THE UNIVERSITY
of EDINBURGH

New Approaches for the Conservation Genomics of the Genus *Cycas* L. in Australia

---

By James A. R. Clugston

In Fulfilment of the Requirements for the Doctor of Philosophy in Molecular Plant Sciences

College of Science and Engineering
School of Biological Sciences

The University of Edinburgh
2019

Acknowledgements

I wish to extend my heartfelt thanks to my incredible supervisors: Nathalie Nagalingum, Greg Kenicer, Murray Henwood and Richard Milne for their support throughout the project. Nathalie for taking me on as a PhD student, and her help in setting me up when we first moved to Australia and her extended support and advice; Greg for unwavering support provided throughout the project no matter where I was in the world; Richard for providing expert advice and support for the research and Murray who agreed to supervise me so late in the process as I moved permanently to Australia and going above and beyond by providing us with a place to stay, helping get us settled and providing such important advice that has been instrumental to the final product of this project.

I wish to acknowledge funding received from the Australian Flora Foundation, the Australasian Systematic Botany Society's Hansjörg Eichler Scientific Research Fund, and The Nature Conservatory – The Thomas Foundation for an Australian Conservation Taxonomy Award. Furthermore, I wish to acknowledge the Biotechnology and Biological Sciences Research Council (BBSRC) UK and the EASTBIO Doctoral Training Partnership for providing me a studentship at the University of Edinburgh.

I thank staff at the Royal Botanic Gardens and Domain Trust Sydney, in particular: Carolyn Connolly for her support in the molecular laboratory, Hannah McPherson for helping in the initial quality testing of data and troubleshooting population genetics scripts, Peter Wilson for his help working with the computer systems and allowing me to run additional RADseq analysis and Maureen Phelan and Simon Goodwin for assistance in obtaining DNA samples from the living collections. I also thank the various botanic gardens including Montgomery Botanic Center, University California Berkley Botanic Garden, Darwin Botanic Garden that provided and allowed me to collect plant tissue samples of many species. At the Royal Botanic Garden Edinburgh, I wish to thank Laura Forrest and Michelle Hart for their continued support throughout the project. And lastly Duncan Reddish at the Royal Botanic Garden Edinburgh for his assistance in setting up a special server to run the *Cycas calcicola* dataset on.

Daren Eaton is acknowledged for assistance with ipyrad: Robert J. Toonen at the University of Hawaii for advice and information about the ezRAD protocol, and Todd McLay at the University of Melbourne for support in quality checking and filtering of RADseq data. Charles Foster at the University of Sydney for his advice and support and Ian Cowie, Nicholas Cuff and David Liddle at the Northern Territory Herbarium for their assistance during my field work and the collection of an additional sample of *Cycas armstrongii* to strengthen our results. Joe Russack at the California Academy of Sciences for his assistance with using Amazon Web Services (AWS). Athena Lam and Boni Cruz for their assistance in the laboratory at the California Academy of Science and in many cases just going above and beyond their jobs!

Joe Perner at Cycad International in Katherine for his help with finding population of *Cycas calcicola* in the Katherine Region. Louisa Murray for her continued support during the length of the project and helping me and my partner out during our arrival to Sydney. Also, to my friends and colleagues in the office Yola Metti and Trevor Wilson who were always there for support and up for helping out.

Markus Ruhsam at the Royal Botanic Garden Edinburgh and Alan Meerow of the United States Department of Agriculture for examining the thesis and providing significant input to help improve the results and outputs of this project.

I must also thank my parents and grandfather for supporting me both financially and mentally during this long period and lastly (but not least) I wish to thank my partner Lyndon for his unwavering support, advice and assistance even during the hardest of times and putting up with me in general.

Declaration

I declare that this thesis has been composed solely by myself, James A. R. Clugston, and that it has not been submitted, in whole or in part, in any previous application for a degree. Except where states otherwise by reference or acknowledgment, the work presented is entirely my own.

Throughout this thesis I use the word (“we or our”) to denote that all future publication coming from this work will be written by myself and co-authored by my supervisors and/or others. For the purposes of this thesis, all work has been carried out by me, and the final product produced by myself only.

Additionally, Chapter - 2 of this final thesis (RADseq as a valuable tool for plants with large genomes—a case study in cycads) has been accepted by Molecular Ecology Resources for publication and available as early access article.

James A. R. Clugston
24 October 2019

Abstract

Cycad species exist as small fragmented populations, therefore understanding their genetic variation is imperative for their conservation to ensure their long-term survival. Genetic data plays a fundamental role in identifying genotypes and detecting populations with the highest genetic diversity. This project uses next generation sequencing (NGS) and restriction associated DNA sequencing (RADseq) to identify thousands of genome-wide polymorphisms from populations of selected cycad species from the Northern Territory, Australia, namely: *Cycas armstrongii*, *Cycas calcicola*, *Cycas maconochiei* ssp. *maconochiei* and the interspecific hybrid *C. armstrongii* x *maconochiei*. RADseq was used to determine intra- and interspecific genetic variation in populations, verify the putative hybrid, recognize populations of conservation priority and determine if botanic garden collections currently represent the genetic diversity inherent in the wild. *Cycas calcicola* showed very low levels of genetic diversity and high inbreeding, and although there was significant geographic partitioning between populations in the Katherine and Litchfield National Park regions, which correlated with genetic differentiation. Additionally, the results showed that *C. calcicola* was not genetically, well represented in ex-situ collections. The genomic diversity of *Cycas armstrongii*, *C. maconochiei* ssp. *maconochiei* and *C. armstrongii* x *maconochiei* differs from that of *C. calcicola* and shows very low levels of genetic diversity yet generally with lower levels of inbreeding. The results show little genetic distance between *Cycas armstrongii* and *C. maconochiei* ssp. *maconochiei*, the most likely explanation is that they represent morphological extremes of a single species. The results from RADseq have far reaching significance for the conservation of cycads. In the case of *C. calcicola*, a far more structured acquisition of genetic material will be required if the full genetic diversity of this species is to be preserved in ex-situ collections.

Lay Summary

More than sixty percent of Cycad species are threatened, with many existing in small and isolated populations. As a result, understanding the diversity of cycads is imperative for their conservation to ensure their long-term survival. In order to understand the diversity of cycads, genetics plays a fundamental role in helping us to identify how populations differ from one another. Australia represents a diversity hotspot for cycads where there are many different species and also many large, clustered and undisturbed populations. This project uses the latest DNA sequencing technologies to understand the genetic diversity of populations of selected cycad species from the Northern Territory in Australia, namely: *Cycas armstrongii*, *C. calcicola*, *C. maconochiei* ssp. *maconochiei* and a hybrid population *C. armstrongii* x *maconochiei*. Our results find that populations of *C. calcicola* showed evidence of inbreeding and low genetic diversity, and geographic distance, populations in the Katherine and Litchfield National Park regions were confirmed based genetic evidence. The results also showed that genetic diversity of *C. calcicola* was not well represented in botanic garden collections, presenting conservation concerns. *Cycas armstrongii* and *C. maconochiei* ssp. *maconochiei* populations also showed low levels of genetic diversity in *C. calcicola*, but less inbreeding. In addition, based on the DNA evidence were found no genetic difference between *C. armstrongii* and *C. maconochiei* ssp. *maconochiei*, despite looking very different. Furthermore, these findings show that the notion of a hybrid between the two species (*C. armstrongii* x *maconochiei*) is invalid. Our results will have far-reaching significance for the conservation of vulnerable populations of cycads. In the case of *C. calcicola*, a far more structured acquisition of seeds from the wild will be required so that the species can be preserved in botanic gardens.

List of Tables and Figures

List of Tables

Table 1 Estimated genome sizes of taxa examined in earlier RADseq studies..... 23

Table 2 Samples of cycad genera obtained for RADseq. 26

Table 3 Results of mapping RADseq reads against reference plastomes. 32

Table 4 RADseq data for Cycas spp., mapped to a reference mitochondrial genome of *Cycas taitungensis*. 33

Table 5 Outputs from de-novo assembly of RADseq data. 33

Table 6 Filtering of loci during de-novo assembly of Cycas calcicola natural populations.. 34

Table 7 Population demographics for Cycas calcicola populations in Northern Territory. 47

Table 8 Summary of population genetic statistics for all populations of *Cycas calcicola*. 53

Table 9 Analysis of molecular variance for *Cycas calcicola* populations. 54

Table 10 Pairwise distance based FST matrix of *Cycas calcicola* populations.. 54

Table 11 Summary of samples collected from natural populations of *Cycas armstrongii* complex in Northern Territory Australia. 72

Table 12 Summary of population genetic statistics for populations of *Cycas armstrongii*. 77

Table 13 Summary of population genetic statistics for populations of *Cycas maconochiei* subsp *maconochiei* and *C. armstrongii* x *maconochiei*. 78

Table 14 Analysis of molecular variance for *Cycas armstrongii* complex. 80

Table 15 Pairwise FST matrix of *Cycas armstrongii* populations. 81

Table 16 Pairwise FST matrix of *Cycas maconochiei* ssp. *maconochiei* populations. ... 81

List of Figures

Figure 1 *Bowenia spectabilis* growing in habitat in the Daintree Rainforest, Queensland 2

Figure 2 *Cycas platyphylla* growing in habitat in the dry Petford District in Queensland . 2

Figure 3 World distribution of cycad genera 3

Figure 4 Loosely arranged megasporophylls of *Cycas maconochiei* subsp. *maconochiei* growing in habitat. 4

Figure 5 Unrooted tree network tree of *C. calcicola* wild populations in Northern Territory, Australia. 36

Figure 6 *Cycas calcicola* populations growing in the wild in the Katherine region and Litchfield National Park 45

Figure 7 Distribution of samples of C. calcicola in Northern Territory 46

Figure 8 Population structure plot for *Cycas calcicola* populations 55

Figure 9 DAPC graph of *Cycas calcicola* populations 56

Figure 10 Map of samples collected for *Cycas armstrongii* and *C. maconochiei* 66
Figure 11 *Cycas armstrongii* populations growing in the Darwin region 67
Figure 12 *Cycas maconochiei* wild populations growing in the Darwin Cox Peninsula.. 68
Figure 13 Population structure plot for *Cycas armstrongii* populations 82
Figure 14 Population structure plot for *Cycas maconochiei* populations..... 83
Figure 15 Population structure plot for *Cycas armstrongii, C. maconochiei* ssp. *maconochiei* and *Cycas armstrongii x maconochiei* 83
Figure 16 DAPC graph of *Cycas armstrongii* populations represented as IBRA subregions..... 84
Figure 17 DAPC graph of *Cycas maconochiei* ssp. *maconochiei* grouped into populations..... ..85
Figure 18 DAPC graph of *Cycas armstrongii, C. maconochiei* ssp. *maconochiei* and *Cycas armstrongii x maconochiei* population group by taxon 85

Table of Contents

Chapter 1 - Introduction 1

1.1 The cycads 1

1.2 Conservation genetics and genomics 10

1.3 Cycad conservation and population genetics 13

1.4 Restriction Site Associated DNA Sequencing 14

1.5 Unresolved questions in conservation genetics 17

1.6 Aims and Objectives 19

Chapter 2 - RADseq as a valuable tool for plants with large genomes—a case study in cycads 21

2.1 Introduction 22

2.2 Materials and methods 25

2.3 Results 31

2.4 Discussion 37

Chapter 3 - Conservation Concerns: Low Genetic Diversity in the Australian cycad, *Cycas calcicola* and the Absence of Key Genotypes in Botanic Gardens 40

3.1 Introduction 41

3.2 Materials and methods 47

3.3 Results 52

3.4 Discussion 57

Chapter 4 - Next-generation population genetics in Australian *Cycas* sheds light on the origins of *C. armstrongii* and *C. maconochiei* ssp. *maconochiei*. 63

4.1 Introduction 64

4.2 Materials and methods 71

4.3 Results 76

4.4 Discussion 86

Chapter 5 - Conclusion 95

References 100

Appendix I. Table of wild collected Cycas calcicola samples for RADseq. 139

Appendix II. Detailed RADseq protocol to support “RADseq as a valuable tool in plants with large genomes—a case study in cycads” 141

Chapter 1 - Introduction

1.1 The cycads

Cycadales have long been considered the most ancient lineage of extant seed plants, as they have the oldest fossil record of all gymnosperms (Norstog & Nicholls 1997). The cycad fossil record dates back to around 300 million years ago (Martínez et al. 2012). Cycads were at their peak of diversity during the Jurassic and Cretaceous periods (199.6-65.5 million years ago) and are often considered 'living fossils' (Norstog & Nicholls 1997). Despite their age as a group, the cycad species of today represent a series of recent radiations no more than 12 million years old (Nagalingum et al. 2011).

Living Cycadales are monophyletic (Rai et al. 2003), although some fossil Bennetitales which were morphologically similar have been erroneously included in the group in the past. The group as defined today consists of around 353 species that are split between two morphologically and phylogenetically differentiated families. Cycadaceae is monotypic, containing only *Cycas* L. (116 spp.). Zamiaceae, the larger family, contains nine genera: *Bowenia* Hook. ex Hook.f. (2 spp.), *Ceratozamia* Brongn. (31 spp.), *Dioon* Lindl (14 spp.), *Encephalartos* Lehm. (65 spp.), *Lepidozamia* Regel (2 spp.), *Macrozamia* Miq. (41 spp.), *Microcycas* (Miq.) A.DC. (1 sp.), *Stangeria* T. Moore (1 sp.), and *Zamia* L. (78 spp.) (Calonje et al. 2019). Salas-Leiva et al. (2013) provided strong support for the backbone of the Cycadales. However, a comprehensive high-level and deep phylogeny of the group is still to be resolved (Nagalingum et al. 2011).

Figure 1 *Bowenia spectabilis* growing in habitat in the Daintree Rainforest, Queensland, Australia.

Figure 2 *Cycas platyphylla* growing in habitat in the dry Petford District in Queensland, Australia.

Cycads are predominantly woody, long-lived and slow-growing perennials found in a variety of habitats from closed forests (Figure 1) to deserts (Figure 2), with most species in tropical wet forests, dry forests and savannah (Norstog & Nicholls 1997). Some fire-prone savannah species form geoxylic suffrutexes, but most species have an aerial caudex and their woody stems are formed from leaf bases, so they typically have a central pith and do not form true secondary wood (Cousins et al. 2013).

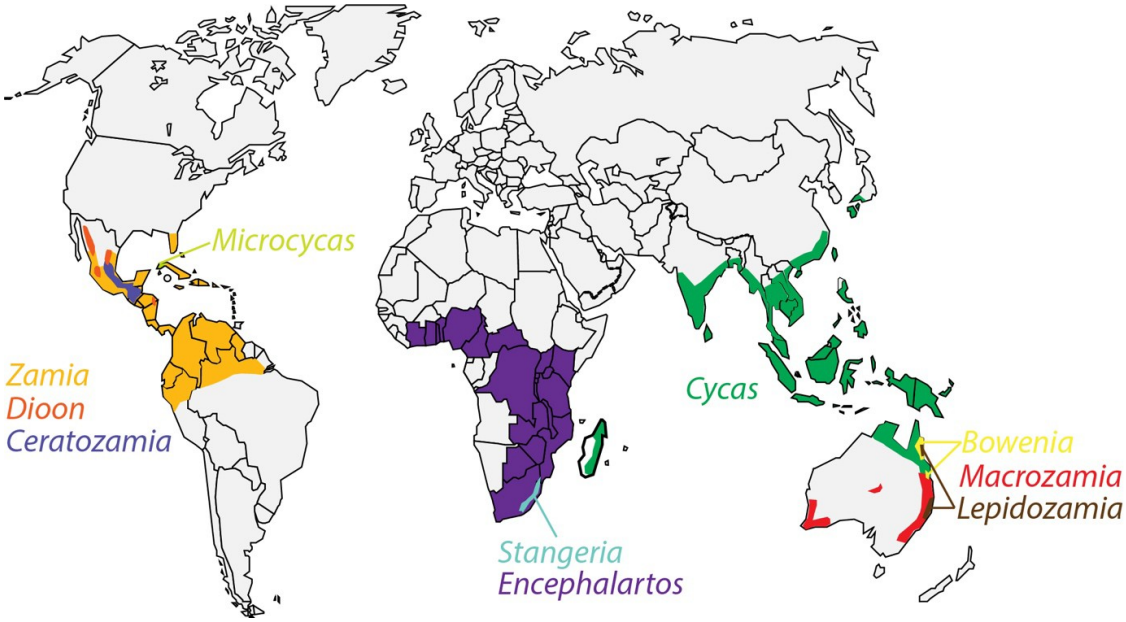

Figure 3 World distribution of cycad genera, image Nagalingum et al., 2011

Cycadales have a pantropical distribution (Figure 3), and are found throughout the tropical Americas, Africa, Asia and Australasia. Cycads offer some of the highest levels of endemism found in any plant group, with endemics found all over the world, including China, Australia, South Africa, Central Africa and India.
Many species are also considered island endemics and, often endemic to individual provinces, especially in Mexico and South America (Cibrián-Jaramillo et al. 2010; Long-Qian et al. 2004; Sosa et al. 1998).

The genus Cycas L.

Figure 4 Loosely arranged megasporophylls of *Cycas maconochiei* growing ssp. *maconochiei* in habitat.

Cycas L. (Cycadaceae) *sensu stricto* is the largest genus of the cycads and represents the earliest divergent lineage from the remainder of the cycads. It consists of 116 extant species and six subspecies (Calonje et al. 2019; Stevenson 1992). Megasporophylls in Cycadaceae are loosely arranged on the main axial spiral of the stem and do not form a distinct megasporangiate cone or strobilus (Figure 4). Comparatively, Zamiaceae megasporophylls are connected to a central axis forming a discrete cone/strobilus. Cycadaceae has a lesser degree of organisation than found in Zamiaceae, and the loose arrangement of the megasporophylls in *Cycas* is considered a more ancestral character (Lindstrom & Hill 2007). Other key identifying characteristics of Cycadaceae (and *Cycas*) that differentiate them from Zamiacae are circinate emerging leaves and leaflets, and platyspermic seeds vs. the radiospermic seeds of Zamiacae (Stevenson 1992). Hill (2003) classified *Cycas* into five taxonomic sections based on morphology: Asiorientales, Cycas, Wadeanae, Indosinenses and Stangerioidies. The sections

within *Cycas* can be differentiated based on the external surface of the ovule, megasporophyll lamina shape, microsporangiate strobilus morphology, terminal bullae of microsporophylls, structure of sclerotesta, and the terminal lamina of megasporophyll. However, more recent molecular data (Liu et al. 2018; Nagalingum et al. 2011; Sangin et al. 2006;) do not support all sections of *Cycas* as some are polyphyletic, and the morphology does not always agree with the molecular relationship among the taxa, with the exception of a few characters. As with many of the other genera in the Cycadales, the phylogenetic relationships within *Cycas* remain to be fully resolved due to the inability of current molecular markers to resolve between species and infra-species levels (Nagalingum et al. 2011).

The leaves and leaflets of *Cycas* can be highly informative characters, particularly in the epidermis, with leaf cuticles potentially able to differentiate taxa within the genus (Mickle et al. 2011). A more recent study by Griffith et al. (2014) found that some leaf micromorphological characters show strong correlations with the phylogenetic relationship of species *sensu* Nagalingum et al. (2011), and these characters are often synapomorphic within a phylogenetic species group.

Distribution. *Cycas* is the most ecologically and morphologically diverse genus within the Cycadales, and has the widest geographic distribution of any cycad (Osborne et al. 2012). *Cycas* can be found throughout Madagascar, India, Sri Lanka to China, Japan (Ryukyu islands, *Cycas revoluta* Thunb.), Thailand, Malaysia, Indonesia, Papua New Guinea (Pacific islands), Australia and New Caledonia (Figure 3) (Chaw et al. 2005). In Australia, *Cycas* is represented by 38 of the 117 cycad species, with most Australian species being endemic (Osborne et al., 2012). The genus is distributed in Western Australia (three species), Northern Territory (16 species) and Queensland (19 species).

Phenology and pollination. The main means of cycad pollination was once assumed to be anemophily, but has subsequently been shown to be entomophily (Terry et al. 2012; Schneider et al. 2002; Norstog & Nicholls 1997). However, the open cone morphology in *Cycas*, where the megasporophylls are loosely arranged and not connected to central cone axis (unlike Zamiaceae, *Cycas* does not form a true cone) means there is some potential for anemophily (Keppel 2001). In comparison with species in closed canopy habitats, anemophily may account for pollination in primarily coastal (e.g. *Cycas micronesica* K.D. Hill and *Cycas seemannii* Braun) and open grassland species (e.g. *Cycas wadei* Merr.) (Terry et al. 2009).

Many species of cycad have a specific pollinator association (Taylor et al. 2012). The array of insects known to pollinate cycads ranges from weevils in the families Brentidae, Curculionidae and Boganiidae (Brookes et al. 2015; Procheş & Johnson 2009; Hall et al. 2004; Vovides et al. 1997; Tang 1987), thrips (Mound and Terry 2001) and some species of bees (Schneider et al. 2002; Forster et al. 1994).

Cycad pollinators are involved in a push-pull pollination system (Terry et al. 2007). This process has been best documented in the South African cycad *Encephalartos villosus* Lem., where beetles swarm and feed upon the microstrobili which emit volatiles (e.g. pyrazines) and heat up to 12°C above ambient temperature in a process called thermogenesis (Suinyuy et al. 2012). The emission of volatiles and thermogenesis act as the initial pollinator push from the microstrobili to the megastrobili where pollination occurs. Once the male cone stops producing volatiles and cools down to ambient temperature, the pollinators return to the microstrobili and the process repeats (Suinyuy et al. 2013). All cycad species are known to have synchronous phenology, the pattern of which can be used to differentiate between species (Clugston et al. 2018.; Ornduff 1992). This is especially important for the push-pull pollination, since the phenology of each species needs to be synchronised for successful pollination (Taylor et al. 2012; Griffith et al. 2012).

Seed dispersal. Today's cycad species have very few seed distributors and low rates of dispersal, mainly due to the size and toxicity of the seeds (Nadarajan et al. 2018). However, the known distributors for seeds of cycads are mostly large mammals such as large rodents, possums, peccaries and other large herbivores (Snow & Walter 2007; Farrera & Vovides 2004). One good example is in northern Australia, where cassowaries have been known to eat the seeds of both *Bowenia* and *Lepidozamia*, thereby acting as dispersal agents for both species (Hall & Walter

2014). But for many species of cycads the seeds rarely distribute greater then 5 m (Hall & Walter 2013). The most likely cycad seed dispersal agents were once the now extinct species of megafauna, which would have been the main dispersal mechanisms for species of *Cycas* and *Macrozamia* in some parts of Australia (Ingham et al. 2013). However, some species of *Cycas,* particularly species in *Cycas rumphii* complex (section Rumphiae) have a spongy air-filled layer within the sclerotesta of the seeds that acts as buoyancy aid (Hill 1994). This allows the seeds to float on water and aid in their dispersal (Nadarajan et al. 2018, Dehgan and Yuen 1983). Species within the *Cycas rumphii* complex have some of the widest distribution of any groups of cycads throughout Indonesia, New Guinea and the western Pacific (Keppel et al. 2008; Hill 1994).

Threats posed to cycad populations. Today, many species of cycad are at risk of extinction, including the monotypic, endemic *Microcycas*, found in western Cuba, some species of *Zamia* in Colombia and Central America, *Ceratozamia* in Mexico, *Cycas* in Indonesia and China, and several species of *Encephalartos* in southern Africa (Osborne et al., 2012). Many cycad species have a restricted geographic distribution. Furthermore, there are typically few individuals participating in a given reproductive event (Da Silva et al. 2012). Usually, few plants produce megasporangiate strobili in a population and are receptive to pollen for a limited time (Clugston et al. 2016). Although they are long-lived plants, they are very slow to reproduce, and populations often have low levels of seedling recruitment. Compounding this is the fact that populations are often small, and this makes them very vulnerable to environmental change. Any significant negative change to the dynamics of a population can take the population a long time to recover - even if conditions do become favourable again (Raimondo & Donaldson 2003).

More than 60% of extant taxa of cycads are on the IUCN Red List of Threatened Species, where they are listed as extinct in the wild, critically endangered, or endangered (Smith et al. 2011; Donaldson 2003). Climate change and human-activities have greatly reduced many habitats for cycads (Donaldson, 2003). Collectors also pose a threat to populations, as they often go to great lengths to obtain plants and illegally collect them from the wild due to their rarity, horticultural appeal and medicinal value (Cousins et al. 2012; Cousins et al. 2011; Kessler 2005; Jones 2002). Plants are often illegally exported from Mexico and South Africa (Donaldson, 2003).

A significant problem in recent years is the introduction of cycad scale insect (*Aulacaspis yasumatsui*) to India (Muniappan et al. 2012), Guam (Marler & Muniappan 2006), USA (Howard et al. 1999), Taiwan (Bailey et al. 2010), and Indonesia (Muniappan et al. 2012). Cycad scale insects feed on the sap, and often cover the leaves causing the plant to constantly produce a new flush of leaves until the plant becomes weakened (Marler & Niklas 2011). Marler and Muniappan (2006) found that the insect has damaged many plants in populations of *Cycas micronesica*, preventing reproduction of infected individuals. Although cycad scale insect has not yet been recorded from Australia, it poses a significant threat to many wild populations.

Cycad diversity and threats in Australia. Australia is one of the world's key areas for biodiversity and has a rich floristic diversity (Sloan et al. 2014). With Australia's diverse flora, the country also has a high number of endemic species (Coates & Atkins 2001). The highest level of cycad diversity exists along the eastern and northern coastal areas, with hotspots ranging from Cape York Peninsula (northern Queensland) to the Northern Territory, and central Queensland to New South Wales (Donaldson 2003).

Australia has the highest cycad genus diversity, comprising both families and four genera. Australian Zamiaceae include the endemic genera: *Bowenia* (far north Queensland), *Macrozamia* (New South Wales, Queensland, Western Australia and the Northern Territory) and *Lepidozamia* (New South Wales and Queensland).

Cycas have adapted to a diverse variety of habitats in Australia ranging from the open or closed wet tropical rainforests of Queensland; coastal, sub-coastal forests and open old-beach dunes near coastal sites in Western Australia; and dry rocky outcrops, open savannah, coastal plains and eucalypt woodlands of the Northern Territory. *Cycas* are often found growing alongside rivers, and steep slopes and mountainsides. Soils range from wet heavy clay-loam soils in tropical Queensland to sandy soils over granite or coarse sandstone, and heavy limestone soils (Hill 1996). This diverse range of conditions can be broadly described as either very wet or dry conditions that are consistently nutrient-poor (Jones 2002). Northern Territory *Cycas* species are considered to be some of the most morphologically diverse, exhibiting a complex geographical pattern of distribution throughout their range. Due to their morphological diversity, they have also developed a wide range of adaptations, allowing them to occur in a range of habitats (Hill, 1996).

Unlike species that occur in Africa and Central America, collectors have targeted fewer species of cycads in Australia. Cycad populations in Australia are often relatively large and show a high level of recruitment. However, few population data are available for the Australian taxa (Donaldson, 2003). There are many threats that affect cycad populations in Australia. Fire is a major threat to seedlings, although many species are well adapted to survive regular burning, due to their monoxylic stems and Crassulacean Acid Metabolism (CAM) pathway (Vovides et al. 2003). Habitat loss is by far the greatest concern for wild cycad populations in Australia, as large-scale clearing of land has removed entire populations of some species. Despite this, many species have viable populations, sometimes with considerable numbers of individuals (Donaldson, 2003), and represent immense conservation value.

1.2 Conservation genetics and genomics

Conservation genetics and genomics uses genetic data, theory and techniques to help understand the genetic diversity and genetic history of a species to better understand the risk of their extinction (Harrisson et al. 2014). The long-term aim of conservation genetics is to retain enough genetic variation in populations so future adaptation, expansion and re-establishment of natural populations is possible, ultimately increasing the survival of a species (Hedrick & Miller 1992). Conservation genetics has also been used to help identify populations that have high levels of genetic diversity and should be the highest priority for conservation; the reasoning for this is that it maximises the potential genetic resilience of a species through the adaptability of populations (Frankham et al. 2004).

The applications of conservation genomics. Conservation genomics is an applied science, with a foundation that first originated in the late 1970s. Through conservation genetics, evolutionary and molecular genetics have been used to aid in biodiversity conservation and species conservation (Frankham 2010). Molecular genetic markers have been important in measuring many detrimental parameters that are imperative to conservation, including reduced population size, past bottlenecks, sex specific gene flow and founder contribution to a population (Hedrick 2001). In addition, conservation genetics applies a number of techniques that directly and indirectly aid conservation. Genetic data have been used to resolve taxonomic uncertainties, such as defining evolutionary divergence between species, and are used to design individual management plans (Frankham 2010). González-Astorga et al. (2003) and Vovides, et al. (2008) used population genetics to differentiate *Dioon edule* var. *angustifolium* from *D. edule* var. *edule*, which was then later recognised as a separate species (*Dioon angustifolium* Miq.) based on further study (González-Astorga et al. 2005).

Genetics can also be used to develop conservation management plans, which helps minimise the level of inbreeding and loss of allelic diversity in wild populations, this is especially important with species that are at risk of extinction (Frankham 2010). Calonje et al. (2013) carried out a conservation genetic assessment of *Zamia lucayana* Britton in the Bahamas using microsatellites. They found populations of the species were once a single population that had become differentiated into multiple sub-populations over time. The results indicated that despite the differentiation between the populations, the species should be managed as a single unit. Additionally, conservation genetics can aid in the management of *ex-situ* conservation collections to minimise the level of inbreeding and decrease the loss of genetic diversity in cultivation (Frankham, 2010). This was demonstrated by Da Silva et al. (2012), who developed a management plan for *Encephalartos latifrons* Lehm. in cultivation using RFLPs, as the F1 generation showed a significant decrease in allelic diversity compared to the parent plants. Due to this reason, they developed a comprehensive breeding programme to prevent inbreeding. In addition, Griffith et al. (2015) addressed whether botanic gardens can represent the genetic diversity of wild populations of cycads. They inferred that the knowledge gained by understanding the genetic diversity of wild populations can be used to assess the optimum number of individuals required to represent the genetic diversity of a species.

Before the use of next generation sequencing (NGS) techniques, applying genomics in conservation genetics had been limited, mainly because the techniques are often seen as being challenging and expensive. However, NGS techniques promise considerable advantages over more traditional methods (Allendorf et al. 2010). Population genomic approaches, unlike traditional population genetics that use around 15-20 markers, use a wide range of markers often covering the entire genome, and often using thousands of informative markers (Narum et al. 2013). Population genomic techniques have proven valuable in understanding the differentiation between neutral markers and non-neutral markers. This information allows more reliable estimates of the demography and history by excluding non-neutral markers from an analysis (Ouborg et al. 2010). This ability to examine thousands of markers would answer many questions that were unanswered until now (Allendorf et al., 2010).

NGS allows researchers to sequence an entire genome and produces gigabases of DNA sequence data at minimum cost. However, full genome sequencing is generally not required for conservation genetics because not all markers will be informative. But, genome-wide marker discovery and analysis of populations is essential for understanding the evolutionary process that effects change (Davey & Blaxter 2010).

Most studies in conservation genetics, population genetics and molecular ecology have focused on a small number of neutral markers, including allozymes, RFLPs, AFLPs and microsatellites. Genotyping-by-sequencing offers major advantages to conservation genetics and genomics by being able to screen for thousands of polymorphisms throughout the genome that are subject to the full evolutionary history of a taxon, including genetic drift, selection, natural selection, recombination, mutation and speciation (Narum et al. 2013).

NGS techniques can be used to answer many new questions with a variety of techniques and methods such as the ability to explore informative genetic makers, that are either neutral (not involved in natural selection) or non-neutral (involved in natural selection) (Freamo et al. 2011). These informative markers can also be used to understand the effects of inbreeding, outbreeding and depression, and levels of heterozygosity and homozygosity across populations (K. R. Andrews et al. 2014). NGS can also be used to eliminate exogenous alleles brought about during hybridisation, which can be problematic for population genetics (Barton 2001). Although next generation sequencing does not need a reference genome, the use of a reference can enable exploration into wider species diversity, target selected regions of the genome to sequence, and aid in the development of new markers (Lister et al. 2009). This can also increase effectiveness of NGS in conservation management in more conserved populations as population-specific markers can be identified (Toro et al. 2014). Additionally, NGS can be used to find markers that link genotypes to phenotypes using the identification of outlier loci or candidate genes under natural selection (Stapley et al. 2010).

1.3 Cycad conservation and population genetics

Population genetics comprises a set of techniques that has allowed a wide range of studies to be done on many organisms – from epidemiology to species, or drug discovery to conservation. Ranging from AFLP and RFLP data of twenty-five years ago, through to the current next-generation platforms, the methods and power of the techniques have been refined immensely through this period (Benestan et al. 2016). Cycads are among the many groups that have been studied using a variety of the available techniques over the years.

Due to the threats posed on cycad populations they are often small and fragmented, especially in Africa (Ekué et al. 2008; Da Silva et al. 2012) and Central America (Cabrera-Toledo et al. 2008). This is reflected in the genetic diversity of their populations, where many species show evidence of genetic drift and bottlenecks due to fragmentation (Zhan et al. 2011; González-Astorga et al. 2003). In summary, cycad population genetic level studies show cycads have a range from high to low levels of interspecific genetic diversity (Calonje et al. 2013). Many species show high levels of intrapopulational differentiation, indicating a lack of gene flow and fragmentation (Long-Qian et al. 2004; Keppel et al. 2002;). Cycad populations are often biogeographically disjunct, and in many cases, species are often rare throughout their range, with small effective population sizes (Donaldson, 2003). The small, disjunct populations are often geographically isolated (Cibrián-Jaramillo et al. 2010), and this explains how gene flow between the population is greatly reduced and could account for the high levels of interpopulational variation (Cibrián-Jaramillo et al., 2010). However, in those species such as *Cycas debaoensis* and *C. panzhihuaensis* from China, levels of interpopulational variation are lower, which indicates high gene flow between the populations, resulting in reduced genetic differentiation between populations due to mixing (Huang et al. 2004; Yang & Meerow 1996). Although populations have medium to high levels of heterozygosity, many species show increased evidence of inbreeding (Meerow et al. 2012; Long- Qian et al. 2004).

In general, cycad populations also show a higher number of polymorphic loci in adult plants, than in seedlings and juveniles due to bottlenecks (Da Silva et al. 2012; Octavio-Aguilar et al. 2009). The long generation times of cycads, with some species taking up to 60 years to reach reproductive maturity (Da Silva et al., 2012), and the dioecious nature of cycads means that populations are often very slow to regenerate. Some species of cycads also have very low seed viability, which further

reduces the regeneration time of populations (Vovides et al. 1997). Fragmentation of ranges will also greatly affect smaller populations that generally have higher levels of inbreeding and evidence of genetic bottlenecks. Cycads are also poor seed dispersers, with recent research indicating that the seeds of many species are rarely dispersed more than 5 m from the maternal parent (Hall & Walter 2014). Low levels of genetic diversity can be detrimental to the long-term survival of a species, as it impairs its adaptability (Da Silva, et al., 2012). Due to the threats posed on cycad populations - now and in the future - there is a definite risk of isolation between populations. Although this leads to an increase in population differentiation and increased divergence, it can lead to increasing levels of inbreeding within populations (James et al. 2018). This could increase the number of deleterious allies in population and, therefore, reduce the future genetic of a species (Pekkala et al. 2014).

1.4 Restriction Site Associated DNA Sequencing

The advantage of restriction site associated DNA sequencing (RADseq) is that the technique sequences a reduced representation of the genome but maintains a deep enough level of sequence coverage near a specific type of restriction site. This is especially useful for organisms with large genomes such as cycads, whereby the large genomes traditionally make the prospect of sequencing the genome intractable and expensive (Zonneveld 2012). RADseq is suitable for organisms with large genomes, as it has been used as an exploratory approach on *Thuja* (Cupressaceae) to generate microsats (Hou et al. 2018) and is, therefore, suitable for cycads. RADseq is one of a new suite of techniques that allow small stretches of the genome to be used for population genotyping by identifying potentially thousands of polymorphisms to be sequenced quickly and cost-effectively (Peterson et al., 2012). These recovered polymorphisms include both neutral and non-neutral markers which are subject to the full evolutionary history of a taxon, including natural selection, mutation, and variation in drift (Narum et al., 2013).

RADseq offers major advantages for the future of population genetics. It does not require a reference genome and is thus an ideal candidate for use on non-model organisms (Andrews et al. 2014). RADseq is different from many other methods of genomic sequencing techniques that use DNA fragments to construct a DNA sequence library by using a restriction enzyme. Enzyme digestion can result in non-random cleavage, where most of the same regions will be sampled across different individuals (Arnold et al. 2013).

The applications of RADseq in biology

RADseq in conservation genetics. RADseq has been used in conservation genetics with great success for organisms with large or small genomes (Hou et al. 2018). Some of the earliest studies employing RADseq aimed to develop and identify large numbers of genome-wide SNPs and SSR markers for population genomic analysis in *Solanum melongena* L. (Barchi et al. 2011) and *Cynara cardunculus* L. var. *scolymus* (Scaglione et al. 2012). Markers developed from RADseq, were used to understand population differentiation and to separate gene pools in populations that can differ in heterozygosity and geography (McCartney-Melstad et al. 2017; Paun et al. 2015; Xu et al. 2012). Additionally, RADseq has been used to understand species introgression in natural populations, and recognise species-specific alleles, which can contribute to the identification of new species (Guo et al. 2014; Wang et al. 2013).

One of the most effective ways RADseq has been used for conservation genetics is to identify genetic differences in populations that correlate with climatic zones where genetic interchange has become reduced (Lexer et al. 2014). Not only can RADseq be used to understand genetic differences between species different climatic zones, it can also be used to determine genetic differentiation of species within the same climatic environment (Cromie et al. 2013). RADseq was used to deduce the genetic difference between species of *Quercus* L., and was able to recover genes that were responsible for drought tolerance in some populations (Oney-Birol et al. 2018). The method can be used to assign individuals of unknown provenance to their original geographic location (Eaton 2014). Significant genetic differences between geographic regions can lead to allopatric speciation. This was demonstrated with the diamondback rattlesnake (*Crotalus atrox* Baird & Girard, 1853), where the species showed significant genetic differentiation between geographic regions that indicated speciation (Schield et al. 2015).

RADseq in phylogenetics. The use of RADseq is an effective method to generate data for reconstructing phylogenetic relationships among closely related species, including younger groups and clades (Rubin et al. 2012). RADseq shows great utility for phylogenetics in correcting confounding effects of sequencing errors, heterozygosity and low genomic coverage (Cariou et al. 2013). Resolving the phylogenetic relationships among recently derived taxa is generally considered to be problematic (Nagalingum et al. 2011). This has been the case for temperate bamboo species, which showed poor resolution and support when using standard approaches, including a complete chloroplast genome phylogeny (Wang et al. 2013). However, the use of RADseq yielded robust phylogenetic inferences of the group, resolving the relationship among the species (Wang et al. 2013). RADseq was able to generate data that could fully resolve the relationship among 18 species of ground beetle, whereas Sanger sequencing failed (Cruaud et al. 2014). A concern with RADseq data and phylogenetics is the presence of missing data (Eaton et al. 2016), but recent studies have shown that even with 90% missing data full phylogenetic support can be obtained (Tripp et al. 2017).

Problems associated with RADseq. RADseq has many benefits, but it does have downsides. Currently there is no standard protocol for RADseq which includes a series of methods including; RAD, ddRAD, GBS, MSG and ezRAD (Puritz et al. 2014). This can make selecting the correct protocol challenging and can lead to an incorrect approach being used to answer the biological questions (Davey & Blaxter 2010). The use of RADseq can lead to problems detecting loci involved in local adaptation and divergence among wild populations, as a large proportion of the genome can be missed, leading linkage disequilibrium and under or over estimation genetic diversity (Lowry et al. 2016).

