Factors shaping major histocompatibility complex diversity in an Australian dragon lizard

by

Jessica Danielle Hacking, BSc (Hons)

Thesis
Submitted to Flinders University
for the degree of

Doctor of Philosophy

College of Science and Engineering

May 2018

Table of contents

Abstract

The major histocompatibility complex (MHC) is a multigene family involved in pathogen recognition and immune response, and is one of the most diverse regions of the vertebrate genome. MHC genes encode cell surface glycoproteins that present self- and pathogen-derived peptides from a peptide binding region to circulating T-lymphocyte cells. Different MHC molecules bind to their own specific range of peptides. Hence, a population with high MHC diversity is less likely to be overcome by a particular pathogen.

The first aim of my thesis was to characterise the MHC in an Australian agamid, the tawny dragon lizard (*Ctenophorus decresii*), and investigate the mechanisms generating MHC polymorphism in Iguanian lizards. Six full-length *C. decresii* MHC class I transcripts were identified and were validated as likely to encode classical MHC molecules. Two partial transcripts were also identified and may represent non-classical MHC class I molecules, although truncation due to transcriptome assembly restrictions is more likely. I found evidence for both birth-and-death and concerted evolution generating MHC diversity within Iguanian lizards.

I used family group data to aid MHC genotyping, explore MHC allele segregation patterns and identify mechanisms maintaining MHC diversity (including MHC-associated mating). To obtain family group data I undertook a captive hatching program for a Flinders Ranges population of *C. decresii*. The second aim of my thesis was to determine the genetic mating system employed by *C. decresii*. A predominantly polygynous genetic mating system with a very low rate of multiple paternity was observed (4%), representing one of the lowest rates among squamates.

The third aim of my thesis was to describe polymorphism for previously characterised *C. decresii* MHC class I transcripts. I genotyped individuals across five populations at the MHC using a recently developed clustering method, which was aided by the analysis of family group data. This method allowed in-depth analysis of type I and II genotyping error rates. Varying numbers of alleles were observed among individuals within populations, and together with allele segregation patterns within family groups, suggested allele dropout and/or natural copy number variation. Sites under selection were identified, allowing the identification of putative peptide binding sites. High allele sharing among populations and low allelic diversity in an island population was uncovered.

Finally, the fourth aim of my thesis was to examine the mechanisms maintaining MHC class I diversity within a *C. decresii* population; specifically, sexual selection and parasite-mediated selection.

I performed model selection analyses and discovered that parasite load was significantly lower in the presence of a specific MHC allele functional group (supertype), indicating that this supertype confers resistance. Furthermore, I uncovered trends suggesting that sexual selection may also play a role in maintaining MHC diversity through MHC-associated mating.

Overall, I was successful in identifying the factors influencing MHC class I diversity within an agamid lizard and addressed broad evolutionary ecology questions. The work presented in this thesis significantly broadens knowledge on squamate MHC class I genes. Future work includes investigating population differentiation using MHC markers and identifying the mechanisms operating to maintain MHC diversity among populations.

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.

Jessica Danielle Hacking

19 December 2017

Acknowledgements

I would like to thank the many people who have made this thesis possible.

Above all, I'd like to thank my primary supervisor, Mike Gardner, for his guidance, support and patience. Mike, you allowed me the freedom to develop my own ideas and you were always there to offer kind words of encouragement and push me to achieve my best work.

Next, I'd like to thank my co-authors. I owe thanks to Tessa Bradford for her support, advice and friendship. Tessa, I relied on you for your incredible knowledge on lab techniques and bioinformatics – I wouldn't have known where to start without your guidance! I feel incredibly lucky to have had you as a sounding-board. Devi Stuart-Fox also played a large role throughout my candidature. Devi, thank-you for teaching me about the tawny dragon, colour analysis and statistics, and for being very generous with your time. I also owe thanks to Terry Bertozzi. Terry, thank-you for your help with transcriptome assembly and for several helpful discussions about the MHC. Adnan Moussalli also helped with transcriptome analysis and provided insights into phylogenetic analysis. I also appreciate the statistical advice provided by Stephanie Godfrey, and the helpful python code provided by Kelly Pierce. Thank-you also to my secondary supervisor Mike Schwarz for his encouragement and support.

I'm grateful for the friendship, support, encouragement and advice provided by past and present PhD students in the LEGS lab and the Stuart-Fox lab. Sarah, thanks for making me feel welcome in the LEGS lab and for introducing me to fieldwork in the Flinders Ranges – I'll always remember that time we had to seek shelter in the car from a huge sand storm! Amy, thanks for your friendship and for helping me with population genetics analysis. Mina, thank-you for introducing me to the MHC and for understanding my frustration with MHC class II. Thanks also to Bonnie, Lucy and Robert for their friendship – I wish you all the best of luck on your PhD adventures. I also owe a huge thank-you to Maddy Yewers for teaching me how to catch tawny dragons and take blood. I remember our time in the field will fondness and I too will never forget our unnerving encounter with a large snake who was probably also trying to catch tawny dragons. I'm indebted to Katrina Rankin for her guidance on tawny dragon captive hatching. Thanks for quickly answering all of my panicky emails! I also owe thanks to Claire McLean for help with fieldwork and microsatellite analysis. My thesis wouldn't have been possible without the previous research completed by the ladies of the Stuart-Fox lab. Your work was an inspiration throughout my PhD journey.

I am incredibly grateful to the army of volunteers who helped with fieldwork and captive hatching; Carly Humphrys, Robbie Brent, Taryn Lourie, Edward Tomczyk, Phoebe Paterson de Heer, Brian Matthews, Amy Tschirn, Hailey Graham, Krysta Simms, Leela Roberts, Patrick Taggart and Shelley Thompson. You were all great company and I truly appreciate your time and effort. I owe a huge thank-you to the Animal house staff, especially Leslie, who had to put up with my slightly neurotic overprotectiveness of captive lizards.

This work was almost completely funded by small grants and I am extremely grateful for the funds awarded by the Holsworth Wildlife Research endowment, Linnean Society of New South Wales, Field Naturalists Society of South Australia, Nature Foundation of South Australia, American Society for the Study of Amphibians and Reptiles, Wildlife Preservation Society of Australia, Royal Society of South Australia and the Flinders University Interfaculty Collaboration Grant.

My friends and family provided love, support and encouragement throughout my candidature. Mum, thank-you for always having confidence in me and always encouraging me to follow my dreams. To my little brother, thank-you for your support and for your help in the field. Emma, Marie and Melly, my kindred spirits, thank-you for your interest in my research, your love and encouragement.

I acknowledge the traditional owners of the land on which lizards were sampled. I am grateful to the Adnyamathanha people of the Flinders Ranges for allowing me to capture lizards on a historical reserve.

I often struggled with the morality of stressing lizards in order to obtain scientific knowledge that didn't directly benefit them, and I feel that it's important that I acknowledge the hundreds of lizards that I stressed out in the name of science.

Last, but not least, I'd like to thank my amazing husband for his unwavering love and patience. Adam, thank-you for being a sounding board, for helping me in the field, entering data and for having confidence in me when I didn't have confidence in myself.

Statement of co-authorship

Author current affiliations

Jessica Hacking (candidate)

College of Science and Engineering, Biological Sciences, Flinders University

Michael Gardner (primary supervisor)

College of Science and Engineering, Biological Sciences, Flinders University

Evolutionary Biology Unit, South Australian Museum

Tessa Bradford

College of Science and Engineering, Biological Sciences, Flinders University

Evolutionary Biology Unit, South Australian Museum

School of Biological Sciences, University of Adelaide

Devi Stuart-Fox

School of BioSciences, University of Melbourne

Terry Bertozzi

Evolutionary Biology Unit, South Australian Museum

School of Biological Sciences, University of Adelaide

Adnan Moussalli

Sciences Department, Museum Victoria

Stephanie Godfrey

Department of Zoology, University of Otago

Kelly Pierce

College of Science and Engineering, Biological Sciences, Flinders University

Details of co-authorship

Chapter 2

Published:

Hacking, J. (1), Bertozzi, T. (2), Moussalli, A. (3), Bradford, T. (4), Gardner, M. (5).

Characterization of major histocompatibility complex class I transcripts in an Australian dragon lizard.

Developmental and Comparative Immunology

The candidate was the primary author and conducted all field work. The candidate and author 3 undertook laboratory work. The candidate undertook data analysis. Authors 2 and 3 undertook transcriptome assembly bioinformatics (data analysis). Authors 2, 4 and 5 provided methodological advice. The candidate and author 5 contributed to the development of main ideas and approach. The candidate and authors 2, 3, 4 and 5 contributed to refining the text.

1: 65%, 2: 10%, 3: 10%, 4: 5%, 5: 10%

Chapter 3

In press:

Hacking, J. (1), Stuart-Fox, D. (2), Gardner, M. (3).

Very low rate of multiple paternity detected in clutches of a wild agamid lizard.

Australian Journal of Zoology

The candidate was the primary author and conducted all field work and laboratory work. The candidate and author 3 undertook data analysis. Authors 2 and 3 provided methodological advice. The candidate and author 3 contributed to the development of main ideas and approach. The candidate and authors 2 and 3 contributed to refining the text.

1: 80%, 2: 5%, 3: 15%

Chapter 4

Submitted for publication:

Hacking, J. (1), Bradford, T. (2), Pierce, K. (3), Gardner, M. (4).

De novo genotyping of the MHC in an Australian dragon lizard aided by family group data.

Transactions of the Royal Society of South Australia

The candidate was the primary author and conducted all field work and laboratory work. The candidate undertook data analysis. Author 3 provided a python script for data manipulation (data analysis). Authors 2 and 4 provided methodological advice. The candidate and author 4 contributed to the development of main ideas and approach. The candidate and authors 2, 3 and 4 contributed to refining the text.

1: 80%, 2: 10%, 3: 5%, 4: 5%

Chapter 5

Submitted for publication:

Hacking, J. D. (1), Stuart-Fox, D. M. (2), Godfrey, S. (3), Gardner, M. G. (4).

Specific MHC class I supertype associated with parasite infection and colour morph in a wild lizard population.

Molecular Ecology

The candidate was the primary author and undertook all data analysis. Authors 2, 3 and 4 provided methodological advice. The candidate and authors 2 and 4 contributed to the development of main ideas and approach. The candidate and authors 2, 3 and 4 contributed to refining the text.

1: 85%, 2: 5%, 3: 5%, 4: 5%

We, the undersigned agree with the above stated "proportion of work undertaken" for each of the above submitted manuscripts or manuscripts in preparation for submission contributing to this thesis:

Jessica Hacking (candidate)

Michael Gardner (primary supervisor)

Tessa Bradford

Devi Stuart-Fox

Terry Bertozzi

Adnan Moussalli

Stephanie Godfrey

Kelly Pierce

Preface

This thesis contains four data chapters (chapters 2 to 5) that contribute to the fulfilment of a doctoral program. A general introduction is provided in chapter 1 and chapter 6 provides a general discussion. Chapters 1 and 6 are brief to avoid unnecessary overlap with data chapters. At the start of each data chapter a 'note to examiners' section provides information on how the chapter relates to previous and/or subsequent chapters. Collaborations were sought to provide advice or to undertake some aspects of the work presented in this thesis. However, I conducted most of the work. Such collaborations resulted in co-authorship of manuscripts. Details of co-authorship are provided in the 'statement of co-authorship' section of the thesis. Each data chapter was written as a separate manuscript for publication in a scientific journal. Therefore, references are provided at the end of each chapter and there is some information overlap among chapters. Chapter 2 has been published and chapter 3 has been accepted for publication and is currently in press. Chapters 4 and 5 have been submitted for publication. Supplementary information that is included in publications is provided after each chapter and two thesis-only appendices are provided at the end of the thesis. Bookmarks have been embedded within this document for easy naviagation.

CHAPTER 1

General introduction

The major histocompatibility complex (MHC) is an important gene family, not only due to its central role in the immune system, but also its role in the fields of conservation genetics and evolutionary ecology. MHC markers are often used as measures of population health or adaptive potential in conservation genetics studies (Ujvari and Belov, 2011) and have been used to aid breeding programs (Grogan et al., 2017). Furthermore, the MHC has been used a tool to investigate several evolutionary ecology questions, such as red queen dynamics (Lighten et al., 2017) and the Hamilton-Zuk 'good genes' hypothesis (Dunn et al., 2013). Despite this, the MHC has not been characterised in most non-avian reptile families. As a result, knowledge on the level of polymorphism present and the mechanisms shaping MHC diversity in non-avian reptiles is conspicuously lacking (Elbers and Taylor, 2016). Parasite-mediated selection is considered to be the primary driver of MHC diversity among jawed vertebrates. However, in most vertebrate orders sexual selection has also been found to play a role via MHC-associated mating. Does sexual selection also play a role in shaping MHC diversity in non-avian reptiles? If so, is there evidence for mating based on honest signals of MHC-based disease resistance, as predicted by the classic Hamilton-Zuk hypothesis (Hamilton and Zuk, 1982)? The tawny dragon lizard, *Ctenophorus decresii*, provides an excellent model system in which to examine the mechanisms generating and maintaining MHC diversity. This species is host to several extracellular and intracellular parasites and is likely under strong sexual selection as it possesses conspicuous secondary sexual colouration .

My primary aim in this thesis is to identify mechanisms shaping MHC diversity in *C. decresii*. Of key interest was determining whether a link exists between the MHC, parasites and secondary sexual colouration. Here, I provide background on the MHC and the tawny dragon lizard to supplement information provided in data chapters. I conclude by providing an outline of the thesis.

An introduction to the major histocompatibility complex

All organisms are challenged by pathogens in their environment. In jawed vertebrates, an essential part of the defence against pathogens is the major histocompatibility complex (MHC). The MHC is a hypervariable multigene family that plays a central role in self-tolerance and pathogen detection (Piertney and Oliver, 2006)., The MHC is one of the best studied gene families in humans and model organisms due to its role in the success of tissue and organ transplantations (Garcia et al., 2012). MHC genes have now been described in many non-model species and are often used in wildlife conservation genetic studies as an indicator of population health and adaptive potential (Eimes et al., 2011, Smith et al., 2009, Sommer, 2005). Furthermore, the extreme diversity observed at the MHC has led many researchers to investigate the evolutionary mechanisms responsible for generating and maintaining this diversity (Evans et al., 2012, Meyer-Lucht et al., 2008, Tollenaere et al., 2008, Wedekind and Furi, 1997).

Structure and function of the MHC

MHC genes are divided into four classes based primarily on differences in gene form and function. Class I and II are further divided into classical and non-classical genes. Classical MHC class I and II genes encode cell surface glycoproteins that bind to both self- and pathogen-derived peptides, presenting them to circulating T-lymphocyte cells (T cells) (Neefjes et al., 2011). Non-classical MHC genes generally exhibit low polymorphism and a restricted pattern of expression (Pratheek et al., 2014). In humans non-classical MHC genes are involved in both cell surface peptide presentation and other immune-related functions (Allen and Hogan, 2013, Pratheek et al., 2014). Although non-classical MHC genes have been characterised for many non-model species, in most cases their function remains unknown (Dirscherl and Yoder, 2014, Glaberman et al., 2009, Harstad et al., 2008, Lei et al., 2015). Class III and IV MHC genes are functionally divergent from class I and II MHC genes and perform other important immune-related functions such as tumour necrosis and complement activation (Gruen and Weissman, 2001).

Classical class I and II MHC (henceforth, simply class I and II) molecules both present peptides to T cells from the surface of cells, although the source of peptides and the pathways involved in peptide presentation differ greatly between class I and II (Neefjes et al., 2011). Class I MHC molecules present intracellularly-derived peptides to cytotoxic T cells from the surface of all nucleated cells except sperm cells and some neurons (Rock et al., 2016). As class I MHC molecules present intracellularly-derived peptides they are mainly associated with defence against intracellular pathogens such as viruses (Rock et al., 2016). Class II MHC molecules, are expressed only on the surface of immune-related cells such as macrophages, lymphocytes and dendritic cells, and present extracellularly-derived peptides to helper T cells (Rock et al., 2016). As a result, MHC class II molecules are primarily associated with

defence against extracellular pathogens such as extracellularly-residing bacteria (Rock et al., 2016). However, cross-presentation and autophagy allow for circumstances in which MHC class I present extracellularly-derived peptides and MHC class II present intracellularly-derived peptides (Neefjes et al., 2011).

