity, F_{IS} inbreeding coefficient. These last parameters were not calculated for spatial collections because they were not technically populations.

*Indicates significant deviation Hardy-Weinberg equilibrium based on 10,000 permutations and Bonferroni corrected *P* value ($P<0.05$).

Population based genetic structure

Mean (and range) pairwise F_{ST} across all populations was 0.087 (0.028 – 0.178) and mean pairwise D_{EST} was 0.215 (0.072 – 0.422) (Table 4.2). D_{EST} values were significantly larger than F_{ST} (P < 0.001), but showed similar patterns of divergence. There was significant genetic divergence for all D_{EST} pair wise comparisons and all F_{ST} comparisons except between NPC1 and NPC2 and between SPC1 and SPC4 (Table 4.2). There was moderate to high genetic differentiation among the pairs of populations from the different creek lines. Furthermore the F_{ST} values were higher for inter-creek population pairs than for intra-creek population pairs (mean = Inter: 0.112; intra: 0.064, P = 0.024), and D_{EST} values showed the same trend (mean = Inter: 0.273; intra: 0.137, P = 0.014). The highest F_{ST} value was between NPC2 and SPC3 (0.178) and the lowest between WPC2 and WPC3, which reflected geographic distance (12.2km and 0.9km, respectively). The highest D_{EST} value was between NPC1 and SPC2 (0.422) and the lowest between NPC1 and NPC2, which also reflected geographic distance (11.7km and 0.3km, respectively).

An AMOVA was used to assess the partitioning of total genetic variation on multiple geographical scales (Table 4.3). A small, but significant, amount of variation was partitioned among the three clusters (4.52%), with a similar amount of variation partitioned among populations within clusters (4.58%) and among individuals within populations. The majority of variation was held within individuals (86.8%). When applied to the creek lines, 89.4% of total genetic variation was partitioned within populations. A significant amount of genetic variation was partitioned among the three creek lines (5.5%, P < 0.001) and among populations within creek lines (5.1%, P >0.001). When AMOVA was applied to downstream and upstream populations, no significant amount of variation was partitioned among them (2.0%, P = 0.02).

Nearly identical ordinations were generated from each of the three distance matrices, therefore only one is shown (Nei's *D)*. The first two principal axes in the PCoA (Fig.4.3) explained 44.6% and 18.43% of the total genetic variation, respectively. The PCoA shows the northern, southern and western populations clusters, and the floodplain populations cluster together in-between creeks. The slope of the regression of pair wise population values of F_{ST}/ (1–F_{ST}) against logarithmic geographic distance was

significant (r^2 = 0.230, P=0.0001; Fig. 4.4), indicating isolation by distance across the sampled populations. Genetic distance versus geographic distance for pairwise comparisons of individuals within creek lines illustrated no significant relationship in the northern, southern or western creek lines (Fig.4.5; r^2 = 0.008; r^2 = 0.0012; r^2 = 0.033; respectively, $P > 0.05$).

Table 4.2 Pairwise D_{EST} and F_{ST} between 10 creek line and 3 flood plain populations of *Grevillea thelemanniana* ssp. Cooljarloo. Upper diagonal: Jost (2008) D_{EST} and lower diagonal: mean F_{ST} values. Values are significantly different from zero for all population comparisons at the 5% level after Bonferroni correction for multiple comparisons except those indicated in bold.

	CFP10	CFP15	CFP34	NPC1	NPC2	NPC3	SPC1	SPC2	SPC3	SPC4	WPC1	WPC2	WPC3
CFP10		0.179	0.094	0.191	0.199	0.185	0.195	0.168	0.230	0.214	0.182	0.078	0.164
CFP15	0.054		0.152	0.292	0.325	0.235	0.097	0.235	0.175	0.211	0.156	0.229	0.201
CFP34	0.030	0.041		0.126	0.156	0.201	0.145	0.193	0.235	0.186	0.250	0.101	0.159
NPC1	0.057	0.067	0.061		0.072	0.177	0.268	0.422	0.357	0.288	0.392	0.165	0.271
NPC2	0.073	0.119	0.080	**0.032**		0.186	0.270	0.384	0.395	0.366	0.359	0.150	0.218
NPC3	0.086	0.098	0.087	0.070	0.093		0.272	0.398	0.320	0.331	0.228	0.184	0.239
SPC1	0.061	0.041	0.051	0.071	0.123	0.113		0.095	0.092	0.091	0.179	0.199	0.187
SPC2	0.057	0.110	0.084	0.146	0.166	0.175	0.046		0.074	0.144	0.200	0.179	0.179
SPC3	0.089	0.093	0.109	0.116	0.178	0.148	0.047	0.060		0.084	0.227	0.297	0.263
SPC4	0.087	0.095	0.094	0.100	0.176	0.150	**0.045**	0.103	0.058		0.308	0.299	0.278
WPC1	0.049	0.063	0.065	0.090	0.112	0.091	0.068	0.073	0.096	0.120		0.227	0.204
WPC2	0.037	0.082	0.055	0.071	0.074	0.097	0.076	0.076	0.138	0.146	0.071		0.078
WPC3	0.057	0.074	0.060	0.077	0.082	0.095	0.081	0.101	0.136	0.147	0.064	0.027	

Table 4.3 Analyses of molecular variance (**AMOVA**) for *Grevillea thelemanniana* ssp. Cooljarloo populations at various hierarchical levels. Clusters correspond to the unique genetic clustering from TESS spatial analyses.

Source of variation	*d.f*	Sum of squares	Percentage of variation	*P*
Among clusters	2	84.8	4.1	<0.001
among populations within clusters	13	124.8	4.52	<0.001
among individuals within populations	349	1072.4	4.58	<0.001
within individuals	365	1014.5	86.8	<0.001
Upstream vs downstream	1	33.4	2.0	0.02
within streams	256	2072.5	98	<0.001

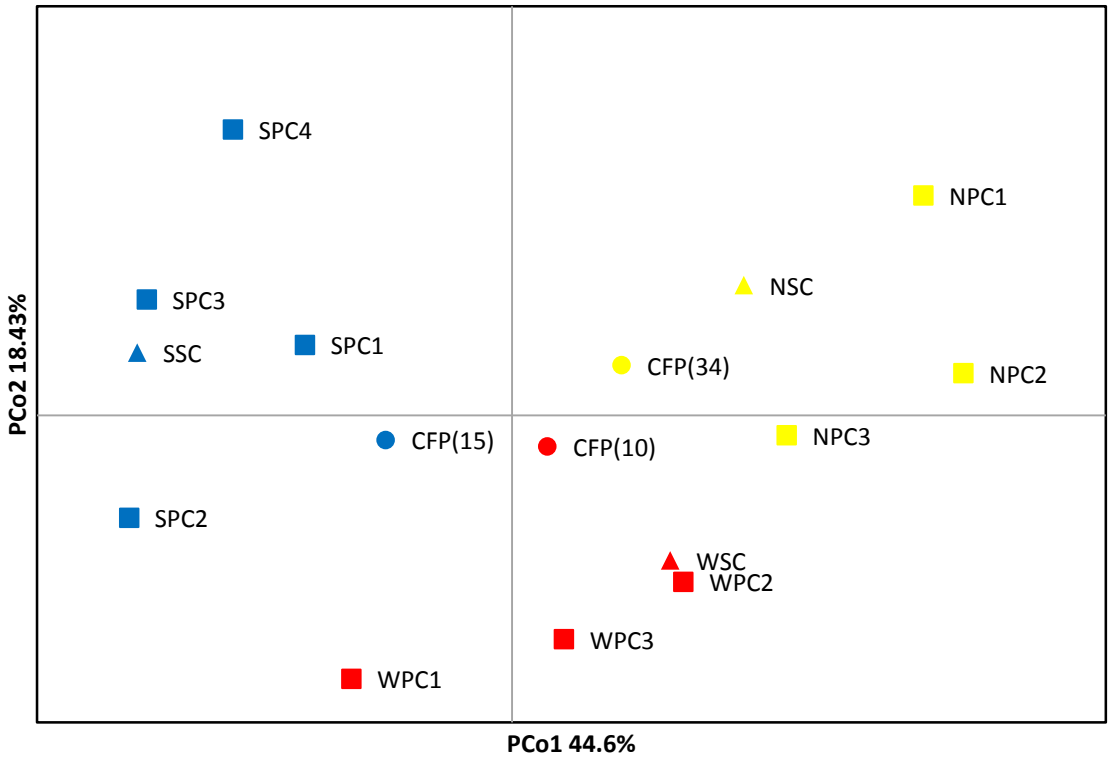

Fig.4.3 A Principal coordinates (PCoA) plot of the first two components calculated using Nei's *D* genetic distance measure between creek line populations, floodplain populations and spatial collections of *Grevillea thelemanniana* ssp. Cooljarloo. Squares, circles and triangles refer to populations, floodplain populations and spatial collections, respectively. Colours correspond to TESS inferred optimal *K* clusters, and labels correspond to populations in Table 4.1.

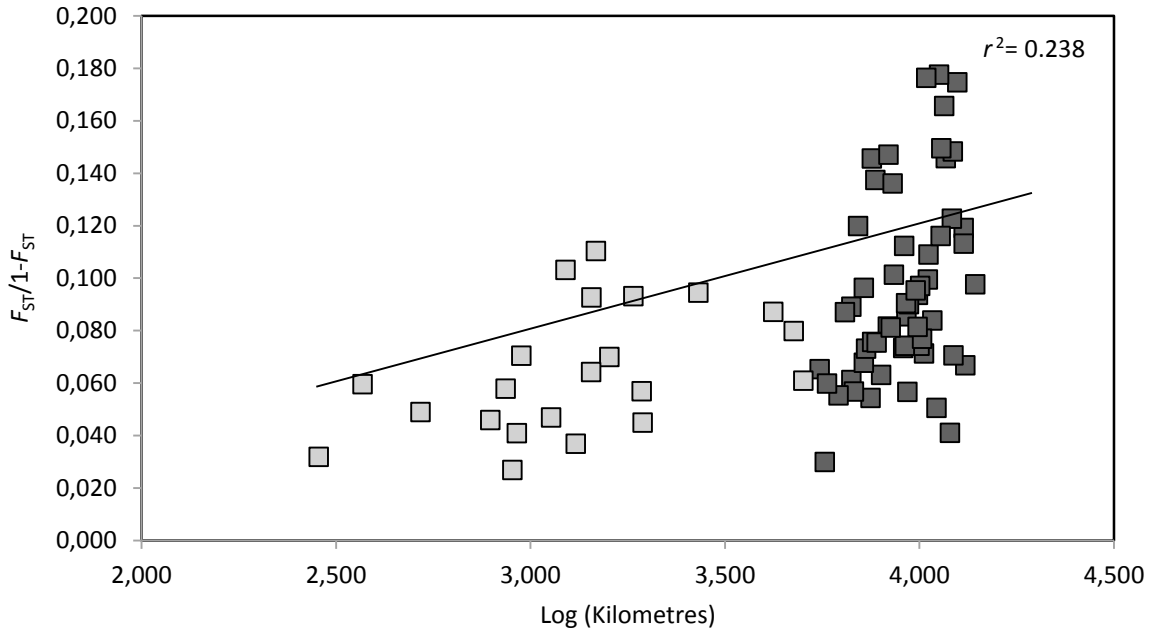

Fig.4.4 Scatter plot of pair wise population genetic differentiation (F_{ST}/(1- F_{ST}) against log geographic distance among populations of *Grevillea thelemanniana* ssp. Cooljarloo. Comparisons between populations within creek lines are illustrated by light grey squares and comparisons between populations among creek lines are illustrated by dark grey squares.

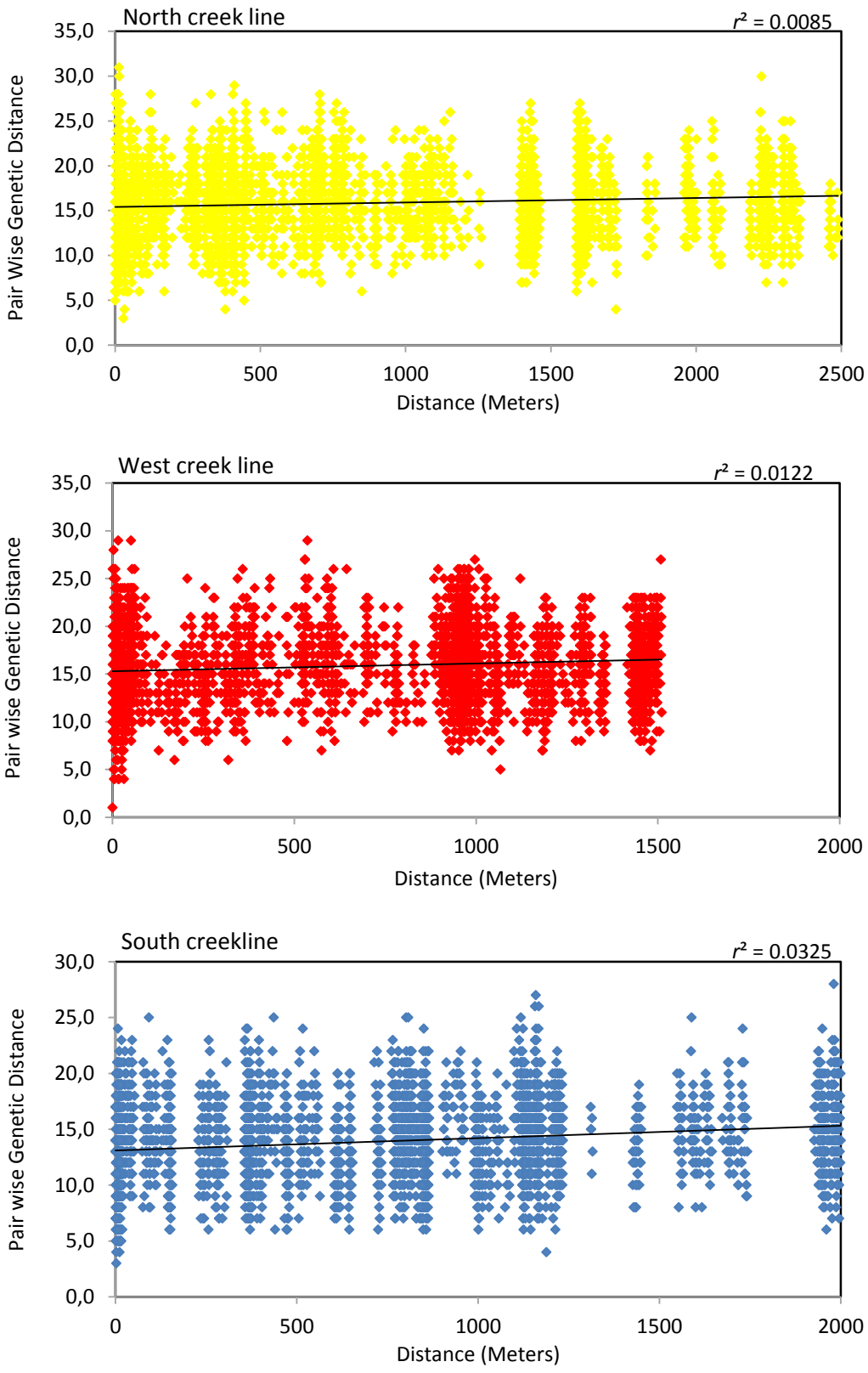

Fig.4.5 Pair wise genetic difference against geographic distance between individuals of *Grevillea thelemanniana* ssp. Coolajrloo within each of three creek lines.

Individual based structure

There was no spatial structure identified within creek lines from the individual based Bayesian analyses but there was evidence for significant structure between the creek lines. The plot of Δ*K* against *K* values from STRUCTURE indicated two genetic clusters were optimal (Fig.4.6a, Fig.4.7a), splitting the populations into a north-south cline. The west and floodplain populations were not resolved into separate clusters, but rather an admixture of the north and south populations' allele frequencies. The peak value for Δ*K* = 2 at 193.00 with a steep drop to *K* = 3 at 69.0 (Fig.4.6b).

The distribution of spatial clusters across the creek lines and floodplains are illustrated in Figure 4.8, as assigned by TESS analyses, whereas the individual membership proportions are illustrated in Figure 4.6c. The deviance information criteria values (DIC) for different numbers of clusters using TESS analysis indicated the optimal number of clusters as three (Fig.4.7b). When using admixture analyses the DIC values for three clusters dropped considerably (noadmixture lowest run = 26,979, admixture lowest run = 20,115) suggesting that the admixture model had a better fit (Row et al 2010). Ball et al (2010) suggested evaluating posterior probabilities when using spatial components in assignment testing, and to base success on ability of the programs to assign number of individuals to a cluster with membership of above 0.8. Under this criterion, TESS assigned 80% of individuals to a cluster with q_i > 0.8 (Table 4.4), outperforming STRUCTURE, which assigned 57% at *K* = 2 and 47 % at *K* = 3 of individuals with q_i > 0.8. Individuals from the southern populations had the greatest mean proportion of membership (q_i = 0.98), secondly the northern populations (q_i = 0.97) and thirdly the western populations (q_i = 0.78) with a degree of association with the southern populations (q_i = 0.19). Two of the floodplain populations were consistently shown to exhibit high levels of admixture in the Bayesian analyses, with low assignment of individuals with > 0.8 posterior probabilities to any cluster. CFP (10) was mostly an admixture of the southern (q_i = 0.42) and western populations (q_i = 0.48) and CFP (34) had a large proportion of membership to the northern populations (q_i = 0.61) with a smaller proportion to the southern (q_i =0.26) and western populations (q_i = 0.14). The exception was CFP (15), the adjacent floodplain to the southern creek line in the TESS analyses, where individuals were assigned with high probability to the south cluster (q_i = 0.99). Genetic clusters identified by STRUCTURE

and TESS were supported by a high similarity coefficient (*h'* > 0.9). Given the incongruence between STRUCTURE and TESS results the posterior probabilities for *K* = 3 are also shown, which parallel the TESS results albeit with a greater degree of admixture among clusters (average q_i = 0.68).

The first-generation migrant test using GeneClass2 assigned 12 individuals to a source other than that from which they were sampled from with δ>1.0. Of the 12 individuals, 8 could not be confidently assigned to any population, including that from which they were sampled, indicating an origin beyond the sampled populations. Two immigrants were detected in the CFP (10) flood plain originating from the western creek line, and conversely one immigrant was detected in the western creek line originating from the CFP (10) flood plain. One immigrant was detected in the CFP (15) flood plain originating from the southern creek line (Fig. 4.9).