RADseq has been known to bias genetic estimation due to allele dropout, because if a polymorphism occurs at a restriction site, a heterozygous genotype could be mistaken for a homozygous genotype (Cariou et al. 2013). Additionally, the choice of RADseq protocol has the potential to further bias results which is the case between single digest (RAD = using a single restriction enzyme) and double digest (ddRAD using two restriction enzymes) methods, which can result in significant differences in the observed heterozygosity and allele frequencies between (Flanagan & Jones 2018).

Finally, and perhaps most important factors is downstream bioinformatic processing as different de novo assembly approaches can results in significant differences in population generic statistics, leading to inconsistencies between assembly methods (Shafer et al. 2016). Some software packages very effective manage RADseq data and can help reduce the biases introduced from sequencing and/or assembly (Eaton 2014). However, although transcriptome sequencing and exome capture could be considered better approaches for reduced genome sequencing (Lowry et al. 2016), these are not viable approaches for capturing genetic diversity of organisms with large and complex genomes. RADseq remains the best and most cost effective approach to full genome SNP capture (Catchen et al. 2017).

1.5 Unresolved questions in conservation genetics

Conservation genetics raises a number of major questions that, so far, have yet to be resolved. Conservation genetics studies have not been able to unambiguously resolve how fitness-related functional genetic variation can affect a population (e.g. what is the correlation between population size and the level of non-neutral genetic variation (Groom et al. 2006)). Many conservation genetics studies have used near-neutral markers, which can be highly conserved and lack important genetic variation, providing an inaccurate picture about the genetic structure of a population (Lesica & Allendorf 1995). Population based studies use a small number of a loci to understand the genetic variation in a population and could be missing important genetic information. Using a genome-wide approach to understand population dynamics could reveal novel amounts of variation within species (Narum et al., 2013). Consequently, this is the approach taken in this research project.

Recently, great progress has been made in conservation genetics by predicting and explaining natural genetic variation in populations (Schield et al. 2015). However, it is still unknown what the underlying mechanisms are that connect dynamics of a population with adaptation and fitness (Frankham 1995). Inbreeding depression is a central problem in most conservation genetic research studies, as it often has unknown genomic causes and detrimental effects (Hedrick & Kalinowski 2000). In order to be able to make accurate predictions of the problems involved with small populations, it is important to understand the genomic causes and effects of inbreeding depression (Charlesworth & Willis 2009). It is important, therefore, to investigate the number of genes or genomic pathways that are involved in the process of inbreeding depression.

A question that has been yet to be explored in cycads is the impact of hybridisation on the diversity of wild populations. Although cycads are thought to readily hybridise in the wild (van der Bank et al. 2001), the effects of this have never been published and it is unknown if these are true hybrids or different morphotypes (Chamberlain 1926). The effects of introgression in plants such as cycads is important, for it promotes genetic mixing and dilution; the deleterious effects of which are further exacerbated in taxa with small effective population sizes (Ellstrand et al. 1990).

1.6 Aims and Objectives

Project overview and aims. This thesis reports the findings of my research into the evolution and genetic diversity of wild *Cycas* populations in the Northern Territory, Australia. The findings also represent case studies that can be more widely applied to the Cycadales as a whole. The thesis report is divided results into three chapters, designed as individual research papers and formatted as such.

Chapter 2. This chapter addresses the challenges presented by cycads in which we test our RADseq approach across all cycad genera as proof of concept for population and conservation genetics. Our goal is to demonstrate the effectiveness of RADseq across large and complex genomes, and to allow others to follow this protocol. Specifically, this chapter aims to: (1) demonstrate that RADseq can be successfully applied to organisms with large, repetitive genomes, such as cycads; (2) generate a sufficient number of loci by using *de novo* assembly for phylogenetic and population genetic analyses; (3) develop an effective method that can be used for genome skimming.

Chapter 3. The geographically widespread species, *C. calcicola* is the subject of this chapter. The chapter explores two areas. Firstly, we investigate the genetic differentiation among and within populations of *C. calcicola* - in particular within the Litchfield and Katherine regions, where the majority of the populations are found. Secondly, we determine if the *ex-situ* botanic garden collections of this species are sufficiently representative of the genetic diversity in the wild populations.

Chapter 4. This chapter aims to understand the population genetics of two closely related species, *Cycas armstrongii* and *C. maconochiei* subsp. *maconochiei*, with parapatric distributions. We also test the validity of a population of putative hybrids between *C. armstrongii* and C. *maconochiei* subsp. *maconochiei* (Hill 1996), based on their morphological intermediate characters. Specifically, we investigate whether the putative hybrids are a new, cryptic species, a morphological variant of one of the species, or a valid hybrid between the two species. Based upon our findings, we provide insights into the formation of conservation assessments for three taxa, and we provide informed guidance on how to target future collections to safeguard wild populations.

Chapter 5. Concludes, compares and contrasts the findings from the preceding three chapters, and illustrates how the genetic diversity of three cycad species can be efficiently determined by using RAD seq techniques. We then provide an informed basis for the generation and implementation of species-specific conservation management plans that will ensure the long-term survival of the study species.

Overall objectives

- Explore new approaches to the conservation genomics of cycads (Chapter 2)
- Develop protocols effective for genome skimming and genome-wide marker discovery for others to apply to cycads and organisms with large and complex genomes (Chapter 2)
- Gain insights into the genetic diversity of *C. calcicola* and find how the disjunction in its range affects this (Chapter 3)
- Understand the diversity and differentiation between *C. armstrongii* and *C. maconochiei* subsp. *maconochiei* (Chapter 4)
- Determine the hybrid status of *C. armstrongii* x *maconochiei* (Chapter 4)
- Determine if ex-situ botanic garden collections represent the genetic diversity for the chosen species of their wild populations (Chapters 3 and 4)
- Understand the effects of geography on the genetic diversity of *Cycas* in Northern Territory (CH?)

Chapter 2 - RADseq as a valuable tool for plants with large genomes—a case study in cycads

Full genome sequencing of organisms with large and complex genomes is intractable and not cost-effective under most research budgets. Cycads (Cycadales) represent one of the oldest lineages of extant seed plants and, partly due to their age, have incredibly large genomes up to ~60Gbp. Restriction site associated DNA sequencing (RADseq) offers an approach to find genome-wide informative markers, and has proven to be effective with both model and non-model organisms. We tested the application of RADseq using ezRAD across all ten genera of the Cycadales including an example dataset of *Cycas calcicola* representing 72 samples from natural populations. Using previously available plastid and mitochondrial genomes as references, reads were mapped recovering plastid and mitochondrial genome regions and nuclear markers for all of the genera. De novo assembly generated up to 138,407 high-depth clusters and up to 1,705 phylogenetically informative loci for the genera, and 4,421 loci for the example assembly of *C. calcicola*. The number of loci recovered by de novo assembly were lower than previous RADseq studies, yet still sufficient for downstream analysis. The number of markers could be increased by relaxing our assembly parameters, especially for the *C. calcicola* dataset. Our results demonstrate the successful application of RADseq across the Cycadales to generate a large number of markers for all genomic compartments, despite the large number of plastids present in a typical plant cell. Our modified protocol was adapted to be applied to cycads and other organisms with large genomes to yield many informative genome-wide markers.

2.1 Introduction

The size of an organism's genome greatly affects the cost of sequencing its genome, which in turn affects the number of organisms for which genomic data are available (Andrews et al. 2016). Large genomes are caused by numerous factors such as tandem repeats, pseudogenes, paralogs, polyploidy or a combination of these factors (Guan et al. 2016). Plant genome sizes are highly plastic (Pellicer et al. 2018), ranging from 13.2 Megabase pairs (Mbp) in the genome of *Ostreococcus lucimarinus*, to over 149 Gigabase pairs (Gbp) in the octoploid *Paris japonica* (Pellicer et al. 2010). As a result of whole genome duplication, gymnosperm genomes are generally larger than those of many angiosperms, ranging from ~8 Gbp in *Microstrobus* to ~72 Gbp in *Pinus* and *Ceratozamia* (Roodt et al. 2017; Zonneveld and Lindstrom 2016; Scott et al. 2016; Zonneveld 2012). Typically, as a result of polyploidy, the large genome size is generally caused by an inefficiency of gymnosperms to eliminate the amplification of repeats in the genome (Pellicer et al. 2018).

Next generation sequencing (NGS) permits sequencing large stretches of a genome to produce DNA sequence data in the Gbp range at relatively low cost. Full genome sequencing may be the best approach for finding informative markers that assist investigating the evolutionary history of a species (Andrews et al., 2016). However, large and complex genomes present problems of cost for existing NGS approaches (Alexeyenko et al., 2014). Further issues include generating enough repeat reads to account for over-representation of highly repeated elements in the genome (Catchen et al. 2017). Additionally, de novo assembly of larger genomes becomes problematic because of repeated elements, making effective repeatability of an assembly difficult (Meyers et al. 2004).

Restriction-site associated DNA sequencing (RADseq), uses restriction enzymes to reduce the proportion of the genome sequenced by cutting DNA into smaller fragments, and a subset of these fragments (typically between 200-600 bp) is then selected for sequencing (Davey and Blaxter, 2010). Thus, RADseq allows the sequencing of a reduced representation of the genome yet still at a deep level of sequence coverage, especially near specific restriction sites. Only a portion of the genome is sequenced (Andrews et al. 2016). Compared to many NGS methods such as shotgun and whole genome sequencing, RADseq is considered quick and economical under most research budgets (Peterson et al., 2012; Toonen et al.,

2013). RADseq has offered new avenues for phylogenetics and population genomics (Table 1) because it does not require the use of a reference genome (Andrews and Luikart, 2014), and has proven to be very effective for population genotyping by identifying thousands of polymorphisms (Mastretta-Yanes, et al., 2015). These polymorphisms include both neutral and non-neutral markers that potentially reflect a large portion of a taxon's genome involved in natural selection and mutation (Narum et al., 2013). RADseq has been applied in population genetics across a range of model plants, such as *Oryza* and *Carex*, as well as non-model plants including *Senecio, Betula, Sisymbrium*, *Mimulus, Passiflora, Psychotria* and *Mangifera* (Warschefsky and von Wettberg 2019; Nazareno et al. 2018; Massatti et al. 2016; Twyford and Friedman, 2015; Guo et al., 2014; Wang et al., 2013; Roda et al. 2013, Vandepitte et al., 2013). It has been used, to a lesser extent, in plant phylogenetics for *Pedicularis*, *Diospyros*, *Quercus*, *Viburnum*, and *Diuris* (Ahrens et al. 2017; Eaton et al. 2016; Eaton et al. 2015; Paun et al. 2015 and Eaton; Ree 2013).

Table 1 Estimated genome sizes of taxa examined in earlier RADseq studies. Gbp= number of billion base pairs and estimation of plant genome size obtained from Plant DNA C-values Database (http://data.kew.org/cvalues/)

Taxon	Est genome size (Gbp)	Type of study	Citation
Carex spp.	0.30-2.36	Phylogenetics	Massatti et al. 2016
Cedrus sp.	16	Phylogenetics	Karam et al. 2015
Sisymbrium austriacum	0.72	Population genetics	Vandepitte et al. 2013
Mimulus spp.	0.74	Population genetics	Twyford & Friedman 2015
Mangifera indica	0.45	Population genetics	Warschefsky and von Wettberg 2019
Betula nana	0.92	Population genetics	Wang et al. 2013
Quercus spp.	1-2	Phylogenetics	Eaton et al. 2015
Oryza sativa	2	Population genetics	Guo et al. 2014
Diospyros spp.	2.40-5.76	Phylogenetics	Paun et al. 2015
Viburnum ssp.	3.8-4	Phylogenetics	Eaton et al. 2016
Senecio lautus	4.90	Population genetics	Roda et al. 2013
Passiflora spinosa	0.92-2.68	Population genetics	Nazareno et al. 2018
Pedicularis spp.	5.68	Phylogenetics	Eaton & Ree 2013
Psychotria lupulina	1.03	Population genetics	Nazareno et al. 2018

Currently published fully-sequenced plastome and mitochondrial genomes for cycads are few, yet this number already appears to provide sufficient evidence to invest in alternative sequencing methods of genomic DNA, such as that of RADseq. Of the ten genera of cycad, eight – *Ceratozamia, Cycas, Dioon, Encephalartos, Macrozamia, Lepidozamia, Stangeria,* and *Zamia* – have documented plastomes (Wu and Chaw 2015 and Wu et al. 2007). Yet a comparison of high GC-biased substitutions, gene conversion, and low sequence variability between both theirs and other published gymnosperm plastomes (e.g. *Pinus thunbergii*, *Abies koreana* and *Araucaria* spp.) indicates that the plastid is not an optimal source of variable markers that are useful for population genetics or phylogenetic studies (Yang et al. 2016; Zhou et al. 2016; Ruhsam et al. 2015; Yi et al. 2015; Jansen et al. 2011; Wu et al. 2007; Tsudzuki et al. 1992). The only full cycad mitochondrial genome that has been sequenced is that of *Cycas* species (Wu et al. 2007). Compared to published mitochondrial genomes of the closest allies of cycads (*Ginkgo biloba* and *Welwitschia mirabilis*), only a few unique and polymorphic sites were found (Guo et al., 2016), which supports the notion that this genomic compartment is as uninformative as the plastome.

In order to test the effectiveness of RADseq for taxa with large genomes, we used a RADseq technique across a cohort of samples representing ten cycad genera. We chose cycads because they have particularly large genomes, ranging from ~25–30 Gbp in *Cycas* L. to ~72 Gbp in *Ceratozamia* (Zonneveld, 2012), which appears to be the result of many tandem repeats, pseudogenes, paralogs, and possibly whole genome duplication (Roodt et al., 2017). In addition to having generally larger genomes, we also chose cycads because there is need for better methods to find more data-rich sequences for systematic and population genomic studies. Therefore, forming part of our larger conservation genomics study targeting cycads, we developed a RADseq protocol that is based on a modification of the ezRAD protocol (Toonen et al., 2013). ezRAD differs from other RADseq approaches as it uses a commercially available library preparation kit and does not require specific restriction enzymes to ligate adapters to cut sites (Andrews et al., 2016). Another advantage of ezRAD when compared to other RADseq protocols is that it requires lower initial setup preparation and costs (Andrews et al., 2014).

The aim of the larger project is to understand the evolution and genetic diversity of wild *Cycas* populations. As a proof of concept, we tested our RADseq approach across all cycad genera. This study aimed to: (1) demonstrate that RADseq can be successfully applied to organisms with large, repetitive genomes, such as cycads, (2) generate a sufficient number of loci using *de novo* assembly for phylogenetic and population genetic analyses, and (3) develop an effective method that can be used for genome skimming.

2.2 Materials and methods

Sampling strategy. Freshly collected silica-dried leaf material was sampled for all ten genera representing 13 species of Cycadales, from Cycadaceae and Zamiaceae (Table 2). Cycadaceae leaf samples were taken from *Cycas taitungensis* at the living collection of the Royal Botanic Garden and Domain Trust, NSW Australia (RBGS), and samples of *C. armstrongii, C. maconochiei,* and *C. calcicola* were collected from wild plants in the Northern Territory, Australia. For Zamiaceae, *Bowenia spectablis*, *Ceratozamia kuesteriana*, *Dioon mejiae*, *Encephalartos lebomboensis*, *Lepidozamia peroffskyana*, *Macrozamia johnsonii*, *Microcycas calocoma*, *Stangeria eriopus,* and *Zamia integrifolia* samples were collected from the living collection of the RBGS (Table 2).

Table 2 Samples of cycad genera obtained for RADseq. Samples obtained from the Royal Botanic Gardens and Domain Trust, Sydney and wild populations in the Northern Territory Australia. Collection No. = individual date assigned number per sample, RBGS Acc No. = Royal Botanic Gardens and Domain Trust living collection data base accession number, Wild/Cul. = of the origin of the samples and if they were collection from the wild or from the RBGS collections and NCBI Acc. = NCBI Sequence Read Achieve, BioSample accession number (BioProject accession: PRJNA526348).

Collection No.	Species	RBGS Acc No.	Wild/Cul.	NCBI Acc.
Bspe-15-39	*Bowenia spectabilis*	862154	Cul.	SAMN11096242
Ckue-15-40	*Ceratozamia kuesteriana*	816444	Cul.	SAMN11096245
Carm15-24-9	*Cycas armstrongii*	N/A	Wild – NT	SAMN11096243
Ccal-16-18-8	*Cycas calcicola*	N/A	Wid – NT	SAMN11096244
Cmac-15-3-5	*Cycas maconochiei*	N/A	Wid – NT	SAMN11096246
Ctai-15-75	*Cycas taitungensis*	816340	Cul.	SAMN11096247
Dmej-15-41	*Dioon mejiae*	816446	Cul.	SAMN11096248
Eleb-15-42	*Encephalartos lebomboensis*	816449	Cul.	SAMN11096249
Lper-15-43	*Lepidozamia peroffskyana*	816455	Cul.	SAMN11096250
Mcal-15-45	*Microcycas calocoma*	816474	Cul.	SAMN11096251
Mjoh-15-44	*Macrozamia johnsonii*	816460	Cul.	SAMN11096252
Seri-15-46	*Stangeria eriopus*	816474	Cul.	SAMN11096253
Zint-15-4	*Zamia integrifolia*	816496	Cul.	SAMN11096254

To test the utility of RADseq at population level, samples were collected from 60 individuals of *Cycas calcicola* from natural populations in the Northern Territory, Australia (Appendix I). The samples included three populations from the Litchfield National Park and three populations in the Katherine region — each population consisted of ten individuals of varying ages. A further 13 samples were sourced from cultivated ex-situ collections of George Brown Darwin Botanic Garden (Darwin, Northern Territory, Australia) and Montgomery Botanical Center (Miami, Florida, USA).

DNA extraction and quantification. Approximately 0.05 g of silica-dried leaf samples were ground to a fine powder using a TissueLyser (Qiagen Inc., Venlo, the Netherlands). When present in large amounts, trichomes were removed to improve extraction quality (specifically in *Cycas calcicola*). High molecular weight genomic DNA was extracted using a DNeasy Plant DNA Extraction Mini Kit (3.0 BR DNA assay; Qiagen, Hilden, Germany). Genomic DNA was inspected using a 2% agarose gel to check for the presence of DNA and impurities. A Qubit fluorometer (3.0 BR DNA assay; Invitrogen, Life Technologies, Carlsbad, CA, USA) was then used to determine the quantity (µg/mL) of the extracted DNA for each sample. The target concentration for samples was (+=) 17 µg/mL; samples that yielded less then this amount was either re-extracted or concentrated using a 1:1 ratio of Agencourt

AMPure XP magnetic purification beads to sample volume (Beckman Coulter, Inc) by combining multiple extractions (For more detailed laboratory methods, please see supplementary data Appendix II).

DNA normalization and double digest reaction. First, genomic DNA was normalized to a concentration of 500 ng in 42 µL total volume (0.01 µg/mL) using a QIAgility liquid handling robot (Qiagen Inc., Venlo, the Netherlands). Second, using the QIAgility, 5 µL of NEB 10x CutSmart buffer and 1 µL of Bovine Serum Albumin (BSA; to help stabilize the enzyme digestion) was added to each well and mixed briefly for five seconds using a plate mixer (although these steps were performed using a liquid handling robot, they can be performed manually). This mix was stored at 4°C for a minimum of 5 hours—our tests showed that this helps to reduce the effect of DNA methylation, improving the cutting action of the restriction enzymes. Next, double digest reactions were set up using 1 µL of each EcoR1-HF and Mse1 restriction enzymes, mixed by pipetting manually. Reactions were run in a thermocycler for 3 hrs at 37°C with a final 20 min deactivation step at 65°C. Using 2% agarose gel, samples were checked for a smear to indicate the quality of digestion. Lastly, double digest reactions were cleaned using 1.8:1.0 ratio of AMPure XP beads to sample (90 µL of AMPure XP beads to 50 µL of digested DNA) and quantified using a Qubit high sensitivity kit (3.0 HS DNA assay; Invitrogen, Life Technologies, Carlsbad, CA, USA).

Library preparation. RADseq libraries were prepared following the ezRAD protocol (Toonen et al. 2013) in which we tested two different Illumina (Illumina Inc., CA, USA) library preparation kits: firstly, an Illumina TruSeq PCR-Free high throughput dual index kit and secondly, an Illumina TruSeq nano high throughput dual index kit (PCR-based, FC-121-4003). Our initial aim was to use the PCR-Free kit to help reduce the probability of PCR amplification bias. However, after multiple attempts the PCR-Free kit resulted in poor final yields when quantified using qPCR, and after multiple troubleshooting steps, it was deemed unfit for our target group (cycads). However, the Illumina TruSeq nano kit proved to be effective when the input of genomic DNA was increased by 5x the recommended input, i.e., from 100ng to 500ng, due to the amount of DNA which is lost during clean-up and size selection. We followed the ezRAD protocol v3 using half of the recommended volumes of an Illumina TruSeq kit to save costs (Toonen et al. 2013).

Several quality control checks were carried out during library preparation on a select number of samples (16-24 samples) using a high performance LabChip and a Qubit fluorometer; more specifically, DNA size and quantity (μg/mL) were checked after digestion and after size selection. During the final step of library preparation, we modified the ezRAD protocol in the final bead clean, using a 0.8:1 ratio of AMPure XP beads to sample for the removal of excess adapters observed using a LabChip. Final Illumina libraries were validated using a LabChip, cleaned using a 0.9:1 ratio of AMPure XP beads to sample, and quantified using a Qubit high sensitivity kit (3.0 HS DNA assay; Invitrogen, Life Technologies, Carlsbad, CA, USA). Final libraries were normalized to 10 nM and pooled for sequencing. For more detailed laboratory methods, please see supplementary data (Appendix 1).

Sequencing. We aimed to capture around 1 gigabyte (Gb) of sequence data per sample (in a run of 95 libraries) to account for overrepresentation of the plastid genome, and to capture as much of the nuclear genome as possible. Genomic sequencing was carried out using an Illumina NextSeq 500 with 150 bp paired-end high throughput (HT) on a single flow cell. The NextSeq 500 HT run can capture up to 120Gb of sequencing data, thereby allowing for our sequencing target of one Gb per sample. The sequencing run was also spiked with 20% PhiX sequencing control V3 (Illumina) to account for low sequence diversity caused by the identical enzymatic digestion cut sites in the ezRAD protocol.

Bioinformatics

Quality control and filtering of sequence reads. The NextSeq 500 generated four fastq files for forward and reverse reads (eight files per sample). The four forward fastq files were concatenated into a single forward fastq file and similarly a single reverse file was created, as required for the downstream RADseq assembly. The concatenated forward and reverse fastq files were screened for quality using PRINSEQ v0.20.4 (Schmieder and Edwards 2011). PRINSEQ allowed the detection of falloff in read quality for a range of samples from each population. The reads were trimmed using Trimmomatic 0.36 (Bolger et al. 2014) using the following settings: 1) the Illumina clip function was used to remove adapters, 2) the first six bases were cropped from the start of all paired-end reads, 3) all reads were cropped to 120 bp in length due to lower quality ends (observed using PRINSEQ), and a sliding window was also used to delete bases with a PhredQ score less than 20 with a sliding window of four, and 4) all reads less than 50 bp were discarded, and only paired reads were retained to improve merging of reads during clustering.

Assembly of RADseq data for cycad genera. De novo assembly of the paired-end reads was performed using ipyrad 0.5.13 (Eaton and Overcast, in prep) on a high-performance cluster based at the Royal Botanic Garden Edinburgh using seven nodes, each with 12 cores and 128 GB of RAM, totalling 84 cores and 896 GB of RAM, running for 21 days. In ipyrad all parameters were set to default, except for the following: data type was set to 'pairgbs' (most closely matches ezRAD), bases with a PhredQ score less than 30 were converted to 'N' and reads with 15 or more uncalled bases were discarded. Reads were further filtered for adapter sequences, trimmed, and reads were discarded if they were less than 40 bp in length. The maximum number of uncalled bases in consensus sequences was set to ten for forward and reserve reads. The maximum heterozygotes in consensus sequences was set at eight for both forward and reverse sequences, and the minimum number of samples per locus for output files was set to 4.

Data assembly followed the general ipyrad workflow. Reads were more stringently filtered for presence of adapters (after initial trimming and filtering earlier in Trimmomatic). Next, clusters were identified within samples and consensus base calls were made. Finally, loci were aligned across all of the samples (four species of *Cycas*, and one species each of the nine other cycad genera) and output files were generated, after applying filters as specified in our parameter settings. These settings also included the minimum samples per locus- for example, a generated site is discarded unless it meets the requirement that it is present in a minimum number of samples. The data is archived to allow reproducibility of the assembly (data available through NCBI Sequence Read Archive BioProject accession: PRJNA526348, Table 2).

Assembly from population data of *Cycas calcicola*. To further demonstrate the utility of our protocol, we carried out de novo assembly for 72 individuals of *C. calcicola* (one sample failed during sequencing). The minimum number of samples per locus was set to 43 (as opposed to 4 for the genus level assembly, above), so that each site would be present across a minimum of ~ 60% of samples, to reduce missing data.

Mapping of reads to published references. Large cycad genomes (25 – 60 Gbp), present potential problems with overrepresentation of repetitive regions, and for this reason it is important to test the genomic sources and distribution of RADseq reads. To test for overrepresentation reads were mapped against the published reference plastomes and the single mitochondrial genome (Wu and Chaw 2015 and Wu et al. 2007) (Tables 3 and 4). The reference plastid and mitochondrial genomes were downloaded from NCBI GenBank and the filtered paired end reads were mapped to these references using CLC Genomics Workbench 11.0 (CLC Genomics, 2019; Qiagen Inc., Venlo, the Netherlands) default parameters: for read alignment mismatch costs = 2, intersection and deletion cost = 3, length fraction= 0.5, similarity fraction = 0.8 and auto detection of paired distances was allowed.

Phylogenetic analysis of *Cycas calcicola* populations. The resulting RADseq data provides the first opportunity to investigate the infraspecific relationships between natural populations of *C. calcicola*. Furthermore, this approach can be used to help demonstrate the effectiveness of RADseq in differentiating natural populations. Phylogenetic reconstruction of *C. calcicola*'s populations was completed using SVDquartet plug-in for PAUP* version 4.0a158 (Swofford 2002) because of its robust approach to analysing short gene sequences from RADseq data (Mirarab et al 2015; Liu and Yu 2010). Phylogenetic trees were estimated from the concatenated gene sequence alignments using SVDquartets analysis. Settings included exhaustive quartet sampling, 100,000 bootstrap replicates, and the multispecies coalescent tree model. We examined results of all analyses using at least three independent runs for multi-species coalescent analysis by allocating samples to their respective populations. The three separate populations are at Litchfield National Park (including Tolmer Falls sites), Daly River, Katherine CDU, and Spirit Hills.

2.3 Results

Number and quality of reads. Sequencing on the Illumina NextSeq 500 platform generated approximately 1.9 to 6.7 million 150 bp paired-end reads per sample (Tables 3, 4 and 5). The number of reads generated varied, with the fewest for *Stangeria eriopus*, and the greatest for *Macrozamia johnsonii.* For *Cycas* (target genus), the number of reads generated showed less variation (1.9 to 2.5 million) and was lowest in *C. taitungensis* and greatest in *C. maconochiei*. The PhredQ Score distribution of the sequencing run measured 75.2% at Q30 or greater, which passed the Illumina sequencing filter. Quality control of reads (measured as PhredQ score in FastQC 0.11.5) indicated that forward reads were of a higher quality with a drop-off after 135 bp, whereas reverse reads were lower quality due to drop-off after 120 bp. Due to this quality drop off, forward and reserve reads were filtered and trimmed to 120 bp.

Table 3 Results of mapping RADseq reads against reference plastomes. RADseq reads from seven cycad genera and four Cycas species were mapped against the plastome of the most closely related, available species. PE=paired end; Gbp= number billion base pairs; Reference GenBank Acc. No.= the reference number of a user submitted to the NCBI GenkBank online database https://www.ncbi.nlm.nih.gov/genbank/. % of reads mapped to reference= indicates the percentage of reads mapped against the total number of reads; Average read depth = the average number of reads forming a cluster on the mapped reference, Max. read depth = the maximum depth of reads assembled into clusters, Length of consensus = the length of consensus formed from the mapped reads and Reference covered = percentage of the reference plastome covered from the reads mapped.

RADseq reads								Reference plastomes		
Species	**Number of PE reads**	**No. reads mapped to reference**	**% reads mapped to reference**	**Average read depth**	**Max. read depth**	**Length of consensus**	**Reference covered**	**Species**	**GenBank Acc. No.**	**Length (bp)**
Bowenia spectabilis	2,245,179	122,695	5.46%	81.65	813	158,285	96%	*B. serrulata*	NC_026036	165,695
Ceratozamia kuesteriana	5,941,118	64,052	1.07%	42.25	846	149,659	89%	*C. hildae*	LC049068	165,733
Cycas armstrongii	2,367,871	189,829	8.01%	131.32	908	159,061	97%	*C. taitungensis*	AP009339	163,403
Cycas calcicola	2,373,959	63,651	2.68%	42.93	365	157,215	96%	*C. taitungensis*	AP009339	163,403
Cycas maconochiei	2,571,329	112,189	4.36%	77.67	552	153,610	94%	*C. taitungensis*	AP009339	163,403
Cycas taitungensis	1,972,585	140,535	7.12%	94.43	802	147,868	90%	*C. taitungensis*	AP009339	163,403
Dioon mejiae	2,061,518	37,437	1.8%	26.02	342	148,002	92%	*D. spinulosum*	LC049070	161,815
Encephalartos lebomboensis	2,026,350	16,292	0.80%	10.74	134	148,986	90%	*E. lehmannii*	LC049336	165,822
Macrozamia johnsonii	6,729,621	59,899	0.89%	67.59	1,460	158,563	95%	*M. mountperriensis*	LC049069	166,341
Lepidozamia peroffskyana	3,799,154	221,486	5.82%	40.59	638	152,953	95%	*L. peroffskyana*	LC049207	165,939
Stangeria eriopus	2,050,790	114,251	5.57%	77.01	771	158,135	97%	*S. eriopus*	LC049067	163,671
Zamia integrifolia	2,600,746	53,004	2.03%	35.2	323	149,292	91%	*Z. furfuracea*	LC040885	164,953

Table 4 RADseq data for Cycas spp., mapped to a reference mitochondrial genome of *Cycas taitungensis*. PE=paired end; Gbp= number billion base pairs; Reference GenBank Acc. Cycas taitungensis reference sequence (AP009381) was the downloaded from the NCBI GenkBank online database https://www.ncbi.nlm.nih.gov/genbank/. % of reads mapped to reference= indicates the percentage of reads mapped against the total number of reads; Average read depth = the average number of reads forming a cluster on the mapped reference, Max. read depth = the maximum depth of reads assembled into clusters, Length of consensus = the length of consensus formed from the mapped reads and Reference covered = percentage of the reference mitochondrial genome covered from the reads mapped.

RADseq reads									
Species	Genome size (Gbp)	Number of PE reads	Length of Reference Mt genome	No. reads mapped to reference	% reads mapped to reference	Average read depth	Max. read depth	Length of consensus	Reference covered
Cycas armstrongii	~ 25-28	2,367,871	414,903	26,402	1.11%	6.99	724	282,883	68%
Cycas calcicola	~ 25-28	2,373,959	414,903	14,672	0.61%	3.76	223	255,979	62%
Cycas maconochiei	~ 25-28	2,571,329	414,903	25,837	1%	6.77	446	285,579	69%
Cycas taitungensis	~ 25-28	1,972,585	414,903	26,616	1.34%	6.73	608	255,598	62%

Table 5 Outputs from de-novo assembly of RADseq data. QC=quality control of reads; Gbp= number billion base pairs; Total paired reads merged= number of paired end reads merged to form a consensus read; No. clusters= number of matching reads greater than one forming a cluster; No. clusters hi-depth= number of clusters with 6 or more reads used for majority rule base calling, No. of loci = the total number of loci removed per sample. Genome sizes estimations obtained from Zonneveld (2012).

Species and genome size		Filtering of reads			Outputs from ipyrad			
Species	Genome size (Gbp)	Number of PE reads	Reads passed QC	% Reads passed QC	Total paired reads merged	No. clusters	No. clusters hi-depth	No. of loci
Bowenia spectablis	~ 42	3,217,808	2,245,179	69.77	2,245,179	1,303,450	37,523	316
Ceratozamia kuesteriana	~ 63	8,558,022	5,941,118	69.42	5,941,118	3,365,724	125,895	233
Cycas armstrongii	~ 25-28	3.210,918	2,367,871	74.74	2,367,871	1,319,958	41,765	1,697
Cycas calcicola	~ 25-28	3,123,683	2,373,959	75.99	2,373,959	1,323,284	43,834	1,641
Cycas maconochiei	~ 25-28	3,601,176	2,571,329	71.40	2,571,329	1,391,947	47,562	1,694
Cycas taitungensis	~ 25-28	**2,905,407**	**1,972,585**	67.89	1,972,585	1,098,674	32,816	1,705
Dioon mejiae	~ 48-50	2,795,366	2,062,518	73.78	2,062,518	1,118,716	39,785	201
Encephalartos lebomboensis	~ 54-58	3,075,899	2,026,350	65.88	2,026,350	1,127,824	39,332	215
Macrozamia johnsonii	~ 54	**9,895,138**	**6,729,621**	68	6,729,621	3,090,462	138,407	362
Microcycas calocoma	~ 41	3,622,719	2,602,759	71.84	2,602,759	1,741,344	36,641	125
Lepidozamia peroffskyana	~ 55	5,519,262	3,799,154	68.83	3,799,154	1,888,907	75,403	337
Stangeria eriopus	~ 38	**3,058,563**	**2,050,790**	67.05	2,050,790	1,252,119	32,114	200
Zamia integrifolia	~ 38	3,653,987	2,600,746	71.18	2,600,746	1,518,302	41,031	192

Mapping of reads to published references. RADseq reads were mapped against published reference mitochondrial and chloroplast (plastid) genomes. Plastomes ranged in size from 161,815 to 166,431 bp (Table 3). The number of reads mapped to the plastomes varied from 16,292 reads (0.80% of total reads) for *Encephalartos lebomboensis* to *Encephalartos lehmannii* and 221,486 reads (5.82% total number of reads) for *M. johnsonii* to *M. mountperriensis* (Table 6). The average read depth (Table 3) also varied between the samples and ranged from 10.74 in *E. lebomboensis* to 131.32 in *C. armstrongii* and demonstrated that no clusters were over-represented. Although the percentage of RADseq reads mapped varied, in all species 89% or greater of the reference was covered and was lowest in *Ceratozamia kuesteriana* (89%) and greatest in *S. eriopus* and *C. armstrongii* (97%).

Table 6 Filtering of loci during de-novo assembly of Cycas calcicola natural populations. RADseq reads representing six natural populations (60 samples) and 12 samples representing ex-situ conservation collections. The final step of de-novo assembly the loci which are generated using are passed though numerus filters; Removing duplicates= removed duplicate loci of which are identical, Max indels per locus= remove loci in clusters that reach the threshold for the maximum number of indels per locus to help reduce missing data, Max SNPs per locus= filter based on the maximum number of SNPs per locus to remove clusters with an excess number of loci which could indicate errors in data, Max shared heterozygotes per locus= by the maximum number of hetrozygotes per locus to filter out an excess heterozygous loci and Min samples per locus= The minimum number of samples per locus was set at 43 indicating that each locus was shard across a minimum of ~60% of the samples.

Filtering steps	No. of filtered loci	No. of retained loci
Removing duplicates	3,368,8	1,395,21
Max indels per locus	8,537	1,309,84
Max SNPs per locus	52,101	115,909
Max shared heterozygotes per locus	279	115,851
Min samples per locus	111,430	4421

Reads for *Cycas* spp. were mapped to the mitochondrial genome of *C. taitungensis* which was 414,903 bp (Table 4). The number of reads mapped ranged from 14,672 (0.61% total number of reads) in *C. calcicola* to 26,616 (1.34% total number of reads) in *C. taitungensis*. The number of reads covering the reference mitochondrial genome only varied somewhat between species and was lowest in *C. calcicola* and C. *taitungensis* (62%) and highest in *C. armstrongii* (68%).

De novo assembly of RADseq data. Initial filtering and trimming of the raw Illumina reads were carried out using TRIMMOMATIC. Approximately 65–75% of paired reads were retained (singletons were removed), each with a minimum PhredQ score of 20 (Table 5). The sample which yielded the lowest number of reads after filtering was *C. taitungensis.* During filtering approximately 1 million reads were discarded for each sample and 3 million reads were removed for *Macrozamia johnsonii*. However, *M. johnsonii* remained the taxon with the greatest number of reads overall (Table 5). The number of clusters obtained from de novo assembly ranged from 1.0 to 3.3 million per sample. The number of high-depth clusters (containing six or more reads) ranged from 32,000 in *S. eriopus* to 38,000 in *M. johnsonii* (Table 5). This lower number of high-depth clusters vs initial clusters indicates that there were a high number of clusters with less than six reads, which were discarded due to a higher likelihood of a base being miscalled. The number of recovered loci varied greatly among genera (Table 5), ranging from 1,641 in *C. calcicola* to 1,705 in C. *taitungensis* C.F. Shen, K.D. Hill, C.H. Tsou & C.J. Chen within *Cycas.* A lower number of loci were recovered for Zamiaceae when compared to Cycadaceae with 125 loci being obtained for *Microcycas calocoma* and 362 for *M.* johnsonii (Table 5).

Example assembly of *Cycas calcicola.* The assembly of 72 samples from natural populations of *C. calcicola* (Table 6)*,* generated 1.7 to 4.7 million reads during sequencing, and most reads passed the ipyrad filter (after trimming). The total number of clusters generated during clustering ranged from 1.3 to 3 million, and the number of high-depth clusters range from 22 to 78 thousand. Overall the assembly generated over three million informative SNPs across the 72 samples, and after final filtering, 4,421 loci were recovered for a minimum of 43 samples per locus (each locus was present for ~60% of samples).

Phylogenetic analysis of *Cycas calcicola.* The unrooted tree (Figure 5) recovered seven well-supported populations/groups. Spirit Hills, Daly River, Litchfield National Park (NP) and Litchfield Tolmer populations received 100% bootstrap support (BS). Katherine Charles Darwin University site (Katherine CDU) recieved 99.3% BS and Katherine population and cultivated samples from Katherine TT (Katherine TT CUL) each recived 90.6% BS. Populations from Katherine and Litchfield national park (NP) were recovered as two separate clades (99.3 and 100%, respectively). Total weight of incompatible quartets was 16.5780 (47.409%), and total weight of compatible quartets was 18.3897 (52.591%).

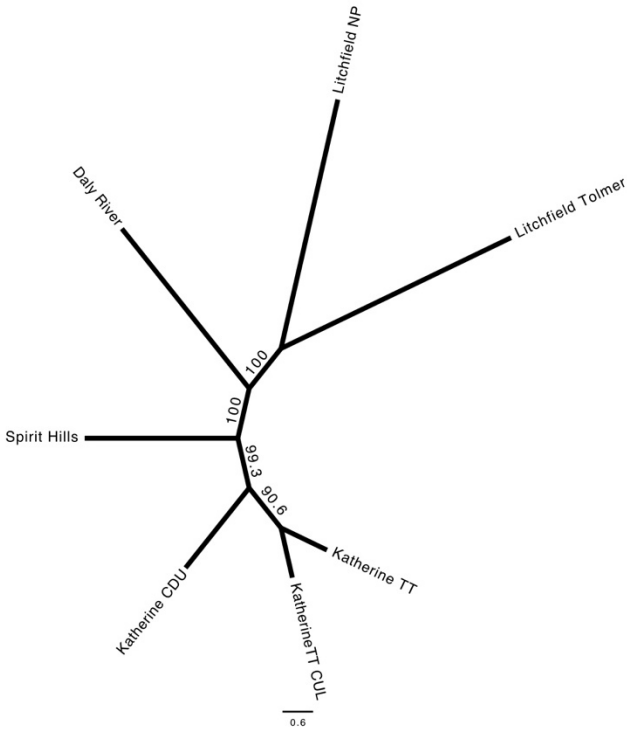

Figure 5 Unrooted tree network tree of *C. calcicola* wild populations in Northern Territory, Australia. The unrooted tree shows seven well-supported populations. Spirit Hills, Daly River, Litchfield National Park (NP) and Litchfield Tolmer populations received 100% bootstrap support (BS). Katherine Charles Darwin University site (Katherine CDU) with 99.3% BS and both Katherine, Katherine TT Katherine TT CUL (cultivated) provided 90.6% BS. Populations from Katherine and Litchfield national park (NP) were recovered as two separate clades (99.5% and 100%).