In terms of structure, both MHC class I and II molecules possess transmembrane immunoglobulin 'stalks' that anchor the molecule to the surface of cells, extracellular domains that form antigen-binding sites, and leader peptides. Class I molecules consist of one polypeptide chain that includes a transmembrane 'stalk' and three extracellular domains (α_1, α_2 and α_3), of which, α_1 and α_2 form the peptide binding region (PBR) (Smith, 2012). The peptide binding region of the MHC gene forms the cleft in which the peptide sits, ready for presentation to T cells. All components of class I molecules are encoded by a single gene. On the other hand, class II molecules consist of two polypeptide chains (α and β) that are encoded by separate genes. Each chain includes a transmembrane 'stalk' and two extracellular domains (either α_1 and α_2 or β_1 and β_2). The α_1 and β_1 domains from the two chains form the antigen binding site (Piertney and Oliver, 2006). Comprehensive genomic maps of the MHC have been completed for mammals (Anzai et al., 2003, Beck and Trowsdale, 2000, Gustafson et al., 2003, Stafuzza et al., 2013), birds (Shiina et al., 2004), amphibians (Flajnik et al., 1999) and fish (Michalová et al., 2000, Sambrook et al., 2002). Comparative genomic analysis revealed that the genomic organisation, number of genes, and level of polymorphism varies greatly among taxa (Kelley et al., 2005).

MHC class I and II diversity

Diversity at the MHC is manifested at multiple levels; it is highly polygenic, often possesses many duplicated loci, and is usually extremely polymorphic within loci (Janeway et al., 2001). Furthermore, the total number of MHC loci may differ among individuals within and among populations; a phenomenon termed copy number variation (CNV; De Groot et al., 2015, Lighten et al., 2014a, Schrider and Hahn, 2010). Yet another layer of MHC diversity is the formation of isotypes via alternative splicing of messenger RNA, resulting in one MHC allele encoding different forms (isoforms) of a MHC molecule that can undertake different functions (Dai et al., 2014, Lillie et al., 2016, Voorter et al., 2016). MHC allelic variants differ primarily within the PBRs, resulting in molecules that differ in peptide binding specificities (Janeway et al., 2001).

In contrast to T cell receptors, MHC molecules do not undergo somatic recombination (Vanhanen et al., 2016). Hence, MHC diversity is fixed within an individual. As a result, the greater MHC diversity an individual possesses, the greater the diversity of pathogens that the immune system can recognise and act upon. However, usually individuals possess an intermediate level of MHC diversity, despite high

MHC diversity within the population (Woelfing et al., 2009). It has been suggested that the advantage of possessing high MHC diversity is counteracted by the effects of high MHC diversity on T cell negative selection in the thymus (Nowak et al., 1992, Wegner et al., 2004). High MHC diversity may reduce the number of naïve T cells allowed though thymic selection, limiting T cell diversity and the ability of the immune system to respond to foreign invaders (Woelfing et al., 2009). However, Borghans et al. (2003) argues that an extreme number of different MHC molecules would be required to have a negative impact on T cell diversity. Other possible explanations for intermediate individual MHC diversity include a reduced chance of autoimmune disorders, and optimal concentration of MHC molecules at the surface of antigen-presenting cells (Woelfing et al., 2009).

As the MHC is characterised in an increasing number of species it is becoming evident that this gene-rich region is extremely plastic in the way diversity is distributed. CNV at the MHC has been confirmed through genomic mapping or segregation analysis in several species (Codner et al., 2012, Doxiadis et al., 2011, Hosomichi et al., 2006, McConnell and Restaino, 2014, Reed et al., 2016) and has been assumed to occur in many others when differing numbers of MHC alleles among individuals within populations are observed (e.g. Alcaide et al., 2014, Bonneaud et al., 2004, Lighten et al., 2014b, Radwan et al., 2014, Radwan et al., 2012, Siddle et al., 2010, Zagalska-Neubauer et al., 2010). A comparison between the class I MHC of rhesus macaques and humans found that although these species have a similar level of diversity, rhesus macaque MHC diversity results primarily from CNV, with little allelic diversity at loci. On the other hand, humans lack CNV in this region and instead possess high allelic diversity (De Groot et al., 2015). The evolution of these alternative strategies of distributing diversity within the population (CNV versus allelic polymorphism) is not well understood and warrants further investigation (De Groot et al., 2015).

Generation and maintenance of MHC diversity

The primary mechanisms generating diversity at the MHC are birth-and-death evolution and concerted evolution, which have often been found to be working in tandem (Edwards and Hedrick, 1998, Eirín-López et al., 2012, Nei et al., 1997). The birth-and-death model of evolution is considered to be the primary mechanism generating and shaping MHC diversity (Eirín-López et al., 2012, Nei et al., 1997). This model of evolution is specific to multigene families and involves the creation of new genes via duplication (Nei and Rooney, 2005). Some duplicated genes are kept within the genome over extended periods of time, often resulting in trans-species polymorphism, and others become non-functional pseudogenes (Edwards and Hedrick, 1998, Nei and Rooney, 2005). As a result, pseudogenes and non-classical MHC genes are often positioned among functional classical MHC genes (Nei et al., 1997). Concerted evolution likely plays a secondary role in generating and shaping MHC diversity. This form of evolution is common among multigene families and involves MHC genes evolving together as a unit;

mutations are spread among MHC genes via repeated unequal crossover and gene conversion events. Consequently, under concerted evolution genes phylogenetically cluster by species rather than gene (Nei and Rooney, 2005).

The proposed mechanisms maintaining MHC diversity include heterozygote or intermediate diversity advantage, rare allele advantage and fluctuating selection (Spurgin and Richardson, 2010). Under heterozygote or intermediate diversity advantage it is proposed that MHC diversity is maintained within the population because individuals with optimally high MHC diversity are better able to recognise and act upon a larger range of pathogens and are therefore at a selective advantage (Sommer, 2005). However, heterozygote or intermediated diversity advantage does not take into account the co-evolutionary relationship between host and pathogen. Rare allele advantage describes the continual cycling or replacement of MHC alleles due to an arms race between host and pathogen (Borghans et al., 2004). An MHC allele that allows individuals to combat a pathogen is selected for but once the pathogen adapts to such an allele it is then selected against and reduces in frequency (perhaps to be selected for in the future; negative-frequency-dependent selection) or is lost (directional selection) (Lighten et al., 2017). An existing rare allele or a new allele that aids in combatting the counter-adapted pathogen is selected for and the cycle continues (Lighten et al., 2017). Under fluctuating selection MHC diversity may be maintained due to spatial and/or temporal heterogeneity in selection pressure from pathogens (Eizaguirre et al., 2012). More than one mechanism may be acting together to maintain diversity and mechanisms may be operating at different levels (i.e. within vs. among populations, Spurgin and Richardson, 2010). Furthermore, selection may be in the form of pathogens (parasite-mediated selection) and/or sexual selection via MHC-associated mate choice (Piertney and Oliver, 2006).

Evidence for parasite-mediated selection maintaining MHC diversity through rare allele advantage (Borghans et al., 2004, Schwensow et al., 2017), fluctuating selection (Jones et al., 2015, Osborne et al., 2017) and heterozygote advantage (Doherty and Zinkernagel, 1975, Takahata and Nei, 1990) or intermediate diversity advantage (Kloch et al., 2010, Lei et al., 2016) has been uncovered across a range of taxa. However, most evidence points towards rare allele advantage and fluctuating selection as the dominant mechanisms by which pathogen-mediated selection acts on the MHC, with little evidence that heterozygote advantage alone can account for the extreme diversity found at MHC loci (De Boer et al., 2004). MHC-associated mate choice has been discovered in most vertebrate classes, including bony fish (Evans et al., 2012, Reusch et al., 2001), amphibians (Bos et al., 2009), reptiles (Miller et al., 2009, Olsson et al., 2003, Pearson et al., 2017), birds (Juola and Dearborn, 2012, Strandh et al., 2012) and mammals (Cutrera et al., 2012, Schad et al., 2012). Sexual selection can play a large role in maintaining MHC diversity in some systems; Winternitz et al. (2013) found that sexual selection explains

more functional variation than parasite-mediated selection in some mammals. Mate choice may be influenced by MHC diversity, mate compatibility for optimal MHC diversity in offspring, and/or specific MHC genotypes (Ejsmond et al., 2014). Spatio-temporal heterogeneity in mate choice for MHC characteristics may also play a role in maintaining diversity (fluctuating selection, Cutrera et al., 2014).

An introduction to the tawny dragon lizard

Natural history

The tawny dragon lizard (*Ctenophorus decresii*) is a small (< 30g) diurnal rock-obligate agamid (fig. 1) that is endemic to rocky ranges throughout South Australia, including the Flinders Ranges, Olary Ranges, Mount Lofty Ranges and Kangaroo Island. *Ctenophorus decresii* is both sexually dimorphic and dichromatic. Males are larger than females in terms of mass, snout-to-vent length (SVL) and head size, and possess secondary sexual ornamentation in the form of colouration (Gibbons, 1979). Males possess bright throat colouration and a black-grey chest patch, both of which are emphasised during male displays and are important signals to conspecifics (Gibbons, 1979, Osborne, 2005, Osborne et al., 2012, Yewers, 2016). Sexual maturity is reached at approximately one year of age and in the wild individuals can live for at least five years (Yewers, 2016). Adult male *C. decresii* are highly territorial and defend territories that likely encompass several female home ranges or territories (Yewers, 2016).

Figure 1. An adult male tawny dragon lizard (*Ctenophorus decresii*) in the Flinders Ranges.

Phylogeography and colour differentiation among populations

Phylogeographic analysis of *C. decresii* populations has uncovered two distinct genetic lineages; a 'northern' lineage encompasses populations of the Flinders and Olary Ranges and a 'southern' lineage includes populations of the Mount Lofty Ranges and Kangaroo Island (McLean et al., 2014b). The northern and southern lineages come together in the Barossa Valley region, likely representing secondary contact (McLean et al., 2014b). Estimation of historical population size revealed population decline in the late Holocene and was attributed to range contraction as a result of increased aridity and climatic instability (McLean et al., 2014b).

Dorsal, lateral and ventral colouration differs between northern and southern lineages (McLean et al., 2014a, McLean et al., 2014b). Populations within the northern lineage are polymorphic for male throat colour and include four discrete throat colour morphs (orange, orange with yellow, yellow and grey), whereas males within the southern lineage are monomorphic and have UV-blue throats (McLean et al., 2014b, Teasdale et al., 2013). Local adaptation to substrate (rock and lichen) colour has played a role in the evolution of male colouration and environmental factors influence the frequency of morph types within polymorphic populations (McLean et al., 2014a, McLean et al., 2015). Despite the absence of UV colouration in the northern lineage, no differences in the expression of UV-related opsin genes or UV spectral sensitivities have been uncovered between lineages (Yewers et al., 2015).

Colour polymorphism

Male *C. decresii* within the northern lineage are polymorphic for throat colour. Male throat morph type is set at maturity and is highly heritable (Rankin et al., 2016). Some females produce small amounts of yellow colouration at the base of the throat but do not express colour associated with male morphs under natural conditions. However, females exposed to testosterone produce throat morphs to the same extent as males, suggesting autosomal rather than sex-linked inheritance and a role for hormones in colour expression (Rankin and Stuart-Fox, 2015). Colouration is variable within male morph types and the brightness of throat colour is affected by stress hormones (Lewis et al., 2017). Male throat morph types don't differ in morphology or microhabitat preferences (Teasdale et al., 2013) but do differ in behaviour and hormone expression (Yewers et al., 2016).

Genetic and biochemical basis of colouration and spectral sensitivities

Recent work has identified the pigments and genes responsible for male *C. decresii* throat colouration (McLean et al., 2017). McLean et al. (2017) found that orange skin has the highest levels of pteridine pigments and genes associated with pteridine synthesis, yellow skin has the highest levels of carotenoid pigments and genes associated with the carotenoid pathway. Grey skin is associated with high levels of melanin pigments and melanin production genes and blue skin is associated with genes linked to structural colouration (McLean et al., 2017).

Work on the spectral sensitivities of *C. decresii* confirmed the existence of three single cones; short wavelength sensitive (SWS), medium wavelength sensitive (MWS) and long wavelength sensitive (LWS) (Yewers et al., 2015). Furthermore, based on spectral sensitivities of the congener *C. ornatus*, it is likely that *C. decresii* also possesses double cones (Barbour et al., 2002, Yewers et al., 2015). Double cones allow detection of signal luminance ('brightness') in birds and lizards (Osorio and Vorobyev, 2005). Although ultraviolet/violet sensitive (UVS/VS) cones were not discovered in *C. decresii,* it is likely that they are present, based on transcriptome analysis of opsin genes (Yewers et al., 2015).

Thesis outline

Here, I have outlined the broad aims of the thesis and introduced the MHC and the tawny dragon lizard. In chapter 2 I characterise MHC class I genes in *C. decresii* and investigate potential mechanisms generating diversity at the Iguanian MHC class I region. Family group data was required for MHC genotyping, analysing allele segregation and testing for MHC-associated mate choice. Therefore, I undertook captive hatching and paternity analysis for a wild *C. decresii* population. Resulting family group data was used to define the genetic mating system employed by *C. decresii* (chapter 3). In chapter 4 I outline the methodology employed for MHC genotyping and describe MHC class I diversity present within and among individuals. In chapter 5 the mechanisms maintaining MHC class I diversity within a wild *C. decresii* population are explored. Finally, in chapter 6 I summarise the thesis and outline directions for future work.

References

ALCAIDE, M., MUNOZ, J., PUENTE, J., SORIGUER, R. & FIGUEROLA, J. 2014. Extraordinary MHC class II B diversity in a non-passerine, wild bird: the Eurasian Coot *Fulica atra* (Aves: Rallidae). *Ecology and Evolution,* 4**,** 688-698.

ALLEN, R. L. & HOGAN, L. 2013. Non-Classical MHC Class I Molecules (MHC-Ib). *eLS.* John Wiley & Sons, Ltd.

ANZAI, T., SHIINA, T., KIMURA, N., YANAGIYA, K., KOHARA, S., SHIGENARI, A., YAMAGATA, T., KULSKI, J., NARUSE, T., FUJIMORI, Y., FUKUZUMI, Y., YAMAZAKI, M., TASHIRO, H., IWAMOTO, C., UMEHARA, Y., IMANISHI, T., MEYER, A., IKEO, K., GOJOBORI, T., BAHRAM, S. & INOKO, H. 2003. Comparative sequencing of human and chimpanzee MHC class I regions unveils insertions/deletions as the major path to genomic divergence. *Proceedings of the National Academy of Sciences of the United States of America,* 100**,** 7708-7713.

BARBOUR, H. R., ARCHER, M. A., HART, N. S., THOMAS, N., DUNLOP, S. A., BEAZLEY, L. D. & SHAND, J. 2002. Retinal characteristics of the ornate dragon lizard, *Ctenophorus ornatus*. *The Journal of Comparative Neurology,* 450**,** 334-344.

BECK, S. & TROWSDALE, J. 2000. The human major histocompatability complex: lessons from the DNA sequence. *Annual review of genomics and human genetics,* 1**,** 117-137.

BONNEAUD, C., SORCI, G., MORIN, V., WESTERDAHL, H., ZOOROB, R. & WITTZELL, H. 2004. Diversity of Mhc class I and IIB genes in house sparrows (*Passer domesticus*). *Immunogenetics,* 55**,** 855-65.

BORGHANS, J. A. M., BELTMAN, J. B. & DE BOER, R. J. 2004. MHC polymorphism under host-pathogen coevolution. *Immunogenetics,* 55**,** 732-739.

BORGHANS, J. A. M., NOEST, A. J. & BOER, R. J. 2003. Thymic selection does not limit the individual MHC diversity. *European Journal of Immunology,* 33.

BOS, D. H., WILLIAMS, R. N., GOPURENKO, D., BULUT, Z. & DEWOODY, J. A. 2009. Condition-dependent mate choice and a reproductive disadvantage for MHC-divergent male tiger salamanders. *Molecular Ecology,* 18**,** 3307-15.

CODNER, G., BIRCH, J., HAMMOND, J. & ELLIS, S. 2012. Constraints on haplotype structure and variable gene frequencies suggest a functional hierarchy within cattle MHC class I. *Immunogenetics,* 64**,** 435-445.

CUTRERA, A. P., FANJUL, M. S. & ZENUTO, R. R. 2012. Females prefer good genes: MHC-associated mate choice in wild and captive tuco-tucos. *Animal Behaviour,* 83**,** 847-856.

CUTRERA, A. P., ZENUTO, R. R. & LACEY, E. A. 2014. Interpopulation differences in parasite load and variable selective pressures on MHC genes in *Ctenomys talarum*. *Journal of Mammalogy,* 95**,** 679-695.

