The distribution of genetic diversity in sleepy lizards

Talat Hojat Ansari

M. Sc. In Animal Biosystematics, Shiraz University

Presented for the degree of Doctor of Philosophy
School of Biological Sciences
Flinders University
Australia

August 2015

Supervisors:

Dr. Michael G. Gardner, School of Biological Sciences, Flinders University. Prof. C. Michael Bull, School of Biological Sciences, Flinders University. Prof. Steve J. B. Cooper, Australian Centre for Evolutionary Biology and Biodiversity, the University of Adelaide

Flinders
UNIVERSITY

Table of contents

Abstract

Understanding the distribution of genetic diversity in and among organisms provides a cornerstone for asking fundamental questions in evolutionary biology. These questions are advanced by connecting the processes affecting genetic diversity within and among individuals, populations and species. Already a strong discipline, molecular ecology has been undergoing a recent resurgence due to dramatic advances in the tools of the trade, the development and availability of molecular markers accessing both neutral and functional areas of genomes. Thanks to these approaches, molecular ecologists have increasing abilities to assess the patterns of genetic variation within and among different levels of life – individuals, populations and species. Multiple markers are necessary to provide a more unbiased view of the many processes and the potential factors that might have contributed to the distribution of current genetic diversity.

Here, I characterise and apply a variety of molecular markers in a widespread monogamous Australian skink *Tiliqua rugosa*. This species is a member of the *Egernia* group of skinks which contains species with many varied social structures. My thesis consists of several marker development chapters for both neutral and functional markers; a chapter on the development of tools for obtaining data from next generation sequencing runs; and culminates in a chapter examining the distribution of current genetic diversity and the inference of the historical factors that have led to the observed pattern. First, I used next generation shotgun sequencing of *T. rugosa* genomic DNA to develop primers for 48 anonymous nuclear loci (ANL).These primers were extensively tested in *T. rugosa*, and the related *Tiliqua*

adelaidensis and *Egernia stokesii*. These polymorphic loci provide a useful source of neutral variation for the study of phylogeography and population genetic structure explored later in my thesis. The next two chapters describe the use of transcriptome sequencing for the characterization of markers for a highly diverse gene region which is involved in the immune response and mate choice, the Major Histocompatability Complex (MHC). I designed specific primers from the resultant sequence assembly and successfully amplified exon 2, 3 and 4 from MHC class I, and exon 2 and 3 in class II. The loci were confirmed to be Mendelianly inherited via the sequencing of several families and the utility of these primers was further tested for the related *T. adelaidensis* and *E. stokesii*. The characterization of these markers is rare in squamates and represents a breakthrough for investigation in the *Egernia* group of social lizards and should provide the basis for much further study.

In the next chapter I used these loci and adapted a newly developed technique to provide a cost effective method to amplify the multiple loci from a large number of individuals. I then used these amplifications, of both the neutral and functional loci from 250 *T. rugosa* individuals, to test bioinformatics software (MITAGSORTER) that I developed to sort the large complicated data sets I obtained from next generation sequencing.

Finally, I used the distribution of genetic variation in samples of *T. rugosa* from across the entire distribution range of the species in Australia to assess historical processes that potentially were responsible for the contemporary pattern of genetic structure in *T. rugosa.* The pattern of genetic diversity was quantified using mitochondrial (sequences from *ND4*) and nuclear (sequences from non-encoding intron 7 of β-fibrinogen, glyceraldehyde-3-phosphate dehydrogenase and nine ANL I developed) markers. This study

revealed three major mitochondrial lineages for *T. rugosa*, although potential secondary admixture was detected, based on nuclear markers. In order to further investigate the potential historical events associated with these phylogeographic patterns, I combined species distribution modelling and a phylogeographic diffusion model. The analyses suggest that the range of *T. rugosa* contracted during the Last Glacial Maxima (LGM) and that the region around the Murray River and the Nullarbor Plain acted as barriers to gene flow since this period. However there is the possibility that other factors, such as selection on mtDNA, may also have contributed to the pattern of variation observed. My study hints at the presence of multiple refugia for this species during previous glacial maxima, with potential secondary contact following climate amelioration.

My thesis provides valuable resources for the further investigation of the distribution of genetic diversity within this iconic lizard species at scales spanning from individual interactions and mate choice through to a thorough investigation of the role of various ecological processes on the patterns of contemporary genetic diversity. In addition the markers developed can be utilized in investigations at similar scales for many other species within this group of lizards, the *Egernia* group, with their unusual and varied social systems.

Declaration

I certify that this thesis does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university; and that to the best of my knowledge and belief it does not contain any material previously published or written by another person except where due reference is made in the text.

Talat Hojat Ansari

Acknowledgements

I would like to sincerely thank my Principle Supervisor Dr. Michael Gardner, for all his support and advice over the past few years. Thanks to Mike for always being supportive and encouraging of my ideas, and helping me to make them possible. I have learnt so much from him that I can't summarise them all here. I would also like to thank my co-supervisors Prof. Michael Bull and Prof. Steve Cooper for all of the comments and suggestions they have provided me. Thank you all for your guidance, patience and advice.

I would like to thank A/Prof. Michael Schwarz for his help and advice. Thank you also to all my co-authors for their help, support and guidance. I appreciated support from the South Australian Museum (SAM) staff for help in sourcing materials, and especially Terry Bertozzi for bioinformatics help. I also would like to thank all of our lab members, in particular Dr. Alison Fitch, Dr. Kym Abrams and Dr. Tessa Bradford for their support and advice in lab work. In addition, thank you to my labmates Dr. Scott Groom, Nahid Shokri, Sarah Pearson, Jessica Hacking, Harry Johnston, Amy Slender, Rebecca Dew and Steffi Graff for their support.

I am very appreciative of all the generous financial support I received during my PhD. Funding for the project was provided by the Sir Mark Mitchell Research Foundation.

Last but certainly not least, I would like to thank my family for their unremitting patience, support and love. My parents, Maryam and Feizollah, always reminded me of the value of education and encouraged me to work

hard to achieve my goals. I am lucky to have such great people in my life, especially my lovely husband, Dr. Mehregan Ebrahimi. He was my motivator, assistant, friend, punching bag, and he was a big support whenever I needed him. This thesis is every bit yours as it is mine, I am not hesitate in saying that without your support and help I could not have made this dream of mine become true.

Organisation of Thesis

This dissertation is formatted as a series of scientific manuscripts and is composed of seven chapters, containing a general introduction (Chapter 1), two published paper (Chapters 2 and 3), three manuscripts prepared for submission to scientific journals (Chapters 4- 6), and a general conclusion (Chapter 7). I have formatted each chapter as per the journal requirements of the journal they are published in or intended to be submitted to. Between each chapter I provide a linking page to indicate where the chapter is published or intends to be submitted to for publication. The linking pages also provide a guide to where the chapter fits into the other material in the thesis.

Chapter 2: Random non-coding fragments of lizard DNA: anonymous nuclear loci for the Australian skink, *Tiliqua rugosa*, and their utility in other *Egernia*-group species.

In this published chapter I used next generation sequencing and de nova approaches to developed 48 anonymous markers from the *T. rugosa*. These developed markers were tested also for two related species *Tiliqua adelaidensis* and *Egernia stokesii*. These loci provides a suite of markers useful for studying population genetic and phylogeography.

Chapter 3: MHC in a monogamous lizard - characterization of Class I MHC genes in the Australian skink *Tiliqua rugosa*.

I characterized MHC class I sequence diversity from *T. rugosa*, using both cDNA and genomic sequence data and also present genomic class I sequences from the related skinks *T. adelaidensis* and *E. stokesii*. This chapter is published.

CHAPTER 4: Characterisation of class II MHC genes in a monogamous Australian skink (*Tiliqua rugosa*).

In this unpublished chapter I used the cDNA sequences sourced from the RNAseq run in chapter 3, but characterized the MHC class II β chain. I developed primers to amplify the exon 2 and 3 of the class II for *T. rugosa* and two other related species *T. adelaidensis* and *E. stokesii*.

CHAPTER 5: Dual indexing with a single PCR for library preparation and development of MITAGSORTER software for de-mulitiplexing sequences data on the Illumina MiSeq platform.

I prepared the MiSeq library by amplifying a large number of samples from *T. rugosa* for both recently developed ANL and MHC markers. After searching for a program to use to sort the MiSeq reads from different individuals into separate files, I discovered that it would be beneficial to organise the development of one myself and I then developed a software program including a manual. Currently unpublished.

CHAPTER 6: Plio-Pleistocene diversification and biogeographic barriers in the Australian arid zone reflected in phylogeography of a widespread and common species.

I used a variety set of markers from mtDNA to nuclear markers including ANL developed in Chapter 2 to assess the level of genetic variation, population genetic and effect of demography and historical event across the whole distribution of *T. rugosa* in Australia. Currently unpublished.

Other published work

During my period as a PhD student I was involved in other publications where I performed the gathering of data and contributed to writing. Both of these articles were also on my study species.

Lancaster, M.L., Gardner, M.G., Fitch, A.J., **Ansari, T.H**. & Smyth, A.K. 2012 A direct benefit of native saltbush revegetation for an endemic lizard (Tiliqua rugosa) in southern Australia. Aust. J. Zool. 60, 192-198.

Godfrey, S.S., **Ansari, T.H.**, Gardner, M.G., Farine, D.R. & Bull, C.M. 2014 A contact based social network of lizards is defined by low genetic relatedness among strongly-connected individuals. Anim. Behav. 97. (doi:doi:10.1016/j.anbehav.2014.08.019).

Statement of co-authorship

The following people contributed to the preparation and publication of work undertaken as part of this thesis:

Talat Hojat Ansari (Candidate), School of Biological Sciences, Flinders University, Adelaide, Australia.

Michael G. Gardner (Supervisor, Author 2), School of Biological Sciences, Flinders University, Adelaide, Australia; and Evolutionary Biology Unit, South Australian Museum, Adelaide, Australia.

C. Michael Bull (Co-Supervisor, Author 3), School of Biological Sciences, Flinders University, Adelaide, Australia.

Steve J. B. Cooper (Adjunct-Supervisor, Author 4), Evolutionary Biology Unit, South Australian Museum, Adelaide, Australia; and Australian Centre for Evolutionary Biology and Biodiversity and the School of Biological Sciences, University of Adelaide, Adelaide, SA 5005, Australia.

Terry Bertozzi (Author 5), Evolutionary Biology Unit, South Australian Museum, Adelaide, Australia; and Australian Centre for Evolutionary Biology and Biodiversity and the School of Biological Sciences, University of Adelaide, Adelaide, SA 5005, Australia.

Michael P. Schwarz (Author 6), School of Biological Sciences, Flinders University, Adelaide, Australia.

Robert D. Miller (Author 7), Center for Evolutionary and Theoretical Immunology, Department of Biology, University of New Mexico, Albuquerque, NM, USA.

Mehregan Ebrahimi (Author 8), Department of Biology, Shiraz University, Shiraz, Fars Province, Iran; and School of Biological Sciences, Flinders University, Adelaide, Australia.

Esmaeil Ebrahimie (Author 9), Department of Genetics and Evolution, School of Biological Sciences, University of Adelaide, Adelaide, South Australia, Australia; and School of Information Technology and Mathematical Sciences, Division of Information Technology, Engineering and the Environment University of South Australia, Adelaide, Australia.

Mahdi Ebrahimi (Author 10), Department of Informatics, Saarland University, Saarbrucken, Germany.

Gaynor Dolman (Author 11), Molecular Systematics Unit, Western Australian Museum; and Australian Centre for Evolutionary Biology and Biodiversity, University of Adelaide, Australia.

Jessica Hacking (Author 12), School of Biological Sciences, Flinders University, Adelaide, Australia

Leah Rheinberger (Author 13), (formally of) School of Animal Biology, University of Western Australia, Crawley 6009 Western Australia, Australia

CHAPTER DETAILS AND RELATIVE AUTHOR CONTRIBUTIONS ARE LISTED BELOW:

Chapter 2 is published as:

Ansari, T. H., Bertozzi, T., Hacking, J., Cooper, S. J. B. and Gardner, M. G. (2015). Random non-coding fragments of lizard DNA: anonymous nuclear loci for the Australian skink, *Tiliqua rugosa*, and their utility in other *Egernia*-group species. *Australian Journal of Zoology* 62:515-518.

The candidate was the primary author. The candidate carried out data analysis in conjunction with author 5. The candidate conducted all laboratory works with assistance of author 12. The candidate and authors 2, 4 and 5 contributed to the development of main ideas and approach, as well as refining the text.

Chapter 3 is published as:

Ansari, T. H., Bertozzi, T., Miller, R. D. and Gardner, M. G. (2015). MHC in a monogamous lizard - characterization of Class I MHC genes

in the Australian skink, *Tiliqua rugosa*. *Developmental & Comparative Immunology* http://dx.doi.org/10.1016/j.dci.2015.07.012.

The candidate was the primary author and conducted all laboratory works. Author 5 assisted with analysis. The candidate as well as authors 2, 5 and 7 contributed to developing the main ideas, and approach as well as in refining the text.

Chapter 4 is in preparation:

Ansari, T. H., Bertozzi, T., Miller, R. D. and Gardner, M. G. (In prep) Characterisation of class II MHC genes in a monogamous Australian skink, *Tiliqua rugosa*.

The candidate was the primary author. The candidate conducted all laboratory work, and also all of the data analysis in conjunction with author 5. The candidate as well as authors 2, 5 and 7 contributed to developing the main ideas, and approach as well as in refining the text.

Chapter 5 is in preparation:

Ansari, T. H., Ebrahimi, M., Ebrahimi, E. and Gardner, M. G. (In prep). Dual indexing with a single PCR for library preparation and development of MITAGSORTER software for de-mulitiplexing sequences data on the Illumina MiSeq platform.

The candidate was the primary author and conducted all laboratory work. Author 10 provided bioinformatics analysis assistance. The candidate as well as author 2 and 9 developed the main ideas and approach, and assisted with refining the text.

Chapter 6 is in preparation:

Ansari, T. H., Cooper, S. J. B., Schwarz, M. P., Ebrahimi, M., Dolman, G., Delphs, L., Bull, C. M., Gardner, M. G. (In prep). Plio-Pleistocene diversification and biogeographic barriers in the Australian arid zone reflected in phylogeography of a widespread and common species.

The candidate was the primary author and conducted all of the laboratory work. The candidate conducted all analysis with assistance of author 8 and 6. Author 13 assisted with sample collection. The candidate as well as author 2, 4 and 6 developed the main ideas and approach. The candidate and authors 2, 3, 4, 6 and 11 assisted with refining the text.

We the undersigned agree with the above stated "proportion of work undertaken" for each of the above published or prepared to submission manuscripts contributing to this thesis:

Signed:

Candidate: Talat Hojat Ansari

Michael G. Gardner

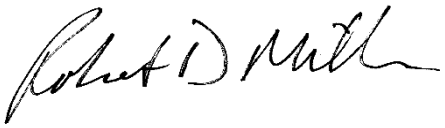
C. Michael Bull

Steve J. B. Cooper

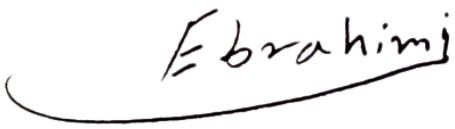
Terry Bertozzi

Michael P. Schwarz

Robert D. Miller



Mehregan Ebrahimi



Esmaeil Ebrahimie

Mahdi Ebrahimi.

Gaynor Dolman

Jessica Hacking

Leah Rheinberger,

Chapter one

1. General introduction

Understanding the distribution of genetic diversity in and among organisms provides a cornerstone for asking fundamental questions in evolutionary biology. These questions are advanced by connecting the processes affecting genetic diversity within and among individuals, populations and species (Lowe, et al., 2009; Hedrick, 2011). Already a strong discipline, molecular ecology has been undergoing a recent resurgence due to dramatic advances in the tools of the trade, the development and availability of molecular markers accessing both neutral and functional areas of genomes (e.g. Gardner, et al., 2011; Bertozzi, et al., 2012). Thanks to these approaches, molecular ecologists have increasing abilities to assess the patterns of genetic

variation within and among different levels of life – individuals, populations and species.

My thesis is primarily concerned with the understanding the forces which lead to the current distribution of genetic diversity across the range of a species. I discuss the potential evolutionary processes which shaped the current genetic structure, and compare the observed data with the proposed theories behind them. As this thesis is presented as a series of published manuscripts or manuscripts in preparation, each chapter provides an introduction and discussion related to the specific question proposed on its own. The following provides a broad overview of topics important to the thesis.

1.1. Population structure

The genetic variation of populations is influenced by gene flow, the movement of migrant genes into a population, and can be affected by biological and demographic factors (Lowe, et al., 2009; Hedrick, 2011). Spatial genetic variation has resulted from distinct barriers to dispersal and isolation by distance (e.g. Berthier, et al., 2005; Gauffre, et al., 2008), reproductive isolation (Butlin, et al., 2014), and migration and dispersal behaviour associated with mating strategies (Stow and Sunnucks, 2004; Moussy, et al., 2013). The dispersal ability of species can be limited, for example, because of demographic processes such as in *Tiliqua adelaidensis*, an endangered Australian skink (Schofield, et al., 2012), and/or due to contemporary landscape discontinuities such as fragmented habitats and agricultural fields (e.g. Stow, et al., 2001;

Berthier, et al., 2005; Fant, et al., 2014; Nowakowski, et al., 2015) or the effects of roads (e.g. Frantz, et al., 2012; Gurrutxaga and Saura, 2014).

Additionally, gene flow and population genetic structure are strongly effected by environmental factors such as physical barriers and landscape structures (Rueness, et al., 2003; Trizio, et al., 2005; Lowe, et al., 2009; Smissen, et al., 2013; Sexton, et al., 2014). Numerous studies have shown the important influence of landscape structures on the gene flow for both invertebrate (Alp, et al., 2012; Phillipsen and Lytle, 2013) and vertebrate species (Shoo, et al., 2008; Mullen, et al., 2010; Row, et al., 2010; Faulks, et al., 2011; Klug, et al., 2011; Frantz, et al., 2012; Peterman, et al., 2013), and also in plants (Young, et al., 1996; Byrne and Hopper, 2008; Cranmer, et al., 2012; Rasic and Keyghobadi, 2012). These longer term landscape structures include globally distributed historical landscape separations such as barriers caused by ice sheets and the subsequent retraction of species ranges into glacial refugia during the Pleistocene (e.g. Hewitt, 2004b; Margraf, et al., 2007; Pepper, et al., 2011; Yannic, et al., 2012; Marín, et al., 2013; Younger, et al., 2015).

Understanding the potential factors which are responsible for changes to how genetic variation is distributed across landscapes can be elucidated by combining genetics, geography, history and ecology, using the well-established phylogeographic method (Avise, 2000).

1.2. Phylogeography

Phylogeography is an approach to study the geographic distribution of genetic lineages, particularly within a single species or among closely related species (Avise, 2000). The field of phylogeography provides detailed species-specific information about how historical processes shaped the current distribution and genetic structure of species (Avise, et al., 1987; Riddle, 1996). It also helps to understand how the genetic variation and differentiation of populations effected by past environmental changes and paleoclimate events (Avise, 2000; Avise, 2003; Kidd and Ritchie, 2006).

Under the advance of phylogeographic studies, a number of scenarios have been proposed regarding the genetic constitution of species and population differentiation as a result of glacial-interglacial cycles (Riddle, 1996; Moritz, et al., 2000; Hewitt and Ibrahim, 2001; Carstens, et al., 2004; Steele and Storfer, 2006). During the Last Glacial Maxima (LGM), as a general concept, reduction in population size was observed due to the unsuitable climate and environments, which forced species into restricted favourable areas - refugia (Webb, 2008; Dubey and Shine, 2010; Coghlan, et al., 2015). Populations, therefore, became isolated through surviving in multiple localised refugia (e.g. Chapple, et al., 2005), or large singular refugia (e.g. Schneider and Moritz, 1999) which restricted gene flow among them. Ultimately new genetic lineages formed through the action of genetic drift and selection, and populations genetically differentiated. However, population expansion during the post glacial cycles has resulted in secondary contact in some species (e.g. Hewitt, 1999; Zamudio and Savage, 2003; Mach, et al., 2011). Most of the

studies have used northern hemisphere species and much less in known about how southern hemisphere species may have been affected.

1.3. Phylogeography in southern Australia

The evolution and distribution of Australian fauna shows evidence for the effect of palaeoclimatic and biogeographic events across the continent since the mid-Tertiary (Byrne, et al., 2008). The growing body of evidence that glacial maxima, which were associated with significant increase aridity across Australia (Byrne, et al., 2008), highlight the importance of studying this area as a priority for informing the conservation of fauna and flora under future climate change.

Although outstanding studies of phylogeographic history have been undertaken within tropical regions in Australia (Hugall, et al., 2002), there has been a much less efforts to understand the effects of Plio-Pleistocene climate change on the present-day diversity of species in the Southern Hemisphere (Byrne, et al., 2008). This is particularly true about the arid and semi-arid region in southern Australia (Byrne, et al., 2008).

Interglacial periods have been associated with increased aridity across the Australian continent (Byrne, 2008). Over the past 5 million years (Myr) southeast Australia experienced substantial transitions in vegetation, where previously widespread forest and rainforest were replaced with more open vegetation and sclerophyllous woodlands (Bowler, 1982; Markgraf, et al., 1995; Gallagher, et al., 2001; Gallagher, et al., 2003; Byrne, et al., 2008). During this period, the climate fluctuated between warm-wet and cool-dry, with an

increased frequency of cooling-drying throughout the Pliocene (Bowler, 1982; Markgraf, et al., 1995; Martin, 2006). The frequency of the climatic oscillations and aridity cycles intensified during the Gelasian and Calabrian (2.6-0.7 Myr), and reached a maximum during the lower Pleistocene (0.7 - 0.01 Myr. Nix, 1982; Markgraf, et al., 1995; Byrne, et al., 2008).

These climate cycles of environmental changes and severe aridity are likely to result in major fluctuations in population size and cause localise extinction events in different parts of a species range (Hughes, 2003; Thomas, et al., 2004). Species currently widespread in the arid and semi-arid zone are likely to be flexible to such events, by surviving in refugial locations and expanding their range when climate ameliorates. So far, the few studies which have been conducted in the arid and semi-arid zone do not indicate a common pattern of genetic structure as a result of historical process in southern Australia. Therefore studies highlight a need for further comprehensive phylogeographic studies or more species to be assessed in this area (Byrne, 2008; Shoo, et al., 2008; Fujita, et al., 2010; Byrne, et al., 2011; Dolman and Joseph, 2012; Edwards, et al., 2012; Lanier, et al., 2013; Rabosky, et al., 2014).

A number of genetic studies that have been done at a micro-geographic scale (e.g. Chapple, et al., 2005; Dubey and Shine, 2010; Chapple, et al., 2011) or across multiple-species in southern Australia (e.g. Oliver, et al., 2009; Dolman and Joseph, 2012; Dolman and Joseph, 2015), have suggested that species' distributions comprise multiple genetically distinct populations. These studies show some common patterns of population structure and signals of expansion indicative of major shifts in distribution associated with climate

change. The source of these expansion events is most likely to be in upland regions associated with the Gawler, Flinders and Mount Lofty Ranges and coastal regions of south-eastern South Australia, where rainfall is likely to have been greater (Williams, et al., 2001). Major range shifts can also have been associated with recent climatic events, such as major droughts or bushfires, and the signature of these events may be evident in the genetic structure of species. What is lacking is the use of a species which has a wide distribution across this range to provide the ability to understand the role of different proposed refugia and genetic barriers simultaneously. Part of the reason for the dearth of studies has been due to the need to develop multiple genetic markers de-novo for new species.

1.4. Molecular genetic markers

An important aspect of phylogeography is to uncover the signatures left by past climates on the distribution and genetic structures of populations by using genetic markers (Avise, 2000; Avise, 2009; Hickerson, et al., 2010). Genetic markers (a gene, DNA sequence, or gene product whose inheritance can be traced through a pedigree or phylogeny) have made it possible to infer the pattern of gene flow, and to investigate the evolutionary processes influencing variation within and among populations (Hare, 2001; Avise, 2009; Hickerson, et al., 2010).For about two decades, mitochondrial DNA (mtDNA) has been a marker of choice in population genetic and phylogeographic studies.

Some of the properties of mtDNA which has made it so popular as a marker are: a lack of, or at least very limited evidence for, recombination,

providing easily interpretable sequence variation; high level of copy numbers in each cells facilitating its amplification; the ease of finding sequences for comparison on databases such as GenBank; and having a rapid evolutionary rate, which increases the likelihood of finding intra-specific patterns (Avise, et al., 1987; Moore, 1995; Avise, 2000). Although very useful and commonly used mtDNA, its use is hampered by several factors. First, the mitochondria is normally maternally inherited so may be a misleading description of evolutionary history of an organism, in particular in cases with sex-biased dispersal or strong mate selection (Hare, 2001; Zhang and Hewitt, 2003; Ballard and Whitlock, 2004). Second, there is only a single locus so presents a snapshot of what might be happening across whole genomes. Third, there is increasing evidence of selection acting on mtDNA due to different environmental conditions (Bazin, et al., 2006; Meiklejohn, et al., 2007; Pavlova, et al., 2013). All these factors initiated the use of multilocus phylogeographies to more rigorously test biogeographical hypotheses (Pavlova, et al., 2013), as has been applied in more recent studies (Fujita, et al., 2010; Edwards, et al., 2012; Welton, et al., 2013; Thome, et al., 2014).

Despite their importance to population genetics and phylogeography, few molecular markers have been available, particularly for non-model organisms (Bertozzi, et al., 2012). Next generation sequencing (NGS) approaches have revolutionised the development of genetic markers (e.g. microsatellites, Gardner, et al., 2011), and so provide the researchers with a growing datasets of different markers (Thomson, et al., 2010). Further benefit boost the developing new genetic markers, is directly related to the advance in bioinformatics analysis. Bioinformatics as an interdisciplinary field, creates and

develops software and protocols, based on a combination of computer science, statistics, mathematics, and engineering, for analysing and interpreting of biological data. The most advance bioinformatics analysis supported sorting the huge amount and complicated data set produced through NGS that facilitate the development of new genetic markers. However, some marker types have been less exploited. One of these less attended marker classes is anonymous nuclear loci (ANL), which can benefit scientists looking for a large number of loci to be fast involving and also independent (Bertozzi, et al., 2012). The ANL markers express random segments of nuclear DNA and so have better representation of variation across the entire genome compare with any other marker (Lee and Edwards, 2008). These markers have multiple loci that confident intervals can be assessed for demographic parameters, such as population divergence time or effective population sizes, with one of them (Brito and Edwards, 2009). In addition, because the majority of non-coding DNA is unlikely to be under selection, therefore, the ANL markers provide a useful source of neutral variation for phylogenetics, phylogeographic and population genetic studies (Rosenblum, et al., 2007; Lee and Edwards, 2008; Thomson, et al., 2008; Bertozzi, et al., 2012). Many of these new markers are chosen for their presumed neutrality.

However, given that selection plays a key role in shaping biodiversity (Funk, et al., 2006), the importance and development of markers for functional loci is also rapidly growing (Hoffmann, et al., 2003; Sommer, 2005a; Nadachowska-Brzyska, et al., 2012; Kyle, et al., 2014). These functional markers provide a direct information about the selective process involved in interaction of individuals and their environment or/and in the capability for

future adaptive changes (Meyers and Bull, 2002; van Tienderen, et al., 2002). In addition, as an evolutionary perspective, the separation of populations can arise on contemporary timescales and might not be detectable at neutral loci (Cohen, 2002; Stockwell, et al., 2003; Sommer, 2005a). This is particularly true when this time scale is too short to leave a signal at neutral loci and therefore the population variation are only detectable at adaptive genes (Cohen, 2002), such as the major histocompatibility complex (MHC).

The MHC genes known as the most polymorphic complex in vertebrates which play a fundamental rule in the adaptive immune system (Hedrick, 1994; Dong, et al., 2013). They are involved in both self/non-self-recognition, disease susceptibility and mate choice (Sommer, 2005a). These substantial biological traits can be affected by the MHC variations that implicate the important of understanding genetic diversity at the MHC level(Sommer, 2005a). Therefore, the MHC genes, because of their important functions and characteristics, are excellent candidates to assess adaptive gene variability (Ujvari and Belov, 2011; Bichet, et al., 2015).

So far, the MHC of many non-model vertebrate taxa are poorly characterised, limiting its application to studying the genetic diversity and to conservation genetics in these groups.

1.5. Case study species

This thesis tackles the need for investigating a common and widespread species for understanding the influence of past climates on southern Australian

arid/semi-arid regions. I use the well-studied sleepy lizard, *Tiliqua rugosa*, to provide the first comprehensive phylogeographic study that covers almost the entire arid and semi-arid zone in southern Australia.

The sleepy lizard is a large, snout-vent length up to 340mm, mainly herbivorous, long lived, monogamous and viviparous Australian skink (Bull, 1987; Dubas and Bull, 1991; Bull and Pamula, 1998). The species is common in semi-arid regions, mostly in southern mainland of Australia (Bull, 1987; Greer, 1989). They occupy undefended, large, stable home ranges, which overlap substantially with other individuals of both sexes (Bull, 1978; Satrawaha and Bull, 1981; Bull, 1987; Dubas and Bull, 1991; Bull, 1994; Freake, 1996), and for most of the year, home range sizes average under 4000 m2 (Bull, 1978; Satrawaha and Bull, 1981; Dubas, 1987; Yeatman, 1988). Average home range size increases during spring when lizards are most active (Bull, 1978; Satrawaha and Bull, 1981).

The sleepy lizard is an appropriate case study model for several reasons. First many aspects of the species' biology and ecology is known due to its having been the subject of a long term study by Prof. C. Michael Bull from Flinders University. Second, it is a common and widespread species, which can form a baseline to understand and compare the population genetic structure and the effect of historical and demographic event on other populations or species (Whiteley, et al., 2006). Common species are ecologically very important as they have a large number of interactions with other biota. Thus the declines in common species cause massive impact across species assemblies (Gaston and Fuller, 2008). Third, partly due to its abundance, samples of this

species are readily available in the Australian Biological Tissue collection held by the South Australian Museum. Finally, the use of common species means that we can utilise the same genetic markers to determine the locations of long term refugia across a large area rather than having to develop further genetic markers for other species.

I consider the potential effect of physical and historical barriers to gene flow across the range of *T. rugosa*. One of the important aspect of this project is using multilocus approach where I develop new markers to answer investigate historical climates on contemporary genetic diversity.

In particular, I use next generation sequencing (NGS) to characterise and develop multiple markers including neutral markers, anonymous nuclear loci (ANL) and functional markers of the major histocompatibility complex (MHC) loci. I also use the Illumina MiSeq platform to amplify the developed markers in a large number of samples from the studied species. Finally, I use genetic data from a set of nuclear markers (including intron 7 of β-fibrinogen, glyceraldehyde-3-phosphate dehydrogenase and anonymous nuclear loci), and mitochondrial DNA to examine the effects of historical glaciation events and ecological barriers on a current genetic structure of the species.

Chapter two

2. Anonymous nuclear loci

This chapter focuses on the de-novo development of multiple neutral genetic markers for use in the validation of the program MiTagSorter (Chapter 5) and the phylogeographic study (Chapter 6).

These markers will be useful to investigate the effects of demographic and historical events in the current study, and are likely to provide an initial source for researchers who wish to investigate the population genetic and phylogeography of other *Egernia* group species.

Australian Journal of Zoology (2015) 62(6) 515-518

Random non-coding fragments of lizard DNA: Anonymous nuclear loci for the Australian skink, Tiliqua rugosa and their utility in other *Egernia* group species

Talat Hojat Ansari, Terry Bertozzi, Jessica Hacking, Steven J. B. Cooper

and Michael G. Gardner

2.1. Abstract

We report the development of 48 anonymous nuclear loci from the Australian skink *Tiliqua rugosa* using 454 sequencing. These loci amplified across a Western Australian lineage (47 loci), a Northern lineage (48 loci) and a Southern lineage (46 loci). We further tested amplification for the related *T. adelaidensis* and *Egernia stokesii* where 37 and 34 loci amplified respectively. The loci showed variability within *T. rugosa* (22 polymorphic loci) and at least 27 loci also exhibiting variation among the three species, highlighting the usefulness of these markers for phylogenetic, phylogeographic and population genetic analyses in *T. rugosa* and related species.

2.2. MAIN TEXT

Anonymous nuclear loci (ANL) are single-copy nuclear markers that are randomly selected across genomes (Balakrishnan, et al., 2010; Bertozzi, et al., 2012). These loci are likely to be from non-coding regions of the genome (Thomson, et al., 2010), and they can contain higher levels of variation than introns (Lee and Edwards, 2008). These ANL regions are probably independent, potentially not affected by selection, and also less affected by ascertainment bias in marker choice (Balakrishnan, et al., 2010), making them a useful source of neutral variation for phylogenetics, phylogeography and population genetics. For example, ANL's have been used in a multilocus phylogenetic reconstruction of the squamate reptiles (Thomson, et al., 2008) and several studies have applied these markers for phylogeographic analyses (Cooper, et al., 1995; Rosenblum, et al., 2007; Lee and Edwards, 2008).

Here we describe the development of 48 ANL's for an Australian skink, the sleepy lizard *Tiliqua rugosa,* from randomly generated sequences produced by the Roche 454 platform. *Tiliqua rugosa* is broadly distributed in semi-arid areas of southern Australia making it a suitable species for contributing to the dearth of phylogeographical studies of Australia's arid biota, as reported by Byrne, et al. (2008). Phylogenetic analyses across the distributional range of *T. rugosa* in Australia, using mitochondrial DNA (NADH Dehydrogenase subunit 4 (*ND4*)) revealed three lineages with a mean pairwise nucleotide difference of 24.67. These geographical lineages correspond to the northern and the southern regions of the Murray River in southern Australia and a western lineage spanning the west and east regions of the Nullarbor Plain (Ansari *et al.*

unpublished data). However the existing nuclear markers, *β-fibrinogen* and glyceraldehyde-3-phosphate dehydrogenase (Dolman and Phillips, 2004), have proved inadequate to identify areas used as long term refugia and to assess the demographic history required to understand the long term effects of past climate change on this species (Ansari *et al.* unpublished data). Therefore the development of further nuclear markers is necessary.

In addition, *T. rugosa* is a member of the *Egernia* group, notable for their high levels of sociality which is unusual for lizards (Gardner, et al., 2008). We therefore evaluated the utility of the developed ANL's in the related species *Tiliqua adelaidensis* and *Egernia stokesii* to facilitate further understanding of this prominent group of lizards.

We obtained high quality DNA (5 µg), using a puregene DNA isolation kit (Gentra systems, Minneapolis MW, USA), from the liver tissue of a single *T. rugosa* individual (SAMR63474) from the Mt Mary region of South Australia (33° 55′S; 139° 21′E). This sample was subjected to shotgun sequencing on a 454 (Roche), following Gardner, et al. (2011), at the Australian Genomic Research Facility, (AGRF Brisbane, Australia). These data have been deposited at the Dryad Digital Repository. We followed the protocol of Bertozzi, et al. (2012) to identify potential ANL loci from the resulting sequences. Briefly, we used the programme 'sffinfo' (Roche GSassembler suite) to extract the FASTA formatted sequences from the GS-FLX output files. Highly repetitive sequences such as microsatellites, known transposable elements, as well as species-specific repeats, were identified using the build-1mer-table script in RepeatScout (Price, et al., 2005). Highly

repetitive sequences were then masked in the FASTA file, with Repeat Masker (Smit, et al., 2010). The resultant library was compared with the Anolis genome using BLASTN V.2.2.21 to identify known coding regions and these were removed. Sequences with substantive hits to mtDNA and those with open reading frames were also removed. We identified 12,332 sequences over 450 bp that were suitable for further processing.

We randomly chose 48 potential ANL's and designed primers for these with target total lengths of 200 bp in Primer 3 (Rozen and Skaletsky, 2000) using the default values for annealing temperature (60 º C), primer length (19-21 bp) and % GC content (50 %). These primers were tested for amplification success with *T. rugosa* samples from North (NA) (ABTC 01276 and ABTC 55308) and South (SA) (ABTC 55271 and ABTC 34301) of the Murray River and also samples from Western Australia (WA) (ABTC 55145 and ABTC 55276) (representing the three geographic lineages) as well as two individuals from each of the related species *T. adelaidensis* (near Burra, SA) and *E. stokesii* (near Hawker, SA). All samples were collected under appropriate permits. The polymerase chain reaction (PCR) amplifications were performed in 25 µL reaction volumes, containing 1× reaction buffer (Applied Biosystems), 20 ng DNA, 0.2 mM of each primer, 0.8 mM each dNTP and 2 mM MgCl2, 0.5 U DNA polymerase (Ampli Taq Gold™; Applied Biosystems). The cycling profiles for all reactions consisted of a single cycle of 9 min at 94 °C, followed by 34 cycles of 45 s at 94 °C, 45 s at 59 °C and 1min at 72 °C, ending with 6 min at 72 °C. PCR amplification products then were purified directly using a multiscreen PCR 384 Filter Plate (Millipore). The purified PCR products were sequenced in the forward direction for one of the two amplicons

from each of the three geographic regions of *T. rugosa* and for *T. adelaidensis* and *E. stokesii*. Sequencing reactions consisted of 3 min at 96 °C, followed by 30 cycles of 30 s at 96 °C, 15 s at 50 °C and 4 min at 60 °C, ending with 1 min at 25 °C. The sequencing reactions were purified using SEQ 384 Filter Plate (Millipore) before sending to the AGRF (Adelaide Australia) for capillary sequencing on an AB 3730*xl*. The resultant sequences were manually checked in MEGA v5 (Tamura, et al., 2011), to ensure heterozygotes and variable sites were correctly scored.

Of the 48 loci tested in *Tiliqua rugosa* from each of three different geographic lineages, 47 successfully amplified for the Western Australian (WA) sample, all 48 for the Northern Australian (NA) and 46 for Southern Australian samples (SA) (Table 1). In *T. adelaidensis* and *E. stokesii*, 37 and 34 of the loci amplified a single clear product respectively (Table 1). We further tested the utility of the loci by examining the sequence variation among the three samples (one sample from each of the three geographic regions) of *T. rugosa.* The standard diversity statistics, including nucleotide (π), number of haplotypes (h), number of polymorphic sites (s), and pairwise sequence divergence (d_x) within *T. rugosa,* were calculated using Arlequin 3.5.1 (Excoffier and Lischer, 2010b) and MEGA v5 (Excoffier and Lischer, 2010b).

Table 1. PCR primers for amplification trials of 48 anonymous nuclear loci in *Tiliqua rugosa*, *T. adelaidensis* and *Egernia stokesii*. Fragment numbers refer to *T. rugosa* 454 shotgun sequences from which the primers were developed (see text). An x indicates no or poor amplification; a ✔ designates loci that successfully amplified with a product at the expected size; and a # signifies that more than one product was amplified. WA, NA and SA refer to the *T. rugosa* samples from Western Australia, North and South of the Murray River, respectively.

Locus ID	Fragment	Primer sequences (5'—3')	*T. rugosa*			*T. adelaidensis*	*E. stokesii*
			WA	NA	SA		
Tiru-1	G6P0TEI01DRJ0N	F:TTTAAAGGGCAGCATTGGTC R:TGGGATGAGAACGTGAGAAAC	✔	✔	✔	✔	✔
Tiru-2	G6P0TEI01B5CD0	F:ACGCACCCATGAGTTAGAGG R:GCTGCTGTTGATGCTGGAG	✔	✔	✔	✔	✔
Tiru-3	G6P0TEI02JFNOC	F:TCCAAGCTGACATCCAAAGG R:GGCTACATTGCACACAGTGG	✔	✔	✔	✔	✔
Tiru-4	G6P0TEI02IAGYT	F:TATGCACAGAGCCTTCAATG R:CCCAAGCAAGTGGACTTCAG	✔	✔	✔	✔	✔
Tiru-5	G6P0TEI02HRJ7L	F:TTCAGTGGTTGGCTCAACTG R:ACACAGTCCCGTGGAAGAAG	✔	✔	✔	✔	✔
Tiru-6	G6P0TEI01CDUFZ	F:GGACACCCAAGGTCTGAGAG R:CTCCCTTCTCGGCAATACAG	✔	✔	✔	✔	✔
Tiru-7	G6P0TEI02JR2D6	F:TCAGGGTACAGCGAACTGC R:ACCATGGCTCTGCATAGGTC	✔	✔	✔	✔	✔
Tiru-8	G6P0TEI02J2505	F:GAATGCAAGGGAAGAACAGC R:TGCAGCACAGGAGATGCTAC	✔	✔	✔	#	✔
Tiru-9	G6P0TEI02G8H36	F:CTAGTGCAGGCTCCTCATCC R:GAGGTGGCAAGCTGAAGAAC	✔	✔	✔	×	✔
Tiru-10	G6P0TEI02HMZMV	F:TGCTTACCCTCTTGGTGTCC R:TGGCTCTACCATGTCACCTTC	✔	✔	✔	✔	✔
Tiru-11	G6P0TEI02JCQ6N	F:CAGCCACCTTGATCCTCAAC R:GGCTTCTTCCACATTGCTTC	✔	✔	✔	✔	×
Tiru-12	G6P0TEI01CO8QB	F:TGGTAGCTGGGTAGTTCAAGG R:AGAGGGACGGAGACAATGTG	✔	✔	✔	×	✔
Tiru-13	G6P0TEI01B1E0M	F:CCTGTGACTGCTCTGAGTTCC R:TCCAGCTCTTTGTGTGTCTCC	✔	✔	✔	×	×
Tiru-14	G6P0TEI01D2AEI	F:TTCATCTCGTTGCCTCTGTG R:TGCATTCGGTCTGCTTACTG	✔	✔	✔	✔	✔
Tiru-15	G6P0TEI01AT2SA	F:CGGAACACATGTCCATTCTG R:TGATCTGAAACCTGCTGCTG	✔	✔	✔	✔	✔
Tiru-16	G6P0TEI02H465A	F:CCTTTGGCAAATGTGGGTAG R:CAGTGGTTGTTGGTTGGTTG	✔	✔	✔	#	✔
Tiru-17	G6P0TEI02GNPMI	F:GTTCACCTACCCACCCACTC R:ACCAATTCAGGATTCCATGC	✔	✔	✔	✔	×
Tiru-18	G6P0TEI02FVFLD	F:TCAGAGCGCAGAGAAGTGAG R:CGTTGCAAGAGTTCCGAATG	✔	✔	✔	×	×
Tiru-19	G6P0TEI02IN8QP	F:TGAGAGAAGGAAGCCCTGAG R:CATTGTCAGTGGTCAACATGG	✔	✔	✔	✔	✔
Tiru-20	G6P0TEI01C5T2H	F:AGAGTCCACAGGCAGCTGAG R:AAGTGACCACACAGCAATGG	✔	✔	✔	✔	✔
Tiru-21	G6P0TEI01ALS3Y	F:TCTGGAGCACGTTGAATCAG R:ATGAGGAATGCAGGGTGAAG	✔	✔	✔	✔	✔
Tiru-22	G6P0TEI02GZ1U4	F:TGACGGCTGTCAGACAATAGAG R:TTTGTGTTTGCCCATGAGAG	✔	✔	✔	✔	✔

Table 1. Continued

Locus ID	Fragment	Primer sequences (5'-3')	*T. rugosa*			*T. adelaidensis*	*E. stokesii*
			WA	NA	SA		
Tiru-23	G6P0TEI02IL409	F:GAAGTGCCAGGAGGTCAATC R:TCACATAGCCCTTGGTTTCC	✔	✔	✔	✔	✔
Tiru-24	G6P0TEI02H1RR9	F:GAACCAGGTGGAAACGACTC R:CCAGGCAGCCTCTTACAGTC	✔	✔	✔	✔	✔
Tiru-25	G6P0TEI02F51K3	F:TGAGGGTACATGCTCCACAC R:TTTCTGGGATTGGCTTGAAC	✔	✔	✔	✔	✔
Tiru-26	G6P0TEI02G6DVW	F:CCTTCCACCTCCAAAGTGTG R:TTGATCCTGCTGTCAAGTGG	✔	✔	✔	✔	✔
Tiru-27	G6P0TEI01DEAKG	F:ACGGAACTTCTCCGTGCTTC R:CCTCATTGGCTCTCGTCTTC	✔	✔	×	✔	×
Tiru-28	G6P0TEI01BNMTW	F:CAGAATCAATCACCGCACAG R:ATGGAAATGCACAGGAAAGG	✔	✔	✔	✔	✔
Tiru-29	G6P0TEI01DFRRG	F:ACAGTGGCCTACCAAAGTCG R:CCTGATGAAGGAGCAGGAAG	✔	✔	✔	×	×
Tiru-30	G6P0TEI01CRFVE	F:CTGAGGACAGTCAGGGCTTC R:CTCCAGAGCATTCTGGGTTC	✔	✔	×	×	×
Tiru-31	G6P0TEI01EHO0M	F:GGCAATGTTGATGTTTGTGC R:GTGTCTCCATGTGTCCATGC	×	✔	✔	✔	×
Tiru-32	G6P0TEI02HXTRM	F:TCCGCTGTTGTAATGTGCTC R:TGCCCTGGAAGTTGTCTTTC	✔	✔	✔	✔	✔
Tiru-33	G6P0TEI02JRT7U	F:GAATCGCACATGCACAGAAC R:TTCACCAAATCCTGGCCTAC	✔	✔	✔	✔	✔
Tiru-34	G6P0TEI02GR3YF	F:GCAAGCAACCAGGAAGAAAC R:CATGGATTCAGATGCCACAC	✔	✔	✔	✔	✔
Tiru-35	G6P0TEI02FQS01	F:TAGGGAAATTGGAGGCACAG R:TGAGCAAATGGAGTCTGGAG	✔	✔	✔	✔	✔
Tiru-36	G6P0TEI02HKXD1	F:ATACGGAAGGGACTGCATTG R:AGGGCTACGAGTGACATTGG	✔	✔	✔	✔	✔
Tiru-37	G6P0TEI02GXNW6	F:GAAGAACCGAGCAAGAGTGG R:GGAGGCTGATCAAGGAACTG	✔	✔	✔	✔	✔
Tiru-38	G6P0TEI02HC181	F:GCTGAGATCCTCCTCATTGC R:TCACCACCATGAAGCTGAAC	✔	✔	✔	×	×
Tiru-39	G6P0TEI02HQK8W	F:GGGAGAGCAGAAAGATGGTG R:CCTCTAGTGGGTGGCAGAAG	✔	✔	✔	✔	×
Tiru-40	G6P0TEI01EFKNO	F:TTCTCCTGACCAATGCACAC R:AGGGCTGTCACACCTTATGC	✔	✔	✔	×	×
Tiru-41	G6P0TEI02I17AF	F:ATGTCTCCATGGCTGATTCC R:TTGCTAGCCTGCCAAATACC	✔	✔	✔	✔	✔
Tiru-42	G6P0TEI02G4H7X	F:GGATGCATATGGCTACAGGAG R:AATTTCACCAGCCTCAGTGG	✔	✔	✔	✔	✔
Tiru-43	G6P0TEI02JP9W6	F:TAATGAACCTGGCCCTTCTG R:TCTTACCGTCAGCCCTTCAG	✔	✔	✔	✔	✔
Tiru-44	G6P0TEI02H4RJU	F:ATTTGGATGCTGTGGGTAGC R:GGCATGGGCAGAGTAACTTC	✔	✔	✔	✔	✔
Tiru-45	G6P0TEI02FSPUT	F:CAGGGCAGTCAATGTTCTTG R:GTGCCATCGGTACTGAAATG	✔	✔	✔	✔	×
Tiru-46	G6P0TEI02FMCXD	F:CACATGCACATTTCCTCTGG R:CTGTGGCACTGCTCTTCTTG	✔	✔	✔	✔	×
Tiru-47	G6P0TEI02FNUCB	F:TGTGCAGACATGGTTCCTTC R:TCAGTCCAGCATCAGTGCTC	✔	✔	✔	✔	✔
Tiru-48	G6P0TEI02F49JR	F:TCCATTTGCATCAGCGTATC R:TGGACTCCTGGGAGTAGGTC	✔	✔	✔	×	×

We found 22 loci exhibited variation among the three *T. rugosa* samples with nucleotide diversity ranging from 0.001 to 0.04 (Table 2). The remaining 26 loci were invariant. Locus Tiru-4 contained an indel in the sequences from SA and NA. However, the levels of variability for these 22 loci within *T. rugosa* are higher in 17 ANL loci compared with the nucleotide diversity of the two previously available nuclear markers, *β-fibronogen* (0.004) and *GAPD* (0.008), which we assessed for the same individuals. For comparison, the nucleotide variation for the mitochondrial gene NADH Dehydrogenase subunit 4 (*ND4*) in the same *T. rugosa* individuals was 0.05. The level of successful amplification in the three studied species (96 %, 77 % and 71 % in *T. rugosa*, *T. adelaidensis* and *E. stokesii*, respectively) is comparable with the amplification success reported in other reptiles' studies using this technique (80 % Gehyra lazelli and 71 % for Hydrophis spiralis in Bertozzi et al. 2012). We found that 30 loci successfully amplified in all three targeted species. We then successfully sequenced a single individual from each species for 28 loci.

The nucleotide diversity among the three species ranged from 0.67 to 1.00 (Table 3). All loci except Tiru-4 exhibit variation among the three *Egernia* group species in our study. Although we tested a small number of individuals, the high levels of nucleotide variation observed in Tables 2 and 3 are consistent with other studies of nuclear variation, particularly in reptiles (Hughes and Mouchiroud, 2001; Rosenblum, et al., 2007).

The high level of genetic variation among the different species suggest that these markers will be useful for phylogenetic, phylogeographic, and

population genetic studies within *T. rugosa* and other members of the enigmatic *Egernia* group.

Table 2 Characteristics of the 22 anonymous nuclear loci that exhibited variation among individuals from the three geographic lineages of *Tiliqua rugosa*. Size (bp) indicates size in base pairs of amplicon. Number of segregating (polymorphic) sites (S), number of haplotypes (H), nucleotide diversity (π) and mean number of pairwise differences (d_x). *=locus containing an indel, metrics are calculated excluding the indel

Locus ID	Size (bp)	S	H	π	d_x
Tiru-2	184	2	2	0.01	1.3
Tiru-4*	70	1	3	0.01	0.6
Tiru-9	144	1	2	0.01	0.67
Tiru-15	161	2	3	0.01	1.33
Tiru-16	115	2	3	0.01	1.33
Tiru-20	107	1	2	0.002	0.004
Tiru-21	130	1	2	0.01	0.67
Tiru-22	192	3	3	0.01	2.00
Tiru-23	121	2	3	0.02	1.33
Tiru-24	182	3	3	0.04	3.00
Tiru-28	209	1	2	0.001	0.002
Tiru-29	300	1	2	0.002	0.67
Tiru-33	184	1	2	0.004	0.67
Tiru-34	111	1	2	0.002	0.01
Tiru-36	150	1	2	0.001	0.003
Tiru-39	140	1	2	0.005	0.67
Tiru-43	170	1	2	0.004	0.67
Tiru-44	155	2	3	0.01	1.3
Tiru-45	166	3	3	0.02	2.00
Tiru-46	173	1	2	0.01	0.67
Tiru-47	161	1	2	0.01	1.00
Tiru-48	116	4	3	0.04	4.00

Table 3. Characteristics of the 27 variable anonymous nuclear loci that successfully sequenced in all three target species (Northern *Tiliqua rugosa* lineage; *T. adelaidensis*; *E. stokesii*). Size (bp) indicates size in base pairs of the largest amplicon. Number of segregating (polymorphic) sites (S), number of haplotypes (H), number of indel bases (I), nucleotide diversity (π) and mean number of pairwise differences (d_x).

Locus ID	Size (bp)	S	H	I	π	d_x
Tiru-2	193	25	3	15	0.09	16.7
Tiru-3	112	6	3	0	0.04	4.3
Tiru-5	172	10	3	0	0.04	6.3
Tiru-6	124	2	3	0	0.01	1.3
Tiru-10	184	36	3	27	0.14	24.00
Tiru-14	109	1	3	0	0.005	0.7
Tiru-15	158	6	3	0	0.03	4.00
Tiru-19	118	5	3	0	0.03	3.3
Tiru-20	91	6	3	0	0.03	3.00
Tiru-21	127	9	3	0	0.05	6.3
Tiru-22	188	3	2	0	0.01	3.00
Tiru-23	122	4	3	0	0.02	2.7
Tiru-24	190	4	2	0	0.03	4.00
Tiru-25	146	4	3	0	0.02	2.7
Tiru-26	131	3	3	0	0.01	2.00
Tiru-28	200	16	3	9	0.05	11.00
Tiru-32	127	1	3	0	0.01	1.3
Tiru-33	183	4	3	0	0.02	2.7
Tiru-34	111	8	3	0	0.04	5.3
Tiru-35	141	12	3	8	0.06	8.00
Tiru-36	139	1	2	0	0.01	1.00
Tiru-37	127	3	2	0	0.02	3.00
Tiru-41	143	5	2	0	0.02	3.33
Tiru-42	131	8	3	7	0.04	5.3
Tiru-43	155	7	3	0	4.7	0.03
Tiru-44	161	8	3	0	0.03	5.3
Tiru-47	140	1	3	0	0.01	0.67

CHAPTER THREE
3. THE MAJOR HISTOCOMPATIBILITY COMPLEX CLASS I

In this chapter (and the following one, Chapter 4), I set out to develop a functional marker set with the intention of using these in my investigation of genetic diversity as a direct comparison to the neutral markers. I chose genes of the major histocompatibility complex (MHC) as these genes would likely provide a reflection on the underlying co-evolutionary relationship with pathogens and parasites of the sleepy lizard and may be different in the different refugial locations and in different lineages. In addition the social interactions of this monogamous species may well be Illuminated using MHC

loci as the loci are also involved directly with mate choice. However, the considerable time it took to fully characterize the MHC meant that there was not enough time to utilize these loci in the context as they were intended. Nonetheless the developed markers have been used in Illumine MiSeq run (Chapter 5), and represent a considerable investment that will be available for future studies.

The development of MHC class I loci in this chapter therefore provides a valuable resource to investigate the genetic diversity, co-evolution with parasites, and effect of selection across the landscape.

Developmental and Comparative Immunology (2015) 53 320-327

MHC in a monogamous lizard-Characterization of class I MHC genes in the Australian skink *Tiliqua rugosa*

Talat Hojat Ansari, Terry Bertozzi, Robert D. Miller and Michael G. Gardner

3.1. ABSTRACT

The major histocompatibility complex (MHC) is a highly variable region of vertebrate genomes that encodes cellular proteins involved in the immune response. In addition to the benefits of MHC research in understanding the genetic basis of host resistance to disease, the MHC is an ideal candidate for studying genetic diversity under strong natural selection. However, the MHC of many non-model vertebrate taxa are poorly characterized, hindering an understanding of disease resistance and its application to conservation genetics in these groups. Squamates (lizards and snakes) remain particularly underrepresented despite their being the most diverse order of non-avian sauropsids. We characterized MHC class I sequence diversity from an Australian skink, the sleepy lizard (*Tiliqua rugosa*), using both cDNA and genomic sequence data and also present genomic class I sequences from the related skinks *Tiliqua adelaidensis* and *Egernia stokesii*. Phylogenetic analysis of *Tiliqua* and other published sqamate MHC class I sequences suggest that MHC diverged very early in *Tiliqua* compared with the other studied squamates. We identified at least 4 classical MHC class I loci in *T. rugosa* and

also shared polymorphism among *T. rugosa*, *T. adelaidensis* and *E. stokesii* in the sequences encoding peptide-binding a1 and a2 domains.

Keywords: MHC, Scincadae, Squamates, CDNA, Genomic DNA, *Tiliqua rugosa*, *Tiliqua adelaidensis*, *Egernia stokesii*

3.2. INTRODUCTION

The major histocompatibility complex (MHC) has long been a major focus of molecular evolution studies since the discovery in the 1930's of its role in regulating responses to transplanted tumours in inbred mouse strains (Gorer, 1936). Indeed, this gene region was an early focus of the human genome project (Nadachowska-Brzyska, et al., 2013) due to its importance in disease resistance. The MHC is an extremely dynamic, gene dense region that encodes cellular proteins involved in the immune response in vertebrates (Hedrick, 1994; Dong, et al., 2013). MHC genes can be highly polymorphic and are linked to important aspects of life history such as survival and reproductive success (Milinski, 2006). Therefore these genes are ideal candidates for studying the action of natural selection in wild populations and many studies have established the role of MHC genes in disease resistance in wildlife (Vigueira, et al., 2013b).

The MHC gene has two main types of receptors, class I and class II (Klein, 1986). MHC Class I molecules are displayed on the surface of the cells of most tissues and every nucleated somatic cell of vertebrates and are responsible for the presentation of intracellular or endogenously derived

antigens (Bjorkman and Parham, 1990). Class I genes encode a single polypeptide chain with structurally conserved protein domains. Two polymorphic exons, exons 2 and 3, encode the α1 and α2 domains that make up the variable peptide binding region (PBR) (Bjorkman, et al., 1987; Sieling, et al., 1995). A third extracellular domain, the α3 domain, associates non-covalently with a β2 -microglobulin molecule, whose presence is necessary for cell surface expression (Klein, 1986; Rodgers and Cook, 2005). There are also two C-terminal domains in many class I molecules: the transmembrane and cytoplasmic domains (Glaberman and Caccone, 2008).

MHC class I molecules have been classified further, based on their structure and function, into classical and non-classical roles. Classical class I have strong and ubiquitous expression across different types of tissues and are generally highly polymorphic. Classical class I molecules present short peptide fragments, derived from pathogens or the self-proteome generated within the host cell, to $CD8^+$ cytotoxic T lymphocytes (CTLs). Recognition of the MHC class I e peptide complex leads to CTL mediated lysis of infected host cells (Klein, 1986; Shawar, et al., 1994; Ploegh and Watts, 1998). Non-classical class I molecules are encoded by genes found either in or outside of the MHC region. Their expression usually is limited to specific tissue types and are typically monomorphic or oligomorphic (two to four alleles). Non-classical MHC molecules perform a wide range of functions that are not necessarily always related to immunity (Braud, et al., 1999; Rodgers and Cook, 2005) such as iron metabolism in humans (Braud, et al., 1999).

In contrast to the extensive MHC research in mammals, other vertebrate taxa, such as amphibians and non-avian sauropsids, are underrepresented, however this imbalance is being redress (Grossberger and Parham, 1992; Radtkey, et al., 1996; Miller, et al., 2005; Miller, et al., 2006; Ohta, et al., 2006; Hauswaldt, et al., 2007; Glaberman and Caccone, 2008; Glaberman, 2008; Murphy, et al., 2009). The order squamata, which is the most diverse reptile order among the four non-avian sauropsids (reptiles) orders and a main amniote lineage, is the most poorly studied of these groups. Analyses of squamate MHC genes will contribute to a better understanding of the evolution of the amniote MHC region as this lineage last shared a common ancestor with mammals more than 300 million years ago (Kumar and Hedges, 1998). In addition, characterization and isolation of MHC class I genes from the squamates will aid comparisons of adaptive genetic diversity and disease interactions in this group. Reptiles provide a fascinating group for study given their general reliance on chemical communication for many aspects of their lives such as prey and predator detection, species and individual identification and mate choice (Mason and Parker, 2010).

One group of squamates, the skinks (Scincidae), could contribute immensely to understanding the influence MHC loci have on mate choice as these animals appear to have advanced chemosensory discrimination that influences mate recognition (Olsson and Shine, 1998). In particular, skinks of the social *Egernia* group (Gardner, et al., 2015) provide an opportunity to understand the role that MHC may have in the formation of social aggregations. Members of this group show varying levels of sociality and these

differences have also been shown to correspond to discrimination of scats by conspecifics (Bull, et al., 1999).

In this study, our major aim was to characterize the MHC class I genes from the Australian sleepy lizard (*Tiliqua rugosa*). In addition, we investigated the genomic MHC class I sequences in the congener pygmy bluetongue lizard (*Tiliqua adelaidensis*) and gidgee skink (*Egernia stokesii*). The three skink species are squamate reptiles (subfamily: Lygosominae) from the well-studied *Egernia* group (Gardner, et al., 2008). *T. rugosa* is a long lived, monogamous and viviparous Australian skink (Bull, 1987). This species is widely distributed in arid to semiarid regions mostly in the southern mainland of Australia (Bull, 1987; Greer, 1989). In contrast, the closely related *T. adelaidensis* is an endangered species that survives in a few isolated fragments on native grassland in the Burra region in mid-South Australia. *Egernia stokesii* is a monogamous and group-living species of the *Egernia* group found primarily in rocky outcrops in arid to semi-arid regions of Australia (Gardner, et al., 2001).

3.3. MATERIAL AND METHODS

3.3.1. SAMPLE COLLECTION AND EXTRACTION OF RNA AND DNA

The spleen, thymus and a muscle sample from a *T. rugosa* individual collected from the Mt Mary region of South Australia (33° 55′S; 139° 21′E) were preserved in 10× the volume of RNA Later (Qiagen, Venlo, Netherlands), kept at 4 °C for 48 h and then stored at -80 °C until required for RNA

extraction. The remainder of the specimen was accessioned into the South Australian Museum herpetology collection (SAMR63474). RNA was isolated from the spleen and thymus (35 ng of mRNA from the thymus and 150 ng from the spleen) using the Qiagen Oligotex Direct mRNA kit (Qiagen, Venlo, Netherlands). The mRNA was fragmented using Mg^{2+} buffer, primed with random hexamers for cDNA synthesis and then sequenced on a single plate of a GS-FLX (Roche Applied Science, Indianapolis, Indiana, USA) at the AGRF (Brisbane, Australia) following standard protocols.