2.4 Discussion

Here we have presented an optimised RADseq protocol used to gain insights into the genetic diversity of cycads. Our results demonstrate that RADseq can successfully be applied across all ten genera of the Cycadales, with sufficient data generated to use this approach for conservation genomics, phylogenetics, and other potential applications.

Assembly of RADseq data. Data was mapped against the reference plastomes and a mitochondrial genome, and showed that less than 8.01% of the total number of reads were mapped. This indicates that neither the plastome or mitochondrial genome were overrepresented in our data, and is confirmed by the average and maximum read depth (Tables 3 & 4). Additionally, large portions of the reference genomes covered up to 97% of the plastome and 69% of the reference mitochondrial genome. These results are expected with RADseq data as reads will rarely cover the entire reference because of the use of restriction enzymes (Liu and Hansen 2017). These results indicate that our RADseq protocol is also effective at recovering large portions of the plastome and mitochondrial genome, without reducing the effectiveness and reliability of RADseq for population genetics or phylogenetic inference (Fitz-Gibbon et al. 2017).

De-novo assembly in ipyrad recovered between 125 (*Macrozamia*) to 1,705 (*Cycas*) informative loci. This is the result of several factors: the number of high-depth clusters generated, the number of genetically similar samples included in the assembly, and the degree of genetic similarity between species and genera (Table 5). A greater number of *Cycas* species were included in the assembly, which are closer genetically (Nagalingum et al. 2011), and is the reason why a greater number of loci were retained for *Cycas,* as with the *Cycas calcicola* example dataset (Table 6). Conversely, fewer loci were recovered for Zamiaceae because of greater genetic distances between genera, and only a single representative species of each genus was included in the assembly. If more samples were included from each genus of Zamiaceae, the resulting number of loci could be greater. Despite the genetic distance among the genera, there was a sufficient number of shared loci recovered between the Zamiaceae and Cycadaceae genera. These results mirror what was found in Myricaceae (Liu et al. 2015) and Diapensiaceae (Hou et al. 2015), as they also found a significant drop in loci recovered in more distantly related taxa, indicating that genetic differences between families would be considerable.
The example assembly of *Cycas calcicola* showed a similar result in clustering to

that found in the generic dataset by having far fewer high-depth clusters than clusters overall. The assembly generated 4,421 markers across 72 samples using a strict minimum number of samples per locus (to reduce missing data), which required that each locus was present in at least 43 samples (~60%). If the minimum samples per locus was reduced to the default of four, this would further increase the number of loci generated, but also the amount of missing data. This demonstrates that with a good number of samples and a high level of generic similarity, an assembly can generate a suitable number of loci even with very large genomes. This approach also provided sufficient data for coalescent-based analysis, since our results recieved high support (>90% BS) for closely related populations of *C. calcicola*.

Sequencing depth and large genomes. Sequencing resulted in 2.7 to 9.8 million paired-end-reads per sample. Although reads needed to be filtered and trimmed, the sequencing quality was high. We aimed to obtain 1 GB per sample to account for the large genome size (25-63 Gbp; Zonneveld 2012) and overrepresentation of the plastome (Wu and Chaw 2015). The amount of data (uncompressed) ranged from 1.2 GB for *Stangeria eriopus* to 3.9 GB in *Macrozamia johnsonii*, thereby meeting our goal.

One of the main considerations in assembling RADseq data is the clustering of reads for calling consensus sequences and SNPs, as this requires numerous repeat reads to be aligned (Eaton 2014). In the third step of assembly in ipyrad, if two or more reads align, they form a cluster. Subsequently, these clusters are further assessed, and six or more reads (depending on minimum depth clustering depth set) are required for a cluster and its constituent SNPs to be considered reliable—these are termed high-depth clusters (Eaton 2014). However, in larger genomes, it is less likely that there will be a sufficient number of repeat reads in the sequence data to generate enough high-depth clusters (except for repetitive regions) (Karam et al. 2015). In our study, we found 1 to 3.3 million clusters in the first clustering step, and 32,000 to 138,000 clusters after selecting only high-depth clusters. This indicates that there were many clusters with fewer than six reads. The number of high-depth clusters, while relatively small compared to the initial number, is nonetheless sufficient for downstream phylogenetic and population genetic purposes, especially given that previous work has used significantly fewer markers (Griffith et al. 2015; Salas-Leiva et al. 2014; Meerow et al. 2012; Nagalingum et al. 2011; Cibrián-Jaramillo et al. 2010).

Thus far, RADseq has been used in phylogenetics and population genetics for a few plant groups with varying genome sizes (Table 1). The taxa with the smallest genomes (<1 Gbp) were *Carex* spp., *Sisymbrium austriacum*, and *Mimulus* spp., whereas those with the largest genomes include *Diospyros* species (2.40-5.76 Gbp), *Senecio lautus* (4.90 Gbp), and *Pedicularis* species (5.68 Gbp). In our study, RADseq was applied to genomes that are 25 to 63 Gbp - i.e. approximately 4 to 11 times larger than all previous studies. Therefore, we have demonstrated that RADseq can successfully be applied to groups of plants with larger genomes and holds a promise for future applications of RADseq to other plant groups, especially non-flowering plants with large genomes such as ferns and gymnosperms.

Conclusions. We have demonstrated that RADseq can be applied to organisms with large genomes, such as cycads. This protocol uses high throughput sequencing to recover informative genome-wide markers. RADseq also offers the ability to multiplex and sequence many individuals simultaneously, at relatively low cost. These markers have the potential to be used for population level and for phylogenetic studies, ultimately helping to resolve the relationships among cycads, obtain a better insight into the genetic diversity among the Cycadales species, and to assist in developing informed conservation management plans for cycads and other groups in the future.

Chapter 3 - Conservation Concerns: Low Genetic Diversity in the Australian cycad, *Cycas calcicola* and the Absence of Key Genotypes in Botanic Gardens

The risk of extinction for many species is increasing, with cycads having the highest risk of extinction among all plants (Nagalingum et al. 2011). As many species become rarer, their conservation is ever more important. Understanding genetic diversity is fundamental for conserving species. Here we obtained insights into the genetic diversity of *Cycas calcicola* L. (Cycadaceae, Cycadales), a species of conservation concern and endemic to the Northern Territory, Australia. Genetic material was collected from wild populations in the Litchfield National Park and the Katherine regions, representing six populations and comprising 60 samples. Botanic garden collections play a vital role in conservation as they offer a repository for the genetic diversity of wild populations, so we also included 12 samples from ex-situ collections. These represented the Spirit Hills Conservation Reserve and Daly River regions. We used next generation sequencing in the form of RADseq following an established ezRAD method modified for cycads. De-novo assembly using ipyrad recovered 2271 informative genome-wide markers. Population genetic statistics revealed very low levels of gene diversity (H_E = 0.023 to 0.116), evidence of inbreeding, and a significant departure from the Hardy-Weinberg equilibrium (Mean F_{IS} = 0.491). The populations also showed some differentiation among regions and populations (6%) and little genetic distance between populations (F_{ST} = ≤ 0.248). Population structure and discriminate analysis confirmed that geographic isolation correlates with genetic differentiation between populations in the Katherine and Litchfield regions. Overall, low levels of genetic diversity and high levels of inbreeding will have a significant impact on the adaptability of the species. We found that the ex-situ collections did not represent the genetic diversity of the natural populations, as the Litchfield National Park populations were unrepresented. We recommend that, despite a lack of genetic differentiation between the two regions, populations of *C. calcicola* should be conserved as separate management units so as to adequately conserve what diversity remains.

3.1 Introduction

The risk of extinction in plant species is increasing worldwide due to habitat fragmentation, climate change, land clearance and, in some cases, over-collection from natural populations (Newbold et al. 2016). The introduction of invasive species adds further pressure to wild populations by sometimes outcompeting native species (Vilà et al. 2011). Because of these pressures imposed upon natural populations, their conservation is becoming ever more important to help preserve biodiversity (Hefley et al. 2016). Cycads have the highest risk of extinction of any group of seed plants, so their conservation is of paramount importance (Donaldson 2003).

Cycads represent the oldest group of extant seed plants (Nagalingum et al. 2011). They consist of 351 accepted species in 10 genera from two families: Zamiaceae and Cycadaceae (Calonje et al. 2019). Cycadaceae is monotypic with the sole genus *Cycas* L. *Cycas* is the largest genus of Cycadales, consisting of 117 extant species. The genus has a pantropical distribution and is found in Madagascar, throughout Asia, the Pacific Islands and Australia. Australia represents a biodiversity hotspot for cycads, where *Cycas* is represented by 38 of the 117 known species. The Australian species are distributed throughout Northern Australia including; Western Australia (three species), Northern Territory (16 species) and Queensland (19 species).

Most parts of cycads are poisonous to livestock, including the leaves, sap, and seeds (Norstog & Nicholls 1997). This toxicity has often caused them to be cleared from arable land in order to limit accidental poisoning (Hall & Walter 2014; Hall & McGavin 1968). Cycads are also highly prized in horticulture, with some species being sold for thousands of US dollars (Donaldson 2003). The ornamental appeal of cycads has generated a great demand, causing over-collection and illegal removal from wild populations (Torgersen 2017; Pérez-Farrera et al. 2006). These factors cause breaks within and between populations leading to fragmentation. Fragmentation can prevent gene flow and further exacerbate isolation, leading to high genetic differentiation between populations, with a risk of increased inbreeding within the fragments (Young et al. 1996).

The increasing range and intensity of threat means that the in-situ conservation of many plant taxa is becoming increasingly urgent (Whitlock et al. 2016). As a result, many botanic gardens have made it their mission to play a critical role in the conservation of species (Nikitsky Botanical Gardens 2017). The living ex-situ plant collections of botanic gardens can harbour a significant amount of genetic diversity, representing that of natural in-situ populations (Cibrian-Jaramillo et al. 2013; Dosmann 2006). Thus, these ex-situ collections can be used to conserve genetic diversity of wild populations (Fant et al. 2016).

It is critical to consider targeted collection of genetic material from natural populations for conservation purposes (Griffith et al. 2015). With careful management, botanic gardens can successfully be used to conserve wild populations (Griffith et al. 2014). Genetic diversity representing natural populations can be introduced into ex-situ conservation collections, and in some cases, has the potential to help replenish the genetic reserves of depleted natural populations (Volis 2017). Seed banks are considered to be a cost-effective method by which to store the seeds of most plant species almost indefinitely (Hamilton 1994), and have the capability of maintaining genetic diversity of species. Seed banks are not, however, an option for plants such as cycads the seeds of within have a very short period of viability (<1 year) under conventional storage regimes (Calonje et al. 2011; Nadarajan et al. 2018; Mondoni et al. 2011). The only way to conserve cycads ex-situ is through living plants, but to do this effectively we need to understand the genetic diversity of the populations (Hurka 1994).

Conservation genetics provides a framework to guide both conservation and restoration to minimise the risk of extinction imposed upon species like cycads (Kramer & Havens 2009; Frankham et al. 2004). The long-term aim of conservation genetics is to understand the genetic variation in wild populations, and to determine if populations contain enough variation for future adaptation, expansion and reestablishment (Paz-Vinas et al. 2018; Yoder et al. 2018; Hedrick & Miller 1992). Many in-situ conservation plans have been informed by measuring genetic factors that affect the overall dynamics of populations, such as decreases in population size, past bottlenecks and sex-specific gene flow (Zhang et al. 2018; Ahrens et al. 2017). Conservation genetics has also been used to identify populations with high levels of genetic diversity, and to help select which of these populations should be prioritised for conservation (Hou et al. 2018; Rodríguez-Rodríguez et al. 2018; Drury et al. 2017).

However, despite the effectiveness of in-situ conservation is not optimal for all plant groups (Fay 2018), and many conservation genetic studies on cycads have focused on species already at high risk of extinction, biasing results (Swart et al. 2018; Feng et al. 2014; Da Silva et al. 2012; Pinares et al. 2009; Long-Qian & Xun 2006). This is the case for many cycads with populations that are already showing declines in genetic diversity (Cabrera-Toledo et al. 2012; Da Silva et al. 2012; Octavio-Aguilar et al. 2009; González-Astorga et al. 2008; Shuguang et al. 2006). The results of these studies showed that cycads exist in small and fragmented populations, especially in Africa (Da Silva et al. 2012; Ekué et al. 2008) and Central America (Cabrera-Toledo et al. 2010). This fragmentation is reflected in the genetic diversity where populations show evidence of drift resulting in a loss of alleles (Zhan et al. 2011; González-Astorga et al. 2008).

Geographic isolation often results in high genetic differentiation between populations (Long-Qian et al. 2004; Keppel et al. 2002). Populations connected by gene flow usually have greater gene diversity but overall lower genetic differentiation. This is because the populations are less genetically differentiated and, therefore, similar genetically (Huang et al. 2004; Yang & Meerow 1996). Over-collection and land clearance reduces the effective population size, increases homozygosity of a population due to inbreeding, and slowly reduces the overall genetic diversity of the species (Meerow et al. 2012; Long-Qian et al. 2004). However, genetic diversity in cycads is not always correlated with the size of the populations, as smaller and isolated populations can contain high levels of genetic diversity, despite population retraction (Gong et al. 2015). This is likely due to historic gene flow caused by the slow reproductive times in cycads, and is often correlated with low variation between populations (James et al. 2018; Cibrián-Jaramillo et al. 2010; González-Astorga et al. 2008).

Australian populations of *Cycas* are often large and less disturbed (Liddle 2009) compared to those of other countries. Of the 38 species of *Cycas* in Australia, only one study into the conservation genetics of an Australian *Cycas* has been conducted (James et al. 2018). This makes Australia a prime location in which to further develop an understanding the population genetics, and evolutionary processes and patterns of cycads.

Cycas calcicola for many is considered an iconic Australian cycad (Figure 6). The species occurs near or on limestone in light bush or rocky outcrops, in large but disjunct populations in the Litchfield National Park, Daly River, and Katherine regions of the Northern Territory (Figure 7) (Jones 2002; Hill 1996). A disjunct population of *C. calcicola* also occurs within the Spirit Hills conservation site (Hill 1996). *Cycas calcicola* has an arborescent caudex ≤ 5 m in height, and is easily distinguished from other Australian *Cycas* by its glossy dark green leaflets having recurved margins that are covered in silvery-grey hairs (Hill 1996). *Cycas calcicola* is insect pollinated (Kono & Tobe 2007), and although the pollinator of *C. calcicola* has not been recorded, the species is said to have no pollinator specificity (Liddle 2009). Some *C. calcicola* populations occur in close proximity to *Cycas armstrongii* Miq., which is known to be pollinated by two species of beetle in the Tenebrionidae (Ornduff 1992). Although *C. calcicola* is not known to hybridize with *C. armstrongii* it is likely that the species share pollinators (Liddle 2009). Seed dispersal for *Cycas* growing in the Northern Territory is rarely greater than 300 cm from a mother plant, and greater dispersal distances have not been recorded (Watkinson & Powell 1997).

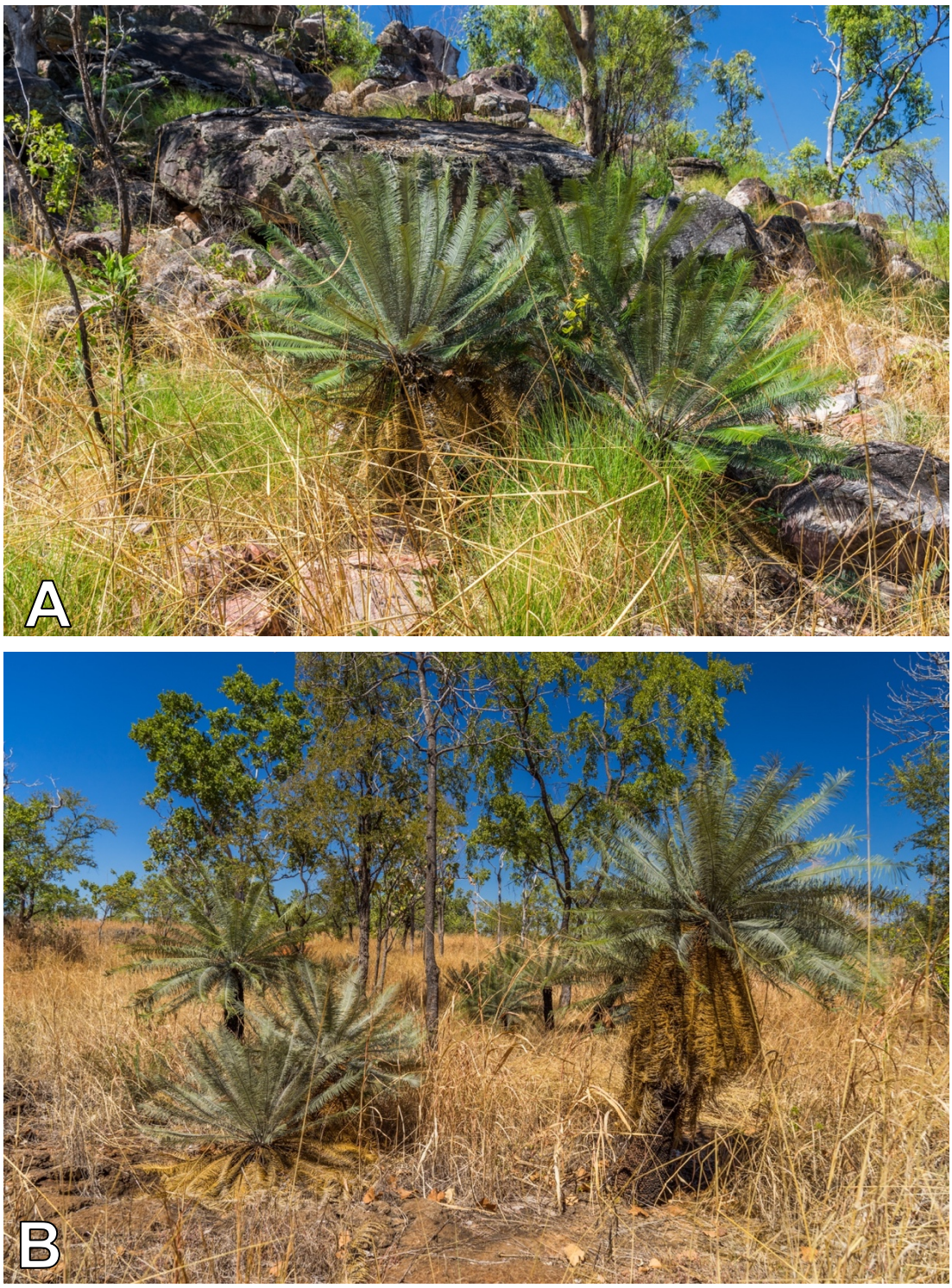

Figure 6 *Cycas calcicola* populations growing in the wild in the Katherine region and Litchfield National Park. (A) Part of a large population of *C. calcicola* growing on sandstone in the Litchfield National Park, Northern Territory. (B) Small group *C. calcicola* growing on limestone in the Katherine region.

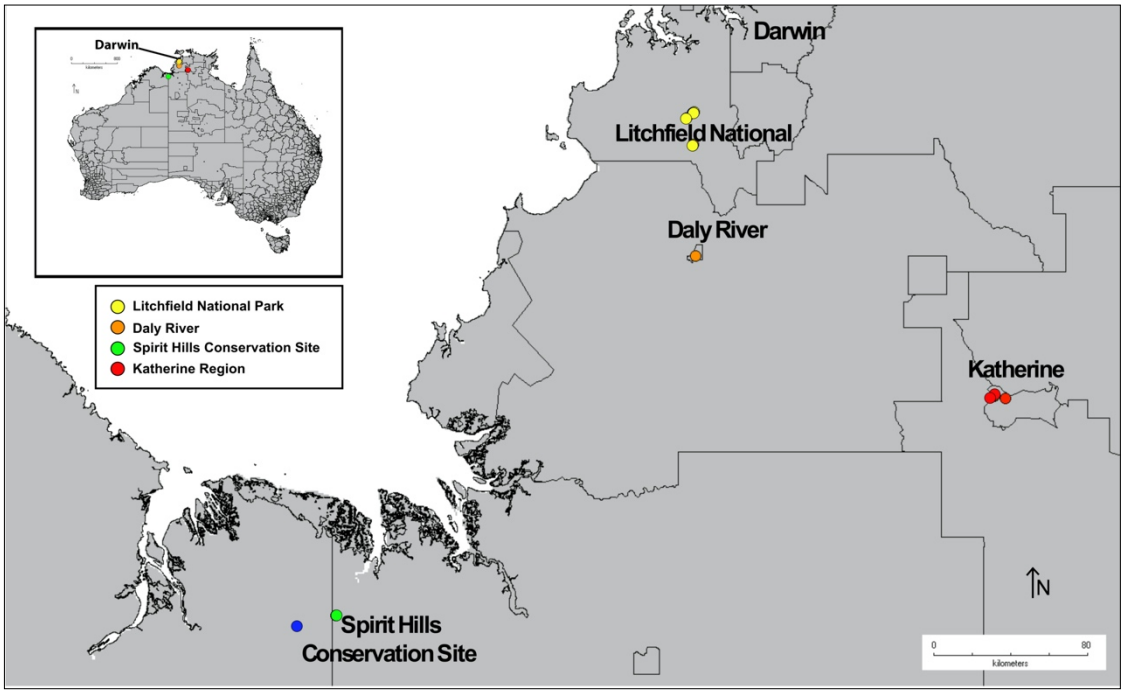

Figure 7 Distribution of samples of *C. calcicola* in Northern Territory. Map of northern region in the Northern Territory, Australia showing sampling sites of wild (Litchfield and Katherine) and ex-situ conservation collections, representing the entire range of the species (Spirit Hills and Daly River).

Cycas calcicola populations are largely undisturbed, yet there is some evidence of recent population contraction due to increased burning, habitat reduction and clearing for farmland or roads (Liddle 2009). The most recent IUCN Red List conservation assessment for *C. calcicola* considered the species as Least Concern (IUCN 2019). The assessment noted that there are significant disjunctions between populations, and that there is evidence of a decline in the number of individuals in the Katherine region due to over-collection and habitat loss (Liddle 2009).

Here we investigated the genetic diversity of *Cycas calcicola* growing in Australia using genetic material collected from natural populations and from ex-situ collections of botanic gardens. To do this we used next generation sequencing in the form of RADseq, which allowed us to find informative markers throughout the genome of *C. calcicola*. The retrieved markers were either neutral or non-neutral, and subject to the full evolutionary history of the species (Andrews et al. 2016). The genetic data were used in two areas: firstly, to investigate the genetic diversity within and between populations of the Litchfield and Katherine regions, where the majority of the populations are found; secondly, to determine if the ex-situ collections adequately represent the genetic diversity of the wild populations.

3.2 Materials and methods

Sampling strategy. Silica-dried leaflets of *Cycas calcicola* were collected from wild populations within Litchfield National Park and the Katherine region in the Northern Territory, Australia (Table 7). Populations were selected based on herbarium specimens and The Australasian Virtual Herbarium (AVH; https://avh.chah.org.au, accessed 12th January 2015). A total of 60 individuals were sampled from six populations: three populations from Litchfield National Park and three from the Katherine region (Figure 7). For each population, ten individuals were sampled from plants of varying ages (juvenile to mature) and bearing microsporangiate or megasporangiate strobili, In addition, a further 12 samples were obtained from cultivated ex-situ collections: George Brown Darwin Botanic Garden (Darwin, Northern Territory, Australia) and Montgomery Botanical Center (Miami, Florida, USA). The ex-situ conservation material came from plants of known wild origin and represented the Katherine, Daly River, and Spirit Hills populations. In addition to the tissue sampling, we also gathered basic population demographics, which were recorded for each population (Table 7).

Table 7 Population demographics for Cycas calcicola populations in Northern Territory. Litchfield 1 and 2 = Litchfield National Park, Litchfield-Tolmer = Tolmer Falls in the Litchfield National Park, Katherine-TT = Katherine population off the Stuart Highway near old used train tracks, Katherine-CDU 1 and 2 = Charles Darwin University campus in Katherine. Katherine-CUL = cultivated samples from botanic garden wild collected from Katherine region. Samples per population = number of individual herbarium vouchers and DNA samples collected at each population, Height of caudex (cm) = height of caudex of each plant vouchered, No. seedlings = number of seedlings observed with caudex height < 2 cm, No. immature = number of immature observed with caudex height ≥ 2cm , No. large plants = number of nature plants observed able to bear strobili with caudex height ≥ 50 cm and No. male (♂) or No. female (♀) = number of plants bearing either microsporangiate or megasporangiate strobili.

Region	Population	Samples per population	Height of caudex (cm)	No. seedlings	No. immature	No. large plants	No. male (♂)	No. female (♀)
Litchfield	Litchfield-1	10	2-170	8	7	20	2	3
	Litchfield-2	10	2-100	6	11	25	0	2
	Litchfield-Tolmer	10	2-220	20	30	200+	6	3
Katherine	Katherine-TT	10	6-230	20	10	70+	3	4
	Katherine-CDU 1	10	20-180	10	20	100+	2	3
	Katherine-CDU 2	10	5-220	10	20	100+	2	3
	Katherine-CUL	7	N/A	N/A	N/A	N/A	N/A	N/A
Daly River	Daly River	4	N/A	N/A	N/A	N/A	N/A	N/A
Spirit Hills	Spirit Hills	2	N/A	N/A	N/A	N/A	N/A	N/A

DNA extraction and quantification. Approximately 0.05 g of silica-dried leaflets were ground to a fine powder using a TissueLyser (Qiagen, Hilden, Germany). When present in large amounts, trichomes were removed using a wire brush to improve extraction yield (common with *C. calcicola*). High molecular weight genomic DNA was extracted using a DNeasy Plant DNA Extraction Mini Kit (Qiagen, Hilden, Germany). Genomic DNA was inspected using a 2% agarose gel to check for the presence of DNA and impurities. DNA extractions were quantified using an Invitrogen Qubit broad range (3.0 BR DNA assay; Invitrogen, Life Technologies, Carlsbad, CA, USA) fluorometer with a target concentration of 17 µg/mL; any sample that yielded less than 17 µg/mL was either re-extracted or concentrated using a 1:1 ratio of Agencourt AMPure XP sample purification beads (Beckman Coulter, Inc.) by combining multiple extractions.

DNA normalisation and restriction digest reaction. For a full protocol, see Clugston et al. (2019). First, genomic DNA was normalised to a concentration of 500 ng in 42 µL total volume (0.01 µg/mL). Second, 5 µL of NEB 10x CutSmart buffer (New England Biolabs, Ipswich, MA) and 1 µL of Bovine Serum Albumin (BSA) was added to each well. Samples were then held at 4°C for a minimum of five hours before adding restriction enzymes—the five hours of incubation helped the cutting action of the restriction enzymes. Next, double digest reactions were carried out using 1 µL each of the restriction enzymes EcoR1-HF and Mse1. Reactions were then placed into a thermocycler for three hours at 37°C with a final 20-minute enzyme deactivation step at 65°C. The reactions were then checked on 2% agarose gel for quality of digestion. Last, reactions were cleaned using 1.8:1 ratio of AMPure XP beads to sample (90 µL of AMPure XP beads to 50 µL of digested DNA) and quantified using a Qubit high sensitivity kit.

Library preparation. Libraries were prepared using an Illumina TruSeq nano high-throughput dual index library preparation kit (Illumina Inc., CA, USA). We followed a modified version of the ezRAD v3 (Toonen et al. 2013), using half of the recommended volumes of the kit to save costs (Clugston et al. 2019). Following the methods by Clugston et al. (2019) the final steps of library preparation, were modified from ezRAD protocol, by using a final bead clean, using a 0.8:1 ratio of AMPure XP beads to remove adapter dimer. Final Illumina libraries were validated using a LabChip, cleaned using a 0.9:1 ratio of AMPure XP beads and quantified using a Qubit high sensitivity kit. Final libraries were then normalised to 10 nM concentration and pooled for sequencing.

Sequencing. We aimed to capture around 1GB of sequence data per sample (in a run of 95 libraries) to: ensure adequate coverage of the large genome of *C. calcicola*, account for overrepresentation of the plastid genome, and capture as much of the nuclear genome as possible. Genomic sequencing was carried out at using an Illumina NextSeq 500 150 bp paired-end high throughput (HT) on a single flow cell. The sequencing run was spiked with 20% PhiX sequencing control V3 to account for low diversity after using enzymatic digestion in the ezRAD protocol.

Bioinformatics

Quality control and filtering of sequence reads. The NextSeq 500 generated eight raw fastq files for each sample: four forward files and four reverse files. The four forward files were combined into a single file and similarly for the reverse files for downstream analysis. Illumina reads were assessed for quality using FastQC 0.11.4 (Andrews et al. 2014). Then using Trimmomatic 0.36 (Bolger et al. 2014), reads were filtered for quality to remove Illumina adapter sequences and the first six base pairs of reads (cut sites) due to quality drop-off and cropped reads to 120 bp in length (reads dropped in quality after 120 bp). A sliding window was used to delete bases with a PhredQ score less than 20, and all reads less than 50 bp were discarded.

Assembly of RADseq data. De novo assembly of the paired-end reads was performed using ipyrad 0.7.18 (Eaton 2017) using a high-performance online instance with Amazon Web Service though the California Academy of Sciences. In ipyrad all parameters were set to default, except “data type” was set to ‘pairgbs’ (most closely matches ezRAD), bases with a ‘PhredQ score’ less than 30 were converted to 'N’ and reads with 15 uncalled bases were discarded. Reads were further filtered for adapter sequences, adapters were trimmed, and reads were discarded if they were less than 40 bp after trimming.

The ‘maximum number of uncalled bases in consensus sequences’ was set to 10 in both forward and reserve reads. The setting for ‘maximum shared heterozygotes per locus’ was left as 0.5 (default) to reduce the effects of paralogs. The ‘maximum heterozygotes in consensus sequences’ was set at eight for both forward and reverse sequences, and the “minimum number of samples per locus” was set to 43, so each SNP would be present across a minimum of 60% of samples. The high minimum samples per locus helps to reduce the amount of missing data, and reduces anomalies that may occur in population level analysis to ensure effective population genotyping (Shafer et al. 2016).

Population genetic analysis. Before comparative statistics could be calculated, samples and populations were assigned to groups referred to as “stratifications” (Table 7). These stratifications were defined as: region (Table 1), sub-region and population (Table 1). ‘Region’ was defined as the broad geographic area from which the samples were collected; Litchfield, Katherine, Daly River and Spirit Hills Conservation Reserve (Spirit Hills). ‘Sub-region’ defined populations grouped into their locality; Litchfield National Park (Litchfield NP), Litchfield National Park Tolmer Falls (Litchfield NP Tolmer), Katherine near train tracks along Stuart Highway (Katherine TT), Katherine Charles Darwin University (Katherine CDU), Katherine cultivated (Katherine CUL), Daly River and Spirit Hills. ‘Population’ defined the collected locality for each population following the designation provided in ‘subregion (Table 1). Stratifications consisted of nine populations from four regions. These stratifications were used to define groups for further downstream statistical analysis.

Descriptive statistics—number of individuals in each population (N), effective number of alleles (N_a), the effective number of alleles per locus (N_e), observed heterozygosity (H_o), expected heterozygosity (H_e), unbiased expected heterozygosity (uH_e), and fixation index (F)—were generated using GenALEx 6.5 (Peakall & Smouse 2012). To test genetic variation among populations and among individuals within a population or region we used an analysis of molecular variance (AMOVA); genetic distance based pairwise F_{ST} was calculated in GenAlEx using 9999 permutations of the dataset, with the ‘Codom-Allelic’ option selected.

STRUCTURE v.2.3.4 (Pritchard et al., 2000) was used to determine the structure of the populations of *C. calcicola* and the degree of admixture in the 72 individuals among 8 populations. STRUCTURE uses a Bayesian algorithm to infer the optimum number of distinct genetic groups *K* (clusters) by minimizing deviations from Hardy–Weinberg and linkage equilibrium within each cluster. The analyses were carried out for *K* = 1–5 using 100,000 MCMC iterations after a burnin of 20,000 steps and were repeated 10 times for each *K*, with the 'Separate Alpha for each Population' option selected.

To visualise the genetic relationships between populations, a discriminant analysis of principal components (DAPC) was carried out using Adegenet 2.1.0 (Jombart & Ahmed 2011) in R (R Core Team, 2019). The optimal number of clusters in the data and the number of principal components (PCAs) to be retained for discriminate analysis were determined using the "find.clusters" command in combination with the optimal *A*-score. A DAPC scatter plot was used to depict the relationship and connectivity of populations.

3.3 Results

Sequencing and de-novo assembly. After filtering raw sequence data, the number of reads that remained per sample ranged from 1,296,034 to 4,650,176. De-novo assembly generated 1,296,034 to 3,037,283 sequence clusters with 22,806 to 78,631 clusters containing six or more reads (referred to as a high depth cluster). The final output from ipyrad generated 2,271 SNPs recovered from 231 unique loci, across a minimum of 36 samples per locus (each locus was present for at least 50% of all individuals).

Population genetic statistics. Population genetic analysis was used to determine the genetic diversity of the species and its populations (Table 8). Gene diversity (H_E) of the populations ranged from 0.023 (± 0.004) in Spirit Hills to 0.116 (± 0.004) in Litchfield NP1, with a mean of 0.080 (± 0.001). The observed heterozygosity (H_O) ranged from 0.028 (± 0.003) in Spirit Hills and Daly River to 0.059 (± 0.004) in Litchfield NP1, with a mean of 0.039 (± 0.001). The differences between the observed and expected heterozygosity indicated that the populations were not in Hardy-Weinberg equilibrium (HWE). The inbreeding coefficient (F_{IS}, Table 8) ranged from -0.244 (-0.278 – -0.210, 95% CI) in Spirit Hills to 0.605 (0.583 – 0.591, 95% CI) in Katherine CDU1 populations. The average F_{IS} across all populations was 0.491 (0.500 – 0.482, 95% CI) indicating they were more related than would be expected under HWE.

Table 8 Summary of population genetic statistics for all populations of *Cycas calcicola*. Mean and standard error (SE) of frequency-based population genetic statistics for populations of *Cycas calcicola*, generated using GenALEx 6.5. Number of individuals in each population (*N*), number of alleles (N_A), the effective number of alleles (N_E), heterozygosity observed (H_O), heterozygosity expected (H_E), unbiased expected heterozygosity (uHe), and Fixation index (F_{IS}). Calculation of the fixation 95% confidence interval (mean±1.96*SE) lower (-95%CI) and upper limits (+95%CI).

Populations		N	N_A	N_E	H_O	H_E	uHe	F_{IS}	-95%CI	+95%CI
Litchfield NP1	Mean	5.613	1.350	1.190	0.059	0.116	0.135	0.463	0.438	0.489
	SE	0.047	0.010	0.007	0.004	0.004	0.004	0.013		
Litchfield NP2	Mean	5.109	1.262	1.143	0.038	0.096	0.111	0.566	0.542	0.590
	SE	0.049	0.010	0.007	0.003	0.003	0.004	0.012		
Litchfield NP Tolmer	Mean	5.134	1.255	1.138	0.035	0.092	0.105	0.568	0.545	0.591
	SE	0.049	0.010	0.007	0.002	0.003	0.004	0.012		
Katherine TT	Mean	5.412	1.288	1.151	0.042	0.098	0.114	0.568	0.543	0.593
	SE	0.050	0.010	0.007	0.003	0.003	0.004	0.013		
Katherine CDU1	Mean	5.210	1.241	1.136	0.032	0.088	0.100	0.605	0.583	0.591
	SE	0.045	0.009	0.006	0.002	0.003	0.004	0.012		
Katherine CDU2	Mean	5.184	1.263	1.138	0.047	0.091	0.104	0.456	0.431	0.481
	SE	0.041	0.010	0.006	0.003	0.003	0.004	0.013		
Katherine CUL	Mean	3.751	1.161	1.094	0.040	0.070	0.085	0.382	0.356	0.407
	SE	0.035	0.009	0.007	0.003	0.003	0.004	0.013		
Daly River	Mean	1.811	0.950	0.931	0.028	0.044	0.061	0.324	0.291	0.357
	SE	0.026	0.010	0.010	0.003	0.003	0.004	0.017		
Spirit Hills	Mean	0.890	0.686	0.680	0.028	0.023	0.037	-0.244	-0.278	-0.210
	SE	0.016	0.012	0.012	0.003	0.002	0.004	0.017		
Total	Mean	4.235	1.162	1.067	0.039	0.080	0.095	0.491	0.500	0.482
	SE	0.018	0.004	0.003	0.001	0.001	0.001	0.004		

Genetic structure and population differentiation. The AMOVA results (Table 9) showed the same level of differentiation at the regional (6%) and population level (6%), with the majority of genetic variation contained within populations (89%). Despite the relatively large geographic distance between the Litchfield and Katherine regions (~250 km between regions) (Figure 7), genetic differentiation is low. Pairwise F_{ST} values (Table 10) indicated low to medium levels of genetic distance between most populations, indicating high levels of gene flow. F_{ST} values ranged from 0.048 between Litchfield NP1 and Litchfield NP2 to 0.248 between Daly River and Spirit Hills. There is also evidence of higher levels of genetic differentiation between populations occurring in the Katherine and Litchfield regions which ranged from 0.082 between Katherine CDU1 and Litchfield NP2 to 0.115 between Katherine CUL and Litchfield NP2, supporting greater geographic distance between regions (Figure 7). The populations with the greatest degree of genetic

differentiation from all other populations were Daly River (F_{ST} = 0.105 - 0.175) and Spirit Hills (F_{ST} = 0.177 - 0.248) (Figure 7).

Table 9 Analysis of molecular variance for *Cycas calcicola* populations. Results for Analysis of Molecular Variance (AMOVA) for populations of *Cycas calcicola*. Df = Degree of Freedom, Among populations = genetic variation among the populations within the region and Within populations = degree of genetic variation within the populations in a region.

Source of Variation	Df	Sum of squares	Estimated Variance	Percentage of variation (%)
Among Regions	3	5090.4	31.748	6
Among populations	5	5351.8	30.836	6
Within populations	135	492.560	492.143	89

Table 10 Pairwise distance based F_{ST} matrix of *Cycas calcicola* populations. Genetic distance based pairwise F_{ST} matrix from AMOVA analysis generated using GenALEx 6.5 for *Cycas calcicola* populations. Values greater ≥ 0.20 are highlighted in bold.

	Litchfield NP1	Litchfield NP2	Litchfield NP Tolmer	Katherine TT	Katherine CDU1	Katherine CDU2	Katherine CUL	Daly River
Litchfield NP2	0.048							
Litchfield NP Tolmer	0.056	0.058						
Katherine TT	0.092	0.101	0.085					
Katherine CDU1	0.082	0.098	0.092	0.058				
Katherine CDU2	0.084	0.104	0.090	0.067	0.052			
Katherine CUL	0.095	0.115	0.097	0.050	0.053	0.078		
Daly River	0.133	0.115	0.105	0.175	0.143	0.158	0.160	
Spirit Hills	0.193	0.189	0.182	**0.216**	0.177	**0.221**	**0.201**	**0.248**

Population structure analysis. Population structure analysis found the most likely number of genetic groups to be *K*=2 (Figure 8). The structure plot shows genetic differentiation between populations in the Litchfield and Katherine region, with some admixture. Spirit Hills and Daly River show populations have a closer genetic relationship Litchfield than the Katherine region, with significant admixture. Overall there was less admixture in the Katherine populations than Litchfield, Spirit Hills and Daly River.