DAI, Z.-X., ZHANG, G.-H., ZHANG, X.-H., ZHU, J.-W. & ZHENG, Y.-T. 2014. A splice variant of HLA-A with a deletion of exon 3 expressed as nonmature cell-surface glycoproteins forms a heterodimeric structure with full-length HLA-A. *Human Immunology,* 75**,** 234-238.

DE BOER, R., BORGHANS, J., VAN BOVEN, M., KESMIR, C. & WEISSING, F. 2004. Heterozygote advantage fails to explain the high degree of polymorphism of the MHC. *Immunogenetics,* 2004**,** 725-731.

DE GROOT, N., BLOKHUIS, J. H., OTTING, N., DOXIADIS, G. G. & BONTROP, R. E. 2015. Co-evolution of the MHC class I and KIR gene families in rhesus macaques: ancestry and plasticity. *Immunological Reviews,* 267**,** 228–245.

DIRSCHERL, H. & YODER, J. A. 2014. Characterization of the Z lineage Major histocompatability complex class I genes in zebrafish. *Immunogenetics,* 66**,** 185-98.

DOHERTY, P. C. & ZINKERNAGEL, R. M. 1975. Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. *Nature,* 256**,** 50-2.

DOXIADIS, G. G., DE GROOT, N., OTTING, N., BLOKHUIS, J. H. & BONTROP, R. E. 2011. Genomic plasticity of the MHC class I A region in rhesus macaques: extensive haplotype diversity at the population level as revealed by microsatellites. *Immunogenetics,* 63**,** 73-83.

DUNN, P. O., BOLLMER, J. L., FREEMAN-GALLANT, C. R. & WHITTINGHAM, L. A. 2013. MHC variation is related to a sexually selected ornament, survival, and parasite resistance in common yellowthroats. *Evolution,* 67**,** 679-87.

EDWARDS, S. & HEDRICK, P. 1998. Evolution and ecology of MHC molecules: from genomics to sexual selection. *Trends in Ecology & Evolution,* 13**,** 305-311.

EIMES, J., BOLLMER, J., WHITTINGHAM, L., JOHNSON, J., VAN OOSTERHOUT, C. & DUNN, P. 2011. Rapid loss of MHC class II variation in a bottlenecked population is explained by drift and loss of copy number variation. *Journal of Evolutionary Biology,* 24**,** 1847-1856.

EIRÍN-LÓPEZ, J. M., REBORDINOS, L., ROONEY, A. P. & ROZAS, J. 2012. The birth-and-death evolution of multigene families revisited. *In:* GARRIDO-RAMOS, M. A. (ed.) *Genome Dynamics: Repetitive DNA.* Basel, Switzerland: Karger.

EIZAGUIRRE, C., LENZ, T. L., KALBE, M. & MILINSKI, M. 2012. Rapid and adaptive evolution of MHC genes under parasite selection in experimental vertebrate populations. *Nature communications,* 3**,** 621.

EJSMOND, M., RADWAN, J. & WILSON, A. 2014. Sexual selection and the evolutionary dynamics of the major histocompatibility complex. *Proceedings of the Royal Society B: Biological Sciences,* 281**,** 1-8.

ELBERS, J. & TAYLOR, S. 2016. Major histocompatibility complex polymorphism in reptile conservation. *Herpetological Conservation and Biology,* 11**,** 1-12.

EVANS, M., DIONNE, M., MILLER, K. & BERNATCHEZ, L. 2012. Mate choice for major histocompatibility complex genetic divergence as a bet-hedging strategy in the Atlantic salmon (*Salmo salar*). *Proceedings of the Royal Society B: Biological Sciences,* 279**,** 379-386.

FLAJNIK, M., OHTA, Y., NAMIKAWA-YOMADA, C. & NONAKA, M. 1999. Insight into the primordial MHC from studies in ectothermic vertebrates. *Immunological Reviews,* 167.

GARCIA, M., YEBRA, B., FLORES, A. & GUERRA, E. 2012. The Major Histocompatibility Complex in Transplantation. *Journal of Transplantation,* 2012**,** 1-7.

GIBBONS, J. 1979. The hind leg pushup display of the *Amphibolurus decresii* species complex (Lacertilia: Agamidae). *Copeia,* 1**,** 29-40.

GLABERMAN, S., DU PASQUIER, L. & CACCONE, A. 2009. Characterization of a Nonclassical Class I MHC Gene in a Reptile, the Galapagos Marine Iguana (*Amblyrhynchus cristatus*). *PLOS ONE,* 3**,** e2859.

GROGAN, K. E., SAUTHER, M. L., CUOZZO, F. P. & DREA, C. M. 2017. Genetic wealth, population health: Major histocompatibility complex variation in captive and wild ring-tailed lemurs (*Lemur catta*). *Ecology and Evolution,* 7**,** 7638-7649.

GRUEN, J. R. & WEISSMAN, S. M. 2001. Human MHC class III and IV genes and disease associations. *Frontiers in Bioscience,* 6**,** D960-72.

GUSTAFSON, A., TALLMADGE, R., RAMLACHAN, N., MILLER, D., BIRD, H., ANTCZAK, D., RAUDSEPP, T., CHOWDHARY, B. & SKOW, L. 2003. An ordered BAC contig map of the equine major histocompatibility complex. *Cytogenetic and genome research,* 102**,** 189-195.

HAMILTON, W. D. & ZUK, M. 1982. Heritable true fitness and bright birds: a role for parasites? *Science,* 218**,** 384-7.

HARSTAD, H., LUKACS, M. F., BAKKE, H. G. & GRIMHOLT, U. 2008. Multiple expressed MHC class II loci in salmonids; details of one non-classical region in Atlantic salmon (*Salmo salar*). *BMC Genomics,* 9**,** 193.

HOSOMICHI, K., SHIINA, T., SUZUKI, S., TANAKA, M., SHIMIZU, S., IWAMOTO, S., HARA, H., YOSHIDA, Y., KULSKI, J. K., INOKO, H. & HANZAWA, K. 2006. The major histocompatibility complex (Mhc) class IIB region has greater genomic structural flexibility and diversity in the quail than the chicken. *BMC Genomics,* 7**,** 322.

JANEWAY, C. J., TRAVERS, P. & WALPORT, M. 2001. *Immunobiology: The Immune System in Health and Disease,* New York, Garland Science.

JONES, M. R., CHEVIRON, Z. A. & CARLING, M. D. 2015. Spatially variable coevolution between a haemosporidian parasite and the MHC of a widely distributed passerine. *Ecology and Evolution,* 5**,** 1045-1060.

JUOLA, F. A. & DEARBORN, D. C. 2012. Sequence-based evidence for major histocompatibility complex-disassortative mating in a colonial seabird. *Proceedings of the Royal Society B: Biological Sciences,* 279**,** 153-62.

KELLEY, J., WALTER, L. & TROWSDALE, J. 2005. Comparative genomics of major histocompatibility complexes. *Immunogenetics,* 56**,** 683-695.

KLOCH, A., BABIK, W., BAJER, A., SIŃSKI, E. & RADWAN, J. 2010. Effects of an MHC-DRB genotype and allele number on the load of gut parasites in the bank vole *Myodes glareolus*. *Molecular Ecology,* 19**,** 255-265.

LEI, W., FANG, W., LIN, Q., ZHOU, X. & CHEN, X. 2015. Characterization of a non-classical MHC class II gene in the vulnerable Chinese egret (*Egretta eulophotes*). *Immunogenetics,* 67**,** 463-472.

LEI, W., ZHOU, X., FANG, W., LIN, Q. & CHEN, X. 2016. Major histocompatibility complex class II DAB alleles associated with intestinal parasite load in the vulnerable Chinese egret (*Egretta eulophotes*). *Ecology and Evolution,* 6**,** 4421-4434.

LEWIS, A. C., RANKIN, K. J., PASK, A. J. & STUART-FOX, D. 2017. Stress-induced changes in color expression mediated by iridophores in a polymorphic lizard. *Ecology and Evolution,* 7**,** 8262-8272.

LIGHTEN, J., OOSTERHOUT, C. & BENTZEN, P. 2014a. Critical review of NGS analyses for *de novo* genotyping multigene families. *Molecular Ecology,* 2014**,** 3957–3972.

LIGHTEN, J., OOSTERHOUT, C., PATERSON, I., MCMULLAN, M. & BENTZEN, P. 2014b. Ultra-deep Illumina sequencing accurately identifies MHC class IIb alleles and provides evidence for copy number variation in the guppy (*Poecilia reticulata*). *Molecular Ecology Resources,* 2014**,** 753–767.

LIGHTEN, J., PAPADOPULOS, A. S. T., MOHAMMED, R. S., WARD, B. J., G. PATERSON, I., BAILLIE, L., BRADBURY, I. R., HENDRY, A. P., BENTZEN, P. & VAN OOSTERHOUT, C. 2017. Evolutionary genetics of immunological supertypes reveals two faces of the Red Queen. *Nature Communications,* 8**,** 1294.

LILLIE, M., CUI, J., SHINE, R. & BELOV, K. 2016. Molecular characterization of MHC class II in the Australian invasive cane toad reveals multiple splice variants. *Immunogenetics,* 68**,** 449-460.

MCCONNELL, S. & RESTAINO, A. D. J., JL 2014. Multiple divergent haplotypes express completely distinct sets of class I MHC genes in zebrafish. *Immunogenetics,* 66**,** 199-213.

MCLEAN, C. A., LUTZ, A., RANKIN, K. J., STUART-FOX, D. & MOUSSALLI, A. 2017. Revealing the biochemical and genetic basis of color variation in a polymorphic lizard. *Molecular Biology and Evolution,* 34**,** 1924-1935.

MCLEAN, C. A., MOUSSALLI, A. & STUART-FOX, D. 2014a. Local adaptation and divergence in colour signal conspicuousness between monomorphic and polymorphic lineages in a lizard. *Journal of Evolutionary Biology,* 27.

MCLEAN, C. A., STUART-FOX, D. & MOUSSALLI, A. 2014b. Phylogeographic structure, demographic history and morph composition in a colour polymorphic lizard. *Journal of Evolutionary Biology,* 27**,** 2123-2137.

MCLEAN, C. A., STUART-FOX, D. & MOUSSALLI, A. 2015. Environment, but not genetic divergence, influences geographic variation in colour morph frequencies in a lizard. *BMC Evol Biol,* 15.

MEYER-LUCHT, Y., OTTEN, C., PÜTTKER, T. & SOMMER, S. 2008. Selection, diversity and evolutionary patterns of the MHC class II DAB in free-ranging Neotropical marsupials. *BMC Genetics,* 9**,** 39.

MICHALOVÁ, V., MURRAY, B., SÜLTMANN, H. & KLEIN, J. 2000. A contig map of the Mhc class I genomic region in the zebrafish reveals ancient synteny. *Journal of immunology (Baltimore, Md. : 1950),* 164**,** 5296-5305.

MILLER, H. C., MOORE, J. A., NELSON, N. J. & DAUGHERTY, C. H. 2009. Influence of major histocompatibility complex genotype on mating success in a free-ranging reptile population. *Proceedings of the Royal Society B: Biological Sciences,* 276**,** 1695-704.

NEEFJES, J., JONGSMA, M., PAUL, P. & BAKKE, O. 2011. Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nature Reviews Immunology,* 11**,** 823-836.

NEI, M., GU, X. & SITNIKOVA, T. 1997. Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *PNAS,* 94.

NEI, M. & ROONEY, A. P. 2005. Concerted and birth-and-death evolution of multigene families. *Annual Review of Genetics,* 39**,** 121-152.

NOWAK, M. A., TARCZY-HORNOCH, K. & AUSTYN, J. M. 1992. The optimal number of major histocompatibility complex molecules in an individual. *PNAS,* 89**,** 10896-10899.

OLSSON, M., MADSEN, T., NORDBY, J., WAPSTRA, E., UJVARI, B. & WITTSELL, H. 2003. Major histocompatibility complex and mate choice in sand lizards. *Proceedings of the Royal Society B: Biological Sciences,* 2**,** S254-6.

OSBORNE, L. 2005. Information content of male agonistic displays in the territorial tawny dragon (*Ctenophorus decresii*). *Journal of Ethology,* 23**,** 189-197.

OSBORNE, L., UMBERS, K. D., BACKWELL, P. R. & KEOGH, J. S. 2012. Male tawny dragons use throat patterns to recognize rivals. *Naturwissenschaften,* 99**,** 869-872.

OSBORNE, M. J., PILGER, T. J., LUSK, J. D. & TURNER, T. F. 2017. Spatio-temporal variation in parasite communities maintains diversity at the major histocompatibility complex class IIβ in the endangered Rio Grande silvery minnow. *Molecular Ecology,* 26**,** 471-489.

OSORIO, D. & VOROBYEV, M. 2005. Photoreceptor sectral sensitivities in terrestrial animals: adaptations for luminance and colour vision. *Proceedings of the Royal Society B: Biological Sciences,* 272**,** 1745-1752.

PEARSON, S. K., GODFREY, S. S., SCHWENSOW, N., BULL, C. M. & GARDNER, M. G. 2017. Genes and Group Membership Predict Gidgee Skink (*Egernia stokesii*) Reproductive Pairs. *Journal of Heredity,* 108**,** 369-378.

PIERTNEY, S. & OLIVER, M. 2006. The evolutionary ecology of the major histocompatibility complex. *Heredity,* 96**,** 7-21.

PRATHEEK, B., NAYAK, T. K., SAHOO, S. S., MOHANTY, P. K., CHATTOPADHYAY, S., CHAKRABORTY, N. G. & CHATTOPADHYAY, S. 2014. Mammalian non-classical major histocompatibility complex I and its receptors: Important contexts of gene, evolution, and immunity. *Indian Journal of Human Genetics,* 20**,** 129.

RADWAN, J., KUDUK, K., LEVY, E., LEBAS, N. & BABIK, W. 2014. Parasite load and MHC diversity in undisturbed and agriculturally modified habitats of the ornate dragon lizard. *Molecular Ecology,* 23**,** 5966-5978.

RADWAN, J., ZAGALSKA-NEUBAUER, M., CICHON, M., SENDECKA, J., KULMA, K., GUSTAFSSON, L. & BABIK, W. 2012. MHC diversity, malaria and lifetime reproductive success in collared flycatchers. *Molecular ecology,* 21**,** 2469-79.

RANKIN, K. & STUART-FOX, D. 2015. Testosterone-induced expression of male colour morphs in females of the polymorphic tawny dragon lizard, *Ctenophorus decresii*. *PLOS ONE,* 10.

RANKIN, K. J., MCLEAN, C. A., KEMP, D. J. & STUART-FOX, D. 2016. The genetic basis of discrete and quantitative colour variation in the polymorphic lizard, *Ctenophorus decresii*. *BMC Evolutionary Biology,* 16**,** 179.

REED, K., MENDOZA, K. & SETTLAGE, R. 2016. Targeted capture enrichment and sequencing identifies extensive nucleotide variation in the turkey MHC-B. *Immunogenetics,* 68**,** 219-229.

REUSCH, T. B. H., HÄBERLI, M. A., AESCHLIMANN, P. B. & MILINSKI, M. 2001. Female sticklebacks count alleles in a strategy of sexual selection explaining MHC polymorphism. *Nature,* 414.

ROCK, K. L., REITS, E. & NEEFJES, J. 2016. Present Yourself! By MHC Class I and MHC Class II Molecules. *Trends in Immunology,* 37**,** 724-737.

SAMBROOK, J., RUSSELL, R., UMRANIA, Y., EDWARDS, Y., CAMPBELL, D., ELGAR, G. & CLARK, M. 2002. Fugu orthologues of human major histocompatibility complex genes: a genome survey. *Immunogenetics,* 54**,** 367–380.

SCHAD, J., DECHMANN, D. K., VOIGT, C. C. & SOMMER, S. 2012. Evidence for the 'good genes' model: association of MHC class II DRB alleles with ectoparasitism and reproductive state in the neotropical lesser bulldog bat, *Noctilio albiventris*. *PLOS ONE,* 7**,** e37101.

SCHRIDER, D. & HAHN, M. 2010. Gene copy-number polymorphism in nature. *Proceedings of the Royal Society B: Biological Sciences,* 2010**,** 3213-3221.

SCHWENSOW, N., MAZZONI, C. J., MARMESAT, E., FICKEL, J., PEACOCK, D., KOVALISKI, J., SINCLAIR, R., CASSEY, P., COOKE, B. & SOMMER, S. 2017. High adaptive variability and virus-driven selection on major histocompatibility complex (MHC) genes in invasive wild rabbits in Australia. *Biological Invasions,* 19**,** 1255-1271.