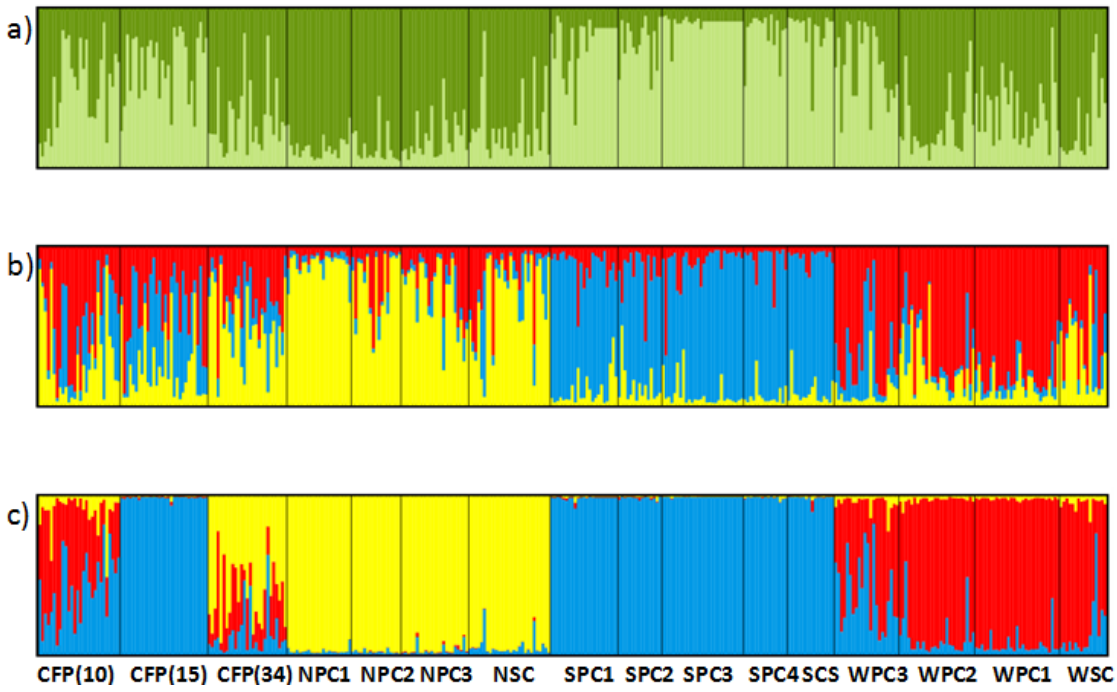

Fig.4.6 Bar plots representing admixture coefficients for *G. thelemanniana* ssp. Cooljarloo populations from assignment tests performed (a) in STRUCTURE with *K* = 2; and (b) in STRUCTURE *K* = 3; and c) with spatial data in TESS with *K* = 3. Each vertical bar represents an individual, and bars are divided into proportions based upon the probability of assignment to each population.

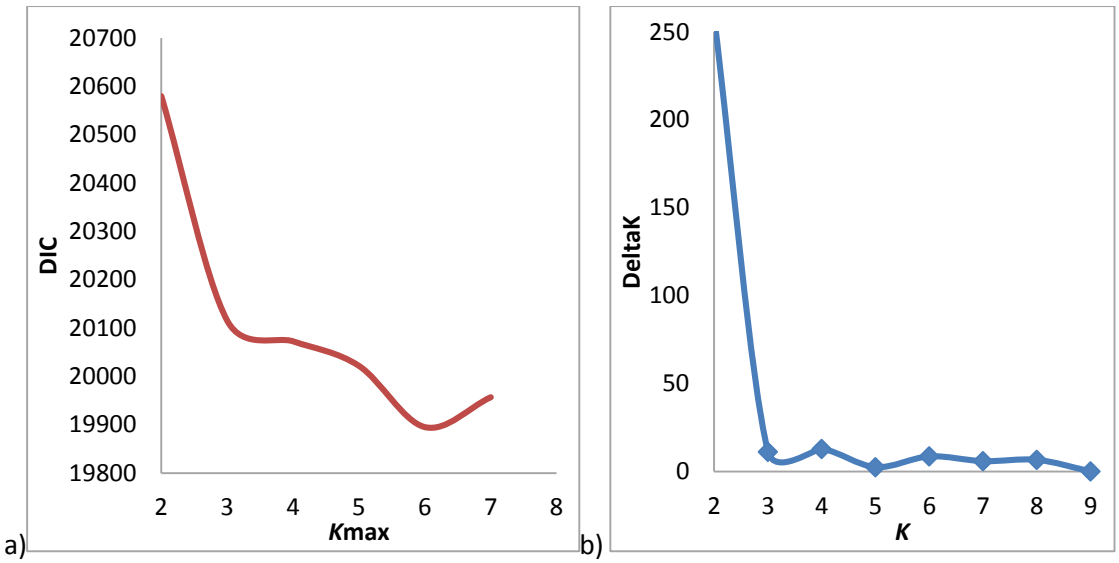

Fig.4.7 Likely number of genetic clusters in *Grevillea telemannaniana* ssp. Cooljarloo dataset from spatial analyses programs based on a) Delta *K* in STRUCTURE, *K* = 2 and b) Deviance Information Criterion (DIC) in TESS, *K* = 3.

Table 4.4 Comparison of assigned (>0.8) and unassigned (<0.8) individuals to inferred genetic clusters in *Grevillea thelemanniana* ssp. Cooljarloo dataset, from Bayesian cluster analyses using non-spatial (STUCTURE) and spatial programs (TESS). Numbers indicate different clusters identiefied and U indicates unassigned individuals. Total is number of assigned individuals with greater than 0.8 posterior probability, in brackets as a percentage of total number of individuals in dataset.

	STRUCTURE (*K*=2)			STRUCTURE (*K*=3)				TESS (*K*=3)			
	1	2	U	1	2	3	U	1	2	3	U
CFP(10)	6	5	17	1	2	5	20	1	4	0	23
CFP(15)	10	0	20	2	0	1	27	30	0	0	0
CFP(34)	2	8	17	0	2	1	24	0	0	7	20
North	1	58	31	0	52	0	37	0	0	89	1
South	74	0	23	65	0	0	32	97	0	0	0
West	10	34	49	2	0	41	50	0	64	0	29
Total		208 (57%)			174 (47%)					292 (80%)	

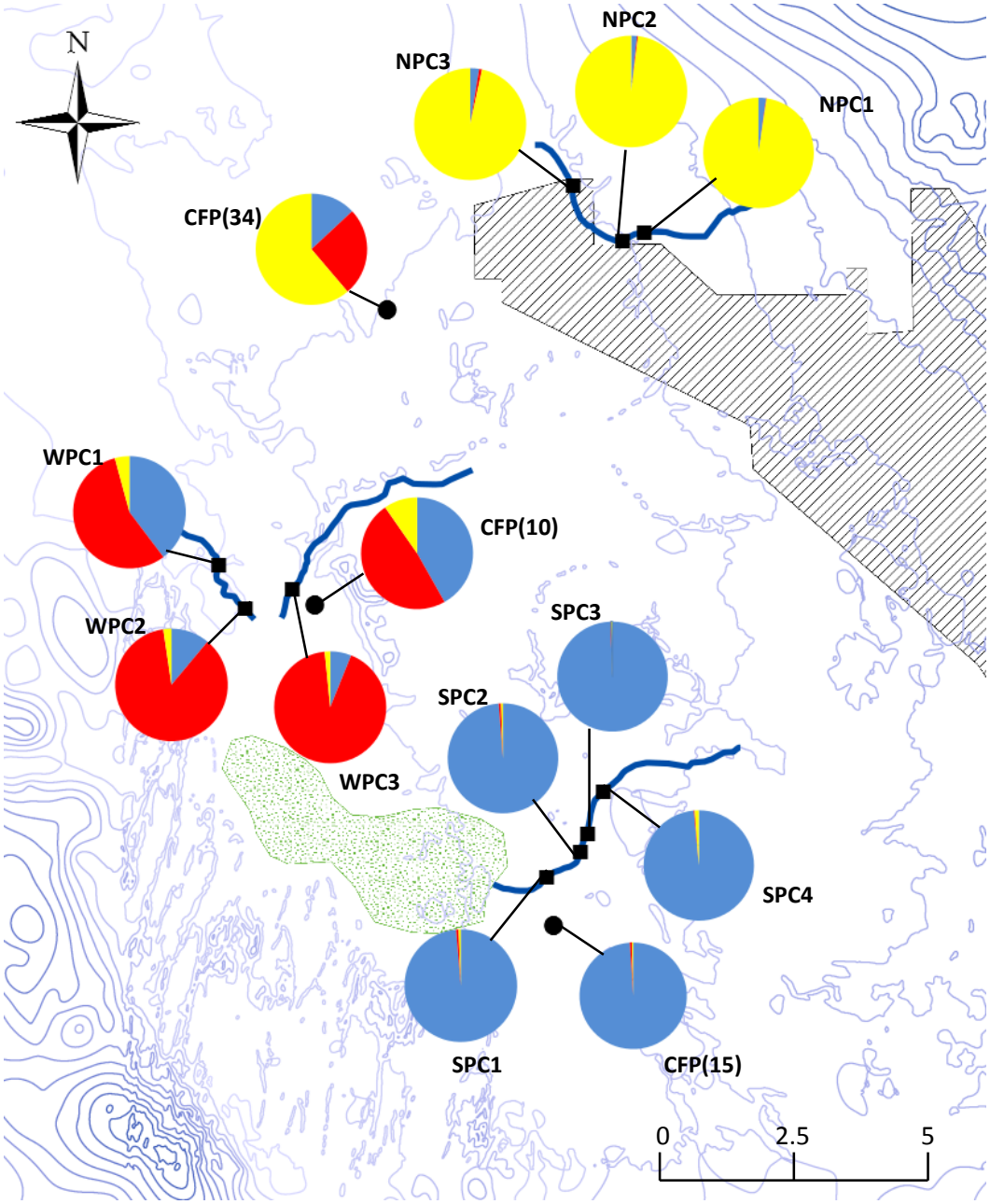

Fig.4.8 Map of the locations of sampled populations of *G. thelemanniana* ssp. Cooljarloo from creeks (black squares) and floodplains (black circles). Pie charts indicate the proportion of assignments of sampled populations to each of three genetic clusters, calculated as the mean proportion of membership values for individuals within that population provided by the TESS analysis. Colours correspond to bar plot in Figure 4.7C. Creek lines are drawn onto map. Contour lines represent elevation, ranging from low (light blue) to high (dark blue) elevation (40-190 meters above sea level). The green stippled area illustrates Cooljarloo Swamp.

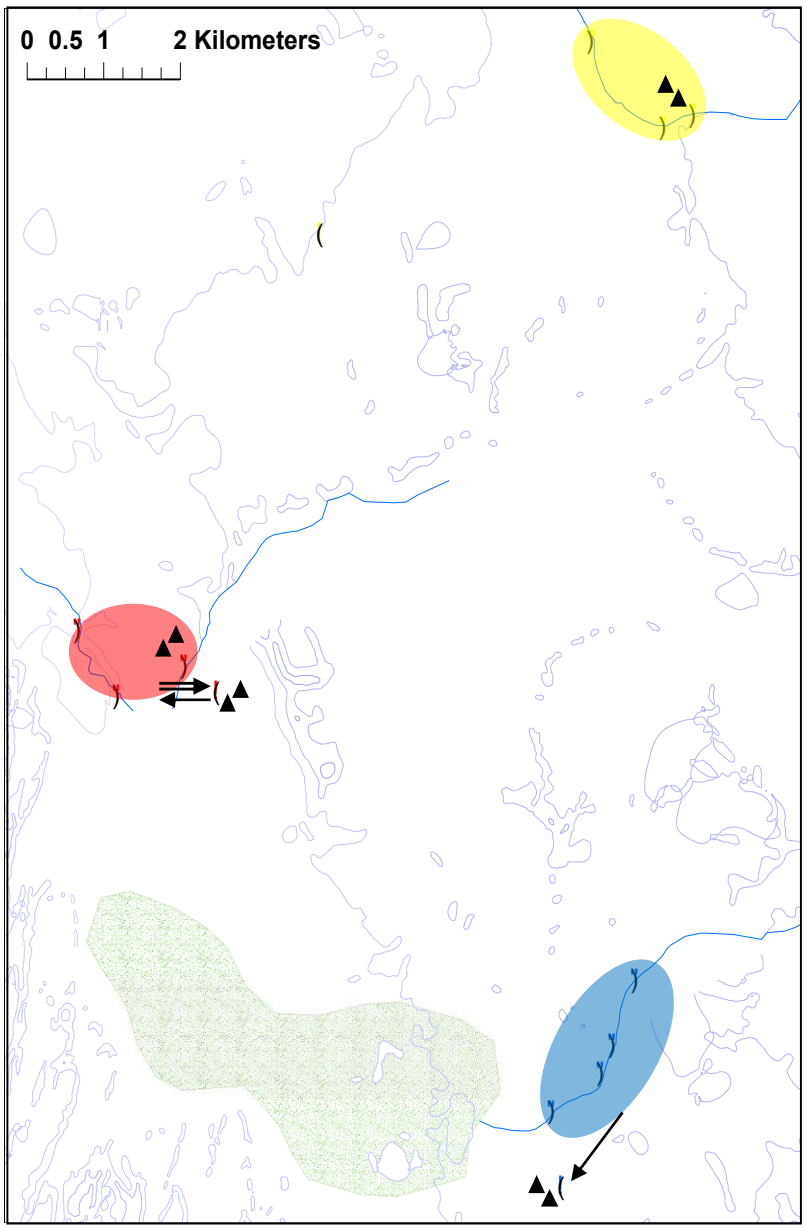

Fig.4.9 Source and sink populations of interpopulation dispersal events inferred from population genetic assignment tests for *G. thelemanniana* ssp. Cooljarloo. Populations within creek lines were treated as a single metapopulation due to low genetic structuring. Each arrow indicates the direction of source and sink of an inferred dispersal event. Single line indicates one migration event and double line indicates two migration events. Triangles indicate genotype sourced from an un-sampled population. Colours correspond to inferred *K* from TESS as in Figure 4.1. Contour lines represent elevation, ranging from low (light blue) to high (dark blue) elevation (40-60 meters above sea level).

Discussion

Spatial genetic analyses revealed strong support for three genetic clusters, corresponding to regional distribution of *G. thelemanniana* ssp. Cooljarloo populations in the northern, southern and western creek lines and flood plains. These results indicate that creek lines in this region are acting as corridors for greater gene flow between *G. thelemanniana* ssp. Cooljarloo individuals and that distance in between creeks is the most likely barrier to gene flow, rather than impermeable to genetic exchange due to a hostile landscape. Figures 4.4 and 4.6 both show that a significant relationship is present between distance and genetic relatedness. The distribution of the species indicates a relationship with lower habitats which are characterised by flood plains and creek lines. This relationship is likely due to the hypothesised dispersal mechanisms of hydrochory in these habitats, and/or greater soil fertility.

Genetic structure and gene flow among populations

This landscape genetic structure revealed by the spatial genetic analyses is supported by the PCoA and high genetic differentiation (mean F_{ST} = 0.110), implies long-term cumulative effects of restricted gene flow among creek line populations. Significant isolation by distance patterns over all populations indicated an association between genetic and geographic distance. However, the magnitude of genetic differentiation among populations within creeks (mean F_{ST} = 0.064) was significantly lower than that among creeks (mean F_{ST} = 0.110), which indicates that gene flow barriers between creek lines inhibiting gene flow relative to within creek lines. Barriers causing differentiation between creek lines are most likely due to distance in this case, rather than due to major landscape features. The TESS spatial analysis supports a relationship between distance and genetic differentiation by strong genetic clustering of the northern and southern creek line populations which are separated by >10 kilometres, whereas the western creek line populations illustrate greater admixture likely due to closer proximity (<10 kilometres) to northern and southern regions. Furthermore, the level of genetic admixture witnessed in northern (CFP 34) and western (CFP 10) floodplain populations is proportional to distance from the creek line populations. These results suggest corridors for dispersal among riparian populations within creek lines, and relative barriers for dispersal among creek lines, largely as a consequence of

distance. In a similar result, proximity of populations as a determining factor in genetic identity has been found at small and large scale riparian distributions of *Eucalyptus camaldulensis* in northern Australia, where genetic relatedness was closest between neighbouring populations, irrespective of river system (Butcher *et al.* 2002).

Genetic diversity and structure within creek lines

Weak IBD across individuals of *G. thelemanniana* ssp. Coolaroo within the three creek lines indicate connectivity within riparian habitats. A similar result was found for riparian plants *Silene tatarica* and *Euptelea pleiospermum*, leading Tero *et al.* (2003) and Wei *et al.* (2013) to reject the "classic stepping stone" and "regional ensemble" models of dispersal. These results, in combination with the Bayesian clustering would indicate that creek lines fit the "spatially extended populations" model, where *G. thelemanniana* ssp. Cooljarloo individuals within a creek line habitat form a single genetically uniform panmictic unit with freely moving gene flow. The relatively even distribution of alleles, allelic richness and moderately high heterozygosity suggest that there is no source population feeding into smaller sink populations, nor were any of these parameters significantly different between populations. Furthermore, there is no accumulation of genetic diversity in downstream populations, indicating that gene flow within creek lines is not unidirectional (Tero *et al.* 2003). Ultimately, these results support previous observations that modelling riparian systems are difficult as consistent patterns are rare (Honnay *et al.* 2010). The patterns of gene flow and structure within creek lines indicates that dispersal in *G. thelemanniana* ssp. Cooljarloo is best explained by the "classic metapopulation" model (Tero *et al.* 2003; Prentis & Mather 2008; Honnay *et al.* 2010; Wei *et al.* 2013). The distribution of *G. thelemanniana* ssp. Cooljarloo metapopulations within creek lines reflects physical separation between groups of plants, where within and among population processes such as pollen and seed dispersal can be distinguished (He *et al.* 2004).