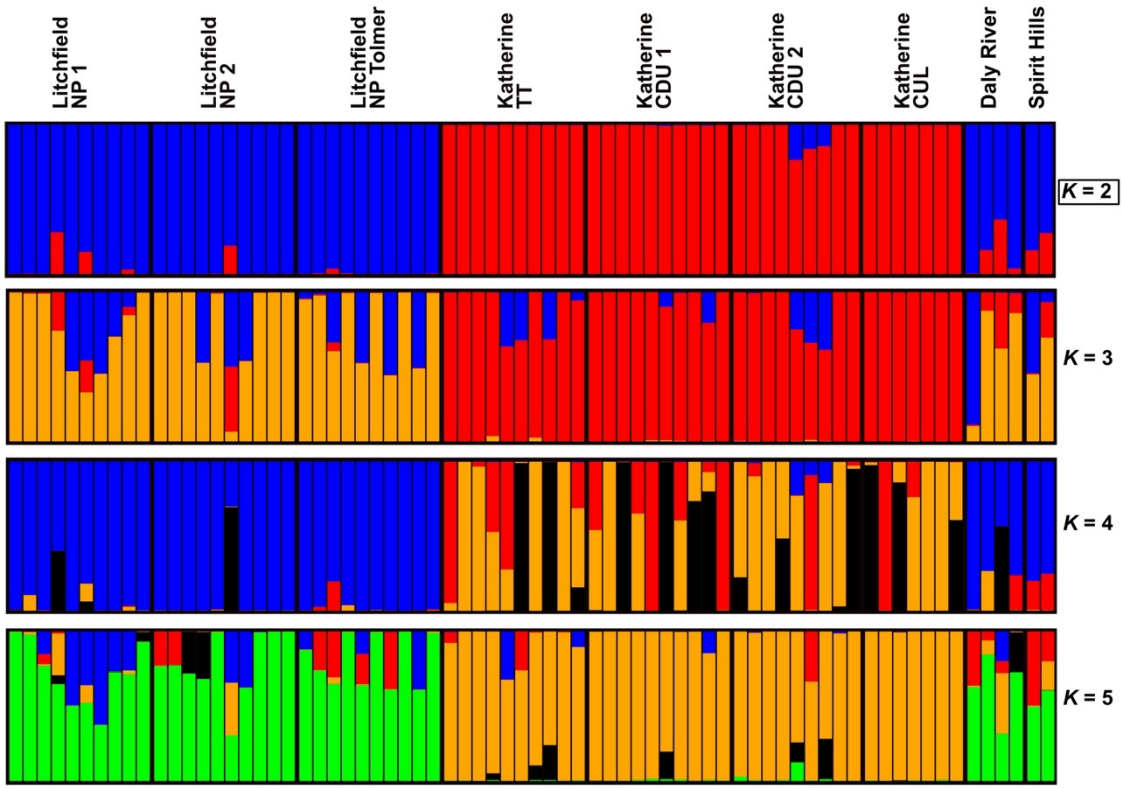

Figure 8 Population structure plot for *Cycas calcicola* populations. Population structure plot represents 72 samples from eight populations of *C. calcicola*. The most likely number of genetic groups for the species was *K*=2 indicating two clusters within the data.

Discriminant analysis of principal components. For the discriminant analysis of principal component (DAPC) (Figure 9), 14 principal components were retained for the PCA, which comprised three genetic groups (*K* = 3) with a proportion of conserved variance = 0.391. These results indicate a close genetic relationship exists among populations in the Katherine and Litchfield region. However, there is evidence genetic differentiation between the Litchfield and Katherine regions. Additionally. The results show that populations occurring in Spirit Hills and Daly River are genetically closer to populations of the Litchfield region.

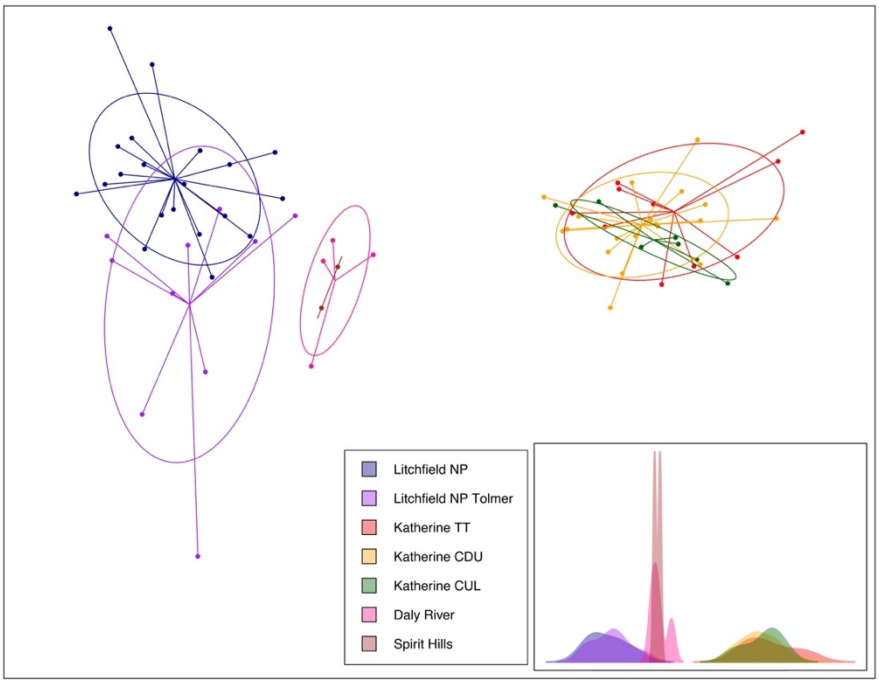

Figure 9 DAPC graph of *Cycas calcicola* populations. Discriminate analysis of principal component shows all seven populations of *C. calcicola*, representing both the Litchfield and Katherine regions, in the Northern Territory, Australia. DAPC is a summary of 14 PCs with three discriminate functions (*K* = 3) and a proportion of conserved variance of 0.391.

3.4 Discussion

Genomic data can play an important role in aiding our understanding of the genetic diversity of species by identifying factors that may affect the genetic fitness and evolutionary potential of populations (Young et al. 1996; Frankham 2003). The aim of this study was to investigate the genetic diversity of *Cycas calcicola* using genomic data, and to assess if ex-situ collections represent the genetic diversity of wild populations. The results of this study indicate low levels of genetic diversity in *C. calcicola* (H_E =0.080 ± 0.001, H_o = 0.039 ± 0.001), and high levels of inbreeding (F_{IS} = 0.491 ± 0.004). This inbreeding is likely to be an effect of small population size and small number of individuals participating in reproductive events (Lönn & Prentice 2002; Frankham 2003; Szczecińska et al. 2016). The effects of inbreeding will have a direct impact on the genetic fitness of the populations, as it will increase in the expression of deleterious recessive alleles and can significantly increase the risk of extinction (Wright et al. 2007).

Genetic diversity. Estimating allelic diversity and heterozygosity allows us to gain an insight into the genetic health of populations (Hughes 2008). These factors can be used to determine the stresses on a population and its adaptive potential to ever changing environmental conditions (Pauls 2013). The few studies on genetic diversity of Australian cycads have focused on mostly *Macrozamia* Miq. and found low genetic diversity and poor population differentiation (Sharma et al. 2004; Sharma et al. 1999; Sharma et al. 1998). Although the populations of *C. calcicola* show evidence of a reduction in population size, the populations are still considered to be large and in good health (Liddle 2009).

The low levels of genetic diversity in *C. calcicola* are likely indicative of numerous factors including small population sizes, habitat fragmentation and poor seed dispersal leading to isolation and fragmentation of populations, as found in other species of cycads (Cibrián-Jaramillo et al. 2010; Octavio-Aguilar et al. 2009; Keppel et al. 2008; Keppel et al. 2002; Vovides 1990). High levels of inbreeding in many populations will play a role in limiting the allelic diversity of the populations and although cycads favour outbreeding, only a few individuals in a population will usually participate in a given reproductive event (Terry et al. 2012; Suinyuy et al. 2009; Vovides et al. 1997). In populations which are small and fragmented, this can contribute to higher levels of inbreeding and would have a dramatic impact on the genetic fitness and adaptive potential in populations of *C. calcicola* (Charlesworth 2003).

Low levels of genetic diversity of cycads in Australia are not restricted to *C. calcicola.* Sharma et al. (2004) found that in the *Macrozamia plurinervia* (L.A.S.Johnson) D.L. Jones complex there was little genetic diversity within populations, poor differentiation between species and inbreeding. Their results indicated that an evident lack of heterozygotes in populations, which would impact the genetic fitness and future ability of populations and species. Low genetic diversity in cycads is not only restricted to Australia. Keppel et al. (2008) found a similar pattern in *Cycas seemannii* A.Br. and *Cycas simplicipinna* (Smitinand) K.D.Hill where the populations of each species had similar genetic diversity to that of *C. calcicola* (Feng et al. 2014). However, for other groups on conifers including *Pinus* L. and *Picea* Mill. high density and genomewide SNPs markers showed much higher levels of expected and observed heterozygosity then found in *C. calcicola* (≥ H_E 0.24 – H_O 0.25) (De La Torre et al. 2019; Plomion et al. 2016; Chen et al. 2013; Namround et al, 2008). Which shows that conifers in generally show much higher levels of genetic diversity to cycads growing in Australia. These indicate that low genetic diversity and inbreeding is a trend in cycad populations and not *gymno*sperms, which is perhaps an effect of population fragmentation resulting in a loss of genetic fitness in their often small and isolated populations (Octavio-Aguilar et al. 2009; Keppel et al. 2008; Hall et al. 1996).

Genetic differentiation. Understanding genetic differentiation between populations and species is important in measuring and understanding gene flow between populations, regions and species (Manel et al. 2003). Low levels of population differentiation among populations in most cases is indicative of recent gene flow (Storfer 1999; Slatkin 1981). Many cycad populations have high levels of genetic differentiation, and is perhaps due to population fragmentation (Calonje 2013; Meerow et al. 2012; Meerow & Nakamura 2007). In *C. calcicola* although populations and regions showed an equal amount of genetic distance (AMOVA = 6%, Table 9), with the majority of the variation being at the population level. There was still twice as much differentiation between regions than between populations within regions (Table 10). For *Cycas calcicola* populations are also geographically disjunct (Figure 7), especially those of the Litchfield and Katherine regions, with nearest populations of each region being separated by ~250 km. This is generally what would be expected, as in many plant groups geographic disjunction of populations is correlated with high levels of genetic differentiation (Muriira et al. 2018; Yang et al. 2016).

This is further supported by population structure (Figure 8), and discriminate analysis (DAPC, Figure 9), which shows evidence of genetic differentiation among population in the Litchfield and Katherine regions. Although, estimates of pairwise genetic distance confirmed low levels of genetic distance among populations (F_{ST} = 0.082 to 0.115, Tables 9 and 10), which is usually correlated with high rates of gene flow in plants (Sork 1999). These results indicate that although there is evidence of gene flow between populations within regions (Tables 9 and 10), the geographic distance backed up by difference by genetic structure and differentiation (Table 10, Figure 8 and 9).

Similar patterns have been recorded for both *Antirrhinum majus* L. and *Euterpe edulis* Mart., where both species showed gene flow between populations despite significant geographic isolation between populations (Pujol et al. 2017; da Silva Carvalho et al. 2015). These results show that the geographic disjunction between regions and among some populations could be very recent and one of the reasons why there is still evidence of gene flow between populations (Table 10). Additionally, due to the poor dispersal mechanism in cycads, in which pollinators of many cycads rarely travel great distances even between close populations (~100 km), thereby reducing the likelihood of pollination over distances of ≥ 100 km (Norstog & Fawcett 1989). Due to these reasons it is unlikely that there is current migration among populations between regions, caused by geographic distance, and the long generation time in cycads, indicates that this geographic disjunction could be a recent occurrence (Segar et al. 2017; Sampson et al. 2016; Cibrián-Jaramillo et al. 2010).

The effects and implications of low genetic diversity and differentiation. Documenting genetic diversity in populations underpins the resilience of species to adapt and evolve in changing environments (Furlan et al. 2012). Isolation of small populations reduces the adaptive potential of a species and can result in lower genetic diversity, with increased the risk inbreeding (Finlay et al. 2017). Although *Cycas calcicola* is considered a species of least concern according the IUCN Red List of Threatened Species (IUCN 2019), the genetic evidence presented by this study suggests there has been significant reductions in the size of its populations. A relative reduction in population size can result in bottlenecks, genetic drift and reduction in genetic fitness (Oakley et al. 2019), which can lead to a loss of evolutionary potential in a population and, thus, increase the risk of extinction (Ellstrand 1993; Reed et al. 2013; de Vere et al. 2009; Kramer & Havens 2009).

For *C. calcicola* the populations show low levels of allelic diversity and high levels of inbreeding. The low levels of genetic diversity is likely the result of multiple factors including habitat fragmentation, low survival, and slow growth rates in seedlings and juvenile plants (Aguilar et al. 2019). This low genetic diversity within and differentiation between populations are likely the result of small population sizes and fragmentation caused by habitat loss (Fischer et al. 2003). Additionally, as only a small number of mature individuals participate in a given reproductive event (Clark & Clark 1987), biased sex ratios will affect the genetic diversity in populations (Rosche et al. 2018). Biased sex ratios combined with low rates of seed dispersal could indicate that a seedling is more likely to establish near a maternal plant. If a seedling bears a strobili, then there is a higher chance of back-crossing, increasing the risk of inbreeding (Furlan et al. 2012). The low levels of allelic diversity in populations could also be historic and the results of recent extinction events (Crisp & Cook 2011), continuous population retraction, a slow reproductive rate, and small population sizes which are all likely to be contributing factors, which are indicative of low genetic fitness of the population, resulting in poor future survivability of *C. calcicola* (Charlesworth 2003).

Do ex-situ collections represent wild diversity? The importance of ex-situ collections – often held by botanic gardens - for the conservation of plant species is critical, as they represent a safety net to conserve the genetic diversity of natural populations (Fant et al. 2016). One aim of this research was to establish if ex-situ conservation collections currently contain the genetic diversity of the natural populations for *C. calcicola*. This study found that populations from the Litchfield National Park are not represented in ex-situ botanic garden collections, and so a unique component of the genetic diversity of *C. calcicola* is not currently being preserved. Although there is little genetic differentiation between the Litchfield and Katherine regions (Figure 8 and Table 9), there is evidence of genetic differentiation among populations in the regions (Table 10). The increasing rarity of cycads in the wild is caused by a combination of habitat fragmentation, over-collection and climate change, making it increasingly important to safeguard natural populations (Kramer & Havens 2009). Ultimately, the only way to safeguard the survival of species is to preserve their genetic diversity in the well-curated living collections of botanic gardens (Griffith et al. 2015).

The future of *C. calcicola* wild populations. Although *C. calcicola* shows little genetic differentiation between populations, there is greater genetic differences between the Lichfield and Katherine regions, likely caused by small population sizes and geographical isolation (Plenk et al. 2019). This close genetic relationship between many geographically disjunct populations (Table 10) can be explained by recent population contractions caused by changes in climate and/or habitat destruction (Salas-Leiva et al. 2017; Feng et al. 2016). A scenario of widespread climate-induced fragmentation might well be exacerbated by the slow generation time of cycads (Roodt et al. 2017; Dehgan & Yuen 1986). Small population sizes would not provide enough genetic diversity for polymorphisms to differentiate amongst regions or populations (Lowe 2008). With low evolutionary rates and slow generation times, it is unlikely that gene flow would reoccur because of limited pollination and dispersal mechanisms in *C. calcicola* (Edwards et al. 2017). This situation is concerning and raises questions about the future and conservation of the species.

The low genetic diversity and high levels of inbreeding in *C. calcicola* have implications for its conservation (Coates et al. 2018). Low genetic diversity and inbreeding means that the genetic fitness of the species is likely to be affected, thereby hindering its potential recovery and/or expansion (Amos & Balmford 2001; Ellstrand 1993). The best approach to conserve *C. calcicola* is to target seed collection for future ex-situ collections. Seeds should be collected from multiple individuals and populations to enhance the representation of the genetic diversity of *C. calcicola* in cultivation, which would be used to safeguard in-situ populations (Amos & Balmford 2001; Ellstrand et al. 1990). Although there is little genetic differentiation between the regions of *C. calcicola,* we would still recommend that each population should be regarded as a separate conservation management unit, and that germplasms from each should be conserved to represent the genetic diversity of wild populations. This, in turn, would facilitate reintroduction from ex-situ collections back into the wild if needed (Cohen et al. 1991).

Considerations for further research. Tissue samples covering the known geographic range of *C. calcicola* (Hill 1996) were obtained from wild populations of Litchfield National Park and Katherine regions (Table 7), and from ex-situ collections representing Spirit Hills Conservation Reserve and the Daly River region. To capture the genetic diversity of the species and its populations we collected 10 samples from each wild population, with additional samples from ex-situ botanic garden collections (Table 7). However, fewer samples were obtained for Spirit Hills

and Daly River which were represented by two and four samples respectively (Table 7). Although RADseq can adequately capture the genetic diversity of populations with ~5 individuals (Nazareno et al. 2017), further samples from Spirit Hills Conservation Reserve and the Daly River might provide a deeper understanding of the genetic diversity of these populations.

For botanic gardens to successfully represent the diversity of *C. calcicola,* populations from the Litchfield National Park need to be in cultivated. Additionally greater genotyping efforts would allow for a better estimation of the optimal number of individuals required to conserve the genetic diversity of *C. calcicola* in cultivation (Griffith et al. 2015). Populations in ex-situ collections have the added benefit of being more accessible for scientific and horticultural research, and would aid in the study of physiological and reproductive factors that may have contributed to the rarity of the species in the wild (Chen et al. 2012).

Conclusions

This study represents the first application of conservation genetics studies using RADseq on cycads and has provided important insights into the genetic history and diversity of *Cycas calcicola* in Australia. By using genetic samples from in-situ wild populations and ex-situ botanic garden collections we were able to in generate thousands of genome-wide informative markers. Our results recorded very low levels of genetic diversity and low genetic differentiation among populations of the species, which has led to inbreeding. These results are indicative of isolation and fragmentation of the population which may result in a loss of genetic fitness and have a significant impediment for the future adaptability of the species. These results are pertinent in the formulation and implementation of conservation strategies for this rare species. However, our results support genetic differentiation between the Litchfield and Katherine regions, where the majority of the populations are found, with correlates with significant geographic disjunction between the regions. The absence of ex-situ collections from populations of the Litchfield National Park is of concern, and we recommend that priority be given to the acquisition of genetically representative material from this region, to aid in the future conservation of the species.

Chapter 4 - Next-generation population genetics in Australian *Cycas* sheds light on the origins of *C. armstrongii* and *C. maconochiei* subsp. *maconochiei*.

Cycads are at the highest risk of extinction of all seed plants. Isolation and fragmentation of natural populations are the foremost drivers of species extinction in this order. Here we present insights into the genetic diversity of two taxa of *Cycas* endemic to the Northern Territory, Australia: *Cycas armstrongii* and *C. maconochiei* ssp. *maconochiei.* Genetic material of each species was collected from 26 wild populations (237 samples) throughout their geographic range. Ten morphologically intermediate individuals from a single putative hybrid population (*C. armstrongii* x *maconochiei*) were also collected. We used next-generation sequencing in the form of restriction site associated DNA sequencing, following the established ezRAD method, modified for cycads. De-novo assembly using ipyrad generated 868 to 3043 informative genome-wide single nucleotide polymorphic markers with 50% minimum samples per locus. Population genetic statistics showed exceptionally low levels of heterozygosity in *Cycas armstrongii* (H_E = 0.009 to 0.033, H_O = 0.005 to 0.057) and in *C. maconochiei* ssp. *maconochiei* (H_O = ≤ 0.049), low inbreeding was found in *C. armstrongii* (F_{IS} = 0.007 to 0.040) and in *C. maconochiei* ssp. *maconochiei* populations show low to moderate levels of inbreeding (F_{IS} = 0.037 to 0.243). AMOVA revealed that both taxa had very low levels of differentiation between populations (6 - 14%) and only 6% differentiation between taxa. The distance-based pairwise F_{ST} matrix recovered very low to high levels of genetic distance between populations of *C. armstrongii* (0 to 23.4%) and of *C*. *maconochiei* ssp. *maconochiei* (4.1% to 9.8%), implying gene flow between populations of both taxa. Discriminant analysis of principal components indicated a lack of spatial differentiation between individuals and taxa. Despite the morphological and geographic differences between the taxa, there was little genetic differentiation between *C. armstrongii* and *C. maconochiei* ssp. *maconochiei*, thereby casting doubt on the taxonomic validity of separating *C. armstrongii* from *C. maconochiei* ssp. *maconochiei*. Furthermore, the lack of interspecific genetic variation provides no evidence of interspecific hybridisation.

4.1 Introduction

Australia is one of the world's key areas of floristic diversity (Sloan et al. 2014), with approximately 92% of vascular plant species being endemic (Chapman 2009). Worldwide, an increasing number of plant species are under threat of extinction from habitat loss and/or fragmentation (Sax et al. 2002). The effects of recent, human-mediated climate change has also caused significant shifts in both the distribution and abundances of many species, including those in Australia (Spooner et al. 2018; Thomas et al. 2004). Thus, the development of a range of effective conservation interventions are critical for the long-term management of threatened and at-risk species (Scheele et al. 2018; Hefley et al. 2016). In order to accomplish appropriate conservation management plans for plant species, it is imperative to understand aspects of genetic diversity and population demographics (Harrisson et al. 2014; Hedrick & Miller 1992). Population genetics has been used to identify and prioritise populations for conservation that have high levels of genetic diversity (Hou et al. 2018; Rodríguez-Rodríguez et al. 2018; Drury et al. 2017). By conserving populations with the highest genetic diversity, species will have the maximum potential for genetic resilience through the adaptability of their populations (Kramer & Havens 2009; Frankham et al. 2004).

Cycads are of global conservation significance, and their populations are threatened by over-collection, habitat loss, and population fragmentation (Donaldson 2003). The situation is further exacerbated by the slow rate of reproduction of many cycads (Raimondo & Donaldson 2003). Population isolation is a factor that drives speciation, but in plants such as cycads it is not always advantageous (Rieseberg 2007). The effects of population isolation in cycads often results in genetic drift and bottlenecks (Zhan et al. 2011; González-Astorga, et al. 2008). The western Pacific species, *Cycas balansae* Warb. and *Cycas seemannii* A.Br., show high levels of inter-population genetic variation, but low levels of genetic variation within populations. These results imply that reduced gene flow has led to greater genetic differences between populations and inbreeding within the populations (Long-Qian et al. 2004; Keppel et al. 2002). Genetic isolation and inbreeding can also be associated with small population sizes which can lead to a reduction in the genetic fitness of a population (Ellstrand 1993). However, genetic diversity in cycads is not always correlated with population size as populations can have a higher level of historic gene flow, but with little variation between geographic regions (James et al. 2018, Cibrián-Jaramillo et al. 2010; González-Astorga, et al. 2008).

Geographic proximity of populations of different species can lead to interspecific hybridisation and outbreeding depression (Alix et al. 2017). Interspecific hybrids among cycads are often morphologically and geographically intermediate between the putative parent species (Chiang et al. 2013). Hybridisation occurs more readily when there are small genetic differences between species (Neri et al. 2018).
The effects of introgression are important in the study of plants such as cycads as it causes genetic mixing which is especially problematic in rare species with small effective population sizes, as it dilutes the genetic identity of a species (Ellstrand et al. 1990). However, interspecific hybridisation can also have its benefits as it can introduce new genetic material into populations by incorporating advantageous alleles, for example by introducing resistance to pathogens (Siemens 2012).

The use of molecular genetics allows the identification of interspecific hybrids, and can quantify the degree to which the putative parent species are genetically related (Schmutzer et al. 2015).The effects of hybridisation on the genetic diversity of wild populations of cycads has yet to be explored (van der Bank et al. 2001). It is, therefore, currently unknown if any formally recognised hybrids between cycad species are biologically valid, or if they are merely intermediate morphotypes of a single morphologically variable species (Chamberlain 1926).

Cycas L. (Cycadaceae) is the largest genus of Cycadales consisting of 117 extant species. The genus has a pantropical distribution, and ranges from coastal Africa to Australia. Australia has around 33% of the world's cycad species with all of species being endemic to Australia (Calonje et al. 2019). *Cycas* species are found in Western Australia (3 species), Northern Territory (16 species) and Queensland (19 species). Despite the species richness of *Cycas* in Australia, little is known of their population genetics with only *Cycas megacarpa* K.D.Hill being studied in any detail (James et al. 2018).

Many population genetics studies of plants have been focused on species thought to be at the greatest risk of extinction (Heywood & Iriondo 2003; Young et al. 1996). This is also the case for cycads, where populations have already experienced bottlenecks and reduced genetic diversity (Cabrera-Toledo et al. 2012; Da Silva et al. 2012; Octavio-Aguilar et al. 2009; González-Astorga, Vovides, et al. 2008; Shuguang et al. 2006). To understand genetic relationships of plants at risk, it is also important to focus on species that are not as threatened as this will better help to contextualize genetic diversity and distinctness of species as well as uncovering instances of introgression (Rodríguez-Quilón et al. 2016).

Cycas populations in the Northern Territory are relatively easily accessible, and many species grow abundantly throughout their distribution (Dixon 2004). *Cycas armstrongii* Miq. and *C. maconochiei* ssp. *maconochiei* K.D.Hill (Figure 10, 11 and 12,) are two of the most common species in the Northern Territory. Their populations are often relatively large (Watkinson & Powell 1997), but are potentially under threat by frequent anthropogenic burning, land clearing for agriculture, and competition from invasive species. Populations of both taxa are considered to be healthy, and show evidence of continuous recruitment (Liddle 2009). Under the Interim Biogeographic Regionalisation for Australia (IBRA7) from Australia's Strategy for the National Reserve System (Australian Government 2000), both taxa occur in biogeographic subregions Darwin Coastal (DAC01), Pine Creek (PCK01), Daly Basin (DAB01), Cobourg (TIW01) and Tiwi (TIW02).

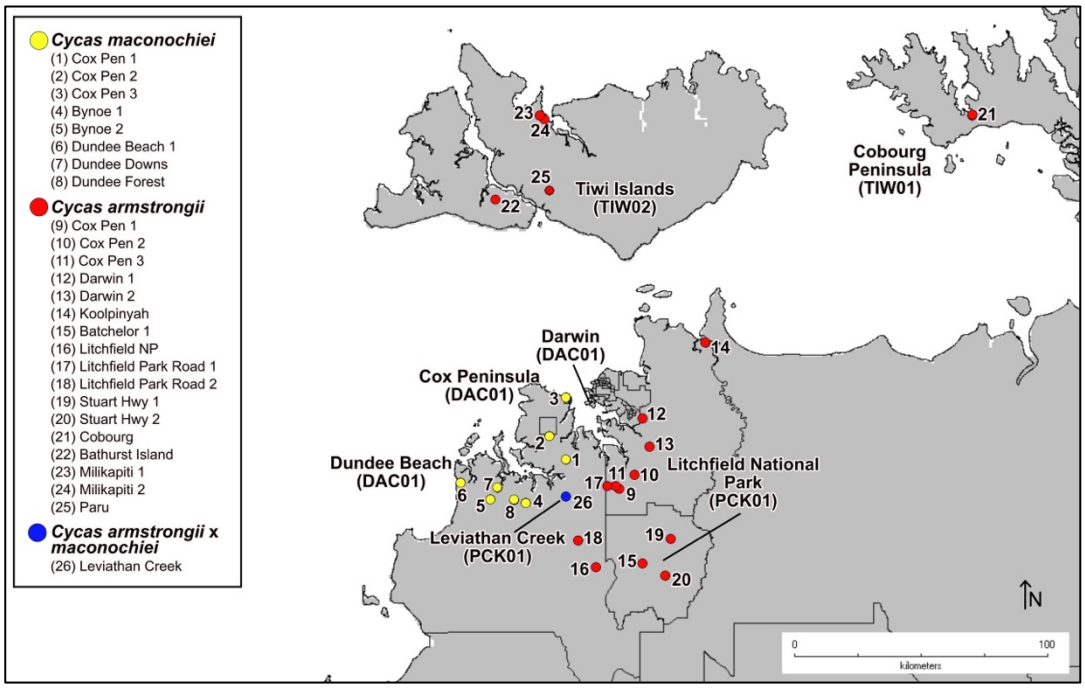

Figure 10 Map of samples collected for *Cycas armstrongii* and *C. maconochiei* ssp. *maconochiei*. Map of Northern Territory, Australia. Showing range of sample collection from *C. armstrongii* (Tiwi Islands, Garig Gunak Barlu National Park, Darwin Coastal Region and Litchfield National Park (Pine Creek)), *C. maconochiei* ssp. *maconochiei* (Cox Peninsula to Dundee Beach, representing most populations expect most southern), representing most populations expect most southern and a hybrid population *C. armstrongii* x *maconochiei*.

Figure 11 *Cycas armstrongii* populations growing in the Darwin Coastal and Pine Creek region, Northern Territory Australia. (A) Small population growing along the Stuart Highway, in the Darwin Region, Northern Territory Australia. (B) Medium sized population consisting of mostly mature specimens growing near the Litchfield National Park.

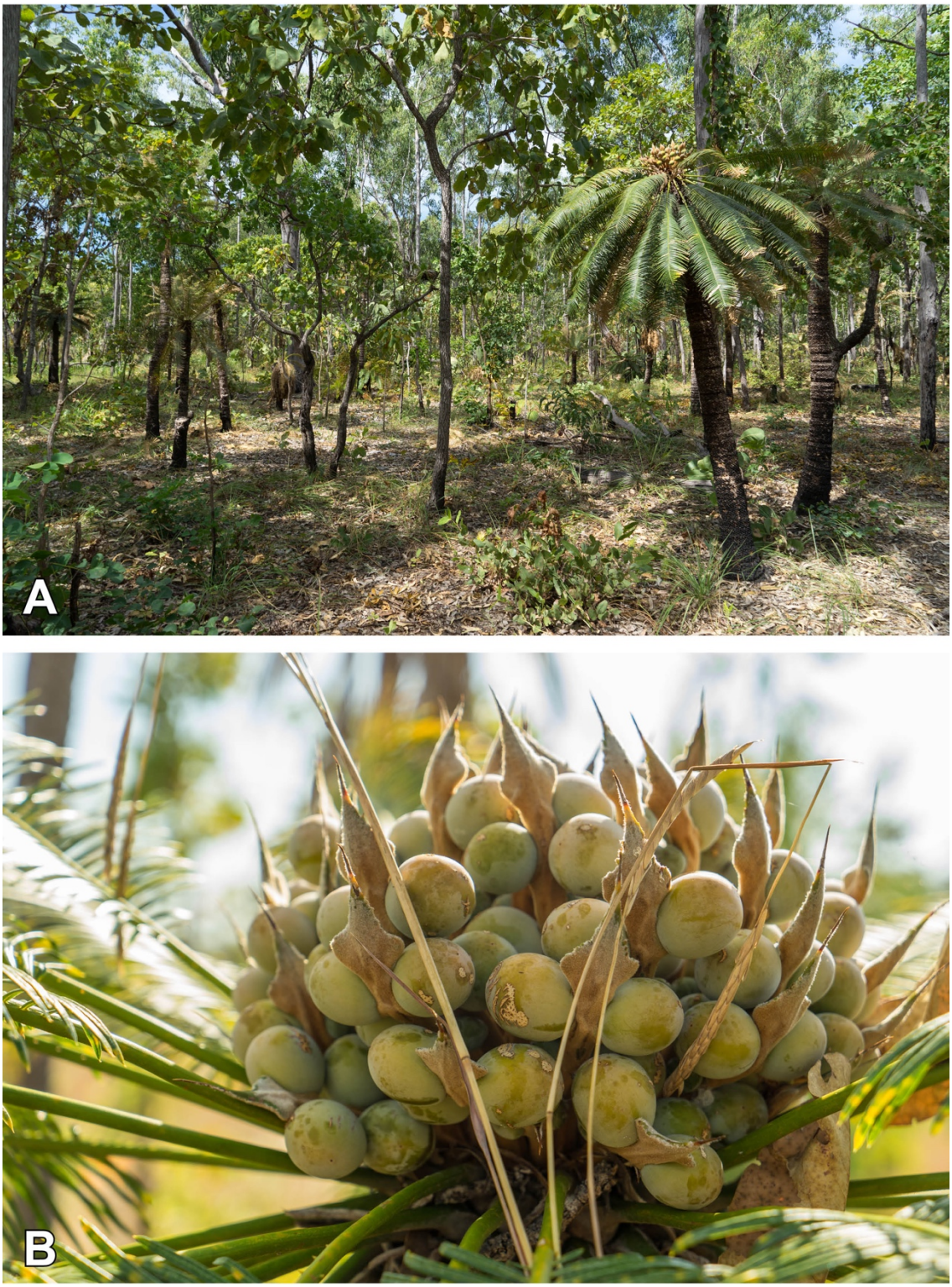

Figure 12 *Cycas maconochiei* ssp. *maconochiei* growing in the Cox Peninsula Northern Territory Australia. (A) Large population of *C. maconochiei* ssp. *maconochiei* (B) Large mature female specimen bearing seeds on megasporophylls.

Cycas armstrongii Miq. (Figure 10 and 11) is considered to be one of the most abundant cycad species in the Darwin Coastal and Pine Creek sub-regions, growing in dense populations in open savannah or semi-closed *Eucalyptus* forests, with scattered populations on the Tiwi Islands (TIW01) and Cobourg Peninsula (TIW02) subregions (Liddle 2009; Watkinson & Powell 1997; Hill 1996). *Cycas maconochiei* ssp. *maconochiei* (Figure 12), like *C. armstrongii*, occurs in large and disjunct populations north-west of Darwin and in the greater Darwin Coastal Region. In addition, fragmented populations matching the morphological characteristics of both taxa occur on the Tiwi Islands (Liddle 2009). *Cycas armstrongii* and *C. maconochiei* share many morphological characteristics although *C. maconochiei* may be differentiated from *C. armstrongii* based on the former's darker green leaves, distinctive recurved leaflet margins, and the lack of a terminal point on the cataphylls (Hill 1996).

Cycas armstrongii and *Cycas maconochiei* subsp. *maconochiei* are thought to be part of a species complex (*sensu* Hill 1996) comprising: *Cycas armstrongii* Miq., *C. canalis* K.D.Hill, *C. conferta* Chirgwin ex Chirgwin & Wigston, *Cycas maconochiei* subsp. *maconochiei* Chirgwin & K.D.Hill and *Cycas maconochiei* ssp. *viridis* K.D.Hill. Species within the complex are considered to hybridise freely (Hill 1996). Only *C. armstrongii* and *C. maconochiei* ssp. *maconochiei* have a continuous distribution through the Darwin Coastal and Pine Creek sub-regions where a single population of the putative interspecific hybrid *C. armstrongii* x *maconochiei* also occurs (Figure 10) (Liddle 2009; Hill 1996). The continuous distribution of *C. armstrongii, C. maconochiei* ssp. *maconochiei*, and the sympatric putative hybrid, presents an opportunity to study the effects of interspecific hybridisation between two closely related species (Hill et al. 2014).

Cycas armstrongii is known to be pollinated by two species of beetle in the family Tenebrionidae and to a lesser extent by wind pollination (Ornduff 1992;Hall & Walter 2018). *Cycas maconochiei* is thought to be insect pollinated (Kono & Tobe 2007), but the pollinator has yet to be identified (Liddle 2009). Based on their parapatric distribution and the putative presence of hybrids, it is possible that *C. maconochiei* and *C. armstrongii* share a pollinator. Seed dispersal in *C. armstrongii* is considered to be less than 1 m from the female plant (Watkinson & Powell 1997). Whilst the contemporary dispersers of Australian cycad seeds are known to be small rodents and marsupials, it is thought that the primary seed dispersers were most likely unknown members of the now extinct Australian Pleistocene megafauna (Hall & Walter 2013).

Here we aim to better understand the genetic diversity of *Cycas armstrongii* and *C. maconochiei* ssp. *maconochiei* by analysing genetic material collected from populations throughout the Northern Territory. We use next generation sequencing (NGS) in the form of restriction-site associated DNA sequencing (RADseq), which allowed us to find hundreds to thousands of markers throughout the genome, which are subjected to the full evolutionary history of the species (Andrews et al. 2016). The data will be used to determine the genetic history and diversity of natural populations of *C. armstrongii* and *C. maconochiei*, and to inform the conservation management of the taxa, to aid in species conservation. The results will provide new insights into the geographic and genetic relationship between the two species, and will allow us to examine the occurrence of the interspecific hybridisation proposed by Hill (1996).

4.2 Materials and methods

Sample preparation

Sampling strategy. Populations were selected based on previously published records by Liddle (2009), Dixon (2004) Hill (1994), herbarium specimens held by The New South Wales National Herbarium (NSW), Northern Territory Herbarium (DNA), and The Australasian Virtual Herbarium (AVH) (https://avh.chah.org.au, accessed 12th January 2015). Silica gel dried leaflets were collected from up to 40 individuals (Average =12) from 21 wild populations (Table 11, Figure 10) for two species: *Cycas armstrongii* and *Cycas maconochiei* ssp. *maconochiei,* and a population of suspected hybrid individuals (= *Cycas armstrongii* x *maconochiei*). The sampled populations represent the following biogeographic subregions as defined by the Interim Biogeographic Regionalisation for Australia (IBRA7; Australian Government 2000): Darwin Costal (DAC01), Pine Creek (PCK01) and Tiwi (TIW01) and Cobourg (TIW02). A population of C. *maconochiei* from the Daly Basin (IBRA subregion DOB01) was not sampled.

Table 11 Summary of samples collected from natural populations of *Cycas armstrongii, C. maconochiei ssp. maconochiei and C. armstrongii* x *maconochiei* in Northern Territory, Australia. Table defines the sampling stratification used for discriminate analysis and AMOVA. Species = The taxonomic identification assigned during collection; Region and Sub-Region are defined as the regions and sub-regions for the Interim Biogeographic Regionalisation for Australia (IBRA7) from Australia's Strategy for the National Reserve System (http://www.environment.gov.au/land/nrs/science/ibra). Locality = name of providence by which groups of populations are referred to in the text; Population = individual populations where collections were carried out and No. samples = the total number of samples collected from each population.

Species	IBRA Region	IBRA Subregion	Locality	Population	No. Samples
C. maconochiei	Darwin Coastal (DAC)	Darwin Coastal (DAC01)	Cox Peninsula	Cox Pen 1	7
				Cox Pen 2	10
				Cox Pen 3	10
			Dundee	Bynoe 1	10
				Bynoe 2	10
				Dundee Beach 1	10
				Dundee Downs	10
				Dundee Forest	10
C. arm X mac	DAC	DAC01	Leviathan Creek	Leviathan Creek/ Bynoe	10
C. armstrongii	Darwin Coastal (DAC)	Darwin Coastal (DAC01)	CA Darwin and Cox	Cox Pen 1	10
				Cox Pen 2	9
				Cox Pen 3	10
				Darwin 1	10
				Darwin 2	10
				Koolpinyah	9
	Pine Creek (PCK)	Pine Creek (PCK01)	Pine Creek	Batchelor 1	10
				Litchfield NP	9
				Litchfield Park Road 1	10
				Litchfield Park Road 2	9
				Stuart Hwy 1	10
				Stuart Hwy 2	9
	Tiwi Cobourg (TIW)	Cobourg (TIW02)	Cobourg	Cobourg	10
		Tiwi (TIW01)	Tiwi	Bathurst Island	2
				Milikapiti 1	10
				Milikapiti 2	10
				Paru	8

DNA extraction and quantification. Approximately 0.05 g of silica-dried leaflets were ground to a fine powder using a TissueLyser (Qiagen, Hilden, Germany). High molecular weight genomic DNA was extracted using a DNeasy Plant DNA Extraction Mini Kit (Qiagen, Hilden, Germany) following a modified protocol (see Chapter 2). Genomic DNA was inspected using a 2% agarose gel to check for the presence of DNA and impurities. DNA extractions were quantified using an Invitrogen Qubit (3.0 BR DNA assay; Invitrogen, Life Technologies, Carlsbad, CA, USA) fluorometer with a target concentration of 17 µg/mL. Any samples that yielded

less than this amount was either re-extracted or concentrated using a 1:1 ratio of Agencourt AMPure XP sample purification beads (Beckman Coulter, Inc.) by combining multiple extractions.