Blood samples were collected onto Whatman® FTA classic cards (GE Healthcare, Buckinghamshire, UK) from additional *T. rugosa* specimens from the Mount Mary region, *T. adelaidensis* from Burra (33° 42′S; 138° 56′E) and *E. stokesii*, from Witchelina Station (30° 00′S; 138° 00′E). Genomic DNA was extracted from the FTA cards using a modified FTA extraction method (Gardner, et al., 2008). In addition, genomic DNA was extracted from the tissue of the same sample (SAMR63474) that was sources for the cDNA library preparation and was used as a positive control, using a puregene DNA isolation kit (Gentra systems, Minneapolis MW, USA).

3.3.2. ASSEMBLY AND DISCOVERY OF MHC CLASS I SEQUENCES

Raw reads were assembled using gsAssembler release 2.5.3, using the -cdna flag, directly from the Standard Flowgram Format (SFF) files output from the GS-FLX. Sequence adapters were removed using the built in adapter removal tools with the quality and clipping information encoded in the SFF files. Assembly was carried out on a quad core Linux workstation with 8 GB of memory. A local BLASTX (2.2.21+) search of all assembled contigs, using

MHC class I amino acid GENBANK sequences from *Iguana iguana* (EU604316 e EU604319); *Amblyrhynchus cristatus*, (EU604308 - EU604312); *Conolophus subcristatus*, (EU604313 - EU604315); *Ameiva ameiva* (M81095, M81097); *Nerodia sipedon* (M81099); *Pelodiscus sinensis* (AB185245); *Sphenodon punctatus* (DQ145788, DQ145789) and *Gallus gallus* (X12780) as queries, was performed with a significance threshold of $e < 10^{-4}$, to identify contigs containing putative MHC class I sequences. The assemblies of the best matching contigs, from the gsAssembler ACE file, were examined in Tablet 1.14.10.21 (Milne et al., 2013) for evidence of misassembly. Reads were mapped back to the longest contig using gsMapper release 2.9 and the sorted BAM alignment visualised in IGV v2.2.11 (Robinson, et al., 2011). Sequence fragments congruent with the putative antigen binding exons of the contig were extracted from the BAM alignment using a combination of samtools 1.1 and AWK. The resultant fragments were clustered using USEARCH v7.0.1090 (Edgar, 2010) at 95% identity. Consensus sequences were generated for each cluster (excluding singleton clusters) and aligned using Geneious v6.1.7 (Drummond, et al., 2012) to investigate sequence diversity. The longest sequence from each resultant sequence group (the bait) was used in an iterative assembly strategy using the mirabait utility from MIRA v4.0.2 (Chevreux, et al., 1999) specifying a minimum of 100 kmers for a sequence to be selected. Following each mirabait run, the selected sequences were remapped to the bait using Geneious and a new longer bait assembled. The process was repeated until sequence length stabilised (approximately five interations). Where two sequences were recovered (likely representing allelic variation), these were separated and run separately. Finally, reads were mapped back to the newly

generated sequences with gsMapper with 75 bp overlap and 90% identity to evaluate the methodology.

3.3.3. PRIMER DESIGN AND AMPLIFICATION OF MHC CLASS I SEQUENCES

We aligned our putative MHC class I cDNA sequences with published class I sequences from other vertebrates in GenBank (Table 3), using Geneious v6.1.7 (Drummond, et al., 2012). Specif-ically, we included: Iguana iguana (EU604317 and EU604319); *Amblyrhynchus cristatus*, (EU604308 and EU604311); and *Conolophus subcristatus*, (EU604313, EU604315 and EU604316). Primers were designed from the aligned sequences (Fig. 1; Table 1), using Primer 3 as implemented in the Geneious software. Initially primers (Tables 1 and 2) for MHC class I exons 2-4 were trialled on genomic DNA obtained from SAMR63474 and *T. adelaidensis* and *E. stokesii* samples. Following successful amplification and sequencing we further tested amplification and sequenced individuals from three family groups of *T. rugosa* comprising mothers, fathers and one or two known offspring. The following primer pairs were used: for exon 2, MHC1α_2 F_2 with MHC1α_2 R_1; and for exon 3, MHC1α_2 F_1 with MHC1α_2 R_2. PCR was performed in 25 mL reactions, containing 20 ng DNA, 1 × reaction buffer (Applied Biosystems), 0.2 mM of each primer, 0.8 mM dNTP, 2 mM $MgCl_2$ and 0.5 unit of DNA polymerase (Ampli Taq Gold™; Applied Biosystems). The cycling profile consisted of a single cycle of 10 min at 94 ˚C, followed by 34 cycles of 30 s at 94 ˚C, 30 s at 60 ˚C and 1:30 min at 72 ˚C, with a final extension for 10 min at 72 ˚C. PCR amplification products were purified using a multiscreen PCR 384 Filter Plate

(Millipore). The purified products were sequenced using BigDye terminator chemistry, consisting of 3 min at 96 ˚ C, followed by 30 cycles of 30 s at 96 ˚ C, 15 s at 50 ˚ C and 4 min at 60 ˚ C. The sequencing products were purified using a SEQ 384 Filter Plate (Millipore) and sent to the AGRF (Adelaide Australia) for capillary sequencing on an AB 3730xl.

Table 1. Sequences of primers designed for amplification of MHC class I genomic sequences in this study. (F)= forward primer, (R) = reverse primer.

	Primer name	Primer Sequence in 5´→ 3´direction
1	**MHC1α 1 F1**	GGYKCCTCBTCCCACTCKGYGMRGTA
2	**MHC1α 1 F2**	ACGGCGGTGTCKGAGCCYRGCCAG
3	**MHC1α 1 R1**	TGTKCHDRTCCCAGWRMTGRGGGT
4	**MHC1 α 2 R**	GTCCAGGTGAGGGTCTCCTT
5	**MHC1α 2 F1**	TCACACBYKGCAGYBSATGTAYGGCTG
6	**MHC1α 2 F4**	ACAAGGAGACCCTCACCTGGA
7	**MHC1α 2 R2**	TCCTCWGCAGDSTCTCCYTCC
8	**MHC1α 2 R3**	CCBTCCCRTASYSCACGTRYTTC
9	**MHC1α 3 F1**	GCCCCCHKHRRTRAARGTGACAC
10	**MHC1α 3 F2**	GADGGCTTGGAGACCCTCCTYTG
11	**MHC1α3 F3**	TCCTYTGCMRGGYCCACGGC
12	**MHC1α 3 R1**	TCRTGYTCCACRYGGCAYHKGTAG
13	**MHC1TMR**	CACTYRTTGASGCTGCTYTGT

Table 2. Primer pairs used for amplification of MHC class I sequences in this study. Numbers in front of each primer name corresponds to the numbering scheme in Table 1. Sl= sleepy lizard, Py= pygmy blue tongue and Gi= gidgee skink.

Primer			
Forward direction	**Reverse direction**	**gDNA Amplification**	**Species**
1 MHC1α$_1$ F$_1$	3 MHC1α$_1$ R$_1$	Exon2	Sl-Py-Gi
1 MHC1α$_1$ F$_1$	4 MHC1α$_2$ R	Exon2-Intron2	Sl
2 MHC1α$_1$ F$_2$	3 MHC1α$_1$ R$_1$	Exon2	Sl-Py-Gi
2 MHC1α$_1$ F$_2$	4 MHC1α$_2$ R	Exon2-Intron2	Sl
5 MHC1α$_2$ F$_1$	7 MHC1α$_2$ R$_2$	Exon3	Sl-Py-Gi
5 MHC1α$_2$ F$_1$-	8 MHC1α$_2$ R$_3$	Exon3	Sl-Py-Gi
6 MHC1α$_2$ F$_4$	7 MHC1α$_2$ R$_2$	Exon3	Sl
6 MHC1α$_2$ F$_4$	8 MHC1α$_2$ R$_3$	Exon3	Sl
9 MHC1α$_3$ F$_1$	12 MHC1α$_3$ R$_1$	Exon4	Sl-Py-Gi
9 MHC1α$_3$ F$_1$	13 MHC1TMR	Exon4	Sl-Py
10 MHC1α$_3$ F$_2$	12 MHC1α$_3$ R$_1$	Exon4	Sl
10 MHC1α$_3$ F$_2$	13 MHC1TMR	Exon4	Sl-Py
11 MHC1α$_3$ F$_3$	12 MHC1α$_3$ R$_1$	Exon4	Sl
11 MHC1α$_3$ F$_3$	13 MHC1TMR	Exon4	Sl

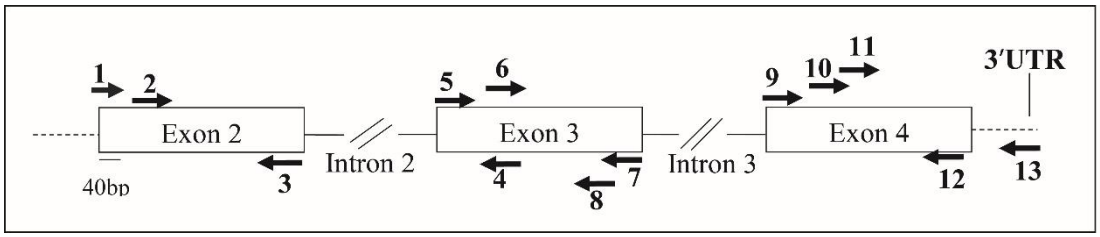


Figure 1. Schematic illustration of part of the MHC class I gene. Numbers above the arrows demonstrates the position of primers used in this study and correspond to the numbering scheme in Table 1.

3.3.4. DATA ANALYSIS

Sanger sequences were edited and assembled (see Table 3 for GenBank submissions) using MEGA v5 (Tamura, et al., 2011); and aligned (both the nucleotide sequences and the translated protein sequences) using Clustal W, as implemented in MEGA v5 (Tamura, et al., 2011) and Geneious v6.1.7 (Drummond, et al., 2012) and then manually checked. We calculated genetic distances between all class I sequences for each exon separately using MEGA v5. Full length amino acid and exon 4 (α - 3 domain) coding sequences derived from *T. rugosa* cDNA and from *T. rugosa*, *T. adelaidensis* and *E. stokesii* gDNA data, were aligned with published sequences used by (Pielou, 1991; Miller, et al., 2006; Glaberman, et al., 2008). We performed Bayesian phylogenetic analyses using MrBayes v3.2 (Ronquist, et al., 2012) using the highly conserved exon 4 nucleotide sequences. Similarly to exon 3, exon 4 is immunoglobulin like in a range of vertebrates. Additionally, we carried out Maximum Likelihood (ML) analysis in RAxML web server (Stamatakis, et al., 2008). We applied the general model GTR + I + Γ in both ML and Bayesian analysis. For Bayesian analyses, we per-formed two independent runs, using the default value of four Markov chains per run. Each chain was run for 10 million generations with a sample frequency of 10,000 and a burn-in period of 100,000 generation. We then measured the effective sample size of each parameter, checked for chain convergence, and visualized the plots using the program tracer v1.5 (Rambaut and Drummond, 2009b).

Table 3. Common name, species and GenBank association numbers, for MHC class I sequences used in our study.

Common name	Species	GenBank association number
Sleepy lizard	*Tiliqua rugosa*	KM515947eKM515965
Great reed warbler	*Acrocephalus arundinaceus*	cN3 AJ005503
Galápagos Marine Iguana	*Amblyrhynchus cristatus*	Amcr-UB*01EU604308, Amcr-UB*02 EU604309, Amcr-UB*03 EU604310, Amcr-UB*0401 EU604311, Amcr-UB*0402 EU604312
Axolotl	*Ambystoma mexicanum*	Amme-3 U83137
Ameiva lizard	*Ameiva ameiva*	LC1 M81094, LC5 M81095, LC13 M81096, LC25 M81097
Mallard	*Anas platyrhynchos*	Du2 AB115242
Green anole	*Anolis carolinensis*	LOC100551708 XM_003230649.1, LOC100559582 XM_003218095.1
Galápagos Land Iguana	*Conolophus subcristatus*	Cosu-UB*0101 EU604313, CosuUB*0102 EU604314, Cosu-UB*02 EU604315, Cosu-UB*03EU604316
Zebrafish	*Danio rerio*	Dare-UBA NM131471
Gidgee skink	*Egernia stokesii*	(Fig. S3)
Racerunner	*Eremias multiocellata*	KJ143572
Chicken	*Gallus gallus*	B-F10 X12780
Human	Human	HLA-B7 U29057, HLA-Cw D50852, HLA-AU07161
Green Iguana	*Iguana iguana*	Igig-UB*0101EU604317, Igig-UB*0102 EU604318, Igig-UB*02 EU604319
Wallaby	*Macropus rufogriseus*	Maru-UB*01 L04952
Mouse	*Mus musculus*	H2-K L36312, H2-D1 NM_010380, H2-Q1 NM_010390, H2-Q10 NM_010391
Water snake	*Nerodia sipedon*	SC1 M81099
Rainbow trout	*Oncorhynchus mykiss*	OnmyUBA AF287487
Platypus	*Ornithorhynchus anatinus*	Oran2-1 AY112715
Chinese soft-shelled turtle	*Pelodiscus sinensis*	AB185243
Tuatara	*Sphenodon punctatus*	Sppu-U*01 DQ145788, Sppu-U*02 DQ145789
Pygmy bluetongue lizard	*Tiliqua adelaidensis*	(Fig. S3)
Possum	*Trichosurus vulpecula*	Trvu-UB AF359509
African clawed frog	*Xenopus laevis*	UAA-1f L20733

3.4. RESULTS

3.4.1. ISOLATION AND CHARACTERIZATION OF MHC CLASS I SEQUENCES FROM T. RUGOSA

The GS-FLX sequencing runs produced a total of 626,996 reads representing 184,465,409 bases. Assembly yielded 43,805 contigs with 63,200 reads remaining as singletons. Basic assembly metrics for contigs greater than 100 bases are shown in Table 4. Blast searches identified several sequences that were significantly similar to MHC class I, however examination of the assemblies indicated that these consensus sequences were comprised of a number of distinct sequences. When the reads were mapped back to a representative contig, coverage of the putative antigen binding exons was not only disproportionately high but many of the reads could only be mapped to this region with extensive hard sequence clipping. Clustering of sequences extracted from this region, produced 45 clusters composed of 1-106 sequences (mean 9.8). Alignment of the consensus sequences from each cluster identified five major sequence groups based on similarity. Iterative assembly using a representative sequence from each sequence group recovered nine sequences (two sequences from four of the groups and a single sequence from the fifth).

Eight of these sequences contained approximately full-length MHC class I sequences encompassing the entire coding region as well as parts of 5′ and 3′ UTRs (Fig. 2). Consequently, we identified four different loci (*MhcTiru-UA*01* _ *MhcTiru-UA*04*) each with two distinct alleles (Fig. 2). The remaining sequence (*MhcTriu-UA*05*) was only partially recovered and composed of

exon 3 through to the 3′ UTR. Interestingly, this sequence would encode shortened transmembrane/cytoplasmic regions (Fig. 2).

Polyadenylation signal sequences (AAAAAA in *MhcTiru-UA*01*; AAGAAA in others) indicating complete 3′ UTR regions were found in the cDNA sequences from the four class I *T. rugosa* loci.

There was greater than 95% nucleotide similarity between al-leles within a locus contrasting with approximately 3% nucleotide identity among loci for the entire region. *MhcTiru.-UA*03* and *MhcTiru.-UA*04* with 70% Identity had the highest nucleotide identity across all loci.

Table 4. Basic assembly metrics for the cDNA library sequences from *Tiliqua rugosa* used in the study.

Metric	**Count**
Number of contigs with>100 bases	39850
Total bases in contigs	19028071
Number of contigs >= 1kbp	2836
Total bases in contigs >= 1kbp	4186544
Max contig length	6528
N50	563

3.4.2. MHC CLASS I VARIATION WITHIN T. RUGOSA

Genomic class I genes from *T. rugosa* were successfully amplified using specific primers designed based on the cDNA sequences. Primers MHC1α_1F$_2$ and MHC1$\alpha_1\alpha_2$ R (Table 1), amplified a 242 bp fragment of exon 2, including most of the coding region plus more than 600 bp of intron2. In total, three different exon 2 sequences including some identical to the cDNA sequences were amplified from *T. rugosa* genomic DNA by using three different primer pairs (Table 2). The primers MHC1α_2 F$_1$ with MHC1α_2 R$_2$, and MHC1α_3 F$_1$ with MHC1α_3 R$_1$ amplified what appear to be a single copy of 274 bp and 276 bp of the exon 3 and 4 of class I, respectively. These products span the majority of exon 3 and 4. The resultant amplicons closely matched the cDNA sequences for loci *MhcTiru-UA*03* and *MhcTiru-UA*04* (Fig. S1). Examination of the Sanger sequences resulting from the amplification of family groups for the three tested primer pairs showed that each pair amplified single loci corresponding to *MhcTiru-UA*03* and *MhcTiru-UA*04* of the cDNA identified loci for exon 2 and 3, respectively (Fig. S2). Based on the analysis of family groups the loci appeared to be inherited in a Mendelian fashion.

3.4.3. COMPARISON OF T. RUGOSA CDNA MHC CLASS I WITH THAT OF OTHER SPECIES

Percent nucleotide identity of the *T. rugosa MhcTiru-UA*0101* sequence with a range of published class I DNA sequences (Table 5) ranged from 50% in Rainbow trout (*Oncorhynchus mykiss*), to 70%nucleotide identity for the Ameiva lizard sequences. The α2 and α3 nucleotide data separately are

most similar to the iguana DNA sequences, but the α - 1 is more similar to the Ameiva lizard (70%nucleotide DNA identity) than it is to the iguana sequences (60-63% nucleotide identity).

We identified many of residues that are important for MHC class I structure and the binding of antigenic peptides in translated *T. rugosa* class I sequences. There are salt bridges formed by histidine (H) residues at positions 5 and 101 and aspartic acid (D) residues at positions 33, and 127. The four cysteine (C) residues at positions 109, 173, 212 and 268 are involved in the formation of intra-domain disulfide bridges (Fig. 2). In addition, the MHC class I genes of most vertebrates encode an NQS or NQT glycosylation site towards the end of the α1 domain. The NQS sequence was present at positions 93-95 in most *T. rugosa* sequences, with just *MhcTiru-UA*0201* and *MhcTiru-UA*0202* containing an NQT glycosylation site (Fig. 2).

Classical class I molecules contain nine highly conserved amino acid positions that bind the C-and N-terminal residues of antigenic peptides (Kaufman et al., 1994; Shum et al., 1999). These highly conserved amino acid positions match with the following types of amino acids and alignment positions: Y9, Y66, R91, F131, T151, K154, W155, Y168 and Y183, in class I sequences of *T. rugosa* (Fig. 2). Phylogenetic reconstruction using exon 4 nucleotide sequences supported the monophyletic clustering of the squamate class I sequences consistent with a common ancestral locus (Fig. 3). There was strong support for the basal position of *T. rugosa* class I sequences relative to other squamate class I sequences included in this analysis. However, many of the class I gene clusters among vertebrates had low to moderate bootstrap

values support, and therefore the evolution of the MHC region and also the relationship among class I genes of all available vertebrates was not well resolved.

Table 5. Percent nucleotide identities between MhcTiru.-U*0101 and published MHC class I sequences from other vertebrates. α1= α-1 (exon 2), α2= α-2 (exon 3) and α3= α-3 (exon 4) of class I gene.

Species	**Percentage nucleotide identity with Sleepy lizard *MhcTiru.-UA*0101***			
	α1	α2	α3	α1, α2, α3
Ameiva lizard (LC1)	70.29	74.1	65.58	69.99
Green Anole (loc100559582)	59.1	71.74	68.11	66.3
Marine Iguana (Amcr-UB*01)	62.68	76.1	89.14	73.94
Land Iguana (Cosu-UB*03)	61.96	75.36	65.94	67.75
Green Iguana (Igig-UB*02)	59.78	69.2	65.58	64.85
Tuatara (Sppu-U)	67.39	66.3	54.94	62.45
Chicken (B-F10)	63.70	65.58	56.52	61.92
Wallaby (Maru-UB*01)	61.95	63.8	52.69	59.26
Possum (Trvu-UB)	65.58	63.1	49.46	59.14
Human (HLA-A)	59.78	64.1	51.1	58.15
Platypus (Oran2-1)	63.77	70.97	50.54	61.53
Warbler (Acar-cN3)	56.3	61.42	57.25	57.66
Axolotl (Amme-3)	64.44	65.22	45.52	58.3
Mouse (H2-K)	60.51	63.8	52.17	58.63
Nurse shark (Gici-UAA)	60.37	58.1	38.1	52
African clawed frog (UAA-1f)	53.70	53.76	43.84	50.42
Rainbow trout (Onmy-UBA)	47.62	57.71	45.39	50.24

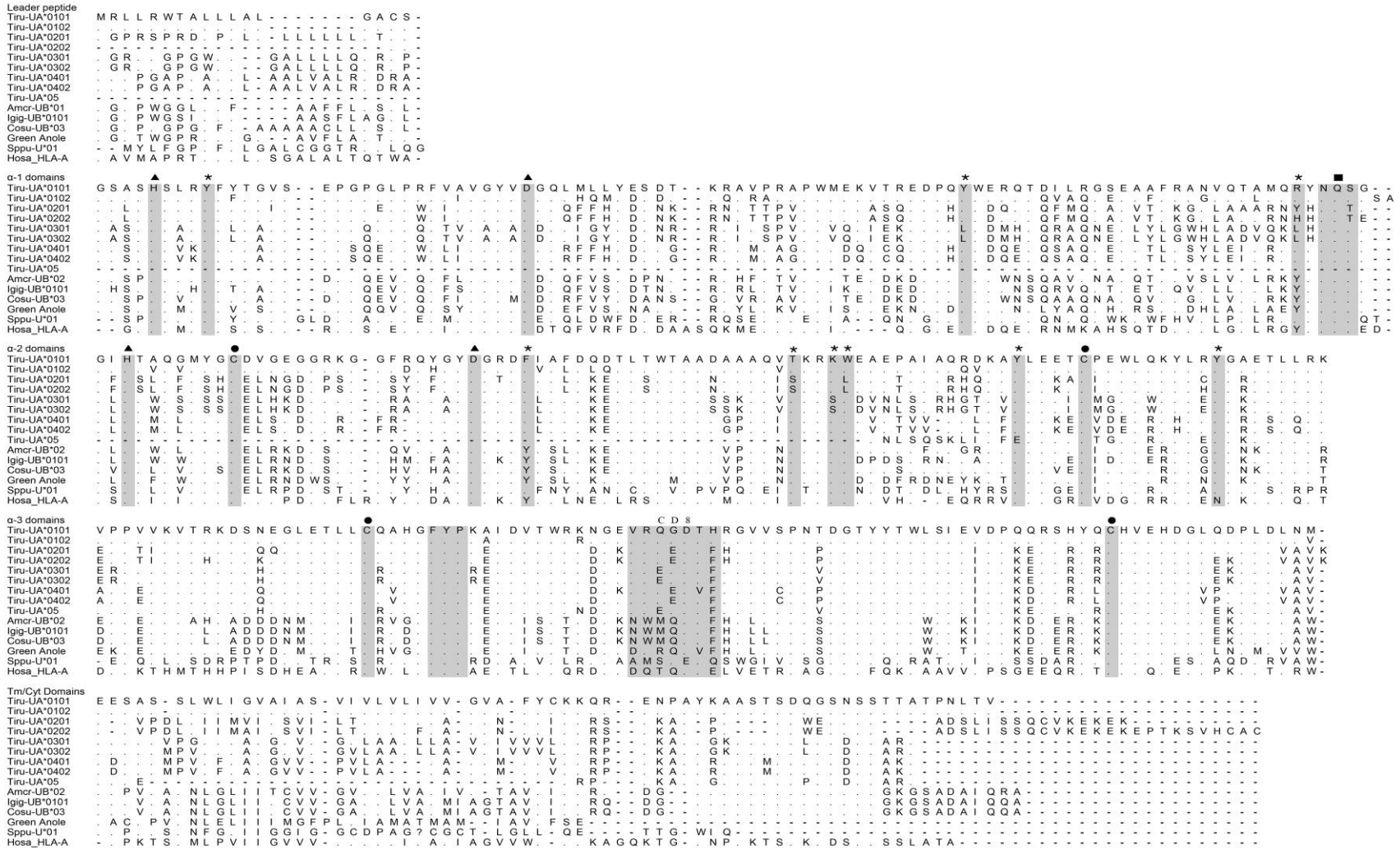


Fig. 2. MHC class I amino acid alignment among *Tiliqua rugosa* and other vertebrates. Coding domains separations are based on Koller and Orr (1985). Shaded amino acid residues show positions which are conserved or have expected functions. Asterisks = conserved peptide-binding residues of antigen N and C termini, triangle = salt bridge-forming residues; circles = disulphide bridge-forming cysteines; squares = N-glycosylation site; CD8 = expected cd8 binding site. Dots indicate identity with the *T. rugosa Tiru-UA*0101* sequence and dashes indicate gaps.

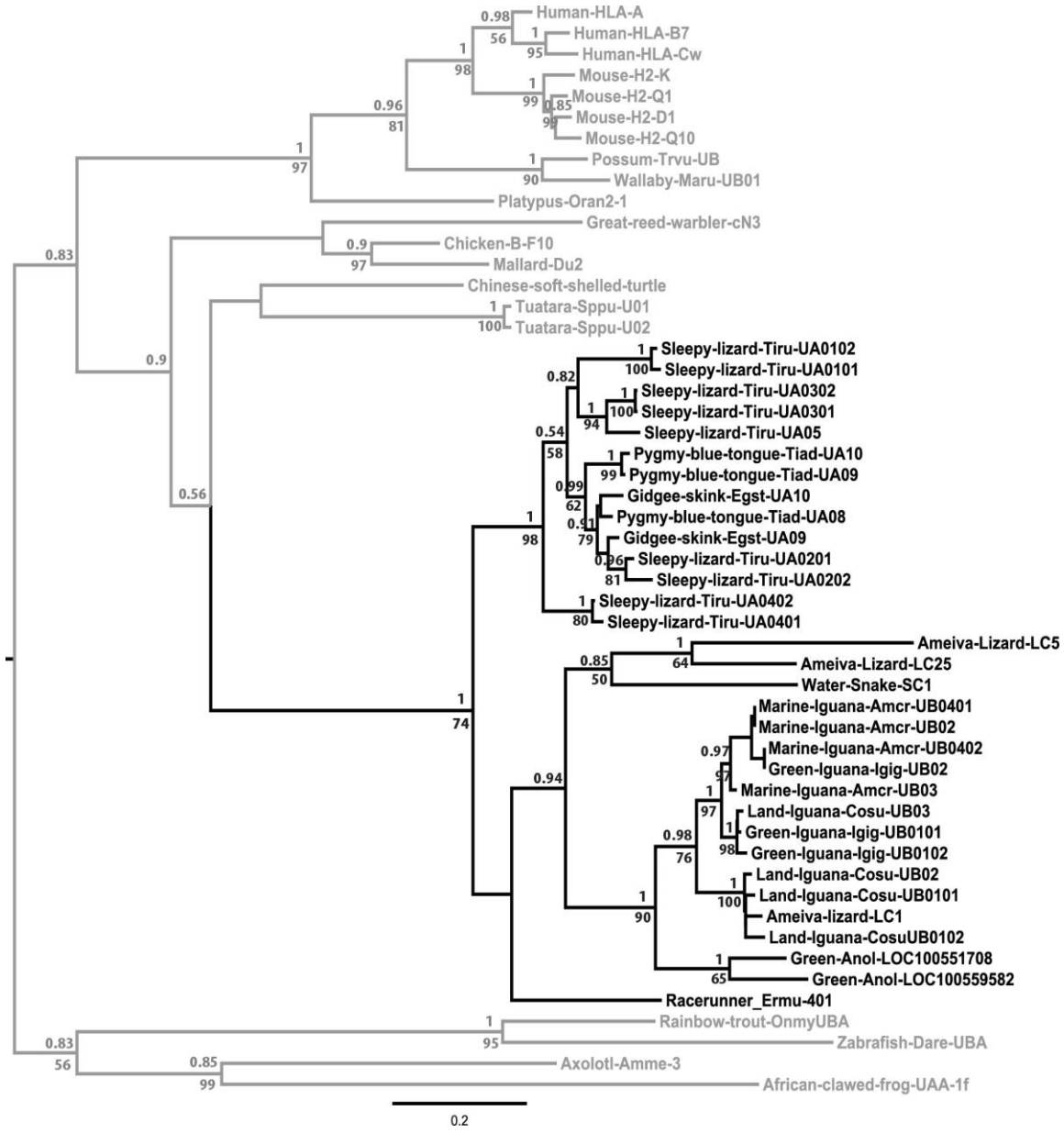


Fig. 3. Phylogenetic relationship of the major vertebrate groups based on published exon 4 sequences. The class I sequences from squamata are shown in black. The tree is midpoint rooted and values in branches are indices of support for the major branches. The top and bottom numbers are Bayesian posterior distribution values, and bootstrap values greater than 50% from neighbour joining analysis, respectively.

3.4.4. GENOMIC SEQUENCES IN RELATED TILIQUA AND EGERNIA GROUP TAXA

Part of the genomic MHC class I spanning exons 2 through 4 was successfully amplified and sequenced from *T. adelaidensis* and *E. stokesii*, using primer pairs specified in Table 2. The primer pairs MHC1α_1 F_1/MHC1α_1 R_1, and MHC1α_1 F_2/MHC1α_1 R_1, for example, amplified 200 bp and 150 bp of the exon 2 from *T. adelaidensis* and *E. stokesii*, respectively. Amplification of the exon 4 region from *E. stokesii* genomic DNA using primer pair MHC1α_3 F_1/MHC1 α_3 R_1 and MHC1 α_3 F_1/MHC1 α_3 R_3 yielded a 170 bp and 200 bp sequence respectively (with more than 90% nucleotide identity) from one individual. Three sequences, which spanned all of the 270 bp of the exon 4 region and the 100 bp of the exons encoding the TM/Cyt domains, were obtained from one individual of *T. adelaidensis* using three different primer pairs, MHC1 α_3 F_1/MHC1 α_3 R_1, MHC1 α_3 F_1/MHC1 TMR and MHC1 α_3 F_2/MHC1 TMR. Exon 2 and 3 of the *T. adelaidensis* and *E. stokesii* sequences shared about 70% nucleotide identity with *MhcTiru.-UA*0101*. However, the exon 4 sequences are the most conserved with more than 78% nucleotide identity between the three species (Fig. S3). Homology of genomic sequences isolated from *T. rugosa*, *T. adelaidensis* and *E. stokesii* to class I MHC genes was confirmed by protein alignments with class I sequences from other vertebrates including the sequences isolated from cDNA library of *T. rugosa*. The protein alignments of amplified sequences from *T. adelaidensis* and *E. stokesii* with *MhcTiru-UA*0101* are presented in Appendix (Fig. S3). Four cysteine residues were present at positions 103, 166, 205 and 261 and two salt

bridges at positions H5 - D31 and H95 - D120, and the FYP motif at positions 210 - 212. An NQS or NQT glycosylation site, which we identified in cDNA sequences from *T. rugosa*, were present at position 89 - 91.

We also confirmed the homology of these genomic sequences by Bayesian and maximum likelihood analyses, which included class I genomic sequences isolated from *T. adelaidensis* and *E. stokesii*, including exon 4 data (Fig. 3).

3.5. DISCUSSION

MHC class I sequences were characterised and isolated from *T. rugosa* using next generation sequencing methods on total RNA from spleen and thymus. cDNA sequences reconstructed from read assembly allowed for the design of specific primers to amplify fragments corresponding to class I exons 2e4 from *T. rugosa* genomic DNA. The primers developed also amplified MHC class I genomic fragments from *T. adelaidensis* and *E. stokesii*. The sequences appear to represent functional genes due to the absence of any premature termination signals and the presence of conserved residues that contribute to the protein domains that form a complete class I molecules. This includes the conserved amino acid residues that bind the terminal ends of antigenic peptides for presentation to CD8+ T cells (Glaberman, et al., 2008).

The presence of sequences encoding the transmembrane and cytoplasmic domains within the coding region is consistent with the gene products being expressed as typical full length cell surface molecules, as are all classical MHC class I. Further evidence that the loci are classical class I loci is provided by

our identification of specific residues at the peptide-binding region of class I sequences (Kaufman, et al., 1994). Confirmation of ubiquitous expression, and high polymorphism are required to be more confident the loci are classical class I. One locus, *MhcTiru-UA*05*, did not contain full transmembrane sequences and was expressed at a low frequency in the studied tissue, indicating that it may not represent a classical locus. The cDNA and genomic sequence data suggests the present of at least four class I loci in *T. rugosa*. Amplification of three different sequences from exon 4 in *T. adelaidensis* is also consistent with the presence of at least two class I loci in this related species. However, as we just amplified a maximum of two genomic sequences per exon from *T. adelaidensis* and *E. stokesii*, the number of class I loci for these two species cannot be currently determined.

Our analysis of exon 4 sequences indicates that *T. rugosa*, *T. adelaidensis* and *E. stokesii* share ancestral polymorphisms (Fig. 3). Shared polymorphisms were observed also between the two *Tiliqua* species for exons 2 and 3. The pattern of our results, where class I sequences are more related among species than within them, is not consistent with the birth and death mode of evolution (Nei, et al., 1997; Glaberman and Caccone, 2008).

Tiliqua and *Egernia* are estimated to have diverged at least 15 million years ago (Estes, 1984). This is the time frame in which orthology begins to be lost for MHC class I genes in the mammals (Nei, et al., 1997). Therefore it may not be unreasonable to consider that the α3 domain has resulted from separately evolving class I lineages in each of the three studied species. However, the observed shared polymorphism could be due to shared ancestral

polymorphism. After species divergence, multiple allelic lineages could be maintained and subsequently this ancestral polymorphism could lead to the current sharing of haplotypes. Alternatively, the shared polymorphism between exon 2 and exon 3 sequences could also have resulted from the convergent evolution of amino acid residues, as suggested in mammalian MHC (Kriener, et al., 2000). Further analysis with a larger number of species and individuals would be required before any firm conclusions could be drawn on the presence of shared ancestral polymorphism.

Generally, MHC class I genes are expressed on all nucleated cells and play a crucial role in expression of peptides generated from intracellular pathogens. Pathogens and parasites are a significant extinction threat to species through increasing mortality (Lyles and Dobson, 1993; Laurenson, et al., 1998; Murray, et al., 1999; Lafferty and Gerber, 2002). In reptiles, susceptibility to a wide range of potentially pathogenic intracellular agents has been reported (Kaschka-Dierich, et al., 1979; Kollias, 1984; Soldati, et al., 2004). *Egernia stokesii*, for example, is parasitized by *Plasmodium circularis* n. sp., and *Plasmodium mackerrasae* (Telford and Stein, 2000). Both T. *rugosa* and *E. stokesii* can be infected by *Hemolivia mariae* (Godfrey, et al., 2006). Characterization of the MHC region will improve our knowledge of the interaction between disease resistance and immune systems and be of benefit to further investigation of intracellular diseases in *Tiliqua* and *Egernia* (e.g. association between sustainability to pathogens and MHC class I polymorphism). In addition, availability of MHC sequences will form the basis of future intraspecific studies examining diversity across the range of *T. rugosa*.

Overall, our study suggests that of all the published sequences, class I sequences from *T. rugosa*, *T. adelaidensis* and *E. stokesii* demonstrate the highest similarity with sequences from iguanine and ameiva lizards. The exon 4 sequence data from major vertebrate groups alone indicates that sequences within reptilian orders cluster together and confirms the monophyly of all available squamate class I sequences. According to the basal position of *T. rugosa* class I sequences in the phylogenetic tree, it appears that this gene diverged very early in the evolution of the squamata.

3.6. ACKNOWLEDGEMENTS

Funding was received from the Sir Mark Mitchell Research Foundation, and support from the South Australian Museum (SAM) staff for helping to source materials, gratefully acknowledged. We thank Dr. Alison Fitch for lab work help, and Dr. Mehregan Ebrahimi for his assistance with improvement of manuscript figures and charts. The research was conducted according to the guidelines of Flinders University Animal Welfare Committee (approval no. E324).

3.7. APPENDIX

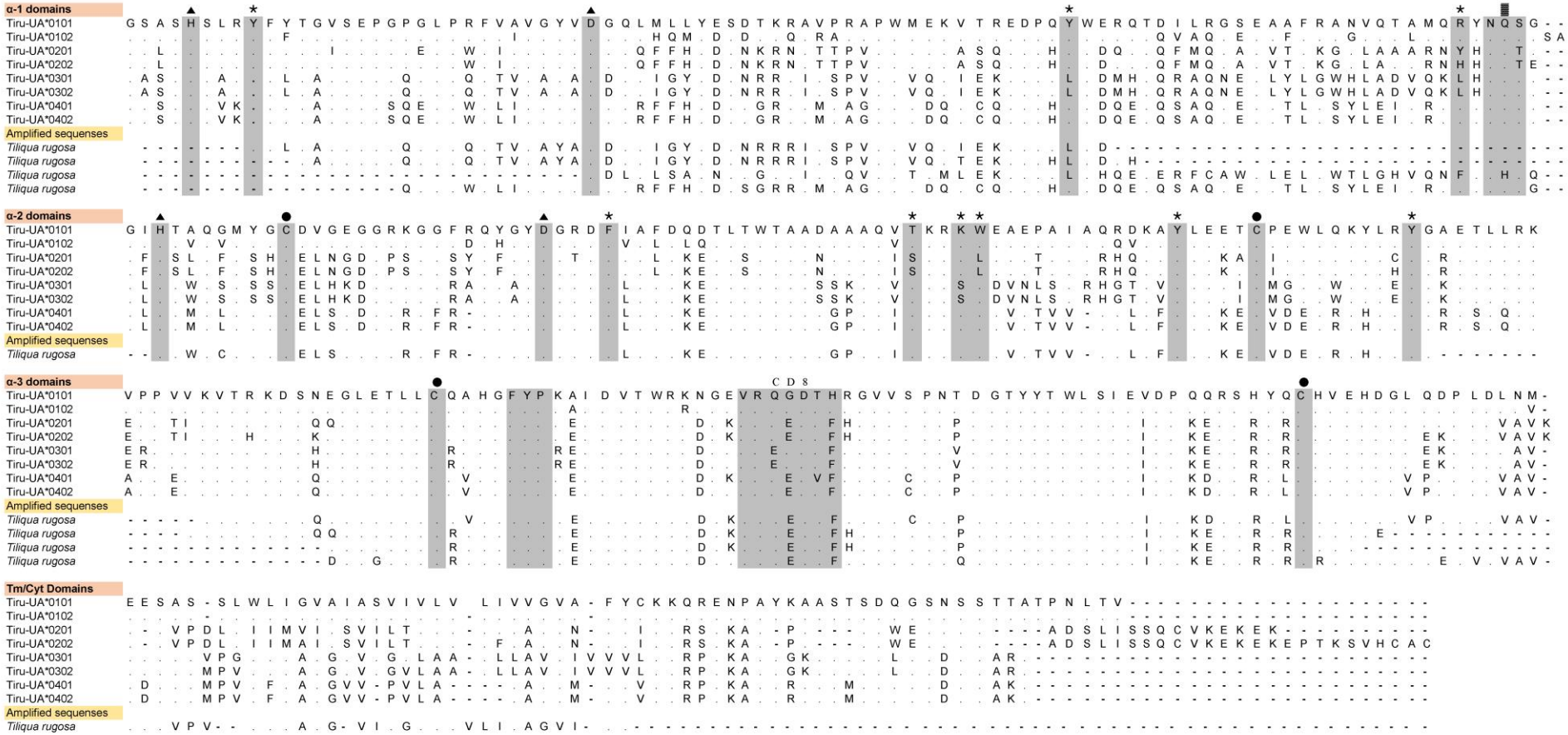


Fig. S1. Amino acid alignment of amplified genomic sequences and isolated cDNA sequences from *T. rugosa*. The conserved sites or positions have expected to be functional highlighted as described in Fig. 2.

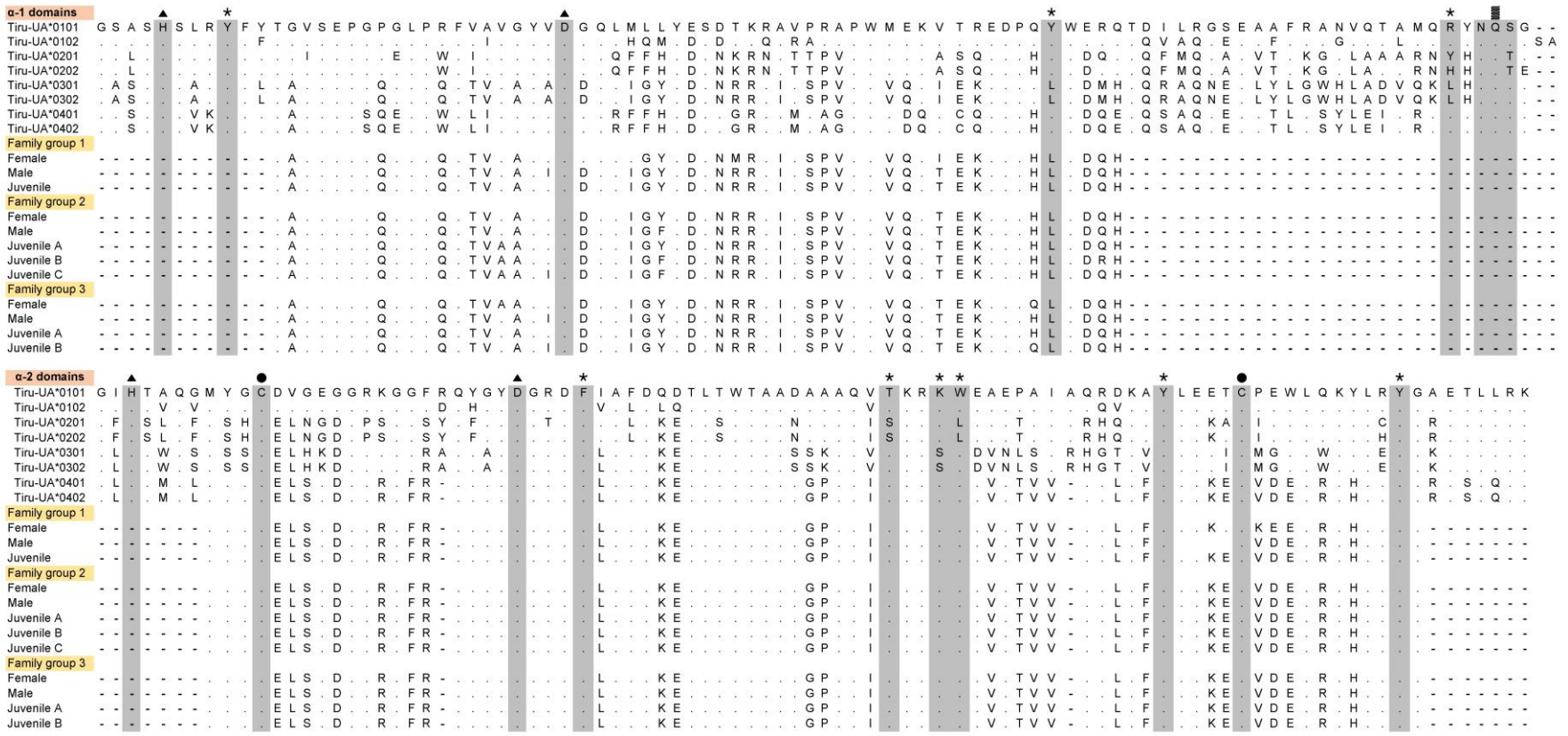


Fig. S2. Amino acid alignment of genomic sequences amplified from three family groups and isolated cDNA sequences from *T. rugosa*. The conserved sites or positions have expected to be functional highlighted as described in Fig. 2.

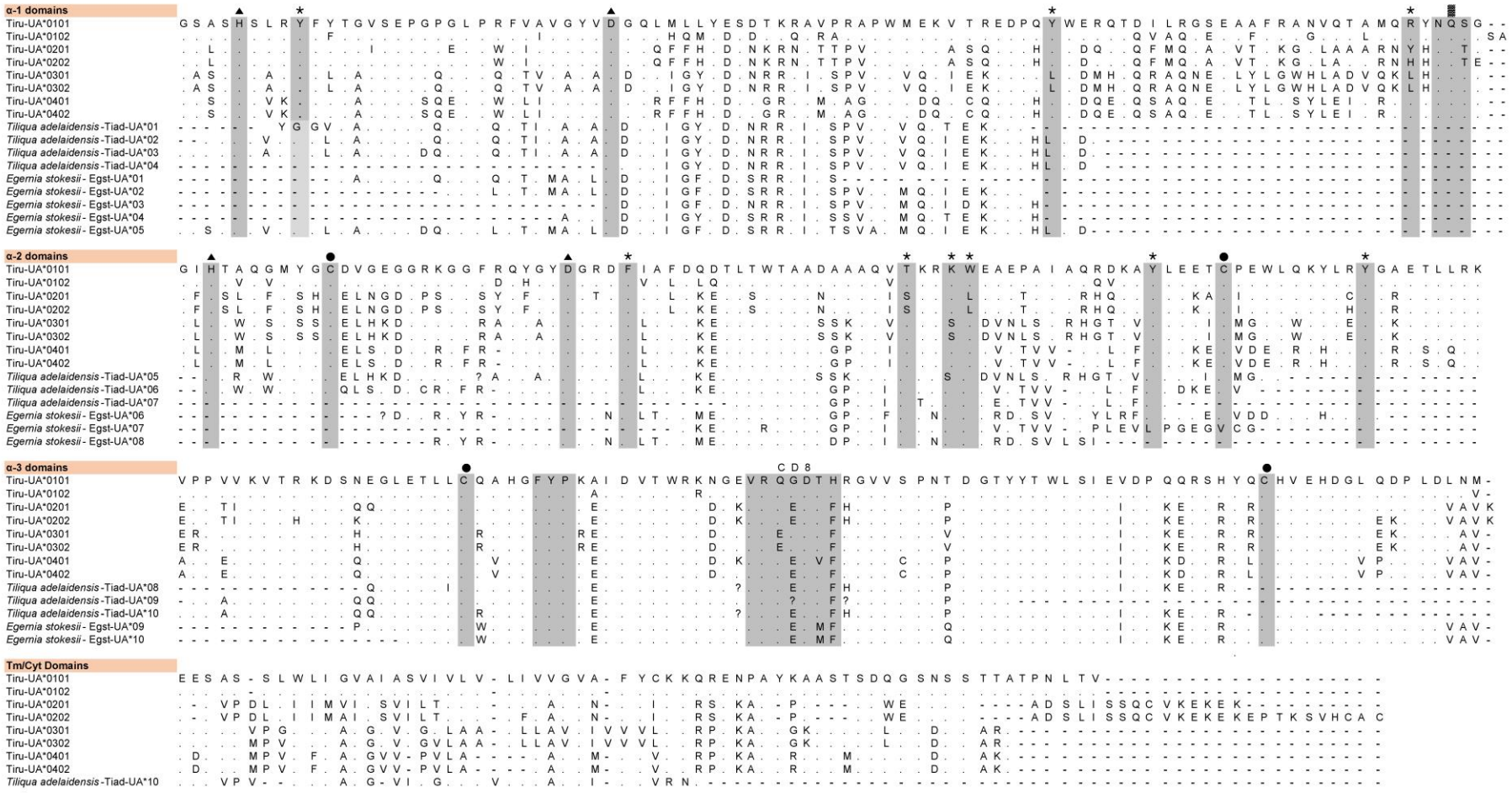


Fig. S3. Amino acid alignment of genomic sequences amplified from *Tiliqua adelaidensis* and *Egernia stokesii* with isolated cDNA sequences from *T. rugosa*. The conserved sites or positions have expected to be functional highlighted as described in Fig. 2.

CHAPTER FOUR

4. THE MAJOR HISTOCOMPATIBILITY COMPLEX CLASS II

This chapter is formatted for submission to the journal Developmental and Comparative Immunology.

This chapter continues the characterization of major histocompatibility complex (MHC) genes that occurred in Chapter 3. Here I develop a functional marker set for class II β chain. Class II loci are involved in the immune response to extracellular pathogens complimenting the intracellular function of class I loci. Having both sets of loci will provide perspectives on the interaction of the

host lizards and their pathogens plus mate choice in wild populations, as well as providing markers for studying genetic diversity under natural selection as opposed to the neutral markers I developed in Chapter 2.

However, the considerable time it took to fully characterize the MHC meant that there was not enough time to utilize these loci in the context as they were aimed within this thesis. Even so they represent a substantial investment regards the evolution of this important gene region in squamates and provide insights into the influence of MHC loci specific to the *Egernia* group.

Developing the MHC class II in current project, therefore, will offer an excellent resource to assess population genetic variation under different environmental conditions as well as extracellular pathogens forces.

Prepared for Developmental and Comparative Immunology

Characterisation of class II MHC genes in a monogamous Australian skink *Tiliqua rugosa*

Talat Hojat Ansari, Terry Bertozzi, Robert D. Miller and Michael G. Gardner

4.1. ABSTRACT

The major histocompatibility complex (MHC) is a gene-dense and highly polymorphic region of the vertebrate genome that plays a crucial role in immune responses. It has been a focus of investigation on the influence of natural selection on genetic diversity in populations and for understanding the genetic pattern of host resistance to pathogens. However, MHC data is largely lacking in many vertebrate taxa, particularly reptiles. Here, we present characterisation of MHC class II loci from *Tiliqua rugosa,* an Australian representative of a major reptile group, the squamates. We describe three unique full-length sequences isolated from *T. rugosa* cDNA and a fourth one, covering exon 2, from genomic DNA. Analyses of the exon 2 and 3 sequences amplified from family groups is consistent with Mendelian inheritance of these loci, supporting their validity. In addition, we present MHC class II exon 2 and 3 genomic sequences from two phylogenetically related species, *Tiliqua adelaidensis* and *Egernia stokesii*, indicating the utility of our designed primers. The characterised loci will form the basis of further studies on mate choice and immune systems within this interesting group of social lizards.

Keywords: MHC class II, Squamates, cDNA, genomic DNA, *Tiliqua rugosa, Tiliqua adelaidensis* and *Egernia stokesii*.

4.2. INTRODUCTION

The major histocompatibility complex (MHC) encodes cellular proteins involved in the immune responses of all vertebrates. The MHC is the most gene-dense and polymorphic region of vertebrate genome (Hedrick, 1994; Dong, et al., 2013). The characterisation of MHC variation has attracted attention given its dual importance in disease resistance and reproductive success (Benacerraf and Germain, 1978; Brown and Eklund, 1994; Penn and Potts, 1999; Milinski, 2006; Havlicek and Roberts, 2008). Variation in genes of the MHC region are thought to be maintained by selective pressure from rapidly evolving pathogens (Wegner, et al., 2003; Sommer, 2005b; Kloch, et al., 2010; Xu, et al., 2011), however there is a non-mutually exclusive role played by disassortative mate choice (Penn and Potts, 1999; Penn, 2002). Indeed some investigators have proposed that the primary role of the MHC was originally a kin-recognition system and its role in immune defence was secondary (Boehm and Zufall, 2006) although this postulation remains highly contentious (Janeway, et al., 1999; Sommer, 2005a). Therefore, these genes are ideal candidates for studying the action of natural selection, the interaction of disease resistance and mate choice in wild populations (Vigueira, et al., 2013a), especially in social animals.

The genes located within the MHC are classified into three main groups: class I, II, and III. (Klein, 1986). Of these, both MHC class I and II molecules

are involved in presentation of antigenic molecules primarily to T cells. While class I molecules are expressed on the surface of most nucleated cells, MHC class II expression is limited to immune cells and antigen presenting cells. MHC class II molecules primarily present antigenic peptides derived from the processing of extracellular pathogens such as bacteria (Strandh, et al., 2011). Molecules of Class II are heterodimers of an α and a β chain which are encoded by separate genes in the MHC (Glaberman, et al., 2009; Xu, et al., 2011). The highly polymorphic sites within class II are the peptide- binding region (PBR) positions within the α_1 and β_1 domains (Xu, et al., 2011). The α_1 and β_1 domains are encoded by exon 2 of the α and β chain genes, respectively. Of the two domains making up the PBR, the β_1 domain has the most direct interaction with the foreign peptide (Madden, 1995).

Characterising MHC class II loci in a variety of species improves our knowledge of the influence immunogenetics has on disease resistance, mate choice, kin recognition, and sociality. A vertebrate lineage for which there is a paucity of MHC information are the squamate reptiles (Kelley, et al., 2005; Miller, et al., 2005; Glaberman, et al., 2009). This contrasts with squamate species' abundance as the order squamata, with nearly seven thousand species, and more than all mammals, is the most diverse reptile order (Vitt, et al., 2003). In addition, the ability of squamate reptiles to use chemical communication for prey and predator detection as well as mate choice (Mason and Parker, 2010) make squamates a fascinating group for MHC studies. Squamates also provide emerging models for understanding the evolution of sociality (Gardner et al., 2015).

One such emerging model is the *Egernia* group of lizards (Gardner, et al., 2015; While, et al., 2015). This group includes species with contrasting life histories. For example, the Australian sleepy lizard *Tiliqua rugosa,* which is a monogamous, viviparous and long lived Australian skink, distributed widely in arid to semiarid regions mostly in the southern mainland of Australia (Bull, 1987; Greer, 1989). A close relative to the sleepy lizard, the endangered pygmy bluetongue lizard, *Tiliqua adelaidensis,* is a promiscuous species that survives in a few isolated fragments on native grassland in the Burra region in mid-South Australia (Schofield, et al., 2014). Another member of the *Egernia* group , the monogamous gidgee skink *Egernia stokesii,* is a group-living species found primarily in rocky outcrops in arid to semi-arid regions of Australia (Gardner, et al., 2001). These three related squamate species provide contrasting life-histories for comparative studies on the role MHC can play in the evolution of sociality and mate-selection.

As a first step to understanding the differences that these contrasting life histories might have on MHC variation, this genomic region needs to be characterized for these lizards. We recently completed a characterisation of the MHC class I loci in this subset of Australian skinks (Ansari, et al., 2015). In this present study, we aimed to characterise genes encoding the MHC class II β chain of *T. rugosa* from exon 2 and 3 cDNA sequences. In addition we present exon 2 and exon 3 genomic sequences from all three Australian skink species described above. These initial characterisations will benefit future investigations addressing questions about the extracellular diseases in *Tiliqua* and *Egernia* such as the association between diseases resistance and MHC class

II polymorphism, and the possible influence that MHC loci have on mate choice and kin recognition.

4.3. Material and methods

4.3.1. Discovery of MHC class II sequences

This study utilised the same next-generation sequencing data and transcriptome assembly from the same *T. rugosa* individual lizard (SAMR634741) that was previously used to characterise MHC class I loci (Ansari, et al. (2015).

MHC class II amino acid sequences were downloaded from GenBank for the marine iguana (*Amblyrhynchus cristatus*), the caiman (*Caiman crocodilus*) and the tuatara (*Sphenodon punctatus*) (Table 2) and used as BLAST queries. We performed a local BLASTX (2.2.21+) with a significance threshold of e ≤ 10-4 to search the assembled contigs to identify those containing potential MHC class II sequences. We then used a PERL script (blast_parse.pl) and parsed significant matches to the query sequences. The parsed report was imported into a spreadsheet and sorted by High Scoring Pair (HSP) rank (based on e-value) and then by HSP length. Following this we used the split_fasta3_1.pl PERL script to retrieve the contigs of all matches with a HSP rank=1 from the assembled contigs and clustered the sequences using cd-hit-454 (Niu, et al., 2010) with the -M 8000 –g 1 options enabled to remove duplicates.

4.3.2. PRIMER DESIGN

We aligned the cDNA sequences from *T. rugosa* with homologous sequences (Table 2) from marine iguana and tuatara using Geneious v 5.6.4 (Drummond, et al., 2012) and designed primers, focussing on exons 2 and 3 (Figure 1; Table 1), using Primer 3 as implemented in the Geneious software.

4.3.3. AMPLIFICATION OF GENOMIC MHC CLASS II SEQUENCES

Initially we tested the designed primers (Table 1) for exons 2 and 3 on *T. rugosa* DNA sourced from the same individual (SAMR634741) used for the cDNA sequencing and subsequent transcriptome assembly, and from the two related species, *T. adelaidensis* and *E. stokesii* (Ansari, et al., 2015). Following successful amplification and Sanger sequencing, we tested Mendelian inheritance for exons 2 and 3 in *T. rugosa* genomic DNA from each of four family groups. These were comprised of the mother and father plus one or two offspring. PCR was performed in 25µL reactions, containing 20ng DNA, 1× reaction buffer (Applied Biosystems), 0.2µM of each primer, 0.8µM dNTP, 2µM $MgCl_2$ and 0.5 U of DNA polymerase (Ampli Taq Gold™; Applied Biosystems). The cycling profile including a single cycle of 10min at 94°C, followed by 34 cycles of 30s at 94°C, 30s at 60°C and 1.5min at 72°C, with a final extension for 10min at 72°C. Amplicons were visualised on 1.5% agarose gels and reactions resulting in a single amplified product were purified using a multiscreen PCR 384 Filter Plate (Millipore). Where multiple products were present on the agarose gels, we excised the PCR fragment corresponding to the anticipated amplicon size directly from the gel and purified it using the QIAquick Gel Extraction Kit (Qiagen). The purified products were sequenced

using BigDye v3.1 terminator chemistry following the reactions volume and concentration recommended by AGRF protocol (half volume 0.25x). The sequence cycling conditions consisted of 3 min at 96°C, followed by 30 cycles of 30s at 96°C, 15s at 50°C and 4min at 60°C. The products were cleaned up using SEQ 384 Filter Plate (Millipore) and sent to the AGRF (Adelaide Australia) for capillary sequencing on an AB 3730*xl*.

The sequencing products were purified using a SEQ 384 Filter Plate (Millipore) and sent to the AGRF (Adelaide Australia) for capillary sequencing on an AB 3730*xl*.

Where Sanger sequencing identified co-amplification of multiple products of the similar lengths, we attempted to separate these via cloning using the pGEM-T Easy Vector Systems (Promega, Madison, WI, USA) kit. The purified PCR products were ligated into the pGEM vectors and then transformed in *Escherichia coli* competent cells (JM109) following manufactures protocols (Promega, Madison, WI, USA). We identified positive clones with blue/white screening and amplified these using primers for the T7 and SP6 vector promotors, following the PCR condition described above. The resultant amplicons were Sanger sequenced with the T7 primer using the conditions outline above.

4.3.4. DATA ANALYSIS

4.3.4.1. cDNA SEQUENCES

To examine the important residues that are either conserved or functional across vertebrates as described by Kaufman, et al. (1994), we aligned the class II β amino acid sequences from *T. rugosa* with human (*Hosa-DRB1*, *Hosa-DQB1*and *Hosa-DOB*) and available sequences for reptile species from GenBank (see Table 2 for accession numbers), using Clustal W implemented in MEGA v6 (Tamura, et al., 2013) and Geneious v5.6.4 (Drummond, et al., 2012) with manual checking. We then defined the peptide binding residues (PBR), following Brown, et al. (1993) and Stern, et al. (1994). We followed the sequence variation nomenclature of Den Dunnen and Antonarakis (2000).

4.3.4.2. THE GENOMIC SEQUENCES

We used Geneious v5.6.4 (Drummond, et al., 2012) to edit the genomic sequences and align the nucleotide and amino acid translated sequences with that of the cDNA sequences characterised from *T. rugosa*. We then examined nucleotide and amino acid identity among the *T. rugosa* isolated sequences (cDNA and amplified sequences) and calculated the exon 2 and 3 nucleotide and amino acid identity between the cDNA sequences from *T. rugosa,* and genomic sequences from *T. adelaidensis,* and *E. stokesii* using Geneious v5.6.4 (Drummond, et al., 2012).

4.3.4.3. PHYLOGENETIC ANALYSIS

We performed Bayesian phylogenetic analyses for both exon 2 (290bp including alignment gaps) and exon 3 (282bp including alignment gaps) nucleotide sequences using MrBayes v3.2 (Ronquist, et al., 2012). We performed four independent runs, using the default value of four Markov chains per run, applying the general model GTR + I + Γ after this was identified as the best model using jModel Test (Guindon and Gascuel, 2003; Darriba, et al., 2012). Each chain was run for 10 million generations with a sample frequency of 10,000 and a burn-in period of 100,000 generation. We then used the program tracer v1.5 (Rambaut and Drummond, 2009b) to measure the effective sample size of each parameter, check for chain convergence, and visualize the plots.