Genetic differentiation among populations within creek lines (mean F_{ST}=0.062) was similar to that of other riparian plant populations across similar distances. In the riparian plant *Sisymbrium austriacum* along the Meuse river, the mean F_{ST} values was 0.086 (Jacquemyn *et al.* 2006), with significant genetic structure between populations separated by less than four kilometres. Similarly Prentis *et al* (2004) found highly

significant differentiation (AFLP markers; ɵ = 0.32) between populations of the lily *Helmholtzia glaberrima* along a three kilometre stretch of creek line and attributed it to genetic drift due to low dispersal. Given that *G. thelemanniana* ssp. Cooljarloo seeds do not have any obvious adaptations for dispersal, restricted interpopulation seed dispersal is likely to be an important driver in differences among creek line populations. Gornall *et al.* (1998) and Prentis and Mather (2008) inferred founder effects to explain genetic differentiation between populations within the same streams, which was likely caused by intermittent flooding causing isolated populations to reconnect. Kitamoto *et al.* (2005) reported a similar pattern of greater differentiation among streamside populations than within streamside populations of *Primula sieboldii,* which was attributed to occasional seed flow along streams due to flooding events. *Grevillea thelemanniana* ssp *Cooljarloo* typically releases seeds mid-to-late spring. In the southwest of Western Australia, high variability of riparian flows in summer is typical (Pettit *et al.* 2001), therefore occasional seed movement during this period is possible. High intensity and short duration storms during summer, derived from tropical airflows, can lead to episodic recruitment events as floodwaters which disperse seed, provide moist seedbeds and create gaps in the vegetation (Pettit *et al.* 2001).

If hydrochory were the predominant mode of gene flow within creek lines, the downstream populations should contain higher genetic diversity than upstream populations due to unidirectional gene flow (Nilsson *et al.* 2010). However, there was no evidence for downstream accumulation of diversity in any of the creek lines. Therefore, gene flow in *G. thelemanniana* ssp. Cooljarloo creek line populations is likely being mediated by animal pollinators which may lead to bidirectional dispersal, counteracting downstream accumulation of genetic diversity (Tero *et al.* 2003; Markwith & Scanlon 2007; Hu *et al.* 2008). Palmer and Bennett (2006) showed that bird assemblages in riparian habitats in south-east Australia were significantly greater in richness, diversity and abundance than adjacent non riparian sites. Their results included numerous honeyeaters, including Western Australian species, which are a main pollinator of *Grevillea* (Ford *et al.* 1979; Celebrezze & Paton 2004). The southwest of Western Australia is known for extensive vertebrate pollination, representing approximately 15% of the flora (Phillips *et al.* 2010). Furthermore, studies in the region

have revealed a trend towards bird pollination leading to greater pollen movement (Krauss *et al.* 2009; Phillips *et al.* 2010). For example, long distance dispersal of pollen has been recorded across distances of 1.5 kilometres to 5 kilometres (Byrne *et al.* 2007; Millar *et al.* 2008). The relationship between pollen and seed dispersal within *G. thelemanniana* ssp. Cooljarloo creek line populations maintains significant genetic differentiation among distant populations but counteracts strong isolation by distance.

Genetic diversity and structure within flood plains

Unlike the creek line populations, the admixture evident in floodplain populations, as well as paths of first generation migrants, indicates a directional pattern of dispersal to these habitats (Fig. 1, 2 & 6). This pattern is indicative of a source-sink relationship between the creek line populations and flood plain populations. Sink populations are however predicted to have lower genetic diversity than source populations (Tero *et al.* 2003), and this is not evident in the flood plain populations which illustrate private alleles and high levels of heterozygosity and allelic richness. Gene flow to flood plains is likely induced by occasional flooding of creek lines and maintained by subsequent animal pollination. He *et al* (2004) hypothesised that due to the diploid nature of seeds, a long distance dispersal event between genetically differentiated populations will result in the individual being allocated to the source and not as an intermediate. Based on this, the identification of migrants (Fig. 6) therefore shows evidence for gene flow through seed dispersal into the southern and western flood plains which is likely driven by flooding. Flooding has been shown to effectively disperse seeds in numerous riparian studies (Gornall *et al.* 1998; Kitamoto *et al.* 2005; Kudoh *et al.* 2006; Van Looy *et al.* 2009). He *et al.* (2004) also hypothesised that given the haploid composition of pollen, a long distance dispersal event between genetically differentiated populations will result in a genotype which is an intermediate between parental genotypes. This is demonstrated in the genetic admixture from Bayesian analyses in the northern and western flood plain populations, where all the individuals are showing extensive admixture (Fig. 1 & 2). Conversely, the southern flood plain did not show any admixture which is likely due to the distance separating it from the northern and western populations being too great.

Landscape factors influencing distribution

The relationship between distribution of *G. thelemanniana* ssp. Cooljarloo populations and elevation indicates that landscape features are affecting spatial genetic structure, however this is confounded by the effect distance is also having on structure. *Grevillea thelemanniana* ssp. Cooljarloo populations seem to berestricted to areas of low elevation (40-80 meters) and are absent at higher elevations (>80 meters). The elevation data suggests an ancient drainage system in the region, and a potentially wider historical distribution for the riparian *G. thelemanniana* ssp. Cooljarloo. Elevational changes do correspond to the habitat changing from creeks and wetlands in low elevation to drier *Eucalyptus* dominated habitats in higher elevations (personal observation). Prentis and Mather (2008) found that size and distribution of stream side populations of *Helmholtzia glaberrima*, was negatively correlated with decreasing altitude. The correlation was due to a preference for higher rain fall areas at higher altitudes as water logged habitats was needed for seed germination (Prentis & Mather 2008). A more detailed study of seed dormancy and germination trials is required to determine which properties of these riparian habitats is limiting the current distribution of *G. thelemanniana* ssp. Cooljarloo.

Conclusions

Ultimately these results indicate that creek lines in this region are acting as corridors for greater gene flow between *G. thelemanniana* ssp. Cooljarloo individuals, and that distance among creeks is the most likely barrier to gene flow, rather than a hostile landscape matrix preventing movement by bird pollinators. The distribution of the species indicates a relationship with lower habitats which are characterised by flood plains and creek lines. This relationship is likely due to the dispersal mechanisms by hydrochory in these habitats, and/or greater soil fertility, which would require further studies. Connectivity relating to hydrochory, especially flood events, is important in regards to current climate change, as the average rainfall in the southwest of Western Australia has been decreasing over the last 50 years (Bates *et al.* 2008). This also suggests the possibility of a contracting distribution for the riparian *G. thelemanniana* ssp. Cooljarloo, from what was historically a potentially much larger distribution than at present.

Chapter 5 - Microsatellites for *Grevillea thelemanniana* ssp. *thelemanniana* and cross-species amplification in the *Grevillea* Thelemanniana group (Proteaceae).

This chapter has been slightly modified from a paper that was published in 2013 by Conservation Genetics Resources.

Hevroy TH, Moody ML, Krauss SL, Gardner MG (2013) Isolation, via 454 sequencing, characterization and transferability of microsatellites for Grevillea thelemanniana subsp. thelemanniana and cross-species amplification in the Grevillea thelemanniana complex (Proteaceae). Conservation Genetics Resources **5**, 887-890.

Introduction

Grevillea thelemanniana ssp*. thelemanniana* is listed as Priority 4 (Smith 2013) and is found only in a fragmented habitat southeast of Perth, Western Australia. The species is part of a larger group comprised of 16 species and 4 subspecies, endemic to Western Australia and found primarily in the coastal sand plains from as far south as Busselton and north to Exmouth. The *Grevillea* Thelemanniana Group has been recognized as a group with high conservation importance and poorly known taxonomic relationships (Makinson 2000). Here we develop 12 microsatellite markers for *Grevillea thelemanniana* ssp. *thelemanniana* that show amplification and polymorphism among other members of the complex. Relatively few microsatellite panels exist for *Grevillea*, and of these all are for eastern Australian species, making this the first library for a Western Australian *Grevillea* (England *et al.* 1999; Hoebee 2011; Mantello *et al.* 2011; James *et al.* 2012).

Materials and Methods

Genomic DNA was extracted from one individual of *G. thelemanniana* ssp. *thelemanniana* collected at Kenwick Nature Reserve. Extraction from freeze-dried, ground leaf material followed the modified CTAB extraction method (Doyle 1987). The DNA solution (4µg) was sent to the Ramaciotti Centre for Gene Function Analysis (http://www.ramaciotti.unsw.edu.au) in Sydney for 454 shotgun sequencing on a Roche GS-FLX Titanium machine (454 Life Sciences, Branford, CT, USA) following the protocol of Gardner *et al.* (2011). Sequences have been deposited in the Dryad Digital

Repository (see Meglécz *et al.* (2012); http://dx.doi.org/10.5061/dryad.jd183) The resulting raw sequences (total length 101,949,487 bases) were searched for unique microsatellite motifs two to six base pairs long with C8 repeats (Gardner *et al.* 2011) in QDD 1.3 software (Meglécz *et al.* 2010). The program Primer3 1.1.1 (Rozen & Skaletsky 1999) within QDD 1.3 was used to design primers, using the basic search defaults. The resulting 501 unique loci were sorted based on PCR product size, repeat class, repeat length, GC content and multiplexing potential. The best 30 (15 dinucleotide and 15 trinucleotide) were chosen for evaluation of amplification and polymorphism. Reactions were performed with 10-50 ng of *G. thelemanniana* ssp. *thelemanniana* DNA in 10 µl reaction volumes containing 50 mM KCl, 20 mM Tris-HCl (pH 8.4), 0.2 mM each dNTP, 0.08 µM forward primer and reverse primer, 0.01 µl *Taq* DNA polymerase and 4 mM $MgCl_2$. Amplification was conducted in Eppendorf™ thermocyclers the following amplification conditions; denaturation for 3 min at 94°C was followed by 30cycles (25 in loci THP17) made up of 40s at the annealing temperature (56-62°C) and 30s at 72°C. The final step was elongation of 15min at 72°C. Of the 30 loci, 13 were selected for further testing and were amplified in two individuals from 3 populations of *G. thelemanniana* ssp. *thelemanniana*. An applied Biosystems DNA Analyzer (Beckman) was used to visualize fragments and genotypes were scored manually using Beckman Fragment Analyzer software. One primer was monomorphic and therefore excluded from further testing.

Genetic diversity of the remaining 12 loci was assessed for 25 individuals from the Kenwick Nature Reserve population (32° 0' 59.6" S 115° 59' 24.3" E (GDA94)). PCR amplification and genotyping was conducted as above. Hardy-Weinberg equilibrium (HWE), linkage disequilibrium (LD) and allelic richness were assessed using exact tests in Genepop 4.1.4 (Raymond & Rousset 1995; Rousset 2008) with Bonferroni correction. GenAlEx 6.5 (Peakall & Smouse 2012) was used to obtain genetic diversity parameters. We used MICROCHECKER 2.2.3 (Van Oosterhout *et al.* 2004) to check each locus for evidence of null alleles, scoring error due to stuttering, and large allele drop out. Primers for all loci were tested for amplification in 8 individuals each of all species of the *Grevillea* Thelemanniana group.

Results and Discussion

There was no evidence for linkage disequilibrium at any locus, however deviation from Hardy Weinberg equilibrium was found for three loci (*P<0.01;* THP4 and THP30, *P<0.05*; THP9; Table 5.1) with heterozgosity deficiency evident. All loci were independent and null alleles were only apparent in THP4 with an estimated frequency of 0.145. All loci showed polymorphism within the population. The number of alleles per locus ranged from two to eight, and effective number of alleles per locus ranged from 1.2 to 3.7. Observed heterozygosity ranged from 0.120 to 0.660 and expected heterozygosity from 0.18 to 0.73. The fixation index ranged from -0.143 to 0.343 for loci in HWE. Cross species amplification to the Thelemanniana group was successful for all loci (12) in 7 species and partially successful in the remaining 5 species (Table 5.2). Of the *Grevillea* Thelemanniana group, allele numbers ranged from one to seven. The microsatellite markers developed here will be used to evaluate taxonomic hypotheses in the *Grevillea* Thelemanniana group and evaluate patterns of landscape genetic structure for populations of *G*. *thelemanniana* ssp. Cooljarloo (B.J. Keighery 28 B) in Western Australia.

Table 5.1 Primer sequences and characteristics of 12 microsatellite loci isolated from *G. thelemannaiana* ssp *thelemanniana*

Locus	Primer Sequence (5'-3')	GenBank no.	Repeat Motif	Temp (C°)	Size range (bp)	N	Ne	*Ho*	*He*	F	PHW	Null allele
THP3	F:AGCTAGCCGACCCCATTTAG R:CAACTTGAATCAGACAAAGATGTG	GLMI75201DP2HD	(GTT)7	56	91-103	3	1.326	0.2	0.246	0.186	0.287	
THP4	F:TCAACATGGGCGTTTAATGA R:GGTCGTGGTACTTGGTGGTC	GLMI75201BFWLD	(GA)8	62	174-200	7	3.165	0.44	0.684	0.357	0.0*	0.145
THP5	F:GAAACCGCGTGCAATAAAAT R:TGGAAATCGATTCGAAGAGG	GLMI75201C8T3I	(TC)8	62	170-172	2	1.625	0.44	0.385	-0.143	0.88	
THP7	F:TCAGAATCGTTTTCTGCGTG R:TACCCAGTTTCTTGGCCTTG	GLMI75202IDULT	(TCT)8	56	100-106	2	1.22	0.12	0.18	0.327	0.198	
THP8	F:CAACAAGCCAACCTTCTTGC R:AGAAGCACCCATTGATTTGG	GLMI75202G91AB	(AAC)8	62	149-158	3	2.163	0.6	0.538	-0.116	0.772	
THP9	F:ACCCCTACCAATAACCACCC R:TTTCCTGGACTCTCCGAATG	GLMI75201CCFRY	(GT)8	62	310-314	3	1.874	0.36	0.466	0.343	0.017*	
THP13	F:GCAATTTCCCACTGTTTCCA R:CCAGAGTGTGAACCTCAGCA	GLMI75201CRJRK	(TA)10	62	98-106	8	3.704	0.667	0.73	0.087	0.209	
THP17	F:GTTCTGATGGTTGGATGCCT R:CATGTTCAAATTCCGTCGAA	GLMI75202IBW27	(CT)12	62	172-208	7	2.723	0.56	0.633	0.115	0.168	
THP20	F:CAGTTCACCCCTGCAAGTTC R:CCTCCTTGGCTGCAATACTC	GLMI75201AV2UW	(TTC)15	62	130-154	7	2.404	0.56	0.584	0.041	0.058	
THP22	F:AGAATCCATGCAGTCAACCC R:CATCTGCTGGTTCGTTTGTG	GLMI75202IBW27	(TC)6	62	154-160	2	1.419	0.28	0.295	0.051	0.583	
THP24	F:CAACAAGCCAACCTTCTTGC R:AGAAGCACCCATTGATTTGG	GLMI75202G91AB	(AG)8	62	152-158	3	2.08	0.56	0.519	-0.079	0.692	
THP30	F:TGTGAAATGAGCGAACTCCA R:GCCTTGTGTACTCACCACAGC	GLMI75202GYCOA	(CT)9	62	116-124	3	1.442	0.28	0.306	0.086	0.021*	

Allele size range, number of alleles (N), number of effective alleles (Ne), observed and expected heterozygosity (Ho, He), fixation index (F), Hardy Weinberg P values (PHW) and null allele frequencies are based on 25 individuals.

* (*P<0.05*) **(*P<0.01*) Departure from Hardy-Weinberg equilibrium

Table 5.2 Cross species amplification in the Thelemanniana group. PCR product size and observed number of alleles (in parenthesis) of 12 microsatellite loci. An X means no amplification was detected.

Locus	THP1	THP3	THP4	THP5	THP7	THP8	THP9	THP13	THP17	THP20	THP22	THP30
G .acropogon	154-166(2)	X	188-196(5)	149(1)	100-115(3)	149-158(3)	318(1)	83-97(3)	183-192(4)	98(1)	153-156(2)	122-124(2)
G .delta	171-174(2)	100-115(5)	182-184(2)	167-171(3)	109-133(7)	153-158(3)	310-314(2)	90-94(3)	172(1)	127-148(6)	154-160(3)	110-136(5)
G .evanescens	166(1)	97-100(2)	174-186(4)	169(1)	106-109(2)	155-158(2)	314(1)	96-106(5)	172-187(2)	130-151(4)	148-163(5)	116-130(4)
G .exposita	162-171(4)	97-100(2)	184-188(3)	165-169(3)	94-112(4)	155-158(2)	312(1)	90(1)	171-204(2)	145-160(4)	151-157(2)	116-132(3)
G .fililoba	100-109(5)	100-109(3)	180-190(3)	167-171(3)	100-115(5)	149-158(3)	310-312(2)	90-104(3)	172-190(5)	127-160(6)	151-160(4)	116-122(2)
G .sp. Gillingara	171(1)	100-106(3)	186(1)	169(1)	100-115(3)	158(1)	312(1)	84-90(2)	171-174(2)	148-151(2)	154-157(2)	132(1)
G .hirtella	171-180(2)	96-99(2)	174-188(6)	167-171(2)	94-115(4)	X	310-312(2)	90-96(2)	171-189(3)	140-158(3)	147-159(4)	116-130(6)
G .humifusa	171-174(2)	94-103(4)	180-184(3)	171(1)	106-118(3)	155-158(2)	310(1)	90(1)	172(1)	137-158(4)	153-159(3)	122(1)
G .maccutcheonii	189(1)	X	X	159-169(3)	112-115(2)	155(1)	X	90-94(2)	171-186(2)	98(1)	156(1)	118-122(2)
G .obtusifolia	166(1)	100-103(2)	176-196(5)	169-171(2)	100-109(3)	155-158(2)	314-318(3)	94-110(5)	171-204(3)	127-142(3)	154-160(3)	116-124(4)
G .olivacea	162-177(4)	100-103(2)	182-186(2)	167(1)	100(1)	158-167(3)	314(1)	90-94(3)	169-175(3)	120-149(4)	154-160(3)	116(1)
G .pinaster	151-174(6)	97-109(6)	165-169(3)	165-173(4)	97-112(5)	155(1)	310-312(2)	90-94(3)	168-192(5)	130-158(5)	148-160(5)	114-126(5)
G. preissii ssp. *preissii*	166(1)	100-115(3)	182-198(3)	167-171(3)	100(1)	155-158(2)	310-314(2)	90-98(4)	171-198(3)	130-160(8)	153(1)	116-122(4)
G. preissii ssp. *glabrilimba*	171-177(3)	100-115(3)	170-196(6)	169-175(3)	100-109(2)	155(1)	310-312(2)	90-102(4)	172(1)	137-164(8)	153(1)	118-124(4)
G. ripicola	165-168(3)	X	181-185(3)	149(1)	100-109(4)	X	317-323(2)	88-100(4)	189(1)	92-98(2)	260(1)	X
G. thelemanniana ssp. Cooljarloo	165-171(2)	97-103(2)	182-196(7)	163-171(4)	106-118(5)	155(1)	310-318(3)	90-96(4)	172(1)	143-158(7)	154-160(3)	116-124(3)
G. variifolia ssp. *variifolia*	150-177(4)	X	156-196(6)	157(1)	109-115(3)	155-158(2)	321-331(4)	90-92(2)	171(1)	120-144(4)	154-160(3)	116-122(2)
G. variifolia ssp. *bundera*	166-169(2)	88-103(5)	191-221(9)	147-159(3)	82-109(5)	177(1)	309-327(7)	88-90(2)	171(1)	123-135(5)	154-157(4)	116-122(2)
G. stenomera	160(1)	91-100(2)	175-187(4)	165-171(3)	94-100(2)	155-158(2)	310(1)	89(1)	177-198(3)	132-141(4)	147(1)	114-136(3)

Chapter 6 – General Discussion

"Evolutionary biology, in contrast with physics and chemistry, is a historical science, the evolutionist attempts to explain events and processes that have already taken place. Laws and experiments are inappropriate techniques for the explication of such events and processes. Instead one constructs a historical narrative, consisting of a tentative reconstruction of the particular scenario that led to the events one is trying to explain" - Ernst Myer

The combination of phylogenetic and population genetic methods with multi-locus data to asses relationships of taxa in an integrated overview is perhaps one of the most effective contribution to help conserve Australia flora (Hopper 2000). The *Grevillea* Thelemanniana group provided a model system for illuminating on diversification patterns in the SWAFR. The long evolutionary past of *Grevillea* in Australia, and the large range of the Thelemanniana group in Western Australia, encompassing the high rainfall province, transitional rainfall province and the arid zone meant that different hypothesis of evolution could be tested. Furthermore, this study adds knowledge to both an understudied genus and globally recognised biodiversity hotspot.