DNA normalization and restriction digest reaction. Genomic DNA was normalised to a concentration of 500 ng in 42 µL total volume (0.01 µg/mL), then 5 µL of NEB 10x CutSmart buffer (New England Biolabs, Ipswich, MA) and 1 µL of Bovine Serum Albumin (BSA) was added to each well. Samples were held at 4°C for a minimum of five hours before adding restriction enzymes (the five hours helps the cutting action of the restriction enzymes). Double digest reactions were carried out using 1 µL each of the restriction enzymes EcoR1-HF and Mse1. Reactions were placed into a thermocycler for three hours at 37°C with a final 20-minute enzyme deactivation step at 65°C. Samples were checked on 2% agarose gel for quality of digestion. The double digest reactions were cleaned using 1.8:1 ratio of AMPure XP beads to sample (90 µL of AMPure XP beads to 50 µL of digested DNA) and quantified using a Qubit high sensitivity kit.

Library preparation. Libraries were prepared using an Illumina TruSeq nano high throughput dual index library preparation kit (Illumina Inc., CA, USA). We followed a modified version ezRAD v3 (Toonen et al. 2013) protocol using half of the recommended volumes (Clugston et al. 2019). During the final step of library preparation, we modified the ezRAD v3 protocol for the final bead clean using a 0.8x ratio of AMPure XP beads to remove adapter dimers, present in the samples after ligation. Final Illumina libraries were validated using a LabChip, cleaned using a 0.9:1 ratio of AMPure XP beads and quantified using a QuBit high sensitivity kit. Final libraries were then normalized to 10 nM concentration and pooled for sequencing.

Sequencing. We aimed to capture a minimum of 1 GB of sequence data per sample (in three sequencing runs of 238 libraries), to ensure adequate coverage of the large genomes, and to help account for overrepresentation of the plastid genome, by capturing as much of the nuclear genome as possible. Genomic sequencing was carried out using an Illumina NextSeq 500 150bp paired-end high throughput (HT) on a single flow cell (x2 runs) and an Illumina HiSeq 400 150bp paired-end single lane. Each sequencing run was spiked with 10% PhiX sequencing control V3 to account for low diversity after using enzymatic digestion in the ezRAD protocol, modified by Clugston et al. (2019).

Bioinformatics

Quality control and filtering of sequence reads. The Illumina NextSeq 500 sequencer generated eight fastq files (four forward and four reverse) which were combined into a single file (one forward and one reserve) using a Unix bash script for compatibility with downstream processing. This was not the case for the Illumina HiSeq 4000 sequencer as only single forward and single reverse reads were generated. Using FastQC 0.11.4 (Andrews et al. 2014), reads were filtered to detect falloff in read quality and then trimmed using Trimmomatic 0.36 (Bolger et al. 2014). To remove Illumina adapter sequences, we removed the first 6 base pairs of reads (cut sites) due to quality drop-off and cropped reads to 120 bp in length (due to lower quality after 120 bp). A sliding window was used to delete bases with a PhredQ score less than 20, and all reads less than 50 bp were discarded.

Assembly of RADseq data. De novo assembly of the paired-end reads was performed using ipyrad 0.7.24 (Eaton and Overcast, in prep) on a high-performance online instance with Amazon Web Service (AWS) though the California Academy of Sciences (California, USA). All parameters were set to default in ipyrad except for data type, which was set to 'pairgbs' (most closely matches ezRAD). Bases with a 'PhredQ score' of less than 30 were converted to 'N', and reads with 15 or more uncalled bases were discarded. Reads were further filtered for adapter sequences and trimmed. Reads were discarded if they were less than 40 bp after trimming. The 'maximum number of uncalled bases in consensus sequences' was set to 10 in both forward and reserve reads. The setting for 'maximum shared heterozygotes per locus was left as 0.5' (default) to reduce the effects of paralogs. The 'maximum heterozygotes in consensus sequences' was set to 8 for both forward and reverse sequences and the 'minimum number of samples per locus' was set to 119, so each SNP would be present across a minimum of 50% of samples. The high 'minimum samples per locus' helps to reduce the amount of missing data and helps prevent anomalies which may occur in population level analysis and to ensure effective population genotyping (Shafer et al. 2016). These settings were used for a combined assembly including *C. armstrongii*, *C*. *maconochiei* and *C. armstrongii* x *maconochiei* as well as two separate assemblies for *C. armstrongii* and *C*. *maconochiei*.

Population genetic analysis. For comparative statistics, samples were assigned to a stratification group (Table 11) depending on the aim of the analysis, namely: Taxon, IBRA region, IBRA subregion, locality and population. Descriptive statistics - Number of individuals in each population (N), effective number of alleles (N_A), the effective number of alleles per locus (N_E), heterozygosity observed (H_O), heterozygosity expected (H_e), unbiased expected heterozygosity (uHe), and the fixation index (F_{IS}) were generated using GenALEx 6.5 (Peakall & Smouse 2012). To test genetic variation among populations and among individuals within a population or region we used an analysis of molecular variance (AMOVA); genetic distance based pairwise F_{ST} was calculated in GenAlEx using 9999 permutations of the dataset, with the 'Codom-Allelic' option selected. However, for *Cycas armstrongii* pairwise F_{ST} was determined using diveRsity 1.9.90 (Keenan et al. 2013) in R 3.8.10 (R core development team), using the pairwise bootstrap function for 100 bootstraps.

To understand population structure in *Cycas armstrongii* and *Cycas maconochiei* ssp. *maconochiei* including the hybrid population, STRUCTURE v.2.3.4 (Pritchard et al., 2000) was used to determine admixture among populations and between species for *Cycas armstrongii* in 150 individuals among 17 populations, for *Cycas maconochiei* ssp. *maconochiei* 76 individuals among 8 populations. Finally, a combined data containing 236 individuals among all taxa including the hybrid populations were tested. STRUCTURE uses Bayesian algorithms to infer the optimum number of distinct genetic groups *K* (clusters) by minimizing deviations from Hardy–Weinberg and linkage equilibrium within each cluster. The analyses were carried out for *K* = 1–5 using 100,000 MCMC iterations after a burnin of 20,000 steps and were repeated 10 times for each *K*, with the 'Separate Alpha for each Population' option selected.

To visualise the genetic relationships between populations, a discriminant analysis of principal component (DAPC) was carried out using Adegenet 2.1.0 in R (Jombart & Ahmed 2011). The optimal number of clusters in the data and number of PCAs to be retained for discriminate analysis was determined using the "find.clusters" command in combination with the optimal *A*-score. The genetic assignment of populations was tested to predict if individuals belonged to the inferred groups in the set stratifications. These results were graphically displayed showing prior group membership for all individuals with group probabilities. The results were visualised with a DAPC scatter plot in order to understand relationships and connectivity of populations.

4.3 Results

Sequencing and de-novo assembly. After filtering raw sequence data, the number of reads for each sample that passed all filters ranged from 1,055,323 to 16,253,580. De-novo assembly of the reads generated sequence clusters from 840,412 to 6,362,790 with 12,241 to 4,677,090 clusters being high-depth containing six or more reads. This resulted in a sequence read consensus of 8,383 to 4,035,099. The final output resulted in 1151 SNPs recovered from 108 unique loci for both species and the putative hybrid population combined. *Cycas armstrongii* had 868 SNPs recovered from 91 unique loci, and 3043 SNPs were recovered for *C. maconochiei* from 315 unique loci.

Population genetic statistics. The results from population genetic analysis were used to determine the genetic diversity of the taxa and their populations (Table 11). *Cycas armstrongii* (Table 12) showed an expected heterozygosity (H_E) that ranged from 0.009 (± 0.002) for the Bathurst Island population to 0.033 (± 0.003) for the Litchfield Park Road 1 population, with a mean of 0.020 (± 0.001). The observed heterozygosity (H_O) ranged from 0.005 (± 0.002) in Bathurst island population to 0.057 (± 0.005) in Milikapiti 1 population, with a mean of 0.036 (± 0.001). The mean H_E (0.020 ± 0.001) and H_O (0.036 ± 0.001) indicates that the populations were within the Hardy-Weinberg equilibrium (HWE). The inbreeding coefficient (F_{IS}) ranged from 0.007 (-0.619 to -0.669 95%CI) in the Bathurst island to 0.040 (0.081 to 0.051 95%CI) in Milikapiti 1 population with a mean of 0.024 (0.021 to 0.014 95%CI). These results indicate that individuals in the Bathurst island population are less related than would be expected under HWE and Cox Peninsula population is more related than would be expected under the HWE. The populations of *C. armstrongii* that were within the HWE (Table 12) were Cox Pen 3 at 0.025 (0.021 to -0.004 95%CI), Cox Pen 1 at 0.017, Litchfield Park Road 2 at 0.028 (0.005 to -0.021 95%CI) and Lichfield NP 1 at 0.018 (0.008 to -0.021 95%CI). The mean F_{IS} for *Cycas armstrongii* was 0.024 (0.021 to 0.014 95%CI) indicating that although some populations are within the HWE the majority deviate significantly.

Table 12 Summary of population genetic statistics for populations of *Cycas armstrongii*. Mean and standard error (SE) of frequency-based population genetic statistics for populations of *C. armstrongii* within the defined IBRA7 regions, generated using GenALEx 6.5. Effective number of individuals in each population (N), effective number of alleles (N_A), the effective number of alleles (N_E), heterozygosity observed (H_O), heterozygosity expected (H_E), unbiased expected heterozygosity (uHe), and Fixation index (F_{IS}). Calculation of the fixation 95% confidence interval (mean±1.96*SE) lower (-95%CI) and upper limits (+95%CI).

Population			N	N_A	N_E	H_O	H_E	uHe	F_{IS}	-95%CI	+95%CI
Tiwi Cobourg (TIW)	Milikapiti 1	Mean	5.471	1.114	1.017	0.057	0.028	0.035	0.040	0.081	0.051
		SE	0.081	0.014	0.009	0.005	0.003	0.003	0.004	0.015	
	Milikapiti 2	Mean	7.199	1.138	1.006	0.051	0.026	0.029	0.031	0.088	0.059
		SE	0.081	0.015	0.007	0.004	0.003	0.002	0.003	0.015	
	Paru	Mean	5.302	1.103	1.007	0.047	0.030	0.028	0.031	-0.085	-0.101
		SE	0.071	0.013	0.007	0.004	0.003	0.003	0.003	0.008	
	Bathurst island	Mean	0.836	0.643	0.642	0.005	0.006	0.004	0.007	-0.619	-0.669
		SE	0.025	0.017	0.017	0.002	0.003	0.001	0.003	0.026	
	Cobourg	Mean	3.907	1.046	0.996	0.037	0.024	0.023	0.027	-0.073	-0.099
		SE	0.052	0.012	0.008	0.004	0.003	0.003	0.003	0.013	
Darwin Coastal (DAC)	Koolpinyah	Mean	6.757	1.138	1.009	0.051	0.029	0.029	0.031	-0.036	-0.054
		SE	0.074	0.014	0.007	0.004	0.003	0.002	0.003	0.009	
	Darwin 1	Mean	5.242	1.070	1.012	0.038	0.014	0.024	0.028	0.198	0.165
		SE	0.076	0.012	0.008	0.004	0.002	0.003	0.004	0.017	
	Darwin 2	Mean	5.977	1.069	1.006	0.037	0.014	0.022	0.025	0.230	0.198
		SE	0.078	0.012	0.007	0.004	0.002	0.003	0.003	0.017	
	CA Cox Pen 3	Mean	5.850	1.048	0.982	0.028	0.015	0.016	0.018	0.021	-0.004
		SE	0.072	0.012	0.007	0.003	0.002	0.002	0.002	0.012	
	CA Cox Pen 2	Mean	4.656	1.033	0.994	0.026	0.014	0.016	0.019	0.085	0.053
		SE	0.071	0.010	0.007	0.004	0.002	0.002	0.003	0.016	
	CA Cox Pen 1	Mean	5.094	1.037	0.996	0.025	0.014	0.015	0.017	-0.001	-0.025
		SE	0.074	0.010	0.006	0.003	0.002	0.002	0.003	0.012	
	Litchfield Park Road 1	Mean	6.783	1.138	1.007	0.055	0.033	0.031	0.035	-0.050	-0.068
		SE	0.086	0.015	0.008	0.004	0.003	0.003	0.003	0.009	
Pine Creek (PCK)	Litchfield Park Road 2	Mean	6.267	1.093	1.002	0.042	0.024	0.024	0.028	0.005	-0.021
		SE	0.081	0.013	0.007	0.004	0.003	0.002	0.003	0.013	
	Stuart Hwy 1	Mean	6.262	1.046	1.003	0.025	0.013	0.015	0.018	0.054	0.022
		SE	0.087	0.010	0.006	0.003	0.002	0.002	0.003	0.016	
	Batchelor 1	Mean	5.492	1.035	0.989	0.027	0.016	0.016	0.019	-0.015	-0.044
		SE	0.087	0.011	0.007	0.004	0.003	0.002	0.003	0.015	
	Litchfield NP 1	Mean	5.301	1.033	0.986	0.025	0.015	0.015	0.018	0.008	-0.021
		SE	0.071	0.011	0.007	0.003	0.002	0.002	0.003	0.015	
	Stuart Hwy 2	Mean	4.820	1.045	0.994	0.028	0.019	0.017	0.020	-0.081	-0.104
		SE	0.080	0.011	0.007	0.004	0.003	0.002	0.003	0.012	
Total		Mean	5.366	1.049	0.979	0.036	0.020	0.021	0.024	0.021	0.014
		SE	0.021	0.003	0.002	0.001	0.001	0.001	0.001	0.003	

Table 13 Summary of population genetic statistics for populations of *Cycas maconochiei* ssp. *maconochiei* and *C. armstrongii* x *maconochiei*. Mean and standard error (SE) of frequency-based population genetic statistics for populations of *C. maconochiei* ssp. *maconochiei* and *C. armstrongii* x *maconochiei*, generated using GenALEx 6.5 in the Darwin Costal Region (DAC01). Number of individuals in each population (N), effective number of alleles (N_A), the effective number of alleles (N_E), heterozygosity observed (H_O), heterozygosity expected (H_E), unbiased expected heterozygosity (uHe), and Fixation index (F_{IS}). Calculation of the fixation 95% confidence interval (mean±1.96*SE) lower (-95%CI) and upper limits (+95%CI). *Cycas armstrongii* x *maconochiei* = Leviathan Creek. Calculation of the fixation 95% confidence interval (mean±1.96*SE) lower (-95%CI) and upper limits (+95%CI).

Population		N	N_A	N_E	H_O	H_E	uHe	F_{IS}	-95%CI	+95%CI
Cox Pen 1	Mean	3.882	1.110	1.029	0.043	0.052	0.063	0.133	0.112	0.154
	SE	0.032	0.008	0.006	0.003	0.002	0.003	0.011		
Cox Pen 2	Mean	5.269	1.157	1.034	0.042	0.055	0.063	0.154	0.136	0.173
	SE	0.045	0.009	0.006	0.002	0.002	0.003	0.009		
Cox Pen 3	Mean	4.823	1.116	1.023	0.034	0.043	0.050	0.120	0.103	0.138
	SE	0.040	0.008	0.005	0.002	0.002	0.002	0.009		
Dundee Beach 1	Mean	5.123	1.179	1.059	0.041	0.059	0.068	0.213	0.193	0.138
	SE	0.044	0.008	0.005	0.002	0.002	0.003	0.010		
Dundee Downs	Mean	4.816	1.166	1.039	0.040	0.056	0.064	0.175	0.157	0.193
	SE	0.043	0.009	0.006	0.002	0.002	0.003	0.009		
Bynoe 2	Mean	5.065	1.168	1.044	0.038	0.059	0.067	0.243	0.225	0.262
	SE	0.042	0.009	0.006	0.002	0.002	0.003	0.010		
Bynoe 1	Mean	5.925	1.179	1.047	0.049	0.053	0.060	0.037	0.021	0.053
	SE	0.046	0.008	0.005	0.002	0.002	0.003	0.008		
Dundee Forest	Mean	5.487	1.141	1.029	0.033	0.045	0.051	0.142	0.126	0.159
	SE	0.043	0.008	0.005	0.002	0.002	0.002	0.008		
Total	Mean	5.049	1.152	1.038	0.040	0.053	0.061	0.154	0.148	0.161
	SE	0.015	0.003	0.002	0.001	0.001	0.001	0.003		
Leviathan Creek	Mean	6.785	1.080	1.009	0.018	0.020	0.022	0.001	-0.018	0.021
	SE	0.062	0.010	0.005	0.002	0.002	0.002	0.010		

Cycas maconochiei (Table 13) had a H_E that ranged from 0.043 (± 0.002) in the Cox peninsula population 3 to 0.059 (± 0.002) in the Dundee Beach 1 population, with a mean of 0.053 (± 0.001). The H_O ranged from 0.033 (± 0.002) in the Dundee Forest population to 0.049 (± 0.002) in the Bynoe 1 population, with a mean of 0.040 (± 0.001). The mean H_E (0.053 ± 0.001) and H_o (0.040 ± 0.001) indicate that all populations of *C. maconochiei* fall outside of the HWE. The inbreeding coefficient (F_{IS}) ranged from 0.037 (0.021 - 0.053 95%CI) in the Bynoe 1 population to 0.243 (0.225 to 0.262 95%CI) in the Bynoe 2 population. The mean F_{IS} value was 0.154 (0.148 to 0.161 95%CI) indicating that individuals within population of *C. maconochiei* are more related than would be expected under a model of random mating. *Cycas armstrongii* x *maconochiei* (Leviathan Creek) (Table 13) had an expected heterozygosity (H_E) of 0.020 (± 0.002) and a H_O of 0.018 (± 0.002). The F_{IS} was 0.001 (-0.018 to 0.021 95%CI) indicating that random mating is occurring in the population.

Genetic structure and population differentiation. Analysis of Molecular Variance (AMOVA, Table 14) showed that there was 6% genetic variation between *C. armstrongii* and *C. maconochiei* There was 14% variation among all populations, with the majority of variation (80%) contained within populations. *Cycas armstrongii* had 3% genetic variation among IBRA subregions, 13% genetic variation between its populations, and with the majority of the genetic variation being within populations (84%). *Cycas maconochiei* had no genetic variation between regions, 6% genetic variation between populations, with most genetic variation being within populations (94%).

Table 14 Analysis of molecular variance for *Cycas armstrongii* and *C. maconochiei* ssp. *maconochiei*. Results for Analysis of Molecular Variance (AMOVA) for populations of *C. armstrongii* and *C. maconochiei* ssp. *maconochiei*. Df = Degree of Freedom, among populations = genetic variation among the populations within the localities and within populations = degree of genetic variation within the populations in a region.

Populations	Df	Sum of squares	Estimated variance	Percentage of variation (%)
Cycas armstrongii* and *C. maconochiei				
Among taxa	1	4128.025	15.684	6
Among populations	23	21285.704	39.176	14
Within populations	427	93882.006	219.864	80
Cycas armstrongii				
Among Regions	2	2476.494	6.664	3
Among population	14	8181.391	24.086	13
Within populations	283	45853.730	162.027	84
Cycas maconochiei				
Among Regions	1	1480.397	1.161	0
Among population	6	8548.108	41.676	6
Within populations	144	90820.514	673.525	94

The distance based (AMOVA) pairwise fixation index (F_{ST}) were calculated for *C. armstrongii* and *C. maconochiei* ssp. *maconochiei* separately. The genetic distance between populations of *C. armstrongii* (Table 15) ranged from 0 (0%) in multiple populations especially populations within the Tiwi Corbourg regions to 0.234 (23.4%) between the Cobourg and Cox Pen 3 population. The pairwise fixation index (F_{ST}) showed genetic distance between IBRA7 sub regions (Table 16) ranged from 0.009 (0.099%) between Darwin Costal and Pine Creek to 0.127 (12.7%) between Cobourg to Pine Creek. These results indicated that populations occurring within Tiwi (TIW) showed evidence of a greater genetic distance from the Darwin Costal and Pine Creek regions (DAC/PCK), which agree somewhat with geography (Figure 1). The genetic distance between populations of *C. maconochiei* ssp. *maconochiei* (Table 17) ranged from 0.041 (4.1%) between the Dundee Forest and the Cox Pen 1 populations to 0.098 (9.8%) between Dundee Beach 1 and Bynoe 1 populations. Overall, no populations of *C. maconochiei* ssp. *maconochiei* had greater than 10% genetic distance.

Table 15 Pairwise FST matrix of *Cycas armstrongii* populations. Genetic distance based pairwise F_{ST} matrix from AMOVA analysis generated using diveRsity 1.9.90 for *Cycas armstrongii* populations within the defined Interim Biogeographic Regionalisation for Australia (IBRA7). Values greater ≥ 0.20 are highlighted in **bold**.

Region	Tiwi Cobourg (TIW)					Darwin Costal (DAC)						Pine Creek (PCK)				
Populations	Milikapiti 1	Milikapiti 2	Paru	Bathurst island	Cobourg	Koolpinyah	Darwin 1	Darwin 2	CA Cox Pen 3	CA Cox Pen 2	CA Cox Pen 1	Litchfield Park Road 1	Litchfield Park Road 2	Stuart Hwy 1	Batchelor 1	Litchfield NP 1
Milikapiti 2	0.035															
Paru	0.037	-0.021														
Bathurst island	-0.119	-0.179	-0.098													
Cobourg	-0.027	0.087	0.080	-0.150												
Koolpinyah	0.055	0.071	-0.009	-0.151	0.095											
Darwin 1	0.130	0.131	0.107	0.013	0.143	-0.021										
Darwin 2	0.101	0.050	0.016	-0.123	0.153	-0.049	-0.042									
CA Cox Pen 3	0.143	0.085	0.071	0.047	**0.234**	0.074	0.123	0.076								
CA Cox Pen 2	0.142	0.001	0.026	-0.037	0.134	0.069	-0.004	-0.004	-0.035							
CA Cox Pen 1	0.129	0.085	0.094	0.076	0.202	0.046	0.129	0.057	0.049	0.017						
Litchfield Park Road 1	0.086	0.053	0.063	-0.072	0.121	0.011	0.025	0.001	0.022	0.016	0.001					
Litchfield Park Road 2	0.128	0.049	0.029	-0.116	0.064	-0.008	0.069	0.015	-0.016	-0.002	0.002	-0.020				
Stuart Hwy 1	0.167	0.095	0.107	0.080	**0.231**	0.031	0.102	0.019	-0.033	0.018	-0.120	-0.007	0.011			
Batchelor 1	0.142	0.048	0.040	0.021	0.165	0.019	-0.109	-0.115	-0.053	0.009	-0.061	-0.027	-0.042	-0.029		
Litchfield NP 1	0.167	0.071	0.099	0.041	**0.229**	0.056	0.095	0.066	0.086	0.045	0.125	-0.015	0.041	0.219	0.035	
Stuart Hwy 2	0.143	0.063	0.019	-0.040	0.210	0.067	0.024	-0.055	0.062	0.027	0.068	0.031	0.019	0.024	-0.038	0.102

Table 16 Pairwise FST matrix between *Cycas armstrongii* regions. Genetic distance based pairwise F_{ST} matrix from AMOVA analysis generated using GenALEx 6.5 to show genetic distance between sub-regions for the Interim Biogeographic Regionalisation for Australia (IBRA7) of *Cycas armstrongii*.

Regions	Tiwi (TIW02)	Cobourg (TIW01)	Darwin Costal (DAC01)
Cobourg (TIW01)	0.041		
Darwin Costal (DAC01)	0.078	0.082	
Pine Creek (PCK01)	0.120	0.127	0.009

Table 17 Pairwise FST matrix of *Cycas maconochiei* ssp. *maconochiei* populations. Genetic distance based pairwise F_{ST} matrix from AMOVA analysis generated using GenALEx 6.5 for *Cycas maconochiei* ssp. *maconochiei* populations in the Darwin Costal Region (DAC01).

	Cox Pen 1	Cox Pen 2	Cox Pen 3	Dundee Beach 1	Dundee Downs	Bynoe 2	Bynoe 1
Cox Pen 2	0.059						
Cox Pen 3	0.050	0.050					
Dundee Beach 1	0.068	0.082	0.057				
Dundee Downs	0.076	0.076	0.056	0.045			
Bynoe 2	0.063	0.066	0.049	0.063	0.046		
Bynoe 1	0.062	0.063	0.066	0.098	0.092	0.074	
Dundee Forest	0.041	0.052	0.052	0.060	0.068	0.059	0.052

Population structure analysis. Population structure results for *Cycas armstrongii* show no defined population structure between regions and populations (Figure 13). *Cycas maconochiei* ssp. *maconochiei* population structure analysis (Figure 14) showed the same results for *Cycas armstrongii* containing only a single genetic group of *K* = 1. The structure plot combining *C. armstrongii*, *C. maconochiei* ssp. *maconochiei* and *C. armstrongii* x *maconochiei* (Figure 15), showed only a single genetic group (*K* = 1), with no differentiation between populations, or between taxa.

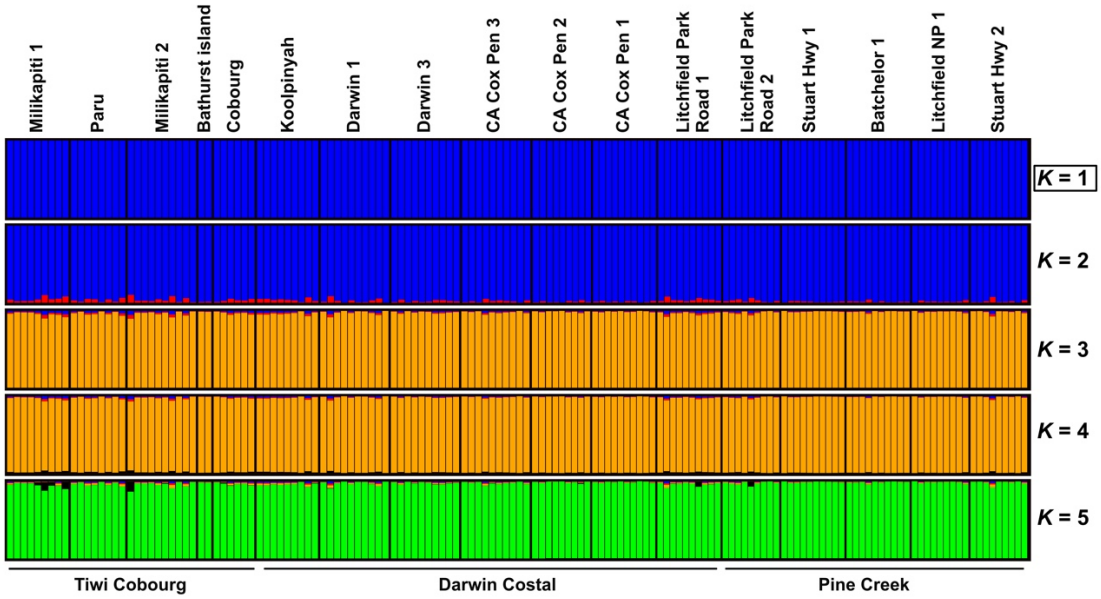

Figure 13 Population structure plot for *Cycas armstrongii* populations. Population structure plot represents 150 samples from 12 populations in 3 defined regions representing the Interim Biogeographic Regionalisation for Australia (IBRA7). The highest model value indicated *K*=1 a single genetic within the data.

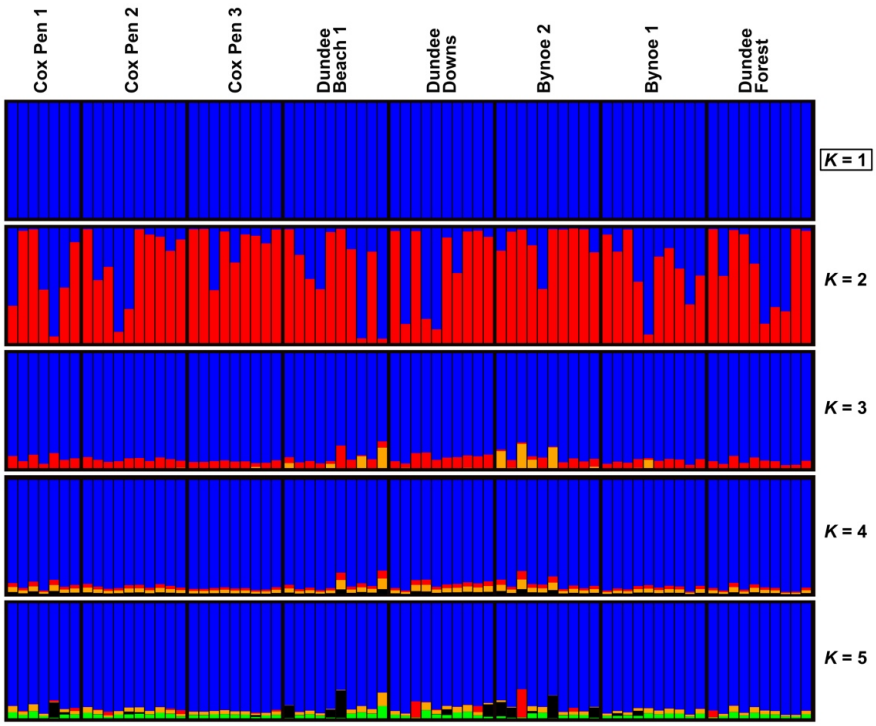

Figure 14 Population structure plot for *Cycas maconochiei* ssp. *maconochiei* populations. Population structure plot represents 77 samples from 8 populations in the IBRA7 defined Darwin Costal Region The highest model value *K*=1 indicating a single genetic group in the data.

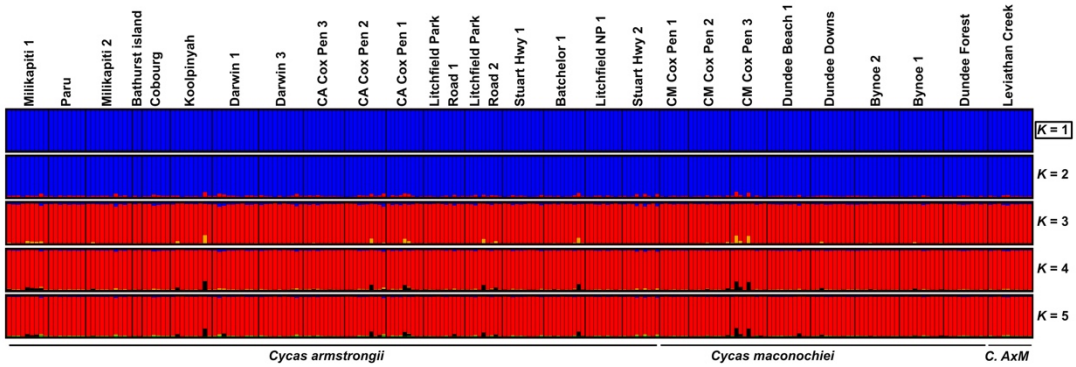

Figure 15 Population structure plot for *Cycas armstrongii, C. maconochiei* ssp. *maconochiei and Cycas armstrongii* x *maconochiei*. Population structure plot represents 247 samples from 22 populations in 3 regions from the defined Interim Biogeographic Regionalisation for Australia (IBRA7). The highest model value *K*=1 indicating a single genetic group in the data.

Discriminant analysis of principal components. The discriminant analysis of principal components (DAPC) was split into four analyses. The DAPC for *Cycas armstrongii* (Figure 16) was examined for IBRA subregions throughout the Darwin coastal (DAC), Pine Creek (PCK), and Tiwi Coburg (TIW) IBRA regions. The DAPC was a summary of 62 PCs from the PCA, with three genetic groups (*K* = 3) and a proportion of conserved variance of 0.618. Figure 16 shows little differentiation between Darwin Coastal and Pine Creek regions, indicating a close genetic connection. However, Tiwi (TIW02) and Coburg subregions (TIW01) show a higher level of genetic differentiation between both subregions and the Darwin coastal (DAC)/Pine Creek (PCK) regions.

The DAPC for *Cycas maconochiei* ssp. *maconochiei* (Figure 17) was representative of populations occurring in the Darwin coast (DAC) IBRA region, and was a summary of 31 PCs for the PCAs with two discriminant functions (*K* = 2) and a proportion of conserved variance of 0.619. However, the DAPC of *C. maconochiei* shows little differentiation between populations, indicating low levels of genetic distance among population, and potential gene flow. The DAPC of *C. armstrongii*, *C. maconochiei* ssp. *maconochiei*, and *C. armstrongii* x *maconochiei* (Figure 18) was a summary of 33 PC's from the PCA with two discriminant functions (*K* = 2), and a proportion of conserved variance of 0.526 (Figure 18). Figure 18 shows very low levels of differentiation between all taxa and provides little evidence to support hybridisation.

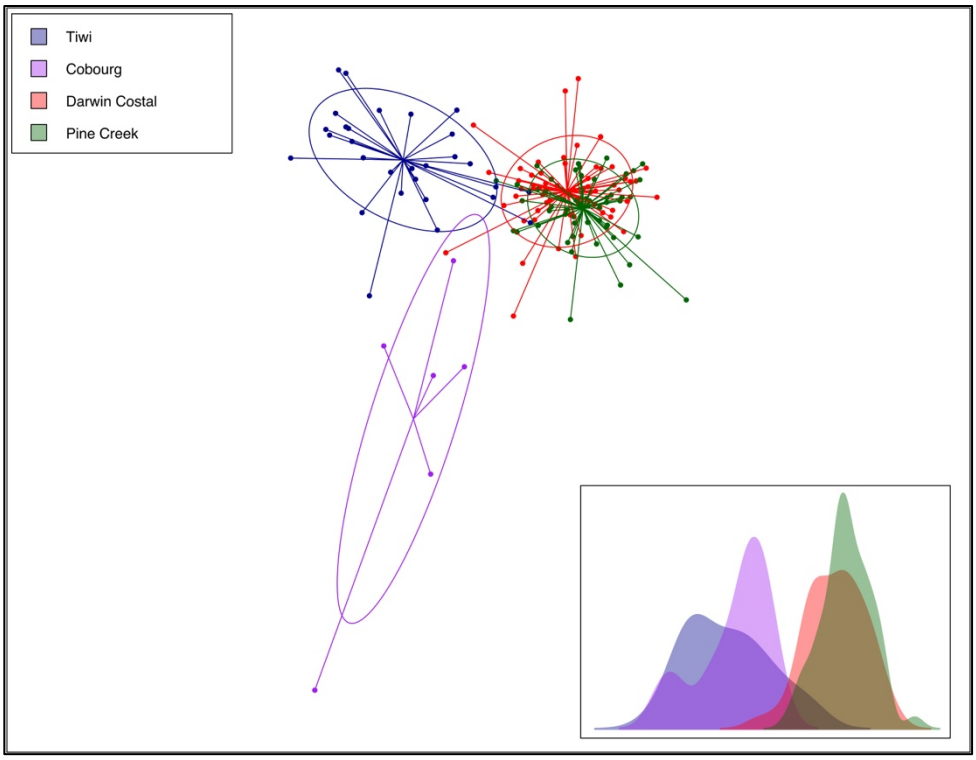

Figure 16 DAPC graph of *Cycas armstrongii* populations represented as IBRA subregions. Discriminate analysis of principal component (DAPC) showing genetic differentiation between *Cycas armstrongii* populations throughout the Darwin Costal, Pine Creek, and Tiwi Cobourg IBRA subregions in the Northern Territory Australia. DAPC is a summary of 62 PC's for the PCA's with three discriminant functions (*K* = 3) and a proportion of conserved variance of 0.618.

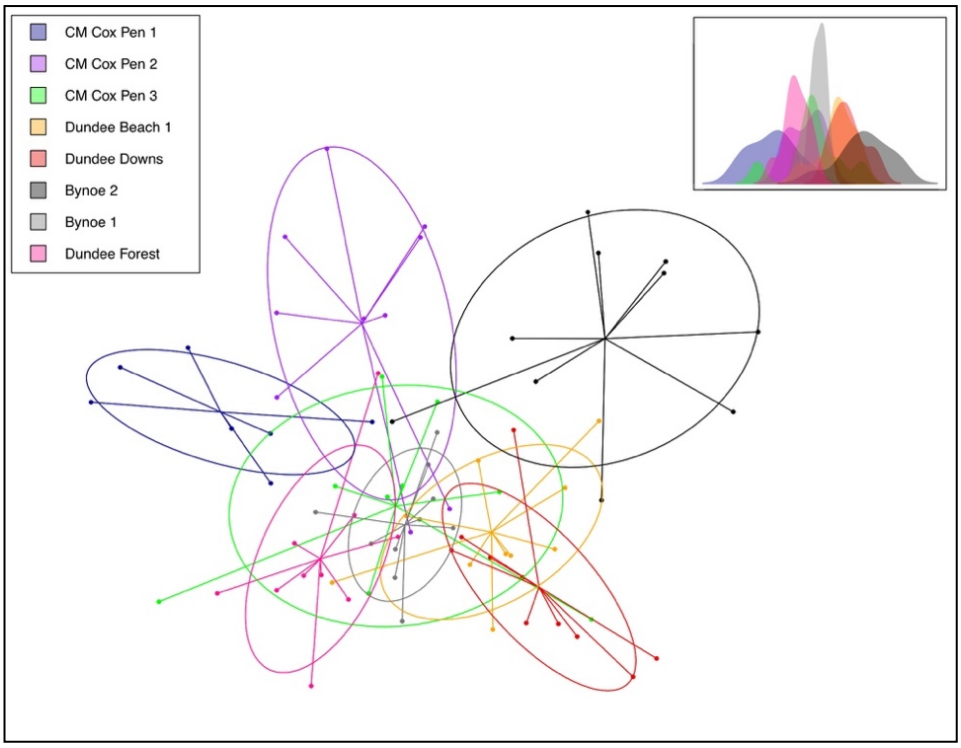

Figure 17 DAPC graph of *Cycas maconochiei* ssp. *maconochiei* grouped into populations. Discriminate analysis of principal component shows genetic differentiation between populations *C. maconochiei* ssp. *maconochiei*. The plot represents populations occurring the Darwin Costal Region, in the Northern Territory Australia. DAPC is a summary of 31 PC's for the PCA's with two discriminant functions (*K* = 2) and a proportion of conserved variance of 0.619.