4.3.4.4. TESTS FOR SELECTION

We performed two tests to determine if the sites within the sequences were under selection. Firstly, we calculated the average number of nucleotide differences per nonsynonymous site (d_N) and per synonymous site (d_S) for the *T. rugosa* cDNA sequences encoded by the β-1 and β-2 domain separately. We used the modified Nei - Gojobori method (Nei and Gojobori, 1986) with Jukes-Cantor correction in MEGA v6 (Tamura, et al., 2013). The relative rate of d_N and d_S substitutions were independently estimated for peptide-binding residues (PBR) and non-peptide-binding residues (non-PBR) in the *T. rugosa* cDNA, exon 2 sequences and also the genomic sequences amplified from family groups. The PBR and non-PBR amino acid positions were defined based on Brown, et al. (1993) and Stern, et al. (1994). Secondly we tested for positive selection

using the codon based Z-test in MEGA, where the P-values were based on the probability of rejecting the null hypothesis of $d_N=d_S$ (Tamura, et al., 2013) with a one-tailed test for the alternative hypotheses of $d_N>d_S$. Standard errors were estimated by 1000 bootstrap replicates.

4.4. RESULTS

4.4.1. IDENTIFICATION AND CHARACTERISATION OF MHC CLASS II SEQUENCES FROM TILIQUA RUGOSA

A total of 626,996 reads representing 184,465,409 bases were produced from GS-FLX sequencing runs using the *T. rugosa* cDNA library. Assembly of the reads resulted in 43,805 contigs with 63,200 reads remaining as singletons. Several sequences were highly similar to MHC class II, were identified from the BLAST search. From these three different, approximately full-length, *T. rugosa* MHC class II sequences were identified containing the entire coding region as well as parts of 5′and 3′ untranslated regions (UTRs). The contigs were 719, 1,092 and 749 nucleotides in length (Figure 2) and differed at both the nucleotide and the translated amino acid sequences (Table 3). As it was not possible to identify if these sequences represent different loci or different alleles, they were named *Tiru-DAB*01* to *Tiru-DAB*03* following the MHC nomenclature rules suggested by Klein, et al. (1990). The three full length coding regions when translated, produced MHC class II β chains of 240 (*Tiru-DAB*01*), 265 (*Tiru-DAB*02*) and 250bp (*Tiru-DAB*03*) amino acids in length. Only the *Tiru-DAB*02* sequence contained what appeared to be a

full length, 297 nucleotide, 3′ UTR that included a polyadenylation signal (ATTAAA).

4.4.2. COMPARISON OF TILIQUA RUGOSA MHC CLASS II WITH OTHER VERTEBRATES

From the amino acid alignment we identified the conserved and functional residues as defined by Kaufman, et al. (1994) (Figure 2). These residues included a pair of cysteines which formed the intra-domain disulfide bridges in both the β-1 and β-2 domain of all presented sequences (Figure 2). The conserved arginine (R) and aspartic acid (D) residues, which form salt bridges (Kaufman, et al., 1994), were present at positions 79 and 83 in the alignment respectively. All sequences contained the residues W68, H88 and N89, with the exception of *Tiru.-DAB*01* which has a phenylalanine (F) instead of tryptophan (W) at position 68. These residues are important for function of peptide binding and the formation of hydrogen bonds with the bound peptide (Brown, et al., 1993).

The RFDS motif, which is a conserved motif for interactions with the CD4 co-receptor facilitating T cell interactions (Auffray and Novotny, 1986), was present in the amino acid alignment at positions 46 to 49 (Figure 2). In the *T. rugosa* sequences the conserved NGT or NGS glycosylation site in the β_1 domain, has different amino acids to all other aligned species with the exception of marine iguana (Fig. 2). At the first position there is a N19D (an aspartic acid, D, instead of asparagine, N at amino acid 19) and in the third position a T21G or S21G (glycine acid, G, instead of a threonine, T, or serine,

S at amino acid 21). The *T. rugosa* sequences also contained a two amino acid insertion after the glycosylation motif compared to the other vertebrates (Figure 2). In addition there were a number of changes in the exon 3 sequences encoding the β_2 domain. Residues here which are known to interact with the CD4 molecules in humans (Cammarota, et al., 1992; Wang, et al., 2001) have a number of substitutions in *T. rugosa* (position 143 to 168, Figure 2). A leucine acid, L167 which forms part of hydrophobic cavity in CD4 binding in mammals (Wang, et al., 2001) is replaced with a glutamine acid, Q167 (Figure 2). However, residues in the Tm domain are implicated in the β and α chain interactions (Cosson and Bonifacino, 1992; Wang, et al., 2001) were conserved in *T. rugosa* (Figure 2).

4.4.3. AMPLIFICATION OF MHC CLASS II FROM TILIQUA RUGOSA GENOMIC DNA

Genomic class II genes from exon 2 and 3 were successfully amplified for *T. rugosa*, using the specific primers designed based on the cDNA sequences for each exon separately (Fig. 1 and Table 1). The sequences were confirmed by direct Sanger sequencing to be class II. All sequences matched the target sequences except for one. This additional class II sequence was amplified with primers MHCIIβ_1F$_1$ and MHCIIβ_1R$_1$ (Fig. 1 and Table 1), spanning 150bp of exon 2 and covering most of the peptide binding region (Fig. S1). This amplified sequence had the highest similarity to the *Tiru-DAB*02* (with 83% nucleotide and 76% amino acid identity). Further amplification with this primer pair, and subsequent cloning and sequencing, determined that this sequence could reliably be isolated in other *T. rugosa* individuals. The sequence

did not have stop codons and included important residues peptide binding sites so we considered it to be a class II sequence (Fig. 2). Due to the difficulty in identifying if the new sequence is from a different locus or is a different allele we named the sequence *Tiru-DAB*04*. Additional sequences were amplified from exon 2, which covered most of the exon and were identical to either *Tiru-DAB*01* or *Tiru-DAB*02*, (Fig. S1 and S2).The combination of forward primer located in the leader peptide (MHCII LPF), with the reverse primer located in in exon 3 (MHCIIβ_2R_1) resulted in multiple products of different sizes (from 300 to more than 1000bp). The cloning and subsequent sequencing of these products resulted in a sequence 280bp in length, including 122bp of intron 2 (the intron between exon 2 and exon 3) and 154bp of exon 3.

We obtained sequence data from the four family groups using the primer pairs MHCIIβ_1F_2 and MHCIIβ_1R_1 for exon 2, and MHCIIβ_2F_1 with MHCIIβ_2R_2 for exon 3. Each of these combinations amplified products consistent with them being Mendelianly inherited single loci corresponding to *Tiru-DAB*01* or *Tiru-DAB*02* (Appendix Figure S2).

4.4.4. AMPLIFICATION OF GENOMIC DNA IN TILIQUA ADELAIDENSIS AND EGERNIA STOKESII

We successfully amplified and sequenced part of the genomic MHC class II regions spaning exon 2 and exon 3 from *T. adelaidensis* and *E. stokesii*. The primer pairs MHCIIβ_1F_1 / MHCIIβ_1 R_1, and MHCIIβ_2F_1 / MHCIIβ_2 R_2, amplified approximately 150bp and 250bp of the exon 2 and exon 3, respectively, from both *T. adelaidensis* and *E. stokesii* (Fig. S1). The amplified

sequences from exon 2 of *T. adelaidensis* and *E. stokesii* had 70% and 64% amino acid identity with *Tiru-DAB*01*, respectively. The percentage identity was lower in exon 3, with 63 and 60 % amino acid identity for *T. adelaidensis* and *E. stokesii*, respectively.

Table 1. Primers designed for amplification of MHC class II genomic sequences in this study. F= forward primer; R = reverse primer. An annealing temperature of 60°C was used for all amplifications.

	Primer name	Primer Sequence in 5´→ 3´direction	Gene
1	MHCII LP-F	AGATGAGGTCTGCTCCAGGC	Exon 1
2	MHCII B1-F1	GAGYWCGYGCGCTTCGACA	Exon 2
3	MHCII B1-R1	WGATKCCRTAGTTGTRCCGGCAG	Exon 2
4	MHCII B1-F2	ACCAGTGGATTGGCGAGTGC	Exon 2
5	MHCII B1-R2	GCTGTCGAAGCGCACGAACT	Exon 2
6	MHCII B2-F1	AAGCCCCAGCTGCAGATCAC	Exon 3
7	MHCII B2-R1	CAGCATCACCTGGATGSRGAAGG	Exon 3
8	MHII B2-R2	CRMGSAGGCTGGDYTGCTCCAC	Exon 3
9	MHCII TM-R	ACCCCYGTCCACATCTTGCT	Exon 4

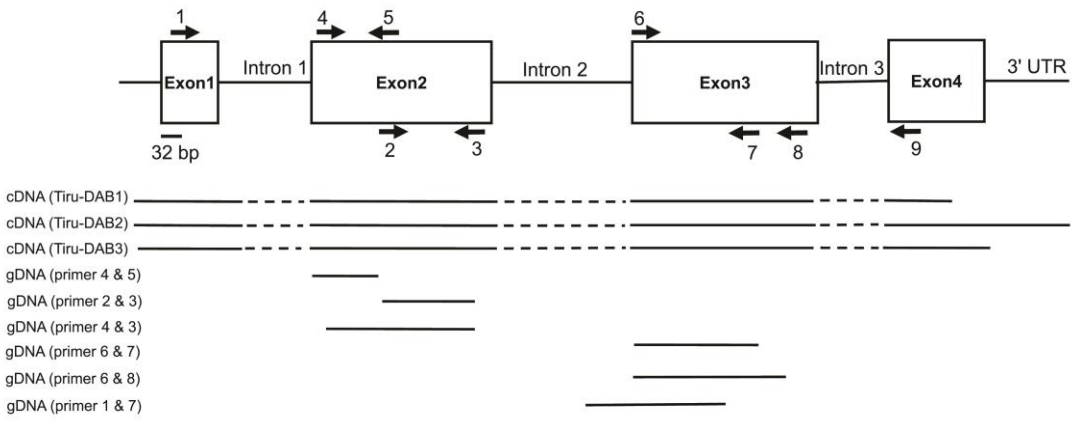


Figure 1. Schematic illustration of MHC class II B gene. Numbers above the arrows show the position of primers designed and used to amplify genomic DNA (gDNA) in

this study, and correspond to the numbering scheme in Table 1. The cDNA sequences demonstrate the region that they covering.

Table 2. Species and their GenBank accession numbers for MHC class II sequences characterised and used in this study.

Species common name	Species scientific name	MHC Class II	GenBank accession numbers
Sleepy lizard	*Tiliqua rugosa*	*Tiru-DAB1-3*	(submitted to GenBank will be supplied upon acceptance)
Pygmy bluetongue lizard	*Tiliqua adelaidensis*	*Tiad-DAB1 &2*	(Fig. S1)
Gidgee skink	*Egernia stokesii*	*Egst-DAB1 &2*	(Fig. S1)
Green anole	*Anolis carolinensis*	*carolinensis-H2*	XM008121064
Marine iguana	*Amblyrhynchus*	*Amcr-DAB1-4*	FJ623746-52
Burmese python	*Python bivittatus*	DR1 & DR	XM007444291-
Massasauga	*Sistrurus catenatus*	*Sica-DAB6 & 18*	HG313840 - LK932131
Caiman	*Caiman crocodilus*	*Croc-B1* & *B2*	AF256651- AF256652
Chinese alligator	*Alligator sinensis*	*Alsi- DLA-DR1*	XM006034608
Tuatara	*Sphenodon punctatus*	*Sppu-DAB1*	DQ124231
Common mallard	*Anas platyrhynchos*	*Anpl*	DQ490139
Great reed warbler	*Acrocephalus arundinaceus*	Acar-cO1	AJ404372
Blue Petrel	*Halobaena caerulea*	Haca-DAB06L	JF276894
Humboldt penguin	*Spheniscus humboldti*	*Sphu-DRB*	AB301947
Kiwi	*Apteryx owenii*	*Apow-DAB02*	HQ639684
Echidna	*Tachyglossus aculeatus*	*Taac-DZB1*	AY288075
Platypus	*Ornithorhynchus anatinus*	*Oran-DZB01*	AY288074
Wallaby	*Macropus rufogriseus*	*Maru-DAB1 & DBB*	M81624- M81625
Cat	*Felis catus*	*Feca-DRB*	U51575
Little brown bat	*Myotis lucifugus*	*Mylu-DRB1*	XM_006108670
Human	*Homo sapiens*	*DRB1, DQB1 & DOB*	M11161- M20432- L29472
Mouse	*Mus musculus*	*Mumu-DQB1*	Mouse_ M13538
Nurse shark	*Ginglymostoma cirratum*	Gici-B1	L20274
Zebrafish	*Danio rerio*	*Brre-DAB1*	L04805
African clawed frog	*Xenopus laevis*	Xela-B3	D13685

Figure 2. MHC class II B gene amino acid alignment among *Tiliqua rugosa* and other vertebrates. Dots indicate identity with the *Tiru-DAB*01* sequence and dashes show gaps. The conserved and functional residues are highlighted with the grey shading. The dashed box shows the NGT glycosylation site and the location of CD4 binding residues (Cammarota, et al., 1992; Wang, et al., 2001) are boxed with a solid line. The peptide binding (PBR) residues (Brown, et al., 1993; Stern, et al., 1994) are shown with an asterisk.

Table 3. Percent nucleotide and amino acid identities among the characterised *T. rugosa* cDNA MHC class II loci. Lp= leader peptide (exon 1), β1= β-1 (exon 2), β2= β-2 (exon 3) and Tm/cyt= transmembrane and cytoplasmic domains (exon 4) of class II gene.

Species	**Percentage nucleotide identity with *Tiru-DAB*01***				**Percentage amino acid identity with *Tiru-DAB*01***			
	Lp	β1	β2	Tm/cyt	Lp	β1	β2	Tm/cyt
Tiru-DAB*02	93.1	89.8	89.5	98.6	93.1	84.8	87.4	100
Tiru-DAB*03	87.36	80.1	90.5	98.6	79.3	67.4	86.3	100

4.4.5. PHYLOGENETIC ANALYSIS

The two Bayesian trees based on exon 2 and 3 nucleotide sequences analysed separately, yielded similar topology of the main branches, but Bayesian posterior probability values (PPV) were lower for many nodes inferences for exon 2 (Fig. 3A and 3B).

Phylogenetic reconstruction using either exon 2 or exon 3 nucleotide sequences supported the monophyletic clustering of all squamate MHC class II β sequences. Within the squamates, the *T. rugosa* MHC class II β sequences formed a monophyletic group with the other *Egernia* group species, *T. adelaidensis* and *E. stokesii*. Diapsida formed a monophyletic group although posterior support for this was low for exon 2 nucleotide sequences (Fig. 3B).

4.4.6. TESTS OF SELECTION

The exon 2 cDNA sequences showed evidence of marginally significant positive selection ($d_N/d_S = 1.017$, $p = 0.054$), while exon 3 showed evidence of purifying selection (Table 4). However, this is just an initial result based on the three sequences developed in this study and would require validating with

a larger sample size. The selection test on the *T. rugosa* sequences from both families and also cDNA sequences indicated strong evidence for positive selection at the PBR in exon 2, but not at the non-PBR (Table 4).

Table 4. The selection analysis of MHC class II in *Tiliqua rugosa*. We present the average number of nonsynonymous substitution per nonsynonymous site (dN) and synonymous substitution per synonymous site (dS) followed by their standard errors (SE). We performed analysis for β-1 and β-2 domains separately, among the described cDNA sequences. The peptide binding (PBR) and non- peptide binding regions (non-PBR), defined following Brown, et al. (1993) and Stern, et al. (1994), were calculated separately for isolated cDNA and family group sequences. Significance P values ($P < 0.05$) are shown by an asterisk. All positions including missing data were eliminated.

Region	Number of codons	$d_N \pm SE$	$d_S \pm SE$	d_N/d_S
		cDNA sequences		
PBR	25	0.508 ± 0.027	0.125 ± 0.024	4.064*
Non-PBR	66	0.138 ± 0.008	0.284 ± 0.014	0.486
β-1 domain all	91	0.238 ± 0.007	0.234 ± 0.009	1.017
β-2 domain	95	0.122 ± 0.005	0.412 ± 0.011	0.296
		Family group		
PBR	15	0.507 ± 0.049	0.103 ± 0.035	4.922*
Non-PBR	40	0.087 ± 0.012	0.167 ± 0.016	0.521

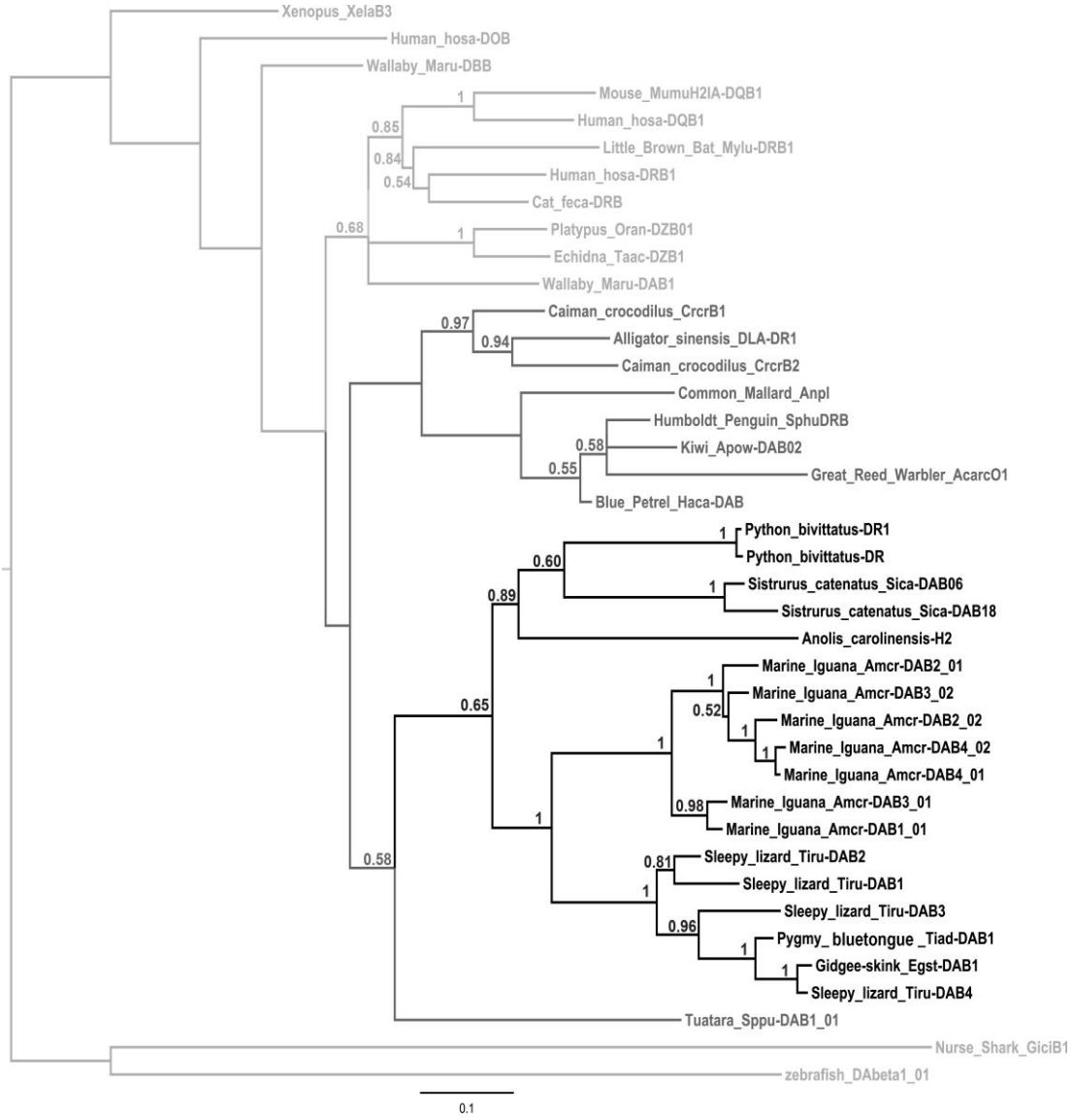


Figure 3A. Bayesian phylogenetic relationship of the major vertebrate groups based on exon 2 nucleotide sequences. The class II B sequences from Squamata are shown with black and all other Diapsida (reptiles includes birds and crocodilians) with dark grey. The values in branches are index of support for the major branches (Bayesian posterior distribution value) and values greater than 50% are represented.

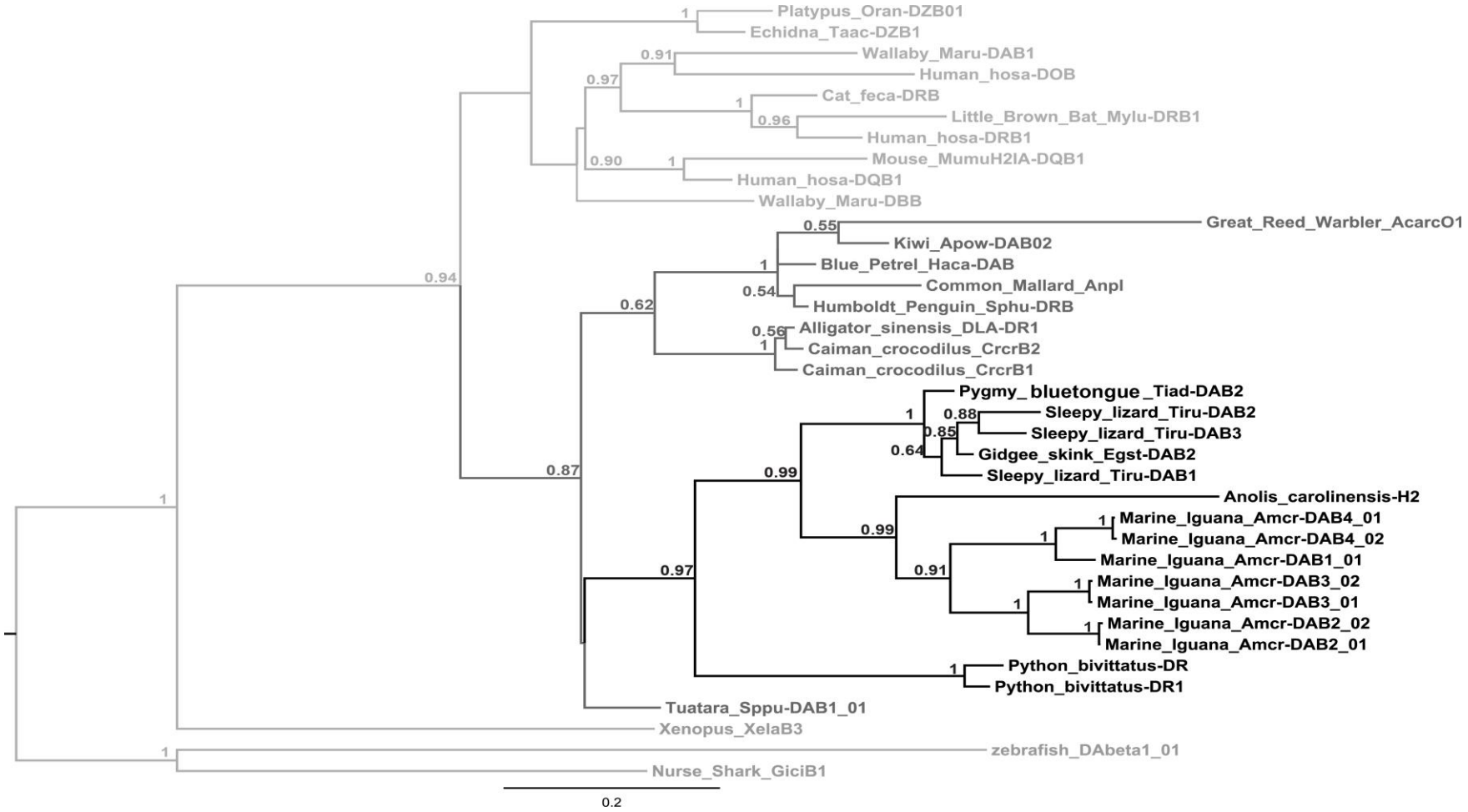


Figure 3B. Bayesian phylogenetic relationship of the major vertebrate groups based on exon 3 nucleotide sequences. The class II B sequences from Squamata are shown with black and all other Diapsida (reptiles includes birds and crocodilians) with dark grey. The values in branches are index of support for the major branches (Bayesian posterior distribution value) and values greater than 50% are represented.

4.5. DISCUSSION

We characterised three full length sequences encoding *T. rugosa* β chains of MHC class II using next generation sequencing data from the cDNA of spleen and thymus total RNA. The full length sequences reconstructed from read assemblies allowed for subsequent design of specific primers for the amplification of genomic class II exon 2 and 3 in *T. rugosa*. From these amplicons we isolated a fourth exon 2 sequence. Uncovering extra exon 2 sequences might indicate the possibility of additional class II β sequences in this species, either because of low expression in the transcriptome, absence from the tissue tested, or due to the technical difficulties during the assembly.

4.5.1. CHARACTERISED CLASS II SEQUENCES IN T. RUGOSA

The *T. rugosa* sequences likely represent functional class II β genes. The sequences did not contain stop codons and encode conserved residues that contribute to protein domains that form a complete MHC class II molecule consistent with the being functional. In addition, our evidence of positive selection acting on the peptide binding residues, providing additional support for the isolated sequences to be encoded by classical MHC class II genes.

Although the *T. rugosa* genes characterised appear to be functional, there are some sites in the translated sequence that are less conserved. The NGT or NGS residues with class II exon 2 are potentially involved in correct protein folding and shaping of the receptor pocket (Wei, et al., 1991; Hammond, et al., 1994; Brown and Pestano, 1998). This motif is usually located at the first part of the β-1 domain (Fig. 2) except for some bony fish

where the motifs are found in the middle of the domain (van Erp, et al., 1996). Nevertheless we did not find this motif in any part of β-1 in *T. rugosa* sequences. In *T. rugosa* MHC class II β_1 domain sequences these normally conserved glycosylation sites have been replaced at the first and third position or are totally absent and instead make the amino acid motif DGG (Fig. 2). Similarly, in some tuatara class II β-1 encoded sequences, an aspartic acid is replaced by an asparagine at the first position of this motif (Miller, et al., 2005). This motif is also absent in class II β-1 sequences of another squamate, the marine iguana (Glaberman, et al., 2009). The N-linked glycans are involved in proper folding and stability of the class II molecules structure in mammals, and therefore the MHC and T cell interaction potentially could be affected by such changes in this normally conserved motif (Wei, et al., 1991; Nag, et al., 1992). Although, the absence of an N-linked glycans site in *T. rugosa* sequences raises the question of the functionality of these sequences, all sequences described from marine iguana (Glaberman, et al., 2009) also lack the site, indicating the potential divergence of these gene groups in squamates. This shared absence of these residues in *T. rugosa* and the marine iguana provides initial evidence that although we did not find the residues, it appears that squamates may have a different residues involved in the proper folding and stability of the class II molecules. On the other hand, two other squamates, an anole and a python, contained the glycosylation residues in the β-1 domain (Figure 2). Given the evolutionary relationships among these species (Pyron, et al., 2013) it is possible that the *Anolis* and *Python* represent a plesiomorphic character states for the N-linked glycans site with the states in Iguanidae and Scincadae representing apomorphic characters. Clearly this is purely

speculation until further sequences from species intermediate to these lineages are obtained. Interestingly, five amino acids (position 25-29, Fig. 2) which were present in the marine iguana and the anole but absent in *T. rugosa* are also absent in all other species characterised to date. Presumably this region evolved after the Iguania lineage, to which the marine iguana and the anole belong (Pyron, et al., 2013), diverged from other squamates. Just when this occurred could be elucidated by the characterisation of the MHC of other species in that lineage including the bearded dragon (*Pogona vitticeps*) which has had a draft genome released (Georges, et al., 2015).

4.5.2. PHYLOGENETIC ANALYSIS

The *T. rugosa* class II sequences formed a monophyletic group with those of the two other *Egernia* group species, *T. adelaidensis* and *E. stokesii*, within the squamates.

Our findings elucidate on the evolution of the class II β gene region. Previous studies have reported that the class II β domain in most of the major mammal groups independently expanded from a small number of ancestral loci with divergent evolution where, following duplication, differences accumulate across loci and there is the maintenance of orthologous loci across species (Nei, et al., 1997; Belov, et al., 2003; Shiina, et al., 2004). However in birds concerted evolution for this gene region has been proposed (Edwards et al., 1999; Edwards et al., 1995; Hess and Edwards, 2002; Wittzell et al., 1999). In that scenario loci (e.g. different MHC class II exon 2 loci) are expected to be more closely related within species than across. Miller, et al. (2005) found high sequence similarity across class II loci, excluding β-1 domain, in tuatara,

providing evidence for concerted evolution. Within *T. rugosa* β-2 sequences we found a more divergent pattern similar to the marine iguana class II sequences (Glaberman, et al., 2009). The evidence for divergent evolution was also previously observed in the class I study of *T. rugosa* based on the conserved exon 4 sequences (Ansari, et al., 2015).

Overall we found that all class II Diapsida sequences were clustered in a monophyletic lineage, with moderate support based on exon 3 but with low support for exon 2. With regards to previous studies that each class II gene region (domain) resulted from different evolutionary forces (Miller and Lambert, 2004; Burri, et al., 2008) our phylogenetic tree indicates that sequences encoded by the β-2 domain from mammals and reptiles (including birds and crocodiles) possibly evolved from different paralogous families of class II genes. However any conclusions on whether *T. rugosa* β-2 genes follow divergent or concerted evolution are premature at this stage due to the small number of sequences obtained.

4.5.3. CONCLUSION

We characterised four different sequences from *T. rugosa* that are consistent with being functional class II genes. However validation of the classical nature of the loci we uncovered requires confirmation of their expression in all immune tissues, and evidence they are highly polymorphic. The utility of the primers designed in this study was supported by successful amplification of the MHC class II fragments in *T. adelaidensis* and *E. stokesii*. We found identical residues in the sequences amplified from *T. adelaidensis* and *E. stokesii* as in the class II *T. rugosa* sequences. These class II sequences

from the three *Egernia* group species form a separate monophyletic cluster within squamates as expected from their close phylogenetic relationship (Gardner, et al., 2008).

The current study is the first to characterise MHC class II in skink and provides insight into the evolution of class II MHC in squamates. We contend that all Squamata class II sequences descended from the same ancestral locus, although each domain possibly has undergone different evolutionary forces. We hope characterizing more sequences from other squamate reptiles would increase the knowledge about the evolution, disease resistance and conservation of these species.

4.6. ACKNOWLEDGEMENTS

We gratefully acknowledge funding from the Sir Mark Mitchell Research. The authors appreciated support from the South Australian Museum (SAM) staff for help in sourcing materials. In addition, they would like to thank Dr. Alison Fitch for her help with the lab work and set up experiments and Dr. Mehregan Ebrahimi for his assistance with figure and charts improvement. The research was conducted according to the guidelines of Flinders University Animal Welfare Committee (approval no. E324).

4.7. APPENDIX FIGURES

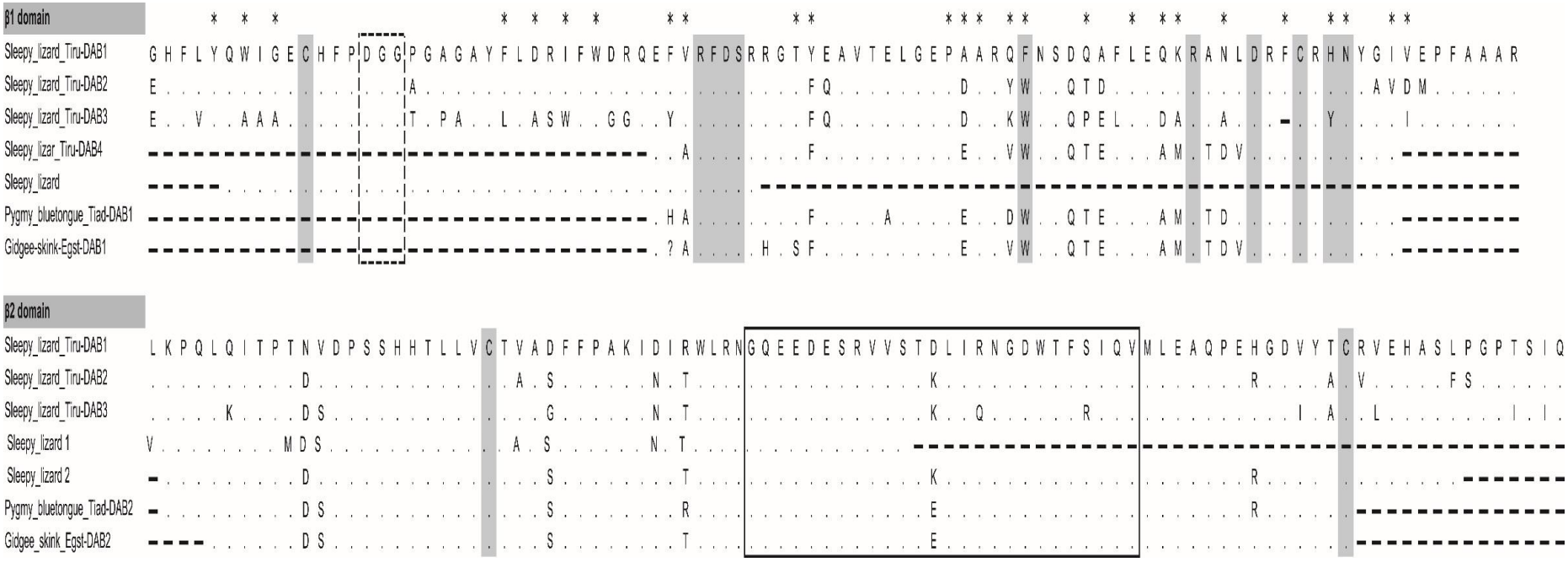


Figure S1: The MHC Class II amino acid alignment among amplified sequences (from *T.* rugosa, *T.* adelaidensis and *E. stokesii*) with the isolated cDNA sequences from *T. rugosa*. The conserved and functional residues highlighted with grey colour. The dashed box shows the NGT glycosylation site and the location of CD4 binding residues (Cammarota, et al., 1992; Wang, et al., 2001) are boxed. The peptide binding (PBR) residues (Brown, et al., 1993; Stern, et al., 1994) are shown with an asterisk.

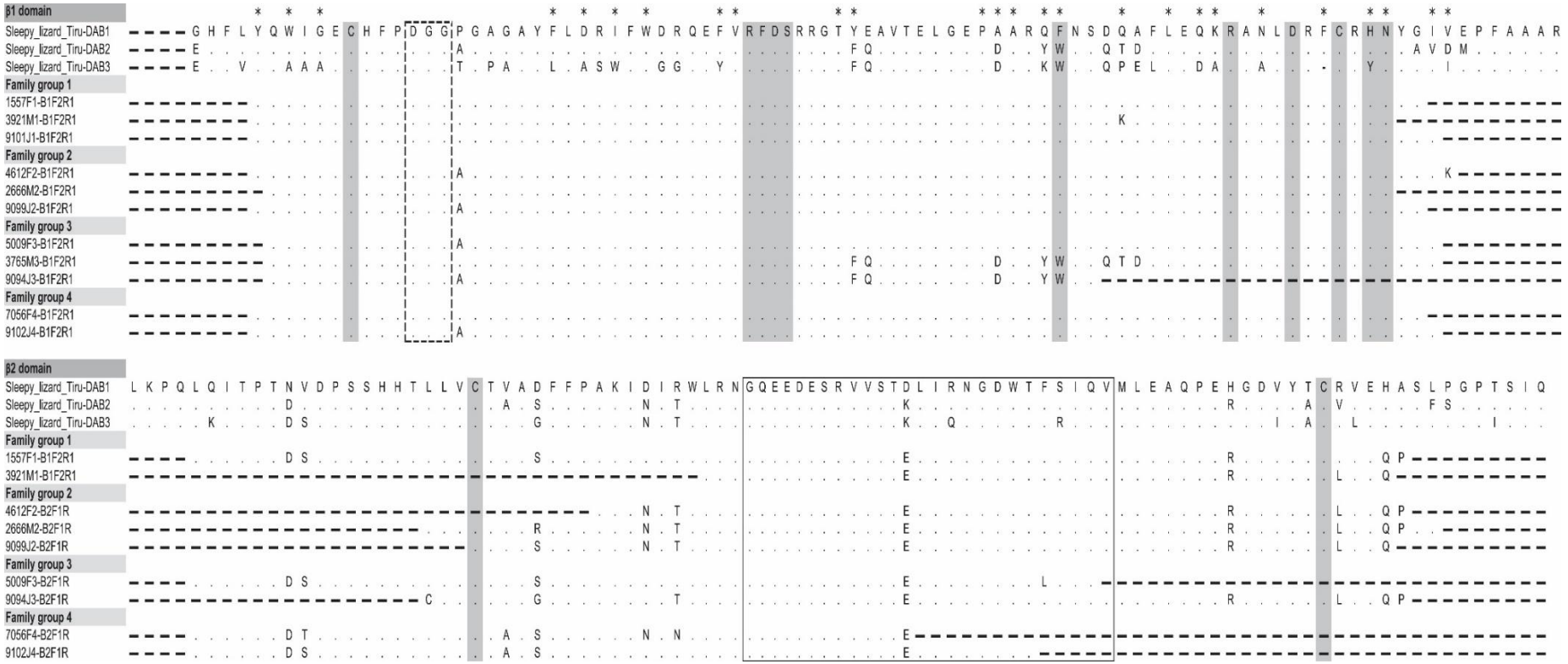


Figure S2: The MHC Class II amino acid alignment of amplified sequences (from family groups in *T.* rugosa) with the isolated sequences from *T.* rugosa. In the individual Id; F means female, M refer to the male and J means Juvenile. The conserved and functional residues highlighted with grey colour. The dashed box shows the NGT glycosylation site and the location of CD4 binding residues (Cammarota, et al., 1992; Wang, et al., 2001) are boxed. The peptide binding (PBR) residues (Brown, et al., 1993; Stern, et al., 1994) are shown with an asterisk.

Chapter five

5. Next generation sequencing & bioinformatics analysis

This chapter has been formatted for submission to the journal of Molecular Ecology Resources.

In this chapter, I adapted a dual indexing strategy involving a single PCR approaches to amplify anonymous nuclear and major histocompatibility complex loci that I developed in Chapters 2-4, sleepy lizard samples. These amplicons were sequence on the Illumina MiSeq platform. Then I developed a bioinformatics software, MITAGSORTER, to sort the huge amount and complicated datasets produced through the next generation sequencing (NGS)

technology. The program facilitates the separation of tagged individuals and multiple loci I used this approach to process nine anonymous nuclear loci and used them in the phylogeographic study of *T. rugosa* (Chapter 6). Time constraints meant that I was unable to utilise the MHC data within my thesis.

However, the approaches I describe in this chapter will potentially boost the advantage of using NGS technology in any study which uses large numbers (hundreds to thousands) of individuals and multiple loci, by improving both sample preparation and bioinformatics analysis to be more time and money efficient.

Prepared for Molecular Ecology Resources

Dual indexing with a single PCR for library preparation and development of MiTagSorter software for de-mulitiplexing sequences data on the Illumina MiSeq platform.

Talat Hojat Ansari, Mahdi Ebrahimi, Esmaeil Ebrahimie and Michael G. Gardner

5.1. Abstract

Population genetic, phylogeographic and phylogenetic studies have benefited from next generation sequencing (NGS) technologies in recent years. The advance of these technologies in rapid and large scale sequencing has increased the popularity of using the DNA MID tags (a.k.a. barcodes or index tags) to allow the pooling of samples into a single run. However, preparation of DNA library is time-consuming and not yet cost-effective for sequencing multiple loci from a large numbers of individuals. In addition, the timely extraction of data of individual samples from a pool of barcoded sequences requires efficient software. Here we present a new library preparation approach and a superfast and comprehensive software, MiTagSorter, to separate sequences of hundreds to thousands of individuals and multiple loci generated using the Illumina MiSeq platform. We evaluated our approach by generating paired-end sequence data from 250 samples of the Australian Skink (*Tiliqua rugosa*) for 21 loci using Illumina MiSeq platform technology. This method significantly reduced the cost and the time involved with sequencing, and

MITAGSORTER software supported de-multiplexing of huge and complicated data sets irrespective of type, size and position of MID tags and will trim off the MID tags and primers.

Keywords: Illumina, MID tags, next generation sequencing, *Tiliqua*, MITAGSORTER

5.2. INTRODUCTION

Next generation sequencing (NGS) technologies are revolutionising genomic studies. These new technologies outperform the Sanger sequencing technology, which was the dominant sequencing technology over two decades (Berglund, et al., 2011), in terms of speed, resolution, throughput and cost (Kircher, et al., 2012; McElhoe, et al., 2014). The NGS technologies can produce over hundreds of billions of base pairs data, currently approximately hundred times more data compared with capillary electrophoresis based Sanger sequencing, at a smaller fraction of time and cost. However, researchers willing to use NGS approaches are facing two major challenges: sample preparation and bioinformatics analysis.

The potential of Illumina MiSeq, introduced in 2011, for sequencing metagenomes, small amplicons, as well as targeted genes (Caporaso, et al., 2012; Quail, et al., 2012; Kozich, et al., 2013; McElhoe, et al., 2014), has benefited the genetic and genomic study of both model and non-model organisms (Wheat, 2010).

Advances in PCR and hybridisation techniques have been introduced to apply targeted sequencing of specific loci through the NGS platform (Briggs,

et al., 2009; Gnirke, et al., 2009; Maricic, et al., 2010). However, the time and cost of preparing samples prior to NGS can be a major challenge, particularly when using the NGS approaches is only cost effective when combining many individuals in a single sequencing run (Glenn, 2011). This multiplexing approach was developed further by using unique (4-10 base pair-bp) DNA MID tags that allow the identification of multiple target sequences in hundreds of individuals in a single sequencing run (Binladen, et al., 2007; Farias-Hesson, et al., 2010; Meyer and Kircher, 2010).

Popular barcoding strategies use a linker sequence at the 5´-tailed of the locus specific primer to attach high throughput NGS (HT-NGS) adaptors and MID tags (Bybee, et al., 2011; de Carcer, et al., 2011), but involve two rounds of PCR, lessening the procedures' cost effectiveness (Figure 1a). To overcome this problem, Clarke, et al. (2014) developed an approach that used just a single PCR to amplify the target locus and attach adaptors and MID tags via a linker (Figure 1b). Here we modify this approach for use on the Illumina MiSeq platform and incorporate a dual indexing strategy. We used our modified technique to sequence a single pool of 250 samples of an Australian Skink (*Tiliqua rugosa*) amplified for 21 markers [including anonymous nuclear loci (ANL) and major histocompatibility complex (MHC) class I and Class II] to test the efficiency of this approach.

Once sequences have been generated, the next problem to overcome is their efficient sorting, often called de-multiplexing, from a pool containing multiple samples and loci. This issue can be complicated by the absence of the MID tags coupled with NGS adapters, or other possible inserted sequences, for example the repetition of the linker used in multiplexing.

Several software packages and tools are available to perform the required bioinformatics analysis for NGS data (Gogol-Döring and Chen, 2012). However, currently available software and freeware in many cases need extraordinary amount of computing time and processing power. In addition, the current available software for de-multiplexing on the Illumina MiSeq platform are either time-consuming or not efficient. This problem is exacerbated in population genetics studies when accurate estimation of parameters requires data from multiple loci and hundreds to thousands of individuals (Carling and Brumfield, 2007; Bertozzi, et al., 2012). The library separation for 21 loci for 250 individuals, for example, requires several steps including separating sequenced reads for each specific locus and then for each individual. This is a very time-consuming task, particularly when loci differ in size which makes it impossible to run the de-multiplexing process for several loci at the same time using the current available commercial software such as CLC genomic workbench. The other alternative method is to separate the sequences based on each individual's mid tags, via the standard illumina post processing if the tags are compliant with this method. Although using this facility is time-efficient, there is still the need to separate the specific sequences for each locus/marker used.

Here we develop a comprehensive, cost-effective, user-friendly and fast bioinformatics tool, MITAGSORTER, to de-multiplex large data sets (2.6 Gb). The program is able to process multiplexed data and complicated reads on modest computer processors such as laptops. We designed MITAGSORTER to provide efficient sorting of raw reads from complex experimental designs by any type, size and position of MID tags. We demonstrate the efficiency of the

software by applying MITAGSORTER to our *Tiliqua rugosa* data set generated from 250 individuals for 21 different markers using Illumina MiSeq platform.

5.3. METHODS

5.3.1. SAMPLE COLLECTION AND DNA EXTRACTION

We obtained tissue and blood samples of *T. rugosa* from across the geographical range of the species in Southern Australia. DNA was extracted from 250 specimens using a puregene DNA isolation kit (Gentra systems, Minneapolis MW, USA) from frozen or alcohol-preserved tissues and blood collections held by the South Australian Museum (SAM) in the Australian Biological Tissue Collection (ABTC).

5.3.2. MULTIPLEXING USING DUAL INDEX STRATEGY IN A SINGLE PCR

We amplified the targeted loci and attached adaptors and MID tags in a single PCR following multiplex-ready PCR technology (MRT) (Hayden, et al., 2008). The MRT method (Hayden, et al., 2008) is a two stage PCR that use the advantage of M13-tailed primers to add a fluorophore for single nucleotide polymorphism (SNPs) and microsatellite (SSRs). We modified the forward and reverse locus specific (LS) primers to include a linker of 15bp (ACGACGTTGTAAAA) and 16bp (CATTAAGTTCCCATTA) respectively, at their 5´ ends. These linkers were used as primer binding sites that allow any MID tags or combination of adaptors to attach in the second stage

of the PCR (Hayden, et al., 2008). The LS primer, with additional linker, facilitated the amplification of the target loci during the first stage of the PCR. In the second stage of the PCR we attached a primer composed of a universal adaptor for Illumina MiSeq (P5 or P7), an 8 bp MID tags, an Illumina sequencing primer (Read 1 and Read 2), and the linker. The lower annealing temperature of the primer made the attachment of this linker possible (Figure 1C). Henceforth we call this combination of adaptor and MID tags: tag F for the forward version; and tag R for the reverse version. We further reduced the price of preparing the library by using a combination of 11 F and 23 R tags, to discriminate amongst 250 individuals with amplifications of each individual lizards' sample consisting of the same tags for all loci (Table 1).

Polymerase chain reactions (PCR) were performed in 12µL volumes, containing 1× reaction buffer, 0.15U of Immolase, 10nM of Ls primer (Forward and Reverse premixed), 75nM of tagF and 75nM of tagR, plus 20ng of DNA. The first stage of the PCR consisted of a 10 min at 95 °C followed by five cycles of, 60 s at 92 °C, 50 s at 90 °C and 1 min at 72 °C; followed by 20 cycles of 30 s at 92 °C, 90 s at 63 °C and 1 min at 72 °C, The second stage of the PCR which immediately followed consisted of 40 cycles of 15 s at 92 °C, 60 s at 54 °C and 1 min at 72 °C, ending with 10 min at 72 °C and 30 s at 25 °C. Each lizard sample was amplified using a specific tag F and tag R primer with the unique MID tag, and the same set of MID tags used for different markers tested here.

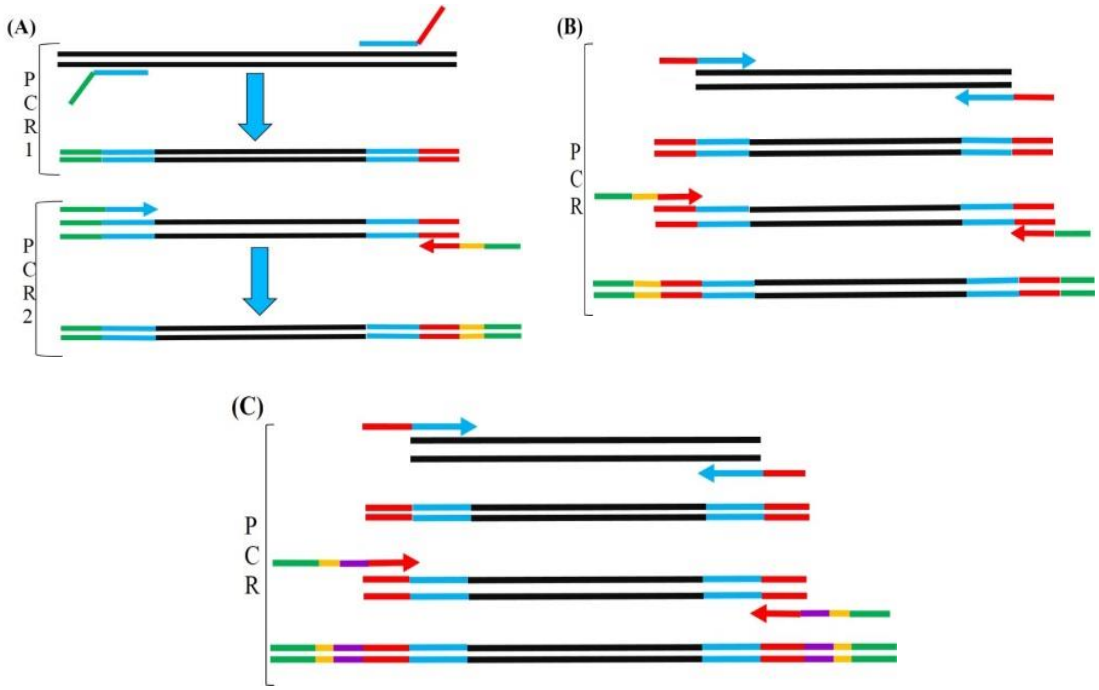


Fig. 1 Schematic representation of individual sample identification strategies. Locus specific primer, is shown with the blue colour. Red representing the linker, and green is the HT-NGS adaptors. The unique MID tags are shown with yellow. (A) The two-step strategy (de Carcer, et al., 2011) using two round of PCR. In the first PCR step, targeted sequences are amplified using a modified locus specific primer+ linker, then the products are barcoded during the Second PCR amplification using the linker sequence composed with the MID tag and appropriate adaptor for 454 technology. (B). Modular tagging protocol (Clarke, et al., 2014) using a single PCR for the Ion Torrent. During the first step of PCR, the linker attach to the target region with the locus specific primer, since then by the use of the linker the unique MID tag and adaptor attached to the amplicons. (C). Our modified single PCR protocol for dual indexing on MiSeq platform. Purple represents the second Illumina adaptor (Read 1 and Read 2). As described above, at the first stage locus specific primer composed with the linker attached to the amplicons and then adaptor and MID tags joined via the linker.

Table1. Primer sequences (5' – 3') employed in this study. MRT tag (linker) sequences are bold and italic.

Illumina adaptors are italic. The position of the MID tags are shown with X.

Name	Locus sp. primer	Primer Sequence (5´-3´)
Tag F (P5+MID tag+Read1+MRT tagF)		*AATGATACGGCGACCACCGAXXXXXXXXACACTCTTTCCCTACACGA CGCTCTTCCGATCT* ***ACGACGTTGTAAAA***
Tag R (P7+MID tag+Read2+MRT tagR)		*CAAGCAGAAGACGGCATACGAXXXXXXXXGTGACTGGAGTTCAGAC GTGTGCTCTTCCGATCT* ***CATTAAGTTCCCATTA***
MRT tagF-MHCIα1F4	α1 MHC classI	MRT tagF-GCTCCTCGTCCCACTCGGTGAAG
MRT tagR-MHCIα1R4		MRT tagR-CTGAGCGTCGCCTCCTCGC
MRT tagF-MHCIα1F2	α1 MHC classI	MRT tagF-ACGGCGGTGTCKGAGCCYRGCCAG
MRT tagR-MHCIα1R1		MRT tagR-TGTKCHDRTCCCAGWRMTGRGGGT
MRT tagF-MHCIα1F2b	α1 MHC classI	MRT tagF-ACGGCGGTGTCGGAGCCTGGCCAG
MRT tagR-MHCIα1R1b		MRT tagR-TGTKCATGTCCCAGAGMTGGGGGT
MRT tagF-MHCIα1F5	α1MHC classI	MRT tagF-GGTGCCTCTTCCCACTCTGCG
MRT tagR-MHCIα1R1b		MRT tagR-TGTKCATGTCCCAGAGMTGGGGGT
MRT tagF-MHCIα2F1	α2 MHC classI	MRT tagF-TCACACBYKGCAGYBSATGTAYGGCTG
MRT tagR-MHCIα2R2		MRT tagR-TCCTCWGCAGDSTCTCCYTCC
MRT tagF-MHCIα2F1b	α2 MHC classI	MRT tagF-TCACACGTGGCAGTGGATGTATGGCTG
MRT tagR-MHCIα2R2b		MRT tagR-TCCTCAGCAGGGTCTCCTTCC
MRT tagF-MHCIIB1F1	β1 MHCclassII	MRT tagF-GAGYWCGYGCGCTTCGACA
MRT tagR-MHCIIB1R1		MRT tagR-WGATKCCRTAGTTGTRCCGGCAG
MRT tagF-TrANM4	ANM4	MRT tagF-TATGCACAGAGCCTTCAATG
MRT tagR-TrANM_4R		MRT tagR-CCCAAGCAAGTGGACTTCAG
MRT tagF-TrANM_9F	ANM9	MRT tagF-CTAGTGCAGGCTCCTCATCC
MRT tagR-TrANM_9R		MRT tagR-GAGGTGGCAAGCTGAAGAAC
MRT tagF-TrANM_15F	ANM15	MRT tagF-CGGAACACATGTCCATTCTG

Table1. Continued.

Name	Locus sp. primer	Primer Sequence (5´-3´)
MRT tagR-TrANM_15R		MRT tagR-TGATCTGAAACCTGCTGCTG
MRT tagF-TrANM_16F	ANM16	MRT tagF-CCTTTGGCAAATGTGGGTAG
MRT tagR-TrANM_16R		MRT tagR-CAGTGGTTGTTGGTTGGTTG
MRT tagF-TrANM_21F	ANM21	MRT tagF-TCTGGAGCACGTTGAATCAG
MRT tagR-TrANM_21R		MRT tagR-ATGAGGAATGCAGGGTGAAG
MRT tagF-TrANM_22F	ANM22	MRT tagF-TGACGGCTGTCAGACAATAGAG
MRT tagR-TrANM_22R		MRT tagR-TTTGTGTTTGCCCATGAGAG
MRT tagF-TrANM_23F	ANM23	MRT tagF-GAAGTGCCAGGAGGTCAATC
MRT tagR-TrANM_23R		MRT tagR-TCACATAGCCCTTGGTTTCC
MRT tagF-TrANM_34F	ANM34	MRT tagF-GCAAGCAACCAGGAAGAAAC
MRT tagR-TrANM_34R		MRT tagR-CATGGATTCAGATGCCACAC
MRT tagF-TrANM_39F	ANM39	MRT tagF-GGGAGAGCAGAAAGATGGTG
MRT tagR-TrANM_39R		MRT tagR-CCTCTAGTGGGTGGCAGAAG
MRT tagF-TrANM_44F	ANM44	MRT tagF-ATTTGGATGCTGTGGGTAGC
MRT tagR-TrANM_44R		MRT tagR-GGCATGGGCAGAGTAACTTC
MRT tagF-TrANM_45F	ANM45	MRT tagF-CAGGGCAGTCAATGTTCTTG
MRT tagR-TrANM_45R		MRT tagR-GTGCCATCGGTACTGAAATG
MRT tagF-TrANM_46F	ANM46	MRT tagF-CACATGCACATTTCCTCTGG
MRT tagR-TrANM_46R		MRT tagR-CTGTGGCACTGCTCTTCTTG
MRT tagF-TrANM_47F	ANM47	MRT tagF-TGTGCAGACATGGTTCCTTC
MRT tagR-TrANM_47R		MRT tagR-TCAGTCCAGCATCAGTGCTC
MRT tagF-TrANM_48F	ANM48	MRT tagF-TCCATTTGCATCAGCGTATC
MRT tagR-TrANM_48R		MRT tagR-TGGACTCCTGGGAGTAGGTC

Approximately 5% of PCR products from each locus were quantified using the Qubit® dsDNA HS with a Qubit® 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA) to obtain an approximate average concentration for each locus. We then combined PCR products in equimolar ratio to create 21 libraries (including 14 ANL loci; six products covering exon 2 and 3 of MHC class I, and one products covering exon 2 of MHC class II). The PCR products were purified for sequencing using a QIAquick® PCR Purification Kit (Qiagen sample and assay technology) and quantified using the Qubit® dsDNA HS with a Qubit® 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA). We further evaluated their size distribution and concentration with an Agilent 2200 Tape Station using High Sensitivity D1K ScreenTape and reagents (Agilent Technologies, Santa Clara, CA, USA). The pooled, cleaned products were then further combined in equimolar ratios to create a single library. We sequenced these pooled amplicons on the Illumina MiSeq platform with a 250 bp paired end read length at the AGRF (Melbourne, Australia) following standard protocols.

5.3.3. SEQUENCE DATA DE-MULTIPLEXING USING MITAGSORTER

To efficiently deal with the large volume and complicated data set we collected from the Illumina MiSeq run, we developed a specific program, MITAGSORTER, to group and simplify our sequences data. This program, written in python, sorts reads for different loci and individuals by using the LS primer sequences and MID tags sequences. A schematic overview of the de-multiplexing is shown in Figure 2.

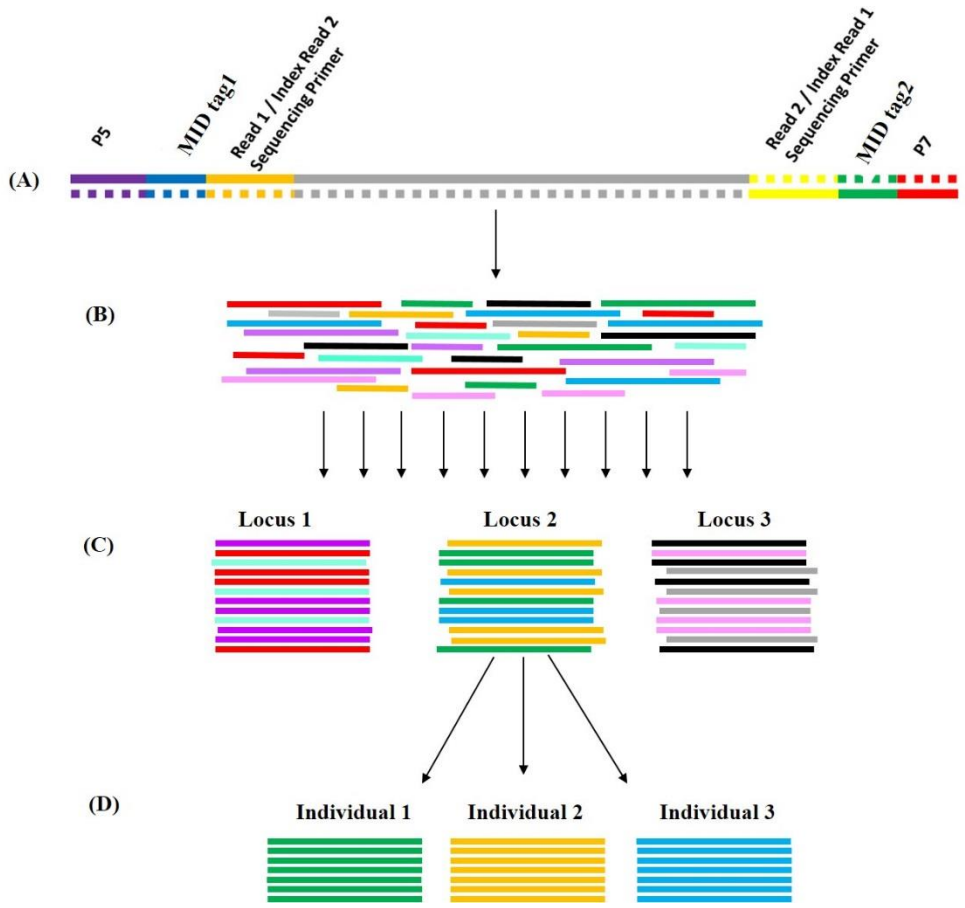


Figure 2. A schematic overview of Illumina MiSeq sequencing, and de-multiplexing. (A). Showing the Adaptors and location of MID tag require for illumine MiSeq sequencing. P5 and P7 are the first Illumina adaptor, and MID tag 1 and 2 are unique tags for forward and revers sequences. Read 1 and read 2 are the second Illumina adaptor. (B). Sequences resulted from MiSeq run including the sequences from large numbers of individuals for several loci with different loci and individuals represented as different colors. (C). Sorting the sequences based on the tested Locus specific Primes using the MiTagSorter. (D). Sorting the individuals sequences via unique MID tags using MiTagSorter.

MITAGSORTER accepts FASTA (.fa) formated files as input and produces output in text (.txt) format. The user can input their own MID tags sets in an excel sheet provided under the data folder of the software. This offers a flexible way to define samples from varied preparation and barcoding methods.

MITAGSORTER is designed to easily and quickly sort reads from complex data set with minimum memory requirements and allows to collect results as output file only by click on the 'Find.py' button. The program will automatically trim off the MID tags and markers whilst assigning the individuals to different groups. In addition an extra text file is provided under the report name that includes the reads that did not have the inputted MID tags or have included error during the de-multiplexing. The software includes a user manual and an example data set that can be find from the following link https://goo.gl/26LjI7 (see CD provided).