In this chapter I bring together the results from each data chapter and discuss how they link together to form a broad perspective of the major issues explored in this thesis. I will outline the major findings and their implications for species diversity and evolutionary patterns in the SWAFR. Lastly, I outline some recommendations for conservation management, including limitations of this study and areas of future research.

Summary

In this thesis I incorporated a multi-locus approach, utilising chloroplast and nuclear data, to examine the genetic relationships within a Western Australian *Grevillea* species complex at three different levels (1) analyses of the evolutionary history across a species complex in Western Australia that has undergone multiple taxonomic revisions; (2) a deeper investigation of underlying causes of incomplete lineage sorting through greater sampling of a subgroup of closely related species and (3) I used a

landscape genetic approach to resolving the spatial genetic structure of restricted populations of the threatened *Grevillea thelemanniana* ssp. Cooljarloo.

This is the first molecular phylogenetic study on *Grevillea* in Western Australia, and only the second for the entire genus, and is important for its contribution to understanding the diversification of the third largest genus of flowring plants in Australia. Specifically, the most important findings of this study are listed as follows:

(1) The data showed that the Thelemanniana group is not monophyletic, but instead is paraphyletic to a number of subgroups outlined by Makinson (2000). The chloroplast data illustrated a deep divergence within the group, which has resulted in two lineages corresponding to a northern and southern geographic distribution in Western Australia (Chapter 2). Makinson (2000) did have concerns about the placement of *G. ripicola, G. acropogon* and *G. mccutcheonii* (southern lineage) in the Thelemanniana group, which was supported by significant divergence in the molecular data. Molecular dating suggests that the two lineages diverged in the early Miocene.

(2) Both the chloroplast and nuclear data illustrated support for the monophyly of phrase named *G. thelemanniana* ssp. Cooljarloo and *G.* sp. Gillingara. The phylogenetic placement of *G. thelemanniana* ssp. Cooljarloo illustrates clearly that this species is not a subspecies of *G. thelemanniana* and warrants reclassification as a unique species. The recognition of these two species has major conservation implications, as *G. thelemanniana* ssp. Cooljarloo has a narrow distribution in the vicinity of an active mine site, and *G.* ssp. Gillingara is currently known from only one roadside population.

(3) The low levels of chloroplast sequence variation, and large amount of haplotype sharing, resulted in patterns of incomplete lineage sorting. Through analyses of haplotype networks and nuclear data of the species complex (Chapter 2) and subgroup (Chapter 3), incomplete lineage sorting was revealed to be due to recent speciation. Although *Grevillea* is an old genus, as recognised from fossil data (Dettmann & Jarzen 1998; Mast *et al.* 2008), the results in this study revealed that the many of the species within the Thelemanniana group have appeared during the late Pliocene/Pleistocene.

(4) The Bayesian analyses of microsatellite data assigned individuals to known species and illustrated very little admixture, strongly supporting the morphologically defined,

and newly proposed, taxa. These data resolved issues of incomplete lineage sorting and strengthened the case for monophyly of many of the morphologically defined species. Overall, the microsatellite data was successful at delimitating species and illustrated the power of using these markers in systematic studies of closely related species.

(5) This study was the first to investigate the spatial genetic structure of a Western Australia *Grevillea* species. This was also one of the few studies in Western Australia undertaking a landscape genetics approach to attempt to explain the drivers of spatial genetic structure of SWAFR flora. The spatial analyses revealed significant genetic structure among populations at three sampled creek lines. The results indicated that creek lines are acting as corridors for greater gene flow and dispersal, maintaining a higher level of gene flow within creek lines than between. Moreover, the level of genetic admixture and isolation by distance revealed a relationship between proximity and gene flow, where distance was likely the main driver of divergence, rather than the landscape matrix between creek lines. The genetic association of flood plain populations was confounding however, although they genetically clustered with the nearest creek line populations, the level of genetic admixture was significantly greater than within creek line populations, indicative of greater gene flow into flood plains. This result reflects a level of landscape affect on the spatial genetic structure of *G. thelemanniana* ssp. Cooljarloo populations.

(6) Microsatellite repeat motifs were detected within DNA sequences isolated from *G. thelemanniana* ssp. *Thelemanniana,* and new primers were designed for amplification of twelve polymorphic loci (Chapter 5). Successful microsatellite transfer across species of the Thelemanniana group supports a close relationship between species. The greater difficulty in amplifying these microsatellites for *G. acropogon, G. mccutcheonii* and *G. ripicola* supported the divergence of these species from the remainder of the group. The overall success of microsatellites in testing and delimitating species is encouraging to future studies of closely related species and species complexes, and supports their recent use in taxonomic studies.

Evolutionary trends in the SWAFR

The early Miocene divergence between the southern and northern lineage in the Thelemanniana followed by slow divergence up until the late Miocene/ early Pliocene corresponds to the patterns from the fossil pollen record of Proteaceae in Western Australia. Dettmann and Jarzen (1998) described a steady climb in number of taxa from the subfamily Grevilleoideae (encompassing *Grevillea*) from the early Eocene, peaking during the early Miocene and then dropping of in the period following. Cardillo and Pratt (2013) also noted slight slowdown in diversification rates after 20 Mya in *Banksia* (also of the Grevilleoideae subfamily). A slight rise in temperature has been noted in the early Miocene, accompanied by renewed formation of laterite and calcrete as well as lake sedimentation (Hopper & Gioia 2004). The resulting downturn of diversification rates around this time could indicate a period extinction as flora adapted to changing landscape and climate. The Miocene is marked as a period of slower speciation rates, after which a period of explosive speciation in the SWAFR arising in the late Tertiary and Quaternary (Hopper and Gioia 2004). Hopper and Gioia (2004) went on to propose a number of hypotheses to explain the high diversity of the flora of the SWAFR, of which the following were addressed in this study;

1) *There is a subtle trend for recently evolved species to occur more frequently in the transitional rainfall province (TRP) compared to the high rainfall province (HRP) due to the expansion and contraction of climatic fluctuations during the Pliocene/Pleistocene.*

Molecular analyses and dating revealed evidence for widespread recent allopatric speciation and formation of numerous narrowly distributed endemics in the *Grevillea* Thelemanniana group. Although it is difficult to determine if many of the species have historically naturally restricted distributions due to natural fragmentation or as a result of major land clearing throughout much of the SWAFR. Furthermore, the divergence times between shared haplotypes, and the extensive sharing of one haplotype between species, makes the dating analyses in this thesis somewhat unreliable and has been taken into consideration when coming to conclussions regarding this species complex. Regardless, recent speciation, resulting in low chloroplast differentiation and haplotype sharing between several species, was prevalent among species in the TRP,

which supports Hopper and Gioia's (2004) original theory of speciation patterns in the TRP.

2) *The high rainfall province has a more stable landscape which should have led to lower extinction rates and accumulation of relict species in the transitional rainfall province.*

There was no evidence for older relict species in the HRP compared to the TRP (Figure 2.1, Chapter 2). The older lineages found in the Thelemanniana group were present in both rainfall provenances. A similar result was found by Cardillo and Pratt (2013), who suggested that fires regimes may have been more frequent and intense in the dense sclerophyllous forest of the HRP, resulting in extinction due to population fragmentation.

3) *Rapid radiation of taxa caused by more diverse topography and climatic stresses of late Tertiary and Quaternary.*

Molecular dating revealed several lineage divergences having occurred during the mid-Miocene and early Pliocene, so although there was evidence for rapid radiation in the northern lineage, the patterns of divergence across the entire Thelemanniana group indicate a pattern of accumulation more similar to that found in *Banksia* (Cardillo & Pratt 2013).

These results illustrate that dating of phylogenies for flora of the south west of Western Australia can greatly enhance our current understandings of speciation and evolution in this region. Furthermore, even though authors often disapprove of dating phylogenies without fossil data, the congruent results between chloroplast and microsatellite data, that showed similar patterns of speciation as the molecular dating results, could be seen as support for successful application ofsecondary calibration pointsin this study. However, a statistical test of secondary calibrations points validity would be recommended for furture studies.

The landscape in which *G. thelemanniana* ssp. Cooljarloo is distributed is particularly interesting. The elevation data illustrates a deep valley system wherein the species resides (Chapter 4). Since the intensified aridification of Western Australia in the late Pliocene and Pleistocene, and the recorded drop in rainfall (Bates *et al.* 2008), many of

the wetter colder environments in this region have disappeared (Hopper 2000). It is possible that during wetter/colder times this valley had creeks with more consistent water flow which may have caused a greater gene flow among populations though hydrochory. This pattern of gene flow could explain greater admixture of flood plain populations and weak divergence among creek line populations. Butcher et al (2009) study of the riparian *Eucalyptus camaldulensis* suggested that current patterns of relatedness may reflect more continuous distributions during wetter periods, and that contraction events may have dissected species distributions, reducing gene flow and prompting genetic drift. Although this theory was applied to populations, at a larger scale it could also be applied to species and subspecies and can explain allopatric speciation of Thelemanniana group and the tendency of species and subspecies within this group to be distributed along creeks and rivers.

Integrative taxonomy

Using multi-locus data has been effective in resolving the Thelemanniana group. The results from this study support recent reviews of molecular taxonomy that argue for the shift from traditional species concepts which seek to define monophyletic species, towards utilising molecular data to describe the evolutionary history of a species, herein focusing more on species as segments of evolutionary lineages (Sites Jr & Marshall 2004; de Queiroz 2007; Shaffer & Thomson 2007; Miralles & Vences 2013). This study highlights that if speciation has taken place a sufficiently long time ago (Miocene divergence; Chapter 2), these chloroplast markers are informative for species delimitation. However, in the case of more recent divergence (Pliocene/Pleistocene; Chapter 3), microsatellite markers were more informative for species delimitation. Moreover a more comprehensive sampling in combination with microsatellite data resolved issues of incongruence initially believed to be hybridization and/or introgression among species of subclade 1 (Chapter 3). When coupled with Bayesian assignment methods, which class individuals in groups, there was no evidence for intermediates which generally indicate hybridization (Duminil & Di Michele 2009).

This study highlights the contribution of different genomes and marker types and integration of both phylogenetic and population genetic concepts to address patterns

of lineage diversification and species delimitation. The analyses support a hypothesis of incomplete lineage sorting due to recent speciation as being the cause of haplotype sharing among species in the phylogenetic analyses of chloroplast sequences. The lack of resolution provided by chloroplast data was overcome by applying population genetic concepts through coalescent analyses of microsatellite data, which supported the majority of species boundaries as defined by morphology. These results illustrate the trade-off between resolving power in taxonomy.

This study also emphasises the benefits of comparative studies (Chapter 3) in investigating the processes which have resulted in incongruence between chloroplast and nuclear data in phylogenetic analyses (Chapter 2). Indeed the corresponding patterns of divergence across different loci and multiple populations identified a pattern of allopatric speciation in subclade 1 (Chapter 1), which can also explain the haplotpye sharing of many species within the Thelemanniana group. Targeting multi-locus data has been shown to be particularly useful in describing the contribution of allopatric processes (Rossetto *et al.* 2012). Furthermore, the inclusion of landscape data and spatial analyses to traditional population genetics lead to a greater understanding of patterns of gene flow, which in turn can be used to inform management strategies (Chapter 4).

Microsatellite markers proved highly informative for resolving the taxonomy and phylogeny of *Grevillea* Thelemanniana group. Very few studies have used microsatellite markers for taxonomic objectives in the SWAFR. This study illustrates their potential, especially given the large number of closely related species complexes in the SWAFR and the often low chloroplast diversity which follows from recent speciation. The use of next-generation DNA sequencing to identify large numbers of microsatellite loci, or simple sequence repeats (SSRs), is increasingly employed by researchers in the fields of ecology, conservation biology and population genetics (Gardner *et al.* 2011). This type of microsatellite development is rapidly becoming more affordable and provides a means for researchers working with modest budgets to identify massive numbers of microsatellite loci from nonmodel organisms (Castoe *et al.* 2012). Of the 30 loci originally trialled on *G. thelemanniana* ssp. *thelemanniana*, 12 were polymorphic (Chapter 5). However, when I tested some of the “monomorphic” primers on other selected species from the Thelemanniana group, polymorphism was

revealed in these loci (Chapter 3). This result highlights the potential for cross amplification of numerous primer pairs, especially given the 454 library provided me with over 500 potential microsatellite regions.

Conservation and taxonomic implications

In Box 6.1 I've listed the conservation codes outlined by the Department of Parks and Wildlife together with the Wildlife Conservation Act 1950 and the IUCN. Furthermore, Coates and Atkins (2001) outlined a priority setting process specifically designed for Western Australia Flora where the majority of species are poorly known but may be of conservation significance. These three categories are ranked in order of priority for survey and evaluation of conservation status so that consideration can be given to their declaration as threatened flora. Species that are adequately known, are rare but not threatened or meet criteria for Near Threatened, or that have been recently removed from the threatened list for other than taxonomic reasons, are placed in Priority 4. The Thelemanniana group is composed of numerous threatened and poorly known taxa (Chapter 2; Table 2.2), which underpins the necessity for resolving species boundaries and genetic diversity of this group. I address each of the threatened and priority listed flora, and the taxonomic and conservation implications that stem from the results of the molecular analyses, below;

Box 6.1 Conservation Codes of Western Australian Flora[1]

T: Threatened species - Specially protected under the Wildlife Conservation Act 1950, which have been adequately searched for and are deemed to be in the wild either rare, in danger of extinction, or otherwise in need of special protection.

Threatened Fauna and Flora are further recognised by the Department according to their level of threat using IUCN Red List criteria.

CR: Critically Endangered - considered to be facing an extremely high risk of extinction in the wild.

EN: Endangered – considered to be facing a very high risk of extinction in the wild.

VU: Vulnerable - considered to be facing a high risk of extinction in the wild.

Definitions for the grouping of rare and poorly known flora in Western Australia[2]

Priority one - poorly known taxa

Taxa which are known from one or a few (generally <5) populations which are under threat due to small population size or lands under immediate threat; road verges, urban areas, farmland, active mineral leases, etc., or the plants are under threat from disease, grazing by feral animals, etc.

May include taxa with threatened populations on protected lands which are under consideration for declaration as 'rare flora' (threatened flora), but are in urgent need of further survey.

Priority two - poorly known taxa

Taxa which are known from one or a few (generally <5) populations, at least some of which are not believed to be under immediate threat (i.e. not currently endangered).

Such taxa are under consideration for declaration as `rare flora' (threatened flora), but are in urgent need of further survey.

Priority three - poorly known taxa

Taxa which are known from several populations, and the taxa are not believed to be under immediate threat (i.e. not currently endangered), either due to the number of known populations (generally >5), or known populations being large, and either widespread or protected. Such taxa are under consideration for declaration as `rare flora' (threatened flora), but are in need of further survey.

Priority four - rare taxa

Taxa which are considered to have been adequately surveyed and which, whilst being rare (in Australia), are not currently threatened by any identifiable factors. These taxa require monitoring every 5±10 years.

[1] Department of Parks and Wildlife .

[2] Coates and Atkins (2001) Priority setting and the conservation of Western Australia's diverse and highly endemic flora.

1) The support for monophyly of Grevillea sp. Gillingara is important for the conservation actions which follow. Firstly, its reclassification from 'phrase named' to a unique species is critical, followed by immediate protection of the severely restricted range it currently occupies. With only one roadside population currently known, which is restricted even further by local agricultural clearing, it is recommended that seeds be collected for *ex situ* conservation. *Grevillea* sp. Gillingara has recently been listed as critically endangered (Smith 2013).

2) *Grevillea acropogon, G. humifusa* and *G. mccutcheonii* are currently listed as threatened taxa. The support for these species from molecular data reinforces their current conservation status. Recovery Plans for these species have been prepared and the remaining populations are marked and protected by the Western Australian Department of Parks and Wildlife. Current conservation efforts on *G. humifusa* are focusing on establishing translocated populations in new areas adjacent to its distribution (Sheehy, N. pers,. comm). Inferences could not be accurately made about the genetic diversity within *G. acropogon* and *G. mccutcheonii* based on the low sample size (Chapter 2). However, greater sampling of *G. humifusa* (Chapter 3) illustrated low chloroplast and microsatellite diversity (heterozygosity and allelic richness). Low starting genetic diversity can impact on the success of species recovery through translocation by genetic erosion and inbreeding (Krauss *et al.* 2002; Hufford & Mazer 2003).

3) *Grevillea thelemanniana* ssp. Cooljarloo and *G. evanescens* are listed as Priority 1. There is strong evidence to support the recognition of *G. thelemanniana* ssp. Cooljarloo as a unique species. Firstly, the chloroplast and microsatellite data analyses (Chapter 2) illustrated a close relationship with *G. delta* and *G. humifusa* rather than with *G. thelemanniana* ssp. *thelemanniana.* Secondly, greater sampling of populations (Chapter 3) illustrated strong support for significant divergence of *G. thelemanniana* ssp. Cooljarloo from *G. delta* and *G. humifusa.* The current status of Priority 1 does not apply to *G. thelemanniana* ssp. Cooljarloo anymore, after surveying its range revealed numerous large continuous populations. This species can therefore be reclassified under the guidelines of Priority 4. Recommendations for monitoring are explained in detail below. The status of *G. evanescens* is still confounding, although unique haplotypes were found between *G. evanescens* and *G. obtusifolia* microsatellite data

did not clearly differentiate the taxa and some admixture between the two may occur. This relationship would need further study before revision of its current status could be made.