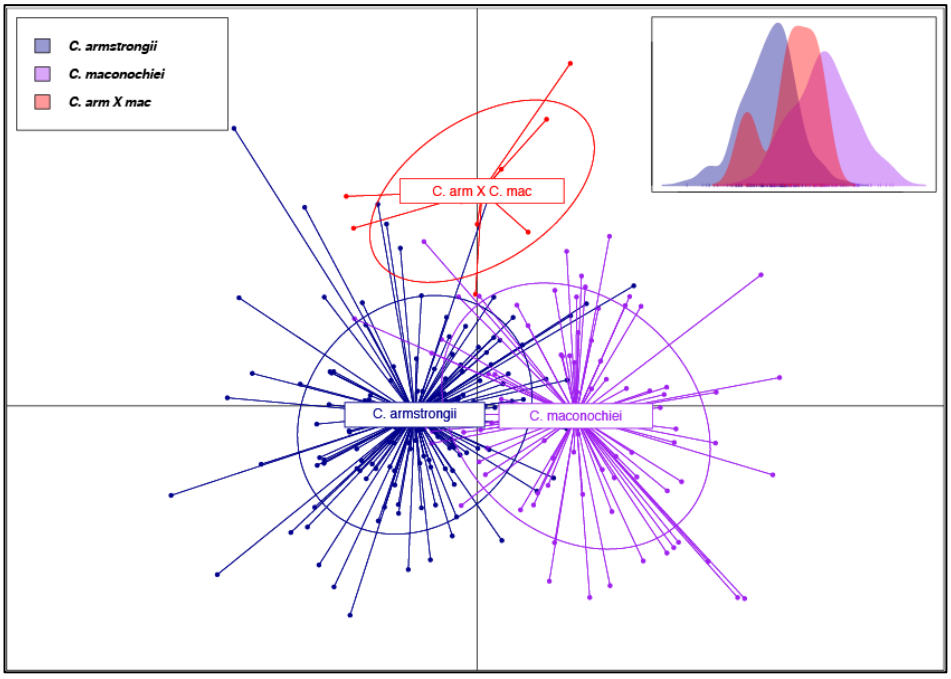

Figure 18 DAPC graph of *Cycas armstrongii, C. maconochiei* ssp. *maconochiei and Cycas armstrongii* x *maconochiei* population group by taxon. Discriminate analysis of principal component shows genetic differentiation between two species *Cycas armstrongii* and *C. maconochiei* ssp. *maconochiei* and a single hybrid population *C. armstrongii* x *maconochiei*. The plot represents all populations throughout the Darwin and greater region, in the Northern Territory Australia. DAPC is a summary of 33 PC's for the PCA's with two discriminant functions and a proportion of conserved variance of 0.526.

4.4 Discussion

This study aimed to obtain a better understanding of the genetic diversity within and between two geographically widespread cycad species with relatively large populations (Liddle 2009): *Cycas armstrongii* and *Cycas maconochiei* ssp. *maconochiei*. By using next generation sequencing (NGS) in the form of RADseq, we were able to generate 868 to 3043 genome-wide SNPs, which were likely subjected to the full evolutionary history of our study taxa (Andrews et al. 2016). In addition, we tested the validity of the formally recognised interspecific hybrid, *Cycas armstrongii* x *maconochiei* (Hill 1996).

Population genetic diversity. Measuring genetic diversity of populations is vital in understanding the life history of a species and detecting its ability to adapt to future challenges, whether this be through biotic or abiotic factors (Brown et al. 1983). Populations of *Cycas armstrongii* show very low levels of heterozygosity (Table 12, H_E = 0.033, H_O = 0.057), and little evidence of inbreeding in populations (F_{IS} = 0.007 to 0.040). Similar results were also recorded for *Cycas maconochiei* ssp. *maconochiei*, where the populations showed a relatively low observed heterozygosity (Table 13, H_E = 0.059, H_O = 0.049). However, the populations of *C. maconochiei* ssp. *maconochiei* also showed higher levels of inbreeding (F_{IS} = 0.037 to 0.243) when compared to *C armstrongii,* and are therefore more likely to express deleterious alleles (Kimura & King 1979). The single population of *Cycas armstrongii* x *maconochiei* (represented by 10 individuals from the Leviathan Creek population), also very low levels of observed heterozygosity and no evidence of inbreeding, which is in contrast with the other two taxa. Overall, these results suggest low levels of genetic diversity in all studied taxa.

The low levels of genetic diversity found in *C. armstrongii* and *C*. *maconochiei* ssp. *maconochiei* will have an impact on the genetic fitness of the populations, and could result in a reduction of individual fitness (Toczydlowski & Waller 2019). Markert et al. (2010) found this was the case in crustaceans where very low genetic diversity reduced the genetic fitness of a population and, under progressive conditions, reduced their adaptive performance which increased the risk of extinction. However, in small populations - like those of many cycads - inbreeding is often unavoidable, and can cause an increase in the accumulation of deleterious mutations (Wright et al. 2007; Donaldson 2003; Whitlock 2000). Deleterious mutations often correlate with low allelic genetic diversity, which can have an impact on the future adaptability of the populations (Szczecińska et al. 2016). Our results for *C. armstrongii* and *C*.

maconochiei ssp. *maconochiei* are similar to those found in *Cycas calcicola* (chapter 3), where the small, isolated populations had equally low levels of genetic diversity, however *Cycas calcicola* populations showed much higher levels of inbreeding. This could be an indicator for lower levels of inbreeding but not necessarily greater gene diversity for species with more widespread populations (e.g. *C. armstrongii*; Liddle 2009, Hill 1994). This is because larger and more diverse populations are considered to be less prone to the effects of inbreeding than small populations (Ellstrand 1993). Therefore, this could provide evidence that cycads with large populations might well be as threatened as cycads with small populations (Keppel 2002).

Population genetic studies of other Australian cycads have mostly concentrated on *Macrozamia* Miq. with a single study into *Cycas* (James et al. 2018), which showed similar results to our own indicating low levels of allelic diversity within populations (Sharma et al. 2004; Sharma et al. 1999; Sharma et al. 1998). However, low levels of genetic diversity is not restricted to cycad species in Australia as this is also the case for species of *Cycas* in the Indo-Malaya, and Palearctic (Long-Qian & Xun 2006; Xiao et al. 2005; Huang et al. 2004; Keppel et al. 2002). This indicates that low levels of genetic diversity characterising cycads, appears to be at odds with that of other plants with a dioecious life history (obligate out-crossing) (Norstog & Nicholls 1997), which by adaptation helps prevent the effects of inbreeding and self-pollination (Barrett 2002).

However, unlike many dioecious plant species, not all reproductive individuals of a population of cycad participate in a given reproductive event (Schmoldt et al. 1975; Ornduff 1991). Male cycad plants typically produce strobili more frequently than females because of the greater energy requirements and the longer development time of the megasporangiate strobili, longer fertilisation process and subsequent seed maturation (Terry et al. 2012; Suinyuy et al. 2009; Vovides et al. 1997). In addition, the poor dispersal system in cycads could make them more prone to inbreeding (Hall et al 2013), because any plant bearing female strobili is likely to be in close proximity to a related male (half or a full sibling) (Vovides et al. 1997). In Cycas although a single arrangement of megasporophylls on a ‘female’ plant may bear hundreds of ovules that might only be fertilised by a single ‘male’ plant (with the ability to be pollinated by many). This is not only because relatively few males are producing pollen (Norstog et al. 1986), but also because cycad pollen is poorly dispersed by wind and does not travel far, therefore there is a high likelihood of subsequent crosses between closely related individuals (Kono & Tobe 2007; Keppel

2001).

The insights gained into the genetic diversity of *Cycas armstrongii* and *C. maconochiei* ssp. *maconochiei* raises questions as to the causes of low genetic diversity within populations. There is most likely not a single answer to this question, as it could be due to a number of factors such as population size, isolation, absence of a primary seed dispersal agent (Ornduff 1990). Additionally, the slow generation times in many cycads species (Griffith et al. 2014), leads to low rates of mutation and substitution, reducing the short term adaptive potential of a population (Ellstrand 1993). Gymnosperms in general (including cycads) have been subjected to high extinction rates and recent post-Eocene radiations (Nagalingum et al. 2011), which might also account for the low genetic diversity of many extant lineages when compared to angiosperms (Crisp & Cook 2011).

Differences within and between populations. Geographic isolation of both species and natural populations is a major limiting factor in gene flow, therefore it contribute to genetic differentiation of species (Séré et al. 2017). Understanding the genetic differences between populations allows us to gain insights into the extent of gene flow and to predict which populations are more or less genetically distant from each other (Rousset 1997; Nei & Roychoudhury 1974). For *C. armstrongii* and *C. maconochiei* ssp. *maconochiei* understanding the genetic distances is important as it provides insight into not only the genetic differentiation between populations but also between taxa.

For *C. armstrongii* there was more genetic variation between populations (Table 14, 13%) then regions (3%). This suggest low levels of genetic differentiation between populations, indicating gene flow and admixture between both populations and regions (Figure 13). Pairwise genetic distance (F_{ST}, Table 15 and 16) are low to high between most populations and regions. However, populations occurring in the Tiwi Cobourg region showed much greater levels of genetic distance between regions (F_{ST} = 0.078 to 0.127) from the Darwin Coastal and Pine Creek regions, indicating evidence of genetic differentiation between Darwin Coastal/Pine Creek and Tiwi Cobourg regions, correlating with geography (Hill 1996; Liddle 2009). These results indicate genetic partitioning within the defined IBRA7 regions, confirming the Cobourg and Tiwi subregions being with the same region, when compared to Darwin Coastal and Pine Creek. This genetic partitioning is also somewhat shown in the discriminate analysis of principal component (DAPC, Figure 16), where the Tiwi Cobourg populations show some differentiation from the Darwin Coastal/Pine Creek populations.

The levels of genetic differentiation between populations of *Cycas maconochiei* ssp. *maconochiei* were similar to that of *C. armstrongii*, with low levels of genetic differentiation among populations (AMOVA = 6%), indicating that the majority of the genetic variation is within populations and suggests gene flow between populations, further confirmed by pairwise genetic distance (F_{ST} = 0.041 to 0.098, Table 17).

The effects of geographic isolation (as found in our taxa) and genetic drift are known to cause geographical differentiation between populations (Proćków et al. 2017). Populations occurring in close geographic proximity should have greater levels of gene flow (lower F_{ST} values) than between those that are more geographically distant (Hellberg 1994). Therefore, it is logical that gene flow would decrease with increased geographic distance, yet this is not always the case. Apart from some populations of *C. armstrongii* (mainly between DAC/PCK and TIW), most populations of *C. armstrongii* and *C. maconochiei* ssp. *maconochiei* showed low levels of genetic distance, indicating gene flow leading to lower differentiation between populations.

Even low levels of gene flow between populations can be beneficial in counteracting the negative effects of mutation isolation and genetic drift. The rates of gene flow in plants can vary significantly, and can act as a major driving force in the process of speciation and evolution (Ellstrand 2014). Tremblay & Ackerman (2001) found that in *Lepanthes* Sw., gene flow was restricted to less than one migrant per generation between populations, which is indictive of high genetic differentiation between populations. Restricted gene flow is known to increase population differentiation and have significant impact on the genetic fitness of a population (Newman & Tallmon 2001). The low levels of genetic divergence in most of *C. armstrongii* and *C. maconochiei* ssp. *maconochiei* populations are indicative of recent gene flow and not ancestral, which suggests a very recent disjunction between populations (Latta and Mitton 1999, Beerli 2004).

Geographic distribution of the taxa. Contemporary long distance dispersal with establishment is uncommon in most cycad species (Ingham et al. 2013). Members of the *Cycas rumphii* complex are known for their long-distance dispersal by water. This method of dispersal is facilitated by the presence of a spongy layer in the sclerotesta of seeds aiding their buoyancy and survival in salt water (Nadarajan et al. 2018). It is thought that hydrochory has accounted for their wide distribution in the Western Pacific (Hill 1994). The majority of *Cycas* do not have a similarly modified sclerotesta, and so rely entirely on zoochory and/or abiotic means of seed

dispersal (Keppel 2002).

The presence of genetically relatively distinct populations of *C. armstrongii* on the Tiwi Islands requires an explanation. Given that members of the now extinct megafauna (e.g. *Genyornis*) have been invoked as biotic vectors of Australian cycad seeds (Hall & Walter 2013; Ingham et al. 2013), the presence of *Cycas* on the Tiwi Islands would have required some means by which herds of large herbivores could move easily between the mainland and the islands. Two possible connections with the mainland might have existed, one linking the Tiwi Islands with the Darwin Coastal subregion via the Vernon island; and the other between the Tiwi Islands (TIW02) and the Cobourg Peninsula (TIW01). Although there is no dated geological evidence to support a terrestrial connection between the Tiwi Islands (TIW) and the mainland (DAC and PCK), it would seem reasonable that mainland source populations should be more genetically similar to Tiwi Island populations. The close genetic relationship between populations of *C. armstrongii* in the Tiwi Islands and Cobourg Peninsula, combined with evidence of a closer genetic relationship between Tiwi and Cobourg sub-regions (Table 16) could help support evidence of land bridge from the Tiwi Islands to the Cobourg Peninsula and Darwin Coastal (DAC01) and Pine Creek (PCK01) subregions.

One explanation is that fragmentation between populations is a recent event and one which hasn't yet left a genetic fingerprint, due to the long generation time in cycads, which could partially explain the results of this research (Liu et al. 2018). Although the seeds of all cycads are toxic to humans, the seeds can be prepared for consumption by leaching out the toxins (Beck 1992). This is documented in *C. armstrongii* where the seeds have been known to have been consumed by indigenous communities in the Darwin region (Beck 1992), and is also likely the case for *C. maconochiei* ssp. *maconochiei.* Unprocessed seeds of both taxa might have been transported by humans as a food source between mainland Australia (DCA01, PCK01 and TIW01) and the Tiwi Islands during a period of around 60 thousand years (Beaton 2014). This is the case for other plant groups where the movement of seeds and fruit by indigenous people had a significant impact on the surrounding plant communities (Hynes & Chase 1982). This might also explain the distribution of *C. armstrongii* and *C. maconochiei* ssp. *maconochiei* and why the disjunct populations still show low levels genetic differentiation caused by mixing.

Differences between taxa and the hybrid population. Understanding genetic differences between populations is important when determining species boundaries (Sousa & Hey 2013). We detected little genetic differentiation between our study taxa, which collates with morphological intermediates between many populations of both *C. armstrongii* from *C. maconochiei* ssp. *maconochiei* (I Cowie 2018, personal communication). Most of the detected genetic variation for each taxon was within populations, and not between them. We saw no evidence for maintaining the formal segregation of *C. armstrongii* from *C. maconochiei* ssp. *maconochiei*. Hybridisation is thought to be widespread in cycads (Calonje et al. 2011; Chamberlain 1926) based on the occurrence of morphologically intermediate individuals expressing synapomorphies of two parent species. The close genetic similarity and relatively recent origin of many cycad species (Nagalingum et al. 2011) means that they could rely on demographic isolation, rather than genetic isolation, for the maintenance of species integrity (Chiang et al. 2013). In the case of the morphological intermediate (*C. armstrongii* x *maconochiei*), hybridisation was not supported by genetic evidence due to no genetic structure between populations and species, and the lack of genetic differentiation between the assumed parental taxa.

Conservation implications. Infraspecific geographic disjunctions are common in many plant species (McHenry & Barrington 2014). Geographic separation often correlates with high levels of genetic differentiation and interpopulational variation caused by reduced gene flow across the disjunction (Feng et al. 2014; Keppel et al. 2002; Keppel 2002). Species that are slow to reproduce and occur only in small, isolated populations (Swart et al. 2018; Niissalo et al. 2017; Szczecińska et al. 2016; Ellstrand 1993) are prone to lower levels of genetic diversity, because they have a higher potential for reduced population fitness (Toczydlowski & Waller 2019; Markert et al. 2010; Calonje et al. 2013). Although *C. armstrongii* and *C. maconochiei* ssp. *maconochiei* are geographically widespread and have relatively large populations (Liddle 2009), they have very low levels of genetic diversity and generally low population differentiation. There also seems to be a trend in both plants and cycads caused by isolation where once a population drops to a critical size, there is a loss of genetic diversity which will eventually become reduced to the point where allelic diversity is permanently lost (Schwartz 2003).

The increased aridification combined with shorter/longer wet seasons in Northern Australia could mean few species are able to survive in their environmental niches, affecting pollinators and seed dispersers (if any) (Glenny et al. 2018). A wetter or drier climate could also have an impact on populations by reducing the number of seedlings that survive yearly (Preece et al. 2007), caused by an increased frequency of bush fires which could be exacerbated by poor dispersal in cycads, although their response to increased fires is unknown (Setterfield 2002). This would further be impaired by the low-genetic representation of *C. armstrongii* and *C. maconochiei* spp. *maconochiei* in ex-situ conservation collections, as they cannot currently be used for reintroduction programs (Parlato & Armstrong 2018; Griffith et al. 2015). Currently the IUCN red list lists *C. armstrongii* as vulnerable and *C. maconochiei* ssp. *maconochiei* as least concern (IUCN 2016), indicating that neither taxon is (currently) of conservation concern in Australia (DEWHA 2019).

From a conservation management perspective, the wide distribution of the taxa could make the in-situ management of individual populations logistically problematic unless they were conserved as larger units e.g. localities (James et al. 2018). But the importance of ex-situ botanic garden collections for conservation of plant species is imperative (Nikitsky Botanical Gardens 2017), because they have the potential to retain up to 90% of the genetic diversity for a species, they are the ideal place to conserve the genetic diversity of cycads to counteract the problem of inherent long term storage problems of cycad seeds (Griffith et al. 2015). Although, there is no genetic differentiation between populations formally recognised as *C. armstrongii* and *C. maconochiei* ssp. *maconochiei* there is still morphological differences between populations considered to be *C. maconochiei* ssp. *maconochiei* when compared to *C. armstrongii*. But although there is no genetic fingerprint between taxa, different morphotypes still need to be represented in ex-situ botanic gardens. However, populations recognised as *C. armstrongii* and *C. maconochiei* ssp. *maconochiei* are mostly absent from many ex-situ conservation collections (BG-Base online search, accessed 19 October 2019), which, indicates that there needs to be a focus on future seed collection to conserve different morphotypes of taxa, to incorporate into botanic gardens. This would then allow botanic gardens to act as a reserve for genetic diversity of wild populations and reduce extinction risk of wild populations

Considerations for further research. This research represents the first comprehensive examination of the genetic diversity of *C. armstrongii* and *C. maconochiei* ssp. *maconochiei*. The evidence provided by this study showed the efficiency of RADseq as a technique and sample sizes required for obtaining genomic level data from cycads. This is important as is can allow us to formally recognise species of conservation value. This technique is cost-effective and causes minimal harm to populations (Clugston et al. 2019). For further research into *C. armstrongii* and *C. maconochiei* ssp. *maconochiei* it is important to understand the genetic relationship of the other subspecies of *C. maconochiei* (*C. maconochiei* ssp. *lanata*) senu Dixon (2004) and other allied taxa including *C. canalis* and *C. conferta* that are considered to freely hybridise with *C. armstrongii* in the wild (Dixon 2004; Hill 1996).

Because of the genetic similarity between *C. armstrongii* and *C. maconochiei* ssp. *maconochiei*, a next step would be to gain further insights into genetic differentiation in the rest of the complex. Do *Cycas canalis* and *C. conferta* also show similarly low levels of genetic differentiation found in *C. armstrongii* and *C. maconochiei* ssp. *Maconochiei*? If this is the case, then it could indicate that although there is morphological differentiation between species in the complex there could be fewer species than suspected (Liddle 2009). By obtaining further insights into the genetic diversity of the complex, we will better understand genetic diversity of the genus *Cycas* in Australia. This will allow us to determine if other species of Cycas show equally low levels of genetic diversity found in our study taxa and provide more explanations to why.

As a result of the number of markers and the number of samples that were sequenced during this project we were able to successfully represent the genetic diversity of populations using seven or more samples per population (Nazareno et al. 2017). However, further genotyping of a greater number of populations for both *C. armstrongii* and *C. maconochiei* ssp. *maconochiei* may help to gain further insights into their genetic diversity, although it is unlikely to show differences between the species.

Conclusions

Cycas armstrongii and *C. maconochiei* spp. *maconochiei* represent two taxa with a wide geographic distribution and in many cases large populations (Liddle 2009; Hill 1996), yet the populations of the taxa have low levels of allelic diversity, with some populations showing significantly higher levels of inbreeding (higher inbreeding in *C. maconochiei* spp. *maconochiei*), indicating historic isolation and population fragmentation are affecting the genetic fitness in both species. Although there is geographic differentiation between population and species, the evidence presented here finds not only low levels of genetic differentiation between populations but also between regions and taxa. The results indicate that although there are morphological (sensu Hill 1994) and geographic differences between *C. armstrongii* and *C. maconochiei* ssp. *maconochiei,* and a supposed recognisable interspecific hybrid population, our results show the opposite. The molecular data suggests that it is likely that *C. armstrongii* and *C. maconochiei* ssp. *maconochiei* should be recognised as a single morphologically diverse species with a wide distribution across the northern range of Northern Territory. We conclude that the contemporary populations represent relictual populations derived from much larger ancestral populations (Laidlaw & Forster 2012), and the lack of genetic differentiation between *C. armstrongii* and *C. maconochiei* ssp. *maconochiei* also means that there is no molecular support for the formal recognition of *C. armstrongii* x *maconochiei*.

Chapter 5 - Conclusion

This study presents a case for exploring new techniques and developing new approaches to be applied to the population genetics of cycads. By using RADseq we have provided important insights into the genetic history and diversity of the genus *Cycas* in Australia. RADseq offers the ability to multiplex and sequence many individuals simultaneously, at relatively low cost. Using RADseq we were able to recover up to 3043 genome wide SNPs for at least 50% of the samples per locus.

While this study focused on selected Australian taxa, the techniques are applicable to all cycad species. The markers generated have the potential to be effective for both population level and phylogenetic studies (Catchen et al. 2017; Tripp et al. 2017; McKinney et al. 2016; Davey & Blaxter 2010). Here we have demonstrated that RADseq can be applied to organisms with large genomes such as cycads, and we have developed a protocol for others to follow. Ultimately, the methodology presented here will help to resolve the relationships among cycads and to help gain deeper insights into the genetic diversity among Cycadales species. This data and relevant techniques are directly applicable when developing informed conservation management plans for cycads and other groups of seed plants.

The research presented here focused on three taxa - *Cycas armstrongii*, *C. calcicola* and *C. maconochiei* ssp. *maconochiei* and a single suspected hybrid population *Cycas armstrongii* x *maconochiei*. In *C. calcicola* the populations are small and geographically disjunct but have a wide distribution with low levels of genetic diversity and low differentiation between populations. However, despite the low levels of differentiation between populations, there was evidence of genetic differences between the Lichfield and Katherine regions which correlates with geographic distances. This indicates that the spatial arrangement of *C. calcicola* populations may be recent. We found high inbreeding in some populations, which could result in lower adaptive potential of the species. This, in turn, raised concerns for the conservation of species in the face of rapid global climate change and anthropogenic threats to habitats.

In order to mitigate the conservation concerns, we determined if ex-situ conservation collections in botanic gardens represented the genetic diversity of the wild populations. Our results found that populations of *C. calcicola* from the Litchfield National Park were currently not represented in well managed and curated ex-situ collections, yet they do occur in National Parks and are afforded some degree of protection. Regardless of the current levels of protection for some populations this species, we were able to identify that a considerable amount of genetic diversity of the species is not represented in ex-situ collections. As a result of these factors the best approach towards conserving *C. calcicola* is to conserve each population in the Litchfield National Park and Katherine regions as separate conservation management units.

Advocating representative ex-situ collection of cycads for the purpose of conservation presents a series of problems that our approach to population genetics might assist with. In particular, our techniques provide insights into what type of material, and how much of it, should be collected in order to represent the optimal genetic diversity of a taxon. The simplest solution would be to collect seed from wild populations, yet while seed production by most cycads is relatively high (~160 viable seeds from a single female strobilus), the number of seeds produced in each season is limited by the number of females and males simultaneously coning in a reproductive event (Ballardie et al. 1986). Seed recruitment in populations of Australian cycads can also be significantly reduced by increased burning frequencies (Hall et al. 2013; Liddle 2009; Ornduff 1991). We have also shown that interpopulational seed dispersal (i.e. migration) in Australia cycads is rare. Thus, the targeted removal of seeds for ex-situ conservation from small populations with low genetic diversity could place further stress upon the long-term survival of some Australian *Cycas* species (Andersen 1989).

Populations of *Cycas armstrongii* and *C. maconochiei* ssp. *maconochiei* had very low levels of allelic diversity and differentiation across the populations with some evidence of inbreeding. Furthermore, low levels of genetic differentiation between the two taxa were observed. These results indicated that *C. armstrongii* and *C. maconochiei* ssp. *maconochiei* might represent a single, morphologically diverse species with a wide distribution across the lower latitudes of Northern Territory. By extension, our results did not recover any support for the formal recognition of the interspecific hybrid, *Cycas armstrongii* x *maconochiei*. Thus, we conclusively demonstrated that RADseq protocols are sensitive enough to enable similar species relationship (and putative hybrids) to be assessed independently of morphological

hypotheses, using genomewide markers (Clugston et al. 2019). However, although studies like our own which used next generation sequencing on cycads have provided a significantly greater number of markers, the resulting genetic statistics have found no greater diversity in cycads in Australia, likely due to historic factors (Sharma et al. 2004; Sharma et al. 1999). Of course, this does not mean that techniques such as RADseq are ineffective on groups like cycads – in fact they are the best option available and have the most potential for uncovering the evolutionary history of the group (Clugston et al. 2019).

The genetic patterns within and between *C. armstrongii* and *C. maconochiei* ssp. *maconochiei*, as revealed by our use of RADseq techniques, provided us with an opportunity to investigate the process that had led to the current genetic architecture and the spatial arrangement of populations. We concluded that the contemporary populations represent relictual populations derived from much larger ancestral populations (Laidlaw & Forster 2012). The fragmentary distribution of contemporary cycads in the Northern Territory of Australia and range retraction could also be a result of the extension of their primary, megafaunal seed dispersal (Hall and Walter 2013). The demise of seed vectors would mean that geographic range extension, and migration between populations, would be significantly curtailed. In turn, the resultant small population sizes and restricted geographic distribution, combined with slow rates of reproduction, could explain the reduction in allelic diversity found in *C. armstrongii*, *C. calcicola,* and *C. maconochiei* ssp. *maconochiei.*

It is conceivable that species of *Cycas* in the Northern territory of Australia could have diversified recently and rapidly (Crisp & Cook 2011; Nagalingum et al. 2011). In particular, the evidence presented by Crisp & Cook (2011) shows that not only cycads have been subjected to multiple extinction and radiation events but also other members of the gymnosperms, which could have accounted for the low levels of genetic diversity in many species, given selective pressures acting on cycad populations (Donaldson 2003). In the case of cycads growing in the Northern Territory of Australia as studied here, warmer, drier periods are likely to cause a range restrictions, whereas cooler drier periods with reduced sea level (e.g. glacial maxima) offer greatly increased habitat areas, with dispersal across land bridges (Preece et al. 2007), allowing periodic mixing of populations (Erwin 2009).

Rapid divergence of closely related plant lineages does not always keep pace with the evolution of breeding barriers, due to a range of biotic and abiotic factors such as pollinators and habitats (Rieseberg and Willis 2007). Additionally, factors such as changes in cone production times and self-fertilisation as a result of mutation can have a significant effect of the genetic variance of population (Macnair 1989). In such cases, demographic barriers serve as isolating mechanisms between closely related taxa. When the demographic barriers are relaxed (Widmer et al. 2009), populations of otherwise segregated lineages become parapatric or sympatric, thereby increasing the likelihood for gene flow between the closely related lineages. Because the effects of habitat fragmentation isolates populations and reduces population size, which can cause genetic erosion and inbreeding, which affects the reproductive potential of a population (Honnay 2002). Although even minor levels of gene flow among populations can help spread advantageous alleles, and can counteract the effects of genetic erosion (Morjan and Rieseberg 2004). Yet for our study taxa although the low interpopulational differentiation, which indicates significant geneflow between populations which are geographically. This means that the isolation is perhaps due to recent fragmentation and exacerbated by poor seed dispersal, which has not left a genetic fingerprint, perhaps due to anthropogenic causes (Leblois et al. 2006). These results demonstrate importance in understanding the demographic history of species with both small and large populations and is a key factor in determining the extinction potential of a species (Oostermeijer et al. 2003).

If the reproductive barriers are not fixed in plants such as cycads, then closely related lineages (species) can hybridise to produce seemingly novel morphological intermediates (Baack et al. 2015). Cycads display this phenomenon repeatedly, which is not consistent with the traditional view of cycads as a slow evolving and ancient group of organisms. The results presented by this study presents evidence from other cycads species with low levels of genetic diversity that may be threatened by extinction. In particular, populations of *C. calcicola* are likely to become more isolated and show lower levels of genetic diversity. This does not look favourable for the genetic fitness of populations and could have a direct effect on the future adaption of species (Mankga & Yessoufou 2017). However, both *C. armstrongii* and *C. maconochiei* ssp. *maconochiei* showed similar patterns found in *C. calcicola* despite having larger populations, indicating a trend for lower genetic diversity of cycads in Australia.

Our work demonstrates despite the rapid and recent diversification of cycads (Nagalingum et al. 2011), RADseq provides a valuable approach to understand the genetic diversity of taxa with large genomes, and used to aid in the conservation of species (Leitch & Leitch 2013) and for investigating similar cases across the cycads despite their very large and complex genomes. As for the future of cycad conservation and survival of the species, well-informed botanic gardens with excellent genetic representation in well-curated collections will play a fundamental role when combined with well-informed conservation management plans in conserving genetic diversity of the species (Griffith et al. 2015). Ultimately, this study presents an excellent case for the application of RADseq not only in cycads, but also for other organisms with large and complex genomes (Clugston et al. 2019).

References

Aguilar, R. et al., 2019. Habitat fragmentation reduces plant progeny quality: a global synthesis. J. M. Gomez, ed. *Ecology letters*, 22(7), pp.1163–1173.

Ahrens, C.W. et al., 2017. Genomic diversity guides conservation strategies among rare terrestrial orchid species when taxonomy remains uncertain. *Annals of Botany*, 119(8), pp.1267–1277.

Alexeyenko, A. et al., 2014. Efficient de novo assembly of large and complex genomes by massively parallel sequencing of Fosmid pools. *Bmc Genomics*, 15(1), p.439.

Alix, K. et al., 2017. Polyploidy and interspecific hybridization: partners for adaptation, speciation and evolution in plants. *Annals of Botany*, 120(2), pp.183–194.

Allendorf, F.W., Hohenlohe, P.A. & Luikart, G., 2010. Genomics and the future of conservation genetics. *Nature Reviews. Genetics*, 11(10), pp.697–709.

Amos, W. & Balmford, A., 2001. When does conservation genetics matter? Heredity, 87(3), pp.257–265.

Andersen, A. N. (1989). How important is seed predation to recruitment in stable populations of long-lived perennials?. *Oecologia*, 81(3), 310-315.

Andersen, A.N., Hoffmann, B.D. & Oberprieler, S., 2018. Diversity and biogeography of a species-rich ant fauna of the Australian seasonal tropics. *Insect science*, 25(3), pp.519–526.

Andrews, K.R. et al., 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews. Genetics*, 17(2), pp.81–92.

Andrews, K.R. et al., 2014. Trade-offs and utility of alternative RADseq methods: Reply to Puritz et al. *Molecular Ecology*, 23(24), pp.5943–5946.

Andrews K. R., Luikart G., 2014. Recent novel approaches for population genomics data analysis. *Molecular Ecology*, 23(7), pp. 1661–1667.

Andrews, S. et al., 2014. FastQC. A quality control tool for high throughput sequence data. *Babraham Bioinformatics*, www.bioinformatics.babraham.ac.uk.

Aristizábal, A., Tuberquia, D.J. & Sanín, M.J., 2018. Conservation Genetics of Two Highly Endangered and Poorly Known Species of Zamia (Zamiaceae: Cycadales) in Colombia. *The Journal of heredity*, 109(4), pp.438–445.

Arnold, B. et al., 2013. RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. *Molecular Ecology*, (11), p.3179.

Australian Government. 2000. 'Interim Biol.-geographic Regionalisation for Australia (IBRA) and development of Version 7.' (Department of Environment and Water Resources: Canberra.) Available at: www.environment.gov.au/parks/nrs/ibra/version5-1/summary-report/summary-report4.html (accessed 24 August 2019).

Baack, E. et al., 2015. The origins of reproductive isolation in plants. *The New Phytologist*, 207(4), pp.968–984.

Bache, S.M. & Wickham, H., 2018. "magrittr" A Forward-Pipe Operator for R. Available at: https://cran.r-project.org/web/packages/magrittr/magrittr.pdf.

Bachman, S.P., Nic Lughadha, E.M. & Rivers, M.C., 2018. Quantifying progress toward a conservation assessment for all plants. *Conservation Biology*, 32(3), pp.516–524.

Bailey, R. et al., 2010. Life table of cycad scale, *Aulacaspis yasumatsui* (Hemiptera: Diaspididae), reared on *Cycas* in Taiwan. *Journal of Asia-Pacific Entomology*, 13(3), pp.183–187.

Ballardie, R.T. & Whelan, R.J., 1986. Masting, seed dispersal and seed predation in the cycad *Macrozamia communis*. *Oecologia*, 70(1), pp.100–105.

Barchi, L. et al., 2011. Identification of SNP and SSR markers in eggplant using RAD tag sequencing. *Bmc Genomics*, 12(1), p.304.

Barton, N.H., 2001. The role of hybridization in evolution. *Molecular Ecology*, 10(3),

pp.551–568.

Barrett, S. C. 2002. Evolution of sex: the evolution of plant sexual diversity. *Nature Reviews Genetics*, 3(4), 274.

Bayzid, M.S., Hunt, T., Warnow, T. 2014. Disk covering methods improve phylogenomic analyses. BMC Genomics, 15(6), p.7.

Beaton, J.M., 2014. Fire and Water: Aspects of Australian Aboriginal Management of Cycads. *Archaeology in Oceania*, 17(1), pp.51–58.

Beck, W., 1992. La preparatión Aboriginál de las semillas deCycas en Australia. *Economic botany*, 46(2), pp.133–147.

Beerli, P. 2004. Effect of unsampled populations on the estimation of population sizes and migration rates between sampled populations. *Molecular Ecology*, 13(4), 827-836.

Benestan, L.M. et al., 2016. Conservation genomics of natural and managed populations: building a conceptual and practical framework. *Molecular Ecology*, 25(13), pp.2967–2977.

BG-Base. 2018. BG-Base Collections Management Software user collections search. Available at: http://rbg-web2.rbge.org.uk/multisite/multisite3.php (accessed 19 October 2019).

Bolger, A.M., Lohse, M. & Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), pp.2114–2120.

Broadhurst, L., & Coates, D. 2017. Plant conservation in Australia: current directions and future challenges. *Plant diversity*, 39(6), 348-356.

Brookes, D. R., Hereward, J. P., Terry, L. I., & Walter, G. H. 2015. Evolutionary dynamics of a cycad obligate pollination mutualism–Pattern and process in extant *Macrozamia* cycads and their specialist thrips pollinators. *Molecular phylogenetics and evolution*, 93, 83-93.

Brown, A. H., & Weir, B. S. 1983. Measuring genetic variability in plant populations. *Isozymes in plant genetics and breeding*, part A, 219-239.

Cariou, M., Duret, L. & Charlat, S., 2013. Is RAD-seq suitable for phylogenetic inference? An in silico assessment and optimization. *Ecology and evolution*, 3(4), pp.846–852.

Cabrera-Toledo, D. et al., 2010. Allozyme diversity levels in two congeneric *Dioon* spp. (Zamiaceae, Cycadales) with contrasting rarities. *Plant Systematics & Evolution*, 290(1-4), pp.115–125.

Cabrera-Toledo, D., Gonzalez-Astorga, J. & Vovides, A.P., 2008. Heterozygote excess in ancient populations of the critically endangered *Dioon caputoi* (Zamiaceae, Cycadales) from central Mexico. *Botanical Journal of the Linnean Society*, 158(3), pp.436–447.

Cabrera-Toledo, D., González-Astorga, J. & Flores-Vázquez, J.C., 2012. Fine-scale spatial genetic structure in two Mexican cycad species *Dioon caputoi* and *Dioon merolae* (Zamiaceae, Cycadales): Implications for conservation. *Biochemical Systematics & Ecology*, 40, pp.43–48.

Calonje M, Stevenson DW, Stanberg L. The World List of Cycads, online edition [Internet]. 2013-2018. [cited 2018 Jul 21]. Available from: http://www.cycadlist.org.

Calonje, M., Meerow, A.W., Knowles, L., Knowles D. Cycad biodiversity in the Bahamas Archipelago and conservation genetics of the threatened *Zamia lucayana* (Zamiaceae). *Oryx*, 47(2), pp.190–198.

Calonje, M., Kay, J. & Griffith, P.M., 2011. Propagation of Cycad Collections from Seed: Applied Reproductive Biology for Conservation. *Sibbaldia: the Journal of Botanic Garden Horticulture*, 9, pp.77–96.

Cantrill, D. J. 2018. The Australasian Virtual Herbarium: Tracking data usage and benefits for biological collections. *Applications in plant sciences*, 6(2), e1026. Available at: https://avh.chah.org.au (accessed 12th January 2015).

Cariou, M., Duret, L. & Charlat, S., 2013. Is RAD-seq suitable for phylogenetic inference? An in silico assessment and optimization. *Ecology and evolution*, 3(4), pp.846–852.

Catchen, J.M. et al., 2017. Unbroken: RADseq remains a powerful tool for understanding the genetics of adaptation in natural populations. *Molecular ecology resources*, 17(3), pp.362– 365.

Chamberlain, C.J., 1926. Hybrids in Cycads. *Botanical Gazette*, 81(4), pp.401–418.

Chapman, A. D. 2009. Numbers of living species in Australia and the world. Report for the Australian Biological Resources Study Canberra, Australia. September 2009.

Charlesworth, D. & Willis, J.H., 2009. The genetics of inbreeding depression. *Nature Reviews. Genetics*, 10(11), pp.783–796.

Charlesworth, D. 2003. Effects of inbreeding on the genetic diversity of populations. Philosophical Transactions of the Royal Society of London. Series B: *Biological Sciences*, 358(1434), 1051-1070.

Chaw, S.-M. et al., 2005. A phylogeny of cycads (Cycadales) inferred from chloroplast matK gene, trnK intron, and nuclear rDNA ITS region. *Molecular Phylogenetics and Evolution*, 37(1), pp.214–234.

Chen, G., Sun, W., 2018. The role of botanical gardens in scientific research, conservation, and citizen science. *Elsevier*, 40(4), pp.181–188.

Chen, C., Mitchell, S. E., Elshire, R. J., Buckler, E. S., & El-Kassaby, Y. A. 2013. Mining conifers' mega-genome using rapid and efficient multiplexed high-throughput genotyping-by-sequencing (GBS) SNP discovery platform. *Tree Genetics & Genomes*, 9(6), 1537-1544.

Chen, S. Y., Zhang, Y. J., Wang, X. L., Sun, J. Y., Xue, Y., Zhang, P., & Qu, L. H. 2012. Extremely low genetic diversity indicating the endangered status of *Ranodon sibiricus* (Amphibia: Caudata) and implications for phylogeography. *PLoS One*, 7(3), e33378.

Chiang, Y.-C. et al., 2013. Asymmetric introgression in the horticultural living fossil cycas sect. Asiorientales using a genome-wide scanning approach. *International journal of molecular sciences*, 14(4), pp.8228–8251.

Chifman, J., Kubatko, L. 2015. Quartet inference from SNP data under the coalescent model. *Bioinformatics*, 30(23), pp. 3317-3324.

Cibrian-Jaramillo, A. et al., 2013. What is the conservation value of a plant in a botanic garden? Using indicators to improve management of ex situ collections. *The Botanical Review,* 79(4), 559-577.

Cibrián-Jaramillo, A. et al., 2010. When North and South don't mix: genetic connectivity of a recently endangered oceanic cycad, *Cycas micronesica*, in Guam using EST-microsatellites. *Molecular Ecology*, 19(12), pp.2364–2379.

CLC Genomics Workbench 11.0 (2019). (https://www.qiagenbioinformatics.com/)

Clark, D.A. & Clark, D.B., 1987. Temporal and environmental patterns of reproduction in *Zamia skinneri*, a tropical rain forest in cycad. *Journal of Ecology*, 75(1), p.135.