5.3.4. QUALITY CONTROL AND TRIMMING

The qualities of reads were assessed using FASTQ (http://www.bioinformatics. Babraham.ac.uk/projects/fastqc/) and also CLC genomic workbench version 7.5.0. The reads were trimmed off for which the per-base quality scores (Phred score) was lower than Q30 (99.9% accuracy) and then data was sorted for each locus and individuals as described above using MITAGSORTER.

The reads for each locus were mapped to a database of reference sequences using CLC Genomics workbench version 7.5.0. In order to generate a database of reference sequences, we used sequences generated through two sources. First, MHC class I and II sequences were sourced from RNA transcriptomic data (Ansari, et al., 2015). Second, sequences for the anonymous nuclear loci were sourced from Ansari, et al. (2014). We culled reads that did not align to the references.

We constructed and checked our ANL haplotype sequences by visual review of the aligned reads through CLC genomic workbench version 7.0 and also by using Geneious version 8.1 (Drummond, et al., 2012). Sequences resulted from both MHC class I and II also checked for potential number of amplified alleles to use in the consequence analysis.

5.4. RESULTS

A total of 5,763,626 reads were obtained from the amplicons of 250 individuals of *T. rugosa* for 21 loci (Table 1) sequenced on the Illumina MiSeq platform. The low quality reads were eliminated for $Q < 30$ which resulted in 5,763,287 reads. The remaining reads were sorted for 21 loci using MITAGSORTER that showed the number of reads per locus ranged from 50,000 to 260,000 paired. Average Q-values for reads were nearly 30 across all of the 21 loci (Figure S1).

The de-multiplexing performed with MITAGSORTER took less than 5 minutes (Dell latitude E5530, CPU Intel i7 (3540), 16 Gb ram), resulted in the recovery of nearly 2.1 Gigabyte (Gb) of sequence data out of 2.3 Gigabyte (Gb), and approximately 200 Megabyte (Mb) of unsorted data. After the first stage of the de-multiplexing, reads assigned to each locus sorted for MID tags in second stage and the MITAGSORTER that allowed us to define reads for each individuals.

Recovery percentage for sequences data were different across the loci where ANL and MHC class I loci had highest amount of contigs compare with MHC class II loci. Therefore different rate of reads between100 to 2000 yielded

per individual per locus. The recovered sequences showed 100% identity for each locus when mapped to our reference database using CLC genomic workbench.

5.5. DISCUSSION

Our work represents an improvement over existing techniques in two ways. First, our dual indexing approach facilitates the generation of a large number of targeted sequences while minimizing time and cost further by adding the adaptors and MID tags in a single PCR run. This reduces clean-up steps and minimizes the possibility of contamination as mentioned by Clarke, et al. (2014). Important benefits are the cost and time reduction of approximately 50%. Our dual indexing approach provides more cost and time-effective protocol compare with previous methods (de Carcer, et al., 2011; Clarke, et al., 2014).

Our approach improves the use of MiSeq platform for population genetics, phylogeography and phylogenetics by allowing multiple loci to be separated for hundreds to thousands of individuals in a single run. This is due to our use of minimal MID tags and just a single PCR run per locus for each individual. For instance, 500 individuals can be MID tags just by using a combination of 50 tag F and 10 tag R, and the same MID tags can be used for the same individual over different loci as loci can be separated by the unique locus specific primers.

The second we advance the de-multiplexing of complicated data sets via the software MITAGSORTER. MITAGSORTER is superfast tool that just took 3

minutes and 10 seconds to sort 2.1 Gb of data for 21 loci on a personal processor laptop with 16GB RAM.

Another significant aspect of MITAGSORTER is in allowing the input of a data excel sheet that permits any kind and size of unique tags or even primer sequences to be easily imported for the further analysis in the software. This is in comparison to programs such as CLC genomic workbench v 7.5.0 which requires prior information in regards to length of linkers or MID tags and doesn't allow the simultaneous processing of tags of significantly increasing the processing time involved. Problematic sequence variants such as sequences with a repetition of the P5 or P7 Illumina tags, base pare deletions from the expected adapter sequence, or sequences that don't reach the end, are also difficult to deal with using CLC workbench v 7.5.0 or to manually check given the enormity of our or typical data sets. Our program has a similar function to SESAME (Meglecz, et al., 2011) although MITAGSORTER is also able to cope with complicated data sets with variable size, type, and position of tags with speed.

Normally, it might be possible to use BLAST and map the sequences with the database of reference sequences to separate the specific sequences for each library. However, this method is not time-efficient and would increase the chance of error in separating libraries. This is especially through in case of population genetics, phylogeography and phylogenetics studies that using a large number of loci. However, MITAGSORTER allowed us to sort our data set based either locus specific primer or unique (8bp) MID tags, negating the problems presented by data variants, or sequences had been pre-sorted for each individual.

An additional important aspect of MITAGSORTER is that it is very easy to setup and run. This software just needs a Python program to be installed prior to run. Running the program only requires a copy of data set in FASTA (.fa) format inside the program and then clicking a button called 'find.py'.

Our approach improves the sequencing of hundreds to thousands of samples for multiple genes in a single run on an extremely sensitive NGS platform (Illumina MiSeq). This method has great potential for any studies that require targeted re-sequencing for a large number of samples) studies because it offer the capability for amplifying the desired gene regions and joins this with the power of Illumina MiSeq sequencing. Although other approaches such as exon capture using sequence baits (Gnirke, et al., 2009) are an efficient way to target many regions, they require large startup costs and time. The benefit of capture approaches is when repeating the assays on numerous occasions. Amplicon sequencings is useful for testing differences at a population genetic, phylogeographic or phylogenetic scales as a single study or a pilot. Further improvements to the library preparation advances could be to perform multiplex PCR so that multiple loci are co-amplified. This would require the careful assessment of primer concentrations to ensure the resultant products from different loci are equi-molar.

5.6. ACKNOWLEDGEMENTS

We appreciatively acknowledge funding from the Sir Mark Mitchell Research. The authors respected support from the South Australian Museum (SAM) staff for help in sourcing materials. In addition, they would like to

thank Dr. Tessa Bradford for her support and advice in the lab work and set up experiments. The research was conducted according to the guidelines of Flinders University Animal Welfare Committee (approval no. E324).

5.7. Appendix

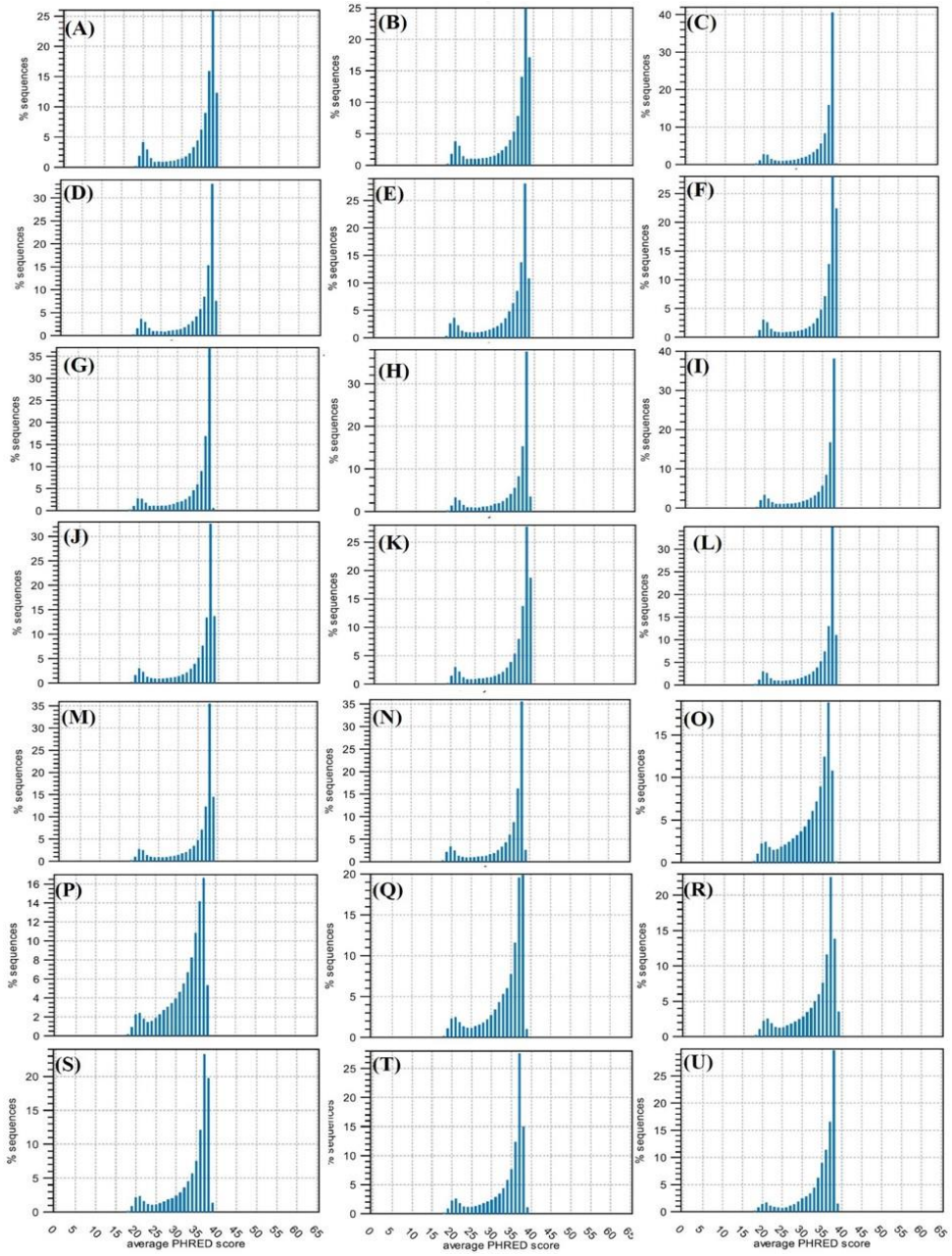


Figure S1: The Average Q-values (Average Phred Score) for 21 loci amplified and sequenced in current study. Each alphabetic character refer to the specific primer pairs, which presented in table1 and used to amplify specific loci. (A). TrANM4, (B). TrANM9, (C). TrANM15, (D). TrANM16, (E). TrANM21, (F). TrANM22, (G). TrANM23, (H). TrANM34, (I). TrANM39, (J). TrANM44, (K). TrANM45, (L). TrANM46, (M). TrANM47, (N). TrANM48, (O). TrIα1F2, (P). TrIα1F2b, (Q). TrIα1F4, (R). TrIα1F5, (S). TrIα2F1, (T). TrIα2F1b, and (U). TrIIB1F1.

5.7.1. MITAGSORTER V1.0 MANUAL

Overview: De- multiplexing is very important tools in the quest to sort the huge and complicated data set generated by next generation sequencing technology. Through de-multiplexing it is possible to sort the sequences base on different locus targeted in the experiment and then assorted them for each individual. MITAGSORTER is a user-friendly software that is easy to install and follow.

Computer requirements and installation: MITAGSORTER runs on a Windows or Linux system as long as you have Python installed on your machine. Python is a free software which could be download and install from the following link; https://www.python.org/downloads/release/python-341/. You can find how to install python in your system in https://www.python.org/.

Installation: To install MITAGSORTER, simply download the software from the following link https://goo.gl/26LjI7 (see CD provided) and extract the zip file then copy the MITAGSORTER software to any directory you desire on your computer. Note the example file is in the same folder.

Input file: The input file includes two types of data:

First, the next generation sequencing data:

The software only accepts sequence files in FASTA format (.fa). Input file must be in the same folder as the software. The paired end sequences results usually are in two separate files. These should be paired into a single file (no need to make consensus sequences but simply put two file together to have a single input data included all sequences, many software available

automatically do this, for example Genious) prior to de-multiplexing (as it shown in example).

Second, the barcode sequences:

The MITAGSORTER assigns reads to loci and individuals, if barcode sequences are available. The barcode sequences are either locus specific primer or MID tag which could be included the next generation adaptors. However, we provide an excel file that you can copy your individuals identity and barcodes sequences into. Although the program does not allow for ambiguity bases in barcodes sequences, there is no limitation for their size or running variant sizes at the same run (for example different locus specific loci might have different sizes but they all can be sorted in a single run).

Start and run the software: Double-click the "find.py" file to start the program, as shown by details in example below.

The program has three option to choose for sorting the sequences.

I. Based on forward barcode sequences

II. Based on reverse barcode sequences

III. Based on both forward and reverse barcode sequences

You can choose any of these analysis based on your barcoding system, and also select to trim the barcode sequences.

Below is an example of how to run the software in detailed screen shots.

Step 1: Preparing input file:

First: Example1.fa file is including the paired end sequences resulted from Illumina MiSeq (Fig. 1A). As we mentioned above both read 1 and read 2 sequences, paired in a single file.

Copy Fasta (.fa) format of the input file (Example 1.fa) in software folder (Fig. 1B).

Second: The data file information: first column is for an individual identity this can be any combination of letters or numbers. In second and third column the specific sequences used for the forward and reverse tag should be pasted (Fig. 2).

Note: You can replace the default sequences with your own barcode sequences in the data file. There is no limitation for the size or the number of barcodes you can add here.

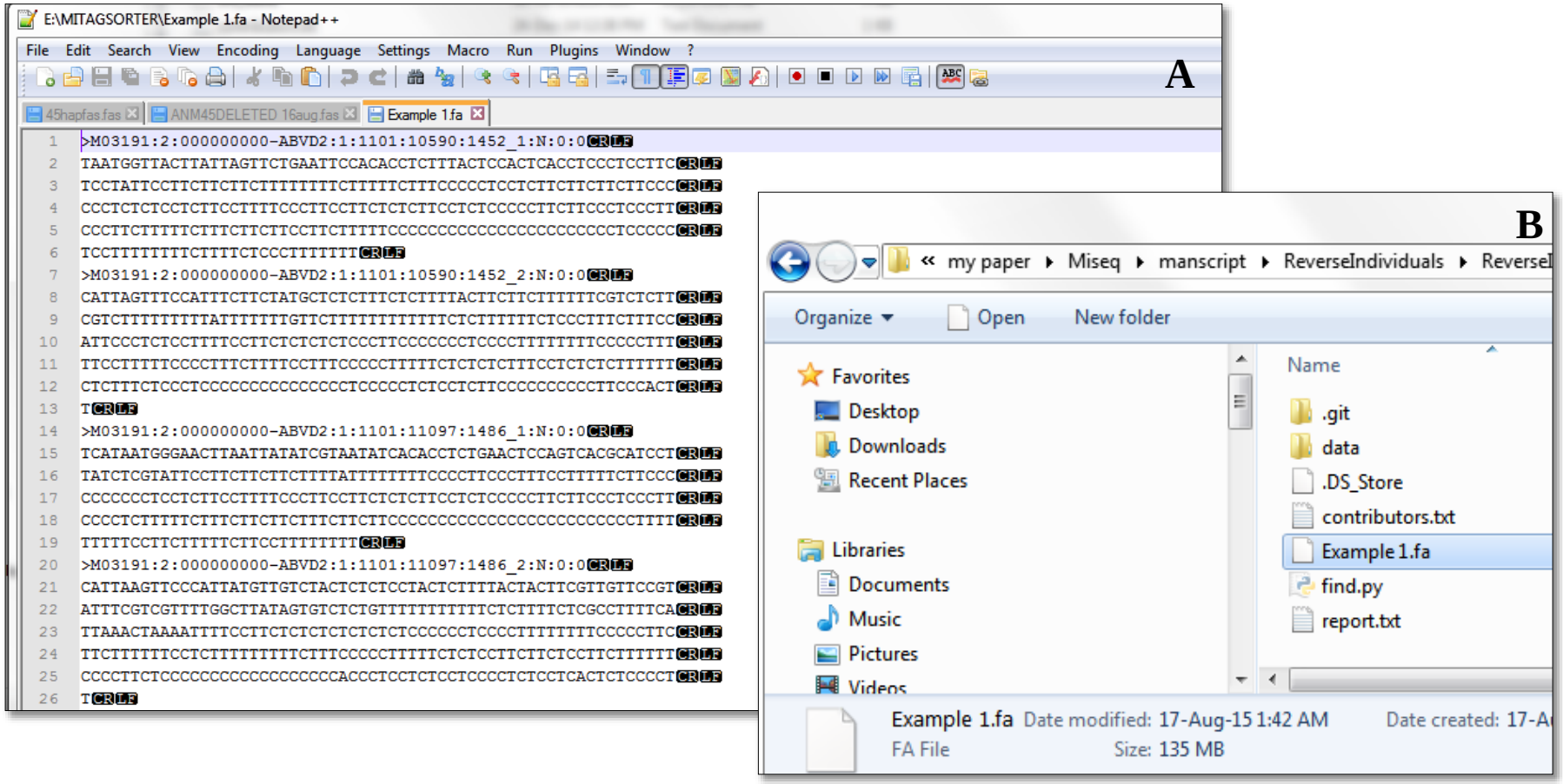
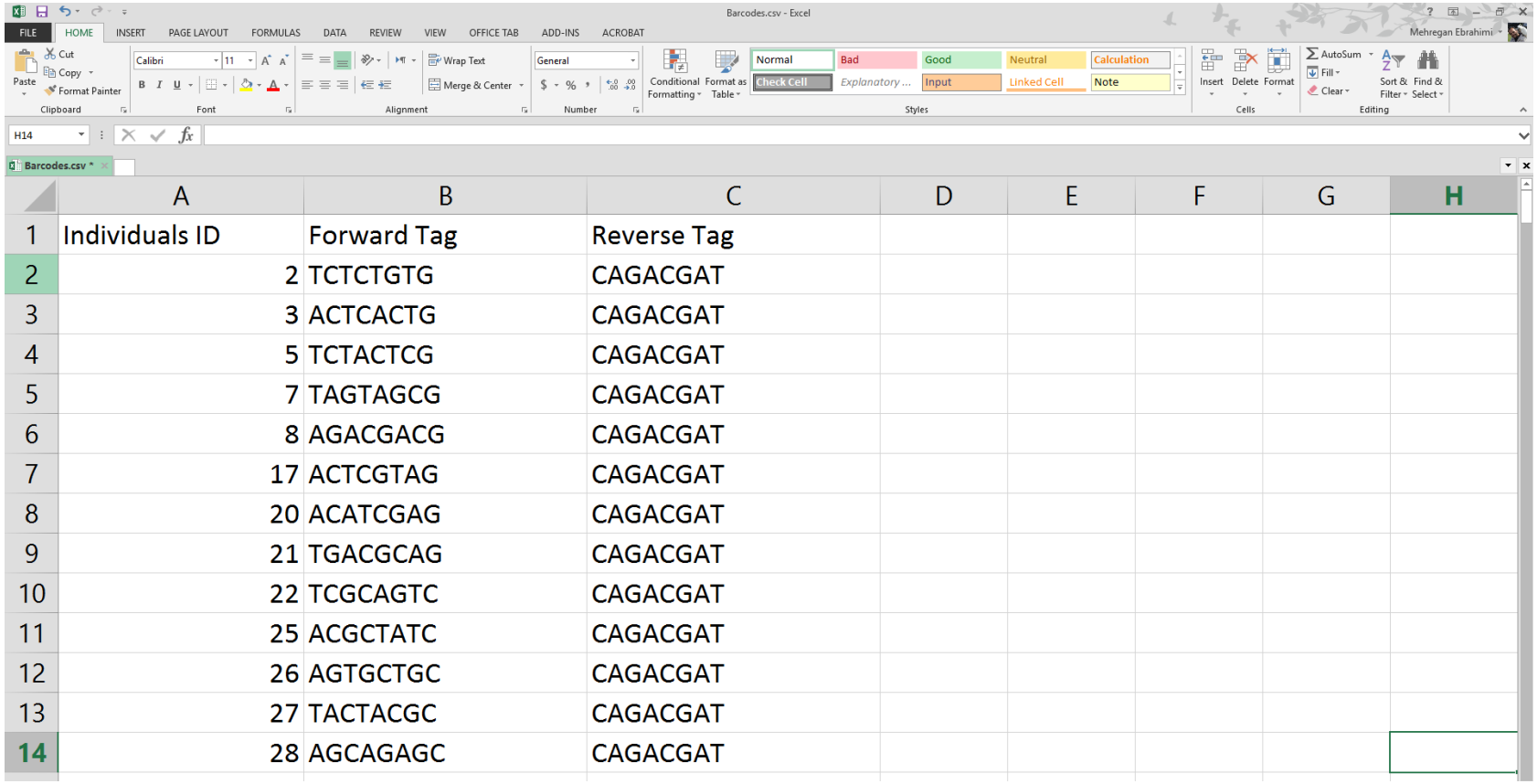


Figure 1. The input files. A: The FASTA file (Example1.fa) included both read 1 and read 2 results. B: Highlight the example file that had been copied inside the software folder.



	A	B	C	D	E	F	G	H
1	Individuals ID	Forward Tag	Reverse Tag					
2	2	TCTCTGTG	CAGACGAT					
3	3	ACTCACTG	CAGACGAT					
4	5	TCTACTCG	CAGACGAT					
5	7	TAGTAGCG	CAGACGAT					
6	8	AGACGACG	CAGACGAT					
7	17	ACTCGTAG	CAGACGAT					
8	20	ACATCGAG	CAGACGAT					
9	21	TGACGCAG	CAGACGAT					
10	22	TCGCAGTC	CAGACGAT					
11	25	ACGCTATC	CAGACGAT					
12	26	AGTGCTGC	CAGACGAT					
13	27	TACTACGC	CAGACGAT					
14	28	AGCAGAGC	CAGACGAT					

Figure 2. The barcode sequences (Example1) include individual identity, forward and reverse tag sequences.

Step 2: Running the program:

To start running the program Double click on the find. py key. If you have problem to run the software follow the steps in troubleshooting (Fig. 3).

Step 2-1: Choosing the available analysis method:

There are three method available for sorting the sequences: 1) Sort the sequences based on the forward tags, 2) Sort the sequences based on the reverse tags, 3) Sort the sequences based on both forward and reverse tags in case of dual indexing approaches (Use this option for paired end sequences). Choose the sorting options by just typing the appropriate number from 1 to 3 (Fig. 4).

Selecting these analysis depends to the barcoding strategy you chosen to specify your individuals. If you used double indexing technique, for example, the third option is the most suitable method.

For the example file you can choose either of these options to test.

Step 2-2: Trim or not to trim:

In this step the program will ask you to choose if the barcodes need to be trimmed off from the sorted sequences per individuals or not (Fig. 5).

Step 3: The output folder

The program makes a folder called "Output" where you can collect the results from output folder. The result can be visualised using notepad. In addition, the program will produce a report file (report.txt) related to the quality of the MITAGSORTER run. This identifies if any of the datasets contained an error during the de-multiplexing and provides which individual id had the error. (Fig. 6).

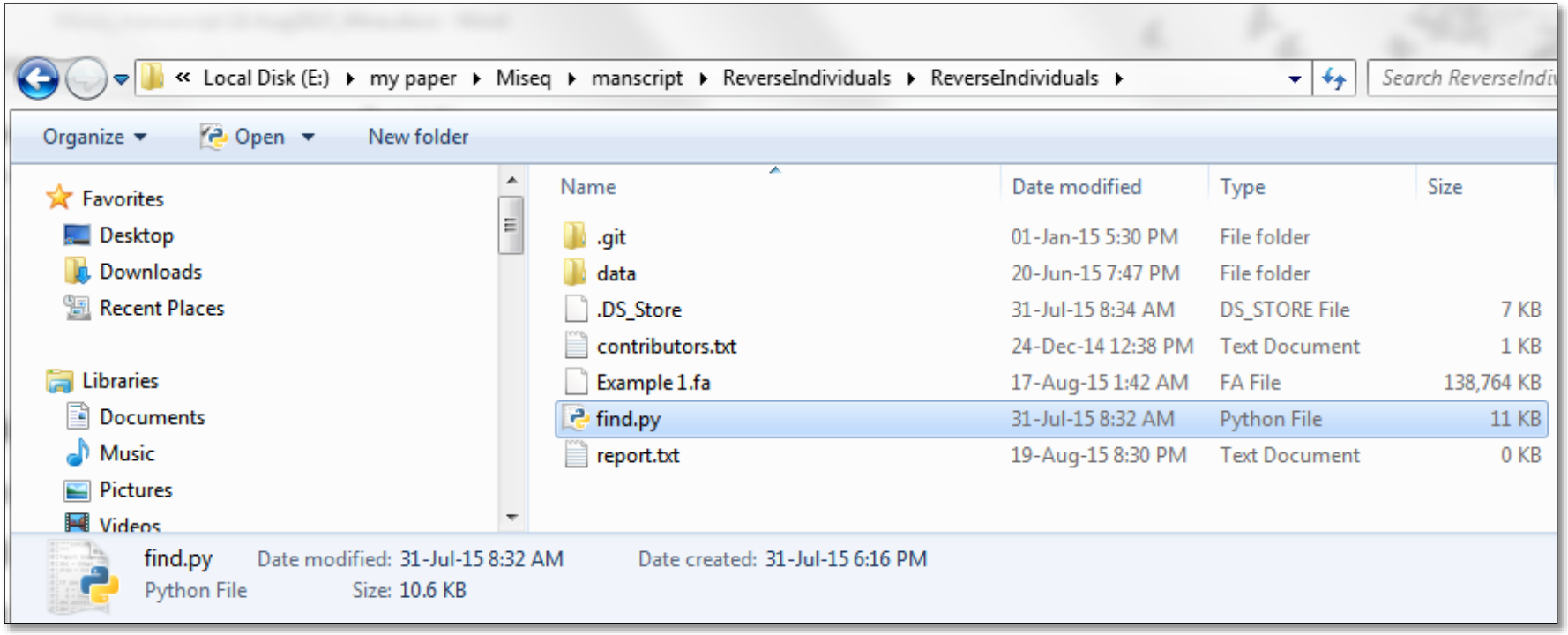


Figure 3. The first step to run the program.

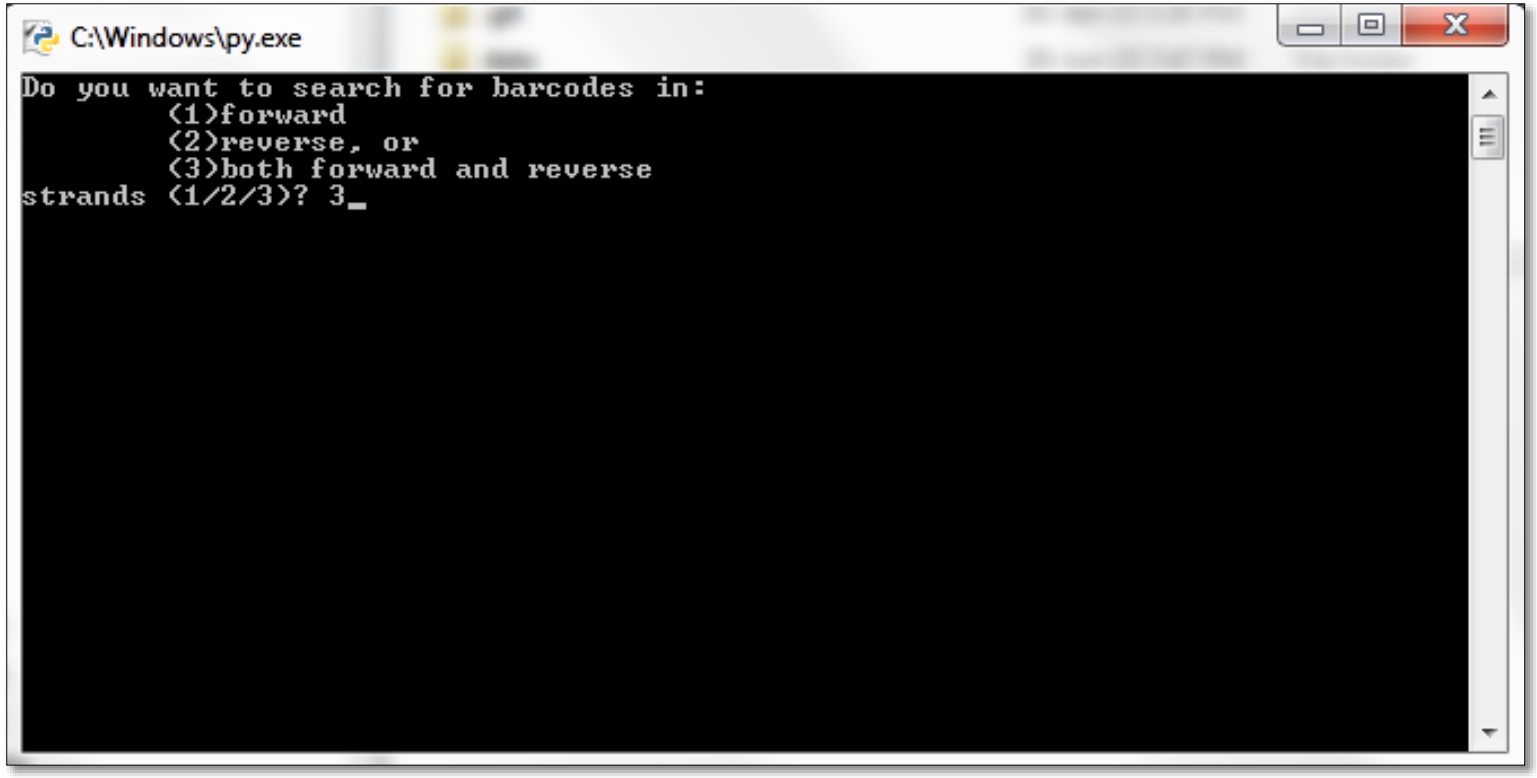


Figure 4. Three different method for sorting the sequences.

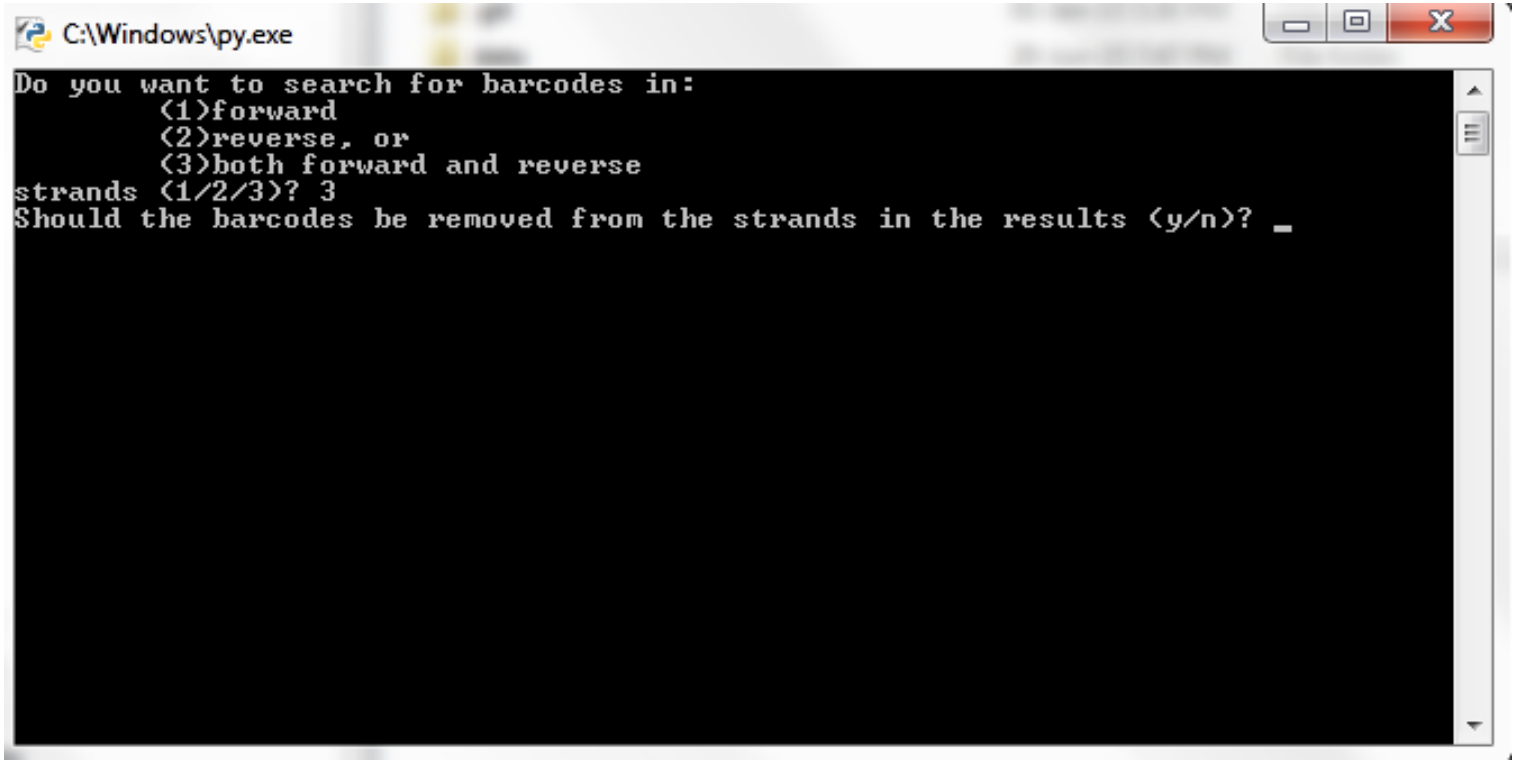


Figure 5. Trimming the barcode sequences from the following reads or run the analysis without trimming.

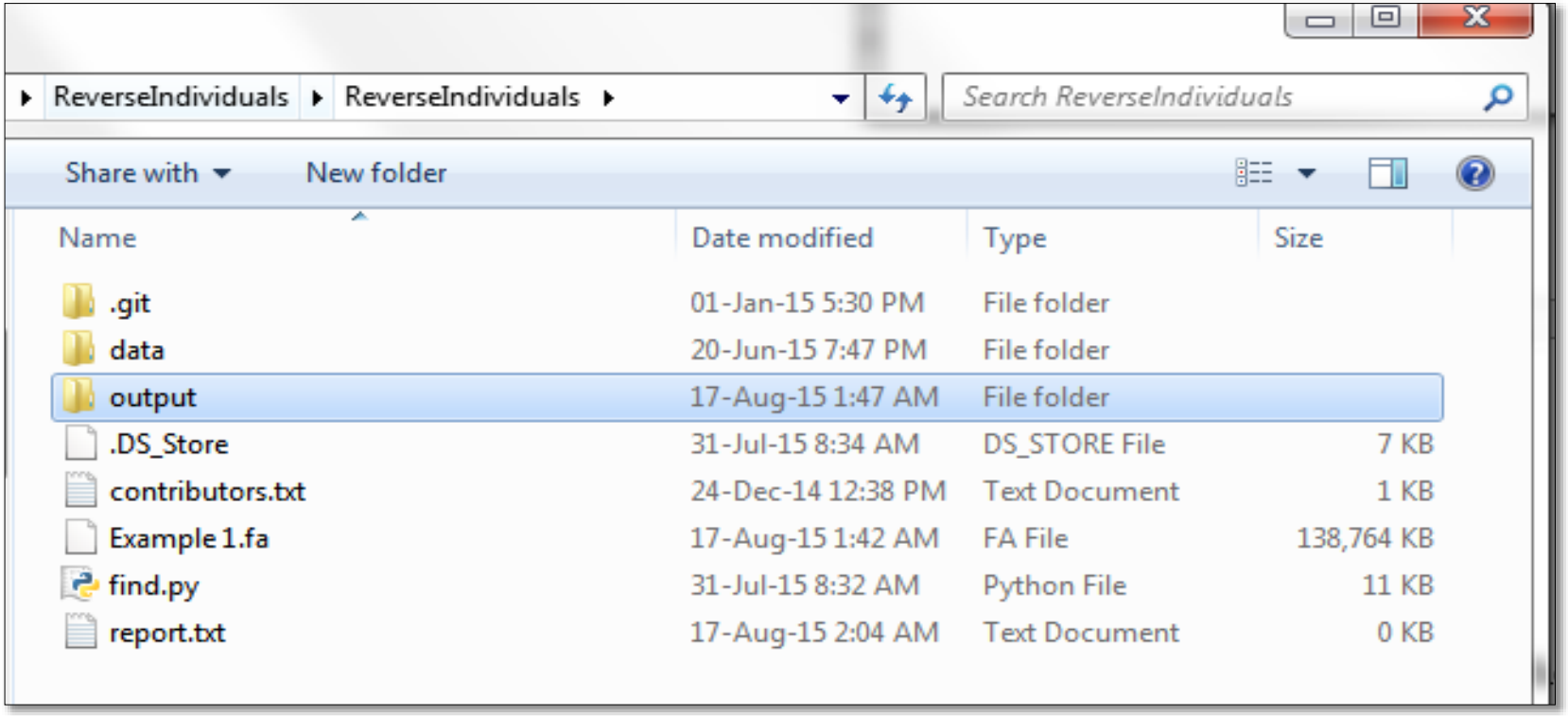


Figure 6: The final results will be available in output folder.

Troubleshooting

If you have difficulty to run the software:

I. Right click on the "find.py" key and choose "Edit with IDLT" option (Fig. 7).

II. Click on run and choose the "Run Module" option (Fig. 8).

III. Finally, follow the steps 2-1 to the end as described above. And collect the data from the output file.

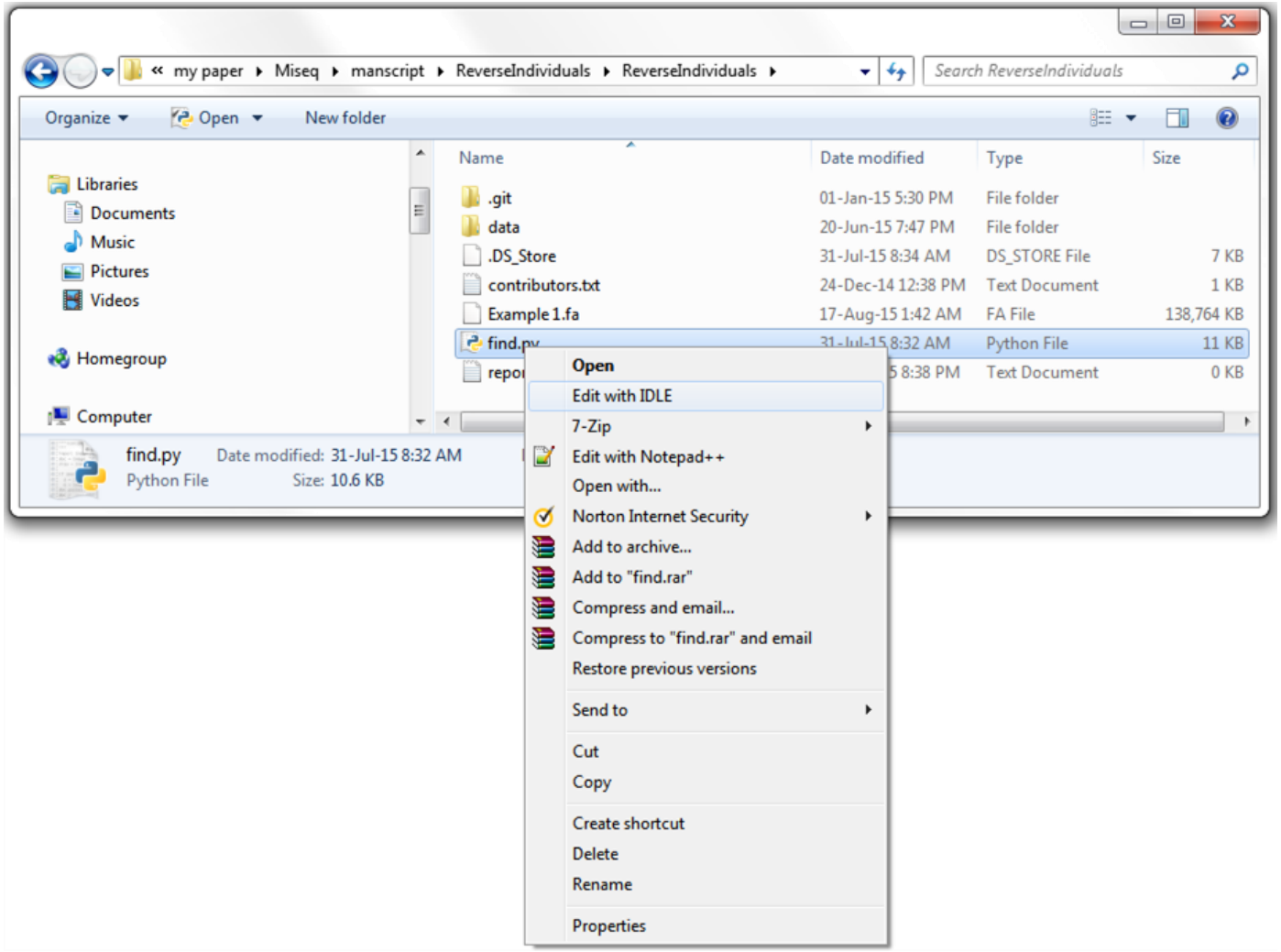


Figure 7: starting software through the EDIT with IDLT option.

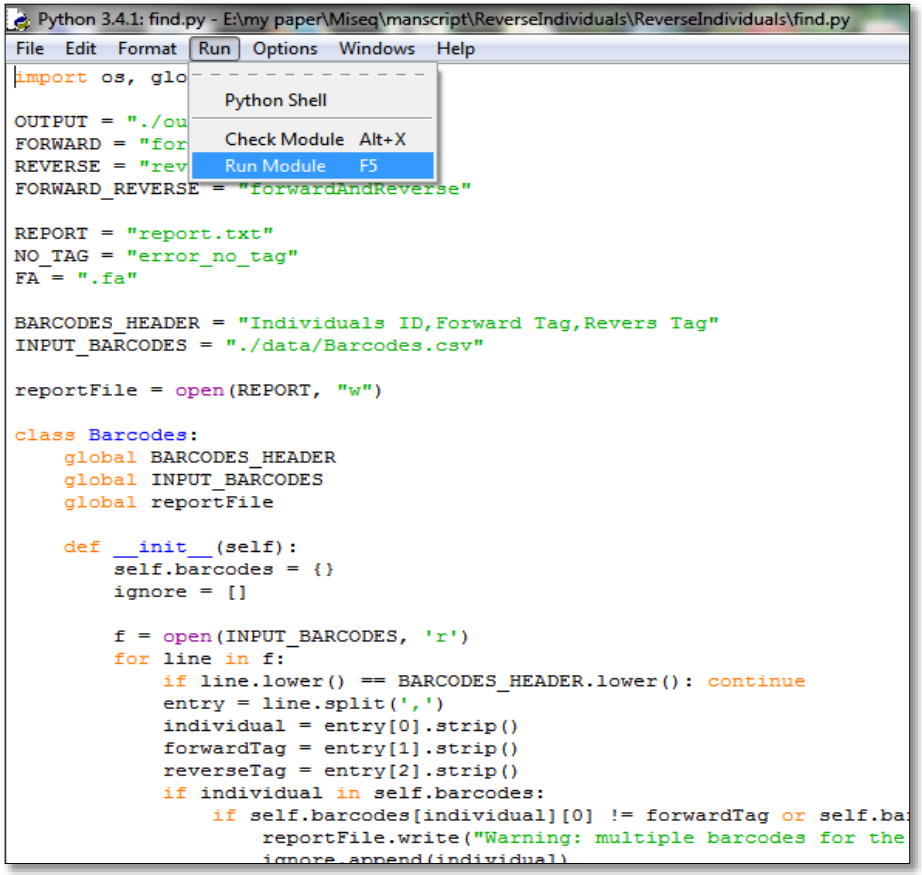


Figure 8: Running the program by selecting the Run Module option.

Chapter six

6. Phylogeography

The chapter has been prepared for submission to the journal Molecular Ecology.

In this chapter I used phylogeographic approaches at a macro-geographic scale to investigate the overall aim of my thesis, which is focused on examining the distribution of genetic variation in response to historical and demographic events. I employed recently described statistical methods, including species distribution modelling and a phylogeographic diffusion model, to further investigate the potential Plio-Pleistocene climatic changes and biogeographic barriers associated with the observed phylogeographic patterns.

The information obtained here help to understand the pattern of genetic structure and the existence of refugia or biogeographical barriers in a widespread and common species, the sleepy lizard, across arid and semi-arid zone of southern Australia., The use of multiple loci (both mitochondrial and neutral nuclear markers) provides insight into the potential common refugia and barriers affecting many species across this landscape, and will be useful for land and conservation management.

Prepared for Molecular Ecology

Plio-Pleistocene diversification and biogeographic barriers in the Australian arid zone reflected in phylogeography of a widespread and common species.

Talat Hojat Ansari, Steven J.B. Cooper, Michael, P. Schwarz, Mehregan Ebrahimi, Gaynor Dolman, Leah Reinberger, C. Michael Bull and Michael G. Gardner

6.1. ABSTRACT

Palaeoclimatic events and biogeographical processes since the mid-Tertiary have played an important role in shaping the evolution and distribution of Australian fauna. However, their impacts on fauna in southern and arid zone regions of Australia are not well understood. Here we investigate the phylogeography of an Australian scincid lizard, *Tiliqua rugosa*, within Australian arid and semi-arid regions. We assessed genetic variability within and among regions across the distributional range of *T. rugosa* in Australia using mitochondrial DNA (mtDNA) and 11 nuclear markers, including nine anonymous nuclear loci (ANL). Phylogenetic analyses revealed three major mtDNA lineages for sleepy lizards, geographically localised north and south of the Murray River in southern Australia and west of the Nullarbor Plain. Molecular variance and population analyses of both mtDNA and nuclear haplotype sequence data revealed significant variation among the three populations associated with each of the mtDNA lineages, although potential

secondary admixture was detected, based on nuclear markers. We used species distribution modelling and a phylogeographic diffusion model to further investigate the potential historical events associated with these phylogeographic patterns. The analyses suggest that the range of *T. rugosa* contracted during the Last Glacial Maxima (LGM) and this contraction appears to coincide with several distinct mtDNA lineages. Overall, our analyses suggest that Plio-Pleistocene climatic changes and biogeographic barriers associated with the Nullabor Plain and Murray River have played a key role in shaping the present-day distribution of genetic diversity in *T. rugosa*.

6.2. INTRODUCTION

Global climates have experienced significant variation over the periods during which present-day species originated and diversified. These fluctuations have affected the evolution and distributions of current species, particularly in the Northern Hemisphere where Ice-Age glaciations led to periodic contractions followed by expansions of the range of species, which in turn led to genetic differentiation and diversification (Pielou, 1991; Cooper, et al., 1995; Taberlet, et al., 1998; Hewitt, 2000; Hewitt, 2004a; Bhagwat and Willis, 2008). However, less is known about the effects of Plio-Pleistocene climate change on the present-day diversity of species in the Southern Hemisphere where the effect of glaciation was less pronounced and where processes may have been more complex (Byrne et al. 2008). This is particularly true of southern Australia (Byrne, et al., 2008), which is the main focus of our current study here.

Phylogeographic studies from southern Australia, especially of arid zone species, have provided an initial understanding the evolution of fauna during the Plio-Pleistocene, however, major deficiencies regarding the historical process shaped biological diversity across the area still exist (Byrne, 2008; Shoo, et al., 2008; Fujita, et al., 2010; Byrne, et al., 2011; Pepper, et al., 2011; Dolman and Joseph, 2012; Edwards, et al., 2012; Lanier, et al., 2013; Rabosky, et al., 2014). Some studies have had a geographically limited focus, with distributions often being relatively narrow and not overlapping (e.g. Chapple, et al., 2005; Dubey and Shine, 2010; Chapple, et al., 2011; Edwards, et al., 2012). Some of the studies have only involved mtDNA analyses (e.g. Symula, et al., 2008; Dolman and Joseph, 2015) although many studies now include nuclear markers (e.g. Cooper, et al., 2000; Shoo, et al., 2008; Kawakami, et al., 2009; Fujita, et al., 2010; Edwards, et al., 2012; Duckett and Stow, 2013; Welton, et al., 2013; Thome, et al., 2014). Additionally, some studies involved multiple species (e.g. Oliver, et al., 2009; Dolman and Joseph, 2012; Edwards, et al., 2012; Dolman and Joseph, 2015) and therefore reflect the influence of different evolutionary forces on patterns of genetic variation, and the vagility of species may have eroded genetic patterns via gene flow. As an general view, although arid zone fauna considered by higher number of researcher in recent years (e.g. Dubey and Shine, 2010; Fujita, et al., 2010; Chapple, et al., 2011; Edwards and Melville, 2011; Dolman and Joseph, 2012; Edwards, et al., 2012; Levy, et al., 2012; Maldonado, et al., 2012; Neaves, et al., 2012; Duckett and Stow, 2013; Lanier, et al., 2013; Welton, et al., 2013; Rabosky, et al., 2014; Thome, et al., 2014; Dolman and Joseph, 2015), most of these studies highlighted the importance of either identifying and characterising refugial

locations (Fujita, et al., 2010), or advocate subjecting a higher number of vertebrate taxa from arid zones to population genetic and phylogeographic studies (Byrne, 2008; Shoo, et al., 2008; Pepper, et al., 2011; Rabosky, et al., 2014).Specifically, there is currently a need to better understand two phylogeographic features of southern Australian arid and semi-arid zone species: firstly the nature of refugia in past climates; and, secondly, the role of biogeographical barriers in shaping the expansion of populations from historical refugia. Three potentially important barriers have been identified, namely the Nullabor plain, Eyrean Barrier (Flinders Ranges–Lake Eyre Basin) and the Murray River. These barriers (Figure 1) have been proposed to play a key role in shaping the present-day diversity of fauna in southern Australia (Cooper, et al., 2000; Byrne, et al., 2008; Kawakami, et al., 2009; Dolman and Joseph, 2012; Dolman and Joseph, 2015). The Nullarbor Plain, which is geographically located in the centre of southern Australia (Figure 1), forms a potentially major biogeographic barrier of unsuitable habitat of approximately 700 km in length between Australia's temperate mesic zones in the south west and south east, during Mid-Miocene (Li, et al., 2004; Crisp and Cook, 2007; Dolman and Joseph, 2012). On the other hand, the Murray River has been identified as a major phylogeographic break between northern and southern populations of some south-eastern Australian species (Cooper, et al., 2000; Kawakami, et al., 2009). An estimation of divergence time between these northern and southern population were during the Pleistocene (Cooper, et al., 2000; Chapple, et al., 2005; Kawakami, et al., 2009). The Murray River is Australia's longest river, stretching from Lake Alexandrina at the South Australian coast up into the start of the highlands of southern NSW and north

eastern Victoria (Figure 1). Finally, the Lake Eyre Basin/Eyrean depression has also been identified as a potential barrier during Plio-Pleistocene in the now-arid regions of northern South Australia (Ford, 1987; Joseph, et al., 2006; Dolman and Joseph, 2015).

The monogamous skink *Tiliqua rugosa*, commonly known as the sleepy or stumpy-tailed lizard, is an excellent model species due to its widespread distribution across arid and semi-arid zone regions of southern Australia (Bull, 1987; Greer, 1989). This distribution encompasses all three of the proposed biogeographic barriers above (the Murray River, Eyrean barrier and Nullabor: see Figure 1), providing an opportunity to further investigate their influence on the biogeographic history of southern Australian arid and semi-arid zone species. We predict that these barriers will leave signatures in the phylogeographic structure of *T. rugosa* and we also predict that the proposed increased aridity during glacial maxima would play a similar role to glaciations in the northern hemisphere, leading to the contraction of the range of species back to one or more refugia located either side of the biogeographic barriers.

We used sequences obtained across the entire range of the species from one mitochondrial gene, NADH Dehydrogenase subunit 4 (*ND4*), and 11 nuclear genes including nine anonymous nuclear loci (ANL). We used both Sanger and next generation sequencing (NGS) technology (Illumina MiSeq) to obtain sequences from these markers. We conduct a range of phylogenetic analyses to investigate genetic structure and infer the likely impact of historical processes on the evolutionary history of *Tiliqua rugosa*. We also use species distribution modelling to determine putative historical refugia. Finally, we use

phylogeographic diffusion model analysis (Lemey, et al., 2010) to determine the geographic location of lineage ancestors, and tested whether they were located in the putative refugial regions identified using climatic modelling analyses.

6.3. MATERIAL AND METHODS

6.3.1. SAMPLING

We obtained tissue and blood samples from 76 specimens of *Tiliqua rugosa* from across the geographical range of the species in Australia (Figure 1). We used frozen or alcohol-preserved tissues and blood collections held by the South Australian Museum (SAM) and the Australian Biological Tissue Collection (ABTC). All samples were collected under appropriate permits (approval no. E324).

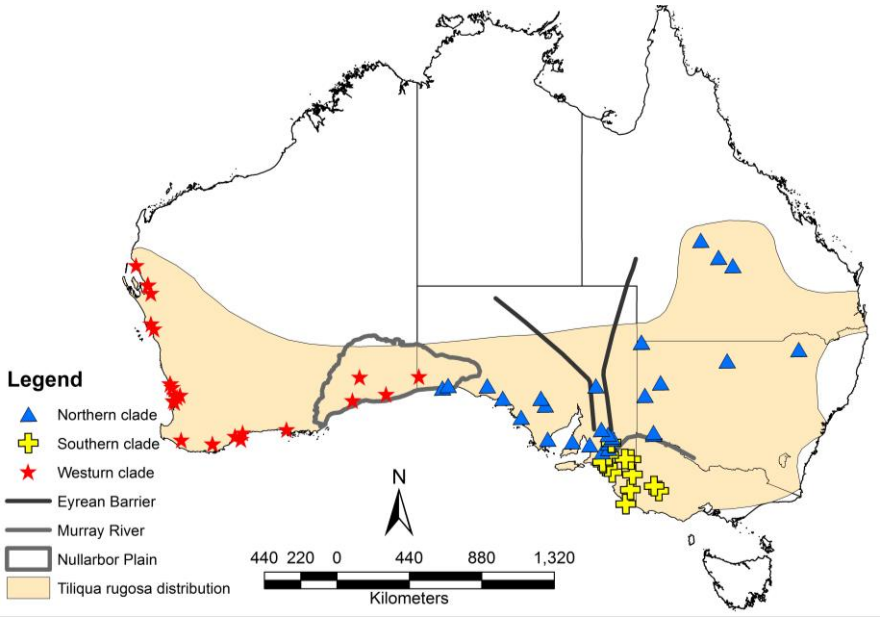


Figure 1: Map showing the collection sites of *Tiliqua rugosa* samples used in this study. The light orange shading shows the distribution of the species across Australia. The distributions of the genetic lineages identified (see Figure 2) are presented with different shapes and colours. The Northern clade is marked with a blue rectangle, the Southern clade is shown with a solid yellow cross and the Western lineage is indicated with a red star. The phylogeographic barriers are also highlighted.

6.3.2. DNA EXTRACTION AND SANGER SEQUENCING

DNA was extracted using a puregene DNA isolation kit (Gentra systems, Minneapolis MW, USA) from frozen or alcohol-preserved tissues and blood samples. We performed DNA amplification of *ND4* using a combination of primers M245, M246 (Gardner, et al., 2008), and M256 (Saint & Smith, unpublished; see supplementary Table S1). The *ND4* mitochondrial gene region has proved to be extremely informative for phylogeographic studies of squamate reptiles (Greaves, et al., 2007; Liggins, et al., 2008a; Liggins, et al., 2008b; O'Neill, et al., 2008). In addition, we sequenced two single-copy nuclear loci, the non-encoding intron 7 of *β-fibrinogen*, and glyceraldehyde-3-phosphate dehydrogenase (*GAPD*), described by Dolman and Phillips (2004). Polymerase chain reactions (PCR) were performed in 25µL volumes, containing 1× reaction buffer, 0.2 mM of each primer, 0.8 mM dNTP and 2 mM $MgCl_2$ for mtDNA loci and 4 mM $MgCl_2$ for nuclear loci, plus DNA polymerase (Ampli Taq Gold™). The cycling profile started with a single cycle of 10 min at 94 °C, followed by 34 cycles of 45 s at 94 °C, 45 s at 55 °C and 1 min at 72 °C, ending with 6 min at 72 °C and 30 s at 25 °C. PCR products were purified for sequencing using a multiscreen PCR 384 Filter Plate (Millipore) and sequenced using ABI Prism Big Dye Terminator Cycle Sequencing chemistry and the manufacturer's recommended protocol. The sequencing programme consisted of 3 min at 96 °C, followed by 30 cycles of 30 s at 96 °C, 15 s at 50 °C and 4 min at 60 °C, ending with 1 min at 25 °C. Reaction products were purified using Multiscreen® 384-SEQ Plates and a vacuum manifold, resuspending the products in 25µL of 0.3mM EDTA. Capillary separation of sequencing products was out-sourced to AGRF

(Australian Genome Research Facility, Adelaide; www.agrf.org.au). We used the software packages Molecular Evolutionary Genetics Analysis (MEGA) version 6 (Tamura, et al., 2013) to edit and align sequences of all individuals for all three markers.

6.3.3. ANONYMOUS NUCLEAR LOCI NEXT GENERATION SEQUENCING

All 76 specimens were sequenced for nine ANL (*Tiru-9*, *Tiru-16*, *Tiru-23*, *Tiru-34*, *Tiru-39*, *Tiru-44*, *Tiru-45*, *Tiru-46* and *Tiru-48*; Ansari *et al.*, 2015) using the Illumina MiSeq platform (Table S1) following the protocol described in Ansari et al. (Chapter 5). Briefly, we amplified the specified ANL loci following multiplex-ready PCR technology (MRT) (Hayden, et al., 2008). Firstly, we modified the forward and reverse locus specific (LS) primers to include a linker of 15bp (ACGACGTTGTAAAA) and 16bp (CATTAAGT TCCCATTA) respectively, at their 5´ ends. Then we developed a tag primer composed of a universal adaptor for Illumina MiSeq (P5 or P7), and 8 bp MID tags (barcodes) plus Illumina sequencing primer (Read 1 and Read 2) and the linker. The linkers allow the tag primer to attach to the LS Primer during the PCR, in order to barcode tag and multiplex our amplicons for a single MiSeq run. Each lizard sample was PCR-amplified using a specific tag primer with the unique MID tag, and the same set of MID tags used across the different markers per individual (PCR conditions are included in Supplementary). Pooled amplicons were sequenced on an Illumina MiSeq Platform for 250bp

paired end read length at the AGRF (Melbourne, Australia) following standard protocols.

We assessed the read qualities using FASTQ (http:// www.bioinformatics.Babraham.ac.uk/projects/fastqc/) and also CLC genomic workbench version 7.5.0 (http://www.clcbio.com). The reads were then trimmed for which the base quality scores (Q) was lower than 30 using the CLC genomic workbench. Finally, we performed de-multiplexing on MITAGSORTER workbench (Ansari et al., Chapter 5), and sorted the sequences for each individual at each targeted locus.

The reads for each locus were mapped to a database of reference sequences (ANL sequences from Ansari et al. 2015) using CLC Genomics workbench version 7.5.0. We culled reads that did not align to the references. We constructed our ANL haplotype sequences and checked them by visual review of the aligned reads through CLC genomic workbench and also by using Geneious version 8.1 (Kearse, et al., 2012). The haplotypes were aligned for all individuals and loci using Geneious version 8.1.

6.3.4. MITOCHONDRIAL PHYLOGENETIC ANALYSIS AND MOLECULAR DATING

To find the most likely model of sequence evolution for our *ND4* dataset we used a Bayesian Information Criterion implemented in the software package MEGA, and this analysis indicated an HKY+ I + Γ model was most appropriate and this model was used for the subsequent phylogenetic analysis. Phylogenetic trees were constructed for *ND4* sequences using neighbour-joining

(Wielstra, et al., 2013) based on an uncorrected "P" distance with tree bisection and reconnection for the heuristic search in PAUP v4.0 b (Swofford, 2002). Additionally, we performed Bayesian phylogenetic analyses for the *ND4* sequences using BEAST 1.6.2 (Drummond and Rambaut, 2007). Additional sequences of species closely related, but forming an external monophyletic clade, to *Tiliqua rugosa* (Gardner et al. 2008) were included in our phylogenetic tree analyses as an outgroup (five species of *Egernia: E. cunninghami,* KM211515; *E. striolata,* KM211516;, *E. saxatilis,* KM211517; *E. hosmeri,* KM211518; and *E. stokesii* KM211519). We also included two *Tiliqua* species for dating of ancestral nodes and to help root the *T. rugosa* haplotype tree (*T. adelaidensis*, KM211511; and *T. scincoides* KM211512). However, for Bayesian analysis, we used two approaches to estimate relative divergence times in *T. rugosa* lineages using the total 83 *ND4* sequence data set. Analyses were performed for 70 million generations, sampling every 10,000 and a burn-in period of 1 million generation. We applied both Birth and Death, and Bayesian Skyline as tree models in separate analyses. We also used the Birth and Death model to calculate the relative divergence time. To assess the Bayesian Skyline plot we performed the Bayesian Skyline model. We used Tracer v1.5 (Rambaut and Drummond, 2009b) to estimate the effective sample size of each parameter, check for model convergence, and to visualize parameter plots.