SHIINA, T., SHIMIZU, S., HOSOMICHI, K., KOHARA, S., WATANABE, S., HANZAWA, K., BECK, S., KULSKI, J. & INOKO, H. 2004. Comparative genomic analysis of two avian (quail and chicken) MHC regions. *Journal of immunology (Baltimore, Md. : 1950),* 172**,** 6751-6763.

SIDDLE, H. V., MARZEC, J., CHENG, Y., JONES, M. & BELOV, K. 2010. MHC gene copy number variation in Tasmanian devils: implications for the spread of a contagious cancer. *Proceedings of the Royal Society B: Biological Sciences,* 277**,** 2001-6.

SMITH, K. 2012. Toward a molecular understanding of adaptive immunity: a chronology, part I. *Frontiers in immunology,* 3**,** 369.

SMITH, S., BELOV, K. & HUGHES, J. 2009. MHC screening for marsupial conservation: extremely low levels of class II diversity indicate population vulnerability for an endangered Australian marsupial. *Conservation Genetics,* 11**,** 269-278.

SOMMER, S. 2005. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Frontiers in Zoology,* 2**,** 16.

SPURGIN, L. G. & RICHARDSON, D. S. 2010. How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proceedings of the Royal Society B: Biological Sciences,* 277**,** 979-88.

STAFUZZA, N., GRECO, A., GRANT, J., ABBEY, C., GILL, C., RAUDSEPP, T., SKOW, L., WOMACK, J., RIGGS, P. & AMARAL, M. 2013. A high-resolution radiation hybrid map of the river buffalo major histocompatibility complex and comparison with BoLA. *Animal genetics,* 44**,** 369-376.

STRANDH, M., WESTERDAHL, H., PONTARP, M., CANBÄCK, B., DUBOIS, M.-P. & MIQUEL, C. 2012. Major histocompatibility complex class II compatibility, but not class I, predicts mate choice in a bird with highly developed olfaction. *Proceedings of the Royal Society B: Biological Sciences,* 279.

TAKAHATA, N. & NEI, M. 1990. Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. *Genetics,* 124.

TEASDALE, L., STEVENS, M. & STUART-FOX, D. 2013. Discrete colour polymorphism in the tawny dragon lizard (*Ctenophorus decresii*) and differences in signal conspicuousness among morphs. *Journal of Evolutionary Biology,* 26.

TOLLENAERE, C., BRYJA, J., GALAN, M., CADET, P., DETER, J., CHAVAL, Y., BERTHIER, K., RIBAS SALVADOR, A., VOUTILAINEN, L., LAAKKONEN, J., HENTTONEN, H., COSSON, J. F. & CHARBONNEL, N. 2008. Multiple parasites mediate balancing selection at two MHC class II genes in the fossorial water vole: insights from multivariate analyses and population genetics. *Journal of Evolutionary Biology,* 21**,** 1307-20.

UJVARI, B. & BELOV, K. 2011. Major Histocompatibility Complex (MHC) Markers in Conservation Biology. *International Journal of Molecular Sciences,* 12.

VANHANEN, R., HEIKKILÄ, N., AGGARWAL, K., HAMM, D., TARKKILA, H., PÄTILÄ, T., JOKIRANTA, T. S., SARAMÄKI, J. & ARSTILA, T. P. 2016. T cell receptor diversity in the human thymus. *Molecular Immunology,* 76**,** 116-122.

VOORTER, C. E. M., GERRITSEN, K. E. H., GROENEWEG, M., WIETEN, L. & TILANUS, M. G. J. 2016. The role of gene polymorphism in HLA class I splicing. *International Journal of Immunogenetics,* 43**,** 65-78.

WEDEKIND, C. & FURI, S. 1997. Body odour preferences in men and women: do they aim for specific MHC combinations or simply heterozygosity? *Proceedings of the Royal Society B: Biological Sciences,* 264**,** 1471-1479.

WEGNER, K. M., KALBE, M., SCHASCHL, H. & REUSCH, T. B. 2004. Parasites and individual major histocompatibility complex diversity - an optimal choice? *Microbes and Infection,* 6**,** 1110-6.

WINTERNITZ, J. C., MINCHEY, S. G., GARAMSZEGI, L. Z., HUANG, S., STEPHENS, P. R. & ALTIZER, S. 2013. Sexual selection explains more functional variation in the mammalian major histocompatibility complex than parasitism. *Proceedings of the Royal Society B: Biological Sciences,* 280.

WOELFING, B., TRAULSEN, A., MILINSKI, M. & BOEHM, T. 2009. Does intra-individual major histocompatibility complex diversity keep a golden mean? *Philosophical Transactions of the Royal Society B: Biological Sciences,* 364**,** 117-128.

YEWERS, M. S., MCLEAN, C. A., MOUSSALLI, A., STUART-FOX, D., BENNETT, A. T. D. & KNOTT, B. 2015. Spectral sensitivity of cone photoreceptors and opsin expression in two colour-divergent lineages of the lizard *Ctenophorus decresii*. *The Journal of Experimental Biology,* 218**,** 1556-1563.

YEWERS, M. S. C. 2016. *The function and evolution of colour polymorphism in the tawny dragon lizard.* Doctor of Philosophy PhD, The University of Melbourne.

YEWERS, M. S. C., PRYKE, S. & STUART-FOX, D. 2016. Behavioural differences across contexts may indicate morph-specific strategies in the lizard *Ctenophorus decresii*. *Anim Behav,* 111.

ZAGALSKA-NEUBAUER, M., BABIK, W., STUGLIK, M., GUSTAFSSON, L., CICHON, M. & RADWAN, J. 2010. 454 sequencing reveals extreme complexity of the class II Major Histocompatibility Complex in the collared flycatcher. *BMC Evolutionary Biology,* 10**,** 395.

CHAPTER 2

Characterisation of major histocompatibility complex class I transcripts in an Australian dragon lizard

Note to examiners

This chapter reports the characterisation of major histocompatibility complex (MHC) class I transcripts in the tawny dragon lizard (*Ctenophorus decresii*) and investigates the evolution of MHC class I in Iguanian lizards. Characterising the MHC in *C. decresii* provides a foundation for future work and allows for comparisons to be made with thoroughly studied MHC genes in other taxa. This chapter has been published in *Developmental and Comparative Immunology* and is therefore formatted according to the specifications of that journal. Please refer to Appendix 1 for additional details regarding transcriptome assembly and primer design, and Appendix 2 for a copy of the published manuscript. Appendix 1 is a thesis-only appendix and is therefore not present in the publication associated with this chapter.

Hacking, J., Bertozzi, T., Moussalli, A., Bradford, T., Gardner, M., (2018). Characterisation of major histocompatibility complex class I transcripts in an Australian dragon lizard. *Developmental & Comparative Immunology, 84, 164-171.*

Characterisation of major histocompatibility complex class I transcripts in an Australian dragon lizard

Jessica Hacking[a] (corresponding author)
jessica.hacking@flinders.edu.au

Terry Bertozzi[b,c]
terry.bertozzi@samuseum.sa.gov.au

Adnan Moussalli[d]
amoussalli@museum.vic.gov.au

Tessa Bradford[a,b,c]
tessa.bradford@samuseum.sa.gov.au

Michael Gardner[a,b]
michael.gardner@flinders.edu.au

[a]College of Science and Engineering, Flinders University, Bedford Park SA 5042, Australia

[b]Evolutionary Biology Unit, South Australian Museum, Adelaide SA 5000, Australia

[c]School of Biological Sciences, University of Adelaide, Adelaide, SA 5005, Australia

[d]Sciences Department, Museum Victoria, Carlton Gardens VIC 3053, Australia

Abstract

Characterisation of squamate major histocompatibility complex (MHC) genes has lagged behind other taxonomic groups. MHC genes encode cell-surface glycoproteins that present self- and pathogen-derived peptides to T cells and play a critical role in pathogen recognition. Here we characterize MHC class I transcripts for an agamid lizard (*Ctenophorus decresii*) and investigate the evolution of MHC class I in Iguanian lizards. An iterative assembly strategy was used to identify six full-length *C. decresii* MHC class I transcripts, which were validated as likely to encode classical class I MHC molecules. Evidence for exon shuffling recombination was uncovered for *C. decresii* transcripts and Bayesian phylogenetic analysis of Iguanian MHC class I sequences revealed a pattern expected under a birth-and-death mode of evolution. This work provides a stepping stone towards further research on the agamid MHC class I region.

Keywords: transcriptome assembly; Iguania, Agamidae; Ctenophorus decresii; MHC class I evolution

Introduction

The major histocompatibility complex (MHC) is a multigene family involved in pathogen recognition and immune response, and is one of the most diverse regions of the vertebrate genome (Piertney and Oliver, 2006). MHC genes encode cell surface glycoproteins that present self- and foreign-derived peptides to circulating T-lymphocyte cells (T cells). The evolution of MHC genes is complex and is thought to be governed primarily by the birth-and-death model of evolution in which loci are duplicated or lost, although concerted evolution via inter-locus gene conversion events may also play a role (Edwards and Hedrick, 1998; Nei and Rooney, 2005; Spurgin et al., 2011). These processes can occur over short time scales and it is apparent that MHC genes have undergone numerous independent expansion and diversification events throughout vertebrate evolution (Nei et al., 1997). The MHC is gene rich and is generally extremely polymorphic within loci (Janeway et al., 2001). Pathogen-mediated natural selection and sexual selection (MHC-associated mating) are considered to be the primary mechanisms maintaining these extraordinary levels of diversity (Edwards and Hedrick, 1998; Ejsmond et al., 2014; Milinski, 2006).

The MHC is divided into four classes based primarily on structural and functional differences (Janeway et al., 2001). Genes belonging to classes I and II are further separated into classical or non-classical genes based primarily on function and expression patterns (Alfonso and Karlsson, 2000; Janeway et al., 2001). The structure of classical MHC class I (hereafter MHC I) molecules is conserved among jawed vertebrates and includes a leader peptide, three α domains, and the transmembrane and cytoplasmic (Tm/Cyt) domains, all of which are encoded by a single gene (Kaufman et al., 1994). The α1 and α2 domains form the peptide binding cleft and contain amino acid positions that are directly involved in peptide binding, termed peptide binding regions (PBR) (Janeway et al., 2001). Classical MHC I molecules are anchored to the surface of somatic cells via the Tm/Cyt domains and display self-peptides and antigenic peptides derived primarily from intracellular pathogens to cytotoxic T cells. When a particular MHC molecule presents an antigenic peptide and is recognized by a T cell, an immune response is initiated, which usually involves lysis of the infected cell (Neefjes et al., 2011). Non-classical MHC class I genes are distinguished from classical genes by low levels of allelic variation and restricted expression (Janeway et al., 2001). In mammals some non-classical MHC class I genes undertake important roles within the immune system, both at the cell surface and in secreted forms (Adams and Luoma, 2013).

The MHC has been thoroughly characterised in humans and model organisms, primarily due to the critical role that the region plays in organ and tissue transplantation (Garcia et al., 2012). MHC genes are also used in conservation genetic studies as a measure of population genetic health and adaptive potential (Sommer, 2005). However, many vertebrate groups, especially non-avian reptiles, are under-

represented within the MHC literature and little is known regarding the mechanisms shaping MHC diversity in these taxa. The tawny dragon (*Ctenophorus decresii*) is a small (<30g) agamid lizard endemic to South Australia and provides a promising model system in which to investigate the mechanisms shaping MHC diversity. Male *C. decresii* use visual cues during social and sexual interactions (Gibbons, 1979; Osborne, 2005a, b; Yewers et al., 2016) and the species is host to external and intracellular parasites (Hacking et al. unpublished results); providing opportunities to investigate the roles of sexual selection and parasite-mediated selection in maintaining MHC diversity. Here, we characterised MHC I transcripts for *C. decresii* and investigated the evolutionary mechanisms playing a role in the generation of MHC I diversity within Iguanian lizards (Iguanidae, Agamidae, Chamaeleonidae, Dactyloidae and related families, Pyron et al., 2013).

Materials and methods

Sample collection

A single *C. decresii* individual was captured in Burra, South Australia (33°40'57.7"S, 138°56'16.8"E) in October 2012 and taken directly to Adelaide to be euthanized for tissue collection. Burra is located just north of the contact zone between the northern and southern clades of this species (McLean et al., 2014). The thymus and spleen were collected immediately after euthanasia and were stored separately in RNA Later (Qiagen, Venlo, Netherlands) at 4°C for 48 h and then at -80°C until required for RNA extraction. The remainder of the specimen was accessioned into the South Australian Museum herpetology collection (SAMAR67384).

Transcriptome sequencing and MHC class I discovery

Total RNA was extracted using the Qiagen RNeasy mini kit. Sequencing libraries were then prepared using the TruSeq RNA Kit v1 using a polyA purification. These were then multiplexed with two other samples and sequenced (100 bp paired end) on a single lane of the HISEQ 2000. Extractions, library preparation and sequencing were carried out by Georgia Genomic Facility (GGF, University of Georgia, USA). Adaptor sequences and low quality reads were removed or trimmed using Trimmomatic ver. 0.22 (Bolger et al., 2014), with a minimum quality Phred score of 25 per 4bp sliding window and a minimum sequence length of 40bp. Assemblies were then constructed from the trimmed and filtered reads for each sample separately using the program Trinity v1 (r2013-02-25) (Haas et al., 2013) with default settings, followed by an assessment of gene completeness using BUSCO v1.22 (Simao et al., 2015) based on the OrthoDB 'vertebrata' database. Lastly, to identify putative *C. decresii* MHC I transcripts we performed local BLASTX (E-value ≤ 1e-10) searches (Altschul et al., 1997; Camacho et al., 2009) using predicted *Pogona vitticeps* MHC gene models (Georges et al., 2015) as our reference. Putative *C.*

decresii MHC I transcripts were aligned manually with published MHC I sequences (table S1) to confirm expected MHC I structure and the presence of conserved sites (Kaufman et al., 1994).

Due to the high diversity and complex structure of the MHC region, traditional assembly methods may not be sufficient to obtain a robust assembly. To refine the MHC I assemblies, 75bp sequence fragments congruent with the putative antigen binding α2 domain for each unique sequence were iteratively re-assembled using the mirabait utility from MIRA v4.0.2 (Chevreux et al., 1999) as presented in Ansari et al. (2015) but using a kmer length of 31 (size of the search string) and requiring 50 matching kmers (number of matching search strings). Sequences were extended until sequence length stabilized or it was no longer possible to uniquely map reads. The resulting contigs were evaluated by concatenating each separated by 200 Ns and remapping cleaned reads using BWA (Li, 2013) with default settings and visualizing the resultant BAM file in IGV (Robinson et al., 2011). Read pairs spanning sequences were checked for accuracy and suspected chimeric reads removed.

Validation of MHC I transcripts and comparison with other vertebrates

Putative *C. decresii* MHC class I transcripts were translated and aligned with a subset of published full length MHC I amino acid sequences of other vertebrates (table S1) using MUSCLE (Edgar, 2004) implemented in MEGA ver. 6.06 (Tamura et al., 2011). The alignment was manually refined to ensure correct alignment of conserved regions. Coding domain boundaries were defined as per Koller and Orr (1985). Aligned *C. decresii* transcripts were validated as likely MHC I sequences by (i) confirming MHC I gene structure (leader peptide, α domains and Tm/Cyt domains), (ii) confirming concordance with known conserved regions and regions with predicted function that are typical of MHC I sequences (Kaufman et al., 1994), and iii) confirming the absence of stop codons within coding regions. Two additional steps needed to further validate transcripts as likely classical MHC I sequences, which were beyond the scope of this study, are confirming polymorphism among individuals, and strong and widespread expression. Pairwise nucleotide and amino acid identity among validated MHC class I transcript sequences was calculated using Geneious ver. 8.1.7 (Kearse et al., 2012). Validated *C. decresii* MHC class I transcripts were named according to Klein et al. (1990); each unique nucleotide sequence was given the species identification prefix (Ctde) followed by U (Uno; class 1) and A (locus group/family designation), and a unique number (e.g. *Ctde-UA*001*). Once full-length MHC I genomic data are available for *C. decresii* locus designations may be defined (i.e. UA1 and UA2).