Clugston, J. A. R., Milne, R. R., Kenicer, G. J., Overcast I., Wilson, T. C., Nagalingum, N. S. RADseq as a valuable tool for plants with large genomes—a case study in cycads. Molecular Ecology Resources. Accepted. 2019.

Clugston, J. et al., 2016. *Zamia* (Zamiaceae) Phenology in a Phylogenetic Context: Does in situ Reproductive Timing Correlate with Ancestry? *Edinburgh Journal of Botany*, 73(3), pp. 345-370.

Coates, D.J. & Atkins, K.A., 2001. Priority setting and the conservation of Western Australia's diverse and highly endemic flora. *Biological conservation*, 97(2), pp.251–263.

Cohen, J.I. et al., 1991. Ex Situ Conservation of Plant Genetic Resources: Global Development and Environmental Concerns. *Science*, 253(5022), pp.866–872.

Cousins, S.R., Williams, V.L. & Witkowski, E.T., 2011. Quantifying the Trade in Cycads (*Encephalartos* Species) in the Traditional Medicine Markets of Johannesburg and Durban, South Africa1. *Economic botany*, 65(4), pp.356–370.

Cousins, S.R., Williams, V.L. & Witkowski, E.T.F., 2013. Sifting through cycads: A guide to identifying the stem fragments of six South African medicinal Encephalartos species. *South African Journal of Botany*, 84, pp.115–123.

Cousins, S.R., Williams, V.L. & Witkowski, E.T.F., 2012. Uncovering the cycad taxa (*Encephalartos* species) traded for traditional medicine in Johannesburg and Durban, South Africa. *South African Journal of Botany*, 78, pp.129–138.

Cotrim, H. et al., 2016. Marked hybridization and introgression in Ophrys sect. Pseudophrys in the western Iberian Peninsula. *American Journal Of Botany*, 103(4), pp.677–691.

Cromie, G.A. et al., 2013. Genomic sequence diversity and population structure of *Saccharomyces cerevisiae* assessed by RAD-seq. *G3: Genes| Genomes| Genetics*, 3(12), pp.2163–2171.

Crisp, M.D. & Cook, L.G., 2011. Cenozoic extinctions account for the low diversity of extant gymnosperms compared with angiosperms. *The New Phytologist*, 192(4), pp.997–1009.

Cruaud, A. et al., 2014. Empirical Assessment of RAD Sequencing for Interspecific Phylogeny. *Molecular biology and evolution*, 31(5), pp.1272–1274.

De La Torre, A. R., Wilhite, B., & Neale, D. B. (2019). Environmental genome-wide association reveal climate adaptation is shaped by subtle to moderate allele frequency shifts in loblolly pine. Genome biology and evolution.

da Silva Carvalho, C. et al., 2015. Contemporary and historic factors influence differently genetic differentiation and diversity in a tropical palm. *Heredity*, 115(3), pp.216–224.

Da Silva, J.M. et al., 2012. Population genetics and conservation of critically small cycad populations: a case study of the Albany Cycad, *Encephalartos latifrons* (Lehmann). *Biological Journal of the Linnean Society*, 105(2), pp.293–308.

Davey, J.L. & Blaxter, M.W., 2010. RADSeq: next-generation population genetics. *Briefings in Functional Genomics*, 9(5/6), pp.416–423.

Dehgan, B., & Yuen, C. K. K. H. (1983). Seed morphology in relation to dispersal, evolution, and propagation of *Cycas* L. *Botanical Gazette*, 144(3), 412-418.

DEWHA—Department of the Environment, Water, Heritage and the Arts. 2009b. EPBC Act List of Threatened Flora. http://www.environment.gov.au/cgi-bin/sprat/public/publicthreatenedlist.pl?wanted=flora [Accessed on 8th September 2019]

de Vere, N. et al., 2009. Population size and habitat quality affect genetic diversity and fitness in the clonal herb *Cirsium dissectum*. *Oecologia*, 159(1), pp.59–68.

Department of the Environment and Energy, 2018. Interim Biogeographic Regionalisation for Australia, Version 7. Available at: www.environment.gov.au.

Dixon, D., 2004. A review of the gymnosperms of the Northern Territory. *Beagle: Records of the Museums and Art Galleries of the Northern Territory*, 20, pp.1–24.

Donaldson, J.S., 2009. Botanic gardens science for conservation and global change. *Trends in Plant Science*, 14(11), pp.608–613.

Donaldson, J.S., 2003. *Cycads: status survey and conservation action plan*, IUCN--the World Conservation Union.

Donaldson, J.S., 1993. Mast-seeding in the cycad genus *Encephalartos*: a test of the predator satiation hypothesis. *Oecologia*, 94(2), pp.262–271.

Dosmann, M.S., 2006. Research in the Garden: Averting the Collections Crisis. *The Botanical Review*, 72(3), pp.207–234.

Drury, C. et al., 2017. Genomic patterns in *Acropora cervicornis* show extensive population structure and variable genetic diversity. *Ecology and evolution*, 7(16), pp.6188–6200.

Eaton, D.A.R. and Overcast, I.A. 2019. ipyrad: interactive assembly and analysis of RADseq data sets. macOS and Linux. Available at: http://ipyrad.readthedocs.io.

Eaton, D.A.R., 2014. PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, 30(13), pp.1844–1849.

Eaton, D.A.R. & Ree, R.H., 2013. Inferring Phylogeny and Introgression using RADseq Data: An Example from Flowering Plants (*Pedicularis*: Orobanchaceae). *Systematic Biology*, 62(5), pp.689–706.

Eaton, D.A.R., Hipp, A.L., González-Rodríguez, A. and Cavender-Bares, J. (2015). Historical introgression among the American live oaks and the comparative nature of tests for introgression. Evolution, 69(10), pp. 2587–2601.

Eaton, D.A.R. et al., 2016. Misconceptions on Missing Data in RAD-seq Phylogenetics with a Deep-scale Example from Flowering Plants. *Systematic Biology*, 66(3), PP.399-412.

Edwards, S.V., 1993. Mitochondrial Gene Genealogy and Gene Flow Among Island and Mainland Populations of a Sedentary Songbird, the Grey-Crowned Babbler (*Pomatostomus Temporalis*). Evolution, 47(4), pp.1118–1137.

Edwards, R.D. et al., 2017. Congruent biogeographical disjunctions at a continent-wide scale: Quantifying and clarifying the role of biogeographic barriers in the Australian tropics. *PLoS ONE*, 12(4), pp.1–17.

Ekué, M.R.M. et al., 2008. Population genetics of the cycad *Encephalartos barteri* ssp. *barteri* (Zamiaceae) in Benin with notes on leaflet morphology and implications for conservation. *Belgian Journal of Botany*, 141(1), pp.78–94.

Ellstrand, N.C., 2014. Is gene flow the most important evolutionary force in plants? *American Journal of Botany*, 101(5), pp.737–753.

Ellstrand, N.C., 1993. Population Genetic Consequences of Small Population Size: Implications for Plant Conservation. *Annual review of ecology and systematics*, 24(1), pp.217–242.

Ellstrand, N.C., Ornduff, R. & Clegg, J.M., 1990. Genetic Structure of the Australian Cycad, *Macrozamia Communis* (Zamiaceae). (5), p.677.

Ellstrand, N.C., Devlin, B. & Marshall, D.L., 1989. Gene flow by pollen into small populations: Data from experimental and natural stands of wild radish. *Proceedings of the National Academy of Sciences*, 86(22), pp.9044–9047.

Engelhardt, K. et al., Effects of genetic diversity on conservation and restoration potential at individual, population, and regional scales. *Elsevier*

Erwin, D.H., 2009. Climate as a Driver of Evolutionary Change. *Current Biology*, 19(14), pp.R575–R583.

Fant, J.B. et al., 2016. What to do when we cant bank on seeds: What botanic gardens can learn from the zoo community about conserving plants in living collections. *American Journal of Botany*, 103(9), pp.1541–1543.

Fay, M.F., 2018. Orchid conservation: how can we meet the challenges in the twenty-first century? *Botanical studies*, 59(1), p.16.

Farrera, M. A. P., & Vovides, A. P. 2004. Spatial distribution, population structure, and fecundity of *Ceratozamia matudai* lundell (Zamiaceae) in El Triunfo Biosphere Reserve, Chiapas, Mexico. *The Botanical Review*, 70(2), 299-311.

Feng, X., Wang, Y. & Gong, X., 2014. Genetic diversity, genetic structure and demographic history of *Cycas simplicipinna* (Cycadaceae) assessed by DNA sequences and SSR markers. *BMC Plant Biology*, 14(1), pp.187–187.

Finlay, C.M.V. et al., 2017. Low genetic diversity and potential inbreeding in an isolated population of alder buckthorn (*Frangula alnus*) following a founder effect. *Scientific reports*, 7(1), p.3010.

Fischer, M., Hock, M. & Paschke, M., 2003. Low genetic variation reduces cross-compatibility and offspring fitness in populations of a narrow endemic plant with a self-incompatibility system. *Conservation Genetics*, 4(3), pp.325–336.

Fitz-Gibbon, S. et al., 2017. Phylogenomic inferences from reference-mapped and de novo assembled short-read sequence data using RADseq sequencing of California white oaks (*Quercus* section *Quercus*). J. P. Jaramillo-Correa, ed. *Genome*, 60(9), pp.743–755.

Flanagan, S.P. & Jones, A.G., 2018. Substantial differences in bias between single-digest and double-digest RAD-seq libraries: A case study. *Molecular Ecology Resources*, 18(2), pp.264–280.

Forster, P. I., Machin, P. J., Mound, L., & Wilson, G. W. 1994. Insects associated with reproductive structures of cycads in Queensland and northeast New South Wales, Australia. *Biotropica*, 26(2), 217-222.

Frankham, R., 2010. Challenges and opportunities of genetic approaches to biological conservation. *Biological conservation*, 143(9), pp.1919–1927.

Frankham, R., 1995. Conservation genetics. *Annual review of genetics*, 29(1), pp.305–327.

Frankham, R., 2003. Genetics and conservation biology. *Comptes Rendus Biologies*, 326, Supplement 1(0), pp.22–29.

Frankham, R., Ballou, J.D. & Briscoe, D.A., 2004. A primer of conservation genetics., *Comptes Rendus Biologies*, 326, Supplement 1(0), pp.22–29.

Freamo, H. et al., 2011. Outlier SNPs show more genetic structure between two Bay of Fundy metapopulations of Atlantic salmon than do neutral SNPs. 11(s1), pp.254–267.

Frichot, E. et al., 2014. Fast and efficient estimation of individual ancestry coefficients. *Genetics*, 196(4), pp.973–983.

Furlan, E. et al., 2012. Small population size and extremely low levels of genetic diversity in island populations of the platypus, *Ornithorhynchus anatinus*. *Ecology and evolution*, 2(4), pp.844–857.

Gatesy, J., Springer, M.S. 2014. Phylogenetic analysis at deep timescales: Unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Molecular Phylogenetics and Evolution, 80, pp. 231–266.

Ge, X.-J. et al., 2005. Genetic diversity and geographic differentiation in endangered *Ammopiptanthus* (Leguminosae) populations in desert regions of northwest China as revealed by ISSR analysis. *Annals of Botany*, 95(5), pp.843–851.

Glenny, W.R., Runyon, J.B. & Burkle, L.A., 2018. Drought and increased CO2 alter floral visual and olfactory traits with context-dependent effects on pollinator visitation. *The New Phytologist*, 220(3), pp.785–798.

Goldstein, P.Z. et al., 2000. Conservation Genetics at the Species Boundary. *Conservation Biology*, 14(1), pp.120–131.

Gong, Y.-Q. et al., 2015. The Historical Demography and Genetic Variation of the Endangered *Cycas multipinnata* (Cycadaceae) in the Red River Region, Examined by Chloroplast DNA Sequences and Microsatellite Markers R. Papa, ed. *PLoS ONE*, 10(2), pp.1–18.

González-Astorga, J. et al., 2005. Allozyme variation in the three extant populations of the narrowly endemic cycad *Dioon angustifolium* Miq. (Zamiaceae) from North-eastern Mexico. *Annals of Botany*, 95(6), pp.999–1007.

González-Astorga, J., Vergara-Silva, F., et al., 2008. Diversity and genetic structure of three species of *Dioon* Lindl. (Zamiaceae, Cycadales) from the Pacific seaboard of Mexico. *Biological Journal of the Linnean Society*, 94(4), pp.765–776.

González-Astorga, J., Vovides, A.P. & Iglesias, C., 2003. Morphological and geographic variation of the cycad *Dioon edule* Lindl. (Zamiaceae): ecological and evolutionary implications. *Botanical Journal of the Linnean Society*, 141(4), pp.465–470.

González-Astorga, J., Vovides, A.P., et al., 2008. Diversity and genetic structure of the endangered cycad *Dioon sonorense* (Zamiaceae) from Sonora, Mexico: Evolutionary and conservation implications. *Biochemical Systematics and Ecology*, 36(12), pp.891–899.

Griffith, M.P. et al., 2015. Can a Botanic Garden Cycad Collection Capture the Genetic Diversity in a Wild Population? *International Journal of Plant Sciences*, 176(1), pp.1–10.

Griffith, M.P., Magellan, T.M. & Tomlinson, P.B., 2014. Variation in Leaflet Structure in *Cycas* (Cycadales: Cycadaceae): Does Anatomy Follow Phylogeny and Geography? *International Journal of Plant Sciences*, 175(2), pp.241–255.

Griffiths, K.E. et al., 2014. Maximizing the phylogenetic diversity of seed banks. *Conservation Biology*, 29(2), pp.370–381.

Griffith, M. P., Calonje, M. A., Stevenson, D. W., Husby, C. E., & Little, D. P. 2012. Time, place, and relationships: cycad phenology in a phylogenetic and biogeographic context. *Memoirs of the New York Botanical Garden*, 106, 59-81.

Groom, M.J., Meffe, G.K. & Carroll, C.R., 2006. *Principles of conservation biology*, Sinauer Associates Sunderland.

Grudinski, M., Wanntorp, L., Pannell, C. M., & Muellner-Riehl, A. N. 2014. West to east dispersal in a widespread animal-dispersed woody angiosperm genus (*Aglaia*, Meliaceae) across the Indo-Australian Archipelago. *Journal of biogeography*, 41(6), 1149-1159.

Guan, R. et al., 2016. Draft genome of the living fossil Ginkgo biloba. *GigaScience*, 5(1), p.49.

Guo, W. et al., 2016. Ginkgo and Welwitschia Mitogenomes Reveal Extreme Contrasts in Gymnosperm Mitochondrial Evolution. *Molecular biology and evolution*, 33(6), pp.1448–1460.

Guo, Y. et al., 2014. An improved 2b-RAD approach (I2b-RAD) offering genotyping tested by a rice (*Oryza sativa* L.) F2 population. *Bmc Genomics*, 15(1), pp.956–956.

Hall, J.A. & Walter, G.H., 2018. Pollination of the Australian cycad *Cycas ophiolitica* (Cycadaceae): the limited role of wind pollination in a cycad with beetle pollinator mutualists, and its ecological significance. *Journal of Tropical Ecology*, 34(02), pp.121–134.

Hall, J.A. & Walter, G.H., 2014. Relative Seed and Fruit Toxicity of the Australian Cycads *Macrozamia miquelii* and *Cycas ophiolitica*: Further Evidence for a Megafaunal Seed Dispersal Syndrome in Cycads, and Its Possible Antiquity. *Journal of Chemical Ecology*, 40(8), pp.860–868.

Hall, J.A. & Walter, G.H., 2013. Seed dispersal of the Australian cycad *Macrozamia miquelii* (Zamiaceae): are cycads megafauna-dispersed “grove forming” plants? *American Journal of Botany*, 100(6), pp.1127–1136.

Hall, J. A., Walter, G. H., Bergstrom, D. M., & Machin, P. 2004. Pollination ecology of the Australian cycad *Lepidozamia peroffskyana* (Zamiaceae). *Australian Journal of Botany*, 52(3), 333-343.

Hall, P., Walker, S., & Bawa, K. 1996. Effect of forest fragmentation on genetic diversity and mating system in a tropical tree, *Pithecellobium elegans*. *Conservation Biology*, 10(3), 757-768.

Hall, W.T. & McGavin, M.D., 1968. Clinical and neuropathological changes in cattle eating the leaves of *Macrozamia lucida* or *Bowenia serrulata* (Family Zamiaceae). *Pathologia veterinaria*, 5(1), pp.26–34.

Hamilton, M.B., 1994. Ex Situ Conservation of Wild Plant Species: Time to Reassess the Genetic Assumptions and Implications of Seed Banks. *Conservation Biology*, 8(1), pp.39–49.

Harrisson, K.A. et al., 2014. Using genomics to characterize evolutionary potential for conservation of wild populations. *Evolutionary Applications*, 7(9), pp.1008–1025.

Hart, C.W.M., 2015. The Tiwi of Melville and Bathurst islands. *Oceania*, 1(2), pp.167–180.

Hedrick, P.W., 2001. Conservation genetics: where are we now? *Trends in Ecology & Evolution*, 16(11), pp.629–636.

Hedrick, P.W. & Kalinowski, S.T., 2000. Inbreeding depression in conservation biology. *Annual review of ecology and systematics*, pp.139–162.

Hedrick, P.W. & Miller, P.S., 1992. Conservation Genetics: Techniques and Fundamentals. *Ecological Applications*, 2(1), pp.30–46.

Hellberg, M.E., 1994. Relationships Between Inferred Levels of Gene Flow and Geographic Distance in a Philopatric Coral, *Balanophyllia Elegans*. Evolution, 48(6), pp.1829–1854.

Hefley, T.J. et al., 2016. When can the cause of a population decline be determined? *Ecology letters*, 19(11), pp.1353–1362.

Heywood, V.H. & Iriondo, J.M., 2003. Plant conservation: old problems, new perspectives. *Biological conservation*, 113(3), pp.321–335.

Hill, K.D., 1994. The Cycas rumphii complex (Cycadaceae) in New Guinea and the western Pacific. Australian Systematic Botany, 7(6), pp.543–567.

Hill, K.D., 1996. A taxonomic revision of the genus *Cycas* (Cycadaceae) in Australia. *Telopea*, 7(1), pp.1–64.

Hill, K.D. et al., 2003. The families and genera of cycads: A molecular phylogenetic analysis of *Cycadophyta* based on nuclear and plastid DNA sequences. *International Journal of Plant Sciences*, 164(6), p.933.

Hill, K. D., Walters, T., & Osborne, R. 2004. Character evolution, species recognition and classification concepts in the Cycadaceae. *Cycad classification, concepts and recommendations*, 23-44.

Honnay, O., Verheyen, K., Butaye, J., Jacquemyn, H., Bossuyt, B., & Hermy, M. 2002. Possible effects of habitat fragmentation and climate change on the range of forest plant species. *Ecology Letters*, 5(4), 525-530.

Hughes, A. R., Inouye, B. D., Johnson, M. T., Underwood, N., & Vellend, M. 2008. Ecological consequences of genetic diversity. *Ecology letters*, 11(6), 609-623.

Hou, L. et al., 2018. Genetic Evaluation of Natural Populations of the Endangered Conifer *Thuja koraiensis* Using Microsatellite Markers by Restriction-Associated DNA Sequencing. *Genes*, 9(4), p.218.

Hou, Y., Nowak, M.D., Mirré, V., Bjorå, C.S., Brochmann, C. and Popp, M. 2015. Thousands of RAD-seq Loci Fully Resolve the Phylogeny of the Highly Disjunct Arctic-Alpine Genus Diapensia (Diapensiaceae). PLoS ONE, 10(10), p. e0140175.

Howard, F.W. et al., 1999. *Aulacaspis yasumatsui* (Hemiptera: Sternorrhyncha: Diaspididae), a Scale Insect Pest of Cycads Recently Introduced into Florida. (1), p.14.

Huang, S. et al., 2004. Patterns of Genetic Variation and Demography of *Cycas taitungensis* in Taiwan. *The Botanical Review*, 70(1), 86-92.

Hudjashov, G. et al., 2007. Revealing the prehistoric settlement of Australia by Y chromosome and mtDNA analysis. *Proceedings of the National Academy of Sciences*, 104(21), pp.8726–8730.

Hurka, H., 1994. Conservation genetics and the role of botanical gardens. In *Conservation Genetics*. Basel: Birkhäuser Basel, pp. 371–380.

Hynes, R. A., & Chase, A. K. 1982. Plants, sites and domiculture: Aboriginal influence upon plant communities in Cape York Peninsula. *Archaeology in Oceania*, 17(1), 38-50.

Ingham, J.A. et al., 2013. Ancient relicts or recent dispersal: how long have cycads been in central Australia? *Diversity and Distributions*, 19(3), pp.307–316.

IUCN, 2016. The IUCN Red List of Threatened Species. *iucnredlist.org*. Available at: http://www.iucnredlist.org/about/summary-statistics [Accessed July 10, 2015].

James, H.E. et al., 2018. Conservation genetics and demographic analysis of the endangered cycad species *Cycas megacarpa* and the impacts of past habitat fragmentation. *Australian Journal of Botany*, 66(2), p.173.

Jansen, R.K. et al., 2011. Complete plastid genome sequences of three Rosids (Castanea, Prunus, Theobroma): evidence for at least two independent transfers of rpl22 to the nucleus. *Molecular biology and evolution*, 28(1), pp.835–847.

Jelinski, D.E., 1997. On genes and geography: a landscape perspective on genetic variation in natural plant populations. *Landscape and Urban Planning*, 39(1), pp.11–23.

Jombart, T. & Ahmed, I., 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, 27(21), pp.3070–3071.

Jones, D.L., 2002. *Cycads of the world* 2nd ed., Washington, D.C.: Smithsonian Institution Press.

Karam, M.-J., Lefèvre, F., Dagher-Kharrat, M.B., Pinosio, S. and Vendramin, G.G. 2015. Genomic exploration and molecular marker development in a large and complex conifer genome using RADseq and mRNAseq. *Molecular Ecology Resources*, 15(3), pp. 601–612.

Kearse, M. et al., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28(12), pp.1647–1649.

Keenan, K., McGinnity, P., Cross, T. F., Crozier, W. W., & Prodöhl, P. A. 2013. diveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods in Ecology and Evolution*, 4(8), 782-788.

Keppel, G., Hodgskiss, P.D. & Plunkett, G.M., 2008. Cycads in the insular South-west Pacific: dispersal or vicariance? *Journal of Biogeography*, 35(6), pp.1004–1015.

Keppel, G., 2002. Low genetic variation in a Pacific cycad: conservation concerns for *Cycas seemannii* (Cycadaceae). *Oryx*, 36(01), pp.41–49.

Keppel, G., Lee, S.-W. & Hodgskiss, P.D., 2002. Evidence for long isolation among populations of a Pacific cycad: genetic diversity and differentiation in *Cycas seemannii* A. Br.(Cycadaceae). *Journal of Heredity*, 93(2), pp.133–139.

Keppel, G., 2001. Notes on the Natural History of *Cycas seemannii* (Cycadaceae). The South Pacific. *Journal of Natural and Applied Sciences*, 19(1), pp.35–41.

Kessler, L., 2005. The Cult of the Cycads. *nytimes.com*. Available at: http://www.nytimes.com/2005/08/28/magazine/28CYCADS.html?pagewanted=print [Accessed March 14, 2012].

Kimura, M. & King, J.L., 1979. Fixation of a deleterious allele at one of two “duplicate” loci by mutation pressure and random drift. *Proceedings of the National Academy of Sciences*, 76(6), pp.2858–2861.

Kleinhans, C. & Willows-Munro, S., 2019. Low genetic diversity and shallow population structure in the endangered vulture, *Gyps coprotheres*. *Scientific reports*, 9(1), p.5536.

Kono, M. & Tobe, H., 2007. Is *Cycas revoluta* (Cycadaceae) Wind- or Insect-Pollinated? *American Journal of Botany* (5), p.847.

Kramer, A.T. & Havens, K., 2009. Plant conservation genetics in a changing world. *Trends in Plant Science*, 14(11), pp.599–607.

Laidlaw, M.J. & Forster, P.I., 2012. Climate predictions accelerate decline for threatened *Macrozamia* cycads from Queensland, Australia. *Biology*, 1(3), pp.880–894.

Latta, R. G., & Mitton, J. B. 1999. Historical separation and present gene flow through a zone of secondary contact in ponderosa pine. *Evolution*, 53(3), 769-776.

Leblois, R., Estoup, A., & Streiff, R. (2006). Genetics of recent habitat contraction and reduction in population size: does isolation by distance matter? *Molecular Ecology*, 15(12), 3601-3615.

Lesica, P. & Allendorf, F.W., 1995. When are peripheral populations valuable for conservation? *Conservation Biology*, 9(4), pp.753–760.

Lexer, C. et al., 2014. Genomics of the divergence continuum in an African plant biodiversity hotspot, I: drivers of population divergence in *Restio capensis* (Restionaceae). *Molecular Ecology*, 23(17), pp.4373–4386.

Li, Y. et al., 2018. Plastomes of nine hornbeams and phylogenetic implications. *Ecology and evolution*, 8(17), pp.8770–8778.

Li, Z.-Z. et al., 2018. Genetic diversity and population structure of the endangered basal angiosperm *Brasenia schreberi* (Cabombaceae) in China. *Peerj*, 6(5), p.e5296.

Liu, J., Zhang, S., Nagalingum, N. S., Chiang, Y. C., Lindstrom, A. J., & Gong, X. (2018). Phylogeny of the gymnosperm genus *Cycas* L. (Cycadaceae) as inferred from plastid and nuclear loci based on a large-scale sampling:

Evolutionary relationships and taxonomical implications. *Molecular Phylogenetics and Evolution*, 127, 87-97.

Liddle, D.R., 2009. Management Program for Cycads in the Northern Territory of Australia. *Northern Territory Department of Natural Resources, Environment, the Arts and Sport, Darwin.*

Lindstrom, A.J. & Hill, K.D., 2007. The genus *Cycas* (Cycadaceae) in India. *Telopea*, 11(4), pp.463–488.

Lister, R., Gregory, B.D. & Ecker, J.R., 2009. Next is now: new technologies for sequencing of genomes, transcriptomes, and beyond. *Current opinion in plant biology*, 12(2), pp.107–118.

Liu, J., Zhang, S., Nagalingum, N. S., Chiang, Y. C., Lindstrom, A. J., & Gong, X. 2018. Phylogeny of the gymnosperm genus *Cycas* L. (Cycadaceae) as inferred from plastid and nuclear loci based on a large-scale sampling: Evolutionary relationships and taxonomical implications. *Molecular Phylogenetics and Evolution*, 127, 87-97.

Liu, S. & Hansen, M.M., 2017. PSMC (pairwise sequentially Markovian coalescent) analysis of RAD (restriction site associated DNA) sequencing data., 17(4), pp.631–641.

Liu, L. et al., 2015. Phylogeny of *Morella rubra* and Its Relatives (Myricaceae) and Genetic Resources of Chinese Bayberry Using RAD Sequencing G. Sun, ed. *PLoS ONE*, 10(10), p.e0139840.

Liu, L., Yu, L., Edwards, S.V. 2010. A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. BMC Evolutionary Biology, 10(1), pp. 302.

Liu, L., Yu, L., Kubatko, L., Pearl, D.K., Edwards, S.V. 2009. Coalescent methods for estimating phylogenetic trees. Molecular Phylogenetics and Evolution, 53(1), pp. 320–328.

Long-Qian, X. & Xun, G., 2006. Genetic differentiation and relationships of populations in the *Cycas balansae* complex (Cycadaceae) and its conservation complications. *Annals of Botany*, 97(5), pp.807–812.

Long-Qian, X. et al., 2004. ISSR Variation in the Endemic and Endangered Plant *Cycas guizhouensis* (Cycadaceae). *Annals of Botany*, 94(1), pp.133–138.

Lönn, M. & Prentice, H.C., 2002. Gene diversity and demographic turnover in central and peripheral populations of the perennial herb *Gypsophila fastigiata*. *Oikos*, 99(3), pp.489–498.

Lowe, P.R., 2008. Species and Subspecies. *Ibis*, 64(1), pp.179–185.

Lowry, D. B., Hoban, S., Kelley, J. L., Lotterhos, K. E., Reed, L. K., Antolin, M. F., & Storfer, A. 2017. Breaking RAD: An evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. *Molecular ecology resources*, 17(2), 142-152.

Lozier, J. D. 2014. Revisiting comparisons of genetic diversity in stable and declining species: assessing genome-wide polymorphism in North American bumblebees using RAD sequencing. Molecular Ecology, 23(4), pp.788-801.

Macnair, M. R. 1989. The potential for rapid speciation in plants. *Genome,* 31(1), 203-210.

Mankga, L.T. & Yessoufou, K., 2017. Factors driving the global decline of cycad diversity. *AoB PLANTS*, 9(4), p.plx022.

Manel, S., Schwartz, M. K., Luikart, G., & Taberlet, P. 2003. Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution*, 18(4), 189-197.

Marler, T. & Muniappan, R., 2006. Pests of *Cycas micronesica* leaf, stem, and male reproductive tissues with notes on current threat status. *MICRONESICA-AGANA-*, 39(1), p.1.

Marler, T.E. & Niklas, K.J., 2011. Reproductive effort and success of *Cycas micronesica* KD Hill are affected by habitat. *International Journal of Plant Sciences*, 172(5), pp.700–706.

Markert, J.A. et al., 2010. Population genetic diversity and fitness in multiple environments. *BMC Evolutionary Biology*, 10(1), p.205.

Martínez, L.C.A., Artabe, A.E.E. & Bodnar, J., 2012. A new cycad stem from the Cretaceous in Argentina and its phylogenetic relationships with other Cycadales. *Botanical Journal of the Linnean Society*, 170(3), pp.436–458.

Massatti, R., Reznicek, A.A. & Knowles, L.L., 2016. Utilizing RADseq data for phylogenetic analysis of challenging taxonomic groups: A case study in Carex sect. Racemosae. *American Journal of Botany*, 103(2), pp.337–347.

Mastretta-Yanes, A. et al., 2015. Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. *Molecular Ecology Resources*, 15(1), pp.28–41.

Matocq, M.D. & Villablanca, F.X., 2001. Low genetic diversity in an endangered species: recent or historic pattern? *Biological conservation*, 98(1), pp.61–68.

McCartney-Melstad, E., Gidiş, M. & Shaffer, H.B., 2017. Population Genomics of the Foothill Yellow-Legged Frog (*Rana boylii*) and RADseq Parameter Choice for Large-Genome Organisms.

McHenry, M.A. & Barrington, D.S., 2014. Phylogeny and biogeography of exindusiate *Andean Polystichum* (Dryopteridaceae). *American Journal of Botany*, 101(2), pp.365–375.

McKinney, G.J. et al., 2016. RADseq provides unprecedented insights into molecular ecology and evolutionary genetics: comment on Breaking RAD by Lowry et al. (2016)., 17(3), pp.356–361.

McMahon, B.J., Teeling, E.C. & Höglund, J., 2014. How and why should we implement genomics into conservation? *Evolutionary Applications*, 7(9), pp.999–1007.

Meerow, A.W. & Nakamura, K., 2007. Ten microsatellite loci from *Zamia integrifolia* (Zamiaceae). *Molecular Ecology Notes*, 7(5), pp.824–826.

Meerow, A.W. et al., 2012. *Zamia* (Cycadales: Zamiaceae) on Puerto Rico: asymmetric genetic differentiation and the hypothesis of multiple introductions. *American Journal of Botany*, 99(11), pp.1828–1839.

Meyers, B.C., Scalabrin, S. & Morgante, M., 2004. Mapping and sequencing complex genomes: let's get physical! *Nature Reviews. Genetics*, 5(8), pp.578–588.

Mickle, J.E. et al., 2011. Scanning electron microscopy studies of cuticle micromorphology in *Cycas* L.(Cycadaceae). *Plant Biosystems - An International Journal Dealing with all Aspects of Plant Biology*, 145(1), pp.191–201.

Mound, L. A., & Terry, I. 2001. Thrips pollination of the central Australian cycad, *Macrozamia macdonnellii* (Cycadales). *International Journal of Plant Sciences*, 162(1), 147-154.

Migdałek, G. et al., 2017. No evidence of contemporary interploidy gene flow between the closely related European woodland violets *Viola reichenbachiana* and *V. riviniana* (sect. *Viola*, Violaceae). J. Keurentjes, ed. *Plant Biology*, 19(4), pp.542–551.

Mirarab, S., Reaz, R., Bayzid, M.S., Zimmermann, T., Swenson, M.S., Warnow, T. 2014. ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics, 30(17), pp. 541–548.

Mirarab, S., Warnow, T. 2015. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics, 31(12), i44-i52.

Mondoni, A. et al., 2011. Seeds of alpine plants are short lived: implications for long-term conservation. *Annals of Botany*, 107(1), pp.171–179.

Morjan, C. L., & Rieseberg, L. H. 2004. How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. *Molecular Ecology*, 13(6), 1341-1356.

Muir, G. & Schlötterer, C., 2004. Evidence for shared ancestral polymorphism rather than recurrent gene flow at microsatellite loci differentiating two hybridizing oaks (*Quercus* spp.). *Molecular Ecology*, 14(2), pp.549–561.

Muniappan, R. et al., 2012. Cycad Aulacaspis Scale, a Newly Introduced Insect Pest in Indonesia. *HAYATI Journal of Biosciences*, 19(3), pp.110–114.

Muriira, N.G. et al., 2018. Genetic Diversity Analysis Reveals Genetic Differentiation and Strong Population Structure in *Calotropis* Plants. *Scientific reports*, 8(1), p.7832.

Nadarajan, J. et al., 2018. Comparative Biology of Cycad Pollen, Seed and Tissue - A Plant Conservation Perspective. *The Botanical Review*, 34(11), pp.20–20.

Nagalingum, N.S., Marshall, C.R., Quental, T.B., Rai, H.S., Little, D.P. and Mathews, S. 2011. Recent synchronous radiation of a living fossil. *Science,* 334(6057), pp. 796–799.

Namroud, M. C., Beaulieu, J., Juge, N., Laroche, J., & Bousquet, J. 2008. Scanning the genome for gene single nucleotide polymorphisms involved in adaptive population differentiation in white spruce. *Molecular Ecology*, 17(16), 3599-3613.

Narum, S.R., Buerkle, C.A., Davey, J.W., Miller, M.R. and Hohenlohe, P.A. 2013. Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology*, 22 (11), pp. 2841–2847.

Nazareno, A.G., Dick, C.W. & Lohmann, L.G., 2018. Tangled banks: A landscape genomic evaluation of Wallace's Riverine barrier hypothesis for three Amazon plant species. *Molecular Ecology*, 76(5), p.1-18

Nazareno, A. G., Bemmels, J. B., Dick, C. W., & Lohmann, L. G. 2017. Minimum sample sizes for population genomics: an empirical study from an Amazonian plant species. *Molecular Ecology Resources*, 17(6), 1136-1147.

Nei, M., & Roychoudhury, A. K. 1974. Sampling variances of heterozygosity and genetic distance. *Genetics*, 76(2), 379-390.

Neri, J., Wendt, T. & Palma-Silva, C., 2018. Natural hybridization and genetic and morphological variation between two epiphytic bromeliads. *AoB PLANTS*, 10(1), p.plx061.

Newbold, T. et al., 2016. Has land use pushed terrestrial biodiversity beyond the planetary boundary? A global assessment. *Science*, 353(6296), pp.288–291.

Newman, D. & Tallmon, D.A., 2001. Experimental Evidence for Beneficial Fitness Effects of Gene Flow in Recently Isolated Populations. *Conservation biology: the journal of the Society for Conservation Biology*, 15(4), pp.1054–1063.

Niissalo, M. A., Leong-Škorničková, J., Khew, G. S., & Webb, E. L. 2017. Very small relict populations suggest high extinction debt of gingers in primary forest fragments of a tropical city. *American journal of botany*, 104(1), 182-189.

Nikitsky Botanical Gardens – National Scientific Centre of RAS et al., 2017. The Role of Botanical Gardens in the Conservation of the Family Cactaceae juss. Species. *Tambov University Reports. Series: Natural and Technical Sciences*, 22(5-1), pp.828–832.

Norstog, K. & Nicholls, T.J., 1997. *The biology of the cycads*, Ithaca, N.Y.: Comstock Pub. Associates.

Norstog, K.J. & Fawcett, P.K.S., 1989. Insect-Cycad Symbiosis and its Relation to the Pollination of Zamia furfuracea (Zamiaceae) by Rhopalotria mollis (Curculionidae). *American Journal of Botany*, 76(9), p.1380.

Norstog, K. J., Stevenson, D. W., & Niklas, K. J. 1986. The role of beetles in the pollination of *Zamia furfuracea* L. fil.(Zamiaceae). *Biotropica*, 300-306.

Nowak, J.J. et al., PopR: Software for Wildlife Managers. *arc.lib.montana.edu*

Oakley, C.G. et al., 2019. Heterosis is common and inbreeding depression absent in natural populations of Arabidopsis thaliana. *Journal of evolutionary biology*, 32(6), pp.592–603.

Octavio-Aguilar, P., González-Astorga, J. & Vovides, A.P., 2009. Genetic diversity through life history of *Dioon edule* Lindley (Zamiaceae, Cycadales). *Plant Biology*, 11(4), pp.525–536.

Oney-Birol, S. et al., 2018. Assessment of shared alleles in drought-associated candidate genes among southern California white oak species (*Quercus* sect. Quercus). *BMC genetics*, 19(1), p.88.

Oostermeijer, J. G. B., Luijten, S. H., & Den Nijs, J. C. M. 2003. Integrating demographic and genetic approaches in plant conservation. *Biological Conservation*, 113(3), 389-398.

Ornduff, R., 1991. Size Classes, Reproductive Behavior, and Insect Associates of *Cycas media* (Cycadaceae) in Australia. (2), p.203.

Ornduff, R., 1992. Features of Coning and Foliar Phenology, Size Classes, and Insect Associates of *Cycas armstrongii* (Cycadaceae) in the Northern Territory, Australia. *Bulletin of the Torrey Botanical Club*, 119(1), p.39.

Osborne, R. et al., 2012. The World List of Cycads. Bronx NY, New York Botanical Garden Press, pp. 480–508.

Ouborg, N.J. et al., 2010. Conservation genetics in transition to conservation genomics. *Trends in Genetics*, 26(4), pp.177–187.

Pauls, S. U., Nowak, C., Bálint, M., & Pfenninger, M. 2013. The impact of global climate change on genetic diversity within populations and species. *Molecular Ecology*, 22(4), 925-946.

Paradis, E. et al., Schliep, K., 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinfromatics*, 20(2), pp. 289–290.

Paradis, E. 2010. pegas: an R package for population genetics with an integrated–modular approach. Bioinformatics, 26(3), 419-420.

Parlato, E.H. & Armstrong, D.P., 2018. Predicting reintroduction outcomes for highly vulnerable species that do not currently coexist with their key threats. *Conservation Biology*, 74(Suppl 2), p.160.

Paun, O., Turner, B., Trucchi, E., Munzinger, J., Chase, M.W. and Samuel, R. 2015. Processes Driving the Adaptive Radiation of a Tropical Tree (*Diospyros*, Ebenaceae) in New Caledonia, a Biodiversity Hotspot. *Systematic Biology*, 65(2), pp. 212–27.

Paz-Vinas, I. et al., 2018. Systematic conservation planning for intraspecific genetic diversity. *Proceedings. Biological sciences*, 285(1877), p.20172746.

Peakall, R. & Smouse, P.E., 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. *Bioinformatics*, 28(19), pp.2537–2539.

Pekkala, N. et al., 2014. The effect of inbreeding rate on fitness, inbreeding depression and heterosis over a range of inbreeding coefficients. *Evolutionary Applications*, 7(9), pp.1107–1119.

Pellicer, J. et al., 2018. Genome Size Diversity and Its Impact on the Evolution of Land Plants. *Genes*, 9(2), p.88.

Pellicer, J., FAY, M.F. & Leitch, I.J., 2010. The largest eukaryotic genome of them all? *Botanical Journal of the Linnean Society*, 164(1), pp.10–15.