In details, to place a timeframe on our phylogenies, we firstly used the estimated divergence time between *Egernia* and *Tiliqua* as a calibration point, based on fossils attributable to *Egernia* that date to the late Oligocene to early–middle Miocene (Hutchinson, 1992; Shea and Hutchinson, 1992; Martin,

et al., 2004). In support of these dates, Estes (1984) reported the presence of both *Tiliqua* and *Egernia* in the middle Miocene (ca. 15 million years ago). To constrain the divergence age priors, we used a lognormal distribution (Drummond, et al., 2006; Ho, et al., 2007; Ho, 2007) with an offset of 15 Ma (million years ago) and standard deviation (SD = 0.907) so that 95% of the sampled divergence date priors fell within the 15-22 Ma time span. However, as the fossils' age is not an exact time and the mutation rate estimate (see below) is based on the divergence rate of mtDNA in other reptilian species, the haplotype divergence times estimated here should be interpreted cautiously.

As a secondary approach for estimating divergence times for the *T. rugosa* phylogeny, we used a strict clock model and applied a rate of 1.0/mutations/site/generation in order to define the resulting branch lengths in units of mean nucleotide substitutions per site. The branch lengths can then be converted to years using mutation rate and generation time. We used Tracer v1.5 (Rambaut and Drummond, 2009a) to reconstruct the Bayesian skyline plot. Based on estimates from other squamate groups, mitochondrial mutation rates range from 0.47% to 2.3% per million years. The lower limit was based on an estimate by Zamudio and Greene (1997) for viper mtDNA. The upper limit of 2.3% per Myr (million years) was derived from analyses of Canary Island skinks and Balearic Islands lacertids (mean rate of 2.05% Myr) reported by Brown and Pestano (1998) and Brown, et al. (2008). However, to convert the branch length to years in this study, we used a divergence rate of 0.65% per million years which is in the range of mitochondrial mutation rates (0.47%-2.3%) as we mentioned above and also had been used in the study of

an agamid lizard (Hugall and Lee, 2004). In order to convert the units given in our result to years, an estimate of the likely number of generations of *T. rugosa* was required. Demographic studies of *T. rugosa* have suggested that this viviparous Australian skink lives for more than 30 years (Bull, 1995) and usually takes approximately three years to reach maturity (Bull, 1987). In addition, not all females are receptive every year and they may skip reproduction in some years due to their reproductive morphology (Egan, 1984), and environmental conditions or body condition (Bull and Pamula, 1996). However, for simplicity, we applied the age at sexual maturity (three years) as the value for the generation time in our analysis.

In addition, we generated a haplotype network, for *T. rugosa* nucleotides and translated amino acid sequences. We used a Maximum Likelihood (ML) tree constructed with Mega v6 (Tamura, et al., 2013) to generate the haplotype networks with HaploViewer (available at http: //www.cibiv.at/~greg/haploviewer).

6.3.5. CONTINUOUS-DIFFUSION PHYLOGEOGRAPHIC ANALYSIS

We defined the geographic origin of *T. rugosa* using a Relaxed Random Walk (RRW) diffusion model in Beast v2.2.1 (Bouckaert, et al., 2014). This model reconstructs the spatial diffusion process across a continuous landscape at any point of time by using the sampling location of the sequences. The RRW model accounts for uncertainty in tree topology and allows variation in diffusion rates along the phylogeny branches (Lemey, et al., 2010). The analyses allow an estimation of the location of the most recent common ancestor (MRCA) of each lineage, which potentially represents the location of

refugia. We included only *T. rugosa* individuals for the RRW model analysis. An initial run of 70 million generations with sampling every 10,000 generations using the RRW model resulted in low values for effective population sizes (ESS). Therefore, we performed three independent runs of 400 million generations, sampling every 10,000 generations, and used LogCombiner v2.2.1 (Bouckaert, et al., 2014) to obtain a combined estimation of the posterior parameters across the three runs. We estimated the maximum clade credibility (MCC) tree using Tree Annotator. We then projected the MCC tree onto geographic coordinates using SPREAD v1.0.6 (Bielejec, et al., 2011) and visualized the generated Kml file in Google Earth.

6.3.6. NUCLEAR DATA ANALYSIS

The software package GENEIOUS v8.1 (Kearse, et al., 2012) was used to edit and align sequences of all individuals and genes. Sequencing of the *β-fibrinogen* intron 7 and *GAPD* indicated that several sites were heterozygous for single nucleotide polymorphisms (SNPs). Therefore haplotype phases were then reconstructed using a Markov chain Monte Carlo (MCMC) algorithm as implemented in PHASE v2.1 (Stephens, et al., 2001; Stephens and Scheet, 2005). This program was run multiple times for each dataset and haplotype frequencies and goodness of fit measures were compared across runs to assess the reliability of results.

Nuclear haplotype networks were generated for each locus with HaploViewer (available at http://www.cibiv.at/~greg/haploviewer) using a Maximum Likelihood (ML) tree constructed with Mega v6 (Tamura, et al., 2013). To calculate the genetic distances for the multiple loci, we constructed

a single genetic distance matrix by combining the allele distance matrix from each locus using POFAD (Joly and Bruneau, 2006). We generated the distance matrix for each locus using uncorrected p genetic distances in MEGA v6 (Tamura, et al., 2013). Consequently, we calculated a mean pairwise distance from the 11 distance matrices using POFAD. In addition, we coded the individuals for the 11 loci based on their haplotypes using the GenAlEx v6.5 (Peakall and Smouse, 2006; Peakall and Smouse, 2012), and then we calculated genetic distances for the multiple loci. In all subsequent analyses based on genetic distances we first used the multilocus genetic distance matrix generated by GenAlEx and the second matrix generated by POFAD.

To test whether there was statistically significant variation among and within mtDNA defined lineages we performed an analysis of molecular variance (AMOVA). The matrix of multilocus genetic distances was also used to perform a principle coordinate analysis (PCoA) in GenAlEx (Peakall and Smouse, 2006; Peakall and Smouse, 2012).

To investigate population structure at the nuclear loci, we also used a Bayesian cluster analysis of allelic data for each of the 11 loci, implemented in STRUCTURE v2.3.3 (Pritchard, et al., 2000). We performed 10 replicate runs for each distinct population (K) and with two independent runs we tested both admixture and non-admixture models with correlated allele frequencies among populations. Each replicate involving 1,000,000 steps, with a burn-in phase of 100,000 steps for a K value ranging from 1 to 6. The most likely value of K was estimated following the method of Evanno *et al.* (2005) using the STRUCTURE Harvester (Earl and vonHoldt, 2012). Furthermore, following

the method described above we performed additional structure and also PCoA analysis, excluding the populations that give strong patterns.

6.3.7. MOLECULAR DIVERSITY AND HISTORICAL DEMOGRAPHY

We calculated standard diversity statistics including nucleotide (π) diversities, number of haplotypes (h), number of polymorphic sites (s), level of variation among and within populations for the populations associated with the mtDNA lineages, for mtDNA and each nuclear marker separately, using Arlequin v3.5.1 (Excoffier and Lischer, 2010a) and MEGA v6 (Tamura, et al., 2013).

To estimate the genetic differentiation among the mtDNA defined lineages, we calculated pairwise population F_{st} values for mtDNA sequences, following Slatkin and Hudson (1991) and using the program Arlequin v3.5.1 (Excoffier and Lischer, 2010a). The significance of F_{st} values were obtained after 1000 permutations. In addition, to test whether there was statistically significant variation among and within the populations, we used an analysis of molecular variance (AMOVA) for mtDNA sequences using Arlequin v3.5.1 (Excoffier and Lischer, 2010a).

Tajima's D (1989), and Fu's F_s (1997) tests were independently calculated for mtDNA and each nuclear marker using DnaSP v5.0 (Librado and Rozas, 2009), in order to identify any possible deviation from neutrality which could be either an effect of natural selection or the result of a past demographic expansion. Recent population expansion would be indicated by a negative value of both D and Fs, and also because both of these tests assume

a historically stable population (Nei and Kumar, 2000) a significant P value would indicate rejection of the null hypothesis of a historically stable population. To detect significant departures from a constant population we simulated Ramos-Onsins and Rozas's R2 parameter in DnaSP v5.0. (Librado and Rozas, 2009). We also used the mtDNA sequences, to calculate Tajima's D, Fu's F_s and R2 for each mtDNA defined lineages separately.

In addition, we further tested for population expansion based on *ND4* sequences by simulating the frequency distribution of pairwise genetic differences (or mismatch distributions) within mtDNA lineages. We compared differences between observed and simulated distributions with the sum of square deviations (SSD) as well as the Harpending's raggedness index (RI) (Harpending, 1994) and compared with the data calculated under the sudden expansion model. A significant P value ($P \leq 0.05$) indicating rejection of the hypothesis of population expansion. All tests used mtDNA sequences and were performed in Arlequin v3.5.1 (Excoffier and Lischer, 2010a). Additionally, in order to test for selection on *ND4* we used McDonald-Kritman (MK) test (Mcdonald and Kreitman, 1991) implimented in DnaSP v5.0 (Librado and Rozas, 2009).

6.3.8. THE MODEL OF SPECIES DISTRIBUTION

We used MAXENT v3.3.3k (Phillips, et al., 2006; Phillips and Dudik, 2008) to generate a distribution model for *T. rugosa* from environmental data and locality records. We projected this model onto climatic layers indicating the glacial maximum and interglacial conditions to assess the possible changes in geographic distribution of the studied species across southern Australia

under Pleistocene climate fluctuations. We used the present-day climatic conditions to represent interglacial climates, and for the reconstruction of the Last Glacial Maximum (LGM) conditions, we used a community climate system model (CCSM) and a model of interdisciplinary research on climate (MIROC). We downloaded the environmental data including six bioclimatic variables with 2.5 arc-min resolution from the WordClim website (www.worldclim.org). Our selected variables included monthly average of maximum and minimum temperature (degrees C * 10), monthly total precipitation (mm), mean temperature of wettest quarter, precipitation of driest quarter, precipitation of wettest quarter and mean temperature of coldest quarter. We used approximately 7000 locality records, from our samples data and Atlas of Living Australian archives for *T. rugosa* (http://www.ala.org.au. accessed 3 June 2015).

We used a linear and hinge feature to generate the MAXENT model. Then we evaluated the predictive performance of the model through fivefold cross-validation and compared the areas under the receiver operating characteristic curves (AUC) to train and test data. The output format was selected that scaled an approximate linear probability of the presence between 0 and 1. However, we could not use this as a direct estimation of probability due to lack of information regarding the sampling effort for locality records (Elith, et al., 2011). Therefore, we interpreted our results initially based on differences in the probability of occurrence between glacial and inter-glacial conditions, instead of using absolute values of probability.

6.4. RESULTS

6.4.1. SEQUENCE VARIABILITY

The mitochondrial sequence data were collected from 76 individuals comprised 438 base pair (bp) of *ND4* and 137 bp of tRNA, which includes tRNA-His, tRNA- ser and tRNA-leu. This 575 bp fragment contained 95 variable sites, of which 66 were parsimony-informative, defining 45 mtDNA haplotypes across the studied populations (Table 1). The number of successfully sequenced individuals was less for the nuclear loci due to technical constraints (Table 1). In most individuals, based on the presence of heterozygous sites, two alleles (haplotypes) were identified for the nuclear loci. The fragment sizes for each nuclear locus ranged between 119 (*Tiru-16*) to 748 bp (*β-fibrinogen*), with a maximum of 50 variable sites and 26 parsimony informative sites (Table 1). The total number of nuclear haplotypes was between six (*Tiru-45* and *Tiru-46*) and 48 (*GAPD)* (Table 1).

Table 1. Standard diversity statistics for mtDNA and nuclear loci examined in the study. The number of individuals amplified for each locus (N); length of the sequences in base pairs (L); number of segregating (polymorphic) sites (S); number of haplotypes (H); nucleotide diversity (π). The significant values of Tajima's D, Fu's Fs, and Ramos-Onsins and Rozas's *R*2 test are shown with asterisk∗.

Locus	N	L	S	H	π	Tajima's D	Fu's F_s	R2
mtDNA	76	574	91	45	0.044	1.12	-4.45	0.13
***β-fib* intron 7**	64	748	50	38	0.002	-2.56*	-17.99*	0.03*
GAPD	68	307	31	48	0.009	-1.83*	-26.93*	0.03*
Tiru-9	67	145	9	11	0.003	-2.09*	-12.28*	0.03*
Tiru-16	68	119	9	13	0.003	-2.07*	-12.05*	0.03*
Tiru-23	62	160	7	9	0.003	-1.38*	-3.41*	0.04
Tiru-34	65	141	6	7	0.003	-1.42*	-4.25*	0.05
Tiru-39	58	145	8	8	0.005	-1.43*	-1.77	0.06
Tiru-44	61	125	4	8	0.005	-0.39	-0.87	0.08
Tiru-45	42	169	5	6	0.004	-1.58*	-7.13*	0.03
Tiru-46	69	185	5	6	0.003	-0.82	-2.03	0.06
Tiru-48	42	128	13	11	0.03	0.58	-1.33	0.12

6.4.2. MITOCHONDRIAL PHYLOGENETIC ANALYSIS AND MOLECULAR DATING

The two phylogenetic methods yielded similar arrangements for the main clades, so we only present the Bayesian phylogenetic tree (Figure 2A). Three main, well supported, mtDNA lineages were evident within our samples of *T. rugosa*: a 'Western' lineage (support value for Neighbour Joining (NJ) and Bayesian Analysis (BA) were respectively 100 and 1.0) that comprised samples from Western Australia (WA); a 'Southern' lineage (support values

of 100 and 1.0 for NJ and BA respectively) that comprised samples from south-east of the Murray River in South Australia (SA) and Victoria; a 'Northern' lineage (support values of 100 and 1.0 for NJ and BA respectively) that comprised samples from New South Wales (NSW), Queensland (QLD) and north of the Murray River in SA (Figure 2A, B and C).

The division between the Northern and Western mtDNA lineages of *T. rugosa* occurred at the Nullarbor Plain in South Australia. We henceforth refer to the three different lineages as the 'Northern', 'Southern' and 'Western' clades.

The entire *T. rugosa* dataset included 45 different nucleotide haplotypes, 24 from the Northern clade, eight from the Southern clade and 13 from the Western clade. Haplotype diversity was highest, 0.98 ± 0.01, in the Northern clade, and lower in the Southern and Western clade (0.82 ± 0.08 and 0.76 ± 0.07 respectively). The Southern clade had the lowest nucleotide diversity (0.005 ± 0.003) than in the other two clades (Western: 0.011 ± 0.006; Northern: 0.02 ± 0.008).

The mean pairwise differences among nucleotide haplotypes from the Northern and Southern clades was 0.7, among Northern and Western lineages it was 0.8, and among haplotypes from the Southern and Western clade was 0.84. In addition, an AMOVA showed significant ($P < 0.005$) variation among the three lineages, with 13.81% of the variation distributed among them and 86.19 % of the variation distributed within clades. These results suggest there is significant genetic structure across the range of *T. rugosa*, based on mtDNA

variation, with population breaks occurring across the Nullarbor Plain and across the Murray River in South Australia.

Using a calibration of the estimated divergence time of 15 Ma between *Egernia* and *Tiliqua*, we obtained a divergence time for the node connecting the western mtDNA lineage to a clade comprising the samples east of the Nullabor of 2.7 Ma (95% HPD = 0.9 - 4.9), and an age of approximately 1.5 Ma (95% HPD = 0.4 - 2.7) for the divergence time of the Northern and Southern mtDNA lineages (Figure 3A). Similar age estimates for these nodes were also obtained using a Bayesian Skyline plot analysis (Figure 3B and C).

Given a generation time of 3 years in *T. rugosa* and a mutation rate of 0.65×10^{-8} per year leads to an estimate of approximately 1,670,000 years ago for the age of the node connecting the Western clades and the clades east of the Nullabor. Following the same procedure, a divergence time of approximately 1,000,000 years ago was obtained for the node connecting the Northern and Southern mtDNA lineages.

6.4.3. CONTINUOUS-DIFFUSION PHYLOGEOGRAPHIC ANALYSIS

We performed analyses for 400 million generations independently three times, and we obtained a generally high value (>100) for the effective sample size (ESS) based on the combined results. We present the median estimated location of the most recent common ancestor (MRCA) of lineages in Figure 2. The Western lineages originated possibly from multiple refugia near the west corner of Western Australia. A southern edge of the Nullarbor Plain near the coastline also defined a potential refugial location for the Western population

(Figure 2A and B). The analysis further indicate that the Northern *T. rugosa* lineage was associated with multiple refugia (Gawler ranges and Mount lofty ranges) in South Australia. The potential refugial location for the Southern lineage was defined as an area close to, but south-east of the Murray River in South Australia (Figure 2D).

6.4.1. NUCLEAR DATA ANALYSIS

We used haplotype networks to visualise the relationships among haplotypes from all nuclear loci, highlighting each of the three clades identified from the mtDNA analyses (Figure 4). The networks for two nuclear loci, *β-fibrinogen* and *GAPD,* were generated after reconstruction of the haplotype phase, with >90% probability. The networks of different loci all showed some degree of haplotype sharing among the three clades, usually for haplotypes with the highest frequency within *T. rugosa* (Figure 4). However, almost all the loci also showed that there were multiple private haplotypes associated with each of the three populations, with the exception of locus *Tiru-23, Tiru-45* and *Tiru-46,* where private haplotypes were absent for the southern population (Figure 4). The PCoA indicated that individuals assigned to the Western clade formed a relatively distinct cluster, separated from Northern/Southern clades, the latter not showing clear differentiation in the analysis of all individuals (Figure 5A and 5B). A small number of individuals from the Western clade that showed admixture with the Northern clade were from the boundary of the two clades across the Nullabor Plain.

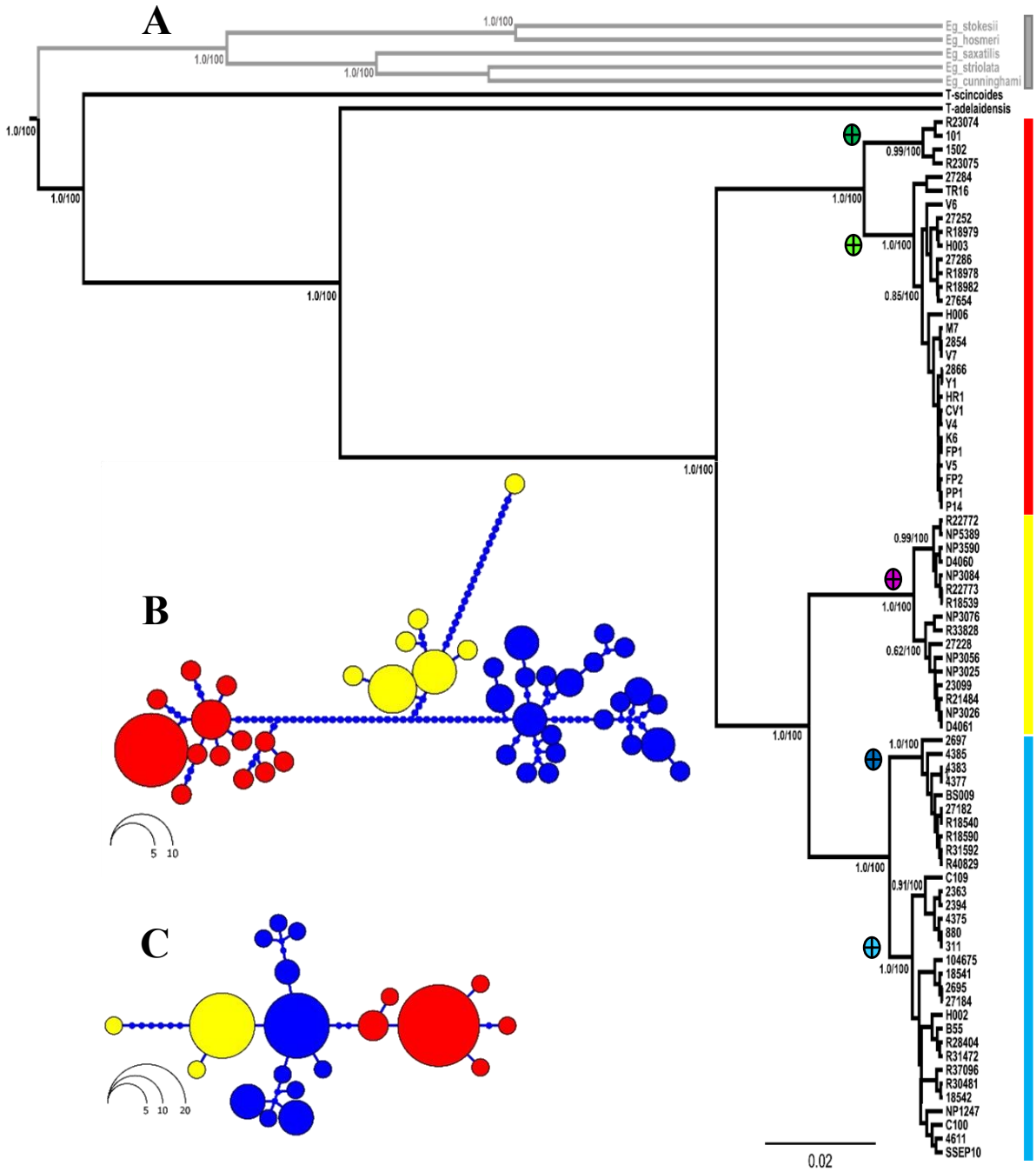


Figure 2 A-C: Phylogenetic tree of mitochondrial DNA (*ND4* region) data from *Tiliqua rugosa* based on Bayesian analysis (BA). Values on branches are indices of support >50% for the major branches for BA and Neighbour-joining bootstrap values. Outgroups are shown by vertical bars with grey colour and western lineages are shown with a red colour. Yellow and blue vertical bars indicate Southern and Northern populations, respectively. The clades that are shown with coloured circles at their nodes in the tree represent the MRCAs as indicated in D. The colour schemes is as describe as D. B: Nucleotide haplotype network for *ND4*. C: Amino acid haplotype network for *ND4*. The colour schemes is as same as A.

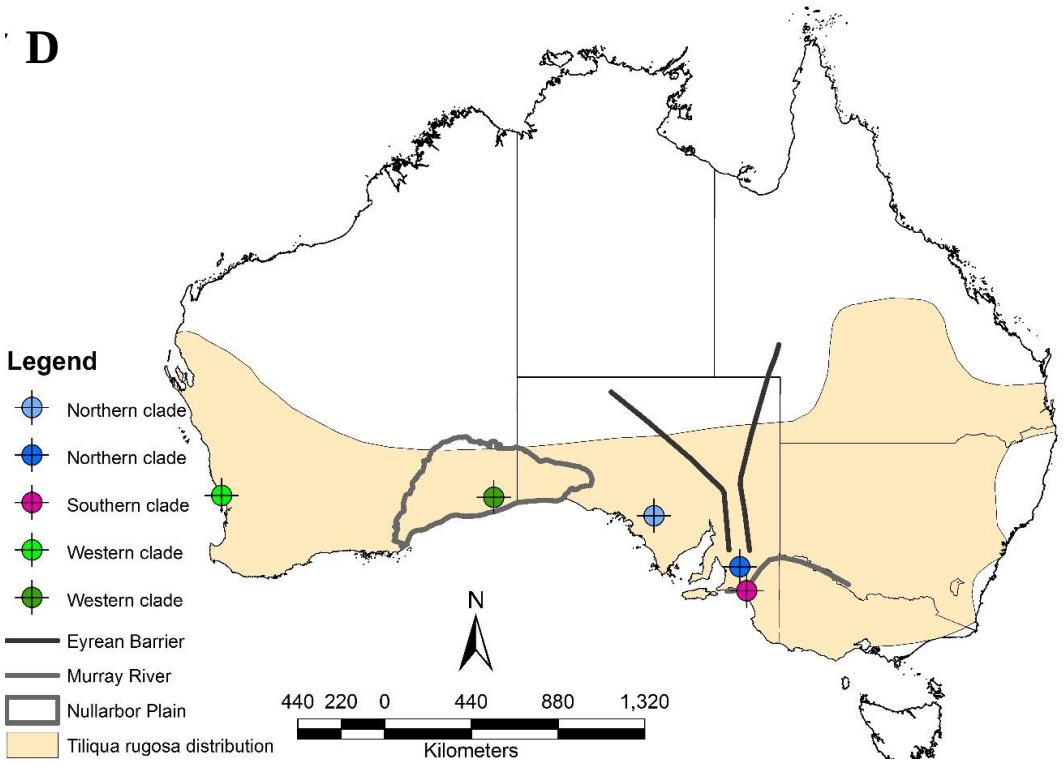


Figure 2D: A map showing the estimated geographic location for the MRCAs of the three main lineages. The mean location for the nodes are shown with coloured circles. The green circles show the estimated location of Western Australia with the darker shade represent the southern edge of Nullarbor Plain. The Blue colour represents the potential geographic location of refugia for the Northern lineages with the Gawler range shown with the lighter blue and the darker blue indicating the Mount Lofty ranges. The mean location for the node for Southern lineages is indicated by a purple circle.

Analysis of molecular variance using the multilocus genetic distance showed significant variation among the three mtDNA lineages with 12 % of the variation distributed among lineages and 88% of variation within lineages.

Results of the STRUCTURE analysis of the nuclear data showed that K=3 was the most likely number of genetic clusters, based on the Ln P (K) and ΔK (Evanno et al. 2005) methods, with clusters broadly corresponding to the three main mtDNA lineages (Figure 5C). There was some degree of admixture between the three genetic clusters, particularly between the

Northern and Southern clades. The admixed individuals from the Western clade were mainly from a contact zone (Nullarbor Plain) with the Northern clade.

Additional STRUCTURE analysis, excluding individuals from the Western clade indicated that K=2 was the most likely splitting the individuals based on the previously defined Northern and Southern mtDNA lineages (Figure 5D).

6.4.2. HISTORICAL DEMOGRAPHIC ANALYSES

Observed mismatch distributions for the Southern clade reasonably matched a unimodal distribution and for both the Western and Northern clades they were multimodal (Figure 6). Additionally, the mismatch distributions did not differ significantly (P> 0.05) from either a demographic or spatial expansion model for all populations except the Western population, which showed a significant difference for the Harpending's raggedness index (RI) model, but not the sum of square deviations (SSD) model (Table 2). Fu's Fs-statistics were negative for each of the three lineages, but the P-value was statistically significant only for the Northern clade. Ramos-Onsin and Rozas' $R2$ statistics did not attain statistical significance only for the Western clade (Table 2). Tajima's D-value was also negative for each population, but only significant for the Northern clade. The Tajima's D and Fu's Fs were significantly negative for most nuclear markers, however R2 statistic was significant just in four nuclear marker out of 11 marker we tested here (Table 2).

The Bayesian skyline plot analyses indicated a relatively stable population size for the majority of the past but at some point there was a dramatic increase in population size possibly during the Middle to the late Pleistocene (0.2-0.02 Ma) with a return to stable in the more recent past (Figure 3B and C). These changes in population size correspond to the emergence of new lineages (Figure 3C).

MK test for selection on *ND4* tests between Western and Northern or between Northern and Southern clades were not significant ($P > 0.05$), and then positive selection can be rejected.

Table 2. Mismatch and neutrality tests of mtDNA (*ND4*), for the combined dataset and each mtDNA lineage. SDD = sum of square deviations; RI= Harpending's raggedness index; hg = Harpending's raggedness index. Tajima's D, Fu's Fs, and Ramos-Onsins and Rozas's *R*2 test An asterisk * signifies when the P values reached the statistical significance based on coalescent simulation (1,000 replications).

Lineage	**Demographic expansion model**		**Spatial expansion model**		**Tajima's D**	**Fu's F_s test**	**R2**
	SDD	RI	SDD	RI			
Combined	0.01	0.01	0.014	0.01	1.12	-4.45	0.13
Western	0.076	0.169*	0.05	0.17	-1.318	-1.01	0.07*
Northern	0.022	0.013	0.02	0.01	-0.208	-8.44*	0.105
Southern	0.017	0.045	0.01	0.04	-0.563	-2.00	0.1*

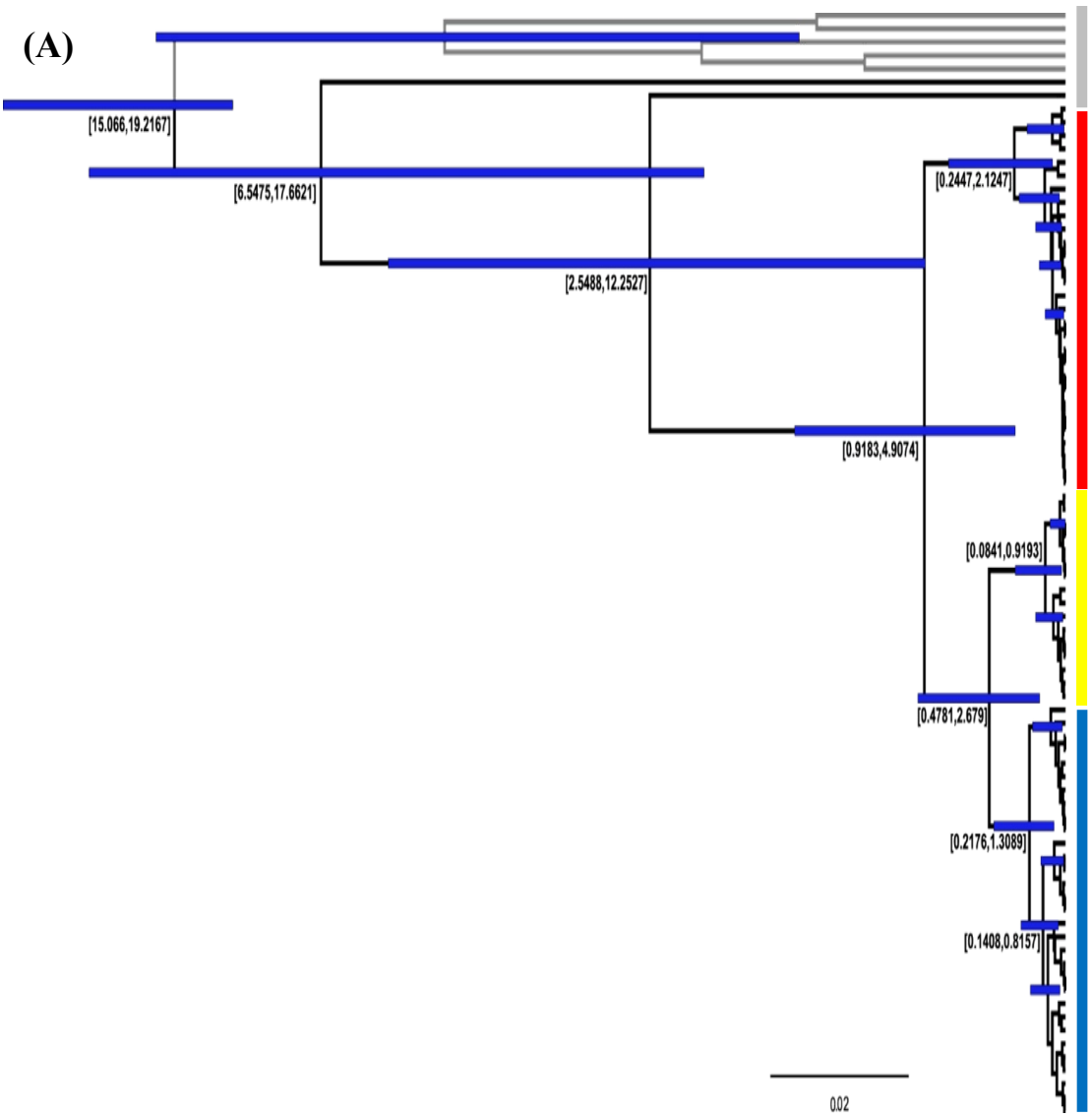


Figure 3A. Age estimate of effective population size for *Tiliqua rugosa* lineages. Colour schemes for lineages in as describe as Figure 2A. Bayesian tree for *Tiliqua rugosa* from fossil calibration (the estimated divergence time between *Egernia* and *Tiliqua*) implemented in Beast. The values represent the 95% HPD for tree height. Vertical dark blue bars show the changes in the height that correspond to features of the phylogenetic tree.

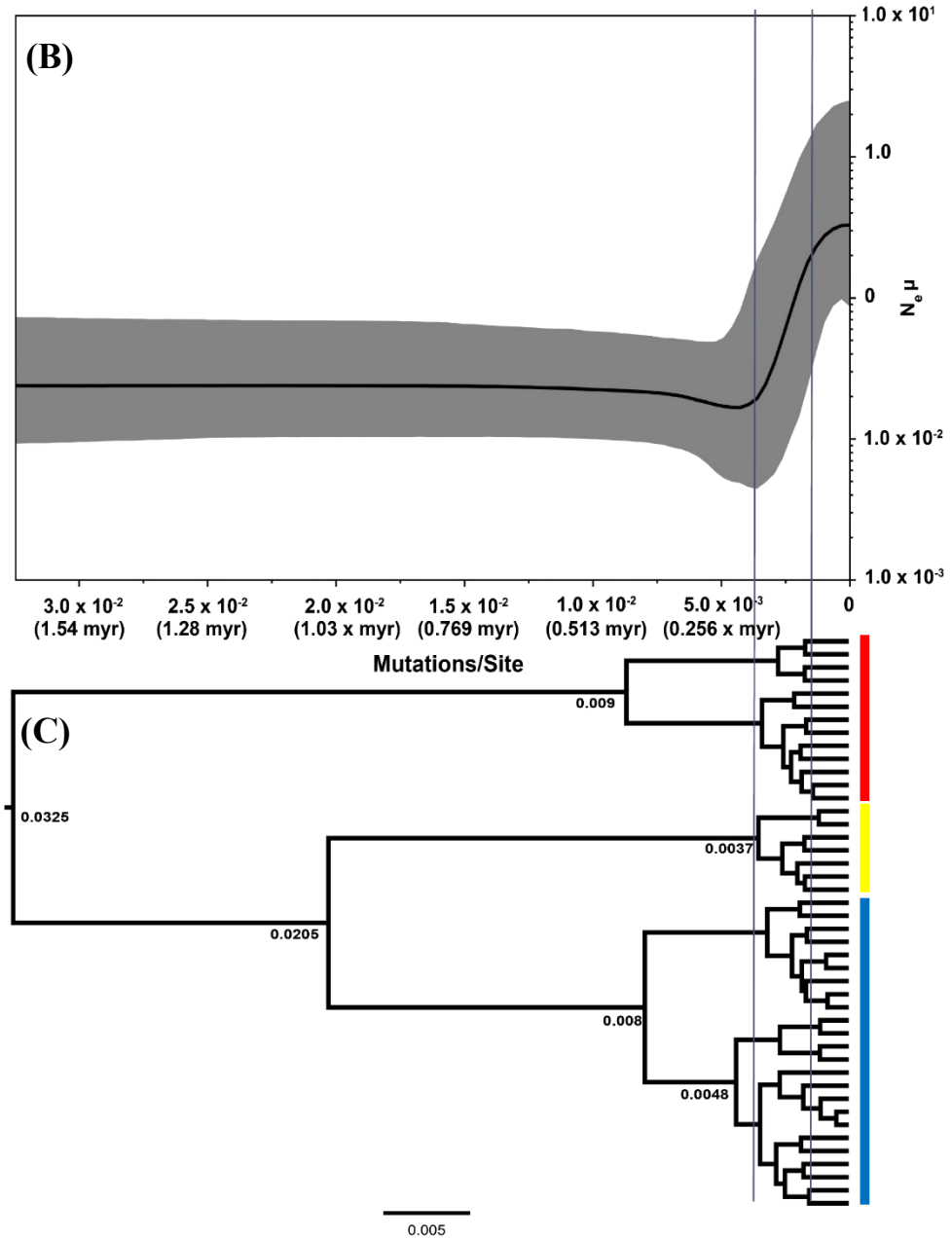


Figure 3B-C. Age estimate of effective population size for *Tiliqua rugosa* lineages. Colour schemes for lineages in as describe as Figure 2A. (B) Bayesian Skyline Plot (BSP) using a strict clock analysis applied in Beast. The black line shows the mean plot values and the shaded area display upper and lower confidence intervals for the mean estimates. The y-axis is the Ne × mutation rate/site, or/and if we assume that the mutation rates are constant, it can be a measure of relative Ne over time. The x-axis is in unit of mutation per site (millions of years before present shown in brackets). (C) Maximum credibility tree from BSB analysis.

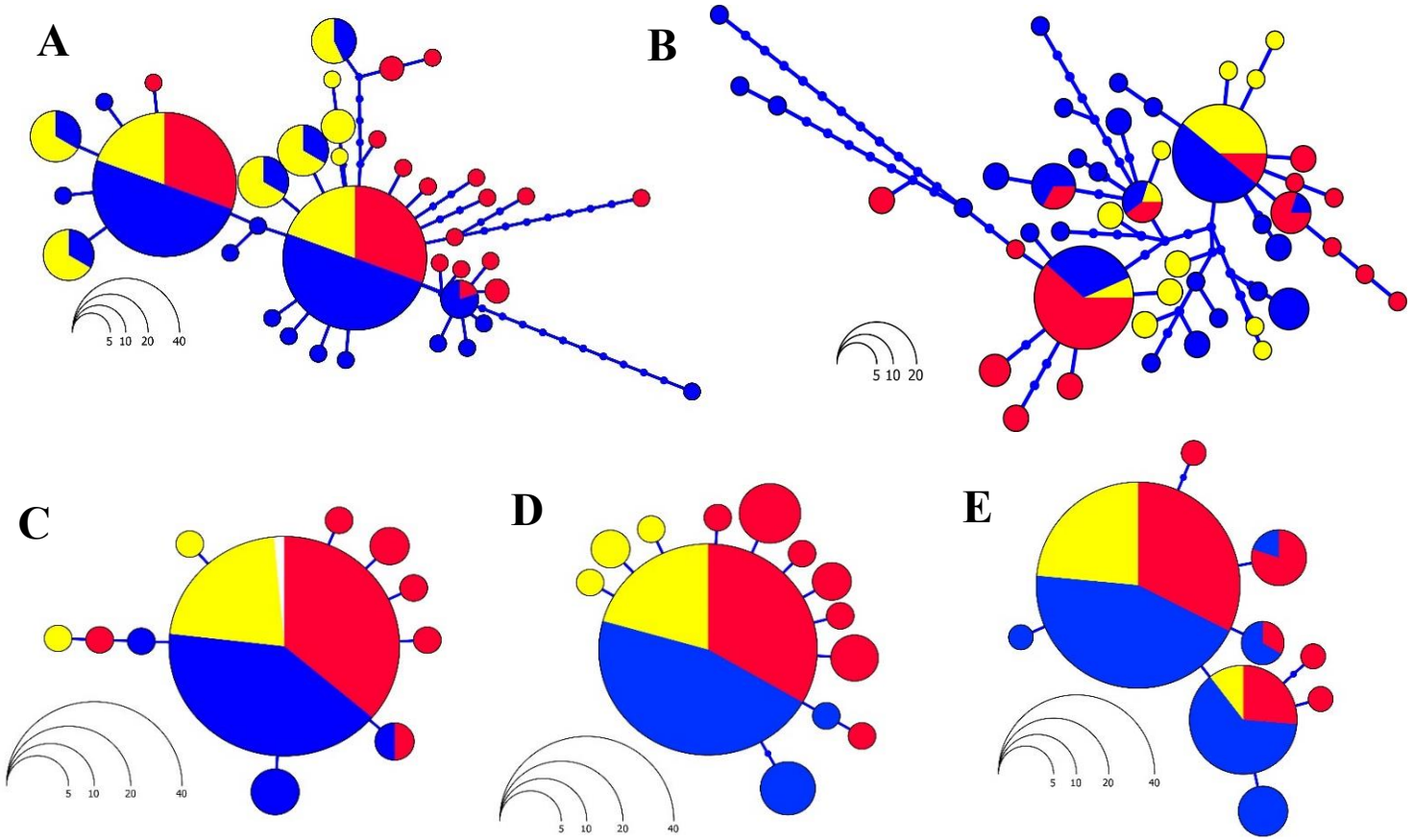


Figure 4A-E. Median- joining network of haplotypes for 11 nuclear markers in *T. rugosa*. Colours represent geographical locations of specimen haplotypes (Red; Western lineage, Blue; Northern lineage and Yellow: Southern lineage). A: *GAPD*, B: *β-fibronogen* intron 7, C: Tiru-9. D: Tiru-16, E: Tiru-23.

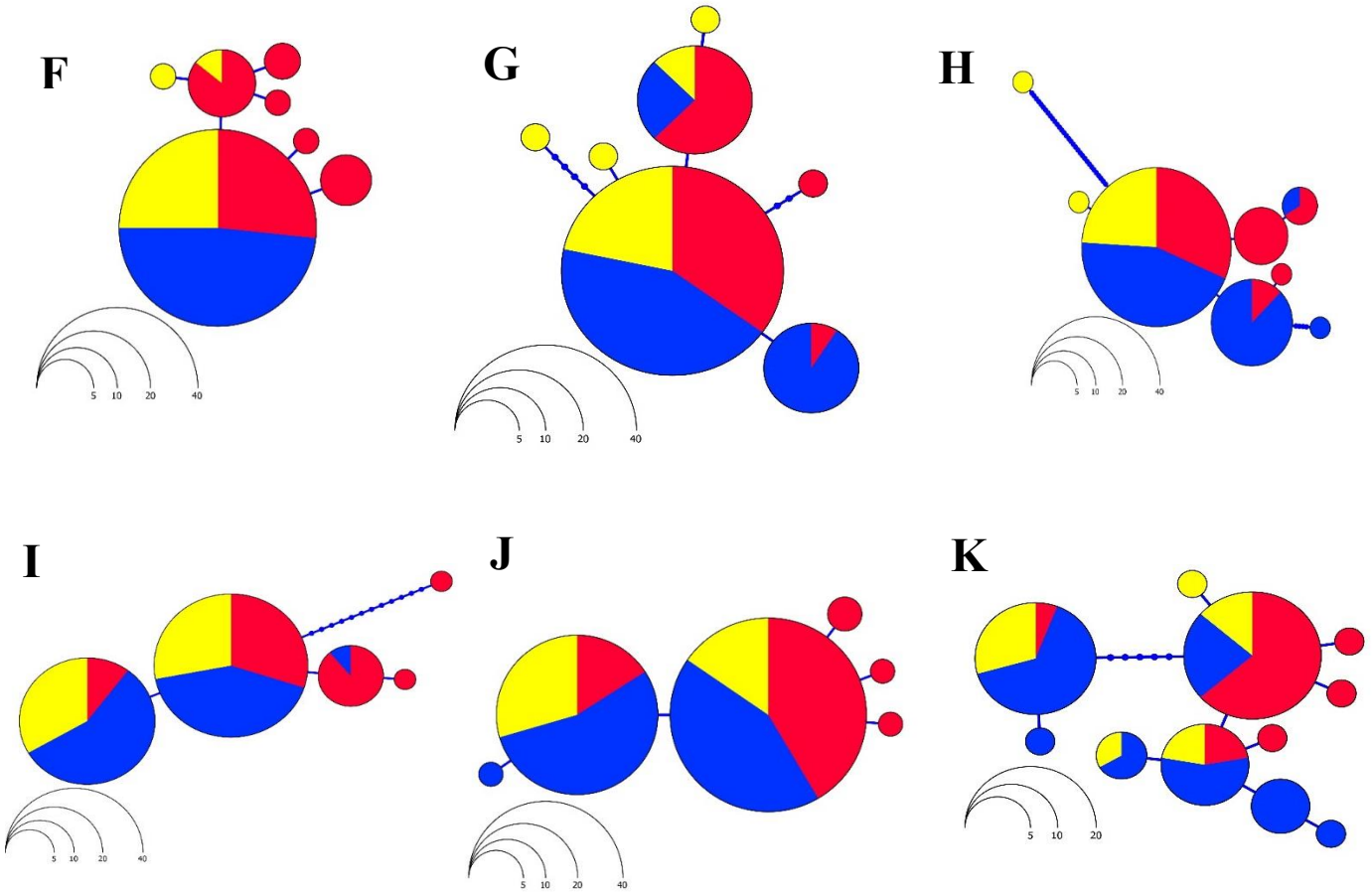


Figure 4F-L. Median- joining network of haplotypes for 11 nuclear markers in *T. rugosa*. Colours represent geographical locations of specimen haplotypes (Red; Western lineage, Blue; Northern lineage and Yellow: Southern lineage). F: Tiru-34, G: Tiru-39, H: Tiru-44, I: Tiru-45, J: Tiru-46, K: Tiru-48 and L: *ND4*. Circle size and pie slices within each circle are proportional to the number of individuals identified with that haplotype and the branch length is proportional to the number of mutations.

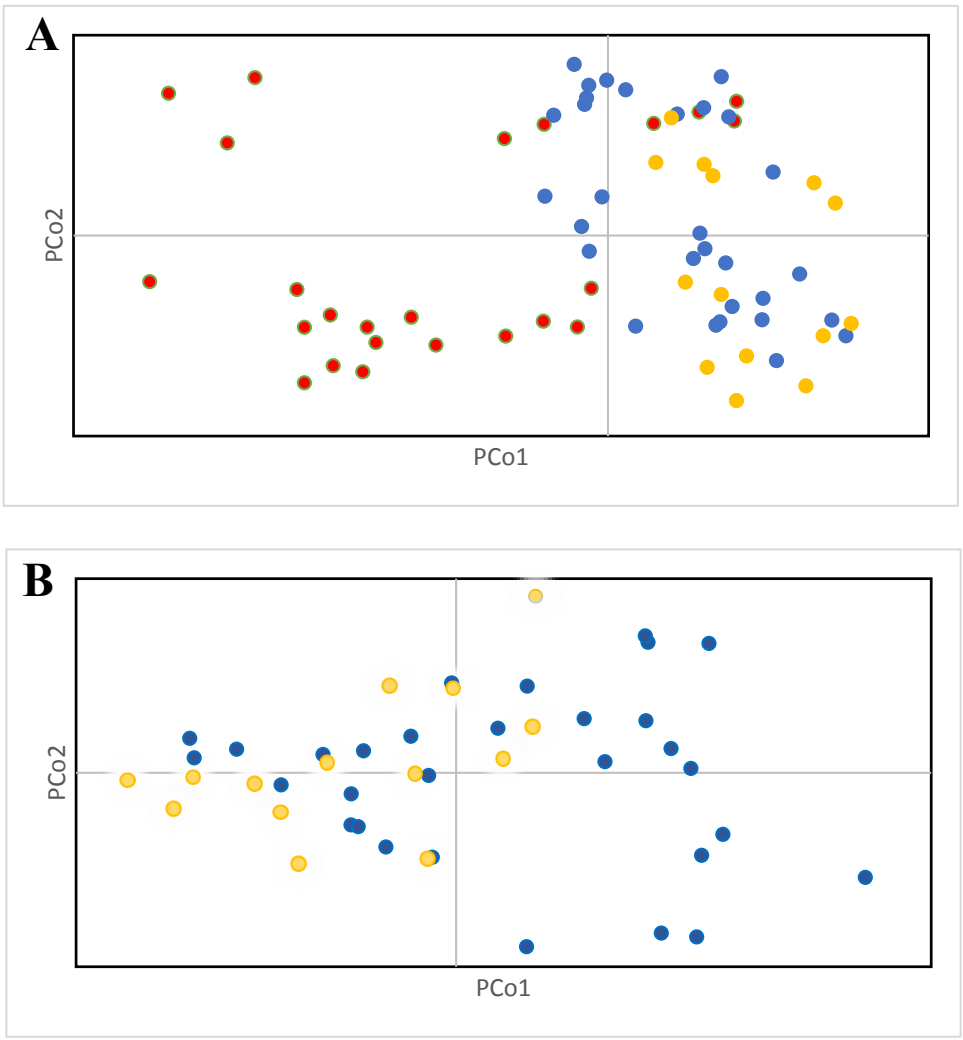


Figure 5A-B. Multilocus analysis of *Tiliqua rugosa* nuclear genetic distances. A: The ordination of individuals along the two principle coordinates (PCo1 and PCo2) based on the principle coordinate analysis for all three mtDNA lineages (Western (Red), Northern (Blue) and Southern (Yellow) clades. B: The principle coordinate analysis excluding the Western clade.

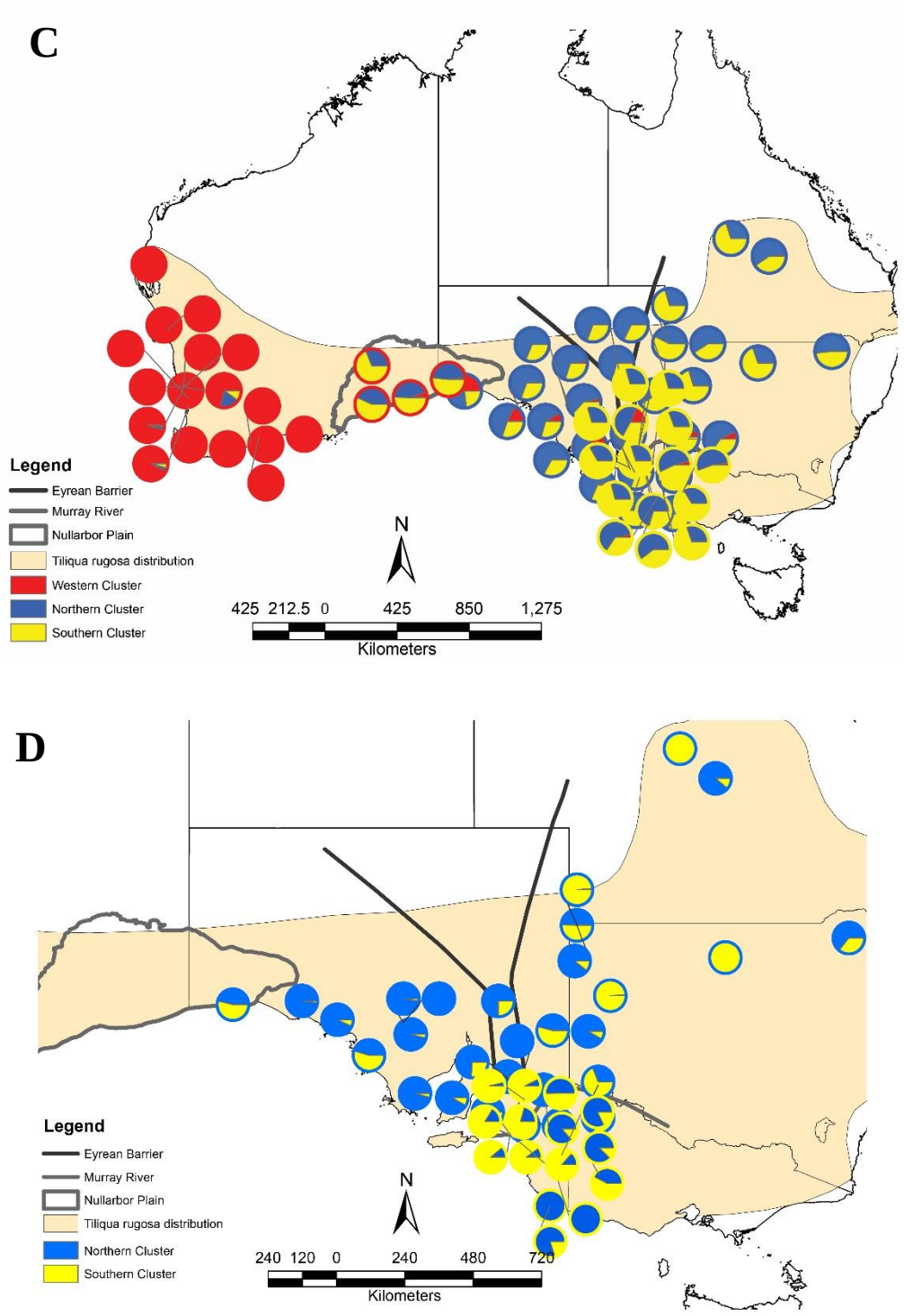


Figure 5C-D. Multilocus analysis of *Tiliqua rugosa* nuclear genetic distances. C and D represent the probabilities of belonging to the each of nuclear genetic clusters in estimated by STRUCTURE. C: The probability of membership to nuclear genetic clusters for all three genetic clusters for K=3. D: The probability of membership to nuclear genetic clusters excluding the Western population for K=2. The thick grey line represents the Murray River.

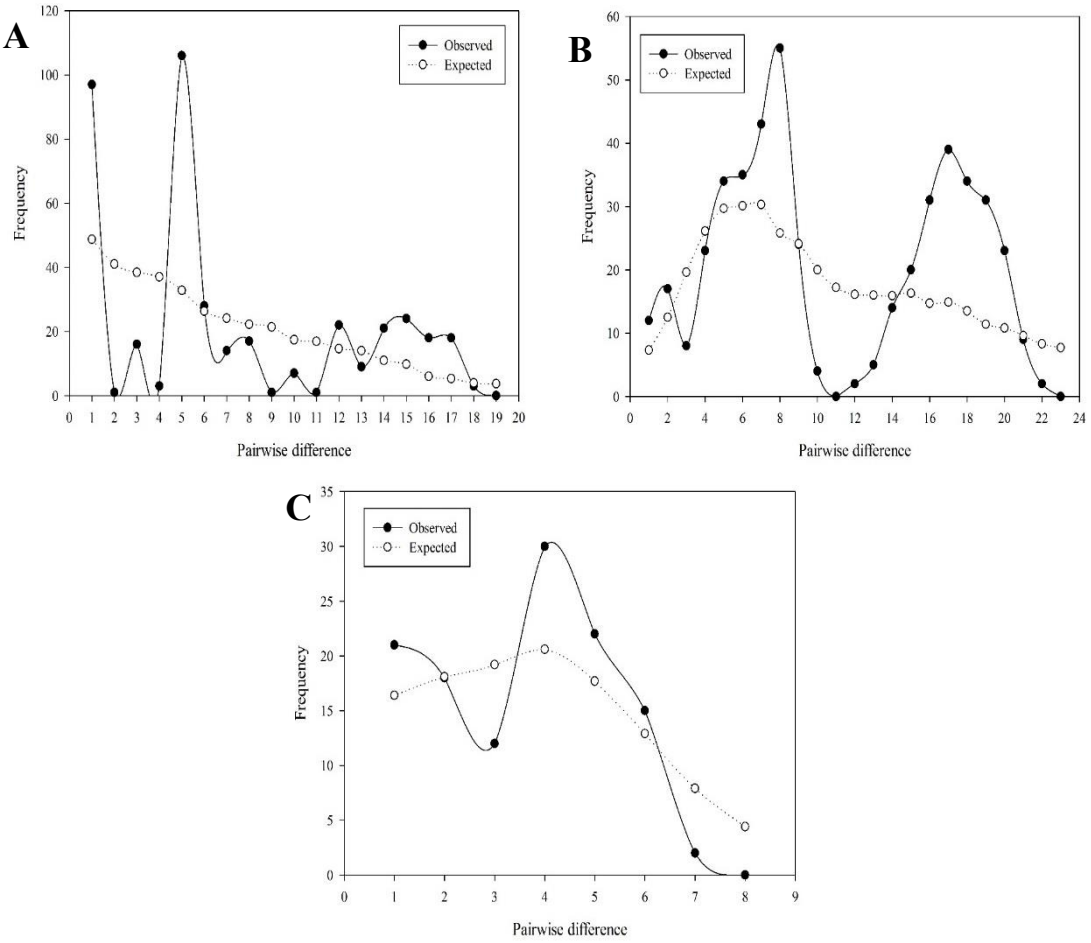


Figure 6. Observed (dotted line) and expected (solid line) mismatch distributions for a sudden expansion of Western (A), Northern (B) and Southern (C) populations of *Tiliqua rugosa*.

6.4.3. Distribution modelling for the T. rugosa

We observed a reasonably high performance of the model across the replications in a five-fold cross validation of AUC (0.824 mean AUC ± 0.02 Standard Error) and a lower power for random prediction of the model (AUC for random prediction = 0.5). The models strongly predicted the current distribution of suitable climate for *T. rugosa* with relatively good match to the known locations used in the model training (Figure 7A and B).

The predictions from the LGM climate models for *T. rugosa,* in general, indicated that there was a reduction of suitable habitat, mainly along the Western Australian coast (indicated by a reduction in the red areas; Figure 7). During the LGM, the coastal regions appeared to have remained suitable for *T. rugosa* across most regions of southern Australia, east of the Nullabor (Figure 7C and D). However, under the CCSM model, there were some breaks in suitable habitat along the Nullabor plain coastline, and east of the Gawler ranges (Eyrean barrier location). The estimated geographic locations for the refugia based on the continuous diffusion phylogeographic analysis (see above), mapped to localities with suitable climate for *T. rugosa* during the LGM (Figure 7C and D). Two refugia in Western Australia were each associated with isolated regions of suitable climate, and, similarly, a refugium mapping to the Eyre Peninsula was also associated with an isolated region of suitable climate during the LGM.

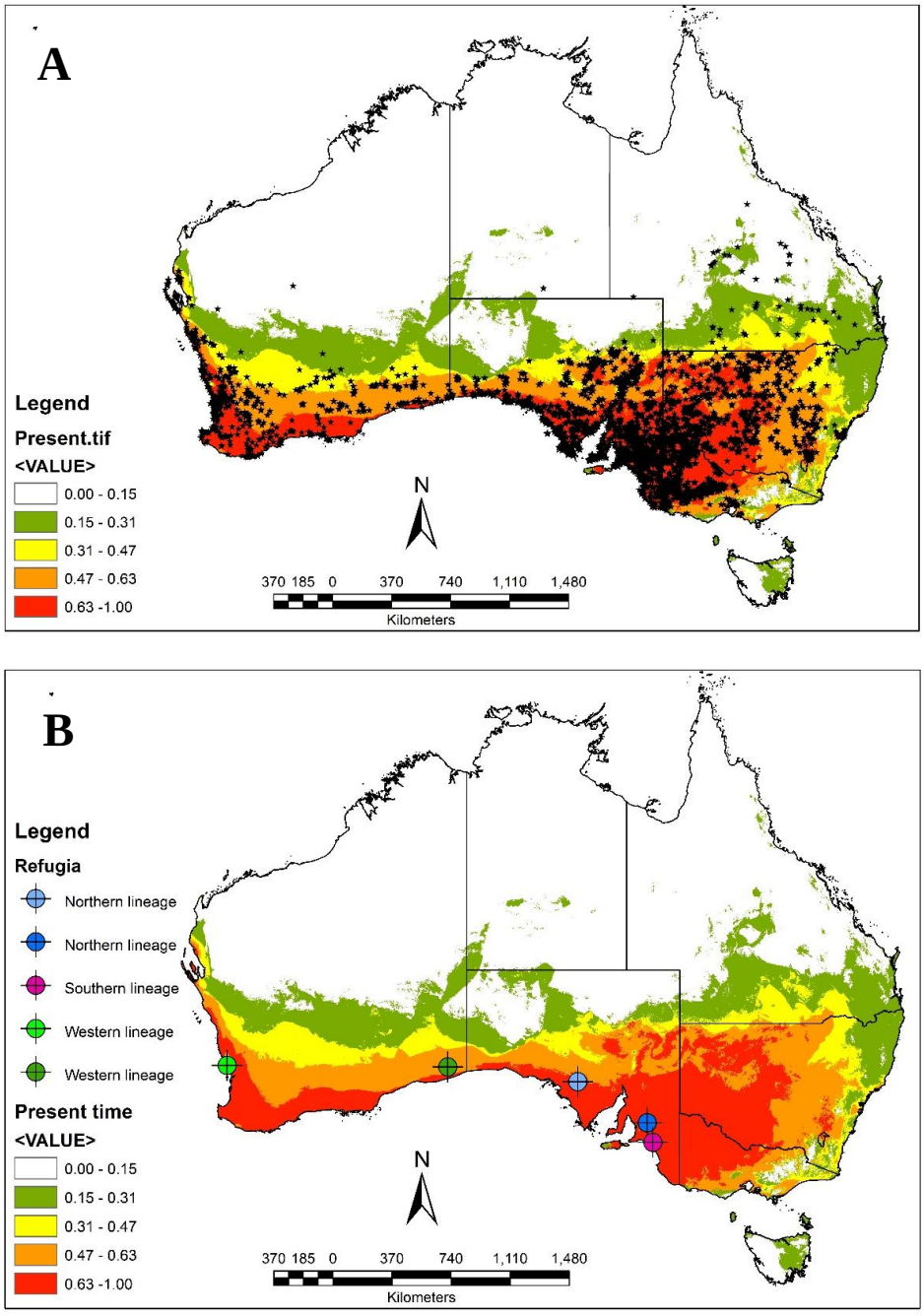


Figure 7A-B. Predicted current and Last Glacial Maximum (LGM) distributions of suitable climate for the Tiliqua rugosa. Different gridded colours in the maps shows the logistic format for outputs from Maxent, that darkest (red) shades had highest probability of occurrence (Probability>0.6). Figures A and B show the prediction of the occurrence of T. rugosa in the current climate. Little black stars (A) show the current locality records of T. rugosa that we used to generate the models.