4) *Grevillea delta* and *G. stenomera* are listed as Priority 2 taxa. Microsatellite data supported genetic clustering of both species (Chapters 2 and 3) and should therefore be regarded as unique species. Although both species have restricted ranges, the majority of their populations exist within two large National Parks; Mt Lesueur and Zuytdorp and are therefore currently protected. The one exception is one population of *G. delta* (Chapter 3; Population CG), which occurs in a small remnent bushland area along Coorow-Greenhead road (2 kilometres x 100 meters). Coates and Atkins (2001) emphasise that the conservation of priority flora is also addressed at ecosystem and ecological community levels, and that each approach has its merits depending upon land tenure. Given the small populations *G. delta* populations found, I recommend that this small patch of remnent bushland be recognised and protected for the conservation of *G. delta*.

5) *Grevillea costata* and *G. hirtella* are listed as Priority 3 taxa. *Grevillea costata* is supported as monophyletic in the phylogenetic analyses, and its continuous distribution along Murchison River in Kalbarri National Park denotes that it remains protected. *Grevillea hirtella* however could not be located at numerous locations in nature from which it was previously collected, which was partially due to large spread fire in the area which has likely killed off numerous plants (Keighery, G. pers., comm.), and partially due to inaccurate GPS recordings. Given that the majority of this species is distributed in small roadside populations, in a highly fragmented agricultural region, the reassessment of its current abundance and range is necessary. A reclassification of this species to Priority 1 is recommended.

6) *Grevillea olivacea, G. ripicola* and *G. thelemanniana* ssp. *thelemanniana* are listed as Priority 4 taxa. Molecular analyses supported the monophyly of all three species. The distribution of *G. olivacea* along the coast is relatively secure since large parts of this range are protected national parks. Similarly, *G. ripicola* is continuously distributed along large sections of the Collie River through Wellington National Park, Mumballup and Collie State Forest, of which large areas are protected. Populations of *G.*

thelemanniana ssp. *thelemanniana* are restricted to a remnent patch of native bush within an urban environment. The remnant patch is severely degraded by weed invasion and pollution from neighbouring developments. However plants appear healthy, and are abundant and reproductive (Personal observation). The taxonomy of *G. thelemanniana* ssp. *thelemanniana* remains somewhat confusing. The molecular data revealed haplotype sharing with *G. obtusifolia* and microsatellite data did not show clear differentiation between the two at K=17 but did at K+20. These data reveal that these taxa should be looked at using a more comprehensive sampling to better assess the conservation status of *G. thelemanniana* ssp. *thelemanniana*.

Genetic monitoring

Recognition of species and populations through molecular data is important for determining conservation actions, and analyses of the population structure aids in setting priorities on how they are conserved. This can be illustrated through the analyses of the spatial genetic structure of *G. thelemanniana* ssp. Cooljarloo.

The level and pattern of genetic differentiation of populations of *G. thelemanniana* ssp. Cooljarloo suggest historically low levels of gene flow among creek lines. Historically low connectivity suggests that the land clearing due to the mine site will not change the pattern of divergence among populations. There were high levels of genetic variation within populations (Chapter 4), which is likely maintained by naturally large and continuous populations to counteract the effects of inbreeding and genetic drift. Consequently, there is conservation concern regarding the northern distribution of *G. thelemanniana* ssp. Cooljarloo, as land clearing for mining activities in this region may have reduced its distribution. Furthermore, the fire which has occurred between the southern and western distributions could have severed an important connection between these regions, as genetic analyses revealed significant admixture between these regions. Hopper (2009) and Sampson *et al* (2013) proposed that rare, long-lived, self-compatible species, with high genetic differentiation between populations may be less susceptible to effects of habitat fragmentation, if their population systems remain intact and provided edge effects are minimal. Maintaining high levels of genetic diversity has been proposed as one of the most critical aspects of species ability to survive environmental change (Petit *et al.* 1998; Bonin *et al.* 2007). These

circumstances apply directly to *G. thelemanniana* ssp. Cooljarloo, and therefore conservation efforts for *G. thelemanniana* ssp. Cooljarloo may be best directed to managing intact habitats and ecosystem processes.

Beyond general concepts and reviews, few studies have practically demonstrated how genetic information can be realistically applied in current conservation projects. Authors have called for a renewed incorporation of adaptive (molecular and/or quantitative) information into conservation planning (Bonin *et al.* 2007; Schwartz *et al.* 2007; Allendorf *et al.* 2010; Laikre 2010). The results from this study will hopefully inform the management practises that Tronox Mining Operations will now tackle to conserve *G. thelemanniana* ssp. Cooljarloo.

Limitations of current study

Selecting which species concept approach to following when attempting to delineate species is an ongling debate. Although I set out to use a coalescent-based approach, I realised that I did not have the appropriate data to apply this method. Although I attempted to incorporate a nuclear gene, I did not get successful amplification with many of the taxa within the group. Therefore, the phylogenetic analyses component of Chapter 3 would benefit from the inclusion of nuclear genomes.

Sampling sizes are often an issue in phylogenetic studies, and recent reviews have highlighted the need to sample multiple populations of species instead of just relying on one representative population of individual for a species (Omland *et al.* 2006). This becomes particularly important when attempting to resolve relationships within recently diversifying lineages where incomplete lineage sorting will be a potential issue. The results from this study would have been more robust if I had sampled a greater number of individuals and populations of each species for the microsatellite dataset in Chapter 2. The analyses of microsatellite would have been more informative with greater sample sizes that would have allowed measurement of genetic parameters such as genetic variation, and inbreeding. The failure to amplify ITS primers was discovered only after sampling of species within the Thelemanniana group had been carried out, and re-visiting sampling locations for all 20 species was beyond the time frame and budget of this study.

Populations from the southern range of *G. pinaster*'s were not sampled in this study because previously recorded populations could not be found due to lack of accurate location data or population loss. Data from the southern end of the range would have provided more robust results discerning whether hybridization and/or introgression or incomplete lineage sorting lead to incongruence in cpDNA and nDNA microsatellite results regardign *G. delta* and *G. pinaster* (Chapter 3). Further sampling of Hakeoides and Linnearifolia groups which exists along the same ranges as the Thelemanniana group may reveal significant relationships that explain this phenomenon.

Future research

Avenues for future research related to both phylogenetic and population genetic studies which would help in resolving some of the confounding patterns in this study are as follows:

1) The *Grevillea* species groups proposed by Makinson (2000) require some reassessment. The paraphyletic relationship between the Thelemanniana group with both Hakeoides and Linnearifolia group species indicates that further sampling of species in these groups would further clarify classifications. The phylogenetic analyses would benefit from inclusion of the nuclear genome and further analyses using a coalescent based species approach. The groupings outlined in this thesis are likely a good starting point for further analyses through different statistical approaches. I recognise the need for a larger dataset to resolve the many complex relationships my data revealed, however both time and resources contraints prevented me from exploring this further.

2) The differentiation of *Grevillea obtusifolia* and *G. evanescens* is not clear. Glass house experiments have shown that species of the Thelemanniana group have a great propensity to hybridize (Growns, D., pers comm.). These results suggest either hybridization and/or introgression could possibly be occurring between *G. obtusifolia* and *G. evanescens,* or they are in fact the same species. Given the overlapping ranges and flowering times, and the morphological cline which has been noted between these species (Keighery, pers., comm.), illustrates that both greenhouse and comprehensive population genetics studies are need to further evaluate the species status.

3) As mentioned above, a large portion of species and subspecies within the Thelemanniana group, and also species from the Hakeoides and Linnearifolia group, were distributed in creek beds, flood plains and along river banks. Makinson (2000) wrote that *Grevillea* tends to show little tolerance for seasonally waterlogged environments, however this does not correspond to the distributions of the species in this study. The influence of hydrochory would be particularly interesting to study in *G. costata* which grows along the Murchison River in Kalbarri, or *G. ripicola* which grows along the Collie River in Collie.

4) The population genetic structure of *G. thelemanniana* ssp. Cooljarloo indicated that it was largely driven by distance between populations, and consequently based on the movements of pollinators and seed. The spatial genetic structure indicates that nearest neighbour mating is prominent, likely through invertebrate pollination, and gravity dispersed seed. In addition, genetic admixture between populations separated by greater distances indicates that pollination by birds is also likely. No firm conclusions can be made on the pollinators of G. *thelemanniana* ssp. Cooljarloo, and a detailed study of the mating system and realised dispersal of seed and pollen is required to assess genetic connectivity directly.

References

Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation genetics. *Nature Reviews Genetics* **11**, 697-709.

Andersson L (1990) The driving force: Species concepts and ecology. *Taxon*, 375-382.

Arnold ML (1992) Natural hybridization as an evolutionary process. *Annual Review of Ecology and Systematics*, 237-261.

Arnold ML (1997) *Natural hybridization and evolution* Oxford University Press, Oxford, UK.

Auld TD, Denham A (1999) The role of ants and mammals in dispersal and post-dispersal seed predation of the shrubs *Grevillea* (Proteaceae). *Plant Ecology* **144**, 201-213.

Auld TD, Denham AJ (2001) The impact of seed predation by mammals on post-fire seed accumulation in the endangered shrub *Grevillea caleyi* (Proteaceae). *Biological Conservation* **97**, 377-385.

Avise JC (2000) *Phylogeography: the history and formation of species* Harvard University Press.

Ayre DJ, Whelan RJ, Reid A (1994) Unexpectedly high levels of selfing in the Australian shrub *Grevillea barklyana* (Proteaceae). *Heredity* **72**, 168-174.

Baack EJ, Rieseberg LH (2007) A genomic view of introgression and hybrid speciation. *Current opinion in genetics & development* **17**, 513-518.

Bacon CD, McKenna MJ, Simmons MP, Wagner WL (2012) Evaluating multiple criteria for species delimitation: an empirical example using Hawaiian palms (Arecaceae: *Pritchardia*). *BMC evolutionary biology* **12**, 23.

Baker AJ, Miskelly CM, Haddrath O (2010) Species limits and population differentiation in New Zealand snipes (Scolopacidae: *Coenocorypha*). *Conservation Genetics* **11**, 1363-1374.

Barker NP, Weston PH, Rourke JP, Reeves G (2002) The relationships of the southern African Proteaceae as elucidated by internal transcribed spacer (ITS) DNA sequence data. *Kew Bulletin* **57**, 867-883.

Barker NP, Weston PH, Rutschmann F, Sauquet H (2007) Molecular dating of the 'Gondwanan'plant family Proteaceae is only partially congruent with the timing of the break-up of Gondwana. *Journal of Biogeography* **34**, 2012-2027.

Barker WJ, Hedderson TA, Dransfield J (2000) Molecular Phylogenetics of Subfamily Calamoideae (Palmae) Based on nrDNA and cpDNA *rps*16 Intron Sequence Data. *Molecular Phylogenetics and Evolution* **14,** 197-217.

Barrett CF, Freudenstein JV (2011) An integrative approach to delimiting species in a rare but widespread mycoheterotrophic orchid. *Molecular Ecology* **20**, 2771-2786.

Bates BC, Hope P, Ryan B, Smith I, Charles S (2008) Key findings from the Indian Ocean Climate Initiative and their impact on policy development in Australia. *Climatic Change* **89**, 339-354.

Beard J, Chapman A, Gioia P (2000) Species richness and endemism in the Western Australian flora. *Journal of Biogeography* **27**, 1257-1268.

Bonin A, Nicole F, Pompanon F, Miaud C, Taberlet P (2007) Population adaptive index: a new method to help measure intraspecific genetic diversity and prioritize populations for conservation. *Conservation Biology* **21**, 697-708.

Briggs JD, Leigh JH (1996) *Rare or threatened Australian plants* CSIRO PUBLISHING.

Broadhurst L, Byrne M, Craven L, Lepschi B (2004) Genetic congruence with new species boundaries in the *Melaleuca uncinata* complex (Myrtaceae). *Australian Journal of Botany* **52**, 729-737.

Broadhurst LM, Coates DJ, Tan BH (2001) Patterns of morphological variation and allelic distribution across a zone of overlap between two subspecies of *Geleznowia verrucosa* (Rutaceae). *Australian Journal of Botany* **49**, 451-458.

Butcher P, Otero A, McDonald M, Moran G (2002) Nuclear RFLP variation in Eucalyptus *camaldulensis* Dehnh. from northern Australia. *Heredity* **88**, 402-412.

Butcher R, Byrne M, Crayn DM (2007) Evidence for convergent evolution among phylogenetically distant rare species of *Tetratheca* (Elaeocarpaceae, formerly Tremandraceae) from Western Australia. *Australian Systematic Botany* **20**, 126-138.

Byrne M (2007) Phylogeography provides an evolutionary context for the conservation of a diverse and ancient flora. *Australian Journal of Botany* **55**, 316-325.

Byrne M, Elliott C, Yates C, Coates D (2007) Extensive pollen dispersal in a bird-pollinated shrub, *Calothamnus quadrifidus*, in a fragmented landscape. *Molecular Ecology* **16**, 1303-1314.

Byrne M, Hines B (2004) Phylogeographical analysis of cpDNA variation in *Eucalyptus loxophleba* (Myrtaceae). *Australian Journal of Botany* **52**, 459-470.

Byrne M, Hopper SD (2008) Granite outcrops as ancient islands in old landscapes: evidence from the phylogeography and population genetics of *Eucalyptus caesia* (Myrtaceae) in Western Australia. *Biological Journal of the Linnean Society* **93**, 177-188.

Byrne M, Macdonald B, Brand J (2003) Phylogeography and divergence in the chloroplast genome of Western Australian Sandalwood (*Santalum spicatum*). *Heredity* **91**, 389-395.

Byrne M, Macdonald B, Coates D (1999) Divergence in the chloroplast genome and nuclear rDNA of the rare Western Australian plant *Lambertia orbifolia* Gardner (Proteaceae). *Molecular Ecology* **8**, 1789-1796.

Byrne M, Macdonald B, Coates D (2002) Phylogeographical patterns in chloroplast DNA variation within the *Acacia acuminata* (Leguminosae: Mimosoideae) complex in Western Australia. *Journal of Evolutionary Biology* **15**, 576-587.

Byrne M, Steane DA, Joseph L, Yeates DK, Jordan GJ, Crayn D, Aplin K, Cantrill DJ, Cook LG, Crisp MD (2011) Decline of a biome: evolution, contraction, fragmentation, extinction and invasion of the Australian mesic zone biota. *Journal of Biogeography* **38**, 1635-1656.

Byrne M, Tischler G, Macdonald B, Coates D, McComb J (2001) Phylogenetic relationships between two rare acacias and their common, widespread relatives in south-western Australia. *Conservation Genetics* **2**, 157-166.

Byrne M, Yeates D, Joseph L, Kearney M, Bowler J, Williams M, Cooper S, Donnellan S, Keogh J, Leys R (2008) Birth of a biome: insights into the assembly and maintenance of the Australian arid zone biota. *Molecular Ecology* **17**, 4398-4417.

Caddy H, Gross C (2006) Population structure and fecundity in the putative sterile shrub, *Grevillea rhizomatosa* Olde and Marriott (Proteaceae) **127**, 11.

Cardillo M, Pratt R (2013) Evolution of a hotspot genus: geographic variation in speciation and extinction rates in *Banksia* (Proteaceae). *BMC Evolutionary Biology* **13**, 155.

Cardoso A, Vogler A (2005) DNA taxonomy, phylogeny and Pleistocene diversification of the *Cicindela hybrida* species group (Coleoptera: Cicindelidae). *Molecular Ecology* **14**, 3531-3546.

Carstens BC, Dewey TA (2010) Species delimitation using a combined coalescent and information-theoretic approach: an example from North American Myotis bats. *Systematic Biology* **59**, 400-414.

Carstens BC, Knowles LL (2007) Estimating species phylogeny from gene-tree probabilities despite incomplete lineage sorting: an example from *Melanoplus* grasshoppers. *Systematic Biology* **56**, 400-411.

Castoe TA, Streicher JW, Meik JM, Ingrasci MJ, Poole AW, Koning A, Campbell JA, Parkinson CL, Smith EN, Pollock DD (2012) Thousands of microsatellite loci from the venomous coralsnake *Micrurus fulvius* and variability of select loci across populations and related species. *Molecular Ecology Resources* **12**, 1105-1113.

Cegelski C, Waits L, Anderson N (2003) Assessing population structure and gene flow in Montana wolverines (*Gulo gulo*) using assignment-based approaches. *Molecular Ecology* **12**, 2907-2918.

Celebrezze T, Paton DC (2004) Do introduced honeybees (*Apis mellifera*, Hymenoptera) provide full pollination service to bird-adapted Australian plants with small flowers? An experimental study of *Brachyloma ericoides* (Epacridaceae). *Australian Ecology* **29**, 129-136.

Chen C, Durand E, Forbes F, François O (2007) Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. *Molecular Ecology Notes* **7**, 747-756.

Christophel DC (1989) Evolution of the Australian flora through the Tertiary. *Plant Systematics and Evolution* **162**, 63-78.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology* **9**, 1657-1659.

Coates DJ (2000) Defining conservation units in a rich and fragmented flora: implications for the management of genetic resources and evolutionary

processes in south-west Australian plants. *Australian Journal of Botany* **48**, 329-339.

Coates DJ, Atkins KA (2001) Priority setting and the conservation of Western Australia's diverse and highly endemic flora. *Biological Conservation* **97**, 251-263.

Coates DJ, Carstairs S, Hamley VL (2003) Evolutionary patterns and genetic structure in localized and widespread species in the *Stylidium caricifolium* complex (Stylidiaceae). *American Journal of Botany* **90**, 997-1008.

Coates DJ, Hamley VL (1999) Genetic divergence and the mating system in the endangered and geographically restricted species, *Lambertia orbifolia* Gardner (Proteaceae). *Heredity* **83**, 418-427.

Coates DJ, Sampson JF, Yates CJ (2007) Plant mating systems and assessing population persistence in fragmented landscapes. *Australian Journal of Botany* **55**, 239-249.