Peterson, B.K. et al., 2012. Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species L. Orlando, ed. *PLoS ONE*, 7(5), pp.1–11.

Pérez-Farrera, M.A. et al., 2016. Anatomy and morphology suggest a hybrid origin of *Zamia katzeriana* (Zamiaceae). *Phytotaxa*, 270(3), p.161.

Pérez-Farrera, M.A. et al., 2006. Demography of the cycad *Ceratozamia mirandae* (Zamiaceae) under disturbed and undisturbed conditions in a biosphere reserve of Mexico. *Plant Ecology*, 187(1), pp.97–108.

Pinares, A. et al., 2009. Genetic diversity of the endangered endemic *Microcycas calocoma* (Miq.) A. DC (Zamiaceae, Cycadales): Implications for conservation. *Biochemical Systematics and Ecology*, 37(4), pp.385–394.

Plenk, K., Bardy, K., Höhn, M., & Kropf, M. 2019. Long-term survival and successful conservation? Low genetic diversity but no evidence for reduced reproductive success at the north-westernmost range edge of *Poa badensis* (Poaceae) in Central Europe. *Biodiversity and conservation*, 28(5), 1245-1265.

Plomion, C., Bartholomé, J., Lesur, I., Boury, C., Rodríguez-Quilón, I., Lagraulet, H., ... & de Miguel, M. 2016. High-density SNP assay development for genetic analysis in maritime pine (*Pinus pinaster*). *Molecular Ecology Resources*, 16(2), 574-587.

Preece, L.D., Duguid, A.W. & Albrecht, D.E., 2007. Environmental determinants of a restricted cycad in central Australia, *Macrozamia macdonnellii*. *Australian Journal of Botany*, 55(6), p.601.

Pritchard, J. K., Stephens, M., & Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics*, 155(2), 945-959.

Procheş, Ş. & Johnson, S.D., 2009. Beetle pollination of the fruit-scented cones of the South African cycad Stangeria eriopus. *American Journal of Botany*, 96(9), pp.1722–1730.

Proćków, M. et al., 2017. Ongoing Speciation and Gene Flow between Taxonomically Challenging *Trochulus* Species Complex (Gastropoda: Hygromiidae). R. Castiglia, ed. *PLoS ONE*, 12(1), p.e0170460.

Pujol, B. et al., 2017. Mountain landscape connectivity and subspecies appurtenance shape genetic differentiation in natural plant populations of the snapdragon (*Antirrhinum majus* L.). *Botany Letters*, 164(2), pp.111–119.

Puritz, J.B. et al., 2014. Demystifying the RAD fad. *Molecular Ecology*, 23(24), pp.5937–5942.

R Core Team, 2019. R: A Language and Environment for Statistical Computing, version 3.4.4. Vienna, Austria: R Foundation for Statistical Computing.

Rai, H.S. et al., 2003. Inference of higher-order relationships in the cycads from a large chloroplast data set. *Molecular Phylogenetics and Evolution*, - 29(- 2), pp.–359.

Raimondo, D.C. & Donaldson, J.S., 2003. Responses of cycads with different life histories to the impact of plant collecting: simulation models to determine important life history stages and population recovery times. *Biological conservation*, 111(3), pp.345–358.

Reed, T.E., Jenouvrier, S. & Visser, M.E., 2013. Phenological mismatch strongly affects individual fitness but not population demography in a woodland passerine A. Roulin, ed. *The Journal of Animal Ecology*, 82(1), pp.131–144.

Rieseberg, L.H., 2007. Genetics of Speciation. *Journal of Heredity*, 98(2), pp.101–102.

Rieseberg, L. H., & Willis, J. H. 2007. Plant speciation. *Science*, 317(5840), 910-914.

Roda, F., Ambrose, L., Walter, G.M., Liu, H.L., Schaul, A., Lowe, A., Pelser, P.B., Prentis, P., Rieseberg, L.H. and Ortiz-Barrientos, D. 2013. Genomic evidence for the parallel evolution of coastal forms in the Senecio lautus complex. *Molecular Ecology*, 22(11), pp. 2941–2952.

Roch, S., Warnow, T. 2015. On the robustness to gene tree estimation error (or lack thereof) of coalescent-based species tree methods. *Systematic Biology*, 64(4), pp. 663–676.

Rodríguez-Quilón, I. et al., 2016. Capturing neutral and adaptive genetic diversity for conservation in a highly structured tree species. *Ecological applications: a publication of the Ecological Society of America*, 26(7), pp.2254–2266.

Rodríguez-Rodríguez, P., Pérez de Paz, P.L. & Sosa, P.A., 2018. Species delimitation and conservation genetics of the Canarian endemic *Bethencourtia* (Asteraceae). *Genetica*, 146(2), pp.199–210.

Roodt, D., Lohaus, R., Sterck, L., Swanepoel, R.L., Van de Peer, Y. and Mizrachi, E. 2017. Evidence for an ancient whole genome duplication in the cycad lineage. PLoS ONE, 12(9), p. e0184454.

Rosche, C. et al., 2018. Sex ratio rather than population size affects genetic diversity in Antennaria dioica. J. Keurentjes, ed. *Plant Biology*, 20(4), pp.789–796.

Rousset, F. 1997. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics*, 145(4), 1219-1228.

Rubin, B.E.R., Ree, R.H. & Moreau, C.S., 2012. Inferring Phylogenies from RAD Sequence Data S.-O. Kolokotronis, ed. *PLoS ONE*, 7(4), pp.1–12.

Ruhsam, M., Rai, H.S., Mathews, S., Ross, T.G., Graham, S.W., Raubeson, L.A., Mei, W., Thomas, P.I., Gardner, M.F., Ennos, R.A. and Hollingsworth, P.M. 2015. Does complete plastid genome sequencing improve species discrimination and phylogenetic resolution in *Araucaria*? *Molecular Ecology Resources*, 15(5), pp. 1067–1078.

Said Gutiérrez-Ortega, J. et al., 2018. Corrigendum: Aridification as a driver of biodiversity: a case study for the cycad genus *Dioon* (Zamiaceae). *Annals of Botany*, 121(1), pp.193–193.

Salas-Leiva, D.E., Meerow, A.W., Francisco-Ortega, J., Calonje, M., Griffith, M.P., Stevenson, D.W. and Nakamura, K. 2014. Conserved genetic regions across angiosperms as tools to develop single-copy nuclear markers in gymnosperms: an example using cycads. *Molecular Ecology* Resources, 14(4), pp. 831–845.

Salas-Leiva D.E., Meerow A.W., Calonje M., Griffith M. P., Francisco-Ortega J., Nakamura K., Stevenson D. W., Lewis C. E., Namoff S. 2013. Phylogeny of the cycads based on multiple single-copy nuclear genes: congruence of concatenated parsimony, likelihood and species tree inference methods. *Annals of Botany*, 112(7), pp. 1263–1278.

Sampson, J.F. et al., 2016. Limiting inbreeding in disjunct and isolated populations of a woody shrub. *Ecology and evolution*, 6(16), pp.5867–5880.

Sangin, P. et al., 2006. Molecular identification of *Cycas* by restriction fragment length polymorphism (RFLP) and random amplified polymorphic DNA (RAPD). *Kasetsarrt J.(Nat Sci.)*, 40, pp.107–120.

Sax, GainesBrown, 2002. Species Invasions Exceed Extinctions on Islands Worldwide: A Comparative Study of Plants and Birds. *The American Naturalist*, 160(6), p.766.

Scaglione, D. et al., 2012. RAD tag sequencing as a source of SNP markers in *Cynara cardunculus* L. *Bmc Genomics*, 13(1), p.3.

Scheele, B.C. et al., 2018. How to improve threatened species management: An Australian perspective. *Journal of environmental management*, 223, pp.668–675.

Schield, D.R. et al., 2015. Incipient speciation with biased gene flow between two lineages of the Western Diamondback Rattlesnake (*Crotalus atrox*). *Molecular Phylogenetics and Evolution*, 83, pp.213–223.

Schmieder, R. & Edwards, R., 2011. Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, 27(6), pp.863–864.

Schmoldt, A., Benthe, H.F. & Haberland, G., 1975. Digitoxin metabolism by rat liver microsomes. *Biochemical pharmacology*, 24(17), pp.1639–1641.

Schmutzer, T. et al., 2015. Species-wide genome sequence and nucleotide polymorphisms from the model allopolyploid plant *Brassica napus*. *Scientific data*, 2, p.150072.

Schneider, D. et al., 2002. Cycads: their evolution, toxins, herbivores and insect pollinators. *Naturwissenschaften*, 89(7), pp.281–294.

Schwartz, M.W., 2003. Assessing Population Viability in Long-Lived Plants. In *Population Viability in Plants*. Ecological Studies. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 239–266.

Scott, A.D., Stenz, N.W.M., Ingvarsson, P.K. and Baum, D.A. 2016. Whole genome duplication in coast redwood (*Sequoia sempervirens*) and its implications for explaining the rarity of polyploidy in conifers. *The New Phytologist*, 211(1), pp. 186–193.

Segar, S.T. et al., 2017. Speciation in a keystone plant genus is driven by elevation: a case study in New Guinean *Ficus*. *Journal of evolutionary biology*, 30(3), pp.512–523.

Séré, M. et al., 2017. Comparison of different genetic distances to test isolation by distance between populations. *Heredity*, 119(2), pp.55–63.

Setterfield, S. A. 2002. Seedling establishment in an Australian tropical savanna: effects of seed supply, soil disturbance and fire. *Journal of Applied Ecology*, 39(6), 949-959.

Shafer, A.B.A. et al., 2016. Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. *Methods in Ecology and Evolution*, 8, pp.907–917.

Sharma, I.K. et al., 1998. The extent and structure of genetic variation in the *Macrozamia pauli-guilielmi* complex (Zamiaceae). *Biochemical Systematics and Ecology*, 26(1), pp.45–54.

Sharma, I.K., Jones, D.L. & Forster, P.I., 1999. Contribution of isozymic analysis in differentiating *Macrozamia moorei* and KD Hill &M. johnsoniiF. Muell (Zamiaceae). *Austrobaileya*, 5(2), pp.363–365.

Sharma, I.K., Jones, D.L. & Forster, P.I., 2004. Genetic differentiation and phenetic relatedness among seven species of the Macrozamia plurinervia complex (Zamiaceae). Biochemical Systematics and Ecology, 32(3), pp.313–327.

Shiga, T. et al., 2017. Genetic diversity and population structure of *Nuphar submersa* (Nymphaeaceae), a critically endangered aquatic plant endemic to Japan, and implications for its conservation. *Journal of plant research*, 130(1), pp.83–93.

Shuguang, J., Yang, Z., Nian, L., Zezheng, G., Qiang, W., Zhenhua, X. and Hal, R. 2006. Genetic variation in the endangered endemic species *Cycas fairylakea* (Cycadaceae) in China and implications for conservation. *Biodiversity & Conservation*, 15(5), pp. 1681–1694.

Siemens, J., 2012. Interspecific Hybridisation between Wild Relatives and *Brassica napus* to Introduce New Resistance Traits iinto the Oilseed Rape Gene Pool. *Czech Journal of Genetics and Plant Breeding*, 38(No. 3-4), pp.155–157.

Siol, M. et al., 2008. How multilocus genotypic pattern helps to understand the history of selfing populations: a case study in *Medicago truncatula*. *Heredity*, 100(5), pp.517–525.

Slatkin, M. 1981. Estimating levels of gene flow in natural populations. *Genetics*, 99(2), 323-335.

Sloan, S. et al., 2014. Remaining natural vegetation in the global biodiversity hotspots. *Biological conservation*, 177, pp.12–24.

Smith, M.J. et al., 2011. Assessing the impacts of international trade on CITES-listed species: Current practices and opportunities for scientific research. *Biological conservation*, 144(1), pp.82–91.

Snow, E.L. & Walter, G.H., 2007. Large seeds, extinct vectors and contemporary ecology: testing dispersal in a locally distributed cycad, *Macrozamia lucida* (Cycadales). *Australian Journal of Botany*, 55(6), pp.592–600.

Sousa, V. & Hey, J., 2013. Understanding the origin of species with genome-scale data: modelling gene flow. *Nature Reviews. Genetics*, 14(6), pp.404–414.

Sosa, V., Vovides, A.P. & Castillo-Campos, G., 1998. Monitoring endemic plant extinction in Veracruz, Mexico. *Biodiversity and Conservation*, 7(11), pp.1521–1527.

Sork, V. L., Nason, J., Campbell, D. R., & Fernandez, J. F. 1999. Landscape approaches to historical and contemporary gene flow in plants. *Trends in Ecology & Evolution*, 14(6), 219-224.

Spooner, F.E.B., Pearson, R.G. & Freeman, R., 2018. Rapid warming is associated with population decline among terrestrial birds and mammals globally. *Global change biology*, 24(10), pp.4521–4531.

Stapley, J. et al., 2010. Adaptation genomics: the next generation. *Trends in Ecology & Evolution*, 25(12), pp.705–712.

Stevens, K., Harrisson, K. A., Hogan, F. E., Cooke, R., & Clarke, R. H. 2018. Reduced gene flow in a vulnerable species reflects two centuries of habitat loss and fragmentation. *Ecosphere*, 9(2), e02114.

Stevenson, D.W., 1992. A formal classification of the extant cycads. *Brittonia*, 44(2), pp.220–223.

Storfer, A. 1999. Gene flow and endangered species translocations: a topic revisited. *Biological Conservation*, 87(2), 173-180.

Suinyuy, T.N., Donaldson, J.S. & Johnson, S.D., 2013. Patterns of odour emission, thermogenesis and pollinator activity in cones of an African cycad: what mechanisms apply? *Annals of Botany*, 112(5), pp.891–902.

Suinyuy, T.N., Donaldson, J.S. & Johnson, S.D., 2012. Geographical variation in cone volatile composition among populations of the African cycad Encephalartos villosus. *Biological Journal of the Linnean Society*, 106(3), pp.514–527.

Suinyuy, T.N., Donaldson, J.S. & Johnson, S.D., 2009. Insect pollination in the African cycad *Encephalartos friderici-guilielmii* Lehm. *South African Journal of Botany*, 75(4), pp.682–688.

Swart, C., Donaldson, J. & Barker, N., 2018. Predicting the distribution of *Encephalartos latifrons*, a critically endangered cycad in South Africa. *Biodiversity and Conservation*, 27(8), pp.1961–1980.

Swofford DL. 2003. PAUP*: Phylogenetic analysis using parsimony (*and other methods). Version 4 4.0a158.

Szczecińska, M. et al., 2016. Genetic Diversity and Population Structure of the Rare and Endangered Plant Species *Pulsatilla patens* (L.) Mill in East Central Europe. D. Gomory, ed. *PLoS ONE*, 11(3), p.e0151730.

Tang, W., 1987. Insect pollination in the cycad *Zamia pumila* (Zamiaceae). *American Journal of Botany*, 74(1), pp.90–99.

Terry, I., Tang, W., & Marler, T. E. 2012. Pollination systems of Island Cycads: predictions based on island biogeography. In Proceedings of Cycad 2008. The 8th International Conference on Cycad Biology, Panama City, Panama, 13-15 January 2008 (pp. 102-132). *New York Botanical Garden Press*.

Terry, I., Tang, W., Taylor, A., Singh, R., Vovides, A., & Cibrián Jaramillo, A. (2012). An overview of cycad pollination studies.

Terry, I., Roe, M., Tang, W. & Marler, T. E. 2009. Cone insects and putative pollen vectors of the endangered cycad, *Cycas micronesica*. *Micronesica*, 41(1), 83-99.

Terry, I., Forster, P. I., Moore, C. J., Roemer, R. B., & Machin, P. J. 2008. Demographics, pollination syndrome and conservation status of *Macrozamia platyrhachis* (Zamiaceae), a geographically restricted Queensland cycad. *Australian Journal of Botany*, 56(4), 321-332.

Terry, I., Walter, G. H., Moore, C., Roemer, R., & Hull, C. 2007. Odor-mediated push-pull pollination in cycads. *Science*, 318(5847), 70-70.

Terry, L.I. et al., 2005. Pollination of Australian *Macrozamia* Cycads (Zamiaceae): Effectiveness and Behavior of Specialist Vectors in a Dependent Mutualism. (6), p.931.

Thomas, C.D. et al., 2004. Extinction risk from climate change. *Nature*, 427(6970), pp.145–148.

Tremblay, R. L., & Ackerman, J. D. 2001. Gene flow and effective population size in *Lepanthes* (Orchidaceae): a case for genetic drift. *Biological Journal of the Linnean Society*, 72(1), 47-62.

Toczydlowski, R.H. & Waller, D.M., 2019. Drift happens: Molecular genetic diversity and differentiation among populations of jewelweed (*Impatiens capensis* Meerb.) reflect fragmentation of floodplain forests. *Molecular Ecology*, 28(10), pp.2459–2475.

Toonen, R.J., Puritz, J.B., Forsman, Z.H., Whitney, J.L., Fernandez-Silva, I., Andrews, K.R. and Bird, C.E. 2013. ezRAD: a simplified method for genomic genotyping in non-model organisms. *Peerj*, 1(14), pp. e203–e203.

Torgersen, J.S., 2017. Crime, culture and collecting: the illicit cycad market in South Africa.

Toro, M.A., Villanueva, B. & Fernandez, J., 2014. Genomics applied to management strategies in conservation programmes. *Livestock Science*, 166, pp.48–53.

Tripp, E.A. et al., 2017. RADseq dataset with 90% missing data fully resolves recent radiation of *Petalidium* (Acanthaceae) in the ultra-arid deserts of Namibia. *Ecology and evolution*, 7(19), pp.7920–7936.

Tsudzuki, J. et al., 1992. Chloroplast DNA of black pine retains a residual inverted repeat lacking rRNA genes: nucleotide sequences of trnQ, trnK, psbA, trnI and trnH and the absence of rps16. *Molecular and General Genetics MGG*, 232(2), pp.206–214.

Twyford, A.D. & Friedman, J., 2015. Adaptive divergence in the monkey flower *Mimulus guttatus* is maintained by a chromosomal inversion. *Evolution*, 69(6), pp.1476–1486.

van der Bank, H. et al., 2001. Allozyme and DNA sequence comparisons of nine species of *Encephalartos* (Zamiaceae). *Biochemical Systematics and Ecology*, 29, pp.241–266.

Vandepitte, K., Honnay, O., Mergeay, J., Breyne, P., Roldán Ruiz, I. and Meyer, T. 2013. SNP discovery using Paired-End RAD-tag sequencing on pooled genomic DNA of *Sisymbrium austriacum* (Brassicaceae). *Molecular Ecology Resources*, 13(2), pp. 269–275.

Vendrami, D.L.J. et al., 2017. RAD sequencing resolves fine-scale population structure in a benthic invertebrate: implications for understanding phenotypic plasticity. *Royal Society open science*, 4(2), p.160548.

Vilà, M. et al., 2011. Ecological impacts of invasive alien plants: a meta-analysis of their effects on species, communities and ecosystems. *Ecology letters*, 14(7), pp.702–708.

Volis, S., 2017. Conservation utility of botanic garden living collections: Setting a strategy and appropriate methodology. *Plant Diversity*, 39(6), pp.365–372.

Vovides, A.P., 1990. Spatial Distribution, Survival, and Fecundity of *Dioon edule* (Zamiaceae) in a Tropical Deciduous Forest in Veracruz, Mexico, with Notes on Its Habitat. *American Journal of Botany*, 77(12), p.1532.

Vovides, A.P. et al., 2003. An outline of our current knowledge on Mexican cycads (Zamiaceae, Cycadales). *Plant Biology*, 4, pp.159–174.

Vovides, A.P. et al., 1997. Pollination of endangered Cuban cycad *Microcycas calocoma* (Miq) A DC. *Botanical Journal of the Linnean Society*, 125(3), pp.201–210.

Wang, N., Thomson, M., Bodles, W.J.A., Crawford, R.M.M., Hunt, H.V., Featherstone, A.W., Pellicer, J. and Buggs, R.J.A. 2013. Genome sequence of dwarf birch (*Betula nana*) and cross-species RAD markers. *Molecular Ecology*, 22(11), pp. 3098–3111.

Warschefsky, E.J. & Wettberg, von, E.J.B. 2019. Population genomic analysis of mango (*Mangifera indica*) suggests a complex history of domestication. *The New Phytologist*, 222(4), 2023-2037.

Watkinson, A.R. & Powell, J.C., 1997. The Life History and Population Structure of C*ycas armstrongii* in Monsoonal Northern Australia. (3), p.341.

Wegrzyn, J. L., Liechty, J. D., Stevens, K. A., Wu, L. S., Loopstra, C. A., Vasquez-Gross, H. A., & Holt, C. 2014. Unique features of the loblolly pine (*Pinus taeda* L.) megagenome revealed through sequence annotation. *Genetics*, 196(3), 891–909.

Werth, S. & Scheidegger, C., 2014. Gene flow within and between catchments in the threatened riparian plant *Myricaria germanica*. *PLoS ONE*, 9(6), p.e99400.

Widmer, A., Lexer, C., & Cozzolino, S. 2009. Evolution of reproductive isolation in plants. *Heredity*, 102(1), 31.

Whitlock, R. et al., 2016. Consequences of in-situ strategies for the conservation of plant genetic diversity. *Biological conservation*, 203, pp.134–142.

Williams, R.J. et al., 1996. Variation in the composition and structure of tropical savannas as a function of rainfall and soil texture along a large-scale climatic gradient in the Northern Territory, Australia. *Journal of Biogeography*, 23(6), pp.747–756.

Wilson, W.D., Hutchinson, J.T. & Ostrand, K.G., 2017. Genetic diversity assessment of in situ and ex situ Texas wild rice (*Zizania texana*) populations, an endangered plant. *Aquatic Botany*, 136, pp.212–219.

Winn, A. A., Elle, E., Kalisz, S., Cheptou, P. O., Eckert, C. G., Goodwillie, C., & Vallejo-Marín, M. 2011. Analysis of inbreeding depression in mixed-mating plants provides evidence for selective interference and stable mixed mating. *Evolution: International Journal of Organic Evolution*, 65(12), 3339-3359.

Woinarski, J.C.Z. et al., 2006. Distributional pattern of plant species endemic to the Northern Territory, Australia. *Australian Journal of Botany*, 54(7), p.627.

Wright, L.I., Tregenza, T. & Hosken, D.J., 2007. Inbreeding, inbreeding depression and extinction. *Conservation Genetics*, 9(4), pp.833–843.

Wu, C.-S. & Chaw, S.-M., 2015. Evolutionary Stasis in Cycad Plastomes and the First Case of Plastome GC-biased Gene Conversion. *Genome biology and evolution*, 7(7), pp.2000–2009.

Wu, C.-S. et al., 2007. Chloroplast genome (cpDNA) of Cycas taitungensis and 56 cp protein-coding genes of *Gnetum parvifolium*: insights into cpDNA evolution and phylogeny of extant seed plants. *Molecular biology and evolution*, 24(6), pp.1366–1379.

Xiao, L. et al., 2005. Comparison of the genetic diversity in two species of cycads. *Australian Journal of Botany*, 53(3), pp.219–223.

Yang, S.-L. & Meerow, A.W., 1996. The *Cycas pectinata* (Cycadaceae) complex: genetic structure and gene flow. *International Journal of Plant Sciences*, 157(4), pp.468–483.

Yang, J., Warnow, T. 2011. Fast and accurate methods for phylogenomic analyses. *BMC Bioinformatics*, 12(9), pp. 4.21.

Yang, J. et al., 2016. High regional genetic differentiation of an endangered relict plant *Craigia yunnanensis* and implications for its conservation. *Plant Diversity*, 38(5), pp.221–226.

Yang, Y. et al., 2016. Comparative Analysis of the Complete Chloroplast Genomes of Five Quercus Species. *Frontiers in plant science*, 07(573), p.803.

Yi, D.-K. et al., 2015. The complete plastid genome sequence of *Abies koreana* (Pinaceae: Abietoideae). *Mitochondrial DNA Part A*, 5, pp. 2351 –3

Yoder, A.D. et al., 2018. Neutral Theory is the Foundation of Conservation Genetics. *Molecular biology and evolution*, 13(8), pp, 1322–1326.

Young, A., Boyle, T. & Brown, T., 1996. The population genetic consequences of habitat fragmentation for plants. *Trends in Ecology & Evolution*, 11(10), pp.413–418.

Zhan, Q.-Q. et al., 2011. Patterns of chloroplast DNA variation in Cycas debaoensis (Cycadaceae): conservation implications. Conservation Genetics, 12(4), pp.959–970.

Zgurski, J.M., Rai, H.S., Fai, Q.M., Bogler, D.J., Francisco-Ortega, J. and Graham, S.W. 2008. How well do we understand the overall backbone of cycad phylogeny? New insights from a large, multigene plastid data set. Molecular *Phylogenetics and Evolution*, 47(3), pp. 1232–1237.

Zhang, Y.-Y. et al., 2018. Development and Application of Genomic Resources in an Endangered Palaeoendemic Tree, *Parrotia subaequalis* (Hamamelidaceae) From Eastern China. *Frontiers in plant science*, 9, p.246.

Zhou, T., Zhao, J., Chen, C., Meng, X. and Zhao, G. 2016. Characterization of the complete chloroplast genome sequence of *Primula veris* (Ericales: Primulaceae). *Conservation Genetics*, 8(4), pp. 455–458.

Zimin, A., Stevens, K.A., Crepeau, M.W., Holtz-Morris, A., Koriabine, M., Marçais, G., Puiu, D., Roberts, M., Wegrzyn, J.L., de Jong, P.J., Neale, D.B., Salzberg, S.L., Yorke, J.A. and Langley, C.H. 2014. Sequencing and assembly of the 22-gb loblolly pine genome. *Genetics*, 196(3), pp. 875–890.

Zonneveld, B.J.M. 2012. Genome sizes for all genera of Cycadales. *Plant Biology*, 14(1), pp. 253–256.

Zonneveld, B.J.M., 2012. Conifer genome sizes of 172 species, covering 64 of 67 genera, range from 8 to 72 picogram. *Nordic Journal of Botany*, 30(4), pp.490–502.

Zonneveld, B.J.M. & Lindstrom, A.J., 2016. Genome sizes for 71 species of *Zamia* (Cycadales: Zamiaceae) correspond with three different biogeographic regions. *Nordic Journal of Botany*, 34(6), pp. 744-751.

Appendix I. Table of wild collected Cycas calcicola samples for RADseq.

72 samples wild collection from population between the Lichfield and Katherine region of the Northern Territory, Australia. Sample ID = assigned dates associated collection number, Region = Samples collection from Lichfield or Katherine region, Popualtion = individual population samples were collection from (CUL = cultivated origin), Cul vs Wild = if the samples were obtained directly from wild population. DBG = George Brown Darwin Botanic Gardens, Northern Territory, Australia and MBC = Montgomery Botanical Center, Flora, USA.

Sample ID	Region	Populations	Cul vs wild
Ccal-15-17-1	Litchfield	Litchfield NP 1	Litchfield NP 1
Ccal-15-17-10	Litchfield	Litchfield NP 1	Litchfield NP 1
Ccal-15-17-2	Litchfield	Litchfield NP 1	Litchfield NP 1
Ccal-15-17-3	Litchfield	Litchfield NP 1	Litchfield NP 1
Ccal-15-17-4	Litchfield	Litchfield NP 1	Litchfield NP 1
Ccal-15-17-5	Litchfield	Litchfield NP 1	Litchfield NP 1
Ccal-15-17-6	Litchfield	Litchfield NP 1	Litchfield NP 1
Ccal-15-17-7	Litchfield	Litchfield NP 1	Litchfield NP 1
Ccal-15-17-8	Litchfield	Litchfield NP 1	Litchfield NP 1
Ccal-15-17-9	Litchfield	Litchfield NP 1	Litchfield NP 1
Ccal-15-18-10	Litchfield	Litchfield NP 2	Litchfield NP 2
Ccal-15-18-1	Litchfield	Litchfield NP 2	Litchfield NP 2
Ccal-15-18-2	Litchfield	Litchfield NP 2	Litchfield NP 2
Ccal-15-18-3	Litchfield	Litchfield NP 2	Litchfield NP 2
Ccal-15-18-4	Litchfield	Litchfield NP 2	Litchfield NP 2
Ccal-15-18-5	Litchfield	Litchfield NP 2	Litchfield NP 2
Ccal-15-18-6	Litchfield	Litchfield NP 2	Litchfield NP 2
Ccal-15-18-7	Litchfield	Litchfield NP 2	Litchfield NP 2
Ccal-15-18-8	Litchfield	Litchfield NP 2	Litchfield NP 2
Ccal-15-18-9	Litchfield	Litchfield NP 2	Litchfield NP 2
Ccal-15-19-10	Litchfield	Litchfield NP - Tolmer	Litchfield NP - Tolmer
Ccal-15-19-1	Litchfield	Litchfield NP - Tolmer	Litchfield NP - Tolmer
Ccal-15-19-2	Litchfield	Litchfield NP - Tolmer	Litchfield NP - Tolmer
Ccal-15-19-3	Litchfield	Litchfield NP - Tolmer	Litchfield NP - Tolmer
Ccal-15-19-4	Litchfield	Litchfield NP - Tolmer	Litchfield NP - Tolmer
Ccal-15-19-5	Litchfield	Litchfield NP - Tolmer	Litchfield NP - Tolmer
Ccal-15-19-6	Litchfield	Litchfield NP - Tolmer	Litchfield NP - Tolmer
Ccal-15-19-7	Litchfield	Litchfield NP - Tolmer	Litchfield NP - Tolmer
Ccal-15-19-8	Litchfield	Litchfield NP - Tolmer	Litchfield NP - Tolmer
Ccal-15-19-9	Litchfield	Litchfield NP - Tolmer	Litchfield NP - Tolmer

Ccal-15-32	Daly River	Daly River	DBG
Ccal-15-34	Daly River	Daly River	DBG
Ccal-15-38	Daly River	Daly River	DBG
Ccal-15-29	Daly River	Daly River	DBG
Ccal-15-30	Spirit Hills	Spirit Hills	DBG
Ccal-15-31	Spirit Hills	Spirit Hills	DBG
Ccal-15-21-10	Katherine	Katherine-TT 1	Katherine-TT 1
Ccal-15-21-1	Katherine	Katherine-TT 1	Katherine-TT 1
Ccal-15-21-2	Katherine	Katherine-TT 1	Katherine-TT 1
Ccal-15-21-3	Katherine	Katherine-TT 1	Katherine-TT 1
Ccal-15-21-4	Katherine	Katherine-TT 1	Katherine-TT 1
Ccal-15-21-5	Katherine	Katherine-TT 1	Katherine-TT 1
Ccal-15-21-6	Katherine	Katherine-TT 1	Katherine-TT 1
Ccal-15-21-7	Katherine	Katherine-TT 1	Katherine-TT 1
Ccal-15-21-8	Katherine	Katherine-TT 1	Katherine-TT 1
Ccal-15-21-9	Katherine	Katherine-TT 1	Katherine-TT 1
Ccal-15-22-10	Katherine	Katherine-TT 1	Katherine-TT 1
Ccal-15-22-1	Katherine	Katherine-CDU 1	Katherine-CDU 1
Ccal-15-22-2	Katherine	Katherine-CDU 1	Katherine-CDU 1
Ccal-15-22-3	Katherine	Katherine-CDU 1	Katherine-CDU 1
Ccal-15-22-4	Katherine	Katherine-CDU 1	Katherine-CDU 1
Ccal-15-22-5	Katherine	Katherine-CDU 1	Katherine-CDU 1
Ccal-15-22-6	Katherine	Katherine-CDU 1	Katherine-CDU 1
Ccal-15-22-7	Katherine	Katherine-CDU 1	Katherine-CDU 1
Ccal-15-22-8	Katherine	Katherine-CDU 1	Katherine-CDU 1
Ccal-15-22-9	Katherine	Katherine-CDU 1	Katherine-CDU 1
Ccal-15-23-10	Katherine	Katherine-CDU 1	Katherine-CDU 1
Ccal-15-23-1	Katherine	Katherine-CDU 2	Katherine-CDU 2
Ccal-15-23-2	Katherine	Katherine-CDU 2	Katherine-CDU 2
Ccal-15-23-3	Katherine	Katherine-CDU 2	Katherine-CDU 2
Ccal-15-23-4	Katherine	Katherine-CDU 2	Katherine-CDU 2
Ccal-15-23-5	Katherine	Katherine-CDU 2	Katherine-CDU 2
Ccal-15-23-6	Katherine	Katherine-CDU 2	Katherine-CDU 2
Ccal-15-23-8	Katherine	Katherine-CDU 2	Katherine-CDU 2
Ccal-15-23-9	Katherine	Katherine-CDU 2	Katherine-CDU 2
Ccal-15-25	Katherine	Katherine-CUL-MBC	MBC
Ccal-15-26	Katherine	Katherine-CUL-MBC	MBC
Ccal-15-28	Katherine	Katherine-CUL-MBC	MBC
Ccal-15-33	Katherine	Katherine-CUL-DBG	DBG
Ccal-15-35	Katherine	Katherine-CUL-DBG	DBG
Ccal-15-36	Katherine	Katherine-CUL-DBG	DBG
Ccal-15-37	Katherine	Katherine-CUL-DBG	DBG

Appendix II. Detailed RADseq protocol to support "RADseq as a valuable tool in plants with large genomes—a case study in cycads"

Preface of protocol

Here we present a RADseq protocol, which we have adapted to the cycads. This approach has been modified from the ezRAD protocol v3.0 (Toonen et al. 2013) developed at the ToBo Lab at the University of Hawaii at Manoa. This protocol has been tested across all 10 genera of the Cycadales including: *Bowenia*, *Ceratozamia*, *Cycas*, *Dioon*, *Encephalartos*, *Lepidozamia*, *Macrozamia*, *Microcycas*, *Stangeria* and *Zamia*. Our modifications are indicated by *****. For more detailed instructions and the full protocol, consult the most recent version of the ezRAD ToBo Lab protocol.

1. Genomic DNA extraction and Quantification

Genomic DNA was obtained from 0.05 g of silica-dried leaf material. Extractions can be carried out using a Qiagen DNeasy plant kit, following the standard protocol (Qiagen document HB-1166-003-1095837) with modifications as follows:

1. *****Once buffer P3 is added, samples should be left at -20°C for a minimum of 20 minutes or overnight.

2. *****Add 55 µL of AE buffer during final elution step and repeated as per Qiagen protocol, to obtain a higher concentration of DNA in the final 110 µL elution.

Following extraction, samples should be tested for the presence of impurities using 2% agarose gel (**Tip: a smear indicates impurities and small fragments of DNA**). Samples which show a significant smear should be cleaned using AMPure XP beads prior to quantification (see section 3). Finally, each sample should be quantified using a Qubit (3.0 BR DNA assay; Invitrogen, Life Technologies, Carlsbad, CA, USA), and samples with a concentration lower than 15 µg/mL should be either re-extracted or concentrated using AMPure XP beads (see below).

2. Concentration of genomic DNA using AMPure XP beads

Samples which do not meet the minimum concentration of 15 µg/mL can be concentrated using AMPure XP beads. However, it is recommended that samples are firstly re-extracted. The method for DNA concentration is described below:

1. Add 1:1 ratio of AMPure XP beads to sample, e.g. for 100 µL sample, add 100 µL beads.
2. Mix well by pipetting until beads and sample appear homogeneous in colour, then incubate at room temperature for 5 minutes.
3. Place plate on magnet for 5 minutes or until liquid is clear.
4. Remove and discard supernatant.
5. Add desired amount of Illumina resuspension buffer (10 mM TRIS), e.g. for 100 µL of starting sample, add 65 µL resuspension buffer to concentrate sample.
6. Remove plate from magnet, mix sample until beads and sample appear homogeneous in colour (**Tip: make sure all beads are fully suspended**), and incubate at room temperature for 5 minutes.
7. Place plate back on magnet, stand for 5 minutes or until liquid is clear.
8. Transfer 60 µL of sample to a new low-binding tube and use the remaining 5 µL to quantify the DNA using a Qubit.

3. Restriction digest and sample standardisation

Double or single digest reactions should be carried out using either a single restriction enzyme (RE) or a combination of two REs to digest the genomic DNA at cut sites, e.g. ACTG (EcoR1-HF) or AAT (Mse1) (**Tip: This protocol does not use adapters that require specific enzymatic cut sites, so any restriction enzyme/s can be used**).

The restriction digest reactions should be set up using the following modifications to the standard NEB protocol:

1. ***Genomic DNA should be normalised to 500 ng using laboratory grade dH_2O in 42 µL volume (43 µL if a single enzyme is used) within a 96 well plate using a liquid handing robot (Qiagen QiaGility, Qiagen Inc., Venlo, the Netherlands) (**Tip: this step can also be performed manually**).

2. Add 5 µL of 10x NEB CutSmart buffer and 1 µL of BSA to each well and mix thoroughly. The mixture should then be placed laboratory grade refrigerator for a minimum of *5 hours (**Tip: the mixture can be left to refrigerate overnight**).

3. 1 µL of each RE should be added into each tube and mixed thoroughly by pipetting. The plate should be placed into a thermocycler for *3 x 1-hour cycles at 37° C, followed by a 20 minute enzyme deactivation stage at 65°C.

4. Double digest clean-up

Digested samples should be cleaned using Agencourt AMPure XP beads (Bechman Coulter). For a more detailed protocol follow the most recently updated Illumina TruSeq protocol with modifications suggested by Toonen et al. (2013). A summary is as follows:

1. Add 90 µL of AMPure XP beads to 50 µL of sample (calculated at 1:1.8 ratio) and follow the standard Illumina clean up protocol. Elute into 36 µL Illumina resuspension buffer (10 mM TRIS).

2. Transfer 30 µL of supernatant to a new plate.

3. 2 µL of remaining supernatant can be used to re-quantify samples using Qubit to check for DNA recovery (**Tip: between 60-80% of digested genomic DNA should be retained**).

4. Use the remaining supernatant 3 µL to visualise samples on a LabChip, Bioanalyzer or agarose gel checking for digestion of genomic DNA and fragment distribution.

5. Library preparation

Library preparation should be carried out using an Illumina TruSeq nano kit (Illumina Inc., CA, USA) or equivalent. Below are our modifications to the ezRAD v.3 protocol (Toonen et al., 2013) tested using 1/2 of the recommended volumes, of library preparation kits to save costs.

Modifications to the ezRAD protocol are as follows:

1. After PCR, carry out a standard bead clean instead using 1:0.8 ratio Ampure XP beads. This is a size selection step that removes fragments below 200 bp (**Tip: this is designed to reduce adapter dimer**).

2. Final libraries should be validated using a LabChip or bioanalyzer.

3. ***If adapter dimer is still present in final libraries, carry out another bead clean using 1:0.9 ratio Ampure XP beads (this is common with ezRAD).

4. Quantify each sample in the final library using a Qubit high sensitivity kit (3.0 HR DNA assay; Invitrogen, Life Technologies, Carlsbad, CA, USA).

5. Once libraries are validated and quantified they should be normalized to 10 nM. 5 µL of each library should be pooled (multiplexed) (**Tip: the concentration can be lowered to 5 nM where samples are lower quality**).

6. Quantify the pooled (multiplexed) final library using a Qubit high sensitivity kit (prior to? sequencing).

7. ***Visualise the final pooled libraries using a LabChip or bioanalyzer to check for presence of adapter dimer and to analyse final fragment size. If adapter dimer is present, carry out another 1:0.9 bead clean on the final pool and re-quantify.

8. Libraries should then be sequenced using an Illumina HiSeq or NextSeq sequencer using high throughput 150 dual index paired read. Aim for 1 Gb sequencing depth per sample for large genomes.

Tip: Sequencing should be spiked with 20% PhiX to help with low diversity associated with RADseq runs (due to the identical enzyme cut sites). However, on newer sequencing platforms e.g. HiSeq 4000 this can be reduced to 10%.