To investigate relationships among *C. decresii* MHC I transcripts and their position relative to other Iguanian lizards a Bayesian phylogenetic tree was constructed based on an alignment of validated full-length *C. decresii* MHC class I transcripts and all full-length Iguanian MHC I sequences available in GenBank (NCBI Resource Coordinators, 2016) (table S1). All sequences obtained from GenBank were

validated as likely MHC I sequences via confirmation of expected MHC I structure and conserved sites. All squamate MHC I nucleotide sequences were translated before aligning with Muscle, implemented in MEGA ver. 6.06 and then untranslated for phylogenetic analysis. Tuatara (*Sphenodon punctatus*) MHC I sequence was used as an outgroup. Only the three α domains (α1, α2 and α3) of these full-length sequences were used in phylogenetic analysis due to extreme variation at leader and Cyt/Tm domains inhibiting alignment. To determine optimal partitioning and the best model of evolution for Bayesian phylogenetic analysis, PartitionFinder2 (Guindon and Gascuel, 2003; Lanfear et al., 2012; Lanfear et al., 2017) was employed. Sequences were split into three data blocks representing codon positions and only models employed by MrBayes were considered, with the best model determined using AICc model selection. Bayesian phylogenetic analyses were undertaken in MrBayes ver. 3.2.6 (Ronquist et al., 2011), with one analysis employing the model of evolution and partitioning identified by PartitionFinder2, and another using mixed models with sequences partitioned by codon position. For both MrBayes analyses, two independent runs were performed, each with four Markov chains run for 20 million generations at a sample frequency of 1000 and a default burn-in period of 25%. Convergence diagnostics, including the standard deviation of split frequencies between runs, the potential scale reduction factor (PSRF) and the average effective sample size (ESS) were examined to confirm run convergence. FigTree ver. 1.4.2 (Rambaut, 2012) was used to annotate trees produced by MrBayes.

Results and Discussion

Transcriptome sequencing and MHC class I discovery

RNA extractions had RIN values of 9.1 and 9.8 for the thymus and spleen samples, respectively. Approximately 49 million paired-end reads were obtained for the thymus and approximately 51 million paired-end reads were obtained for the spleen. Of these reads, approximately 86% survived Trimmomatic filtering. Based on the Busco analysis, 73% and 76% of the vertebrata orthologous gene set were recovered and complete for the thymus (C:73%[D:15%],F:7.7%,M:19%,n:3023) and spleen (C:76%[D:13%],F:6.4%,M:16%,n:3023) transcriptomes, respectively. Combing the two transcriptome assemblies resulted in 82% of the vertebrata gene set being complete (C:82%[D:43%],F:5.1%,M:12%,n:3023). In total, eight putative *C. decresii* MHC I transcripts (*Ctde-UA*001 – Ctde-UA*008*) were discovered, based on unique α2 and α3 domains, indicating the presence of at least four different loci. We were unable to confidently assign transcripts to specific loci based on data from a single individual. As a result, when naming transcripts all sequences were designated the letters 'UA' (MHC I locus group 'A') and no specific locus numbers were assigned (i.e. UA1, UA2, UA3 etc).

Validation of C. decresii MHC I transcripts

Of the eight putative *C. decresii* MHC I transcripts, six were validated as likely MHC I sequences after confirming (i) normal MHC I gene structure, (ii) concordance with known conserved regions and regions with predicted function that are typical of MHC I sequences, and (iii) the absence of stop codons within coding regions (*Ctde-UA*001 – Ctde-UA*006*). Each of these transcripts contained complete leader peptide, α domains and Tm/Cyt domains and did not contain any premature stop codons. An amino acid alignment of putative *C. decresii* MHC I transcripts with published squamate, tuatara (*Sphenodon punctatus*) and human sequences, confirmed concordance with conserved regions and regions with predicted functions that are typical of MHC I molecules (table S2, fig. 1, Kaufman et al., 1994). Specifically, nine amino acid positions that bind C- and N-terminal residues of antigenic peptides are highly conserved among classical MHC class I molecules (Kaufman et al., 1994). All nine of these positions were identified within *C. decresii* MHC I transcripts and displayed conserved amino acids or low (≤2 changes) amino acid variability (site numbers: 45, 101, 126, 167, 187, 190, 191, 204 and 219). Similarly, amino acid positions involved in salt bridge formation within α1 and α2 were found to be conserved, with the histidine (H) residues at site 41 and 137, and the aspartic acid (D)/glutamic acid (E) residues at sites 69 and 163. The four cysteine (C) residues involved in intra-domain disulphide bridge formation within α2 and α3 (site numbers: 145, 209, 249 and 305) were conserved across all taxa included in the amino acid alignment. Finally, most vertebrates possess an NQS or NQT nitrogen-linked glycosylation acceptor site near the end of the α1 domain (Kaufman et al., 1994). Nitrogen-linked glycans play an important role in the folding and stability of classical, and likely also non-classical, MHC class I molecules (Ryan and Cobb, 2012, 2015). The amino acid sequence for nitrogen-linked glycosylation site consists of an asparagine (N), any other amino acid except proline, followed by serine (S) or threonine (T) (Ryan and Cobb, 2015). Most of the *C. decresii* transcripts encoded NQS glycosylation sites, with three transcripts (*Ctde-UA*005* and *Ctde-UA*006*) encoding NHS instead. Together, these findings suggest that transcripts *Ctde-UA*001 – Ctde-UA*006* produce molecules that undertake classical MHC I functions at the cell surface.

Classical MHC I gene structure (leader peptide, α domains and Tm/Cyt domains) could not be confirmed for two *C. decresii* putative MHC I transcripts; *Ctde-UA*007* and *Ctde-UA*008*. Both *Ctde-UA*007* and *Ctde-UA*008* ended prematurely; *Ctde-UA*007* was missing the 5' end of the α3 domain and all of the Tm/Cyt domains and *Ctde-UA*008* was missing both the α3 and the Tm/Cyt domains. Both *Ctde-UA*007* and *Ctde-UA*008* possessed the expected conserved sites associated with peptide binding at the α1 and α2 domains and did not contain any premature stop codons (fig. 1). These sequences could not be unambiguously resolved during transcriptome assembly, probably due to similarity with the other transcripts. This is not surprising given that MHC loci arise by periodic gene

duplication and rearrangement events (Nei and Rooney, 2005). It is also possible that these transcripts are naturally truncated and encode non-classical MHC I molecules that are secreted and undertake peptide presentation away from the cell surface, given the lack of Tm/Cyt domains and the presence of conserved sites associated with peptide presentation (Carlini et al., 2016; Donadi et al., 2011; Glaberman et al., 2009). Given that *Ctde-UA*007* and *Ctde-UA*008* are likely truncated due to incomplete transcriptome assembly they were not assigned to a new locus group and were included in the 'UA' MHC I locus group.

Average nucleotide percent identity among the validated full-length *C. decresii* MHC I transcripts was 79.8%, with this differentiation primarily driven by variation at the α1 and α2 domains (table S2). Nucleotide identity within α1 and α2 was 81% and 82%, respectively, whereas nucleotide identity within α3 was 88%. Sequences *Ctde-UA*001* and *Ctde-UA*002* were identical only at the α1 domain, suggesting the possibility for exon shuffling recombination in the *C. decresii* MHC I region. Exon shuffling recombination occurs between entire exon regions with breaks within introns and has been found to occur in a range of vertebrates (Holmes and Parham, 1985; Wang et al., 2010; Zhao et al., 2013). Overall percent nucleotide identity between *Ctde-UA*001* and *Ctde-UA*002* is 95.6% and between *Ctde-UA*003* and *Ctde-UA**004 it is 95.1%. Nucleotide identity is slightly lower between *Ctde-UA*005* and *Ctde-UA*006* at 93.7%. The high similarity between these pairs of transcripts indicates that they may be alleles at the same loci, suggesting five loci in total. Bayesian phylogenetic analysis of *C. decresii* transcripts (see section 3.3 for further details) suggest the presence of five or six loci. *Ctde-UA*001* and *Ctde-UA*002,* and *Ctde-UA*003* and *Ctde-UA**004 each clustered together but *Ctde-UA*005* and *Ctde-UA*006* didn't, with *Ctde-UA*006* located between *Ctde-UA*005* and a *P. vitticeps* sequence.

Phylogenetic analysis of Iguanian MHC class I genes

One of the Bayesian phylogenetic analyses of MHC I sequences was conducted with no partitioning of codon positions and a GTR+G model of evolution, according to PartitionFinder2 results. The other analysis was conducted using data partitioned by codon position and allowing mixed models of evolution. For both Bayesian analyses, convergence diagnostics indicated run convergence and high posterior probability values were obtained across most of the tree. The two resulting trees had identical topology but the tree resulting from the PartitionFinder2-informed analysis had slightly better overall posterior probability values and was therefore retained. Posterior probabilities were high (>0.95) for most branch nodes, but low for some outer nodes (fig. 2).

Full-length MHC I sequences were only available for three Iguania species (*Amblyrhynchus cristatus*, *Conolophus subcristatus* and *Iguana iguana*), one Dactyloidae species (*Anolis carolinensis*) and two Agamidae species (C*. decresii* and *Pogona vitticeps*, fig. 2, table S1). No orthologous relationships were

observed within the Iguanidae clade, representing three genera that diverged 10 to 20 million years ago (Rassmann, 1997). Instead, MHC I sequences clustered by species, suggesting loss of ancestral diversity, recent gene duplications and potentially gene conversion events (concerted evolution, Glaberman and Caccone, 2008). In contrast, Agamidae MHC I sequences (*C. decresii* and *P. vitticeps*) displayed orthologous relationships despite similar divergence times to the iguanid species (Hugall et al., 2008). Given divergence of approximately 20 million years between *P. vitticeps* and *C. decresii* (Hugall et al., 2008), the separation of the agamid MHC I sequences into three orthologous clades suggests that at least two gene duplication events occurred greater than 20 million years ago (fig. 2). The conservation of MHC class I loci over such a time scale is not unusual; some primate MHC I loci have been conserved for at least 46-66 million years (Piontkivska and Nei, 2003). The phylogenetic relationships within the Iguanidae and Agamidae families are consistent with the birth-and-death model of evolution; the continual gain and loss of genes (Edwards and Hedrick, 1998; Nei and Rooney, 2005). Concerted evolution via gene conversion may also play a role in generating diversity within Iguanidae and Agamidae, as evidenced by the patterns observed within Iguanidae and putative exon shuffling recombination in *C. decresii*.

Conclusions

A total of eight MHC I transcripts were isolated and characterised for *C. decresii* from a single individual using HiSeq next generation sequencing of thymus and spleen total RNA. Due to the complex nature of the MHC, the initial assembly was refined using an assembly technique capable of distinguishing highly similar transcripts. Six of the putative MHC I transcripts were validated as likely to encode classical MHC I molecules based on three criteria; (i) normal MHC I gene structure, (ii) concordance with known conserved regions and regions with predicted function, and (iii) the absence of stop codons within coding regions. Two of the putative MHC I transcripts ended prematurely either due to transcriptome assembly restrictions (i.e. due to high sequence similarity among transcripts) or non-classical functionality (naturally truncated). Bayesian phylogenetic analysis indicated that a birth-and-death model of evolution is likely the main mechanism shaping MHC I diversity within Iguanian lizards. MHC I sequences from a wider range of squamates are required to obtain a clearer view of the mechanisms responsible for creating MHC I diversity within Squamata. This work provides a foundation for future work examining the mechanisms shaping diversity at the MHC class I region of agamid lizards.

Acknowledgments

We thank Talat Hojat Ansari and Sarah Pearson for assistance with laboratory work and data analysis, and Aaron Fenner for assistance with fieldwork. This work was funded by the Holsworth Wildlife Research Endowment and the Nature Foundation of South Australia, awarded to JH. Specimen collection and euthanasia were conducted according to the Flinders University Animal Welfare Committee (approval no. E379) and the South Australian Department of Environment, Water and Natural Resources regulations (permit no. U26225).

References

Adams, E.J., Luoma, A.M., 2013. The adaptable major histocompatibility complex (MHC) fold: Structure and function of nonclassical and MHC class I–like molecules. Annual Review of Immunology 31, 529-561.

Alfonso, C., Karlsson, L., 2000. Nonclassical MHC Class II Molecules. Annual Review of Immunology 18, 113-142.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research 25, 3389-3402.

Ansari, T., Bertozzi, T., Miller, R., Gardner, M., 2015. MHC in a monogamous lizard - Characterization of class I MHC genes in the Australian skink *Tiliqua rugosa*. Developmental and Comparative Immunology 53, 320-327.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114-2120.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. BMC Bioinformatics 10, 421.

Carlini, F., Ferreira, V., Buhler, S., Tous, A., Eliaou, J.-F., René, C., Chiaroni, J., Picard, C., Di Cristofaro, J., 2016. Association of HLA-A and Non-Classical HLA Class I Alleles. PLoS ONE 11, e0163570.

Chevreux, B., Wetter, T., Suhai, S., 1999. Genome Sequence Assembly Using Trace Signals and Additional Sequence Information, Proceedings of the German Conference on Bioinformatics (GCB) pp. 45-56.

Donadi, E.A., Castelli, E.C., Arnaiz-Villena, A., Roger, M., Rey, D., Moreau, P., 2011. Implications of the polymorphism of HLA-G on its function, regulation, evolution and disease association. Cellular and Molecular Life Sciences 68, 369-395.

Edgar, R., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32, 1792-1797.

Edwards, S., Hedrick, P., 1998. Evolution and ecology of MHC molecules: from genomics to sexual selection. Trends in Ecology & Evolution 13, 305-311.

Ejsmond, M., Radwan, J., Wilson, A., 2014. Sexual selection and the evolutionary dynamics of the major histocompatibility complex. Proceedings of the Royal Society B: Biological Sciences 281, 1-8.

Garcia, M., Yebra, B., Flores, A., Guerra, E., 2012. The Major Histocompatibility Complex in Transplantation. Journal of Transplantation 2012, 1-7.

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., Matsubara, K., Waters, P., Graves, J.A.M., Sarre, S.D., Zhang, G., 2015. High-coverage sequencing and annotated assembly of the genome of the Australian dragon lizard *Pogona vitticeps*. GigaScience 4, 1-11.

Georges, A., O'Meally, D., Genomics@UC, 2016. The *Pogona vitticeps* genome browser (pvi1.1 Jan 2013). Institute for Applied Ecology, Canberra: University of Canberra.

Gibbons, J., 1979. The hind leg pushup display of the *Amphibolurus decresii* species complex (Lacertilia: Agamidae). Copeia 1, 29-40.

Glaberman, S., Caccone, A., 2008. Species-specific evolution of class I MHC genes in iguanas (order: Squamata; subfamily: Iguaninae). Immunogenetics 60, 371-382.

Glaberman, S., Du Pasquier, L., Caccone, A., 2009. Characterization of a Nonclassical Class I MHC Gene in a Reptile, the Galapagos Marine Iguana (*Amblyrhynchus cristatus*). PLOS ONE 3, e2859.

Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., MacManes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., LeDuc, R.D., Friedman, N., Regev, A., 2013. *De novo* transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. Nature Protocols 8, 1494–1512.

Holmes, N., Parham, P., 1985. Exon shuffling *in vivo* can generate novel HLA class I molecules. The EMBO Journal 4, 2849-2854.

Hugall, A.F., Foster, R., Hutchinson, M., Lee, M.S.Y., 2008. Phylogeny of Australasian agamid lizards based on nuclear and mitochondrial genes: implications for morphological evolution and biogeography. Biological Journal of the Linnean Society 93.

Janeway, C.J., Travers, P., Walport, M., 2001. Immunobiology: The Immune System in Health and Disease, 5th ed. Garland Science, New York.

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

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., Mentjies, P., Drummond, A., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647-1649.

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

Koller, B.H., Orr, H.T., 1985. Cloning and complete sequence of an HLA-A2 gene: analysis of two HLA-A alleles at the nucleotide level. Journal of immunology (Baltimore, Md. : 1950) 134, 2727-2733.

Lanfear, R., Calcott, B., Ho, S.Y., Guindon, S., 2012. Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Molecular Biology and Evolution 29, 1695-1701.

Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B., 2017. Partitionfinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Molecular Biology and Evolution 34, 772-773.

Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM, arXiv:1303.3997v2 [q-bio.GN].

McLean, C.A., Stuart-Fox, D., Moussalli, A., 2014. Phylogeographic structure, demographic history and morph composition in a colour polymorphic lizard. Journal of Evolutionary Biology 27, 2123-2137.

Milinski, M., 2006. The Major Histocompatibility Complex, Sexual Selection, and Mate Choice. Annual Review of Ecology, Evolution, and Systematics 37.

NCBI Resource Coordinators, 2016. Database resources of the National Center for Biotechnology Information. Nucleic Acids Research 44, D7-D19.

Neefjes, J., Jongsma, M., Paul, P., Bakke, O., 2011. Towards a systems understanding of MHC class I and MHC class II antigen presentation. Nature Reviews Immunology 11, 823-836.

Nei, M., Gu, X., Sitnikova, T., 1997. Evolution by the birth-and-death process in multigene families of the vertebrate immune system. PNAS 94.