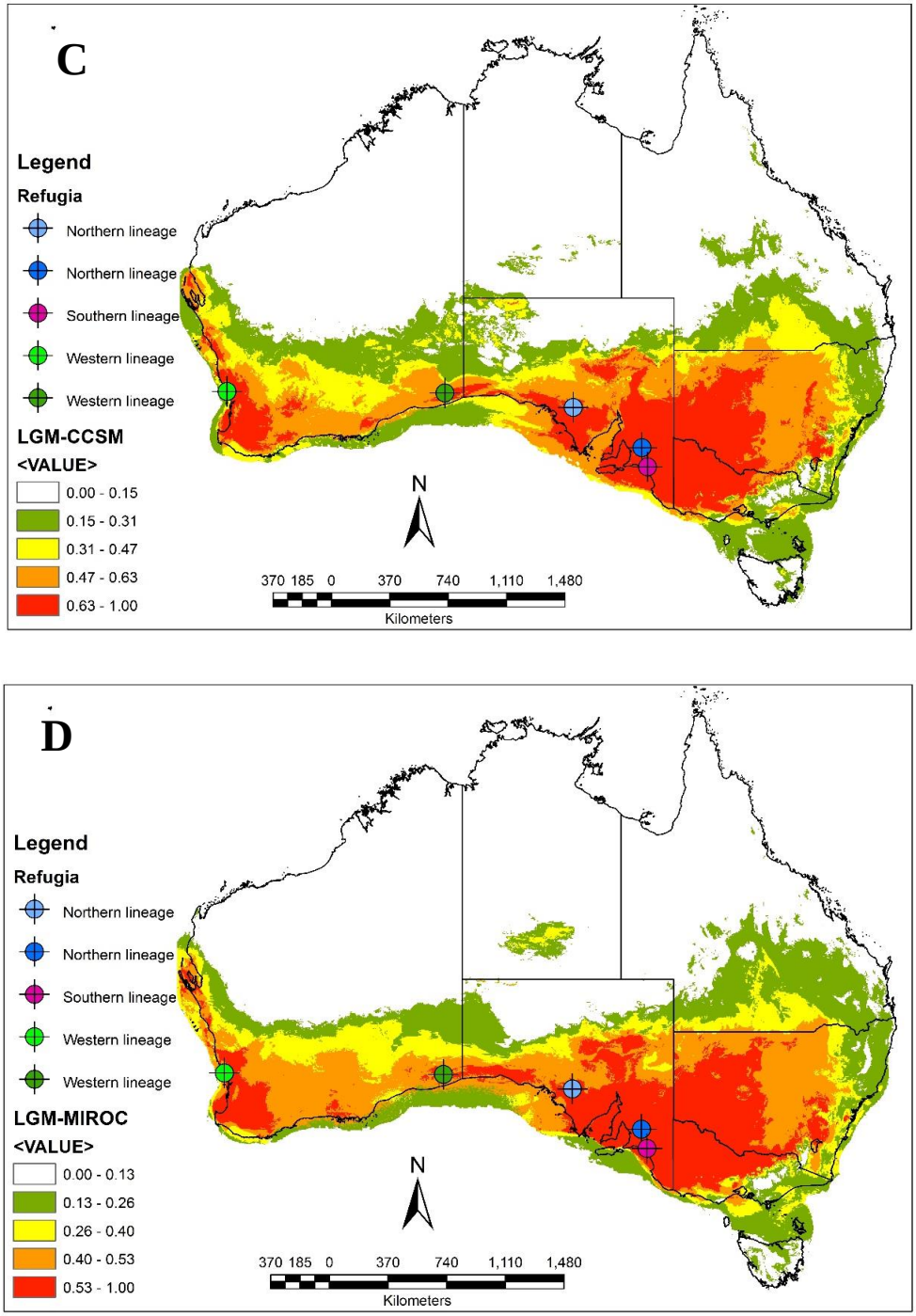


Figure 7C-D. Predicted current and Last Glacial Maximum (LGM) distributions of suitable climate for the *Tiliqua rugosa*. Different gridded colours in the maps shows the logistic format for outputs from Maxent, that darkest (red) shades had highest probability of occurrence (Probability>0.6). C and D represent the presence of suitable climate during the LGM according to CCSM and MIROC models, respectively. Big circles in B, C and D represent the estimated median of refugial location based on phylogeographic analysis in the current study as described in Figure 2D.

6.5. DISCUSSION

We found a pattern of strong mtDNA divergence with potential ongoing gene flow at nuclear DNA within the distribution of the common and widespread southern Australian squamate reptile, *T.rugosa*. Our phylogeographic analyses based on mtDNA suggests that *T. rugosa* comprises at least three genetically distinct lineages: A 'Western' clade, which occurs to the west of the Nullarbor Plain; a 'Northern' clade which is distributed throughout NSW, Queensland and north of the Murray River in South Australia to east of the Nullabor plain; and a 'Southern' clade which comprises individuals from Victoria and south-east of the Murray River in South Australia. This genetic structure was also supported by analyses of 11 nuclear loci, with strong differentiation between lineages west and east of the Nullabor plain (Figure 5, PCoA and STRUCTURE analyses). However, samples from the three mtDNA lineages shared haplotypes at all nuclear loci.

This pattern has several potential explanations: 1) male biased dispersal; 2) a selective sweep at mtDNA; 3) stochastic mtDNA lineage sorting; and 4) vicariance with barriers limiting gene flow. We deal with each of these scenarios in following discussions.

First, male biased dispersal could lead to this pattern. However, we believe this is unlikely as previous studies in this species have not found major differences in the dispersal between sexes (Godfrey *et al.* 2014). In fact the dispersal is potentially female biased if there is any difference. Another facet that minimises male biased dispersal as a factor producing the pattern we see is the presence of shared haplotypes for each of the nuclear loci (Figure 3) is

most likely due to the retention of ancestral haplotypes, rather than through ongoing gene flow among the lineages. This is because each shared haplotype was usually internal in the haplotype network and connected by one or two nucleotide substitutions to numerous private haplotypes. If gene flow was ongoing, shared haplotypes would not be internal.

The second explanation is that the pattern could be explained by a selective sweep at mtDNA such as has been discovered in eastern yellow robin *Eopsaltria australis* (Pavlova *et al.* 2015). Although the investigation of this scenario would involve further in depth analysis outside the scope of the study, we do note that a preliminary investigation of the presence of positive selection in the mtDNA sequences was non-significant. The lack of ongoing gene flow we discuss above also makes this scenario less likely. However, mtDNA is a single locus and we just testing a small part of that in current study, therefore further analysis would be required to be confident selection at mtDNA has not played a role.

Third, stochastic mtDNA lineage sorting following contraction to refugia could be responsible. Our study also provides evidence for the existence of long-term refugia in southern Australia. Predictions for the distribution of *T. rugosa* during the LGM based on both MIROC and CCSM models showed a reduction in the favourable climatic range for the species across southern Australia compared with its current distribution, particularly in coastal regions of Western Australia (Figure 7). The majority of the central east area of southern Australia, and the Adelaide Geosyncline (a geological province comprising the Mount Lofty and Flinders Ranges in South Australia) remained

favourable during the LGM (Figure 7). The Adelaide Geosyncline has been hypothesised to represent a major refugium for Australian arid zone taxa (Byrne, 2008; Byrne, et al., 2008; Guerin and Lowe, 2013), and our findings, based on the phylogeographic diffusion analysis and from greater mtDNA diversity in the Northern clade lend further support to this hypothesis.

The estimated geographic location for the MRCAs of the three populations based on phylogeographic diffusion analysis corresponded well with the modelling predictions for climatically suitable areas for *T. rugosa* during the LGM. On the other hand, the small sampling size might have affected the defined MRCAs locations, therefore the identified refugia locations in this study are preliminary. However, based on available data in current study, two estimated refugial locations were identified in WA, one located in the south-west and a second refugium in the eastern part of WA, just west of the Nullabor plain in a predicted climatically suitable area. Each of these refugia were associated with distinct and divergent mtDNA clades, suggesting that they are likely to have been refugia over multiple glacial cycles, long enough to lead to reciprocal monophyly of mtDNA haplotypes. In addition to the Mount Lofty Ranges, an estimated refugium was also identified near the Gawler Ranges on the Eyre Peninsula in SA. The Gawler Ranges have also been proposed to represent a likely refugium for other species (Byrne et al. 2008). However, considering the sampling number and location in current study, particularly regards to the Southern clade where we have a limited number of samples from Victoria, the specified refugia location still might be rejected and needs to be interpreted with caution.

In support of our scenario of past refugial isolation, our demographic analyses, based on mtDNA, (Table 3) suggested a relatively recent population expansion may have occurred in all lineages. The presence of several haplotypes that have very broad distributions in NSW and southern QLD, which are regions containing low levels of variation compared to populations in the south, further suggests that *T. rugosa* may have recently expanded its range into these regions from refugia in the south. In addition, a multimodal mismatch distribution for mtDNA shows an ancient splitting of haplotype clades at that locus (Reich, et al., 1999). This modal pattern in Western and Northern clades lends further support to the proposal that there were multiple refugial regions for each of these populations.

Coming back to the evidence for stochastic variation of mtDNA lineage sorting, the evidence from the pattern of nuclear variation in *T. rugosa* across the Nullabor plain indicated several genetically intermediate individuals that are from a region representing a potential contact zone between mtDNA lineages (Figure 6 and Figure 7). These individuals likely represent cases of nuclear introgression between the two lineages. The LGM model predictions suggest that the western population may have contracted to a refugium located along the coast west of the Nullabor plain and was isolated by climatically unsuitable habitats from populations to the east of the Nullabor plain. These predictions suggest that the patterns of introgression most likely resulted from secondary contact after the LGM (Figure 5). This potential reconnection left a clear signature in the nuclear genome of *T. rugosa* in the area of the contact zone, but further sampling is necessary to determine whether a hybrid zone between lineages is currently evident within the species.

The pattern of nuclear variation in *T. rugosa* across the Murray River also showed many individuals with genetically intermediate between two different nuclear groups. Although, there is evidence for a quite few private haplotypes in each population, the presence of ancestral haplotype at a high frequency in each population are more likely to indicate that there has been insufficient time for lineage sorting to lead to reciprocal monophyly of their nuclear haplotypes (Toews & Brelsford 2012). It is expected that mtDNA will complete the process of lineage sorting faster than nuDNA, because mitochondrial genome is haploid and uniparentally inherited in most animals, and, therefore, it has a fourfold smaller effective population size than nuclear genes (Palumbi, et al., 2001; Funk and Omland, 2003; Hudson and Turelli, 2003; Zink and Barrowclough, 2008). Clearly there has been a period of time where the lineages have been separated with enough time for mtDNA lineages to have sorted but where nuclear DNA may not have had time to become completely differentiated.

The last potential process that could lead to our observed patter is that it is due to vicariance with barriers limiting gene flow. This is not mutually exclusive to the previous explanation and both may have acted in concert.

The significant genetic differentiation we found in *T. rugosa* on either side of the Nullarbor Plain in Australia is consistent with this region representing a major historical environmental barrier to gene flow (Fujioka, et al., 2005; Toon, et al., 2007; Dolman and Joseph, 2012). The Nullarbor Plain occupies an area of about 200,000 square kilometres (Webb and James, 2006) and is fringed by xeric woodlands and shrublands (Beard, 1975) between

Australia's temperate mesic zones in the south west and south east. Our analyses suggest that the divergence among the western and eastern mtDNA lineages located on either side of the Nullarbor Plain dated to approximately the Plio-Pleistocene (95% HPD= 4.9 – 0.9 Ma). This time period coincides with the progression of aridification on the Australian continent, which started in the mid-Miocene and increased in intensity during the Plio-Pleistocene approximately 1-4 Ma (Bowler, 1976; Martin, 1978; Byrne, et al., 2008). A study of ten bird species with population structuring on either side of the Nullabor also estimated similar divergence times for populations of several species, although they also found evidence for later (Pleistocene; < 1 Ma) periods of diversification (Dolman and Joseph, 2012). An earlier time period of diversification, associated with the uplift of the Nullabor plain and the onset of aridity around 13-14 Ma, has been identified for 23 sister species of plants (Crisp and Cook, 2007), but this hypothesis is unlikely to explain the patterns identified within *T. rugosa*.

A second phylogeographical break in *T. rugosa* occurs at the Murray River which separates the northern and southern populations. Here too we found introgression of nuclear DNA with a fairly abrupt discordance of mtDNA. The Murray River is Australia's longest river, which flows to the northwest from the Snowy Mountains, and then turns to the south into South Australia, reaching the ocean at Lake Alexandrina. An estimation of the coalescence time of mtDNA haplotypes from the northern and southern populations suggests that population divergence likely occurred during the Pleistocene (95% HPD= 0.4 - 2.7 Ma). Studies on other species have suggested approximately similar timescales for divergences between lineages from the

north and south of the Murray River. For example, cladogensis in *Sminthopsis crassicaudata*, was estimated at sometimes during the mid-Pleistocene to late Pliocene, around 1.3-2.1 Ma (Cooper, et al., 2000). This time period coincides with a series of warming and cooling cycles (show fluctuation between glacial and interglacial climates) that resulted in changing habitat boundaries (Byrne, et al., 2008). At this time period sea levels fluctuated leading to isolation and then reconnection of island populations (Byrne, et al., 2008). However, the predicted LGM distributions from the modelling analyses suggest that the Murray basin region remained climatically suitable for *T. rugosa* during the LGM, further suggesting that the Northern and Southern populations had the potential for a long history of contact. In addition, the potential refugial location corresponding to the Southern lineage was suggested to be close to the Murray River. However, considering the sampling number and location in current study, particularly in regards to the Southern lineage where we have limited number of samples from Victoria, the specified refugia location needs to be interpreted with caution.

Cooper, et al. (2000) suggested that Lake Bungunnia, which was an ancient mega-lake occupying the western Murray basin about 3.5 – 0.7 Myr (Stephenson, 1986; Bowler, et al., 2006), was a possible long-term physical barrier to gene flow in *S. crassicaudata*. In addition, the Murray River itself has continued to actively flow during both glacial maxima and interglacial periods, with its outlet during the Last Glacial Maximum being in the Sprigg Canyon, on the edge of the continental shelf (Page, et al., 1996; Hill, et al., 2011), providing a further long-term barrier to gene flow. This Lake Bungunnia/ Murray River barrier hypothesis was further supported in a study

of the grasshopper *V. viatica* (Kawakami, et al., 2009) and it may have contributed to population isolation of additional ground-dwelling animals such as southern brown bandicoots (Li, et al., 2014). Our finding of a similar phylogeographical break in *T. rugosa* lends additional support to the hypothesis and further suggests that the barrier has had a broad impact on the phylogeographic structure of many ground-dwelling animal taxa distributed across southern Australia.

6.5.1. CONCLUSION

Our molecular phylogeographic study of the Australian arid /semi-arid zone skink, *T. rugosa*, provides evidence for significant population structuring associated with two major genetic barriers, the Murray River and Nullarbor Plain in southern Australia. Our data did not allow a complete test of the Eyrean barrier. The current study hints at the presence of multiple refugia for this species during previous glacial maxima based on both phylogeographic diffusion and species distribution models, with secondary contact most likely occurring in the Nullabor plain region following climate amelioration. These findings suggest that phylogeographic breaks are likely to be found in many other arid zone species with broad distributions across southern Australia.

6.6. ACKNOWLEDGMENTS

Funding was received from the Sir Mark Mitchell Research Foundation, and support from the South Australian Museum (SAM) staff for help with tissue sampling is gratefully acknowledged. The authors would like to thank Dr. Alison Fitch for help with lab work. The research was conducted according

to the guidelines of Flinders University Animal Welfare Committee (approval no. E324).

6.7. APPENDIX TABLE

Table S1. Primers sequences (5' – 3') employed in this study. For next generation sequencing we used dual multiplexing approach which MRT tag (linker) sequences are shown bold and italic. Illumina adaptors are italic. The position of the MID tags are shown with X and the location of locus specific primers (Ansari, et al., 2014) shown with ---.

Locus	**Primer name**	**Primer sequence (5' – 3')**
ND4	*ND4* (L) (Forstner & Arévalo,1995)	TGACTACCAAAAGCTCATGTAGAAGC
	M246 (H) (Gardner *et al*., 2008)	TTTTACTTGGATTTGCACCA
	M256 (H) (Saint & Smith, unpublished)	CCTTTTATRTTRATTGTKGGKGG
***β-fib* intron 7**	G375 (F) (Gardner *et al*., 2008)	GACAGAGACAATGATGGA
	G376 (R) (Gardner *et al*., 2008)	GTGAGGAATAATRCACAAAG
GAPD	*GAPD* L890 (Lyons *et al*., 1997)	ACCTTTAATGCGGGTGCTGGCATTGC
	GAPD H950 (Lyons *et al*., 1997)	CATCAAGTCCACAACACGGTTGCTGTA
Tag F (P5+MIDtag+Read1+MRT tagF)	Ansari *et al*. (Chapter 5)	*AATGATACGGCGACCACCGA*XXXXXXXX *ACACTCTTTCCCTACACGACGCTCTTCCG ATCT* ***ACGACGTTGTAAAA*---**
Tag R (P7+MIDtag+Read2+MRT tagR)	Ansari *et al*. (Chapter 5)	*CAAGCAGAAGACGGCATACGA*XXXXXXXX *XGTGACTGGAGTTCAGACGTGTGCTCTTC CGATCT* ***CATTAAGTTCCCATTA*---**

Chapter seven

7. Conclusion

In order to better understand historical evolutionary processes, it is constructive to assess the potential mechanisms that influence patterns of gene flow and their effects on population genetic structure. In this thesis, I investigated the potential evolutionary processes that led to the distribution and structure of genetic diversity across the arid and semi-arid zones of southern Australia. I characterised and developed a variety of molecular markers and used a combination of mitochondrial (mtDNA) and the developed neutral nuclear markers (nuDNA) in a phylogeographic approach, to assess the genetic structure of the Australian widespread skink, and understand the evolutionary process behind the pattern of its variation.

Intraspecific studies insight into genetic variation requires the use of molecular markers, of which mtDNA has been the marker of choice (Avise, et al., 1987; Wan, et al., 2004; Karachanak, et al., 2012; Brandt, et al., 2013; Kawamoto, et al., 2013). However, recent studies have indicated the importance of using multiple loci to obtain a comprehensive overview (Dolman and Moritz, 2006; Pavlova, et al., 2013). The limited resources on non-model species has previously restricted the availability of genetic markers for particular groups (Bertozzi, et al., 2012). It was for this reason I characterised different types of molecular markers using next generation sequencing (NGS) approaches and used them in order to assess contemporary and historical genetic variation of a common skink *Tiliqua rugosa*, widespread across the arid and semi-arid region of southern Australia.

First I developed 48 anonymous nuclear loci (ANL), for *T. rugosa* using 454 shotgun sequencing, among which 37 of them amplified in other related *Egernia* group species and potentially might be useful in the study inside into the architecture of genetic variation and population genetics in this group. Application of next generation sequencing in this study confirmed the higher success of using these approaches in molecular ecology as already stated by Ekblom and Galindo (2011). The characterisation and initial test of these ANL in *T. rugosa* and related species (*Tiliqua adelaidensis* and *Egernia stokesii*) shows evidence of variation within *T. rugosa* and among the three species. This result highlighted the practicality of these markers for phylogenetic, phylogeographic and population genetic analyses in *T. rugosa* and related species. In my thesis these markers provided me with the benefit to investigate

the genetic structure of populations and the historical forces that formed them across southern Australia.

Secondly, I utilized a transcriptome of *T. rugosa* to characterise MHC genes, both class I and class II. I then developed genetic markers for class I and class II independently and successfully amplified exon 2 to 4 in class I and exon 2 and 3 in MHC class II. Although the information contained in MHC genes provide insight into conservation and contributes to an understanding of the evolution of species and populations, limited number of studies have focused on the order Squamata (See, Miller, et al., 2005; Miller, et al., 2007a; Glaberman and Caccone, 2008; Glaberman, et al., 2009), and few for skinks (only a single other class I paper exists, Murphy, et al., 2009). Therefore the information provided here is excellent initial data for skinks, which will add to better understanding the evolution of adaptive gene in squamates and so help in conservation plans for members of this order of reptiles. The developed markers which successfully tested also in two other related *Egernia* species (*T. adelaidensis* and *E. stokesii*), support the potential benefit of using these markers to study other *Egernia* group species and potentially other skinks.

A rapidly growing body of work indicates the importance use of adaptive genes in population genetic (Hoffmann, et al., 2003; Sommer, 2005a; Nadachowska-Brzyska, et al., 2012; Kyle, et al., 2014). Therefore, I planned to use these newly developed markers from MHC class I and II to investigate the genetic variation structure across the landscape at different scales. I used these markers in an Illumina MiSeq run, which provided me with a large number of sequences. However, the amount of time that I had to spend developing the loci and then analysing the huge amount of data produced through the

sequencing platform meant I had little time to utilize the loci in this thesis. However analysis of these data to assess the level of variation, and historical or contemporary forces including potential differences across landscapes in selective forces from pathogens will be the subject of my future research.

Following my aim, using phylogeographic approaches to test the historical and contemporary structure of genetic variation, I used the application of NGS (Illumina MiSeq platform) to sequence my loci for large number of samples. The advantages and limitations of using next generation approached for phylogeographic studies had been discussed in the review by McCormack, et al. (2013). So far many studies focused on NGS approaches and developed exceptional methods and analysis to make it accessible to use for molecular ecologist and population geneticists (Binladen, et al., 2007; Farias-Hesson, et al., 2010; Meyer and Kircher, 2010; Bybee, et al., 2011; de Carcer, et al., 2011; Meglecz, et al., 2011; Clarke, et al., 2014). However, I developed a new application that made it possible to use NGS approaches to be more time and cost effective. Out of huge amount of data set I obtained by using my new developed approaches, I used and analysed sequences from nine ANL in my phylogeographic assessment, but it indicated that there is potential benefit of using such approaches for comprehensive population genetic studies that require multiple loci.

Other techniques exist that enable the simultaneous development and genotyping of loci which could have been an alternative source of markers. These approaches are complexity reduction techniques which aim to produce a set of homologous markers across individuals. For example, restriction-site associated DNA sequencing (RAD-Seq) and related approaches, such as

genotype-by-sequencing (GBS) and double digest RAD-Seq (ddRAD), use one or more restriction enzymes to target homologous loci across a set of samples and no prior genomic resources is required (Miller, et al., 2007b), or (Davey and Blaxter, 2010; Narum, et al., 2013; Heffelfinger, et al., 2014).

A lack of initial funds prohibited my using these techniques. My approach allowed me to obtain data and continue development and utilize these as funding became available.

Another highly useful approach would have been to use a gene capture technique (see, Lemmon and Lemmon, 2013). This technology allows targeting, enriching for, and sequencing large number of genes and often their associated introns (Bi, et al., 2012; Hancock-Hanser, et al., 2013). The gene capture method includes some restrictions in comparison within and among species with narrow divergence or/and comparisons among highly divergent sequences and taxa (Li, et al., 2013), however, recently developed approaches which modify the power of this method have overcome the limitation (Li, et al., 2013).

The development costs and time associated with techniques of this sort precluded their use in my thesis but they could provide a way to screen for variation in the future in this group, perhaps utilizing the markers I developed and other markers that could be gleaned from the transcriptome.

My thesis culminated in a test of the hypothesis that the pattern of genetic variation and existing population genetic structuring was due to glacial isolation and the presence of physical barriers over the southern Australia. At a large geographic scale, the determination of fundamental population

structure was accomplished through Bayesian approaches, cluster analysis and spatial principal components analysis (PCoA).

Overall, I detected evidence for significant population structuring in *T. rugosa* across the distribution of the species in arid /semi-arid zone of Australia which was coincident with physical barriers, the Murray River and Nullarbor Plain. The Murray River and Nullarbor barrier have been previously reported as factors in shaping genetic variation (e.g. Cooper, et al., 2000; Kawakami, et al., 2009; Dolman and Joseph, 2012; Dolman and Joseph, 2015), and my finding supports the influence of the barrier on the genetic structure of species across southern Australia.

Multiple refugial locations were also detected for *T. rugosa* across the region, using species distribution modelling and a phylogeographic diffusion model (Lemey, et al., 2010). My findings supported the hypothesis of the presence of arid zone refugia in Australian upland regions, Gawler-Flinders and Mt Lofty Ranges (Byrne, 2008; Byrne, et al., 2008; Guerin and Lowe, 2013). My studies indicate that these glacial refugia were followed by secondary contact when climate emolliated, are associated with three distinct mtDNA lineages and the presence of a high frequency of nuclear ancestral haplotypes in each lineage. Another potential factor I tested for was the possibility of selection on mtDNA, as proposed by Pavlova, et al. (2013), that might explain the strong mtDNA structuring but not evident in the nuclear loci. Although I did not detect any signature of positive selection in the mtDNA sequences there was amino acid differences among the three mtDNA lineages and this intriguing possibility warrants further exploration.

There has been a call for studies to assess the presence of possible common genetic patterns, and contractions to refugia or the presence of biogeographical barriers, across the arid/semi-arid zones of southern Australia (Byrne, 2008). Therefore, my study provides a general overview of genetic structure related to contractions to refugia or/and the presence of biogeographical barriers across the arid/semi-arid zones of southern Australia. My findings suggest that phylogeographic breaks across the barriers are likely to be found in many other arid zone species with broad distributions across southern Australia.

7.1. Future research

This research provides an understanding on how the distribution of genetic variation affected by glacial climate changes as well as biogeographical barriers across the southern Australia.

In this study, a strong population structure was detected related to the mtDNA, and potential secondary contact between distinct lineages based on nuclear markers at a large special scale. As stated above although no significant selection was detected from mtDNA in this current study, the possibility still exists as we just tested a small proportion of mtDNA gene. Therefore using next generation sequencing approaches, such as shotgun sequencing, might help to recover a greater proportion of the mtDNA or even obtain whole mtDNA genomes in order to have a comprehensive overview of selection on mtDNA. Additional selection tests such as a G test or Fisher's exact test (Sokal and Rohlf, 1981), or testing the mitochondrial DNA coding for transmembrane

proteins, which is expected to be more affected than regions coding for surface proteins, are alternative methods that could be investigated.

Future research would benefit from studying the intraspecific genetic variations of population in a fine special scale which are needed to better understand the effect of geographic barrier, such as Murray River, on population genetic structure of non-model species. As different type of markers are now available, these could be used to test the effect that these barriers might influence the genetic diversity throughout different genomic regions. In particular, assaying variation at MHC loci would be beneficial to test different level of diversity that might maintained by selection. In addition, genetic variation within and among population could be further tested using MHC loci to assess the effect of parasites and pathogens in shaping such variation across different ecological environments or either side of the tick parapatric boundary (Andrews, et al., 1982) under intensive study at Flinders University.

8. REFERENCES

Alp, M., Keller, I., Westram, A. M. and Robinson, C. T. (2012). How river structure and biological traits influence gene flow: a population genetic study of two stream invertebrates with differing dispersal abilities. Freshwater Biology 57:969-981.

Andrews, R. H., Petney, T. N. and Bull, C. M. (1982). Reproductive interference between three parapatric species of reptile tick. Oecologia 52:281-286.

Ansari, T. H., Bertozzi, T., Hacking, J., Cooper, S. J. B. and Gardner, M. G. (2014). Random non-coding fragments of lizard DNA: anonymous nuclear loci for the Australian skink, Tiliqua rugosa, and their utility in other Egernia-group species. Australian journal of zoology 62:515-518.

Ansari, T. H., Bertozzi, T., Miller, R. D. and Gardner, M. G. (2015). MHC in a monogamous lizard - characterization of Class I MHC genes in the Australian skink (Tiliqua rugosa). Developmental & Comparative Immunology http://dx.doi.org/10.1016/j.dci.2015.07.012.

Auffray, C. and Novotny, J. (1986). Speculations on Sequence Homologies between the fibronectin cell-attachment site, major histocompatibility antigens, and a putative AIDS virus polypeptide. Human Immunology 15:381-390.

Avise, J. C. (2000). Phylogeography: the history and formation of species:1 - 447.

Avise, J. C. (2003). Catadromous Eels of the North Atlantic: A Review of Molecular Genetic Findings Relevant to Natural History, Population Structure, Speciation, and Phylogeny, in: K. Aida, et al. (Eds.), Eel Biology, Springer Japan. pp. 31-48.

Avise, J. C. (2009). Phylogeography: retrospect and prospect. Journal of biogeography 36:3-15.

Avise, J. C., Arnold, J., Ball, R. M., Bermingham, E., Lamb, T., Neigel, J. E., Reeb, C. A. and Saunders, N. C. (1987). Intraspecific phylogeography - the Mitochondrial-DNA Bridge between Population-Genetics and Systematics. Annual Review of Ecology and Systematics 18:489-522.

Balakrishnan, C. N., Lee, J. Y. and Edwards, S. V. (2010). Phylogeography and phylogenetics in the nuclear age. , in: G. B. Grant PR (Ed.), In Search of the Causes of Evolution: From Field Observations to Mechanisms, Princeton University Press, Priceton. pp. 65–88.

Ballard, J. W. O. and Whitlock, M. C. (2004). The incomplete natural history of mitochondria. Mol Ecol 13:729 - 744.

Bazin, E., Glémin, S. and Galtier, N. (2006). Population Size Does Not Influence Mitochondrial Genetic Diversity in Animals. Science 312:570-572.

Beard, J. S. (1975). The Vegetation of the Nullarbor Area. , Vegetation Survey of Western Australia,, University of Western Australia

Belov, K., Lam, M. K., Hellman, L. and Colgan, D. J. (2003). Evolution of the major histocompatibility complex: Isolation of class II β cDNAs from two monotremes, the platypus and the short-beaked echidna. Immunogenetics 55:402-411.

Benacerraf, B. and Germain, R. N. (1978). The Immune Response Genes of the Major Histocompatibility Complex. Immunological Reviews 38:70-119.

Berglund, E. C., Kiialainen, A. and Syvanen, A. C. (2011). Next-generation sequencing technologies and applications for human genetic history and forensics. Investig Genet 2:23.

Berthier, K., Galan, M., FoltÊTe, J. C., Charbonnel, N. and Cosson, J. F. (2005). Genetic structure of the cyclic fossorial water vole (Arvicola terrestris): landscape and demographic influences. Molecular ecology 14:2861-2871.

Bertozzi, T., Sanders, K. L., Sistrom, M. J. and Gardner, M. G. (2012). Anonymous nuclear loci in non-model organisms: making the most of high-throughput genome surveys. Bioinformatics 28:1807-1810.

Bhagwat, S. A. and Willis, K. J. (2008). Species persistence in northerly glacial refugia of Europe: a matter of chance or biogeographical traits? Journal of biogeography 35:464-482.

Bi, K., Vanderpool, D., Singhal, S., Linderoth, T., Moritz, C. and Good, J. M. (2012). Transcriptome-based exon capture enables highly cost-effective comparative genomic data collection at moderate evolutionary scales. BMC Genomics 13:403.

Bichet, C., Moodley, Y., Penn, D. J., Sorci, G. and Garnier, S. (2015). Genetic structure in insular and mainland populations of house sparrows (Passer domesticus) and their hemosporidian parasites. Ecol Evol 5:1639-1652.

Bielejec, F., Rambaut, A., Suchard, M. A. and Lemey, P. (2011). SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. Bioinformatics 27:2910-2912.

Binladen, J., Gilbert, M. T. P., Bollback, J. P., Panitz, F., Bendixen, C., Nielsen, R. and Willerslev, E. (2007). The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. PLoS One 2:e197.

Bjorkman, P. J. and Parham, P. (1990). Structure, function, and diversity of class I major histocompatibility complex molecules. Annu Rev Biochem 59:253-88.

Bjorkman, P. J., Saper, M. A., Samraoui, B., Bennett, W. S., Strominger, J. L. and Wiley, D. C. (1987). The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. Nature 329:512-518.

Boehm, T. and Zufall, F. (2006). MHC peptides and the sensory evaluation of genotype. Trends in Neurosciences 29:100-107.

Bouckaert, R., Heled, J., Kuhnert, D., Vaughan, T., Wu, C. H., Xie, D., Suchard, M. A., Rambaut, A. and Drummond, A. J. (2014). BEAST 2: A

Software Platform for Bayesian Evolutionary Analysis. PLoS Comput Biol 10:e1003537.

Bowler, J. M. (1976). Aridity in Australia: Age, origins and expression in aeolian landforms and sediments. Earth-Science Reviews 12:279-310.

Bowler, J. M. (1982). Aridity in the late Tertiary and Quaternary of Australia, in: W. R. Barker and P. J. M. Greenslade (Eds.), Evolution of the flora and fauna of arid Australia, Peacock, Adelaide. pp. 35-44.

Bowler, J. M., Kotsonis, A. and Lawrence, C. R. (2006). Environmental evolution of the Mallee region, western Murray Basin. Proceedings of the Royal Society of Victoria 118:161-210.

Brandt, G., Haak, W., Adler, C. J., Roth, C., Szecsenyi-Nagy, A., Karimnia, S., Moller-Rieker, S., Meller, H., Ganslmeier, R., Friederich, S., Dresely, V., Nicklisch, N., Pickrell, J. K., Sirocko, F., Reich, D., Cooper, A. and Alt, K. W. (2013). Ancient DNA reveals key stages in the formation of central European mitochondrial genetic diversity. Science 342:257-261.

Braud, V. M., Allan, D. S. and McMichael, A. J. (1999). Functions of nonclassical MHC and non-MHC-encoded class I molecules. Curr Opin Immunol 11:100-8.

Briggs, A. W., Good, J. M., Green, R. E., Krause, J. and Maricic, T. (2009). Targeted retrieval and analysis of five Neandertal mtDNA genomes. Science 325:318.

Brito, P. H. and Edwards, S. V. (2009). Multilocus phylogeography and phylogenetics using sequence-based markers. Genetica 135:439-455.

Brown, J. H., Jardetzky, T. S., Gorga, J. C., Stern, L. J., Urban, R. G., Strominger, J. L. and Wiley, D. C. (1993). Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. Nature 364:33-39.

Brown, J. L. and Eklund, A. (1994). Kin Recognition and the Major Histocompatibility Complex: An Integrative Review. The American Naturalist 143:435-461.

Brown, R. P. and Pestano, J. (1998). Phylogeography of skinks (Chalcides) in the Canary Islands inferred from mitochondrial DNA sequences. Molecular ecology 7:1183-1191.

Brown, R. P., Terrasa, B., Pérez-Mellado, V., Castro, J. A., Hoskisson, P. A., Picornell, A. and Ramon, M. M. (2008). Bayesian estimation of post-Messinian divergence times in Balearic Island lizards. Molecular phylogenetics and evolution 48:350-358.

Bull, C. M. (1978). Dispersal of the Australian Reptile Tick Aponomma Hydrosauri by Host Movement. Australian journal of zoology 26:305-322.

Bull, C. M. (1987). A Population Study of the Viviparous Australian Lizard, Trachydosaurus-Rugosus (Scincidae). Copeia 3:749-757.

Bull, C. M. (1994). Population dynamics and pair fidelity in sleepy lizards. Lizard ecology: historical and experimental perspectives. Princeton University Press, Princeton, NJ:159-174.

Bull, C. M. (1995). Population ecology of the sleepy lizard, Tiliqua rugosa, at Mt Mary, South Australia. Australian Journal of Ecology 20:393-402.

Bull, C. M., Griffin, C. L. and Johnston, G. R. (1999). Olfactory discrimination in scat-piling lizards. Behavioral Ecology 10:136-140.

Bull, C. M. and Pamula, Y. (1996). Sexually dimorphic head sizes and reproductive success in the sleepy lizard Tiliqua rugosa. Journal of Zoology 240:511-521.

Bull, C. M. and Pamula, Y. (1998). Enhanced vigilance in monogamous pairs of the lizard, Tiliqua rugosa. Behavioral Ecology 9:452-455.

Burri, R., Hirzel, H. N., Salamin, N., Roulin, A. and Fumagalli, L. (2008). Evolutionary Patterns of MHC Class II B in Owls and Their Implications for the Understanding of Avian MHC Evolution. Molecular Biology and Evolution 25:1180-1191.

Butlin, R. K., Saura, M., Charrier, G., Jackson, B., Andre, C., Caballero, A., Coyne, J. A., Galindo, J., Grahame, J. W., Hollander, J., Kemppainen, P., Martinez-Fernandez, M., Panova, M., Quesada, H., Johannesson, K. and Rolan-Alvarez, E. (2014). Parallel evolution of local adaptation and reproductive isolation in the face of gene flow. Evolution 68:935-49.

Bybee, S. M., Bracken-Grissom, H., Haynes, B. D., Hermansen, R. A., Byers, R. L., Clement, M. J., Udall, J. A., Wilcox, E. R. and Crandall, K. A. (2011). Targeted Amplicon Sequencing (TAS): A Scalable Next-Gen Approach to Multilocus, Multitaxa Phylogenetics. Genome Biology and Evolution 3:1312-1323.

Byrne, M. (2008). Evidence for multiple refugia at different time scales during Pleistocene climatic oscillations in southern Australia inferred from phylogeography. Quaternary Science Reviews 27:2576-2585.

Byrne, M. and Hopper, S. D. (2008). Granite outcrops as ancient islands in old landscapes: evidence from the phylogeography and population genetics of

Eucalyptus caesia (Myrtaceae) in Western Australia. Biological journal of the Linnean Society 93:177-188.

Byrne, M., Steane, D. A., Joseph, L., Yeates, D. K., Jordan, G. J., Crayn, D., Aplin, K., Cantrill, D. J., Cook, L. G., Crisp, M. D., Keogh, J. S., Melville, J., Moritz, C., Porch, N., Sniderman, J. M. K., Sunnucks, P. and Weston, P. H. (2011). Decline of a biome: evolution, contraction, fragmentation, extinction and invasion of the Australian mesic zone biota. Journal of biogeography 38:1635-1656.

Byrne, M., Yeates, D. K., Joseph, L., Kearney, M., Bowler, J., Williams, M. A., Cooper, S., Donnellan, S. C., Keogh, J. S., Leys, R., Melville, J., Murphy, D. J., Porch, N. and Wyrwoll, K. H. (2008). Birth of a biome: insights into the assembly and maintenance of the Australian arid zone biota. Mol Ecol 17:4398-417.

Cammarota, G., Scheirle, A., Takacs, B., Doran, D. M., Knorr, R., Bannwarth, W., Guardiola, J. and Sinigaglia, F. (1992). Identification of a CD4 binding site on the [beta]2 domain of HLA-DR molecules. Nature 356:799-801.

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S. M., Betley, J., Fraser, L., Bauer, M., Gormley, N., Gilbert, J. A., Smith, G. and Knight, R. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. The ISME journal 6:1621-1624.

Carling, M. D. and Brumfield, R. T. (2007). Gene Sampling Strategies for Multi-Locus Population Estimates of Genetic Diversity ☙ PLoS One 2:e160.

Carstens, B. C., Stevenson, A. L., Degenhardt, J. D. and Sullivan, J. (2004). Testing Nested Phylogenetic and Phylogeographic Hypotheses in the Plethodon vandykei Species Group. Systematic biology 53:781-792.

Chapple, D. G., Chapple, S. N. J. and Thompson, M. B. (2011). Biogeographic barriers in south-eastern Australia drive phylogeographic divergence in the garden skink, Lampropholis guichenoti. Journal of biogeography 38:1761-1775.

Chapple, D. G., Keogh, J. S. and Hutchinson, M. N. (2005). Substantial genetic substructuring in southeastern and alpine Australia revealed by molecular phylogeography of the Egernia whitii (Lacertilia: Scincidae) species group. Mol Ecol 14:1279-92.

Chevreux, B., Wetter, T. and Suhai, S. (1999). Genome sequence assembly using trace signals and additional sequence information, German conference on bioinformatics. pp. 45-56.

Clarke, L. J., Czechowski, P., Soubrier, J., Stevens, M. I. and Cooper, A. (2014). Modular tagging of amplicons using a single PCR for high-throughput sequencing. Mol Ecol Resour 14:117-121.

Coghlan, B. A., Goldizen, A. W., Thomson, V. A. and Seddon, J. M. (2015). Phylogeography of Eastern Grey Kangaroos, Macropus giganteus, Suggests a Mesic Refugium in Eastern Australia. PLoS One 10:e0128160.

Cohen, S. (2002). Strong positive selection and habitat-specific amino acid substitution patterns in MHC from an estuarine fish under intense pollution stress. Molecular Biology and Evolution 19:1870-1880.

Cooper, S. J. B., Adams, M. and Labrinidis, A. (2000). Phylogeography of the Australian dunnart Sminthopsis crassicaudata (Marsupialia : Dasyuridae). Australian journal of zoology 48:461-473.

Cooper, S. J. B., Ibrahim, K. M. and Hewitt, G. M. (1995). Postglacial expansion and genome subdivision in the European grasshopper Chorthippus parallelus. Molecular ecology 4:49-60.

Cosson, P. and Bonifacino, J. S. (1992). Role of transmembrane domain interactions in the assembly of class II MHC molecules. Science 258:659-62.

Cranmer, L., McCollin, D. and Ollerton, J. (2012). Landscape structure influences pollinator movements and directly affects plant reproductive success. Oikos 121:562-568.

Crisp, M. D. and Cook, L. G. (2007). A congruent molecular signature of vicariance across multiple plant lineages. Molecular phylogenetics and evolution 43:1106-1117.

Darriba, D., Taboada, G. L., Doallo, R. and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. Nature methods 9:772-772.

Davey, J. W. and Blaxter, M. L. (2010). RADSeq: next-generation population genetics. Briefings in Functional Genomics 9:416-423.

de Carcer, D. A., Denman, S. E., McSweeney, C. and Morrison, M. (2011). Strategy for modular tagged high-throughput amplicon sequencing. Applied and environmental microbiology 77:6310-6312.

Den Dunnen, J. T. and Antonarakis, S. E. (2000). Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. Human mutation 15:7-12.

Dolman, G. and Joseph, L. (2012). A species assemblage approach to comparative phylogeography of birds in southern Australia. Ecol Evol 2:354-69.

Dolman, G. and Joseph, L. (2015). Evolutionary history of birds across southern Australia: structure, history and taxonomic implications of mitochondrial DNA diversity in an ecologically diverse suite of species. Emu 115:35-48.

Dolman, G. and Moritz, C. (2006). A multilocus perspective on refugial isolation and divergence in rainforest skinks (Carlia). Evolution 60:573-82.

Dolman, G. and Phillips, B. (2004). Single copy nuclear DNA markers characterized for comparative phylogeography in Australian wet tropics rainforest skinks. Molecular Ecology Notes 4:185-187.

Dong, Z. D., Zhao, Y., Zeng, Q. F., Fu, Y., Zhou, F. N., Ji, X. S. and Wang, H. (2013). Molecular cloning and polymorphism of the major histocompatibility complex (MHC) class IIB gene of grass carp (Ctenopharyngodon idella). Biochem Genet 51:139-46.

Drummond, A., Ho, S., Phillips, M. and Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. PLoS Biology 4:e88.

Drummond, A. J., Ashton, B., Buxton, S., Cheung, M., Cooper, A., Duran, C., Field, M., Heled, J., Kearse, M. and Markowitz, S. (2012). Geneious Pro v5. 6, Available at: http://www.geneious.com.

Drummond, A. J. and Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. BMC evolutionary biology 7:214.

Dubas, G. (1987). Biotic determinants of home range size of the scincid lizard Trachydosaurus rugosus (Gray). Flinders University, Adelaide.

Dubas, G. and Bull, C. M. (1991). Diet Choise and Food Availability in the Omnivorous Lizard, Trachydosaurus rugosus. Wildlife Research 18:147-155.

Dubey, S. and Shine, R. (2010). Plio-Pleistocene diversification and genetic population structure of an endangered lizard (the Blue Mountains water skink, Eulamprus leuraensis) in south-eastern Australia. Journal of biogeography 37:902-914.

Duckett, P. E. and Stow, A. J. (2013). Higher genetic diversity is associated with stable water refugia for a gecko with a wide distribution in arid Australia. Diversity and Distributions 19:1072-1083.

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

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics.

Edwards, D. L., Keogh, J. S. and Knowles, L. L. (2012). Effects of vicariant barriers, habitat stability, population isolation and environmental features on species divergence in the south-western Australian coastal reptile community. Molecular ecology 21:3809-3822.

Edwards, D. L. and Melville, J. (2011). Extensive phylogeographic and morphological diversity in Diporiphora nobbi (Agamidae) leads to a taxonomic review and a new species description. Journal of Herpetology 45:530-546.

Ekblom, R. and Galindo, J. (2011). Applications of next generation sequencing in molecular ecology of non-model organisms. Heredity (Edinb) 107:1-15.

Elith, J., Phillips, S., Hastie, T., Dudik, M., Chee, Y. and Yates, C. (2011). A statistical explanation of MaxEnt for ecologists. Diversity and Distributions 17:43 - 57.

Estes, R. (1984). Fish, amphibians and reptiles from the Etadunna Formation, Miocene of South Australia. Australian Zoologist 21:335-343.

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

Excoffier, L. and Lischer, H. (2010a). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10:564 - 567.

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

Fant, J. B., Havens, K., Keller, J. M., Radosavljevic, A. and Yates, E. D. (2014). The influence of contemporary and historic landscape features on the genetic structure of the sand dune endemic, Cirsium pitcheri (Asteraceae). Heredity (Edinb) 112:519-530.

Farias-Hesson, E., Erikson, J., Atkins, A., Shen, P., Davis, R. W., Scharfe, C. and Pourmand, N. (2010). Semi-automated library preparation for high-throughput DNA sequencing platforms. BioMed Research International 2010.

Faulks, L. K., Gilligan, D. M. and Beheregaray, L. B. (2011). The role of anthropogenic vs. natural in-stream structures in determining connectivity and genetic diversity in an endangered freshwater fish, Macquarie perch (Macquaria australasica). Evolutionary Applications 4:589-601.

Ford, J. (1987). Minor Isolates and Minor Geographical Barriers in Avian Speciation in Continental Australia. Emu 87:90-102.

Frantz, A. C., Bertouille, S., Eloy, M. C., Licoppe, A., Chaumont, F. and Flamand, M. C. (2012). Comparative landscape genetic analyses show a Belgian motorway to be a gene flow barrier for red deer (Cervus elaphus), but not wild boars (Sus scrofa). Molecular ecology 21:3445-3457.

Freake, M. J. (1996). Orientation in the Australian sleepy lizard Tiliqua rugosa. Flinders University, Adelaide.

Fujioka, T., Chappell, J., Honda, M., Yatsevich, I., Fifield, K. and Fabel, D. (2005). Global cooling initiated stony deserts in central Australia 2–4 Ma, dated by cosmogenic 21Ne-10Be. Geology 33:993-996.

Fujita, M. K., McGuire, J. A., Donnellan, S. C. and Moritz, C. (2010). Diversification and persistence at the arid–monsoonal interface: Australia-wide biogeography of the Bynoe's gecko (Heteronotia binoei; Gekkonidae). Evolution 64:2293-2314.

Funk, D. J., Nosil, P. and Etges, W. J. (2006). Ecological divergence exhibits consistently positive associations with reproductive isolation across disparate taxa. Proceedings of the National Academy of Sciences of the United States of America 103:3209-3213.

Funk, D. J. and Omland, K. E. (2003). Species-level Paraphyly and Polyphyly: Frequency, Causes, and Consequences, with insights from Animal Mitochondrial DNA. Annual Review of Ecology, Evolution, and Systematics 34:397-423.

Gallagher, S. J., Greenwood, D. R., Taylor, D., Smith, A. J., Wallace, M. W. and Holdgate, G. R. (2003). The Pliocene climatic and environmental evolution of southeastern Australia: evidence from the marine and terrestrial realm. Palaeogeography, Palaeoclimatology, Palaeoecology 193:349-382.

Gallagher, S. J., Smith, A. J., Jonasson, K., Wallace, M. W., Holdgate, G. R., Daniels, J. and Taylor, D. (2001). The Miocene palaeoenvironmental and palaeoceanographic evolution of the Gippsland Basin, Southeast Australia: a record of Southern Ocean change. Palaeogeography, Palaeoclimatology, Palaeoecology 172:53-80.

Gardner, M. G., Bull, C. M., Cooper, S. J. and Duffield, G. A. (2001). Genetic evidence for a family structure in stable social aggregations of the Australian lizard Egernia stokesii. Mol Ecol 10:175-83.

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

Gardner, M. G., Hugall, A. F., Donnellan, S. C., Hutchinson, M. N. and Foster, R. (2008). Molecular systematics of social skinks: phylogeny and taxonomy of the Egernia group (Reptilia: Scincidae). Zoological Journal of the Linnean Society 154:781-794.

Gardner, M. G., Pearson, S. K., Johnston, G. R. and Schwarz, M. P. (2015). Group living in squamate reptiles: a review of evidence for stable aggregations. Biological Reviews:n/a-n/a.

Gaston, K. J. and Fuller, R. A. (2008). Commonness, population depletion and conservation biology. Trends in Ecology & Evolution 23:14-19.

Gauffre, B., Estoup, A., Bretagnolle, V. and Cosson, J. F. (2008). Spatial genetic structure of a small rodent in a heterogeneous landscape. Molecular ecology 17:4619-4629.

Georges, A., Li, Q., Lian, J., O'Meally, D., Deakin, J., Wang, Z., Zhang, P., Fujita, M., Patel, H. R., Holleley, C. E., Zhou, Y., Zhang, X., Matsurbara, K.,

Waters, P., Graves, J. A. M., Sarre, S. D. and Zhang, G. (2015). High-coverage sequencing and annotated assembly of the genome of the Australian dragon lizard *Pogona vitticeps.* GigaScience submitted.

Glaberman, S. and Caccone, A. (2008). Species-specific evolution of class I MHC genes in iguanas (order: Squamata; subfamily: Iguaninae). Immunogenetics 60:371-82.

Glaberman, S., Du Pasquier, L. and Caccone, A. (2008). Characterization of a nonclassical class I MHC gene in a reptile, the Galapagos marine iguana (Amblyrhynchus cristatus). PLoS One 3:e2859.

Glaberman, S., Moreno, M. A. and Caccone, A. (2009). Characterization and evolution of MHC class II B genes in Galapagos marine iguanas (Amblyrhynchus cristatus). Dev Comp Immunol 33:939-47.

Glaberman, S. R. (2008). Phylogeography and major histocompatibility complex evolution in the Galápagos Marine Iguana (Amblyrhynchus Cristatus) Yale University.

Glenn, T. C. (2011). Field guide to next-generation DNA sequencers. Mol Ecol Resour 11:759-769.

Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E. M., Brockman, W., Fennell, T., Giannoukos, G., Fisher, S., Russ, C., Gabriel, S., Jaffe, D. B., Lander, E. S. and Nusbaum, C. (2009). Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat Biotech 27:182-189.

Godfrey, S., Bull, C. M., Murray, K. and Gardner, M. (2006). Transmission mode and distribution of parasites among groups of the social lizard *Egernia stokesii.* Parasitology Research 99:223-230.

Gogol-Döring, A. and Chen, W. (2012). An Overview of the Analysis of Next Generation Sequencing Data, in: J. Wang, et al. (Eds.), Next Generation Microarray Bioinformatics, Humana Press. pp. 249-257.

Gorer, P. (1936). The detection of antigenic differences in mouse erythrocytes by the employment of immune sera. British journal of experimental pathology 17:42-50.

Greaves, S. N., Chapple, D. G., Gleeson, D. M., Daugherty, C. H. and Ritchie, P. A. (2007). Phylogeography of the spotted skink (Oligosoma lineoocellatum) and green skink (O. chloronoton) species complex (Lacertilia: Scincidae) in New Zealand reveals pre-Pleistocene divergence. Mol Phylogenet Evol 45:729-39.

Greer, A. (1989). The biology and evolution of Australian lizards. Surrey Beatty and Sons Pty Limited, Chipping Norton. New South Wales.

Grossberger, D. and Parham, P. (1992). Reptilian class I major histocompatibility complex genes reveal conserved elements in class I structure. Immunogenetics 36:166-174.

Guerin, G. R. and Lowe, A. J. (2013). Multi-species distribution modelling highlights the Adelaide Geosyncline, South Australia, as an important continental-scale arid-zone refugium. Austral Ecology 38:427-435.

Guindon, S. and Gascuel, O. (2003). A simple, fast and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic biology 52:696 - 704.

Gurrutxaga, M. and Saura, S. (2014). Prioritizing highway defragmentation locations for restoring landscape connectivity. Environmental Conservation 41:157-164.

Hammond, C., Braakman, I. and Helenius, A. (1994). Role of N-linked oligosaccharide recognition, glucose trimming, and calnexin in glycoprotein folding and quality control. Proceedings of the National Academy of Sciences 91:913-917.

Hancock-Hanser, B. L., Frey, A., Leslie, M. S., Dutton, P. H., Archer, F. I. and Morin, P. A. (2013). Targeted multiplex next-generation sequencing: advances in techniques of mitochondrial and nuclear DNA sequencing for population genomics. Mol. Ecol. Res. 13:254.

Hare, M. P. (2001). Prospects for nuclear gene phylogeography. Trends in Ecology & Evolution 16:700-706.

Harpending, H. C. (1994). Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. Human biology 66:591-600.

Hauswaldt, J. S., Stuckas, H., Pfautsch, S. and Tiedemann, R. (2007). Molecular characterization of MHC class II in a nonmodel anuran species, the fire-bellied toad Bombina bombina. Immunogenetics 59:479-91.

Havlicek, J. and Roberts, S. C. (2008). MHC-correlated mate choice in humans: A review. Psychoneuroendocrinology 34:497-512.

Hayden, M., Nguyen, T., Waterman, A. and Chalmers, K. (2008). Multiplex-Ready PCR: A new method for multiplexed SSR and SNP genotyping. BMC Genomics 9:80.

Hedrick, P. W. (1994). Purging inbreeding depression and the probability of extinction: full-sib mating. Heredity (Edinb) 73 (Pt 4):363-72.

Hedrick, P. W. (2011). Genetics of populations. 4th ed. Jones and Bartlett Sudbury, Mass.

Heffelfinger, C., Fragoso, C. A., Moreno, M. A., Overton, J. D., Mottinger, J. P., Zhao, H., Tohme, J. and Dellaporta, S. L. (2014). Flexible and scalable genotyping-by-sequencing strategies for population studies. BMC Genomics 15:979.

Hewitt, G. M. (1999). Post-glacial re-colonization of European biota. Biological journal of the Linnean Society 68:87-112.

Hewitt, G. M. (2000). The genetic legacy of the Quaternary ice ages. Nature 405:907-13.

Hewitt, G. M. (2004a). Genetic consequences of climatic oscillations in the Quaternary. Philos Trans R Soc Lond B Biol Sci 359:183-95.

Hewitt, G. M. (2004b). The structure of biodiversity–insights from molecular phylogeography. Frontiers in zoology 1:1-16.

Hewitt, G. M. and Ibrahim, K. M. (2001). Inferring glacial refugia and historical migrations with molecular phylogenies. Integrating Ecology and Evolution in a Spatial Context:271 - 294.

Hickerson, M. J., Carstens, B. C., Cavender-Bares, J., Crandall, K. A., Graham, C. H., Johnson, J. B., Rissler, L., Victoriano, P. F. and Yoder, A. D. (2010). Phylogeography's past, present, and future: 10 years after Avise, 2000. Molecular phylogenetics and evolution 54:291-301.

Ho, S., Shapiro, B., Phillips, M., Cooper, A. and Drummond, A. (2007). Evidence for time dependency of molecular rate estimates. Systematic biology 56:515 - 522.

Ho, S. Y. M. (2007). Calibrating molecular estimates of substitution rates and divergence times in birds. Journal of Avian Biology 38:409-414.

Hoffmann, A. A., Hallas, R. J., Dean, J. A. and Schiffer, M. (2003). Low potential for climatic stress adaptation in a rainforest Drosophila species. Science 301:100-102.

Hudson, R. R. and Turelli, M. (2003). Stochasticity Overrules the "Three-Times Rule": Genetic Drift, Genetic Draft, and Coalescence Times for Nuclear Loci Versus Mitochondrial DNA. Evolution 57:182-190.

Hugall, A., Moritz, C., Moussalli, A. and Stanisic, J. (2002). Reconciling paleodistribution models and comparative phylogeography in the Wet Tropics rainforest land snail Gnarosophia bellendenkerensis (Brazier 1875). Proc Natl Acad Sci U S A 99:6112-7.

Hugall, A. F. and Lee, M. S. (2004). Molecular claims of Gondwanan age for Australian agamid lizards are untenable. Molecular Biology and Evolution 21:2102-10.

Hughes, L. (2003). Climate change and Australia: Trends, projections and impacts. Austral Ecology 28:423-443.

Hughes, S. and Mouchiroud, D. (2001). High evolutionary rates in nuclear genes of Squamates. Journal of Molecular Evolution 53:70-76.

Hutchinson, M. N. (1992). Origins of the Australian scincid lizards: A preliminary report on the skinks of Riversleigh, Beagle, Records of the Northern Territory Museum of Arts and Sciences. pp. 61-69.

Janeway, C. A., Travers, P., Walport, M. and Capra, J. D. (1999). Immunobiology: the immune system in health and disease Current Biology Publications New York, NY;.

Joly, S. and Bruneau, A. (2006). Incorporating allelic variation for reconstructing the evolutionary history of organisms from multiple genes: an example from Rosa in North America. Systematic biology 55:623 - 636.

Joseph, L., Wilke, T., Ten Have, J. and Terry Chesser, R. (2006). Implications of mitochondrial DNA polyphyly in two ecologically undifferentiated but morphologically distinct migratory birds, the masked and white-browed woodswallows Artamus spp. of inland Australia. Journal of Avian Biology 37:625-636.

Karachanak, S., Carossa, V., Nesheva, D., Olivieri, A., Pala, M., Hooshiar Kashani, B., Grugni, V., Battaglia, V., Achilli, A., Yordanov, Y., Galabov, A. S., Semino, O., Toncheva, D. and Torroni, A. (2012). Bulgarians vs the other European populations: a mitochondrial DNA perspective. International Journal of Legal Medicine 126:497-503.

Kaschka-Dierich, C., Nazerian, K. and Thomssen, R. (1979). Intracellular State of Marek's Disease Virus DNA in Two Tumour-derived Chicken Cell Lines. Journal of General Virology 44:271-280.

Kaufman, J., Salomonsen, J. and Flajnik, M. (1994). Evolutionary conservation of MHC class I and class II molecules—different yet the same. Seminars in immunology 6:411-424.

Kawakami, T., Butlin, R. K., Adams, M., Saint, K. M., Paull, D. J. and Cooper, S. J. (2009). Re-examination of a proposed case of stasipatric speciation: phylogeography of the Australian morabine grasshoppers (Vandiemenella viatica species group). Mol Ecol 18:3429-42.

Kawamoto, Y., Takemoto, H., Higuchi, S., Sakamaki, T., Hart, J. A., Hart, T. B., Tokuyama, N., Reinartz, G. E., Guislain, P., Dupain, J., Cobden, A.

K., Mulavwa, M. N., Yangozene, K., Darroze, S., Devos, C. and Furuichi, T. (2013). Genetic Structure of Wild Bonobo Populations: Diversity of Mitochondrial DNA and Geographical Distribution. PLoS One 8:e59660.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P. and Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647-1649.

Kelley, J., Walter, L. and Trowsdale, J. (2005). Comparative genomics of major histocompatibility complexes. Immunogenetics 56:683 - 695.

Kidd, D. M. and Ritchie, M. G. (2006). Phylogeographic information systems: putting the geography into phylogeography. Journal of biogeography 33:1851-1865.

Kircher, M., Sawyer, S. and Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic acids research 40:e3.

Klein, J. (1986). Natural history of the major histocompatibility complex Wiley, New York.

Klein, J., Bontrop, R. E., Dawkins, R. L., Erlich, H. A., Gyllensten, U. B., Heise, E. R., Jones, P. P., Parham, P., Wakeland, E. K. and Watkins, D. I. (1990). Nomenclature for the major histocompatibility complexes of different species: a proposal. Immunogenetics 31:217-219.

Kloch, A., Babik, W., Bajer, A., Sinski, E. and Radwan, J. (2010). Effects of an MHC-DRB genotype and allele number on the load of gut parasites in the bank vole Myodes glareolus. Mol Ecol 19:255 - 265.

Klug, P. E., Wisely, S. M. and With, K. A. (2011). Population genetic structure and landscape connectivity of the Eastern Yellowbelly Racer (Coluber constrictor flaviventris) in the contiguous tallgrass prairie of northeastern Kansas, USA. Landscape Ecology 26:281-294.

Kollias, G. V., Jr. (1984). Immunologic Aspects of Infectious Diseases, in: G. Hoff, et al. (Eds.), Diseases of Amphibians and Reptiles, Springer US. pp. 661-691.

Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K. and Schloss, P. D. (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Applied and environmental microbiology 79:5112-5120.

Kriener, K., O'HUigin, C., Tichy, H. and Klein, J. (2000). Convergent evolution of major histocompatibility complex molecules in humans and New World monkeys. Immunogenetics 51:169-178.

Kumar, S. and Hedges, S. B. (1998). A molecular timescale for vertebrate evolution. Nature 392:917-920.

Kyle, C. J., Rico, Y., Castillo, S., Srithayakumar, V., Cullingham, C. I., White, B. N. and Pond, B. A. (2014). Spatial patterns of neutral and functional genetic variations reveal patterns of local adaptation in raccoon (Procyon lotor) populations exposed to raccoon rabies. Molecular ecology 23:2287-98.

Lafferty, K. D. and Gerber, L. R. (2002). Good Medicine for Conservation Biology: the Intersection of Epidemiology and Conservation Theory Buena Medicina para la Biología de la Conservación: la Intersección de la Epidemiología y la Teoría de Conservación. Conservation Biology 16:593-604.