Cook LG, Morris DC, Edwards RD, Crisp MD (2008) Reticulate evolution in the natural range of the invasive wetland tree species *Melaleuca quinquenervia*. *Molecular Phylogenetics and Evolution* **47**, 506-522.

Cooper SJ, Harvey MS, Saint KM, Main BY (2011) Deep phylogeographic structuring of populations of the trapdoor spider *Moggridgea tingle* (Migidae) from southwestern Australia: evidence for long-term refugia within refugia. *Molecular Ecology* **20**, 3219-3236.

Corander J, Sirén J, Arjas E (2008) Bayesian spatial modeling of genetic population structure. *Computational Statistics* **23**, 111-129.

Cornuet J-M, Piry S, Luikart G, Estoup A, Solignac M (1999) New methods employing multilocus genotypes to select or exclude populations as origins of individuals. *Genetics* **153**, 1989-2000.

Cowling RM, Lamont BB (1998) On the nature of Gondwanan species flocks: diversity of Proteaceae in Mediterranean south-western Australia and South Africa. *Australian Journal of Botany* **46**, 335-355.

Cowling RM, Rundel PW, Lamont BB, Kalin Arroyo M, Arianoutsou M (1996) Plant diversity in Mediterranean-climate regions. *Trends in Ecology and Evolution* **11**, 362-366.

Cracraft J (1983) Species concepts and speciation analysis. *Current Ornithology*, pp. 159-187. Springer.

Crawford NG (2010) SMOGD: software for the measurement of genetic diversity. *Molecular Ecology Resources* **10**, 556-557.

Crisp M, Cook L, Steane D (2004) Radiation of the Australian flora: what can comparisons of molecular phylogenies across multiple taxa tell us about the evolution of diversity in present–day communities? *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **359**, 1551-1571.

Crisp MD, Cook LG (2007) A congruent molecular signature of vicariance across multiple plant lineages. *Molecular Phylogenetics and Evolution* **43**, 1106-1117.

de Queiroz K (1998) The general lineage concept of species, species criteria, and the process of speciation: A conceptual unification and terminological recommendations. Pages 57–75 in Endless forms: Species and speciation. Oxford University Press.

de Queiroz K (2007) Species concepts and species delimitation. *Systematic Biology* **56**, 879-886.

Degnan JH, Rosenberg NA (2009) Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology and Evolution* **24**, 332-340.

Dettmann ME, Jarzen DM (1998) The early history of the Proteaceae in Australia: the pollen record. *Australian Systematic Botany* **11**, 401-438.

Doyle JJ (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bullitin* **19**, 11-15.

Doyle JJ (1990) Isolation of plant DNA from fresh tissue. *Focus* **12**, 13-15.

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* **7**, 214.

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* **29**, 1969-1973.

Duminil J, Di Michele M (2009) Plant species delimitation: a comparison of morphological and molecular markers. *Plant Biosystems* **143**, 528-542.

Duminil J, Kenfack D, Viscosi V, Grumiau L, Hardy OJ (2012) Testing species delimitation in sympatric species complexes: The case of an African tropical tree, *Carapa* spp.(Meliaceae). *Molecular Phylogenetics and Evolution* **62**, 275-285.

Earl D, von Holdt B (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output ans implementing the Evanno method. . *Conservation Genetics Resources* **4**, 359-361.

Edwards C, Soltis D, Soltis P (2008) Using patterns of genetic structure based on microsatellite loci to test hypotheses of current hybridization, ancient hybridization and incomplete lineage sorting in *Conradina* (Lamiaceae). *Molecular Ecology* **17**, 5157-5174.

Edwards DL (2007) Biogeography and speciation of a direct developing frog from the coastal arid zone of Western Australia. *Molecular Phylogenetics and Evolution* **45**, 494-505.

Edwards RD, Craven LA, Crisp MD, Cook LG (2010) *Melaleuca* revisited: cpDNA and morphological data confirm that *Melaleuca* L.(Myrtaceae) is not monophyletic. *Taxon*, 744-754.

Elliott C, Byrne M (2004) Phylogenetics and the conservation of rare taxa in the Eucalyptus angustissima complex in Western Australia. *Conservation Genetics* **5**, 39-47.

Elliott CP, Yates CJ, Ladd PG, Coates DJ (2002) Morphometric, genetic and ecological studies clarify the conservation status of a rare *Acacia* in Western Australia. *Australian Journal of Botany* **50**, 63-74.

England P, Whelan R, Ayre D (2003) Effects of seed bank disturbance on the fine-scale genetic structure of populations of the rare shrub *Grevillea macleayana*. *Heredity* **91**, 475-480.

England PR, Ayre DJ, Whelan RJ (1999) Microsatellites in the Australian shrub *Grevillea macleayana* (Proteaceae). *Molecular Ecology* **8**, 689-690.

England PR, Beynon F, Ayre DJ, Whelan RJ (2001) A Molecular Genetic Assessment of Mating-System Variation in a Naturally Bird-Pollinated Shrub: Contributions from Birds and Introduced Honeybees. *Conservation Biology* **15**, 1645-1655.

Epps CW, Wehausen JD, Bleich VC, Torres SG, Brashares JS (2007) Optimizing dispersal and corridor models using landscape genetics. *Journal of Applied Ecology* **44**, 714-724.

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

Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**, 564-567.

Ford HA, Paton DC, Forde N (1979) Birds as pollinators of Australian plants. *New Zealand Journal of Botany* **17**, 509-519.

Franck P, Cameron E, Good G, RASPLUS JY, Oldroyd B (2004) Nest architecture and genetic differentiation in a species complex of Australian stingless bees. *Molecular Ecology* **13**, 2317-2331.

François O, Durand E (2010) Spatially explicit Bayesian clustering models in population genetics. *Molecular Ecology Resources* **10**, 773-784.

Fujita MK, Leaché AD, Burbrink FT, McGuire JA, Moritz C (2012) Coalescent-based species delimitation in an integrative taxonomy. *Trends in Ecology and Evolution* **27**, 480-488.

Gaggiotti O, Lange O, Rassmann K, Gliddon C (1999) A comparison of two indirect methods for estimating average levels of gene flow using microsatellite data. *Molecular Ecology* **8**, 1513-1520.

Gardner MG, Fitch AJ, Bertozzi T, Lowe AJ (2011) Rise of the machines–recommendations for ecologists when using next generation sequencing for microsatellite development. *Molecular Ecology Resources* **11**, 1093-1101.

George AS (1998) Proteus in Australia. An overview of the current state of taxonomy of the Australian Proteaceae. *Australian Systematic Botany* **11**, 257-266.

Gibson N, Keighery G, Keighery B (2000) Threatened plant communities of Western Australia. 1. The ironstone communities of the Swan and Scott Coastal Plains. *Journal of the Royal Society of Western Australia* **83**, 1-11.

Goldstein DB, Linares AR, Cavalli-Sforza LL, Feldman MW (1995) An evaluation of genetic distances for use with microsatellite loci. *Genetics* **139**, 463-471.

Goldstein PZ, Desalle R, Amato G, Vogler AP (2000) Conservation genetics at the species boundary. *Conservation Biology* **14**, 120-131.

Gornall RJ, Hollingsworth PM, Preston CD (1998) Evidence for spatial structure and directional gene flow in a population of an aquatic plant, *Potamogeton coloratus*. *Heredity* **80**, 414-421.

Griffin E, Hopkins A, Hnatiuk R (1983) Regional Variation in Mediterranean-type shrublands near Eneabba, south-western Australia. *Vegetation* **52**, 103-127.

Guillot G, Estoup A, Mortier F, Cosson JF (2005) A spatial statistical model for landscape genetics. *Genetics* **170**, 1261-1280.

Guillot G, Leblois R, Coulon A, Frantz AC (2009) Statistical methods in spatial genetics. *Molecular Ecology* **18**, 4734-4756.

Harriss F, Whelan R (1993) Selective fruit abortion in *Grevillea barklyana* (Proteaceae). *Australian Journal of Botany* **41**, 499-509.

Hart MW (2011) The species concept as an emergent property of population biology. *Evolution* **65**, 613-616.

Hausdorf B, Hennig C (2010) Species delimitation using dominant and codominant multilocus markers. *Systematic Biology* **59**, 491-503.

He T, Krauss SL, Lamont BB, Miller BP, Enright NJ (2004) Long-distance seed dispersal in a metapopulation of *Banksia hookeriana* inferred from a population allocation analysis of amplified fragment length polymorphism data. *Molecular Ecology* **13**, 1099-1109.

Hedges BS, Kumar S (2004) Precision of molecular time estimates. *TRENDS in Genetics* **20**, 242-247.

Heled J, Drummond AJ (2010) Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution* **27**, 570-580.

Heller R, Siegismund HR (2009) Relationship between three measures of genetic differentiation GST, DEST and G'ST: how wrong have we been? *Molecular Ecology* **18**, 2080-2083.

Herbarium WA (1998) FloraBase—The Western Australian Flora. Kensington, Western Australia, Department of Parks and Wildlife, florabase.dpaw.wa.gov.au.

Hermanutz L, Innes D, Denham A, Whelan R (1998) Very low fruit: flower ratios in *Grevillea* (Proteaceae) are independent of breeding system. *Australian Journal of Botany* **46**, 465-478.

Hevroy TH, Moody ML, Krauss SL, Gardner MG (2013) Isolation, via 454 sequencing, characterization and transferability of microsatellites for *Grevillea thelemanniana* subsp. *thelemanniana* and cross-species amplification in the *Grevillea thelemanniana* complex (Proteaceae). *Conservation Genetics Resources* **5**, 887-890.

Hey J, Pinho C (2012) Population genetics and objectivity in species diagnosis. *Evolution* **66**, 1413-1429.

Hey J, Waples RS, Arnold ML, Butlin RK, Harrison RG (2003) Understanding and confronting species uncertainty in biology and conservation. *Trends in Ecology and Evolution* **18**, 597-603.

Hoebee S (2011) Development and cross-species amplification of microsatellite markers from the endangered Wee Jasper *Grevillea* (*Grevillea iaspicula*, Proteaceae). *Muelleria* **29**, 93-96.

Hoebee SE, Young AG (2001) Low neighbourhood size and high interpopulation differentiation in the endangered shrub *Grevillea iaspicula* McGill (Proteaceae). *Heredity* **86**, 489-496.

Holderegger R, Buehler D, Gugerli F, Manel S (2010) Landscape genetics of plants. *Trends in Plant Science* **15**, 675-683.

Holmes GD, James EA, Hoffmann AA (2009) Divergent levels of genetic variation and ploidy among populations of the rare shrub, *Grevillea repens* (Proteaceae). *Conservation Genetics* **10**, 827-837.

Honnay O, Jacquemyn H, Nackaerts K, Breyne P, Van Looy K (2010) Patterns of population genetic diversity in riparian and aquatic plant species along rivers. *Journal of Biogeography* **37**, 1730-1739.

Hoot SB, Douglas AW (1998) Phylogeny of the Proteaceae based on atpB and atpB-rbcL intergenic spacer region sequences. *Australian Systematic Botany* **11**, 301-320.

Hopper S, Chappill J, Harvey M, George A (1996) *Gondwanan Heritage: Past, Present and Future of the Western Australian biota* Surrey Beatty & Sons, Chipping Norton, NSW, Australia

Hopper SD (1979) Biogeographical aspects of speciation in the southwest Australian flora. *Annual Review of Ecology and Systematics* **10**, 399-422.

Hopper SD (2000) How well do phylogenetic studies inform the conservation of Australian plants? *Australian Journal of Botany* **48**, 321-328.

Hopper SD (2009) OCBIL theory: towards an integrated understanding of the evolution, ecology and conservation of biodiversity on old, climatically buffered, infertile landscapes. *Plant and Soil* **322**, 49-86.

Hopper SD, Gioia P (2004) The southwest Australian floristic region: evolution and conservation of a global hot spot of biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, 623-650.

Hopper SD, Smith RJ, Fay MF, Manning JC, Chase MW (2009) Molecular phylogenetics of Haemodoraceae in the Greater Cape and Southwest Australian floristic regions. *Molecular Phylogenetics and Evolution* **51**, 19-30.

Hu LJ, Uchiyama K, Shen HL, Saito Y, Tsuda Y, Ide Y (2008) Nuclear DNA microsatellites reveal genetic variation but a lack of phylogeographical structure in an endangered species, *Fraxinus mandshurica*, across North-east China. *Ann Bot* **102**, 195-205.

Hudson RR, Coyne JA (2002) Mathematical consequences of the genealogical species concept. *Evolution* **56**, 1557-1565.

Hufford KM, Mazer SJ (2003) Plant ecotypes: genetic differentiation in the age of ecological restoration. *Trends in Ecology and Evolution* **18**, 147-155.

Jacquemyn H, Honnay O, Galbusera P, Roldan-Ruiz I (2004) Genetic structure of the forest herb *Primula elatior* in a changing landscape. *Molecular Ecology* **13**, 211-219.

Jacquemyn H, Honnay O, Van Looy K, Breyne P (2006) Spatiotemporal structure of genetic variation of a spreading plant metapopulation on dynamic riverbanks along the Meuse River. *Heredity* **96**, 471-478.

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23**, 1801-1806.

James E, Brown G, Citroen R, Blacket M (2012) Microsatellite marker development for two species of holly-leafed *Grevillea* and cross-species amplification in the Aspleniifolia/Hookeriana Subgroup (Proteaceae). *Conservation Genetics Resources* **4**, 137-140.

Janes JK, Steane DA, Vaillancourt RE (2012) What does population structure analysis reveal about the *Pterostylis longifolia* complex (Orchidaceae)? *Ecology and Evolution* **2**, 2631-2644.

Jarne P, Lagoda PJ (1996) Microsatellites, from molecules to populations and back. *Trends in Ecology and Evolution* **11**, 424-429.

Johnson LA, Briggs BG (1975) On the Proteaceae—the evolution and classification of a southern family*. *Botanical Journal of the Linnean Society* **70**, 83-182.

Jones RC, Steane DA, Lavery M, Vaillancourt RE, Potts BM (2012) Multiple evolutionary processes drive the patterns of genetic differentiation in a forest tree species complex. *Ecology and Evolution* **3**, 1-17.

Jost L (2008) GST and its relatives do not measure differentiation. *Mol Ecol* **17**, 4015-4026.

Kalinowski ST (2005) hp-rare 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes* **5**, 187-189.

Kane NC, King MG, Barker MS, Raduski A, Karrenberg S, Yatabe Y, Knapp SJ, Rieseberg LH (2009) Comparative genomic and population genetic analyses indicate highly porous genomes and high levels of gene flow between divergent *Helianthus* species. *Evolution* **63**, 2061-2075.

Kim C, Jung J, Choi H-K (2012) Molecular identification of *Schoenoplectiella* species (Cyperaceae) by use of microsatellite markers. *Plant Systematics and Evolution* **298**, 811-817.

Kitamoto N, Honjo M, Ueno S, Takenaka A, Tsumura Y, Washitani I, Ohsawa R (2005) Spatial genetic structure among and within populations of *Primula sieboldii* growing beside separate streams. *Molecular Ecology* **14**, 149-157.

Knowles LL (2009) Estimating species trees: methods of phylogenetic analysis when there is incongruence across genes. *Systematic Biology* **58**, 463-467.

Knowles LL, Carstens BC (2007) Delimiting species without monophyletic gene trees. *Systematic Biology* **56**, 887-895.

Koch MA, Dobeš C, Mitchell-Olds T (2003) Multiple hybrid formation in natural populations: concerted evolution of the internal transcribed spacer of nuclear ribosomal DNA (ITS) in North American *Arabis divaricarpa* (Brassicaceae). *Molecular Biology and Evolution* **20**, 338-350.

Krak K, Caklová P, Chrtek J, Fehrer J (2013) Reconstruction of phylogenetic relationships in a highly reticulate group with deep coalescence and recent speciation (*Hieracium*, Asteraceae). *Heredity* **110**, 138-151.

Kramer AT, Fant JB, Ashley MV (2011) Influences of landscape and pollinators on population genetic structure: examples from three *Penstemon* (Plantaginaceae) species in the Great Basin. *American Journal of Botany* **98**, 109-121.

Krauss S, He T, Barrett L, Lamont B, Enright N, Miller B, Hanley M (2009) Contrasting impacts of pollen and seed dispersal on spatial genetic structure in the bird-pollinated *Banksia hookeriana*. *Heredity* **102**, 274-285.

Krauss SL, Dixon B, Dixon KW (2002) Rapid genetic decline in a translocated population of the endangered plant *Grevillea scapigera*. *Conservation Biology* **16**, 986-994.

Kudoh H, Shimamura R, Takayama K, Whigham DF (2006) Consequences of hydrochory in *Hibiscus*. *Plant Species Biology* **21**, 127-133.

Kudoh H, Whigham D (1997) Microgeographic genetic structure and gene flow in *Hibiscus moscheutos* (Malvaceae) populations. *American Journal of Botany* **84**, 1285-1285.

Laikre L (2010) Genetic diversity is overlooked in international conservation policy implementation. *Conservation Genetics* **11**, 349-354.

Larkin M, Blackshields G, Brown N, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948.

Leblois R, Rousset F, Estoup A (2004) Influence of spatial and temporal heterogeneities on the estimation of demographic parameters in a continuous population using individual microsatellite data. *Genetics* **166**, 1081-1092.

Lepais O, Petit R, Guichoux E, Lavabre J, Alberto F, Kremer A, Gerber S (2009) Species relative abundance and direction of introgression in oaks. *Mol Ecol* **18**, 2228-2242.

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

Linder HP (2008) Plant species radiations: where, when, why? *Philosophical Transactions of the Royal Society B: Biological Sciences* **363**, 3097-3105.

Linder HP, Hardy CR, Rutschmann F (2005) Taxon sampling effects in molecular clock dating: an example from the African Restionaceae. *Molecular Phylogenetics and Evolution* **35**, 569-582.

Liu Y, Wang Y, Huang H (2006) High interpopulation genetic differentiation and unidirectional linear migration patterns in *Myricaria laxiflora* (Tamaricaceae), an endemic riparian plant in the Three Gorges Valley of the Yangtze River. *American Journal of Botany* **93**, 206-215.

Llorens T, Ayre D, Whelan R (2004) Evidence for ancient genetic subdivision among recently fragmented populations of the endangered shrub *Grevillea caleyi* (Proteaceae). *Heredity* **92**, 519-526.

Lowe A, Harris S, Ashton P (2009) *Ecological Genetics: Design, Analysis, and Application*. Blackwell, Oxford, UK.

Maddison WP (1997) Gene trees in species trees. *Systematic Biology* **46**, 523-536.