Nei, M., Rooney, A.P., 2005. Concerted and birth-and-death evolution of multigene families. Annual Review of Genetics 39, 121-152.

Osborne, L., 2005a. Information content of male agonistic displays in the territorial tawny dragon (*Ctenophorus decresii*). Journal of Ethology 23, 189-197.

Osborne, L., 2005b. Rival recognition in the territorial tawny dragon (*Ctenophorus decresii*). Acta Ethologica 8, 45-50.

Pearson, S.K., Bull, C.M., Gardner, M.G., 2017. *Egernia stokesii* (gidgee skink) MHC I positively selected sites lack concordance with HLA peptide binding regions. Immunogenetics 69, 49-61.

Piertney, S., Oliver, M., 2006. The evolutionary ecology of the major histocompatibility complex. Heredity 96, 7-21.

Piontkivska, H., Nei, M., 2003. Birth-and-death evolution in primate MHC class I genes: divergence time estimates. Mol Biol Evol 20.

Pyron, R.A., Burbrink, F.T., Wiens, J.J., 2013. A phylogeny and revised classification of Squamata, including 4161 species of lizards and snakes. BMC Evolutionary Biology 13, 93.

Rambaut, A., 2012. FigTree, 1.4.2 ed, Institute of Evolutionary Biology, University of Edinburgh.

Rassmann, K., 1997. Evolutionary Age of the Galápagos Iguanas Predates the Age of the Present Galápagos Islands. Molecular Phylogenetics and Evolution 7, 158-172.

Reche, P.A., Reinherz, E.L., 2003. Sequence variability analysis of human class I and class II MHC molecules: functional and structural correlates of amino acid polymorphisms. Journal of Molecular Biology 331, 623-641.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., Mesirov, J.P., 2011. Integrative genomics viewer. Nature biotechnology 29, 24-26.

Ronquist, F., Teslenko, P., Ayres, D., Darling, S., Hohna, S., Larget, B., Liu, L., Suchard, M.H., JP, 2011. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biology 6, 539-542.

Ryan, S.O., Cobb, B.A., 2012. Roles for major histocompatibility complex glycosylation in immune function. Seminars in Immunopathology 34, 425-441.

Ryan, S.O., Cobb, B.A., 2015. Major Histocompatibility Complex: N-Glycosylation Form and Function, in: Taniguchi, N., Endo, T., Hart, G.W., Seeberger, P.H., Wong, C.-H. (Eds.), Glycoscience: Biology and Medicine. Springer Japan, Tokyo, pp. 643-648.

Simao, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210-3212.

Sommer, S., 2005. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. Frontiers in Zoology 2, 16.

Spurgin, L.G., van Oosterhout, C., Illera, J.C., Bridgett, S., Gharbi, K., Emerson, B.C., Richardson, D.S., 2011. Gene conversion rapidly generates major histocompatibility complex diversity in recently founded bird populations. Molecular ecology 20, 5213-5225.

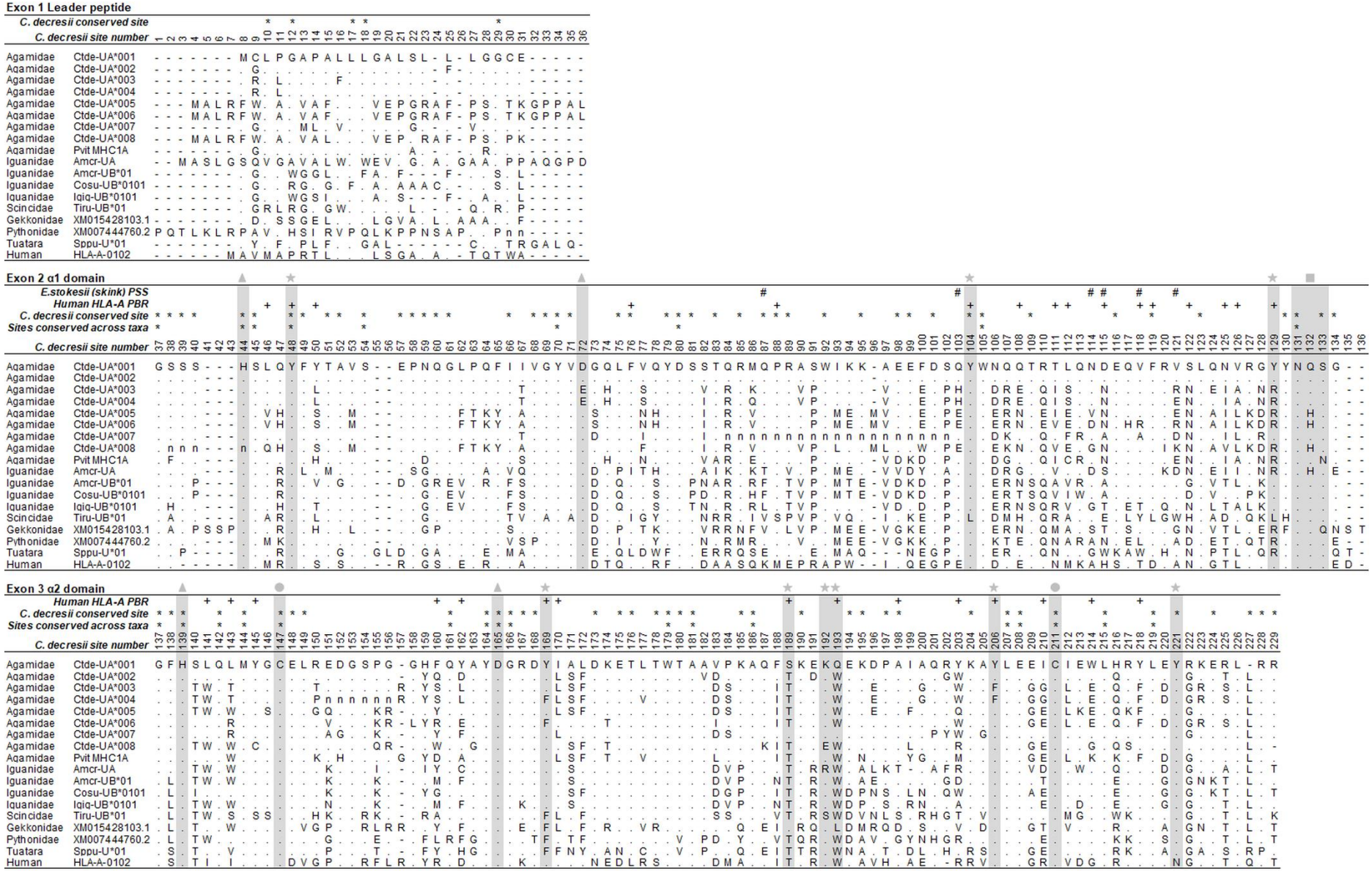
Tamura, K., Peterson, D., Petersoin, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution 28, 2731-2739.

Wang, D., Zhong, L., Wei, Q., Gan, X., He, S., 2010. Evolution of MHC class I genes in two ancient fish, paddlefish (*Polyodon spathula*) and Chinese sturgeon (*Acipenser sinensis*). FEBS letters 584, 3331-3339.

Yewers, M.S.C., Pryke, S., Stuart-Fox, D., 2016. Behavioural differences across contexts may indicate morph-specific strategies in the lizard *Ctenophorus decresii*. Animal Behaviour 111, 329-339.

Zhao, M., Wang, Y., Shen, H., Li, C., Chen, C., Luo, Z., Wu, H., 2013. Evolution by selection, recombination, and gene duplication in MHC class I genes of two Rhacophoridae species. BMC Evolutionary Biology 13, 113.

Tables and figures


Exon 1 Leader peptide
C. decresii conserved site
C. decresii site number
Agamidae Ctde-UA*001
Agamidae Ctde-UA*002
Agamidae Ctde-UA*003
Agamidae Ctde-UA*004
Agamidae Ctde-UA*005
Agamidae Ctde-UA*006
Agamidae Ctde-UA*007
Agamidae Ctde-UA*008
Agamidae Pvit MHC1A
Iguanidae Amcr-UA
Iguanidae Amcr-UB*01
Iguanidae Cosu-UB*0101
Iguanidae Iqig-UB*0101
Scincidae Tiru-UB*01
Gekkonidae XM015428103.1
Pythonidae XM007444760.2
Tuatara Sppu-U*01
Human HLA-A-0102
Exon 2 α1 domain
E.stokesii (skink) PSS
Human HLA-A PBR
C. decresii conserved site
Sites conserved across taxa
C. decresii site number
Exon 3 α2 domain
Human HLA-A PBR
C. decresii conserved site
Sites conserved across taxa
C. decresii site number

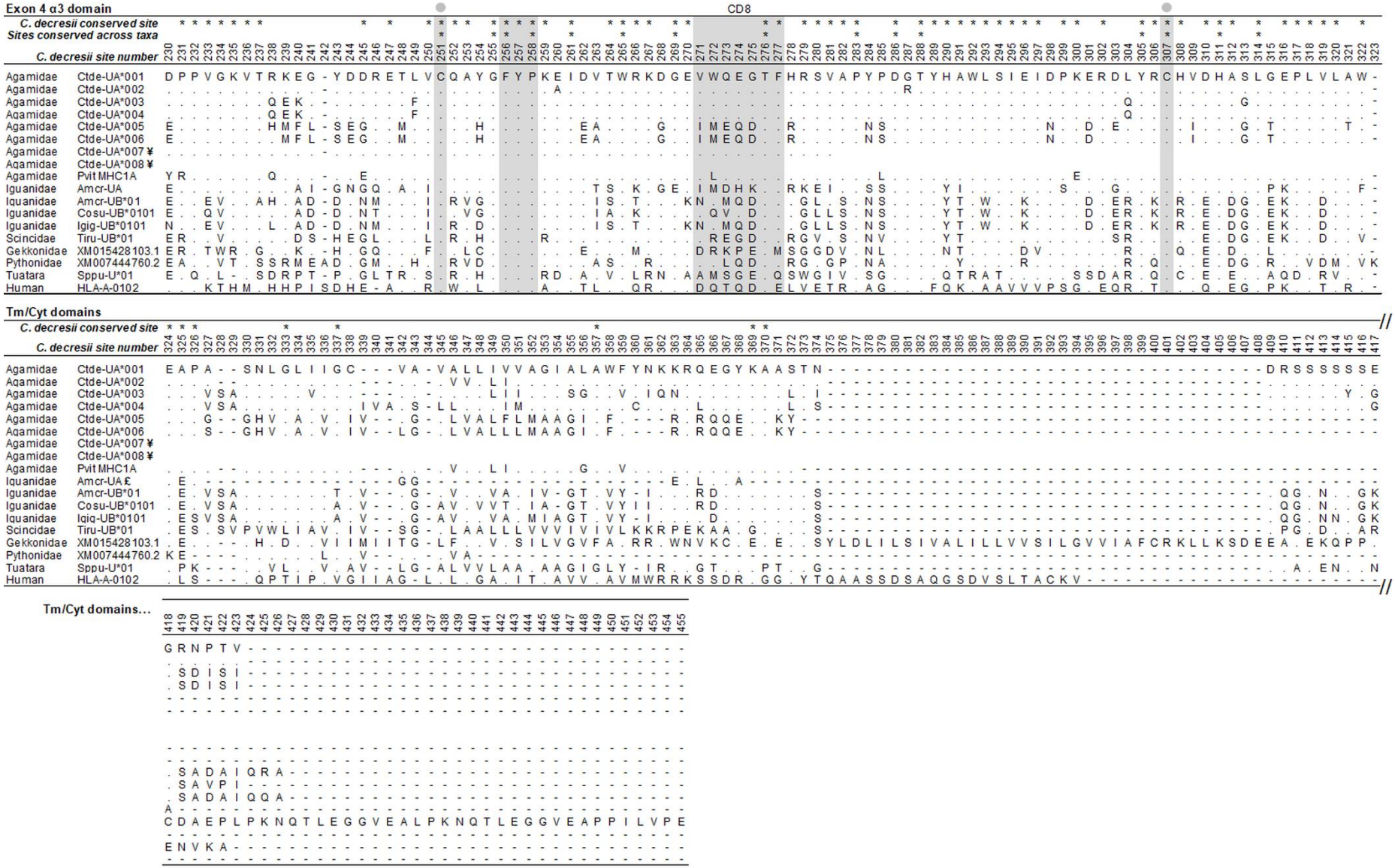

Exon 4 α3 domain
CD8
C. decresii conserved site
Sites conserved across taxa
C. decresii site number
Tm/Cyt domains
Tm/Cyt domains...
Agamidae Ctde-UA*001
Agamidae Ctde-UA*002
Agamidae Ctde-UA*003
Agamidae Ctde-UA*004
Agamidae Ctde-UA*005
Agamidae Ctde-UA*006
Agamidae Ctde-UA*007¥
Agamidae Ctde-UA*008¥
Agamidae Pvit MHC1A
Iguanidae Amcr-UA
Iguanidae Amcr-UB*01
Iguanidae Cosu-UB*0101
Iguanidae Igig-UB*0101
Scincidae Tiru-UB*01
Gekkonidae XM015428103.1
Pythonidae XM007444760.2
Tuatara Sppu-U*01
Human HLA-A-0102
Iguanidae Amcr-UA£

Figure 1. Amino acid alignment of tawny dragon (*Ctenophorus decresii*) MHC I transcripts with other squamates, tuatara (*Sphenodon punctatus*) and human. Coding domain separations are based on Koller and Orr (1985). Dots indicate identity with *Ctde-UA*001* and dashes indicate alignment gaps. ¥ = partial coding sequence. Residues with expected functions as per Kaufman et al. (1994) are shaded grey; *stars* = conserved peptide-binding residues of antigen N and C termini, *triangles* = salt bridge-forming residues, *circles* = disulphide bridge-forming cysteines, *square* = N-glycosylation site, CD8 = expected CD8 binding site. *Asterisks* represent conserved sites, across *C. decresii* sequences and across all taxa. Gidgee skink (*Egernia stokesii*) positively selected sites (PSS, putative peptide binding regions) (Pearson et al., 2017) and human peptide binding regions (PBR) (Reche and Reinherz, 2003) are indicated with *hashes* and *crosses*, respectively.

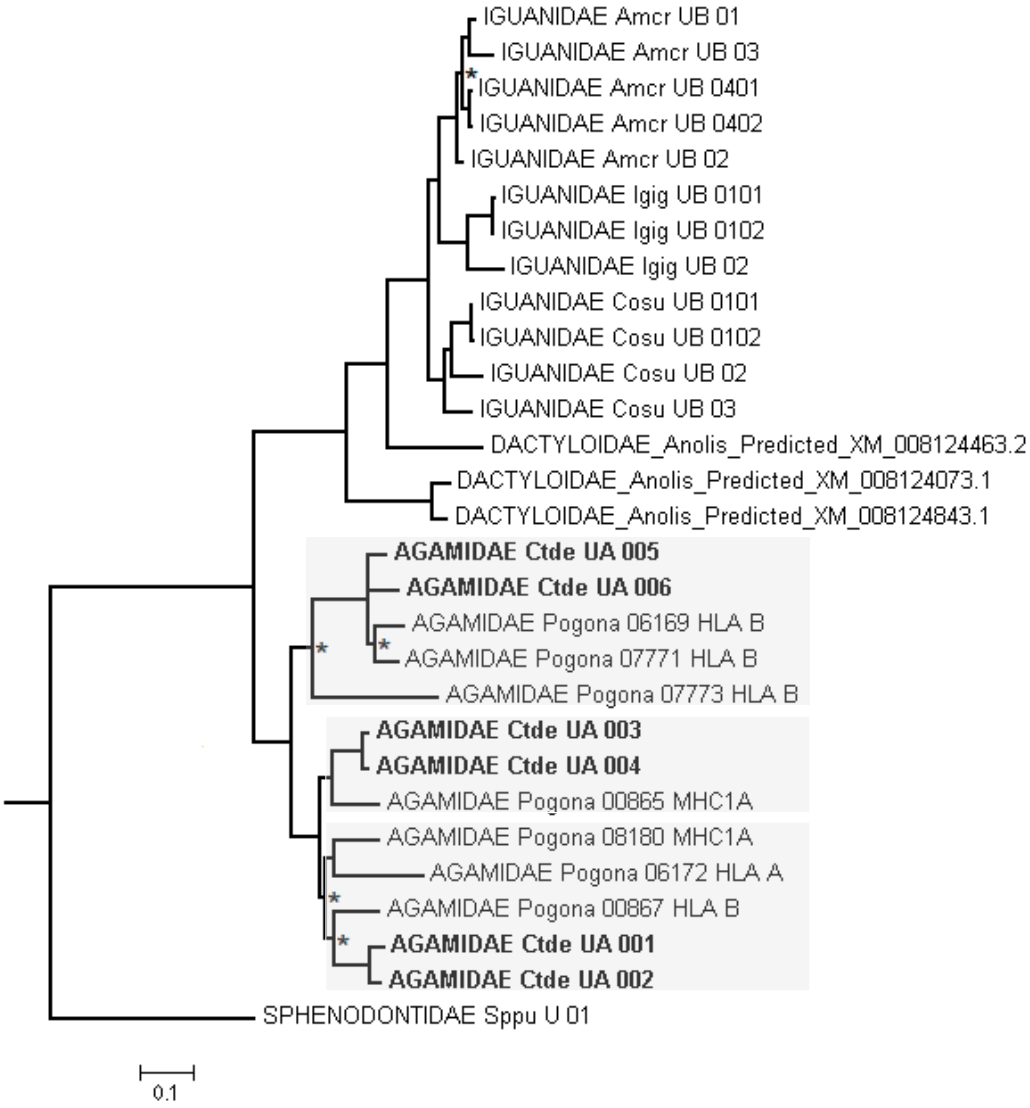


Figure 2. Bayesian phylogenetic tree of the α1, α2 and α3 domains of available full length Iguania (here, Iguanidae, Agamidae and Dactyloidae) MHC class I nucleotide sequence. Validated full length tawny dragon (*Ctenophorus decresii*) MHC class I transcripts are highlighted by bold text. The tree is rooted using tuatara (*Sphenodon punctatus*) MHC class I sequence. Three orthologous Agamidae clades are shaded grey. Nodes for which the Bayesian posterior probability <0.95 are indicated with an asterisk. The scale bar indicates the number of expected nucleotide substitutions per site. *Anolis* sequences were predicted by NCBI automated computational searches of genomic sequence (NCBI Resource Coordinators, 2016). The *Pogona* sequences were obtained from the Pogona Genome Project (Georges et al., 2015; Georges et al., 2016). Refer to table S1 for further information on sequences.