Lanier, H. C., Edwards, D. L. and Knowles, L. L. (2013). Phylogenetic structure of vertebrate communities across the Australian arid zone. Journal of biogeography 40:1059-1070.

Laurenson, K., Sillero-Zubiri, C., Thompson, H., Shiferaw, F., Thirgood, S. and Malcolm, J. (1998). Disease as a threat to endangered species: Ethiopian wolves, domestic dogs and canine pathogens. Animal Conservation 1:273-280.

Lee, J. Y. and Edwards, S. V. (2008). Divergence across Australia's Carpentarian barrier: statistical phylogeography of the red-backed fairy wren (Malurus melanocephalus). Evolution 62:3117-34.

Lemey, P., Rambaut, A., Welch, J. J. and Suchard, M. A. (2010). Phylogeography Takes a Relaxed Random Walk in Continuous Space and Time. Molecular Biology and Evolution 27:1877-1885.

Lemmon, E. M. and Lemmon, A. R. (2013). High-Throughput Genomic Data in Systematics and Phylogenetics. Annual Review of Ecology, Evolution, and Systematics 44:99-121.

Levy, E., Kennington, W. J., Tomkins, J. L. and LeBas, N. R. (2012). Phylogeography and population genetic structure of the ornate dragon lizard, Ctenophorus ornatus.

Li, C., Hofreiter, M., Straube, N., Corrigan, S. and Naylor, G. J. (2013). Capturing protein-coding genes across highly divergent species. BioTechniques 54:321-6.

Li, Q., Simo, J. A., McGowran, B. and Holbourn, A. (2004). The eustatic and tectonic origin of Neogene unconformities from the Great Australian Bight. Marine Geology 203:57-81.

Li, Y., Lancaster, M. L., Carthew, S. M., Packer, J. G. and Cooper, S. J. B. (2014). Delineation of conservation units in an endangered marsupial, the southern brown bandicoot (Isoodon obesulus obesulus), in South Australia/western Victoria, Australia. Australian journal of zoology 62:345-359.

Librado, P. and Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451-2.

Liggins, L., Chapple, D. G., Daugherty, C. H. and Ritchie, P. A. (2008a). Origin and post-colonization evolution of the Chatham Islands skink (Oligosoma nigriplantare nigriplantare). Molecular ecology 17:3290-3305.

Liggins, L., Chapple, D. G., Daugherty, C. H. and Ritchie, P. A. (2008b). A SINE of restricted gene flow across the Alpine Fault: phylogeography of the New Zealand common skink (Oligosoma nigriplantare polychroma). Molecular ecology 17:3668-3683.

Lowe, A., Harris, S. and Ashton, P. (2009). Ecological genetics: design, analysis, and application John Wiley & Sons.

Lyles, A. M. and Dobson, A. P. (1993). Infectious disease and intensive management: population dynamics, threatened hosts, and their parasites. Journal of Zoo and Wildlife Medicine 24:315-326.

Mach, M. E., Sbrocco, E. J., Hice, L. A., Duffy, T. A., Conover, D. O. and Barber, P. H. (2011). Regional differentiation and post-glacial expansion of the Atlantic silverside, Menidia menidia, an annual fish with high dispersal potential. Marine Biology 158:515-530.

Madden, D. R. (1995). The three-dimensional structure of peptide-MHC complexes.

. Annu Rev Immunol 13:587-622.

Maldonado, S. P., Melville, J., Peterson, G. N. L. and Sumner, J. (2012). Human-induced versus historical habitat shifts: identifying the processes that shaped the genetic structure of the threatened grassland legless lizard, Delma impar. Conservation genetics 13:1329-1342.

Margraf, N., Verdon, A., Rahier, M. and Naisbit, R. E. (2007). Glacial survival and local adaptation in an alpine leaf beetle. Molecular ecology 16:2333-2343.

Maricic, T., Whitten, M. and Pääbo, S. (2010). Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. PLoS One 5:e14004.

Marín, J. C., Varas, V., Vila, A. R., López, R., Orozco-terWengel, P. and Corti, P. (2013). Refugia in Patagonian fjords and the eastern Andes during the Last Glacial Maximum revealed by huemul (Hippocamelus bisulcus) phylogeographical patterns and genetic diversity. Journal of biogeography 40:2285-2298.

Markgraf, V., McGlone, M. and Hope, G. (1995). Neogene paleoenvironmental and paleoclimatic change in southern temperate ecosystems — a southern perspective. Trends in Ecology & Evolution 10:143-147.

Martin, H. A. (1978). Evolution of the Australian flora and vegetation through the Tertiary: evidence from pollen. Alcheringa: An Australasian Journal of Palaeontology 2:181-202.

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

Martin, J. E., Hutchinson, M. N., Meredith, R., Case, J. A. and Pledge, N. S. (2004). The oldest genus of scincid lizard (Squamata) from the tertiary etadunna formation of South Australia. Journal of Herpetology 38:180-187.

Mason, R. and Parker, M. R. (2010). Social behavior and pheromonal communication in reptiles. Journal of Comparative Physiology A 196:729-749.

McCormack, J. E., Hird, S. M., Zellmer, A. J., Carstens, B. C. and Brumfield, R. T. (2013). Applications of next-generation sequencing to phylogeography and phylogenetics. Molecular phylogenetics and evolution 66:526-38.

Mcdonald, J. H. and Kreitman, M. (1991). Adaptive protein evolution at the Adh locus in Drosophila. Nature 351:652-654.

McElhoe, J. A., Holland, M. M., Makova, K. D., Su, M. S., Paul, I. M., Baker, C. H., Faith, S. A. and Young, B. (2014). Development and assessment of an optimized next-generation DNA sequencing approach for the mtgenome using the Illumina MiSeq. Forensic Science International: Genetics 13:20-29.

Meglecz, E., Piry, S., Desmarais, E., Galan, M., Gilles, A., Guivier, E., Pech, N. and Martin, J. F. (2011). SESAME (SEquence Sorter & AMplicon Explorer): genotyping based on high-throughput multiplex amplicon sequencing. Bioinformatics 27:277-8.

Meiklejohn, C. D., Montooth, K. L. and Rand, D. M. (2007). Positive and negative selection on the mitochondrial genome. Trends in Genetics 23:259-263.

Meyer, M. and Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harbor Protocols 2010:pdb. prot5448.

Meyers, L. A. and Bull, J. J. (2002). Fighting change with change: adaptive variation in an uncertain world. Trends in Ecology & Evolution 17:551-557.

Milinski, M. (2006). The major histocompatibility complex, sexual selection, and mate choice. Annual Review of Ecology, Evolution, and Systematics 37:159-186.

Miller, H. and Lambert, D. (2004). Gene duplication and gene conversion in class II MHC genes of New Zealand robins (Petroicidae). Immunogenetics 56:178 - 191.

Miller, H. C., Andrews-Cookson, M. and Daugherty, C. H. (2007a). Two patterns of variation among MHC class I loci in tuatara (Sphenodon punctatus). Journal of Heredity 98:666-677.

Miller, H. C., Belov, K. and Daugherty, C. H. (2005). Characterization of MHC class II genes from an ancient reptile lineage, Sphenodon (tuatara). Immunogenetics 57:883-91.

Miller, H. C., Belov, K., Daugherty, C. H. and Investigators, S. T.-N. Y. (2006). Proceedings of the SMBE Tri-National Young Investigators' Workshop 2005. MHC Class I genes in the Tuatara (Sphenodon spp.): evolution of the MHC in an ancient reptilian order. Molecular Biology and Evolution 23:949-56.

Miller, M. R., Dunham, J. p., Amores, A., Cresko, W. A. and Johnson, E. A. (2007b). Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Res 17:240-248.

Moore, W. S. (1995). Inferring Phylogenies from mtDNA Variation: Mitochondrial-Gene Trees Versus Nuclear-Gene Trees. Evolution 49:718-726.

Moritz, C., Patton, J., Schneider, C. and Smith, T. (2000). Diversification of Rainforest Faunas: an integrated molecular approach. Annual Reviews of Ecology and Systematics 31:533 - 563.

Moussy, C., Hosken, D. J., Mathews, F., Smith, G. C., Aegerter, J. N. and Bearhop, S. (2013). Migration and dispersal patterns of bats and their influence on genetic structure. Mammal Review 43:183-195.

Mullen, L. B., Arthur Woods, H., Schwartz, M. K., Sepulveda, A. J. and Lowe, W. H. (2010). Scale-dependent genetic structure of the Idaho giant salamander (Dicamptodon aterrimus) in stream networks. Molecular ecology 19:898-909.

Murphy, B. F., Thompson, M. B. and Belov, K. (2009). Evolution of viviparity and the maternal immune system: major histocompatibility complex (MHC) class I genes in skinks. Orbit: University of Sydney Undergraduate Research Journal 1(1).

Murray, D. L., Kapke, C. A., Evermann, J. F. and Fuller, T. K. (1999). Infectious disease and the conservation of free-ranging large carnivores. Animal Conservation 2:241-254.

Nadachowska-Brzyska, K., Burri, R., Olason, P. I., Kawakami, T., Smeds, L. and Ellegren, H. (2013). Demographic Divergence History of Pied Flycatcher and Collared Flycatcher Inferred from Whole-Genome Re-sequencing Data. PLoS Genet 9:e1003942.

Nadachowska-Brzyska, K., ZieliŃSki, P., Radwan, J. and Babik, W. (2012). Interspecific hybridization increases MHC class II diversity in two sister species of newts. Molecular ecology 21:887-906.

Nag, B., Passmore, D., Kendrick, T., Bhayani, H. and Sharma, S. D. (1992). N-linked oligosaccharides of murine major histocompatibility complex class II

molecule. Role in antigenic peptide binding, T cell recognition, and clonal nonresponsiveness. Journal of Biological Chemistry 267:22624-22629.

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 22:2841-2847.

Neaves, L. E., Zenger, K. R., Prince, R. I. T. and Eldridge, M. D. B. (2012). Impact of Pleistocene aridity oscillations on the population history of a widespread, vagile Australian mammal, Macropus fuliginosus. Journal of biogeography 39:1545-1563.

Nei, M. and Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Molecular Biology and Evolution 3:418-426.

Nei, M., Gu, X. and Sitnikova, T. (1997). Evolution by the birth-and-death process in multigene families of the vertebrate immune system. Proc Natl Acad Sci U S A 94:7799-806.

Nei, M. and Kumar, S. (2000). Molecular evolution and phylogenetics Oxford University Press, New yourk.

Niu, B., Fu, L., Sun, S. and Li, W. (2010). Artificial and natural duplicates in pyrosequencing reads of metagenomic data. BMC Bioinformatics 11:187.

Nix, H. A. (1982). Environmental determinants of biogeography and evolution in Terra Australis. Evolution of the flora and fauna of arid Australia 47.

Nowakowski, A. J., DeWoody, J. A., Fagan, M. E., Willoughby, J. R. and Donnelly, M. A. (2015). Mechanistic insights into landscape genetic structure of two tropical amphibians using field-derived resistance surfaces. Molecular ecology 24:580-595.

O'Neill, S. B., Chapple, D. G., Daugherty, C. H. and Ritchie, P. A. (2008). Phylogeography of two New Zealand lizards: McCann's skink (Oligosoma maccanni) and the brown skink (O. zelandicum). Molecular phylogenetics and evolution 48:1168-1177.

Ohta, Y., Goetz, W., Hossain, M. Z., Nonaka, M. and Flajnik, M. F. (2006). Ancestral organization of the MHC revealed in the amphibian Xenopus. J Immunol 176:3674-85.

Oliver, P. M., Adams, M., Lee, M. S. Y., Hutchinson, M. N. and Doughty, P. (2009). Cryptic diversity in vertebrates: molecular data double estimates of species diversity in a radiation of Australian lizards (Diplodactylus, Gekkota).

Olsson, M. and Shine, R. (1998). Chemosensory mate recognition may facilitate prolonged mate guarding by male snow skinks, Niveoscincus microlepidotus. Behavioral Ecology and Sociobiology 43:359-363.

Palumbi, S. R., Cipriano, F. and Hare, M. P. (2001). Predicting nuclear gene coalescence from mitochondrial data: the three-times rule. Evolution 55:859-68.

Pavlova, A., Amos, J. N., Joseph, L., Loynes, K., Austin, J. J., Keogh, J. S., Stone, G. N., Nicholls, J. A. and Sunnucks, P. (2013). Perched at the mito-nuclear crossroads: divergent mitochondrial lineages correlate with environment in the face of ongoing nuclear gene flow in an Australian bird. Evolution 67:3412-28.

Peakall, R. and Smouse, P. E. (2006). genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6:288-295.

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

Penn, D. J. (2002). The Scent of Genetic Compatibility: Sexual Selection and the Major Histocompatibility Complex. Ethology 108:1-21.

Penn, D. J. and Potts, W. K. (1999). The Evolution of Mating Preferences and Major Histocompatibility Complex Genes. The American Naturalist 153:145-164.

Pepper, M., Ho, S. Y. W., Fujita, M. K. and Scott Keogh, J. (2011). The genetic legacy of aridification: Climate cycling fostered lizard diversification in Australian montane refugia and left low-lying deserts genetically depauperate. Molecular phylogenetics and evolution 61:750-759.

Peterman, W. E., Feist, S. M., Semlitsch, R. D. and Eggert, L. S. (2013). Conservation and management of peripheral populations: Spatial and temporal influences on the genetic structure of wood frog (Rana sylvatica) populations. Biological conservation 158:351-358.

Phillips, S. J., Anderson, R. P. and Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. Ecological Modelling 190:231-259.

Phillips, S. J. and Dudik, M. (2008). Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 31:161-175.

Phillipsen, I. C. and Lytle, D. A. (2013). Aquatic insects in a sea of desert: population genetic structure is shaped by limited dispersal in a naturally fragmented landscape. Ecography 36:731-743.

Pielou, E. C. (1991). After the Ice Age: The return of life to glaciated North America University of Chicago Press, Chicago.

Ploegh, H. and Watts, C. (1998). Antigen recognition. Curr Opin Immunol 10:57-8.

Price, A. L., Jones, N. C. and Pevzner, P. A. (2005). De novo identification of repeat families in large genomes. Bioinformatics 21:i351-i358.

Prideaux, G. J., Long, J. A., Ayliffe, L. K., Hellstrom, J. C., Pillans, B., Boles, W. E., Hutchinson, M. N., Roberts, R. G., Cupper, M. L., Arnold, L. J., Devine, P. D. and Warburton, N. M. (2007). An arid-adapted middle Pleistocene vertebrate fauna from south-central Australia. Nature 445:422-425.

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

Pyron, R. A., Burbrink, F. T. and Wiens, J. J. (2013). A phylogeny and revised classification of Squamata, including 4161 species of lizards and snakes. BMC evolutionary biology 13:93.

Quail, M. A., Smith, M., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., Bertoni, A., Swerdlow, H. P. and Gu, Y. (2012). A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics 13:341.

Rabosky, D. L., Hutchinson, M. N., Donnellan, S. C., Talaba, A. L. and Lovette, I. J. (2014). Phylogenetic disassembly of species boundaries in a widespread group of Australian skinks (Scincidae: Ctenotus). Molecular phylogenetics and evolution 77:71-82.

Radtkey, R. R., Becker, B., Miller, R. D., Riblet, R. and Case, T. J. (1996). Variation and evolution of class I Mhc in sexual and parthenogenetic geckos. Proc Biol Sci 263:1023-32.

Rambaut, A. and Drummond, A. (2009a). FigTree v1. 3.1. Program distributed by the author. Institute of Evolutionary Biology, University of Edinburgh. Edinburgh, United Kingdom.

Rambaut, A. and Drummond, A. (2009b). Tracer, version 1.4. Computer program and documentation distributed by the author, website http://beast. bio. ed. ac. uk/Tracer.

Rasic, G. and Keyghobadi, N. (2012). From broadscale patterns to fine-scale processes: habitat structure influences genetic differentiation in the pitcher plant midge across multiple spatial scales. Molecular ecology 21:223-236.

Reich, D. E., Feldman, M. W. and Goldstein, D. B. (1999). Statistical properties of two tests that use multilocus data sets to detect population expansions. Molecular Biology and Evolution 16:453-466.

Riddle, B. R. (1996). The molecular phylogeographic bridge between deep and shallow history in continental biotas. Trends in Ecology & Evolution 11:207-211.

Robinson, J. T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G. and Mesirov, J. P. (2011). Integrative genomics viewer. Nat Biotech 29:24-26.

Rodgers, J. R. and Cook, R. G. (2005). MHC class Ib molecules bridge innate and acquired immunity. Nat Rev Immunol 5:459-471.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Hohna, S., Larget, B., Liu, L., Suchard, M. A. and Huelsenbeck, J. P. (2012).

MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol 61:539-42.

Rosenblum, E. B., Belfiore, N. M. and Moritz, C. (2007). Anonymous nuclear markers for the eastern fence lizard, Sceloporus undulatus. Molecular Ecology Notes 7:113-116.

Row, J. R., Blouin-Demers, G. and Lougheed, S. C. (2010). Habitat distribution influences dispersal and fine-scale genetic population structure of eastern foxsnakes (Mintonius gloydi) across a fragmented landscape. Molecular ecology 19:5157-5171.

Rozen, S. and Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132:365-86.

Rueness, E. K., Stenseth, N. C., O'Donoghue, M., Boutin, S., Ellegren, H. and Jakobsen, K. S. (2003). Ecological and genetic spatial structuring in the Canadian lynx. Nature 425:69-72.

Satrawaha, R. and Bull, C. M. (1981). The Area Occupied by an Omnivorous Lizard, Trachydosaurus rugosus. Wildlife Research 8:435-442.

Schneider, C. and Moritz, C. (1999). Rainforest refugia and Australia's Wet Tropics. Proceedings of the Royal Society of London. Series B: Biological Sciences 266:191-196.

Schofield, J. A., Fenner, A. L., Pelgrim, K. and Bull, C. M. (2012). Male-biased movement in pygmy bluetongue lizards: implications for conservation. Wildlife Research 39:677-684.

Schofield, J. A., Gardner, M. G., Fenner, A. L. and Bull, C. M. (2014). Promiscuous mating in the endangered Australian lizard Tiliqua adelaidensis: a potential windfall for its conservation. Conservation genetics 15:177-185.

Sexton, J. P., Hangartner, S. B. and Hoffmann, A. A. (2014). Genetic isolation by environment or distance: which pattern of gene flow is most common? Evolution 68:1-15.

Shawar, S. M., Vyas, J. M., Rodgers, J. R. and Rich, R. R. (1994). Antigen presentation by major histocompatibility complex class I-B molecules. Annu Rev Immunol 12:839-80.

Shea, G. M. and Hutchinson, M. N. (1992). A new species of lizard (Tiliqua) from the Miocene of Riversleigh, Queensland. Memoirs of the Queensland Museum 32:303-310.

Sheard, M. J. and Smith, P. C. (1995). Karst and mound spring deposits, J.F. Drexel, W.V. Preiss (Eds.), The Geology of South Australia, , The Phanerozoic,, Department of Mines and Energy, Adelaide. pp. 257-261.

Shiina, T., Shimizu, S., Hosomichi, K., Kohara, S., Watanabe, S., Hanzawa, K., Beck, S., Kulski, J. K. and Inoko, H. (2004). Comparative Genomic Analysis of Two Avian (Quail and Chicken) MHC Regions. The Journal of Immunology 172:6751-6763.

Shoo, L. P., Rose, R., Doughty, P., Austin, J. J. and Melville, J. (2008). Diversification patterns of pebble-mimic dragons are consistent with historical disruption of important habitat corridors in arid Australia. Molecular phylogenetics and evolution 48:528-542.

Sieling, P. A., Chatterjee, D., Porcelli, S. A., Prigozy, T. I., Mazzaccaro, R. J., Soriano, T., Bloom, B. R., Brenner, M. B., Kronenberg, M., Brennan, P. J. and et al. (1995). CD1-restricted T cell recognition of microbial lipoglycan antigens. Science 269:227-30.

Smissen, P. J., Melville, J., Sumner, J. and Jessop, T. S. (2013). Mountain barriers and river conduits: phylogeographical structure in a large, mobile lizard (Varanidae: Varanus varius) from eastern Australia. Journal of biogeography 40:1729-1740.

Smit, A. F. A., Hubley, R. and Green, P. (2010). RepeatMasker open-3.0. Available from: http://repeatmasker.org.

Sokal, R. R. and Rohlf, F. J. (1981). Biometry, Freeman San Francisco, CA: Freeman.

Soldati, D., Foth, B. J. and Cowman, A. F. (2004). Molecular and functional aspects of parasite invasion. Trends in Parasitology 20:567-574.

Sommer, S. (2005a). The importance of immune gene variability (MHC) in evolutionary ecology and conservation. Front Zool 2:16.

Sommer, S. (2005b). The importance of immune gene variability (MHC) in evolutionary ecology and conservation. Front Zool 2:16.

Stamatakis, A., Hoover, P. and Rougemont, J. (2008). A rapid bootstrap algorithm for the RAxML web servers. Systematic biology 57:758-771.

Steele, C. A. and Storfer, A. (2006). Coalescent-based hypothesis testing supports multiple Pleistocene refugia in the Pacific Northwest for the Pacific giant salamander (Dicamptodon tenebrosus). Molecular ecology 15:2477-2487.

Stephens, M. and Scheet, P. (2005). Accounting for Decay of Linkage Disequilibrium in Haplotype Inference and Missing-Data Imputation. The American Journal of Human Genetics 76:449-462.

Stephens, M., Smith, N. J. and Donnelly, P. (2001). A new statistical method for haplotype reconstruction from population data. The American Journal of Human Genetics 68:978-989.

Stephenson, A. E. (1986). Lake Bungunnia — A Plio-Pleistocene megalake in southern Australia. Palaeogeography, Palaeoclimatology, Palaeoecology 57:137-156.

Stern, L. J., Brown, J. H., Jardetzky, T. S., Gorga, J. C., Urban, R. G., Strominger, J. L. and Wiley, D. C. (1994). Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. Nature 368:215-221.

Stockwell, C. A., Hendry, A. P. and Kinnison, M. T. (2003). Contemporary evolution meets conservation biology. Trends in Ecology & Evolution 18:94-101.

Stow, A. J. and Sunnucks, P. (2004). Inbreeding avoidance in Cunningham's skinks (Egernia cunninghami) in natural and fragmented habitat. Molecular ecology 13:443-447.

Stow, A. J., Sunnucks, P., Briscoe, D. A. and Gardner, M. G. (2001). The impact of habitat fragmentation on dispersal of Cunningham's skink (Egernia cunninghami): evidence from allelic and genotypic analyses of microsatellites. Mol Ecol 10:867-78.

Strandh, M., Lannefors, M., Bonadonna, F. and Westerdahl, H. (2011). Characterization of MHC class I and II genes in a subantarctic seabird, the blue petrel, Halobaena caerulea (Procellariiformes). Immunogenetics 63:653-666.

Swofford, D. L. (2002). PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods), Sinauer Associates Sunderland.

Symula, R., Keogh, J. S. and Cannatella, D. C. (2008). Ancient phylogeographic divergence in southeastern Australia among populations of

the widespread common froglet, Crinia signifera. Molecular phylogenetics and evolution 47:569-580.

Taberlet, P., Fumagalli, L., Wust-Saucy, A. G. and Cosson, J. F. (1998). Comparative phylogeography and postglacial colonization routes in Europe. Mol Ecol 7:453-64.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. (2011). MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Molecular Biology and Evolution 28:2731-2739.

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

Telford, S. R. and Stein, J. (2000). Two malaria parasites (Apicomplexa: Plasmodiidae) of the Australian skink *Egernia stokesii*. Journal of Parasitology 86:395-406.

Thomas, C. D., Cameron, A., Green, R. E., Bakkenes, M., Beaumont, L. J., Collingham, Y. C., Erasmus, B. F., De Siqueira, M. F., Grainger, A., Hannah, L., Hughes, L., Huntley, B., Van Jaarsveld, A. S., Midgley, G. F., Miles, L., Ortega-Huerta, M. A., Peterson, A. T., Phillips, O. L. and Williams, S. E. (2004). Extinction risk from climate change. Nature 427:145-8.

Thome, M. T., Zamudio, K. R., Haddad, C. F. and Alexandrino, J. (2014). Barriers, rather than refugia, underlie the origin of diversity in toads endemic to the Brazilian Atlantic Forest. Molecular ecology 23:6152-6164.

Thomson, R. C., Shedlock, A. M., Edwards, S. V. and Shaffer, H. B. (2008). Developing markers for multilocus phylogenetics in non-model organisms: A test case with turtles. Molecular phylogenetics and evolution 49:514-525.

Thomson, R. C., Wang, I. J. and Johnson, J. R. (2010). Genome-enabled development of DNA markers for ecology, evolution and conservation. Molecular ecology 19:2184-2195.

Toon, A., Mather, P. B., Baker, A. M., Durrant, K. L. and Hughes, J. M. (2007). Pleistocene refugia in an arid landscape: analysis of a widely distributed Australian passerine. Molecular ecology 16:2525-2541.

Trizio, I., Crestanello, B., Galbusera, P., Wauters, L. A., Tosi, G., Matthysen, E. and Hauffe, H. C. (2005). Geographical distance and physical barriers shape the genetic structure of Eurasian red squirrels (Sciurus vulgaris) in the Italian Alps. Molecular ecology 14:469-481.

Ujvari, B. and Belov, K. (2011). Major histocompatibility complex (MHC) markers in conservation biology. International journal of molecular sciences 12:5168-5186.

van Erp, S. H., Egberts, E. and Stet, R. J. M. (1996). Characterization of class IIA andB genes in a gynogenetic carp clone. Immunogenetics 44:192-202.

van Tienderen, P. H., de Haan, A. A., Van der Linden, C. G. and Vosman, B. (2002). Biodiversity assessment using markers for ecologically important traits. Trends in Ecology & Evolution 17:577-582.

Vigueira, C. C., Rauh, B., Mitchell-Olds, T. and Lawton-Rauh, A. L. (2013a). Signatures of Demography and Recombination at Coding Genes in Naturally-Distributed Populations of <italic>Arabidopsis Lyrata</italic> Subsp. <italic>Petraea</italic>. PLoS ONE 8:e58916.

Vigueira, C. C., Rauh, B., Mitchell-Olds, T. and Lawton-Rauh, A. L. (2013b). Signatures of Demography and Recombination at Coding Genes in Naturally-Distributed Populations of *Arabidopsis Lyrata* Subsp. *Petraea*. PLoS One 8:e58916.

Vitt, L. J., Pianka, E. R., Cooper, W. E., Jr. and Schwenk, K. (2003). History and the global ecology of squamate reptiles. Am Nat 162:44-60.

Wan, Q. H., Wu, H., Fujihara, T. and Fang, S. G. (2004). Which genetic marker for which conservation genetics issue? ELECTROPHORESIS 25:2165-2176.

Wang, J. H., Meijers, R., Xiong, Y., Liu, J. H., Sakihama, T., Zhang, R., Joachimiak, A. and Reinherz, E. L. (2001). Crystal structure of the human CD4 N-terminal two-domain fragment complexed to a class II MHC molecule. Proceedings of the National Academy of Sciences 98:10799-10804.

Webb, J. A. and James, J. M. (2006). Karst evolution of the Nullarbor Plain, Australia, 'Perspectives on Karst Geomorphology, Hydrology, and Geochemistry – A Tribute Volume to Derek C. Ford and William B. White'. (Eds R. S. Harmon and C. Wicks.), SPECIAL PAPERS-GEOLOGICAL SOCIETY OF AMERICA. pp. 65-78.

Webb, S. (2008). Megafauna demography and late Quaternary climatic change in Australia: A predisposition to extinction. Boreas 37:329-345.

Wegner, K. M., Reusch, T. B. H. and Kalbe, M. (2003). Multiple parasites are driving major histocompatibility complex polymorphism in the wild. Journal of Evolutionary Biology 16:224-232.

Wei, B. Y., Buerstedde, J. M., Bell, M., Chase, C., Nilson, A., Browne, A., Pease, L. and McKean, D. J. (1991). Functional effects of N-linked

oligosaccharides located on the external domain of murine class II molecules. The Journal of Immunology 146:2358-66.

Welton, L. J., Siler, C. D., Oaks, J. R., Diesmos, A. C. and Brown, R. M. (2013). Multilocus phylogeny and Bayesian estimates of species boundaries reveal hidden evolutionary relationships and cryptic diversity in Southeast Asian monitor lizards. Molecular ecology 22:3495-3510.

Wheat, C. W. (2010). Rapidly developing functional genomics in ecological model systems via 454 transcriptome sequencing. Genetica 138:433-451.

While, G. M., Chapple, D. G., Gardner, M. G., Uller, T. and Whiting, M. J. (2015). Egernia lizards. Current Biology 25:R593-R595.

Whiteley, A. R., Spruell, P. and Allendorf, F. W. (2006). Can common species provide valuable information for conservation? Molecular ecology 15:2767-2786.

Wielstra, B., Crnobrnja-Isailovi, J., Litvinchuk, S. N., Reijnen, B. T., Skidmore, A. K., Sotiropoulos, K., Toxopeus, A. G., Tzankov, N., Vukov, T. and Arntzen, J. W. (2013). Tracing glacial refugia of Triturus newts based on mitochondrial DNA phylogeography and species distribution modeling. Front Zool 10:13.

Williams, M., Prescott, J. R., Chappell, J., Adamson, D., Cock, B., Walker, K. and Gell, P. (2001). The enigma of a late Pleistocene wetland in the Flinders Ranges, South Australia. Quaternary International 83–85:129-144.

Xu, T., Sun, Y., Shi, G., Cheng, Y. and Wang, R. (2011). Characterization of the Major Histocompatibility Complex Class II Genes in Miiuy Croaker. PLoS One 6:e23823.

Yannic, G., Pellissier, L., Dubey, S., Vega, R., Basset, P., Mazzotti, S., Pecchioli, E., Vernesi, C., Hauffe, H. C., Searle, J. B. and Hausser, J. (2012). Multiple refugia and barriers explain the phylogeography of the Valais shrew, Sorex antinorii (Mammalia: Soricomorpha). Biological journal of the Linnean Society 105:864-880.

Yeatman, E. M. (1988). Resource partitioning by three congeneric species of skink (Tiliqua) in sympatry in South Australia. Flinders University, .

Young, A., Boyle, T. and Brown, T. (1996). The population genetic consequences of habitat fragmentation for plants. Trends in Ecology & Evolution 11:413-418.

Younger, J. L., Clucas, G. V., Kooyman, G., Wienecke, B., Rogers, A. D., Trathan, P. N., Hart, T. and Miller, K. J. (2015). Too much of a good thing: sea ice extent may have forced emperor penguins into refugia during the last glacial maximum. Global Change Biology 21:2215-2226.

Zamudio, K. R. and Greene, H. W. (1997). Phylogeography of the bushmaster (Lachesis muta: Viperidae): implications for neotropical biogeography, systematics, and conservation. Biological journal of the Linnean Society 62:421-442.

Zamudio, K. R. and Savage, W. K. (2003). Historical isolation, range expansion, and secondary contact of two highly divergent mitochondrial lineages in spotted salamanders (Ambystoma maculatum). Evolution 57:1631-52.

Zhang, D. X. and Hewitt, G. M. (2003). Nuclear DNA analyses in genetic studies of populations:practice, problems and prospects. Molecular ecology 12:563 - 584.

Zink, R. M. and Barrowclough, G. F. (2008). Mitochondrial DNA under siege in avian phylogeography. Mol Ecol 17:2107-21.

9. APPENDIX

9.1. APPENDIX I

Animal Behaviour 97 (2014) 35–43

Contents lists available at ScienceDirect

Animal Behaviour

journal homepage: www.elsevier.com/locate/anbehav

A contact-based social network of lizards is defined by low genetic relatedness among strongly connected individuals

Stephanie S. Godfrey [a, b, *], Talat Hojat Anasri [a], Michael G. Gardner [a, c], Damien R. Farine [d, e, f], C. Michael Bull [a]

[a] School of Biological Sciences, Flinders University, Adelaide, Australia
[b] School of Veterinary and Life Sciences, Murdoch University, Perth, Australia
[c] Evolutionary Biology Unit, South Australian Museum, Adelaide, Australia
[d] Edward Grey Institute of Field Ornithology, Department of Zoology, University of Oxford, Oxford, U.K.
[e] Smithsonian Tropical Research Institute, Panama
[f] Department of Anthropology, University of California Davis, Davis, CA, U.S.A.

ARTICLE INFO

Article history:
Received 3 December 2013
Initial acceptance 5 February 2014
Final acceptance 8 August 2014
Published online
MS. number: 13-01005R

Keywords:
lizard
male–male interactions
pair bond
relatedness
Scincidae
social network
space use

Social organization is widespread; even largely solitary species must organize themselves to enable contacts with mates and reduce competition with conspecifics. Although the forms of social structure can be subtle in solitary species, understanding the factors that influence them may be important for understanding how different forms of social organization evolved. We investigated the influence of genetic relatedness and spatial structure on social associations in a solitary living Australian scincid lizard, *Tiliqua rugosa*. We derived the genetic relatedness of 46 lizards from analysis of genotypes at 15 microsatellite DNA loci, and described social networks from GPS locations of all the lizards every 10 min for 81 days during their main activity period of the year. We found that connected male dyads were significantly more related than expected by chance, whereas connected male–female and female–female dyads had lower relatedness than expected. Among neighbouring male–male and male–female dyads, the strongest social relationships were between lizards that were the least related. Explanations of this pattern may include the avoidance of inbreeding in male–female dyads, or the direction of aggressive behaviour towards less related individuals in male–male dyads. Observed social associations (inferred through synchronous spatial proximity) were generally lower than expected from null models derived from home range overlap, and many close neighbours did not make social contact. This supports our hypothesis for the presence of deliberate avoidance between some neighbouring individuals. We suggest that lizards can discriminate between different levels of relatedness in their neighbours, directing their social interactions towards those that are less related. This highlights differences in how social associations are formed between species that are solitary (where associations form between unrelated conspecifics) and species that maintain stable social groups structured by kinship.

Animal species range from solitary to eusocial in their social organization (Linksvayer, 2010; Michener, 1969), but all interact socially with conspecifics at some times, in some of their activities. An ongoing question is how genetic relatedness influences these social associations (Wilson, 1975). There are two main mechanisms by which genetic relatedness may influence social interactions. First, if individuals have limited opportunity for dispersal, they may avoid inbreeding by reducing social contact with related individuals of the opposite sex (Pusey & Wolf, 1996). Second, indirect fitness accrued through interactions with kin may be favoured when the benefits from cooperating exceed the costs associated with close living (Alexander, 1974). Even in reptiles, benefits can be gained from kin-structured social interactions. For example, the gidgee skink, *Egernia stokesii*, lives in highly related groups that include a breeding pair and one or more cohorts of their offspring (Gardner, Bull, Cooper, & Duffield, 2001), with related individuals benefiting from the resulting enhanced vigilance against predators (Lanham & Bull, 2004). The alternative ways in which genetic relatedness may influence social interactions among individuals may shape social network structure.

Increasingly, social networks are being used to explore the structure of social associations within populations and within aggregations (Krause, Croft, & James, 2007; Sih, Hanser, & McHugh,

* Correspondence: S. S. Godfrey, School of Veterinary and Life Sciences, Murdoch University, 90 South Street, Murdoch 6150, WA, Australia.
E-mail address: s.godfrey@murdoch.edu.au (S. S. Godfrey).

http://dx.doi.org/10.1016/j.anbehav.2014.08.019
0003-3472/

2009). They provide a framework for quantifying associations among individuals on a dyadic level, by representing a population as a series of nodes (representing individuals) connected by edges (representing associations) and are particularly useful for testing hypotheses about the factors influencing social structure (Wey, Blumstein, Shen, & Jordán, 2008). For instance, network analysis has shown consistent social associations among members of fission–fusion aggregations (Croft et al. 2012), with these associations sometimes stronger in one sex than the other (Carter, Brand, Carter, Shorrocks, & Goldizen, 2013; Farine, 2014; Stanley & Dunbar, 2013). However, there is conflicting evidence about whether these social associations are influenced by genetic relatedness (Lukas, Reynolds, Boesch, & Vigilant, 2005). On the one hand, several studies suggest that relatedness can influence social structure. For example, Wisniewski, Lusseau, and Moller (2010) showed that related female dolphins, *Tursiops aduncus*, form stable coalitions in the fission–fusion dynamics of pod formation, and Best, Seddon, Dwyer, and Goldizen (2013) found that social groupings of female kangaroos, *Macropus giganteus*, had higher relatedness than average for the population. Similarly, Chiyo et al. (2011) reported stronger associations among related than unrelated male elephants, *Loxodonta africana*, and Kurvers et al. (2013) found that foraging barnacle geese, *Branta leucopsis*, preferentially associated with related and familiar individuals. On the other hand, Croft et al. (2012) found no evidence that related individuals associated more strongly in shoals of wild guppies, *Poecilia reticulata*, although their results did not suggest avoidance of related individuals.

In subsocial or solitary species, patterns of association might be affected by relatedness in different ways, with fewer benefits from cooperative behaviours. Among these species, social network structure is primarily shaped by contacts during courtship and mating, by aggressive encounters to maintain territory boundaries (Lattanzio & Miles, 2014) or by contacts while foraging at a common source (Hamede, Bashford, McCallum, & Jones, 2009). For example, in solitary living woodchucks, *Marmota monax*, nonaggressive interactions were more frequent among pairs of individuals with higher genetic relatedness (Maher, 2009). Similarly, normally solitary mountain brush-tailed possums, *Trichosurus cunninghami*, preferred to share tree hollow dens with kin when tree hollows were limited (Banks et al. 2011). In contrast, Hirsch, Prange, Hauver, and Gehrt (2013) reported no influence of relatedness in social networks of solitary living racoons, *Procyon lotor*. Although solitary species are less likely to show cooperative behaviours, kin selection should normally favour higher tolerance and thus stronger associations between related individuals. But avoidance of kin competition should reduce associations among related individuals (or increase aggression) when resources are indivisible (e.g. Foster & Briffa, 2014). Selection to reduce the degree of inbreeding should also favour associations of less related individuals for mating activity. Thus, social structure should still be influenced by relatedness in predominantly solitary species. We tested this hypothesis by comparing social network associations among individuals of known genotype in a population of a largely solitary living Australian scincid lizard. We predicted that tolerance of kin would lead to stronger social associations among more related individuals of the same sex, whereas avoidance of inbreeding would lead to stronger associations among less related individuals of the opposite sex.

The Australian sleepy lizard, *Tiliqua rugosa*, is a large, long-lived, Australian scincid lizard that occupies stable, overlapping home ranges (Bull, 1994; Kerr & Bull, 2006). Although it has a largely solitary life, each spring adult lizards form monogamous pair bonds for up to 10 weeks before they mate, and individual pairs of lizards often re-establish these partnerships in subsequent years (Bull, 1988, 1994, 2000; Bull & Burzacott, 2006; Bull, Cooper, & Baghurst, 1998; Leu, Bashford, Kappeler, & Bull, 2010). The use of onboard activity and GPS loggers (Kerr, Bull, & Cottrell, 2004; Kerr, Bull, & Mackay, 2004; Leu et al. 2010) has allowed us to describe more cryptic and infrequent aspects of their social system beyond pair associations, which cannot be captured from snapshot observations.

Social networks based on synchronous spatial proximity among active lizards have shown that individuals associate with some neighbours and avoid others, and that this social structure remains stable both within a year and over multiple years (Godfrey, Sih, & Bull, 2013; Leu et al. 2010). Our current study builds upon this previous research by exploring the genetic relationships between adult lizards in a social network. We tested the hypothesis that social associations among lizards were influenced by relatedness by determining whether the strength of social connections among lizards in the social network were positively (or negatively) correlated with relatedness to each other, compared with whether there had been random associations. Given that individuals that live closer together will have more opportunities to interact than those living further apart, we used a null model to generate expected rates of interaction based on shared space use. By controlling for spatial proximity (using the null model), we could specifically ask whether there was a social influence of genetic relatedness on network structure.

METHODS

The study was conducted from October to December 2010, in a 1.0×1.5 km area of chenopod shrubland (33° 54′ S, 139° 20′ E), near Bundey Bore Station in the mid-north region of South Australia. The study period was during the austral spring and early summer, the time when these lizards are most active each year (Kerr & Bull, 2006; Kerr, Bottema, & Bull, 2008). All 60 adult lizards resident in the area (30 males, 30 females) were captured by hand in September 2010 and fitted with data loggers that were attached to the dorsal surface of the tail with surgical tape. Handling time was normally no longer than 30 min, and usually only 10–15 min. The 60 lizards were part of a larger continuous population inhabiting similar habitat surrounding the study area.

The data loggers recorded synchronous GPS locations for each lizard every 10 min when it was active (determined by a step-counter attached to the lizard), for the duration of the study (Kerr, Bull, & Cottrell, 2004; Kerr, Bull, & Mackay, 2004; Leu et al. 2010). GPS loggers were manufactured at Flinders University, Adelaide, Australia (Kerr, Bull, & Cottrell, 2004; Kerr, Bull, & Mackay, 2004). For our analyses, we considered all locations collected over the period 1 October–20 December 2010 (81 days), when the majority (ca. 90%) of lizards in the study area had data loggers attached. A radiotransmitter (Sirtrack, Havelock North, New Zealand) with unique frequency allowed us to identify, locate and hand-capture each lizard every 12 days to download data and to change batteries. Each data logger plus radio unit weighed 37 g, or 4.5% of the average body weight of an adult lizard, and 5.6% of the body weight of the lightest lizard in our study. Data downloads were conducted at times before or after the diurnal period of activity, to avoid interfering with normal behaviours and to reduce the impact of handling on lizard behaviour (Kerr, Bull, & Cottrell, 2004; Kerr, Bull, & Mackay, 2004). The lizards did not grow substantially during the season, and for any lizards that had noticeably gained (or lost) weight between captures (12 days), we completely refitted the logger. Lizards foraged and mated normally with the loggers on (S. Godfrey, personal observation), and maintained weight levels throughout the study comparable to other lizards without loggers attached in adjacent areas. At the end of the study, all lizards were recaptured and we removed the units and released

the lizards. We found no skin damage or irritation where the units had been attached and lizards naturally shed their skin in the following months. The lizards were treated using procedures formally approved by the Flinders University Animal Welfare Committee in compliance with the Australian Code of Practice for the Use of Animals for Scientific Purposes and conducted with a Permit to Undertake Scientific Research from the South Australian Department of the Environment, Water and Natural Resources.

Network Structure

We developed a social network from incidents of spatial proximity of dyads of active lizards. These were derived from synchronous GPS locations every 10 min for each of the 60 lizards. Following Leu et al. (2010), we considered that two lizards within 2 m of each other at any of the GPS location times had probably made, or would soon make, social contact. Allowing for a median GPS precision of 6 m, we included each pair of GPS-derived locations within 14 m of each other at the same time as a record of social contact. To construct the social network we calculated the simple ratio index (SRI) for each dyad, as the number of recorded contacts divided by the number of observations when both lizards were active. This was a measure of association strength, which on a biological level represents the amount of time two lizards spent together when they were active. This is an appropriate association measure because sleepy lizards have infrequent contact with other lizards, which in some cases (especially male–male contacts) can be brief. Our SRI measure does not represent either the frequency or the duration of interactions. Two equivalent SRI measures could be derived from two lizards either if they spent one long period together or if they spent several shorter periods together. Instead, the SRI measure represents the total relative time that lizards spent in proximity to each other across the entire activity season. The network consisted of 60 nodes, representing the 60 lizards. An edge was included for each pair of lizards that was recorded in contact at least once (SRI > 0) over the study period. Edge weight was determined by the SRI, with a higher weighting for pairs of lizards that spent more time in proximity. The network was nondirectional in that both lizards were assumed to have equal roles in a contact interaction. Although that might not have been the case, for example if a more aggressive lizard was more likely to initiate contacts, the data did not allow any inference of directionality.

Home Ranges and Expected Associations among Lizards

We included all GPS locations to derive, using Ranges 6 (Kenward, South, & Walls, 2003), the 95% minimum convex polygon home range, the home range centre for each lizard, the area of home range overlap and the distance between home range centres for each dyad. To account for the influence of spatial proximity on social associations, we developed an expected association network, assuming random encounters among lizards, using the ideal gas model (Hutchinson & Waser, 2007). This model estimated expected association rates based on individual lizards moving randomly within their home ranges. For each dyad, we calculated f, the expected encounter rate per day, using the formula derived from Leu et al. (2010):

$$f = \frac{8v(14)o}{\pi \mathrm{hr}_i \mathrm{hr}_j} \quad (1)$$

where v is the mean velocity of the two lizards (average distance (m) travelled/day), o is the area of home range overlap between the two lizards, and hr_i and hr_j are the home range areas of individual i and individual j, respectively. We used the estimated f as encounter rates to determine edge weights and to derive expected association networks. The distance travelled per day for each lizard was estimated by calculating the total number of steps taken in each day of the study, and multiplying it by the average stride length for a sleepy lizard (148 mm, Kerr, Bull, & Cottrell, 2004; Kerr, Bull, & Mackay, 2004), to derive the total distance travelled per day (m). This value was averaged across all days for each lizard to derive an 'average distance travelled per day' for each lizard, and, for each pairwise combination of lizards, we calculated an 'average velocity' (the mean of the average distance travelled for both lizards) which was used in equation (1) following Leu et al. (2010).

Microsatellite DNA Genotypes

We collected blood onto a 3 mm^2 area of an FTA card by clipping the tip of one toe of each lizard in the social network using a pair of sterilized, sharp, dog nail clippers. An analgesic (Meloxicam) was administered orally prior to toe clipping to reduce pain and discomfort. Lizards were gently restrained by hand during the procedure. About 30% of lizards flinched briefly during the procedure with limb movement, but became calm again within 1 min. We ensured bleeding had ceased before the lizard was released, and all lizards behaved normally upon release. The persons conducting the surgery had at least a full year's experience in conducting the procedure. We recaptured all lizards 12 days later, and in all cases the clipped area had healed and there were no signs of infection. We recaptured all toe-clipped lizards throughout the study, and the method had no observable impact on survival, movement or body condition, compared with other studies, or with other conspecifics we encountered in adjacent sites. Sleepy lizards do not use their claws for digging or climbing so toe clipping should not affect their ability to seek refuge. They are slow-moving reptiles so it should also not affect their locomotor performance to the detriment of the individual movement speed. Many lizards are found with natural toe loss, and with no obvious loss of body condition. In other studies of the same species, several hundred toe-clipped individuals, with several toe tips removed for individual recognition, have been recaptured over periods of up to 20 years (Bull & Burzacott, 2006) with no apparent loss of body condition compared with unmarked animals. Thus we consider there were no short- or long-term adverse effects of removing the tip of a single toe on the lizards in this study. Alternative methods of DNA collection are unreliable (caudal vein blood sampling), impractical (tail tipping) or untested (buccal swabs) in this species. In particular, caudal vein sampling can extend handling time because the vein is difficult to find in this species.

We extracted DNA from blood samples on 3 mm^2 squares of the FTA cards following the Whatman FTA Elute card procedure (GE Healthcare, Buckinghamshire, U.K.). We then used the procedures described by Gardner, Delphs, Smith, Dudaniec, and Sanchez (2008) to determine lizard genotypes at 15 microsatellite DNA loci (Trl1, Trl3, Trl9, Trl10, Trl12, Trl14, Trl16, Trl19, Trl21, Trl22, Trl27, Trl30, Trl32, Trl36 and Trl37). Genotypes were successfully scored for 46 lizards (26 males and 20 females) using GENEMAPPER v4.0 (Applied Biosystems, Inc., Foster City, CA, U.S.A.) and were checked manually.

We used the program Coancestry (Wang, 2011) to calculate coefficients of relatedness (r) between pairs of individuals with a moments estimator that assumes there is no inbreeding (Wang, 2002). Allele frequencies used in the calculations were simulated from all genotyped individuals in the sample. Pairwise coefficients of relatedness were used in our comparisons of relatedness detailed in the following section, and to calculate mean relatedness in comparisons of different dyadic groups.

Analyses of Social, Spatial and Genetic Relationships

Because our data consisted of GPS tracks, we derived a novel randomization procedure that is a combination of data-stream and node-based permutation techniques. This procedure maintained biologically valid GPS tracks for each individual while reducing the potential for type II error associated with node permutations (see Farine, 2014). The method randomizes the identity of each daily track, assigning new identities to entire tracks within each day, but assigning different identities on different days. Randomizations were repeated 1000 times, and the *P* values were calculated by comparing the observed test statistic to the statistic calculated for each permutation.

We first asked how relatedness of the 46 genotyped lizards in our study population varied in respect to sex (to detect signals of sex-biased dispersal) and space, to understand the factors influencing general patterns of relatedness in our study population. We estimated mean relatedness values separately for all possible male–male, female–female and male–female combinations in the population. Additionally, we calculated mean relatedness among neighbouring (home range centres within 200 m) and non-neighbouring (home range centres more than 200 m apart) lizards to determine whether there was an effect of distance on relatedness among lizards. We then repeated these analyses for pairs of lizards that were connected in the social network to determine what factors influenced relatedness among connected lizards (SRI > 0).

We then structured our analyses into three components. First we asked, across the entire study social network, what most influenced social association strength among connected lizards: genetic relationships or spatial relationships. The analysis we used, in this and subsequent components, was the multiple regression quadratic assignment procedure (MRQAP) following Dekker, Krackhardt, and Snijders (2007). The dependent matrix in this procedure contained the observed association strengths (SRI values) in the social network. The two predictor matrices came from the expected association rates among dyads in the ideal gas model, allowing us to explore the influence of spatial arrangements in the population, and from the genetic relatedness values among dyads, allowing us to explore how genetic relationships influence network structure.

Second, because lizards further apart were less likely to encounter each other and form social contacts, we restricted our analysis to lizards that had home range centres within 200 m of each other (that is, analysing a subset of the data set used in the first analysis). We called these lizards neighbours, because 200 m is within the distance across a normal home range for this species (Bull & Freake, 1999; Kerr & Bull, 2006), and asked whether genetic relationships or spatial relationships influenced association strength among all neighbouring lizards. Third, we conducted similar analyses separately for three subgroups of neighbouring lizards: male–male, female–female and male–female dyads. Within neighbouring male–female dyads, we also performed separate analyses for dyads we had previously defined as paired (those with an SRI > 0.1) and for dyads we had previously defined as having formed an extrapair association (0 < SRI < 0.1) (Godfrey, Bradley, Sih, & Bull, 2012; Leu et al. 2010), allowing us to distinguish between strong pair bonds and weaker links among males and females. These analyses were performed to address our hypothesis about how relatedness influences social associations, but with particular emphasis on differences between different subgroups and different behavioural interactions in the population, such as intrasexual associations and male–female partnerships. We performed all the network analyses, including permutations and MRQAP tests, using the package asnipe (Farine, 2013) in R (R Core Team, 2013).

RESULTS

General Patterns of Relatedness in the Population

Mean relatedness differed significantly between different dyadic combinations, with male–male dyads more related than male–female dyads, and with female–female dyads the least related (Fig. 1). Relatedness did not differ significantly between neighbouring lizards and non-neighbours (mean difference = 0.006, 95% confidence interval, CI = −0.011 to 0.011, $P = 0.118$).

All Connected Dyads in the Social Network

Mean values of network edge weight (SRI), distance between home range centres, percentage home range overlap and relatedness (R) among the connected dyads of the 46 genotyped lizards (those with SRI > 0) are shown in Table 1. Connected male–male dyads were more related, and both male–female and female–female connected dyads were less related than expected from the randomized data ($P < 0.001$ for all cases; Fig. 1). Among connected dyads, those that were not neighbours had a higher mean relatedness (mean $R = 0.09 \pm 0.018$ SE) than those that were neighbours (mean $R = 0.049 \pm 0.006$) resulting in a significant difference between neighbouring and non-neighbouring dyads (mean difference = 0.041, 95% CI = −0.024 to 0.026, $P = 0.001$).

The MRQAP analysis showed a significant effect of spatial pattern, but not of genetic relatedness, on the observed SRI of all connected dyads (Table 2). Observed SRI was positively related to expected association strength (which assumed lizards moved randomly), but values were lower than predicted (Fig. 2a).

Among Neighbouring Dyads (<200 m Apart)

Neighbouring lizards were more likely to be connected in the observed social network than expected by chance (64% of all dyads

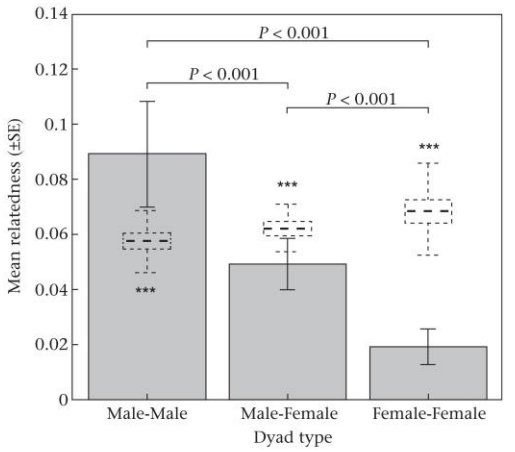


Figure 1. Mean genetic relatedness among male–male dyads, male–female dyads and female–female dyads (filled bars and solid error bars), for all genotyped adult lizards in our study area. *P* values correspond to comparisons of means between different dyadic combinations. Hollow bars with dashed error bars represent the expected relatedness within each type of dyad based on 1000 permutations of the data stream. Asterisks indicate a significant difference between the observed patterns of relatedness of different dyads and the population-level patterns of relatedness.

Table 1
Summary of mean values of network edge weight, distance between home range centres, home range overlap and relatedness among dyads of lizards connected in the social network

	N	Network edge weight (SRI)		Distance between home range centres (m)		Home range overlap (%)		Relatedness (*R*)	
		Mean (SE)	Range	Mean (SE)	Range	Mean (SE)	Range	Mean (SE)	Range
All dyads	303	0.023 (0.005)	0.0006–0.495	149 (5.9)	0–385.1	19.2 (1.25)	0–83.2	0.059 (0.007)	0–0.661
Neighbouring dyads									
All dyads	194	0.029 (0.007)	0.0006–0.495	111.9 (4)	0–198.8	23.4 (1.5)	0–83.2	0.049 (0.006)	0–0.618
Male–male dyads	69	0.005 (0.0008)	0.0006–0.024	121.6 (6.8)	7.1–195.8	18.9 (2.7)	0.7–56.5	0.079 (0.013)	0–0.618
Female–female dyads	30	0.007 (0.002)	0.0008–0.063	110.0 (8.7)	5.8–182.5	19.7 (2.5)	0–61.9	0.015 (0.008)	0–0.199
Male–female dyads	95	0.048 (0.012)	0.0008–0.495	106.5 (5.9)	0–198.8	27.1 (2.3)	0–83.2	0.045 (0.007)	0–0.247

with SRI > 0 were between neighbours; 25% of all possible dyads were between neighbours; randomization test: $P < 0.001$). Among neighbours, observed SRI was also lower than expected by the ideal gas model, but SRI was positively correlated with the expected association rates (Fig. 2b, Table 2). Close proximity did not necessarily mean high SRI, as 91% of dyadic edges between neighbouring lizards (<200 m apart) had an SRI below 0.1, and 13.7% of neighbouring dyads were never observed to associate (SRI = 0). In these analyses, we consistently found no effect of genetic relatedness on observed SRI (Table 2).

Neighbouring Lizards of the Same Sex

Among genotyped neighbouring lizards, there were 69 male–male dyads and 30 female–female dyads (Table 1). The mean SRIs of the two forms of same-sex dyads were not significantly different (Table 1, mean difference = 0.002, $t = -0.1735$, 95% CI = −0.025 to 0.0207, $P = 0.863$). Nor were there spatial differences between these two dyadic groups. For instance, mean home range overlap was similar between neighbouring male–male dyads and female–female dyads (Table 1, mean difference = 0.031, 95% CI = 0–0.055, $P = 0.219$). However, neighbouring males were significantly more related to each other than were neighbouring females (Fig. 1; mean difference = 0.062, $t = 2.999$, 95% CI = 0.021–0.103, $P = 0.004$).

The MRQAP analyses showed that within each of these groups observed SRI values were positively related to (but lower than) expected association rates (Table 3, Fig. 3a, b). Genetic relatedness had a significant negative effect on male–male SRI, but no significant effect on female–female SRI (Table 3). Males were more strongly associated with less related neighbouring males in the observed social network (Fig. 3c).

Neighbouring Lizards of the Opposite Sex

Among genotyped neighbouring lizards, there were 95 male–female dyads (Table 1). Neighbouring male–female dyads had significantly greater observed SRI (mean SRI difference = 0.018, $t = 4.0134$, 95% CI = 0.009–0.026, $P < 0.001$), and had a significantly higher percentage of home range overlap (mean overlap difference = 18.4%, 95% CI = 14.2–18.1%, $P = 0.012$) than for other neighbouring dyad types (male–male and female–female dyads). However, relatedness among neighbouring male–female dyads was not significantly different from the mean for other neighbouring dyad types (mean difference = 0.002, $t = 0.178$, 95% CI = −0.015 to 0.018, $P = 0.858$).

In the MRQAP analysis, SRI was positively correlated with expected association rates, both overall and separately for pairs

Table 2
Results of MRQAP analyses of the effects of spatial relationships (expected associations) and genetic relatedness on observed association strength in the social network among all connected lizards, and among those within 200 m of each other

	All lizards		Within 200 m	
	Regression coefficient	*P*	Regression coefficient	*P*
Expected associations	0.276	**<0.001**	0.269	**<0.001**
Relatedness	−0.024	0.263	−0.026	0.216

Significant *P* values are shown in bold.

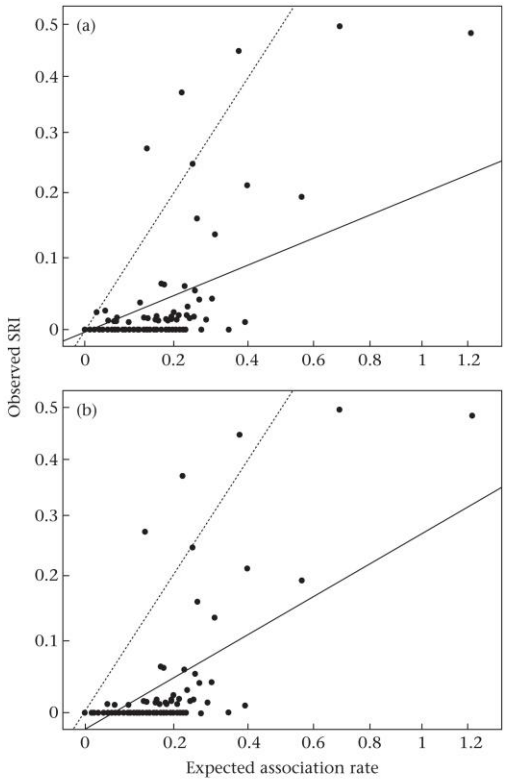


Figure 2. Relationships between the observed social association strength for each connected dyad in the network (SRI) and expected association rates (calculated using the ideal gas model) showing separately (a) all possible dyads among all lizards and (b) all possible dyads among neighbouring lizards (<200 m apart). Solid lines represent the regression between expected association rate and observed SRI, and the dotted lines represent the equivalence line (x = y). Note that both axes are represented on a log (10) scale for visualization purposes.

Table 3
Results of MRQAP analyses of the effects of spatial relationships (expected associations) and relatedness on observed association strength in the social network among males and among females within 200 m of each other

	Among males		Among females	
	Regression coefficient	*P*	Regression coefficient	*P*
Expected associations	0.033	**<0.001**	0.050	**<0.001**
Relatedness	−0.002	**0.025**	0.000	0.954

Significant *P* values are shown in bold.

(SRI > 0.1; Fig. 4a) and for extrapair (SRI < 0.1; Fig. 4b) associations (Table 4). In each case, the slope of the regression line showed lower SRIs than the expected association rates from the random gas model, although three male–female pairs associated more strongly than expected by chance (Fig. 4a, b). Genetic relatedness also had a significant effect on SRI among pairs (Fig. 4c) and among extrapair associations (Fig. 4d, Table 4). In each case, less related individuals had higher SRI (Fig. 4c, d).

DISCUSSION

We noted three major results from our study population. The first concerned genetic structure within the population. Among the adult lizards in our social network, there were generally low levels of relatedness, and only a small number of dyads had relatedness values that exceeded 0.25, a level that would indicate close familial relatives. This implies that the social structure among adult lizards was not based on associations of close kin. Rather, the results suggest that some dispersal to separate related individuals is the normal strategy in this species. Furthermore, both at the level of the whole study population and at the level of neighbouring lizards, female–female dyads were significantly less related to each other than were male–male dyads (Fig. 1). Although both sexes probably disperse, since no full sib dyads were detected, an implication is that females disperse further from their natal sites and from related females than do males. Sex-biased dispersal has been widely reported among many animal groups, with a common explanation that inbreeding is avoided if one sex disperses more than the other (Greenwood, 1980; Pusey, 1987). Our results confirm that most neighbouring males and females that are potential mating partners are only distantly related.