Maddison WP, Knowles LL (2006) Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology* **55**, 21-30.

Maddison WP, Maddison D (2004) *Mesquite: a modular system for evolutionary analysis.* Version 1.01, http://mesquiteproject.org

Makinson R (2000) *Grevillea*. *Flora of Australia* **17A**, 1–460.

Mallet J (1995) A species definition for the modern synthesis. *Trends Ecol Evol* **10**, 294-299.

Manel S, Holderegger R (2013) Ten years of landscape genetics. *Trends in Ecology and Evolution* **28**, 614-621.

Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology and Evolution* **18**, 189-197.

Mantello C, Kestring DR, Sousa VA, Aguiar AV, Souza AP (2011) Development and characterization of microsatellite loci in *Grevillea robusta*. *BMC Proceedings* **5**, P16.

Markwith SH, Scanlon MJ (2007) Multiscale analysis of *Hymenocallis coronaria* (Amaryllidaceae) genetic diversity, genetic structure, and gene movement under the influence of unidirectional stream flow. *American Journal of Botany* **94**, 151-160.

Marshall JC, Arévalo E, Benavides E, Sites JL, Sites JW (2006) Delimiting species: comparing methods for Mendelian characters using lizards of the *Sceloporus grammicus* (Squamata: Phrynosomatidae) complex. *Evolution* **60**, 1050-1065.

Martin H (2006) Cenozoic climatic change and the development of the arid vegetation in Australia. *Journal of Arid Environments* **66**, 533-563.

Mast AR, Givnish TJ (2002) Historical biogeography and the origin of stomatal distributions in *Banksia* and *Dryandra* (Proteaceae) based on their cpDNA phylogeny. *American Journal of Botany* **89**, 1311-1323.

Mast AR, Jones EH, Havery SP (2005) An assessment of old and new DNA sequence evidence for the paraphyly of *Banksia* with respect to *Dryandra* (Proteaceae). *Australian Systematic Botany* **18**, 75-88.

Mast AR, Milton EF, Jones EH, Barker RM, Barker WR, Weston PH (2012) Time-calibrated phylogeny of the woody Australian genus *Hakea* (Proteaceae) supports multiple origins of insect-pollination among bird-pollinated ancestors. *American Journal of Botany* **99**, 472-487.

Mast AR, Willis CL, Jones EH, Downs KM, Weston PH (2008) A smaller *Macadamia* from a more vagile tribe: inference of phylogenetic relationships, divergence times, and diaspore evolution in *Macadamia* and relatives (tribe Macadamieae; Proteaceae). *American Journal of Botany* **95**, 843-870.

Mayr E (1942) *Systematics and the origin of species, from the viewpoint of a zoologist* Harvard University Press.

McCormack JE, Huang H, Knowles LL (2009) Maximum likelihood estimates of species trees: how accuracy of phylogenetic inference depends upon the divergence history and sampling design. *Systematic Biology* **58**, 501-508.

McGillivray DJ, Makinson RO (1993) *Grevillea, Proteaceae: a taxonomic revision* Melbourne University Press.

McLoughlin S, Hill R (1996) The succession of Western Australian Phanerozoic terrestrial floras. In: *Gondwanan heritage: Past, Present and Future of the Western Australian biota*, pp. 61-80.

Meglécz E, Costedoat C, Dubut V, Gilles A, Malausa T, Pech N, Martin J-F (2010) QDD: a user-friendly program to select microsatellite markers and design primers from large sequencing projects. *Bioinformatics* **26**, 403-404.

Meglécz E, Nève G, Biffin E, Gardner MG (2012) Breakdown of phylogenetic signal: a survey of microsatellite densities in 454 shotgun sequences from 154 non model eukaryote species. *PloS One* **7**, e40861.

Meirmans PG, Hedrick PW (2011) Assessing population structure: FST and related measures. *Molecular Ecology Resources* **11**, 5-18.

Merwin L, He T, Lamont BB (2012) Phylogenetic and phenotypic structure among *Banksia* communities in south-western Australia. *Journal of Biogeography* **39**, 397-407.

Meudt HM, Bayly MJ (2008) Phylogeographic patterns in the Australasian genus *Chionohebe* (*veronica,* Plantaginaceae) based on AFLP and chloroplast DNA sequences. *Molecular Phylogenetics and Evolution* **47**, 319-338.

Millar M, Byrne M, O'Sullivan W (2011) Corrigendum to: Defining entities in the *Acacia saligna* (Fabaceae) species complex using a population genetics approach. *Australian Journal of Botany* **59**, 507-507.

Millar M, Coates D, Byrne M (2013) Genetic connectivity and diversity in inselberg populations of *Acacia woodmaniorum*, a rare endemic of the Yilgarn Craton banded iron formations. *Heredity* **11**, 437-444.

Millar MA, Byrne M, Nuberg I, Sedgley M (2008) High outcrossing and random pollen dispersal in a planted stand of *Acacia saligna* subsp. *saligna* revealed by paternity analysis using microsatellites. *Tree Genetics & Genomes* **4**, 367-377.

Miralles A, Vences M (2013) New metrics for comparison of taxonomies reveal striking discrepancies among species delimitation methods in *Madascincus* lizards. *PloS One* **8**, e68242.

Mitsui Y, Isagi Y, Setoguchi H (2010) Multiple spatial scale patterns of genetic diversity in riparian populations of *Ainsliaea faurieana* (Asteraceae) on Yakushima Island, Japan. *Gondwanan heritage: Past, Present and Future of the Western Australian biota* **97**, 101-110.

Morgan MJ, Roberts JD, Keogh JS (2007) Molecular phylogenetic dating supports an ancient endemic speciation model in Australia's biodiversity hotspot. *Molecular Phylogenetics and Evolution* **44**, 371-385.

Myers N, Mittermeier RA, Mittermeier CG, Da Fonseca GA, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature* **403**, 853-858.

Nevill PG, Bradbury D, Williams A, Tomlinson S, Krauss SL (2014) Genetic and palaeo-climatic evidence for widespread persistence of the coastal tree species *Eucalyptus gomphocephala* (Myrtaceae) during the Last Glacial Maximum. *Ann Bot* **113**, 55-67.

Nilsson C, Brown RL, Jansson R, Merritt DM (2010) The role of hydrochory in structuring riparian and wetland vegetation. *Biological Reviews* **85**, 837-858.

Olde P, Marriott N (1994) *The Grevillea Book, Volume 3* Kangaroo Press, Kenthurst, NSW, Australia.

Oliver PM, Adams M, Lee MS, Hutchinson MN, Doughty P (2009) Cryptic diversity in vertebrates: molecular data double estimates of species diversity in a radiation of Australian lizards (*Diplodactylus*, Gekkota). *Proceedings of the Royal Society B: Biological Sciences* **276**, 2001-2007.

Omland KE, Baker JM, Peters JL (2006) Genetic signatures of intermediate divergence: population history of Old and New World Holarctic ravens (*Corvus corax*). *Molecular Ecology* **15**, 795-808.

Padial JM, Miralles A, De la Riva I, Vences M (2010) Review: The integrative future of taxonomy. *Frontiers in Zoology* **7**, 1-14.

Palmer GC, Bennett AF (2006) Riparian zones provide for distinct bird assemblages in forest mosaics of south-east Australia. *Biological Conservation* **130**, 447-457.

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* **28**, 2537-2539.

Pessoa EM, Alves M, Alves-Araujo A, Palma-Silva C, Pinheiro F (2012) Integrating different tools to disentangle species complexes: A case study in *Epidendrum* (Orchidaceae). *Taxon* **61**, 721-734.

Petit RJ, El Mousadik A, Pons O (1998) Identifying populations for conservation on the basis of genetic markers. *Conservation Biology* **12**, 844-855.

Petit RJ, Excoffier L (2009) Gene flow and species delimitation. *Trends in Ecology and Evolution* **24**, 386-393.

Pettit N, Froend R, Davies P (2001) Identifying the natural flow regime and the relationship with riparian vegetation for two contrasting western Australian rivers. *Regulated Rivers: Research & Management* **17**, 201-215.

Pharmawati M, Yan G, McFarlane I (2004) Application of RAPD and ISSR markers to analyse molecular relationships in *Grevillea* (Proteaceae). *Australian Systematic Botany* **17**, 49-61.

Phillips RD, Hopper SD, Dixon KW (2010) Pollination ecology and the possible impacts of environmental change in the Southwest Australian Biodiversity Hotspot. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**, 517-528.

Piry S, Alapetite A, Cornuet J-M, Paetkau D, Baudouin L, Estoup A (2004) GENECLASS2: a software for genetic assignment and first-generation migrant detection. *Journal of Heredity* **95**, 536-539.

Posada D (2008) jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution* **25**, 1253-1256.

Prentis PJ, Mather PB (2008) Fine-scale patterns of genetic variation indicate non-equilibrium gene frequency divergence in the stream lily, *Helmholtzia glaberrima*. *Freshwater biology* **53**, 973-980.

Prentis PJ, Vesey A, Meyers N, Mather P (2004) Genetic structuring of the stream lily *Helmholtzia glaberrima* (Philydraceae) within Toolona Creek, south-eastern Queensland. *Australian Journal of Botany* **52**, 201-207.

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

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

Rambaut A (2009) FigTree version 1.3. 1. *Computer program distributed by the author, website: http://tree. bio. ed. ac. uk/software/figtree/*

Rambaut A, Drummond A (2003) Tracer: a program for analysing results from Bayesian MCMC programs such as BEAST & MrBayes. Oxford, UK. http://evolve. zoo. ox. ac. uk/software. html.

Rannala B, Mountain JL (1997) Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences* **94**, 9197-9201.

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86**, 248-249.

Razafinarivo NJ, Guyot R, Davis AP, Couturon E, Hamon S, Crouzillat D, Rigoreau M, Dubreuil-Tranchant C, Poncet V, De Kochko A (2013) Genetic structure and diversity of coffee (*Coffea*) across Africa and the Indian Ocean islands revealed using microsatellites. *Annals of Botany* **111**, 229-248.

Renner SS (2005) Relaxed molecular clocks for dating historical plant dispersal events. *Trends in Plant Science* **10**, 550-558.

Rice WR (1989) Analyzing tables of statistical tests. *Evolution*, 223-225.

Richard M, Thorpe RS (2001) Can Microsatellites Be Used to Infer Phylogenies? Evidence from Population Affinities of the Western Canary Island Lizard (*Gallotia galloti*). *Molecular Phylogenetics and Evolution* **20**, 351-360.

Richardson MB, Ayre DJ, Whelan RJ (2000) Pollinator behaviour, mate choice and the realised mating systems of *Grevillea mucronulata* and *Grevillea sphacelata*. *Australian Journal of Botany* **48**, 357-366.

Rieseberg LH (1998) Molecular ecology of hybridization. *NATO ASI SERIES A LIFE SCIENCES* **306**, 243-266.

Rieseberg LH, Carney SE (1998) Plant hybridization. *New Phytologist* **140**, 599-624.

Ritz L, Glowatzki-Mullis ML, MacHugh D, Gaillard C (2000) Phylogenetic analysis of the tribe Bovini using microsatellites. *Animal genetics* **31**, 178-185.

Rix MG, Harvey MS (2012) Phylogeny and historical biogeography of ancient assassin spiders (*Araneae*: Archaeidae) in the Australian mesic zone: evidence for Miocene speciation within Tertiary refugia. *Molecular Phylogenetics and Evolution* **62**, 375-396.

Rosenberg NA (2002) The probability of topological concordance of gene trees and species trees. *Theoretical Population Biology* **61**, 225-247.

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes* **4**, 137-138.

Rossetto M, Allen CB, Thurlby KA, Weston PH, Milner ML (2012) Genetic structure and bio-climatic modeling support allopatric over parapatric speciation along a latitudinal gradient. *BMC Evolutionary Biology* **12**, 149.

Rossetto M, Crayn D, Ford A, Ridgeway P, Rymer P (2007) The comparative study of range-wide genetic structure across related, co-distributed rainforest trees reveals contrasting evolutionary histories. *Australian Journal of Botany* **55**, 416-424.

Rossetto M, Kooyman R, Sherwin W, Jones R (2008) Dispersal limitations, rather than bottlenecks or habitat specificity, can restrict the distribution of rare and endemic rainforest trees. *American Journal of Botany* **95**, 321-329.

Rousset F (2000) Genetic differentiation between individuals. *Journal of Evolutionary Biology* **13**, 58-62.

Rousset F (2008) genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources* **8**, 103-106.

Rozen S, Skaletsky H (1999) Primer3 on the WWW for general users and for biologist programmers. In: *Bioinformatics Methods and Protocols*, pp. 365-386. Springer.

Safner T, Miller MP, McRae BH, Fortin M-J, Manel S (2011) Comparison of Bayesian clustering and edge detection methods for inferring boundaries in landscape genetics. *International Journal of Molecular Sciences* **12**, 865-889.

Sampson J, Byrne M, Yates C, Gibson N, Thavornkanlapachai R, Stankowski S, MacDonald B, Bennett I (2013) Contemporary pollen-mediated gene immigration reflects the historical isolation of a rare, animal-pollinated shrub in a fragmented landscape. *Heredity* **112**, 172-181.

Schwartz MK, Luikart G, Waples RS (2007) Genetic monitoring as a promising tool for conservation and management. *Trends in Ecology and Evolution* **22**, 25-33.

Seddon G (2004) Sense of place: a response to an environment. The Swan Soastal Plain, Western Australia. In: *Sense of place: a response to an environment. The Swan coastal plain, Western Australia*. Bloomings Books, Melbourne.

Segelbacher G, Cushman SA, Epperson BK, Fortin M-J, Francois O, Hardy OJ, Holderegger R, Taberlet P, Waits LP, Manel S (2010) Applications of landscape genetics in conservation biology: concepts and challenges. *Conservation genetics* **11**, 375-385.

Shaffer HB, Thomson RC (2007) Delimiting species in recent radiations. *Systematic Biology* **56**, 896-906.

Shaw J, Lickey EB, Beck JT, Farmer SB, Liu W, Miller J, Siripun KC, Winder CT, Schilling EE, Small RL (2005) The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany* **92**, 142-166.

Shaw J, Lickey EB, Schilling EE, Small RL (2007) Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *American Journal of Botany t* **94**, 275-288.

Shepherd KA, Perkins A, Collins J, Byrne M, Thiele KR (2013) Morphological and molecular evidence supports the recognition of a new subspecies of the critically endangered *Pityrodia scabra* (Lamiaceae). *Australian Systematic Botany* **26**, 1-12.

Sinclair EA, Bezy RL, Bolles K, Crandall KA, Sites Jr JW (2004) Testing species boundaries in an ancient species complex with deep phylogeographic history: genus *Xantusia* (Squamata: Xantusiidae). *The American Naturalist* **164**, 396-414.

Sites Jr JW, Marshall JC (2003) Delimiting species: a Renaissance issue in systematic biology. *Trends in Ecology and Evolution* **18**, 462-470.

Sites Jr JW, Marshall JC (2004) Operational criteria for delimiting species. *Annual Review of Ecology, Evolution and Systematics* **35**, 199-227.

Slatkin M (1985) Gene flow in natural populations. *Annual Review of Ecology, Evolution and Systematics*, 393-430.

Smith J, Gross C (2002) The pollination ecology of *Grevillea beadleana* McGillivray, an endangered shrub from northern New South Wales, Australia. *Annals of Botany* **89**, 97-108.

Smith M (2013) Threatened and Priority Flora List for Western Australia. Department of Parks and Wildlife. Kensington, Western Australia.

Soltis PS, Gitzendanner MA (1999) Molecular systematics and the conservation of rare species. *Conservation Biology* **13**, 471-483.

Spinks PQ, Shaffer HB (2009) Conflicting mitochondrial and nuclear phylogenies for the widely disjunct *Emys* (Testudines: Emydidae) species complex, and what they tell us about biogeography and hybridization. *Systematic Biology* **58**, 1-20.

Storfer A, Murphy MA, Evans JS, Goldberg CS, Robinson S, Spear SF, Dezzani R, Delmelle E, Vierling L, Waits LP (2007) Putting the "landscape" in landscape genetics. *Heredity* **98**, 128-142.

Storfer A, Murphy MA, Spear SF, Holderegger R, Waits LP (2010) Landscape genetics: where are we now? *Molecular Ecology* **19**, 3496-3514.

Taberlet P, Gielly L, Pautou G, Bouvet J (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant molecular biology* **17**, 1105-1109.

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

Takezaki N, Nei M (2008) Empirical tests of the reliability of phylogenetic trees constructed with microsatellite DNA. *Genetics* **178**, 385-392.

Takezaki N, Nei M, Tamura K (2010) POPTREE2: Software for constructing population trees from allele frequency data and computing other population statistics with Windows interface. *Molecular Biology and Evolution* **27**, 747-752.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution* **30**, 2725-2729.

Tapper SL, Byrne M, Yates C, Keppel G, Hopper S, Van Niel K, Schut A, Mucina L, Wardell-Johnson G (2014) Isolated with persistence or dynamically connected? Genetic patterns in a common granite outcrop endemic. *Diversity and Distributions*.

Tautz D, Arctander P, Minelli A, Thomas RH, Vogler AP (2003) A plea for DNA taxonomy. *Trends in Ecology and Evolution* **18**, 70-74.

Tero N, Aspi J, Siikamäki P, Jäkäläniemi A (2005) Local genetic population structure in an endangered plant species, *Silene tatarica* (Caryophyllaceae). *Heredity* **94**, 478-487.

Tero N, Aspi J, Siikamäki P, Jäkäläniemi A, Tuomi J (2003) Genetic structure and gene flow in a metapopulation of an endangered plant species, *Silene tatarica*. *Molecular Ecology* **12**, 2073-2085.

Torres E, Iriondo JM, Escudero A, Pérez C (2003) Analysis of within-population spatial genetic structure in *Antirrhinum microphyllum* (Scrophulariaceae). *American Journal of Botany* **90**, 1688-1695.

Trénel P, Hansen MM, Normand S, Borchsenius F (2008) Landscape genetics, historical isolation and cross-Andean gene flow in the wax palm, *Ceroxylon echinulatum* (Arecaceae). *Molecular Ecology* **17**, 3528-3540.

Van Looy K, Jacquemyn H, Breyne P, Honnay O (2009) Effects of flood events on the genetic structure of riparian populations of the grassland plant *Origanum vulgare*. *Biological Conservation* **142**, 870-878.

Van Oosterhout C, Hutchinson WF, Wills DP, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* **4**, 535-538.