Supplementary Information

Characterisation of major histocompatibility complex class I transcripts in an Australian dragon lizard

Hacking, J., Bertozzi, T., Moussalli, A., Bradford, T., Gardner, M.

Developmental & Comparative Immunology, 84, 164-171, (2018).

Table S1. Details of the full-length MHC class I sequences used in this study.

Species	Common name	Sequence name and/or GenBank accession
Ctenophorus decresii	Tawny dragon	*Ctde-UA*001 – Ctde-UA*008* (KY905239 – KY905246)
Pogona vitticeps	Central bearded dragon	Pvit 06069, Pvit 07771, Pvit 07773, Pvit 00867, Pvit 00865, Pvit 08180, Pvit 06172¥
Amblyrhynchus cristatus	Galápagos marine iguana	*Amcr-UB*01 – Amcr-UB*0402* (EU604308.1 – EU604312.1), *Amcr-UA* (EU839663.1)
Conolophus subcristatus	Galápagos land iguana	*Cosu-UB*0101 – Cosu-UB*03* (EU604313.1 – EU604316.1)
Iguana iguana	Green iguana	*Igig-UB*0101 – Igig-UB*02* (EU604317.1 – EU604319.1)
Anolis carolinensis£	Green anole	XM_008124843.1, XM_008124463.2, XM_008124073.1
Tiliqua rugosa	Sleepy lizard	*Tiru-UB*01* (Ansari et al., 2015)
Gekko japonicus£	Schlegel's Japanese gecko	XM_015428103.1
Python bivittatus£	Burmese python	XM_007444760.2
Boiga irregularis	Brown tree snake	GBSH01001602.1 – GBSH01001604.1
Thamnophis sirtalis£	Garter snake	XM_014062154.1, XM_014063303.1, XM_014067447.1, XM_014071192.1, XM_014074231.1
Sphenodon punctatus	Tuatara	*Sppu-U*01* (DQ145788.1)
Homo sapiens	Human	*HLA-A*0102* (U07161.1)

¥ sourced from the Pogona Genome Project (Georges et al., 2016)

£ predicted to be MHC class I sequence by automated computational analysis (NCBI Resource Coordinators, 2016)

Table S2. Tawny dragon lizard (*Ctenophorus decresii*) MHC I transcript percent pairwise nucleotide and amino acid (in brackets) identity with *Ctde-UA*001*.

Transcript	Complete	α1	α2	α3
*Ctde-UA*001*	-	-	-	-
*Ctde-UA*002*	95.6 (93.0)	100 (100)	88.4 (81.5)	98.6 (97.8)
*Ctde-UA*003*	84.1 (75.1)	81.2 (69.6)	78.9 (65.6)	94.9 (93.5)
*Ctde-UA*004*	84.1 (75.4)	82.6 (70.7)	76.2 (62.1)	96.0 (94.6)
*Ctde-UA*005*	67.1 (56.7)	72.8 (57.0)	78.1 (67.4)	74.6 (69.6)
*Ctde-UA*006*	67.9 (57.9)	71.0 (54.8)	81.4 (71.7)	76.1 (72.8)
*Ctde-UA*007*	-	88.0 (78.4)	89.5 (83.7)	-
*Ctde-UA*008*	-	70.4 (56.2)	81.9 (73.6)	-
Average	*79.8 (71.6)*	*80.9 (69.5)*	*82.1 (72.2)*	*88.0 (85.7)*

References

Ansari, T., Bertozzi, T., Miller, R., Gardner, M., 2015. MHC in a monogamous lizard - Characterization of class I MHC genes in the Australian skink *Tiliqua rugosa*. Developmental and Comparative Immunology 53, 320-327.

Georges, A., O'Meally, D., Genomics@UC, 2016. The *Pogona vitticeps* genome browser (pvi1.1 Jan 2013). Institute for Applied Ecology, Canberra: University of Canberra.

NCBI Resource Coordinators, 2016. Database resources of the National Center for Biotechnology Information. Nucleic Acids Research 44, D7-D19.

CHAPTER 3

Very low rate of multiple paternity detected in clutches of a wild agamid lizard

Note to examiners

This chapter outlines the methods used to obtain family group data for *C. decresii* and describes the genetic mating system employed by this species. Family group data will be used in subsequent chapters to i) aid in genotyping at the MHC, ii) investigate segregation patterns of MHC alleles, and iii) test hypotheses regarding MHC-associated mating. This chapter has been accepted for publication in the *Australian Journal of Zoology* and is therefore formatted according to the specifications for that journal.

Hacking, J., Stuart-Fox, D., Gardner, M., (In Press). Very low rate of multiple paternity detected in clutches of a wild agamid lizard. *Australian Journal of Zoology.*

Very low rate of multiple paternity detected in clutches of a wild agamid lizard

Jessica Hacking[a] (corresponding author)

jessica.hacking@flinders.edu.au

Devi Stuart-Fox[c]

d.stuart-fox@unimelb.edu.au

Michael Gardner[a,b]

michael.gardner@flinders.edu.au

[a]College of Science and Engineering, Flinders University, Bedford Park SA 5042, Australia

[b]Evolutionary Biology Unit, South Australian Museum, Adelaide SA 5000, Australia

[c]School of BioSciences, University of Melbourne, Parkville VIC 3010, Australia

Abstract

Genetic mating systems described for squamate reptiles range from primarily monogamous to completely polygynandrous. The presence of female multiple mating is almost ubiquitous among squamates and even occurs, albeit at a low rate, in socially monogamous species. Here we examine the genetic mating system of the territorial tawny dragon lizard (*Ctenophorus decresii*). Paternity was assigned to captive-born hatchlings using eight microsatellite loci, revealing a 4% rate of multiple paternity. One quarter of males sired more than one clutch, although multiple mating by males is likely underestimated. The rate of multiple paternity in *C. decresii* represents one of the lowest among squamates and may be a result of successful male territoriality. However, the observed low rate of multiple paternity does not eliminate the possibility of widespread female multiple mating due to the potential for sperm storage and sperm competition. We conclude that the tawny dragon lizard employs a predominantly polygynous genetic mating system.

Keywords: genetic mating system, polygyny, *Ctenophorus decresii*

Short title: Limited multiple paternity in an Australian lizard

Introduction

The evolution of mating systems involves a delicate balance between the costs and benefits of mating to each sex (Birkhead and Møller 1998). Male multiple mating is expected in most taxa due to the direct benefits involved, whereas the benefits of multiple mating by females are often not immediately apparent (Barbosa, Connolly *et al.* 2012; Uller and Olsson 2008). Male multiple mating has been detected to varying degrees in all squamate reptiles where male mating patterns have been examined (table S2). Furthermore, the presence of female multiple mating is almost ubiquitous among squamates (table S2). This is true even within socially monogamous species, although rates of male and female multiple mating are relatively low (Bull, Cooper *et al.* 1998; Gardner, Bull *et al.* 2002). The only squamate species for which female multiple mating has not been detected is the great desert skink, *Liopholis kintorei* (McAlpin, Duckett *et al.* 2011), some true sea snake species (Lukoschek and Avise 2011) and a species of garter snake (Wusterbarth, King *et al.* 2010). However, the studies investigating female multiple mating in snakes (Lukoschek and Avise 2011; Wusterbarth, King *et al.* 2010) comprised very small sample sizes, ranging from one to four individuals per species (table S2) and that of *L. kintorei* did not include full clutches but wild caught animals only. The high incidence of female multiple mating among squamates implies that females may benefit from multiple mating. Indeed, there are a number of examples of indirect benefits associated with female multiple mating (Eizaguirre, Laloi *et al.* 2007; Frère, Chandrasoma *et al.* 2015; Madsen, Shine *et al.* 1992; Noble, Keogh *et al.* 2013). However, the sexual conflict hypothesis predicts that this may not always be the case; the occurrence of female multiple mating is sometimes governed by the cost of mating versus the cost of resisting male sexual advances (Chapman, Arnqvist *et al.* 2003; Keogh, Umbers *et al.* 2013; McLean, Moussalli *et al.* 2010; York and Baird 2015).

Female multiple mating within and between reproductive cycles can lead to multiple paternity within clutches or litters in squamate reptiles (Uller and Olsson 2008). However, the relationship between multiple mating and multiple paternity is not straight-forward in lizards and snakes; the potential for cryptic female post-copulatory choice and sperm competition (Friesen, Mason *et al.* 2014; Tolley, Chauke *et al.* 2014; Uller, Schwartz *et al.* 2013) means that multiple mating may not always result in multiple paternity. As a result, the genetic mating system cannot be used to infer the behavioural mating system in squamate reptiles. The rate of multiple paternity is highly variable among squamates, ranging from 2.6% in Cunningham's skinks (*Egernia cunninghami*) to 100% in Cape dwarf chameleons (*Bradypodion pumilum*) (table S2). Uller and Olsson (2008) argue that the main driver of this variation is mate encounter rate. Factors that are likely to influence mate encounter rate, and hence rates of multiple mating, include population density (Jensen, Abreu-Grobois *et al.* 2006), social pair formation

(Bull, Cooper *et al.* 1998; Gardner, Bull *et al.* 2002; Stow and Sunnucks 2004), territoriality (York and Baird 2015) and mate-seeking behaviours (Glass, Klein *et al.* 2016).

Here, we investigated the genetic mating system of a wild population of colour polymorphic tawny dragon lizards (*Ctenophorus decresii*), a small (<30 g) rock-obligate agamid lizard endemic to South Australia. We used a robust captive hatching method that ensured all maternal relationships were known and all members of a clutch were sampled. Sampling of the male population allowed paternity assignment using a maximum likelihood approach. We hypothesise that *C. decresii* exhibits a predominantly polygynous genetic mating system, with a relatively small portion of clutches containing multiple paternity, as is the case in other members of the *Ctenophorus* genus (Lebas 2001; Olsson, Healey *et al.* 2007). Defining the genetic mating system employed by *C. decresii* will provide a better understanding of the basic biology of the species and will direct future work investigating *C. decresii* reproductive biology and factors influencing mate choice.

Materials and methods

Sample collection

During spring and summer between 2013 and 2015 individual tawny dragons were collected from a single location near Hawker in the central Flinders Ranges, South Australia (-31.954865, 138.373987). Males were released at the point of capture after a blood sample (≤50 µL) was taken. During the spring of 2014, 23 gravid females were captured from the Hawker field site and taken to Flinders University, Adelaide, to lay their eggs. Females were subsequently released at the point of capture at the Hawker field site after a blood sample (≤50 µL) was taken. Both males and females were marked with a unique number using a non-toxic paint pen to ensure they were not recaptured and could be identified by sight. Hatchlings were housed individually at the Flinders University Animal House. At 12 weeks of age, a blood sample (≤15 µL) was taken from hatchlings before release near their mothers' capture locations. Refer to supplementary material (text S1) for future details on captive hatching methodology. Blood was taken from adults and hatchlings via venipuncture of the *sinus angularis*, located inside the mouth. Whole blood was preserved on Whatman® Classic FTA cards (GE Healthcare, Buckinghamshire, UK) and stored at room temperature until DNA extraction. Five hatchlings died of natural causes and were frozen at -20 °C until a tissue sample (5mm tail tip) was taken for DNA extraction.

Microsatellite amplification

Hatchling paternity was determined from the genotypes of eight microsatellite loci. All mothers (n = 23) and their offspring (n = 139), and all adult males (n = 90, >70 mm snout-vent length) within 100 m

of mothers were genotyped using eight previously published microsatellite markers designed for *C. decresii* (Ctde03, Ctde05, Ctde08, Ctde12, Ctde21 and Ctde45; McLean et al., 2014) or the closely related *Ctenophorus pictus* (CP10 and CP11; Schwartz, Warner *et al.* 2007). DNA was extracted from blood stored on FTA cards according to Smith and Burgoyne (2004) and tissue DNA extraction was undertaken using a Gentra ® Puregene ® (Gentra Systems) method. Microsatellite amplification reaction protocols followed McLean, Stuart-Fox *et al.* (2014). Briefly, all loci were amplified singly and PCR reactions were 12.5 µL in total volume, containing 1x MRT buffer (4 mM dNTP, 0.25 mg/ml BSA), 0.4 µM of forward and reverse primers, 0.5 U of Immolase enzyme and approximately 20 ng of template DNA. Thermal cycling conditions for Ctde03, Ctde21, Ctde45, CP10 and CP11 consisted of initial denaturisation at 95 °C for 10 minutes followed by 42 cycles of 95 °C for 30 seconds, 48 °C for 30 seconds and 72 °C for 30 seconds, with a final extension at 72 °C for 10 minutes. For the remaining loci (Ctde05, Ctde08 and Ctde12) thermal cycling conditions consisted of initial denaturisation at 95 °C for 10 minutes followed by four different annealing steps. These annealing steps included two cycles each of 95 °C for 30 seconds, annealing temperatures of 60 °C, 55 °C and 50 °C for 30 seconds followed by 72 °C for 45 seconds. The last annealing step consisted of 35 cycles of 95 °C for 30 seconds, 45 °C for 30 seconds followed by 72 °C for 45 seconds. A final extension of 72 °C for 10 minutes was used. After diluting and pooling per individual, PCR products were sent to the Australian Genome Research Facility (AGRF) for fragment visualization. Fragment sizes were called using the GS500(-250)LIZ size standard in GeneMapper® ver. 4 (Applied Biosystems).

Paternity assignment and determining the genetic mating system

Relatedness values among all adults were estimated using Coancestry ver. 1.0.1.5 (Wang 2007). All microsatellite loci were tested for deviation from Hardy-Weinberg equilibria (HWE) and the presence of linkage disequilibrium using GenePop ver. 4.5.1 (Rousset 2008) based on a subset of unrelated adults (< 0.25 relatedness value, n = 23). The number of alleles, number of individuals typed, the observed and expected heterozygosities, and polymorphic information content were calculated for each locus using Cervus ver. 3.0.7 (Kalinowski, Taper *et al.* 2007), based on all adults (n=113). Cervus was also used to determine locus-specific error rates (including null alleles) based on known mother-offspring mismatches, with the average probability of detecting a mismatch taken into account (Kalinowski, Taper *et al.* 2007).

To estimate the rate of multiple paternity within *C. decresii* clutches we counted the number of paternal alleles at each locus for each clutch. Greater than two paternal alleles at two or more loci within a clutch was considered evidence of multiple paternity (Fitzsimmons 1998; Gardner, Bull *et al.* 2002; Todd, Blair *et al.* 2013).