Our second result was the strongly significant influence of expected association rates on observed association strength (SRI) of connected dyads. Our expected association rates were generated from an ideal gas model which estimates encounter rates based on the extent of home range overlap between members of each dyad, and which assumes that lizards moved randomly within their home ranges. Thus, this suggests that a component of lizard social interactions can be explained by their spatial ecology. Lizards living closer to each other are more likely to interact. However, even among neighbouring lizards with home range centres less than 200 m apart, or with overlapping home ranges, a proportion of dyads showed very low levels of social association. This observation confirms previous analyses from this species showing social structure is characterized by individuals apparently deliberately avoiding contact with some close neighbours (Godfrey et al. 2013; Leu et al. 2010). In our current analyses, observed SRI was consistently lower than the expected association rates, suggesting that, among all connected dyads, lizards (even some pairs) spend less time in proximity than expected if they were moving randomly within their home ranges. These results indicate that lizards do not contact each other randomly. There is a specific structure to the social network that is based on avoiding some close neighbours, even though the strengths of the social associations within the network, the SRI values, are strongly influenced by spatial overlap.

Our third result, and the result that directly addressed the questions we asked in this study, was that genetic relatedness influenced the strength of social associations among neighbouring male–male dyads and male–female dyads. But, contrary to the kin selection hypothesis which predicts stronger associations among

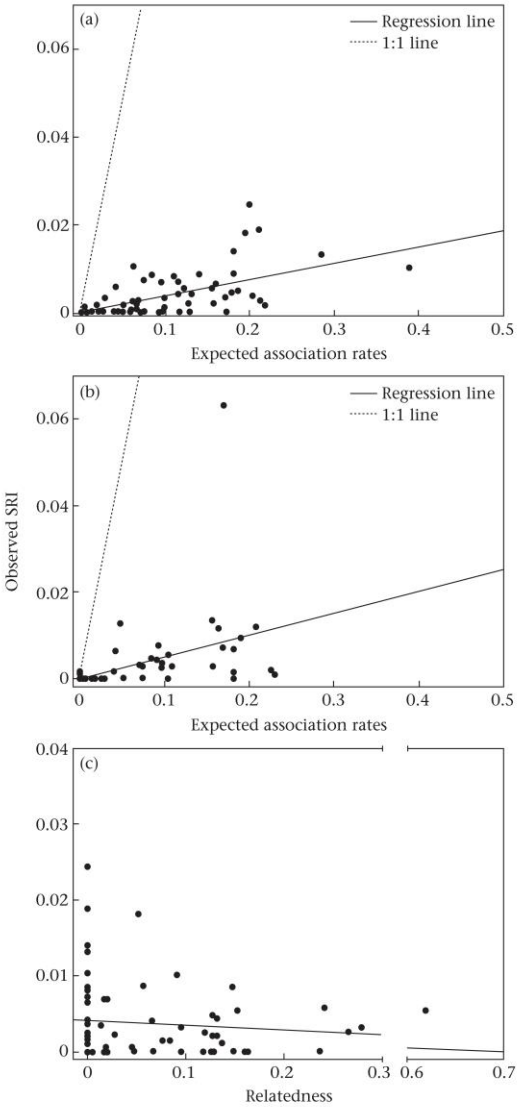


Figure 3. Relationships between social association strength in the network (SRI) and expected association rates for (a) neighbouring male–male dyads and (b) neighbouring female–female dyads, and (c) relatedness among neighbouring male–male dyads. The solid lines represent the regression line and the dashed lines represent the equivalence line (x = y).

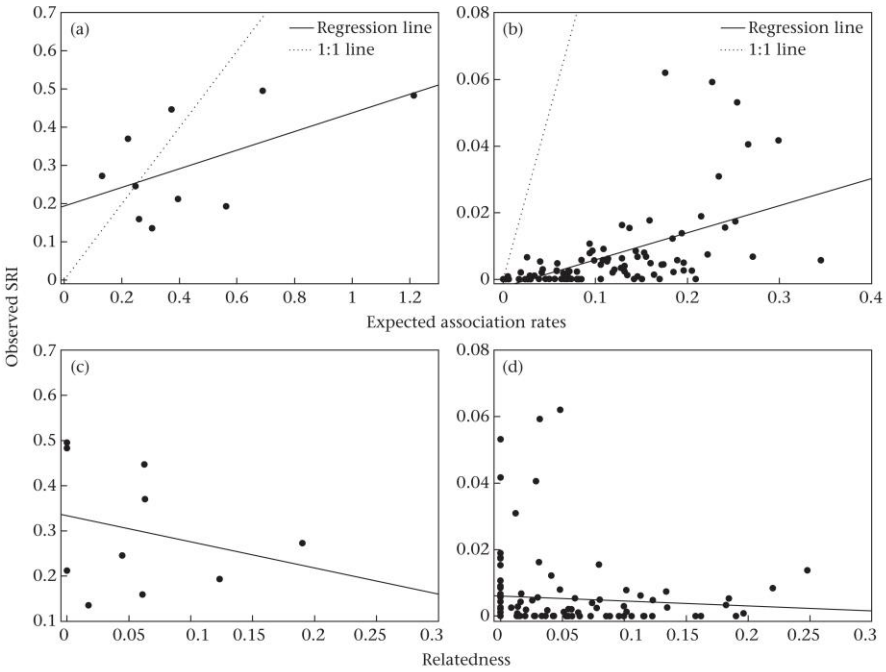


Figure 4. Relationships between social association strength in the network (SRI) and expected association rates for (a) pairs (SRI > 0.1) and (b) extrapair associations (SRI < 0.1), and social association strength (SRI) and relatedness for (c) pairs and (d) extrapair associations. Solid lines represent the regression and dashed lines represent the equivalence line (x = y).

Table 4
Results of MRQAP analyses of the effects of spatial relationships (expected associations) and relatedness on observed association strength in the social network among males and females, and separately for pair associations and extrapair associations, for dyads within 200 m of each other

	Males and females		Among pairs		Extrapair associations	
	Regression coefficient	*P*	Regression coefficient	*P*	Regression coefficient	*P*
Expected associations	0.269	**<0.001**	0.033	**<0.001**	0.050	**<0.001**
Relatedness	−0.032	0.272	−0.022	**0.031**	−0.041	**0.011**

Significant *P* values are shown in bold.

more related individuals, in our study the strongest associations in each case were among the least related individuals. Even with the low level of genetic relatedness that we recorded in our study population, close neighbours were more likely to associate if they were less related to each other, and the neighbours that were avoided were the ones that were genetically more related.

We considered four possible explanations for this pattern. One is that the absence of association among more related individuals reflects the lack of any general cooperative behaviour in this species. We have observed neither cooperative foraging nor collaboration in defending resources in this species, so there would be little opportunity for kin selection to favour associations of closer relatives, as reported in other species that form social aggregations (e.g. Krakauer, 2005). And while kin selection may favour higher tolerance of related lizard individuals, for instance by a greater level of overlap of home ranges, this would not necessarily result in more social contacts.

A second explanation concerns parasite transmission. We have already shown for this species that gut bacteria are transmitted along social network connections rather than among spatially adjacent individuals (Bull, Godfrey, & Gordon, 2012). Similarly, ectoparasitic ticks are transmitted along network pathways (Leu et al. 2010; Wohlfiel, Leu, Godfrey, & Bull, 2013). Thus more socially connected lizards are more likely to transmit parasitic infections among themselves. Other studies in other species have demonstrated that higher genetic variability, particularly at MHC (major histocompatibility complex) loci, confers higher resistance to pathogens (Bonneaud, Chastel, Federici, Westerdahl, & Sorci, 2006; Coltman, Pilkington, Smith, & Pemberton, 1999; Penn, 2002). If infection is influenced by host resistance genotype, then transmission from one host to another is likely to be more successful if host genotypes are similar (Shykoff & Schmid-Hempel, 1991). Thus to reduce the risk of infection from parasites that are transmitted along network pathways, it would be advantageous to prefer social contacts with more distantly related individuals.

The third explanation comes from our analyses of male–male dyads. Males often interact aggressively with each other (Godfrey et al. 2012; Kerr & Bull, 2002; Murray & Bull, 2004), so that the

social contacts we detected between dyads of males may have been primarily agonistic. Bull (1990) previously suggested that younger males may fight for home range positions and access to females. The inclusive fitness of an individual male may be increased by directing aggressive encounters, and thus stronger social associations, towards less related males. This result for males may be further enhanced by the generally higher levels of relatedness among males than females in sleepy lizards, meaning that differentiating between related and unrelated individuals may have more impact on inclusive fitness for males than for females.

In a fourth explanation, social contact among male–female dyads may be predominantly related to courtship, with a sustained association between monogamous partners that extends over some weeks before mating (Bull, 2000). Even with the low levels of relatedness among individuals in the study population, sleepy lizards appeared to spend the most time with less related potential partners, a result that confirms previous analyses of inbreeding avoidance in these lizards (Bull & Cooper, 1999).

Independent of the mechanisms that might explain why there are stronger social interactions among less genetically related lizards, the results suggest a remarkable ability in this species for individual lizards to detect small differences in the degree of relatedness. We have previously suggested that olfactory signals are used by scincid lizards to differentiate between familiar and unfamiliar adult (Bull, Griffin, & Johnston, 1999; Bull, Griffin, Lanham, & Johnston, 2000) or neonate conspecifics (Main & Bull, 1996) and between siblings and nonsiblings (Bull, Griffin, Bonnett, Gardner, & Cooper, 2001). The current result extends these conclusions to suggest that differentiation of the degree of relatedness even among distantly related conspecific lizards can form the basis of a social structure. Wolf and Trillmich (2008) reached a similar conclusion in their study of Galapagos sea lions, *Arctocephalus galapagoensis*, in which individuals associated more strongly with genetically more similar conspecifics, even with low levels of relatedness among neighbours in a colony.

In sleepy lizards, other analyses have shown that this social network structure remains stable across time and across a range of ecological conditions (Godfrey et al., 2013). Here we have shown that social structures in this lizard population are not random with respect to genetic relatedness, but are based largely on avoidance of genetic relatives, particularly between males and between males and females. Avoidance occurs both through dispersal, so that near relatives are separated in space, and by behaviourally avoiding contact among spatially close relatives. Social organization in sleepy lizards is developed around each individual occupying both a social and a spatial area of low relatedness. We suggest that this might be a common form of social structure in more solitary species in which cooperative behaviours are infrequent and selection favours directing attention away from more related individuals, for instance in acts of aggression or mating. This represents an alternative social structure to the more commonly reported kin associations in species in which individuals tend to aggregate.

Acknowledgments

Our sleepy lizard research was funded by the Australian Research Council and the Holsworth Wildlife Research Endowment. D.R.F. was funded by grants from BBSRC (BB/L006081/1) to Ben C. Sheldon and the NSF (NSF-IOS1250895) to Margaret C. Crofoot. We thank Ron and Leona Clark, and Chris Mosey for allowing us access to their land, and the use of the homestead at Bundey Bore Station. We thank Jana Bradley, Dale Burzacott, Emilie Chavel and Caroline Wohlfeil for assistance with field work and Alecia Carter for constructive comments on the manuscript. S.S.G. conducted the field data collection, data processing and analysis and wrote the manuscript. M.T.A. and M.G.G. conducted the molecular analyses and derived the relatedness estimates. D.R.F. suggested new permutation methods for our analyses, developed script to load our data into asnipe and adjusted script to rerun our analyses using data-stream permutation methods. C.M.B. contributed to the design of the project, acquired the funding to run it and provided advice on the execution of the project, data analyses and the structure of the manuscript. All authors contributed towards writing and editing the final manuscript.

References

Alexander, R. D. (1974). The evolution of social behaviour. *Annual Review of Ecology and Systematics, 5*, 325–383.
Banks, S. C., Lindenmayer, D. B., McBurney, L., Blair, D., Knight, E. J., & Blyton, M. D. J. (2011). Kin selection in den sharing develops under limited availability of tree hollows for a forest marsupial. *Proceedings of the Royal Society B: Biological Sciences, 278*, 2768–2776.
Best, E. C., Seddon, J. M., Dwyer, R. G., & Goldizen, A. W. (2013). Social preference influences female community structure in a population of wild eastern grey kangaroos. *Animal Behaviour, 86*, 1031–1040.
Bonneaud, C., Chastel, O., Federici, P., Westerdahl, H., & Sorci, G. (2006). Complex Mhc-based mate choice in a wild passerine. *Proceedings of the Royal Society B: Biological Sciences, 273*, 1111–1116.
Bull, C. M. (1988). Mate fidelity in an Australian lizard *Trachydosaurus rugosus*. *Behavioral Ecology and Sociobiology, 23*, 45–49.
Bull, C. M. (1990). Comparisons of displaced and retained partners in a monogamous lizard. *Australian Wildlife Research, 17*, 135–140.
Bull, C. M. (1994). Population dynamics and pair fidelity in sleepy lizards. In L. J. Vitt, & E. R. Pianka (Eds.), *Lizard ecology: Historical and experimental perspectives* (pp. 159–174). Princeton, NJ: Princeton University Press.
Bull, C. M. (2000). Monogamy in lizards. *Behavioural Processes, 51*, 7–20.
Bull, C. M., & Burzacott, D. A. (2006). The influence of parasites on the retention of long-term partnerships in the Australian sleepy lizard, *Tiliqua rugosa*. *Oecologia, 146*, 675–680.
Bull, C. M., & Cooper, S. J. B. (1999). Relatedness and avoidance of inbreeding in the lizard *Tiliqua rugosa*. *Behavioral Ecology and Sociobiology, 46*, 367–372.
Bull, C. M., Cooper, S. J. B., & Baghurst, B. C. (1998). Social monogamy and extra-pair fertilization in an Australian lizard, *Tiliqua rugosa*. *Behavioral Ecology and Sociobiology, 44*, 63–72.
Bull, C. M., & Freake, M. J. (1999). Home range fidelity in the Australian sleepy lizard, *Tiliqua rugosa*. *Australian Journal of Zoology, 47*, 125–132.
Bull, C. M., Godfrey, S. S., & Gordon, D. M. (2012). Social networks and the spread of *Salmonella* in a sleepy lizard population. *Molecular Ecology, 21*, 4386–4392.
Bull, C. M., Griffin, C. L., Bonnett, M., Gardner, M. G., & Cooper, S. J. B. (2001). Discrimination between related and unrelated individuals in the Australian lizard *Egernia striolata*. *Behavioral Ecology and Sociobiology, 50*, 173–179.
Bull, C. M., Griffin, C. L., & Johnston, G. R. (1999). Olfactory discrimination in scat-piling lizards. *Behavioral Ecology, 10*, 136–140.
Bull, C. M., Griffin, C. L., Lanham, E. J., & Johnston, G. R. (2000). Recognition of pheromones from group members in a gregarious lizard, *Egernia stokesii*. *Journal of Herpetology, 34*, 92–99.
Carter, K. D., Brand, R., Carter, J. K., Shorrocks, B., & Goldizen, A. W. (2013). Social networks, long-term associations and age-related sociability of wild giraffes. *Animal Behaviour, 86*, 901–910.
Chiyo, P. I., Archie, E. A., Hollister-Smith, J. A., Lee, P. C., Poole, J. H., Moss, C. J., & Alberts, S. C. (2011). Association patterns of African elephants in all-male groups: the role of age and genetic relatedness. *Animal Behaviour, 81*, 1093–1099.
Coltman, D. W., Pilkington, J. G., Smith, J. A., & Pemberton, J. M. (1999). Parasite mediated selection against inbred Soay sheep in a free-living, island population. *Evolution, 53*, 1259–1267.
Croft, D. P., Hamilton, P. B., Darden, S. K., Jacoby, D. M. P., James, R., Bettaney, E. M., & Tyler, C. R. (2012). The role of relatedness in structuring the social network of a wild guppy population. *Oecologia, 170*, 955–963.
Dekker, D., Krackhardt, D., & Snijders, T. A. B. (2007). Sensitivity of MRQAP tests to collinearity and autocorrelation conditions. *Psychometrika, 72*, 563–581.
Farine, D. R. (2013). Animal social network inference and permutations for ecologists in R using *asnipe*. *Methods in Ecology and Evolution, 4*, 1187–1194.
Farine, D. R. (2014). Measuring phenotypic assortment in animal social networks: weighted associations are more robust than binary edges. *Animal Behaviour, 89*, 141–153.
Foster, N. L., & Briffa, M. (2014). Familial strife on the seashore: aggression increases with relatedness in the sea anemone *Actinia equina*. *Behavioural Processes, 103*, 243–245.
Gardner, M. G., Bull, C. M., Cooper, S. J. B., & Duffield, G. A. (2001). Genetic evidence for a family structure in stable social aggregations of the Australian lizard *Egernia stokesii*. *Molecular Ecology, 10*, 175–183.
Gardner, M. G., Delphs, L., Smith, A., Dudaniec, R., & Sanchez, J. (2008). *Tiliqua rugosa* microsatellites: isolation via enrichment and characterisation of loci for

multiplex PCR in *T. rugosa* and the endangered *T. adelaidensis*. *Conservation Genetics, 9*, 233–237.
Godfrey, S. S., Bradley, J. K., Sih, A., & Bull, C. M. (2012). Lovers and fighters in sleepy lizard land: where do aggressive males fit in a social network? *Animal Behaviour, 83*, 209–215.
Godfrey, S. S., Sih, A., & Bull, C. M. (2013). The response of a sleepy lizard social network to altered ecological conditions. *Animal Behaviour, 86*, 763–772.
Greenwood, P. (1980). Mating systems, philopatry and dispersal in birds and mammals. *Animal Behaviour, 28*, 1140–1162.
Hamede, R. K., Bashford, J., McCallum, H., & Jones, M. (2009). Contact networks in a wild Tasmanian devil (*Sarcophilus harrisii*) population: using social network analysis to reveal seasonal variability in social behaviour and its implications for transmission of devil facial tumour disease. *Ecology Letters, 12*, 1147–1157.
Hirsch, B. T., Prange, S., Hauver, S. A., & Gehrt, S. D. (2013). Genetic relatedness does not predict racoon social network structure. *Animal Behaviour, 85*, 463–470.
Hutchinson, J. M. C., & Waser, P. M. (2007). Use, misuse and extensions of 'ideal gas' models of animal encounter. *Biological Reviews, 82*, 335–359.
Kenward, R. E., South, A. B., & Walls, S. S. (2003). *Ranges 6*. Wareham, U.K.: Anatrack Ltd.
Kerr, G. D., Bottema, M. J., & Bull, C. M. (2008). Lizards with rhythm? Multi-day patterns in total daily movement. *Journal of Zoology, 275*, 79–88.
Kerr, G. D., & Bull, C. M. (2002). Field observations of aggressive encounters between male sleepy lizards (*Tiliqua rugosa*). *Herpetological Review, 33*, 24–25.
Kerr, G. D., & Bull, C. M. (2006). Exclusive core areas in overlapping ranges of the sleepy lizard, *Tiliqua rugosa*. *Behavioral Ecology, 17*, 380–391.
Kerr, G. D., Bull, C. M., & Cottrell, G. R. (2004). Use of an 'on board' datalogger to determine lizard activity patterns, body temperature and microhabitat use for extended periods in the field. *Wildlife Research, 31*, 171–176.
Kerr, G. D., Bull, C. M., & Mackay, D. (2004). Human disturbance and stride frequency in the sleepy lizard (*Tiliqua rugosa*): implications for behavioral studies. *Journal of Herpetology, 38*, 519–526.
Krakauer, A. H. (2005). Kin selection and cooperative courtship in wild turkeys. *Nature, 434*, 69–72.
Krause, J., Croft, D. P., & James, R. (2007). Social network theory in the behavioural sciences: potential applications. *Behavioral Ecology and Sociobiology, 62*, 15–27.
Kurvers, R. H. J. M., Adamczyk, V. M. A. P., Kraus, R. H. S., Hoffman, J. I., van Wieren, S. E., van der Jeugd, H. P., et al. (2013). Contrasting context dependence of familiarity and kinship in animal social networks. *Animal Behaviour, 86*, 993–1001.
Lanham, E. J., & Bull, C. M. (2004). Enhanced vigilance in groups in Egernia stokesii, a lizard with stable social aggregations. *Journal of Zoology, 263*, 95–99.
Lattanzio, M. S., & Miles, D. B. (2014). Ecological divergence among colour morphs mediated by changes in spatial network structure associated with disturbance. *Journal of Animal Ecology*. http://dx.doi.org/10.1111/1365-2656.12252.
Leu, S. T., Bashford, J., Kappeler, P. M., & Bull, C. M. (2010). Association networks reveal social organization in the sleepy lizard. *Animal Behaviour, 79*, 217–225.
Linksvayer, T. A. (2010). Subsociality and the evolution of eusociality. In M. D. Breed, & J. Moore (Eds.), *Encyclopedia of animal behavior* (Vol. 3, pp. 358–362). Oxford, U.K.: Academic Press.
Lukas, D., Reynolds, V., Boesch, C., & Vigilant, L. (2005). To what extent does living in a group mean living with kin? *Molecular Ecology, 14*, 2181–2196.
Maher, C. R. (2009). Effects of relatedness on social interaction rates in a solitary marmot. *Animal Behaviour, 78*, 925–933.
Main, A. R., & Bull, C. M. (1996). Mother-offspring recognition in two Australian lizards, *Tiliqua rugosa* and *Egernia stokesii*. *Animal Behaviour, 52*, 193–200.
Michener, C. D. (1969). Comparative social behaviour of bees. *Annual Review of Entomology, 14*, 299–342.
Murray, K., & Bull, C. M. (2004). Aggressiveness during monogamous pairing in the sleepy lizard, *Tiliqua rugosa*: a test of the mate guarding hypothesis. *Acta Ethologica, 7*, 19–27.
Penn, D. J. (2002). The scent of genetic compatibility: sexual selection and the major histocompatibility complex. *Ethology, 108*, 1–21.
Pusey, A. E. (1987). Sex-biased dispersal and inbreeding avoidance in birds and mammals. *Trends in Ecology & Evolution, 2*, 295–299.
Pusey, A., & Wolf, M. (1996). Inbreeding avoidance in animals. *Trends in Ecology & Evolution, 11*, 201–206.
R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
Shykoff, J. A., & Schmid-Hempel, P. (1991). Parasites and the advantage of genetic variability within social insect colonies. *Proceedings of the Royal Society B: Biological Sciences, 243*, 55–58.
Sih, A., Hanser, S. F., & McHugh, K. A. (2009). Social network theory: new insights and issues for behavioral ecologists. *Behavioral Ecology and Sociobiology, 63*, 975–988.
Stanley, C. R., & Dunbar, R. I. M. (2013). Consistent social structure and optimal clique size revealed by social network analysis of feral goats, *Capra hircus*. *Animal Behaviour, 85*, 771–779.
Wang, J. (2002). An estimator for pairwise relatedness using molecular markers. *Genetics, 160*, 1203–1215.
Wang, J. (2011). COANCESTRY: a program for simulating, estimating and analysing relatedness and inbreeding coefficients. *Molecular Ecology Resources, 11*, 141–145.
Wilson, E. O. (1975). *Sociobiogy: The new synthesis*. Cambridge, MA: Harvard University Press.
Wey, T., Blumstein, D. T., Shen, W., & Jordán, F. (2008). Social network analysis of animal behaviour: a promising tool for the study of sociality. *Animal Behaviour, 75*, 333–344.
Wisniewski, J., Lusseau, D., & Moller, L. M. (2010). Female bisexual kinship ties maintain social cohesion in a dolphin network. *Animal Behaviour, 80*, 895–904.
Wohlfiel, C. K., Leu, S. T., Godfrey, S. S., & Bull, C. M. (2013). Testing the robustness of transmission network models to predict ectoparasite loads. One lizard, two ticks and four years. *International Journal for Parasitology: Parasites and Wildlife, 2*, 271–277.
Wolf, J. B. W., & Trillmich, F. (2008). Kin in space: social viscosity in a spatially and genetically substructured network. *Proceedings of the Royal Society B: Biological Sciences, 275*, 2063–2069.

9.2. APPENDIX II

CSIRO PUBLISHING
Australian Journal of Zoology
http://dx.doi.org/10.1071/ZO12063

A direct benefit of native saltbush revegetation for an endemic lizard (*Tiliqua rugosa*) in southern Australia

Melanie L. Lancaster[A,B], *Michael G. Gardner*[A,B,C,G], *Alison J. Fitch*[B,C], *Talat H. Ansari*[B,C] *and Anita K. Smyth*[D,E,F]

[A]School of Earth and Environmental Sciences, The University of Adelaide, Adelaide, SA 5005, Australia.
[B]South Australian Regional Facility for Molecular Ecology and Evolution, South Australian Museum, Adelaide, SA 5000, Australia.
[C]School of Biological Sciences, Flinders University, Adelaide, SA 5001, Australia.
[D]CSIRO, PMB 2, Glen Osmond, SA 5064, Australia.
[E]Terrestrial Ecosystem Research Network – Adelaide Node, The University of Adelaide, Adelaide, SA 5005, Australia.
[F]Future Farm Industries Cooperative Research Centre, The University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia.
[G]Corresponding author. Email: michael.gardner@flinders.edu.au

Abstract. Land alteration for intensive agriculture has been a major cause of species decline and extinction globally. In marginal grazing regions of southern Australia, native perennial shrubs are increasingly being planted to supplement pasture feeding of stock. Such revegetation has the benefits of reducing erosion and salinity, and importantly, the potential provision of habitat for native fauna. We explored the use of revegetated native saltbush by the sleepy lizard (*Tiliqua rugosa*) an endemic Australian species common in the region. We repeatedly sampled revegetated saltbush throughout 2010 and 2011 for adults ($n = 55$) and juveniles ($n = 26$). Using genotypes from eight microsatellite loci, parents were assigned to half of all juveniles with high statistical confidence. Parents were sampled in the same patch of revegetated saltbush as their offspring, thus supporting the observation that juvenile sleepy lizards remain within the home range of their parents before dispersal. Most importantly, our findings indicate that revegetated saltbush provides important habitat for *T. rugosa* at significant life stages – before and during breeding for adults, and before dispersal for juveniles. We conclude that revegetation using simple, monoculture plantations provides beneficial habitat for *T. rugosa* and may also be beneficial habitat for other native species in human-altered agricultural landscapes.

Additional keywords: agriculture, biodiversity, microsatellite, parentage, reptile.

Received 19 June 2012, accepted 10 September 2012, published online 12 October 2012

Introduction

Agro-ecological landscapes are essential for our food security (Stokes and Howden 2008). Much is already known about the loss of biodiversity through agricultural land use and the associated destruction and degradation of native habitats (Saunders *et al.* 1993; McIntyre *et al.* 2002). Despite an acute awareness of the need for guiding principles, the design of landscape networks to mitigate loss of biodiversity is poorly understood (Menninger and Palmer 2006). Given also the uncertainty of climate change impacts on food production and biodiversity, an understanding of how agro-ecological landscapes can be productive and manipulated to maintain biodiversity persistence is an important ecological question (Morton *et al.* 2009).

In recent times, broad-scale revegetation has been implemented in an attempt to rehabilitate agricultural landscapes and improve their 'biodiversity value' (Munro *et al.* 2007). Such actions can have indirect and direct benefits for native species. Revegetation can protect native vegetation from wind and water erosion and incursion of ground water and nutrients. It can also benefit fauna directly if it provides cover, food sources and habitat (Lefroy *et al.* 2005). Whilst the replication or recreation of complex indigenous habitats that existed before land clearing is desirable, simple revegetation in managed agricultural landscapes is more feasible and has the potential to contribute to biodiversity conservation (Collard and Fisher 2010).

In Mediterranean climates the planting of halophytic shrubs such as native saltbush (*Atriplex* spp.) has been demonstrated to be one of the most effective ways of restoring land to production (Houérou 1992). Increasingly, land managers in marginal farming areas of Australia are also using native perennial saltbush

plantings as fodder to supplement their pastures during summer and autumn months (Hobbs *et al*. 2009). This revegetation offers direct economic benefits to landholders by reducing the costs of supplementary feeding, allowing higher stocking rates and ultimately improving animal production per hectare (Monjardino *et al*. 2010). In effect, saltbush revegetation halts the spread of salinity, reduces soil erosion and 'captures carbon' while allowing grazing on previously degraded areas (Norman 2008; Collard and Fisher 2010).

As of 2002, ~454 000 ha of Australia was occupied by revegetated salt-tolerant native perennials (Lefroy *et al*. 2005), and in South Australia revegetation by native saltbush (mostly *A. nummularia*) occupies ~7000 ha of agricultural land. In the South Australian Murray Mallee (SAMM) region it has been proposed that revegetated, native saltbush (termed 'revegetated saltbush' hereafter) has the potential to provide shrub layer vegetation and associated resources for biodiversity, as well as supplementing understorey structure in existing stands of native vegetation (Collard and Fisher 2010). If this is the case, these agricultural systems could be highly important for the conservation of South Australia's native fauna, which has experienced local and landscape-scale range contractions and extinctions since the 1800s as a result of the conversion of native vegetation to farmland (Hobbs 1993).

Our study explores the direct value of revegetated saltbush for the sleepy lizard (*Tiliqua rugosa*) in the SAMM region by examining whether the species breeds in revegetated, native saltbush. *Tiliqua rugosa* is a medium-sized, long-lived scincid lizard that is broadly distributed across arid, semiarid and temperate landscapes in southern Australia (Cogger 2000). Adults are monogamous and have small, stable, overlapping home ranges (200–1000 m^2: Bull 1988). Females produce litters of 2–3 live young in late summer but juveniles generally do not disperse in their first year of life, instead remaining within the same home range as their mothers (Bull and Baghurst 1998). *Tiliqua rugosa* shelters in leaf litter under shrubs and in burrows constructed by rabbits and wombats (Kerr *et al*. 2003). It is a mainly herbivorous species (Dubas and Bull 1991). Population studies conducted over 30 years on this species near Mt Mary in the extensive rangelands ~60 km north-west of our study area show that it is not a species of conservation concern (Bull 1987, 1995). However, anecdotal observations by local farmers suggest that this species may have declined in the past 70 years in the SAMM. Adult *T. rugosa* have few predators, but many are killed on roads by vehicles and some are likely to die from consuming rabbit and fox bait during baiting periods in late summer and early autumn. Juveniles are highly susceptible to predation by large venomous snakes, foxes and cats (61–86% mortality: Bull 1987). High juvenile mortality, combined with a lack of recruitment in drought years, and the increasing incidence of very dry years, makes the conservation status of this species a concern.

Methods

Study site

This study took place in the fragmented landscapes of the SAMM wheatbelt in southern Australia. The wheatbelt extends from Blanchtown (34°20′9.90″S, 139°37′4.95″E) south to Murray Bridge (35°7′29.84″S, 139°16′47.27″E), east to Pinaroo (35°35′56.91″S, 140°42′32.96″E) and north to Sturt Highway on the Victorian border (34°16′42.21″S, 140°57′38.13″E). Land use in the region is predominantly cereals and sheep broadacre farming interspersed with discontinuous, linear roadside or patch-like remnants of mallee woodland (dominated by *Eucalyptus socialis* F.Muell. ex.Miq.) and revegetated saltbush (*Atriplex nummularia* Lindl. subsp. *nummularia*) on unproductive cropping areas. Eighty percent of native vegetation has been intensively cleared in the region since the 19th century for farming (Willoughby 2006). Since the late 1990s, 5% of the region has been planted with saltbush as a fodder reserve.

Adult *T. rugosa* were captured for marker evaluation and population genetic structure analysis from a large area (68 km × 84 km or 571 200 ha) across the SAMM (Fig. 1). Blood was collected for parentage analyses from *T. rugosa* adults and juveniles at a site (23.5 ha) within a replicate of severely modified landscape (LS3, 54 000 ha) (Fig. 2). Site LS3 was chosen for several reasons. First, the maturity and vigour of two nearby patches of revegetated saltbush provided complex habitat for food, basking and shelter from predators and summer heat. The site (34°49′14.85″S, 140°17′15.36″E) also consisted of two small revegetated saltbush patches (locality A, 6.9 ha; locality B, 12.7 ha) and a nearby native remnant patch (locality C, 3.6 ha) and a small strip of planted *Acacia* spp. (locality D, 0.3 ha), which made it practical to resample intensively until blood samples were obtained from most individuals at the site for parentage analysis. All revegetated saltbush plantings were the same age (10 years old). Most importantly, the site was surrounded by cereal crops on all sides and isolated from any linear strips of roadside native vegetation – a spatial arrangement that we predicted would limit the flow of new migrants and thus improve our chances of saturation sampling of lizards (i.e. close to 100% recapture rate, assuming a closed population). Surveys during the study period in cereal crop fields *en route* to revegetated saltbush in LS1, 2 and 3 resulted in no sightings of individuals despite many being observed in native remnants, roadside vegetation and revegetated saltbush (Fig. 1). Cereal crops at this time of year have rapid growth due to the winter rains and form virtually a closed canopy, not providing any suitable basking sites for ground-dwelling lizards but also impeding their detection.

Lizards were surveyed (multiple times) by employing a single reptile visual encounter survey (Manley *et al*. 2005) using a randomised line transect method in all revegetated saltbush and small remnant localities. Teams of two observers systematically searched neighbouring rows of monocultures of saltbush (~8–10 m apart) and a 10-m-wide strip between observers in remnants at a consistent slow pace, ensuring that both sides of individual saltbush rows and the 10-m-wide path in the remnant were surveyed thoroughly for lizards. Multiple sampling occurred over a five-day period for five weeks during November and December 2010 and in February 2011 when the ground temperature was <28°C and ideally until no new captures were recorded. Captures were grouped into two categories: juveniles (SVL < 200 mm) and subadults/adults (SVL ≥200 mm) following Smallridge and Bull (2000). Immediately after capture, animals were sexed via eversion of hemipenes in males (see below), microchipped (PIT tag, TROVAN® ID 100) for recapture identification, and sampled for genetic analysis by taking 0.4–0.6 mL blood from the caudal tail vein by

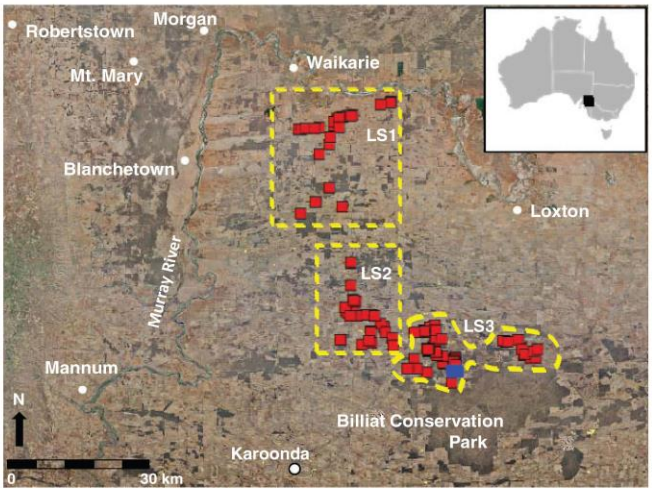


Fig. 1. Location of genotypes from 274 individuals (red squares) for analysis of population structure across three cereal cropland landscapes (LS1, LS2, LS3) with site of parentage analysis (blue rectangle) in the South Australian Murray Mallee (SAMM) region of southern Australia. Dark grey-brown area, native mallee vegetation; pale brown area south of Murray River, cropping landscapes.

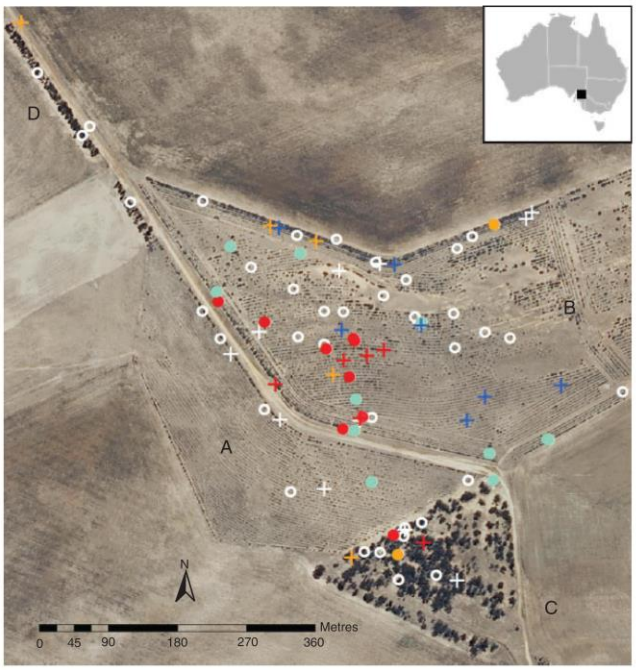


Fig. 2. Location of *T. rugosa* adults (circle, $n = 63$) and juveniles (cross, $n = 28$) in saltbush planting localities A (6.9 ha) and B (12.7 ha), native mallee remnant locality C (3.6 ha) and planted *Acacia* spp. locality D (0.3 ha) surrounded by cereal crops of LS3 of Fig. 1, South Australian Murray Mallee region, Australia. Colours correspond with parentage assignment results: white, no parents assigned; red, parents assigned with 95% confidence; blue, parents assigned with 80% confidence; orange, parents and offspring 'most likely').

venipuncture with preheparinised disposable Terumo® 23G needle per 1-mL syringe for adult/subadults (≥20 cm) or Terumo® 21G needle per 1-mL syringe for juveniles (≤20 cm) (Jacobson 1993). Blood was stored on Whatman® FTA Elute cards at room temperature with silica until required for DNA extraction. The method followed the conditions outlined in the CSIRO Human Nutrition Animal Ethics Committee 2.0 Permit (756–06/11) and the Government of South Australia Scientific Research Permit (Q25820–1).

DNA extraction and microsatellite genotyping

DNA was extracted from 3-mm^2 blood samples from 55 adult and 26 juvenile *T. rugosa* sampled in the LS3 site in accordance with the Whatman® FTA Elute card procedure (GE Healthcare, Buckinghamshire, UK). Samples from a further 219 adult *T. rugosa* were utilised for analysis of population genetic structure and marker suitability, and were extracted using the same method. Following the final wash, DNA was eluted in 50 µL of sterile distilled water and stored at −20°C. All individuals were genotyped at a panel of 10 existing microsatellite loci (Est1, TrL1, TrL3, TrL10, TrL14, TrL16, TrL21, TrL27, TrL30, TrL32: Gardner *et al*. 1999, 2008). PCR conditions (Supplementary Material 1) and visualisation followed Gardner *et al*. (2008). One negative control per PCR was run and 8.3% of individuals were amplified, genotyped and scored blindly twice to generate genotyping error rates for subsequent use in parentage analysis.

To test the performance of all microsatellites as population genetic markers, we used the expanded dataset shown in Fig. 1. All microsatellites were checked for deviations from Hardy–Weinberg Equilibrium proportions and genotypic disequilibrium using GENEPOP ver. 3.4 (Raymond and Rousset 1995). Null allele frequencies (*r*) were estimated using MICRO-CHECKER ver. 2.2.3 (Van Oosterhout *et al*. 2004). We tested the strength of our microsatellite panel for parentage analysis by calculating the probability of identity (in GenAlEx 6. 0: Peakall and Smouse 2006) using genotypes from all adult and juvenile lizards. This approach estimates the average probability that two unrelated individuals drawn from the same population have the same genotype.

Parentage analysis

Parentage was inferred using the software program CERVUS ver. 3.0.3 (Kalinowski *et al*. 2007), which uses a likelihood-based approach to assign parentage to individuals based on codominant markers and several user-defined variables. For all analyses, allele frequencies were generated from the larger dataset of *T. rugosa* individuals ($n = 219$) described above. To ensure that these allele frequencies reflected those within the smaller site used for parentage analysis (localities A–D), we used STRUCTURE ver. 2.3.3 (Pritchard *et al*. 2000) to test for population differentiation across the landscape, with the following parameters: admixture model and correlated allele frequencies, lambda = 1. 0, $K = 1–5$ with 100 000 burn-ins, 900 000 repetitions post-burn-in and 10 iterations of each *K*. Results were examined using STRUCTURE HARVESTER ver. 0.6.8 (Earl and vonHoldt 2012). We also compared pair-wise genetic relatedness between adults in the smaller site (Localities A–D) with adults at the landscape level to determine what proportions were genuinely related by pedigree and how that may affect our parentage analysis in the smaller site. We used KINGROUP ver. 2.0 (Konovalov *et al*. 2004) and compared the mean, median and range between the two sites.

A simulation of parent-pair analysis with neither parent nor sex known was run to generate log-likelihood (LOD) ratios of true parents to arbitrarily assigned parents with statistical confidence. On the basis of recapture rates, we assumed that 50% of all candidate parents were sampled and we simulated parentage for 10 000 offspring with 1000 parents, 0.86 of individuals genotyped and a 0.005 error rate (based on calculated genotyping error rates). This simulation was then used in our analysis of parentage. Only trios (offspring plus both parents) that were assigned with statistical confidence (80% and 95%) and with no allele mismatches were included in the results.

Results

Presence and density of T. rugosa *in saltbush plantations*

We achieved a total of 136 captures, comprising 91 individuals, 63 adults and 28 juveniles (of which 55 and 26 were suitable for genetic analysis) over ~700 h (Fig. 3). Thirteen juveniles and 32 adults were recaptured at least once, giving a 50% recapture rate of adults and juveniles in the last week of sampling. Although this rate fell well short of the ideal of saturation sampling, it was sufficiently robust for parentage analysis. No individual was recaptured on consecutive days during any of the five sampling weeks. During consecutive sampling weeks in November, individual lizards were recaptured at a mean interval of 7.8 days, which made it very difficult to derive a reasonable estimate of distance moved since last capture. Juveniles and adults showed a similar pattern of time between captures. Results of the population differentiation analysis identified a single genetic cluster.

Densities of *T. rugosa* were highest in saltbush locality B (5 ha^{-1}), followed by remnant locality C (3.6 ha^{-1}) and saltbush locality A, and the planted strip locality D (1.3 and 1.5 ha^{-1}) (Fig. 3). Juvenile density was higher in saltbush plantings

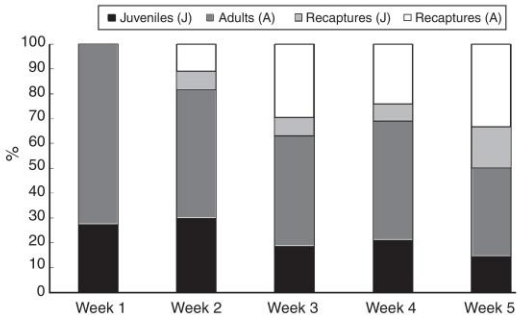


Fig. 3. Captures and recaptures of juvenile (<200 mm, $n = 28$) and adult (≥200 mm, $n = 63$) individuals over five weeks of sampling in saltbush plantings of the South Australian Murray Mallee (SAMM) region in southern Australia. (Week 1, 21 September 2010; Week 2, 13–15 November 2010; Week 3, 20–22 November 2010; Week 4, 26–29 November 2010; Week 5, 5–9 February 2011.)

(localities A and B combined) than in the other habitat types. One adult moved from remnant locality C to the south-east of saltbush locality B and another adult moved from saltbush locality B into the eastern edge of saltbush locality A. Recaptures did not show individuals moving from saltbush localities A, B or the planted-strip locality D into remnant locality C during sampling. Most individuals appeared resident in saltbush locality B during the study. A single juvenile was observed once on a track near locality D ~300 m north-west of saltbush locality B (Fig. 2). Successful captures of adults and juveniles within saltbush plantations and from the native remnant into revegetated saltbush indicate that this habitat is important for multiple life stages of the species and from one breeding season to the next for juveniles.

Hardy–Weinberg equilibrium proportions, null alleles and probability of identity

The 10 microsatellites were highly polymorphic (8–49 alleles, mean = 25.9: Table 1). All except two (TrL16 and TrL30) PCR-amplified reliably and showed no evidence of null alleles. Loci TrL16 and TrL30 deviated from Hardy–Weinberg proportions as a result of homozygous excess, most likely due to null alleles, and were removed from further analyses. The average probability that two unrelated individuals drawn from the same population had the same genotype in our samples was low (3.86×10^{-14} for unrelated individuals and 1.27×10^{-4} for siblings), indicating that our locus panel was sufficient to assign unique genotypes to even closely related individuals. All individuals within the study had unique multilocus genotypes. The raw genotypes are presented in Supplementary Material 2.

Evidence of breeding in revegetated saltbush

Of the 28 *T. rugosa* juveniles sampled from revegetated saltbush, 26 were suitable for parentage analysis. Pair-wise genetic relatedness (*R*) between adults in the smaller site (localities A–D) ranged from –0.503 to 0.644, indicating that some adults were highly related to one another, but the median value of –0.032 suggests that adults were generally unrelated to one another overall. Thirteen were assigned a parent pair (mother and father) with statistical confidence (80 or 95%). Of the remaining 13, five were assigned parent pairs with high likelihoods (LOD scores 3.53–7.37) but low delta scores (difference in LOD between the first and second most likely parent pair). For two of these five, this was due to one of the second most-likely parents being genuinely related to at least one true parent ($r = 0.46$ and 0.25). For the remaining three, however, the same parent was allocated twice in the parentage analysis, once as the most likely parent, and then again as one of the second most likely parents, thus explaining the low delta score between the two pairs. In all cases, the other second most likely parent had a low LOD score and multiple allele mismatches, indicating that the most likely pair assignment was probably genuine. The remaining eight juveniles had negative LOD scores for assignment to their most likely parent pair. This implies that for these individuals the field-assigned parents were less likely to be the true parents than a pair of arbitrary unrelated parents. This proportion of unresolved parentage assignments is not surprising given that only 50−60% of parents in the region were likely to have been sampled.

Table 1. Microsatellite variation in all 274 adult *Tiliqua rugosa* genotyped

N_a, no. of alleles; N_e, effective no. of alleles; H_e, expected heterozygosity; H_o, observed heterozygosity. ***, statistical significance at the 99.9% confidence interval following Hardy–Weinberg exact test

Locus	N_a	N_e	H_e	H_o
Est1	45	23.5	0.96	0.89
TrL1	49	31.4	0.97	0.93
TrL10	8	2.9	0.65	0.63
TrL3	29	16	0.94	0.93
TrL14	22	11.1	0.91	0.95
TrL16	17	8.1	0.88	0.64***
TrL21	20	9	0.89	0.85
TrL27	13	5.6	0.82	0.82
TrL30	53	9.56	0.9	0.86***
TrL32	21	10.3	0.9	0.88

Discussion

Planting of native perennial monocultures in agricultural landscapes can improve agricultural production and reduce overall environmental impacts of native vegetation clearance. This study has shown that native saltbush planted in the SAMM region for stock fodder also provides habitat for the endemic Australian sleepy lizard, *T. rugosa*. Results of field sampling and molecular analysis of parentage indicate that saltbush plantations are used by adult lizards before and during the breeding season, and by their young before dispersal. While complexity and structure of revegetation have been flagged as key indicators of biodiversity retention, our findings suggest that even simple revegetation can have direct conservation benefits for some native fauna.

Using molecular genetic data, we assigned parentage to 50% (13/26) of all *T. rugosa* juveniles sampled in saltbush plantations. An additional 19% ($n = 5$) were assigned parents with high likelihood scores and no allele mismatches but low statistical confidence; however, the assignments were potentially genuine. Pair-wise relatedness between adults in the study site was generally low but ranged from unrelated to having values indicative of parent/offspring or full sibling relationships. This is potentially explained by aspects of the species' behavioural ecology: *T. rugosa* takes 3–5 years to reach maturity, is long lived (up to 50 years), and has low juvenile dispersal and high adult survivorship (80−90%) (Bull 1995). A previous study of *T. rugosa* in South Australia noted that, in their first year, juveniles establish small home ranges largely within the home range of their mother (Bull and Baghurst 1998). Contact between parents and their adult offspring may therefore be quite common (Bull and Cooper 1999). If we include the five additional assignments in our results, over two-thirds of juveniles sampled in the revegetated saltbush were born to adults found in the same plantations. Although we did not track these lizards individually, we can conclude that revegetated saltbush forms at least part of the home ranges of breeding adult and juvenile sleepy lizards in the SAMM region.

Revegetated saltbush in the SAMM region often lies adjacent to roadside remnants of native vegetation, which presumably provide important habitat for *T. rugosa*. A previous study found

bird abundance and species diversity to be significantly lower in revegetated saltbush than in native vegetation sites, although some bird species used the saltbush for nesting (Seddon *et al*. 2009; Collard *et al*. 2011). On the basis of our finding that sleepy lizards use revegetated saltbush, it must be assumed that plantations provide them with resources. In agricultural landscapes, revegetated saltbush effectively re-establishes mid-storey vegetation in habitats that are otherwise devoid of complexity (Seddon *et al*. 2009). Further, in eastern Australia, a wide variety of native grasses and forbs grow under and around revegetated saltbush plants, thereby improving structure and composition of the habitat (Seddon *et al*. 2009). Although further study into fine-scale habitat use and diet of *T. rugosa* is required, it is probable that revegetated saltbush provides direct resources in the form of food and shelter to both adult and juvenile sleepy lizards. Certainly, individuals were observed foraging on plant species (flowers, leaves) that have been reported in *T. rugosa* diet ~60 km away at Mt Mary (C. M. Bull, pers. comm.).

The success of revegetation efforts in halting species declines across Australia has been mixed and appears to be linked to the level of ecological specialisation of different species (e.g. Kanowski *et al*. 2006; Collard *et al*. 2011). Overwhelmingly, remnant vegetation contains the highest abundance and species diversity when compared with plantations or cleared land for birds, mammals and reptiles (Cunningham *et al*. 2007; Munro *et al*. 2007; Collard *et al*. 2011). However, plantations are not without their benefits. A previous study on the effects of revegetation on reptiles found three-toed skinks (*Hemiergis decresiensis talbingoensis*), olive legless lizards (*Delma inornata*) and eastern blue-tongue lizards (*Tiliqua scincoides*) to be approximately three times more abundant in farms with plantings than in farms without plantings (Cunningham *et al*. 2007). Furthermore, in the same region as used in this study, threatened bird species were observed in revegetated saltbush but not in adjacent farmland sites (Collard *et al*. 2011). These findings suggest that although revegetation in the form of simple plantations is unlikely to be equivalent to remnant vegetation in terms of its conservation value, for many species it may be a better option than no revegetation at all.

In intensively farmed regions, such as the SAMM region, revegetating large blocks of land is likely to compromise farm productivity through the loss of land for farming. Economically, planting 10% of total farm area with native perennial shrubs has been shown to increase farm profitability by an average of 24% in some regions; however, planting beyond this threshold results in reduced profitability (Monjardino *et al*. 2010). In these cases, planting fodder shrubs such as saltbush may be the only means of reintroducing structural complexity into the landscape. Our most significant finding from this study was that revegetated saltbush is used by sleepy lizards in multiple life stages – as adults, before, during and after the breeding season, and by juveniles, before dispersal. Based on this, we put forward the following recommendations for managing saltbush revegetation in this and comparable regions: first, it may be beneficial for landholders to manage the timing and intensity of saltbush grazing by livestock in order to maintain a mosaic of saltbush habitats for *T. rugosa*. This may be achieved by having multiple plantations that are grazed at different times of the year, or by planting multiple species of native saltbush that vary in their tolerance to grazing. Second, the configuration of revegetated saltbush is likely to be important on both a broad scale (for establishing functional connectivity (dispersal and gene flow) across the landscape) and on a fine scale (to ensure that lizards have access to revegetated saltbush as well as remnant vegetation within their home ranges). While there are few large reserves of remnant woodland left in the SAMM region, there are many stretches of remnant roadside vegetation that could potentially act as corridors and/or suitable habitat for *T. rugosa*. Given the lack of population genetic structure across the area studied, gene flow appears to be uninhibited at the scale measured. However, the long-lived nature of this species may mean that the effects of reduced connectivity are yet to manifest in these data. Therefore, planting saltbush adjacent to these remnants would increase connectivity and the overall area of habitat that could be utilised by lizards. Finally, we acknowledge the potential conservation value of alley farming, an alternative to conventional crop-pasture rotation systems that has been shown to improve the structure, function and composition of vegetation at the site and paddock scale (Seddon *et al*. 2009). While this method of farming has the potential to integrate successful agricultural practices with conservation, further study is required to determine the value of strips of revegetated saltbush (alley farming) compared with larger, contiguous patches of revegetated saltbush (currently used in the SAMM region). For now we can conclude that revegetation using simple, monoculture plantations has some benefit for preserving native fauna in human-altered agricultural landscapes.

Acknowledgements

We thank Liz Smee and the many volunteers for helping with lizard sampling, and the Proud family for allowing access to their property during sampling. Dr Ken Wallace of the Future Farms Industry CRC and Professor Mike Bull provided very helpful comments on the manuscript. The work was funded by the Future Farms Industry CRC and CSIRO. The University of Adelaide, Flinders University and the South Australian Regional Facility for Molecular Ecology and Evolution were contracted by CSIRO to undertake the genetic analyses.

References

Bull, C. M. (1987). A population study of the viviparous Australian lizard, *Trachydosaurus rugosus* (Scincidae). *Copeia* **1987**, 749–757. doi:10.2307/1445669

Bull, C. M. (1988). Mate fidelity in an Australian lizard *Trachdosaurus rugosus*. *Behavioral Ecology and Sociobiology* **23**, 45–49. doi:10.1007/BF00303057

Bull, C. M. (1995). Population ecology of the sleepy lizard, *Tiliqua rugosa*, at Mt Mary, South Australia. *Australian Journal of Ecology* **20**, 393–402. doi:10.1111/j.1442-9993.1995.tb00555.x

Bull, C. M., and Baghurst, B. C. (1998). Home range overlap of mothers and their offspring in the sleepy lizard, *Tiliqua rugosa*. *Behavioral Ecology and Sociobiology* **42**, 357–362. doi:10.1007/s002650050448

Bull, C. M., and Cooper, S. J. B. (1999). Relatedness and avoidance of inbreeding in the lizard, *Tiliqua rugosa*. *Behavioral Ecology and Sociobiology* **46**, 367–372. doi:10.1007/s002650050631

Cogger, H. C. (2000). 'Reptiles and Amphibians of Australia.' (Reed New Holland: Sydney.)

Collard, S. J., and Fisher, A. M. (2010). Shrub-based plantings of woody perennial vegetation in temperate Australian agricultural landscapes: what benefits for biodiversity? *Ecological Management & Restoration* **11**, 31–35. doi:10.1111/j.1442-8903.2010.00510.x

Collard, S. J., Fisher, A. M., and McKenna, D. J. (2011). Planted saltbush (*Atriplex nummularia*) and its value for birds in farming landscapes of the South Australian Murray Mallee. *Ecological Management & Restoration* **12**, 37–45. doi:10.1111/j.1442-8903.2011.00568.x

Cunningham, R. B., Lindenmayer, D. B., Crane, M., Michael, D., and MacGregor, C. (2007). Reptile and arboreal marsupial response to replanted vegetation in agricultural landscapes. *Ecological Applications* **17**, 609–619. doi:10.1890/05-1892

Dubas, G., and Bull, C. M. (1991). Diet choice and food availability in the omnivorous lizard, *Trachydosaurus rugosus. Wildlife Research* **18**, 147–155. doi:10.1071/WR9910147

Earl, D. A., and vonHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualising STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4**, 359–361. doi:10.1007/s12686-011-9548-7

Gardner, M. G., Cooper, S. J. B., Bull, C. M., and Grant, W. N. (1999). Isolation of microsatellite loci from a social lizard, *Egernia stokesii*, using a modified enrichment procedure. *The Journal of Heredity* **90**, 301–304. doi:10.1093/jhered/90.2.301

Gardner, M. G., Sanchez, J. J., Dudaniec, R. Y., Rheinberger, L., Smith, A. L., and Saint, K. M. (2008). *Tiliqua rugosa* microsatellites: isolation via enrichment and characterisation of loci for multiplex PCR in *T. rugosa* and the endangered *T. adelaidensis. Conservation Genetics* **9**, 233–237. doi:10.1007/s10592-007-9316-0

Hobbs, R. J. (1993). Effects of landscape fragmentation on ecosystem processes in the Western Australian wheatbelt. *Biological Conservation* **64**, 193–201. doi:10.1016/0006-3207(93)90321-Q

Hobbs, T., Bennell, M., Huxtable, D., Bartle, J., Neumann, C., George, N., O'Sullivan, W., and McKenna, D. (2009). Potential agroforestry species and regional industries for lower rainfall Southern Australia. Florasearch 2. Available from: http://rirdc.infoservices.com.au/items/07-082.

Houérou, H. N. (1992). The role of saltbushes (*Atriplex* spp.) in arid land rehabilitation in the Mediterranean Basin: a review. *Agroforestry Systems* **18**, 107–148. doi:10.1007/BF00115408

Jacobson, E. R. (1993). Blood collection techniques in reptiles: laboratory investigations. In 'Zoo and Wild Animal Medicine: Current Therapy 3'. (Ed. M. E. Fowler.) pp. 144–152. (WB Saunders: Philadelphia.)

Kalinowski, S., Taper, M., and Marshall, T. (2007). Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology* **16**, 1099–1106. doi:10.1111/j.1365-294X.2007.03089.x

Kanowski, J. J., Reis, T. M., Catterall, C. P., and Piper, S. D. (2006). Factors affecting the use of reforested sites by reptiles in cleared rainforest landscapes in tropical and subtropical Australia. *Restoration Ecology* **14**, 67–76. doi:10.1111/j.1526-100X.2006.00106.x

Kerr, G. D., Bull, C. M., and Burzacott, D. (2003). Refuge sites used by the scincid lizard *Tiliqua rugosa. Austral Ecology* **28**, 152–160. doi:10.1046/j.1442-9993.2003.01268.x

Konovalov, D. A., Manning, C., and Henshaw, M. T. (2004). KINGROUP: a program for pedigree relationship reconstruction and kin group assignments using genetic markers. *Molecular Ecology Notes* **4**, 779–782. doi:10.1111/j.1471-8286.2004.00796.x

Lefroy, E. C., Flugge, F., Avery, A., and Hume, I. (2005). Potential of current perennial plant-based farming systems to deliver salinity management outcomes and improve prospects for native biodiversity: a review. *Australian Journal of Experimental Agriculture* **45**, 1357–1367. doi:10.1071/EA04160

Manley, P. N., Van Horne, B., Roth, J. K., Zielinski, W. K., McKenzie, M. M., Weller, T. J., Weckerly, F. W., and Vojta, C. (2005). Multiple species inventory and monitoring technical guide. Available from: http://www.fs.fed.us/psw/programs/snrc/featured_topics/msim/documents/msim_preprint.pdf

McIntyre, S., McIvor, J. G., and Heard, K. M. (2002). 'Managing and Conserving Grassy Woodlands.' (CSIRO Publishing: Melbourne.)

Menninger, H. L., and Palmer, M. A. (2006). Restoring ecological communities: from theory to practice. In 'Foundations of Restoration Ecology'. (Eds D. A. Falk, M. A. Palmer and J. B. Zedler.) pp. 88–112. (Island Press: Washington, DC.)

Monjardino, M., Revell, D., and Pannell, D. J. (2010). The potential contribution of forage shrubs to economic returns and environmental management in Australian dryland agricultural systems. *Agricultural Systems* **103**, 187–197. doi:10.1016/j.agsy.2009.12.007

Morton, S. R., Hoegh-Guldberg, O., Lindenmayer, D. B., Harriss Olson, M., Hughes, M. T., McCulloch, M. T., McIntyre, S., Nix, H. A., Prober, S. M., Saunders, D. A., Anderson, A. N., Burgman, M. A., Lefroy, E. C., Lonsdale, W. M., Lowe, I., McMichael, A. J., Parslow, J. S., Steffen, W., Williams, J. E., and Woinarski, J. C. Z. (2009). The big ecological questions inhibiting effective environmental management in Australia. *Austral Ecology* **34**, 1–9. doi:10.1111/j.1442-9993.2008.01938.x

Munro, N. T., Lindenmayer, D. B., and Fischer, J. (2007). Faunal response to revegetation in agricultural areas of Australia: a review. *Ecological Management & Restoration* **8**, 199–207. doi:10.1111/j.1442-8903.2007.00368.x

Norman, H. C. (2008). Sustainable grazing on saline lands WA1 research project final report. Available from URL: http://lwa.gov.au/products/pk071404.

Peakall, R., and Smouse, P. E. (2006). GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**, 288–295. doi:10.1111/j.1471-8286.2005.01155.x

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

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

Saunders, D. A., Hobbs, R. J., and Ehrlich, P. R. (1993). 'Nature Conservation 3: Reconstruction of Fragmented Ecosystems, Global and Regional Perspectives.' (Surrey Beatty: Sydney.)

Seddon, J., Doyle, S., Bourne, M., Maccallum, R., and Briggs, S. (2009). Biodiversity benefits of alley farming with old man saltbush in central western New South Wales. *Animal Production Science* **49**, 860–868. doi:10.1071/EA08280

Smallridge, C. J., and Bull, C. M. (2000). Prevalence and intensity of the blood parasite *Hemolivia mariae* in a field population of the skink *Tiliqua rugosa. Parasitology Research* **86**, 655–660. doi:10.1007/PL00008547

Stokes, C. J., and Howden, S. M. (2008). Climate change and Australian livestock systems: impacts, research and policy issues. *Australian Journal of Experimental Agriculture* **48**, 780–788. doi:10.1071/EA08033

Van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M., and Shipley, P. (2004). Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* **4**, 535–538. doi:10.1111/j.1471-8286.2004.00684.x

Willoughby, N. (2006). A vegetation strategy for the Murray Mallee. Towards a habitat restoration strategy for the Murray Mallee. Department for Environment and Heritage, South Australia.

Handling Editor: Steve Cooper