Van Valen L (1976) Ecological species, multispecies, and oaks. *Taxon*, 233-239.

Vaughton G (1998) Soil seed bank dynamics in the rare obligate seeding shrub, *Grevillea barklyana* (Proteaceae). *Australian Journal of Ecology* **23**, 375-384.

Wei X, Meng H, Jiang M (2013) Landscape Genetic Structure of a Streamside Tree Species *Euptelea pleiospermum* (Eupteleaceae): Contrasting Roles of River Valley and Mountain Ridge. *PloS One* **8**, e66928.

Wendel JF, Doyle JJ (1998) Phylogenetic incongruence: window into genome history and molecular evolution. In: *Molecular Systematics of Plants II*, pp. 265-296. Springer.

Wheeler M, Byrne M (2006) Congruence between phylogeographic patterns in cpDNA variation in *Eucalyptus marginata* (Myrtaceae) and geomorphology of the Darling Plateau, south-west of Western Australia. *Australian Journal of Botany* **54**, 17-26.

Wheeler M, Byrne M, McComb J (2003) Little genetic differentiation within the dominant forest tree, *Eucalyptus marginata* (Myrtaceae) of South-Western Australia. *Silvae Genetica* **52**, 254-258.

Wiens JJ (2007) Species delimitation: new approaches for discovering diversity. *Systematic Biology* **56**, 875-878.

Willis SC, Macrander J, Farias IP, Ortí G (2012) Simultaneous delimitation of species and quantification of interspecific hybridization in Amazonian peacock cichlids (genus *cichla*) using multi-locus data. *BMC Evolutionary Biology* **12**, 96.

Yates CJ, Ladd PG, Coates DJ, McArthur S (2007) Hierarchies of cause: understanding rarity in an endemic shrub *Verticordia staminosa* (Myrtaceae) with a highly restricted distribution. *Australian Journal of Botany* **55**, 194-205.

Appendix 1

Isolation, via 454 sequencing, characterization and transferability of microsatellites for *Grevillea thelemanniana* subsp. *thelemanniana* and cross-species amplification in the *Grevillea thelemanniana* complex (Proteaceae).

Tanya H. Hevroy[1], Michael L. Moody[2], Siegfried L. Krauss[3] and Michael G. Gardner[4]

[1]School of Plant Biology, University of Western Australia, 35 Stirling Ave, Crawley, Western Australia, 6009, Australia.

[2]Department of Biological Sciences, University of Texas at El Paso, 500 W. University Ave., El Paso, TX 79968, USA

[3]Kings Park and Botanic Garden, Botanic Gardens and Parks Authority, Fraser Ave, West Perth, Western Australia, 6005, Australia.

[4]School of Biological Sciences, Flinders University, GPO Box 2100, Adelaide 5001 South Australia, Australia; Australian Centre for Evolutionary Biology and Biodiversity, University of Adelaide, 5005 Adelaide; and Evolutionary Biology Unit South Australian Museum, Nth Tce Adelaide, South Australia Australia 5001.

Abstract

A set of 12 polymorphic microsatellite markers were developed for *Grevillea thelemanniana* subsp*. thelemanniana*, a rare and threatened shrub, endemic to southwest Western Australia. Genomic sequences were obtained from next generation (454) sequencing in the target species. Primer pairs for a total of 30 microsatellite loci were designed from these of which 13 were successfully amplified in 25 individuals from Kenwick Nature Reserve, Perth. Twelve loci were polymorphic with observed heterozygosity ranging from 0.12 to 0.6 and the number of alleles per locus ranged from 2 to 8. These markers were trialed in all species within the *Grevillea thelemanniana* species complex. Amplification success and level of polymorphism varied among loci and taxa, but most (96%) were successful. These loci will be useful in understanding the genetic variation, molecular ecology and phylogenetic relationships in the *Grevillea thelemanniana* complex.

Grevillea thelemanniana Endl. subsp. *thelemanniana* is a priority listed taxon in Western Australia (Priority 4 - Rare, near threatened and in need of monitoring) (Smith 2012). It is found only in a fragmented habitat southeast of Perth, Western Australia. The species is part of a larger complex comprised of 16 species and 4 subspecies, endemic to Western Australia and found primarily in the coastal sand plains from as far south as Busselton and north to Exmouth. The *Grevillea thelemanniana* complex has been recognized as a group with high conservation importance and poorly known taxonomic relationships (Makinson 2000). Here we develop 12 microsatellite markers for *Grevillea thelemanniana* subsp. *thelemanniana* that show amplification and polymorphism among other members of the complex. Relatively few microsatellite panels exist for *Grevillea*, and of these all are for eastern Australian species, making this the first library for a western Australian *Grevillea* (England et al 1999,Hoebee et al 2011, Mantello et al 2011, James et al 2012).

Genomic DNA was extracted from one individual of *G. thelemanniana* subsp. *thelemanniana* collected at Kenwick Nature Reserve. Extraction from freeze-dried, ground leaf material followed the modified CTAB extraction method (Doyle and Doyle 1987). The DNA solution (4µg) was sent to the Ramaciotti Centre for Gene Function Analysis (http://www.ramaciotti.unsw.edu.au) in Sydney for 454 shotgun sequencing on a Roche GS-FLX Titanium machine (454 Life Sciences, Branford, CT, USA) following the protocol of Gardner et al (2011). Sequences have been deposited in the Dryad Digital Repository (see Meglécz et al. 2012; http://dx.doi.org/10.5061/dryad.jd183) The resulting raw sequences (total length 101,949,487 bases) were searched for unique microsatellite motifs two to six base pairs long with C8 repeats (Gardner et al 2011) in QDD 1.3 software (Meglécz et al 2010). The program Primer3 1.1.1 (Rozen and Skaletsky 2000) within QDD 1.3 was used to design primers, using the basic search defaults. The resulting 501 unique loci were sorted based on PCR product size, repeat class, repeat length, GC content and multiplexing potential. The best 30 (15 dinucleotide and 15 trinucleotide) were chosen for evaluation of amplification and polymorphism. Reactions were performed with 10-50 ng of *G. thelemanniana* subsp. *thelemanniana* DNA in 10 µl reaction volumes containing 50 mM KCl, 20 mM Tris-HCl (pH 8.4), 0.2 mM each dNTP, 0.08 µM forward primer and reverse primer, 0.01 µl *Taq* DNA polymerase and 4 mM $MgCl_2$. Amplification was conducted in Eppendorf™

thermocyclers the following amplification conditions; denaturation for 3 min at 94°C was followed by 30cycles (25 in loci THP17) made up of 40s at the annealing temperature (56-62°C) and 30s at 72°C. The final step was elongation of 15min at 72°C. Of the 30 loci, 13 were selected for further testing and were amplified in two individuals from 3 populations of *G. thelemanniana* subsp. *thelemanniana*. An applied Biosystems DNA Analyzer (Beckman) was used to visualize fragments and genotypes were scored manually using Beckman Fragment Analyzer software. One primer was monomorphic and therefore excluded from further testing.

Genetic diversity of the remaining 12 loci was assessed for 25 individuals from the Kenwick Nature Reserve population (32° 0' 59.6" S 115° 59' 24.3" E (GDA94)). PCR amplification and genotyping was conducted as above. Hardy-Weinberg equilibrium (HWE), linkage disequilibrium (LD) and allelic richness were assessed using exact tests in Genepop 4.1.4 (Raymond and Rousset 1995; Rousset 2008) with Bonferroni correction. GenAlEx 6.4.1 (Peakall and Smouse 2006) was used to obtain genetic diversity parameters. We used MICROCHECKER 2.2.3 (Van Oosterhout et al. 2004) to check each locus for evidence of null alleles, scoring error due to stuttering, and large allele drop out. Primers for all loci were tested for amplification in 8 individuals each of all species of the *Grevillea thelemanniana* complex (Table 2).

There was no evidence for linkage disequilibrium at any locus, however deviation from Hardy Weinberg equilibrium was found for three loci (*P<0.01;* THP4 and THP30, *P<0.05*; THP9; Table1) with heterozgosity deficiency evident. All loci were independent and null alleles were only apparent in THP4 with an estimated frequency of 0.145. All loci showed polymorphism within the population. The number of alleles per locus ranged from two to eight, and effective number of alleles per locus ranged from 1.2 to 3.7. Observed heterozygosity ranged from 0.120 to 0.660 and expected heterozygosity from 0.18 to 0.73. The fixation index ranged from -0.143 to 0.343 for loci in HWE. Cross species amplification was successful in all species for 7 of 12 (58%), loci of the *Grevillea thelemanniana* complex and allele number ranged from one to seven (Table 2). The microsatellite markers developed here will be used to evaluate taxonomic hypotheses in the *Grevillea thelemanniana* complex and evaluate patterns of landscape genetic structure for populations of *G. thelemanniana* ssp. Cooljarloo (B.J. Keighery 28 B) in Western Australia.

Table 1 Primer sequences and characteristics of 12 microsatellite loci isolated from *G. thelemannaiana subsp thelemanniana*

Locus	Primer Sequence (5'-3')	GenBank no.	Repeat Motif	Temp (C°)	Size range (bp)	N	Ne	*Ho*	*He*	F	PHW	Null allele
THP3	F:CAACTTGAATCAGACAAAGATGTG R:CAACTTGAATCAGACAAAGATGTG	GLMI75201DP2HD	(GTT)7	56	91-103	3	1.326	0.200	0.246	0.186	0.287	
THP4	F:GGTCGTGGTACTTGGTGGTC R:GGTCGTGGTACTTGGTGGTC	GLMI75201BFWLD	(GA)8	62	174-200	7	3.165	0.440	0.684	0.357	0.0*	0.145
THP5	F:TGGAAATCGATTCGAAGAGG R:TGGAAATCGATTCGAAGAGG	GLMI75201C8T3I	(TC)8	62	170-172	2	1.625	0.440	0.385	-0.143	0.880	
THP7	F:TACCCAGTTTCTTGGCCTTG R:TACCCAGTTTCTTGGCCTTG	GLMI75202IDULT	(TCT)8	56	100-106	2	1.220	0.120	0.180	0.327	0.198	
THP8	F:AGAAGCACCCATTGATTTGG R:AGAAGCACCCATTGATTTGG	GLMI75202G91AB	(AAC)8	62	149-158	3	2.163	0.600	0.538	-0.116	0.772	
THP9	F:TTTCCTGGACTCTCCGAATG R:TTTCCTGGACTCTCCGAATG	GLMI75201CCFRY	(GT)8	62	310-314	3	1.874	0.360	0.466	0.343	0.017*	
THP13	F:CCAGAGTGTGAACCTCAGCA R:CCAGAGTGTGAACCTCAGCA	GLMI75201CRJRK	(TA)10	62	98-106	8	3.704	0.667	0.730	0.087	0.209	
THP17	F:CATGTTCAAATTCCGTCGAA R:CATGTTCAAATTCCGTCGAA	GLMI75202IBW27	(CT)12	62	172-208	7	2.723	0.560	0.633	0.115	0.168	
THP20	F:CCTCCTTGGCTGCAATACTC R:CCTCCTTGGCTGCAATACTC	GLMI75201AV2UW	(TTC)15	62	130-154	7	2.404	0.560	0.584	0.041	0.058	
THP22	F:CATCTGCTGGTTCGTTTGTG R:CATCTGCTGGTTCGTTTGTG	GLMI75202IBW27	(TC)6	62	154-160	2	1.419	0.280	0.295	0.051	0.583	
THP24	F:AGAAGCACCCATTGATTTGG R:AGAAGCACCCATTGATTTGG	GLMI75202G91AB	(AG)8	62	152-158	3	2.080	0.560	0.519	-0.079	0.692	
THP30	F:GCCTTGTGTACTCACCACAGC R:GCCTTGTGTACTCACCACAGC	GLMI75202GYCOA	(CT)9	62	116-124	3	1.442	0.280	0.306	0.086	0.021*	

Allele size range, number of alleles (N), number of effective alleles (Ne), observed and expected heterozygosity (Ho, He), fixation index (F), Hardy Weinberg P values (PHW) and null allele frequencies are based on 25 individuals. * (P<0.05) **(P<0.01) Departure from Hardy-Weinberg equilibrium.

Table 2 PCR product size and observed number of alleles (in parenthesis) for cross species comparison of 12 microsatellite loci. An X means no amplification was detected.

Locus	THP1	THP3	THP4	THP5	THP7	THP8	THP9	THP13	THP17	THP20	THP22	THP30
G.acropogon	154-166(2)	X	188-196(5)	149(1)	100-115(3)	149-158(3)	318(1)	83-97(3)	183-192(4)	98(1)	153-156(2)	122-124(2)
G.delta	171-174(2)	100-115(5)	182-184(2)	167-171(3)	109-133(7)	153-158(3)	310-314(2)	90-94(3)	172(1)	127-148(6)	154-160(3)	110-136(5)
G.evanescens	166(1)	97-100(2)	174-186(4)	169(1)	106-109(2)	155-158(2)	314(1)	96-106(5)	172-187(2)	130-151(4)	148-163(5)	116-130(4)
G.exposita	162-171(4)	97-100(2)	184-188(3)	165-169(3)	94-112(4)	155-158(2)	312(1)	90(1)	171-204(2)	145-160(4)	151-157(2)	116-132(3)
G.fililoba	100-109(5)	100-109(3)	180-190(3)	167-171(3)	100-115(5)	149-158(3)	310-312(2)	90-104(3)	172-190(5)	127-160(6)	151-160(4)	116-122(2)
*G.*sp. Gillingara	171(1)	100-106(3)	186(1)	169(1)	100-115(3)	158(1)	312(1)	84-90(2)	171-174(2)	148-151(2)	154-157(2)	132(1)
G.hirtella	171-180(2)	96-99(2)	174-188(6)	167-171(2)	94-115(4)	X	310-312(2)	90-96(2)	171-189(3)	140-158(3)	147-159(4)	116-130(6)
G.humifusa	171-174(2)	94-103(4)	180-184(3)	171(1)	106-118(3)	155-158(2)	310(1)	90(1)	172(1)	137-158(4)	153-159(3)	122(1)
G.maccutcheonii	189(1)	X	X	159-169(3)	112-115(2)	155(1)	X	90-94(2)	171-186(2)	98(1)	156(1)	118-122(2)
G.obtusifolia	166(1)	100-103(2)	176-196(5)	169-171(2)	100-109(3)	155-158(2)	314-318(3)	94-110(5)	171-204(3)	127-142(3)	154-160(3)	116-124(4)
G.olivacea	162-177(4)	100-103(2)	182-186(2)	167(1)	100(1)	158-167(3)	314(1)	90-94(3)	169-175(3)	120-149(4)	154-160(3)	116(1)
G.pinaster	151-174(6)	97-109(6)	165-169(3)	165-173(4)	97-112(5)	155(1)	310-312(2)	90-94(3)	168-192(5)	130-158(5)	148-160(5)	114-126(5)
G. preissii subsp *preissii*	166(1)	100-115(3)	182-198(3)	167-171(3)	100(1)	155-158(2)	310-314(2)	90-98(4)	171-198(3)	130-160(8)	153(1)	116-122(4)
G. preissii subsp *glabrilimba*	171-177(3)	100-115(3)	170-196(6)	169-175(3)	100-109(2)	155(1)	310-312(2)	90-102(4)	172(1)	137-164(8)	153(1)	118-124(4)
G.ripicola	165-168(3)	X	181-185(3)	149(1)	100-109(4)	X	317-323(2)	88-100(4)	189(1)	92-98(2)	260(1)	X
G. thelemanniana subsp Cooljarloo	165-171(2)	97-103(2)	182-196(7)	163-171(4)	106-118(5)	155(1)	310-318(3)	90-96(4)	172(1)	143-158(7)	154-160(3)	116-124(3)
G.variifolia subsp *variifolia*	150-177(4)	X	156-196(6)	157(1)	109-115(3)	155-158(2)	321-331(4)	90-92(2)	171(1)	120-144(4)	154-160(3)	116-122(2)
G.variifolia subsp *bundera*	166-169(2)	88-103(5)	191-221(9)	147-159(3)	82-109(5)	177(1)	309-327(7)	88-90(2)	171(1)	123-135(5)	154-157(4)	116-122(2)
G.stenomera	160(1)	91-100(2)	175-187(4)	165-171(3)	94-100(2)	155-158(2)	310(1)	89(1)	177-198(3)	132-141(4)	147(1)	114-136(3)

References

Doyle J, Doyle J (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:11-15

England PR, Ayre DJ, Whelan RJ (1999) Microsatellites in the Australian shrub *Grevillea macleayana* (Proteaceae). Mol Ecol 8:685-702

Gardner MG, Fitch AJ, Bertozzi T, Lowe AJ (2011) Rise of the machines-recommendations for ecologists when using next generation sequencing for microsatellite development. Mol Ecol Resour 11:1093-1101

Hoebee SE (2011) Development and cross-species amplification of microsatellites markers from the endangered Wee Jasper Grevillea (*Grevillea iaspicula,* Proteaceae). Muelleria 29(1):93-96

James EA, Brown GK, Citroen R, Blacket MJ (2012) Microsatellite development for two species of holly-leafed *Grevillea* and cross-species amplification in the *Asplendiifolia/Hookeriana* subgroup (Proteaceae). Cons Gen Res 4:137-140

Mantello C, Kestring DR, Sousa VA, Aguiar AV, Sousa AP (2011) Development and characterization of microsatellite loci in *Grevillea robusta.* BMC Proceedings 5:16

Makinson RO (2000) *Grevillea. Flora of Australia* ***17A***. 1-460

Meglécz E, Costedoat C, Dubut V, Gilles A, Malausa T, Pech N, Martin J-F (2010) QDD: a user-friendly program to select microsatellite markers and design primers from large sequencing projects. Bioinformatics 26:403–404

Meglécz, E, Nève, G, Biffin, E, and Gardner, MG (2012) Breakdown of Phylogenetic Signal: A Survey of Microsatellite Densities in 454 Shotgun Sequences from 154 Non Model Eukaryote Species. PLoS ONE 7(7), e40861

Oosterhout CV, Hutchinson WF, Wills DPM, Shipley P (2004) Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Notes 4:535–538

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes 6:288–295

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J Hered 86:248–249

Rousset F (2008) Genepop '007: a complete re-implementation of the GENEPOP software for Windows and Linux. Mol Ecol Resour 8:103–106

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics methods and protocols: methods in molecularbiology. Humana Press, Totowa, pp 365–386

Smith, M.G. (2012). *Threatened and Priority Flora List for Western Australia, 27 April 2012*. Dept. of Environment and Conservation. Como, W.A.