Paternity assignment was conducted using Colony ver. 2.0.6.2 (Jones and Wang 2010). Colony takes a maximum likelihood approach to assigning both sibship and parentage relationships and deals well with deviations from HWE, null alleles and other stochastic errors (Jones and Wang 2010). Males captured in the season preceding (2013-2014) the collection of gravid females were included as potential fathers in paternity analyses in addition to males captured during the same season as females (2014-2015). Male *C. decresii* live for at least five years in this population (Yewers 2016) and some males may not have been captured, despite being present during the 2014-2015 season, given the low recapture rate reported by Yewers (2016). Another reason for including males from the previous season is the potential for between-season female sperm storage. Between-season sperm storage has been reported for other lizards (Uller, Stuart-Fox *et al.* 2010) and sperm storage has been reported in congeners *C. fordi* and *C. pictus* (Olsson, Schwartz *et al.* 2007; Olsson, Schwartz *et al.* 2009; Uller, Schwartz *et al.* 2013). However, in contrast to *C. decresii*, both *C. fordi* and *C. pictus* are short-lived and most individuals die after their first reproductive season (Olsson, Schwartz *et al.* 2007; Olsson, Schwartz *et al.* 2009; Uller, Schwartz *et al.* 2013), limiting the ability to test for between-season sperm storage in these species. In a captive *C. decresii* population only a single case of within-season sperm storage and no cases of between-season sperm storage occurred in 25 clutches (23 sires and 17 dams, Rankin, McLean *et al.* 2016). Hence, although between-season female sperm storage is unlikely to occur in *C. decresii* it cannot be discounted and remains a possibility.

Colony requires information on the mating system employed (male and female multiple mating) when assigning paternity. We discovered evidence of multiple paternity based on the number of paternal alleles at loci within clutches. Therefore, to infer paternity for *C. decresii* we assumed male multiple mating and undertook two Colony runs; one with the mating system set to polygyny (only male multiple mating) and the other set to polygynandry (both male and female multiple mating). For both runs the run length was set to 'very long', which included a high number of configurations in the simulated annealing algorithm used by Colony (Wang 2017). All maternal sibship relationships were known, no males were excluded as fathers and the probability that a father was included among the candidates was set to 0.75. Relationships (sibship and paternity) with a probability value above 0.95 were accepted. Microsatellite error rates, including estimated null allele frequency, were taken into account during Colony analyses (Wang 2004). The spatial position of fathers relative to mothers was plotted using ArcMaps ver. ® (ArcGIS Ersi) and was used to validate paternity assignment.

Results

The polymorphic information content was high for all loci except for CP11 due to the low number of alleles at this locus (table 1). However, this locus was retained as the error rate was acceptable and it did not deviate from HWE. The error rates for all loci were acceptable. For all loci, except CP10 and CP11, most or all of the mismatches were likely caused by null alleles (table 1). Mismatches associated with CP10 and CP11 may be due to mistyping or mutations. None of the microsatellite loci pairs were in linkage disequilibrium, although three loci (Ctde03, Ctde05 and Ctde21) were found to deviate from HWE. The slight deviations from HWE for Ctde03 and Ctde21 were likely caused by null alleles (Pemberton, Slate *et al.* 1995). As there were no mismatches associated with Ctde05 the source of the deviation of this locus from HWE is unknown (Kalinowski, Taper *et al.* 2007). As Colony analyses are error-tolerant and are able to deal with some loci deviating from HWE (Wang 2004), all eight microsatellite loci were considered suitable to use for paternity assignment. All candidate males and mothers had unique genotypes.

We uncovered evidence for multiple paternity in a single clutch (1/23). This clutch had three paternal alleles at four different loci. No other clutches had more than two paternal alleles at any loci (table S3). The polygyny Colony run assigned paternity with high probability values for 21 of the 23 clutches. Paternity probability values were greater than or equal to 0.95 for all but one clutch, which had a paternity probability of 0.83 (clutch of female 987 and male 700, table S3). This paternity assignment was accepted because the lower probability was likely due to the presence of null alleles and missing data in the male and the assigned male was in close proximity to the mother (12 m). A total of 16 males were identified as fathers and a quarter of those males sired more than one clutch. Three fathers were captured only in the season previous to captive hatching, all of which were captured in autumn (March). The polygynandry Colony run produced the same paternity assignments as the polygyny run but paternity probability values were slightly lower for some hatchlings, which was likely due to the presence of null alleles in those hatchlings (Jones and Wang 2010). For both the polygyny and polygynandry Colony runs single paternity was assigned to the clutches that were previously identified as lacking multiple paternity (≤ two alleles at loci, table S3). Paternity was not assigned to two clutches, one of which was the clutch that was sired by multiple males (three paternal alleles at four loci). It is therefore likely that the fathers contributing to that clutch were not sampled. The other clutch for which paternity was not assigned was likely sired by a single unsampled male. All fathers were within close proximity to mothers (mean = 28 m, range = 5 m to 83 m; fig. 1). Overall, paternity was assigned with high confidence and a 4% (1/23) rate of multiple paternity was uncovered.

Discussion

Tawny dragons display a territorial social structure with males occupying home ranges that exclude other males (Yewers 2016) and are likely to encompass several females. We tested the hypothesis that *C. decresii* employs a predominantly polygynous genetic mating system, including instances of multiple paternity. Paternity assignment produced 21 complete family groups, with a single male siring each clutch. However, upon inspection of the two clutches not assigned paternity, one indicated that at least two males contributed to the clutch (multiple paternity). Hence, out of 23 clutches, one (4%) consisted of multiple paternities. A total of 16 sampled males contributed to the 21 clutches for which paternity was assigned and at least 18 males contributed to the total sample of 23 clutches. Of the sampled males, 25% (4/16) sired more than one clutch. However, male multiple mating is probably underestimated as it is likely that some males also mated with females that were not sampled. Consequently, we accept our hypothesis that a predominantly polygynous genetic mating system is employed by *C. decresii*, with at least a quarter of fathers siring more than one clutch and a single case of multiple paternity detected. The extremely low rate of multiple paternity observed here is the third lowest reported for squamates; only Cunningham's skinks, *Egernia cunninghami* (3%, Stow and Sunnucks 2004), the great dessert skink, *Liopholis kintorei* (0%, McAlpin, Duckett *et al.* 2011) and potentially some snake species (0%, Lukoschek and Avise 2011; Wusterbarth, King *et al.* 2010) have lower reported rates of multiple paternity. However, it is possible that the rate of multiple paternity in these studies is underestimated. In the case of both *E. cunninghami* and *L. kintorei* multiple paternity was determined based on sampling juveniles within the population. Therefore, some clutch members may have been missed during sampling. Furthermore, the very small sample sizes in the snake studies (one to four individuals per species) limits the ability to detect cases of multiple paternity.

Uller and Olsson (2008) predicted that the low rates of multiple paternity in congeners *C. pictus* and *C. ornatus* (18% and 25%, respectively) are due to highly territorial behaviour. However in *C. ornatus,* Lebas (2001) found that 65% of clutches were sired completely or partially by a male that was not the main territory-holder, which may indicate that a mating strategy other than strict territoriality is employed, or territoriality is not an effective strategy for *C. ornatus*. Recent work suggests that male *C. decresii* are highly territorial and defend areas of 213 m^2, on average (Yewers 2016). Hence, the extremely low rate of multiple paternity observed for *C. decresii* may be associated with successful male territoriality. However, the factors influencing male territory size and the effectiveness of territoriality have not been investigated in *C. decresii*. Other factors may contribute to the lack of multiple paternity in most clutches. For instance, females may reject males other than the territory-holder or reject additional mates after the first mating (McLean, Moussalli *et al.* 2010). Sperm storage and sperm competition has been uncovered in congeners *C. pictus* (Olsson, Schwartz *et al.* 2009) and *C. fordii*

(Uller, Schwartz *et al.* 2013). Therefore, another possibility is that female multiple mating in conjunction with cryptic post-copulatory mate choice or sperm competition results in single paternity clutches (Moore, Daugherty *et al.* 2009). The fact that some fathers were captured only in the season previous to captive hatching suggests that females may store sperm between-seasons (from autumn to spring). However, another possibility is that these males mated with females in the same season as captive hatching but evaded capture.

A combination of behavioural and molecular data is required to gain a complete picture of the mating system employed by *C. decresii*. If successful male territoriality and/or female rejection tactics are main driver of the low rate of multiple paternity in *C. decresii*, the genetic mating system likely reflects the behavioural mating system. However, due to the potential for sperm storage and sperm competition, cryptic female multiple mating cannot be ruled out. The nature of both the genetic and behavioural mating system used by a species has important implications for several aspects of its biology. For example, mating systems and associated behaviours influence connectivity among individuals, which can affect rates of parasite transmission (van Schaik and Kerth 2017; White, Richard *et al.* 2011). Furthermore, polygyny associated with territoriality or harem structure can severely depress effective population size due to skew in male reproductive success within the population (Nunney 1993). Reduced effective population size can limit the evolutionary potential of populations to deal with environmental change (Franklin 1980), accelerate the rate of accumulation of deleterious mutations (Lande 1995), and inhibit the fixation of beneficial alleles (Whitlock and Bürger 2004).

The type of mating system employed by a species may differ spatially or temporally (Moore, Daugherty *et al.* 2009; Natoli, Phillips *et al.* 2017; Phillips, Jorgensen *et al.* 2013). Therefore, the genetic mating system described for this *C. decresii* population may not necessarily be employed throughout the entire season, across years or across all populations. For instance, the mating system employed by a tuatara population was examined over three years; in two of those years multiple paternity was observed but in one year there was no genetic evidence of female multiple mating (Moore, Daugherty *et al.* 2009). Population density is likely to influence mate encounter rates and hence rates of multiple paternity. Therefore, spatial or seasonal variation in population density may result in varying rates of multiple paternity (Natoli, Phillips *et al.* 2017). The population on which this study is based has a relatively high population density (McLean, C. and Stuart-Fox, D., unpublished data). Therefore, in the case of *C. decresii*, geographic variation in population density is unlikely to cause higher than observed levels of multiple paternity. However, further work is required to determine whether other factors influence rates of multiple paternity in *C. decresii*.

A predominantly polygynous genetic mating system was described for *C. decresii,* with a single clutch sired by more than one male. This result is consistent with other members of the *Ctenophorus*

genus, although the rate of multiple paternity was much lower than expected. This extremely low rate of multiple paternity may be due to mating behaviour such as male territoriality, or post-copulatory mechanisms such as sperm competition. Behavioural observations and experiments will provide further insights into the behavioural mating system employed by *C. decresii*, and wider spatial and temporal sampling will determine whether our results are representative of other *C. decresii* populations and whether there is seasonal variation in the genetic mating system employed by *C. decresii*. Our results provide a foundation for future work examining the factors influencing mate choice and reproductive success in the tawny dragon lizard.

Acknowledgements

We thank Tessa Bradford for assistance with laboratory work and Claire McLean for advice on microsatellite markers. We are grateful for the guidance on captive hatching methodology provided by Katrina Rankin. The Flinders University Animal House staff and volunteers provided assistance with animal husbandry, for which we are thankful. We also thank volunteers that helped with field work. This work was funded by the Holsworth wildlife research endowment, Linnean Society of New South Wales and Field Naturalists Society of South Australia, awarded to JH, and an Australian Research Council grant (DP1092908) awarded to DSF. All methods were approved by Flinders University Animal Welfare Committee (E379) and the South Australian Department of Environment, Water and Natural Resources (permit number U26225). We acknowledge the traditional owners of the land on which lizards were sampled and all field work at the Yourambulla Caves Aboriginal Heritage site was approved by the Adnyamathanha Traditional Lands Association.

References

Barbosa, M., Connolly, S.R., Hisano, M., Dornelas, M., and Magurran, A.E. (2012) Fitness consequences of female multiple mating: A direct test of indirect benefits. *BMC Evolutionary Biology* **12**(1), 185.

Birkhead, T.R., and Møller, A.P. (1998) 'Sperm competition and sexual selection.' (Academic Press: New York)

Bull, C.M., Cooper, S.J.B., and Baghurst, B.C. (1998) Social monogamy and extra-pair fertilization in an Australian lizard, *Tiliqua rugosa*. *Behavioral Ecology and Sociobiology* **44**(1), 63-72.

Chapman, T., Arnqvist, G., Bangham, J., and Rowe, L. (2003) Sexual conflict. *Trends in Ecology & Evolution* **18**(1), 41-47.

Eizaguirre, C., Laloi, D., Massot, M., Richard, M., Federici, P., and Clobert, J. (2007) Condition dependence of reproductive strategy and the benefits of polyandry in a viviparous lizard. *Proceedings of the Royal Society B: Biological Sciences* **274**(1608), 425-430.

Fitzsimmons, N.N. (1998) Single paternity of clutches and sperm storage in the promiscuous green turtle (*Chelonia mydas*). *Molecular Ecology* **7**(5), 575-584.

Franklin, I.R. (1980) 'Evolutionary change in small populations.' (Sinauer: Sunderland , MA)

Frère, C.H., Chandrasoma, D., and Whiting, M.J. (2015) Polyandry in dragon lizards: inbred paternal genotypes sire fewer offspring. *Ecology and Evolution* **5**(8), 1686-1692.

Friesen, C.R., Mason, R.T., Arnold, S.J., and Estes, S. (2014) Patterns of sperm use in two populations of Red-sided Garter Snake (*Thamnophis sirtalis parietalis*) with long-term female sperm storage. *Canadian Journal of Zoology* **92**, 33+. [In English]

Gardner, M.G., Bull, C.M., and Cooper, S.J.B. (2002) High levels of genetic monogamy in the group-living Australian lizard *Egernia stokesii*. *Molecular Ecology* **11**(9), 1787-1794.

Glass, G.E., Klein, S.L., Norris, D.E., and Gardner, L.C. (2016) Multiple paternity in urban norway rats: Extended ranging for mates. *Vector Borne Zoonotic Diseases* **16**(5), 342-8. [In eng]

Jensen, M.P., Abreu-Grobois, F.A., Frydenberg, J., and Loeschcke, V. (2006) Microsatellites provide insight into contrasting mating patterns in arribada vs. non-arribada olive ridley sea turtle rookeries. *Molecular Ecology* **15**(9), 2567-2575.

Jones, O., and Wang, J. (2010) COLONY: a program for parentage and sibship inference from multilocus genotype data. *Molecular Ecology Resources* **10**(3), 551–555.

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**(5), 1099-1106.

Keogh, J.S., Umbers, K.D.L., Wilson, E., Stapley, J., and Whiting, M.J. (2013) Influence of alternate reproductive tactics and pre- and postcopulatory sexual selection on paternity and offspring performance in a lizard. *Behavioral Ecology and Sociobiology* **67**(4), 629-638.

Lande, R. (1995) Mutation and Conservation. *Conservation Biology* **9**(4), 782-791.

Lebas, N.R. (2001) Microsatellite determination of male reproductive success in a natural population of the territorial ornate dragon lizard, *Ctenophorus ornatus*. *Molecular Ecology* **10**(1), 193-203. [In eng]

Lukoschek, V., and Avise, J.C. (2011) Genetic monandry in 6 viviparous species of true sea snakes. *Journal of Heredity* **102**(3), 347-351.

Madsen, T., Shine, R., Loman, J., and Hakansson, T. (1992) Why do female adders copulate so frequently? *Nature* **355**(6359), 440-441.

McAlpin, S., Duckett, P., and Stow, A. (2011) Lizards cooperatively tunnel to construct a long-term home for family members. *PLOS ONE* **6**(5), e19041.

McLean, C.A., Moussalli, A., and Stuart-Fox, D. (2010) The predation cost of female resistance. *Behavioral Ecology* **21**(4), 861-867.

McLean, C.A., Stuart-Fox, D., and Moussalli, A. (2014) Phylogeographic structure, demographic history and morph composition in a colour polymorphic lizard. *Journal of Evolutionary Biology* **27**(10), 2123-2137.

Moore, J.A., Daugherty, C.H., Godfrey, S.S., and Nelson, N.J. (2009) Seasonal monogamy and multiple paternity in a wild population of a territorial reptile (tuatara). *Biological Journal of the Linnean Society* **98**(1), 161-170.

Natoli, A., Phillips, K.P., Richardson, D.S., and Jabado, R.W. (2017) Low genetic diversity after a bottleneck in a population of a critically endangered migratory marine turtle species. *Journal of Experimental Marine Biology and Ecology* **491**, 9-18.

Noble, D.W.A., Keogh, J.S., and Whiting, M.J. (2013) Multiple mating in a lizard increases fecundity but provides no evidence for genetic benefits. *Behavioral Ecology* **24**(5), 1128-1137.

Nunney, L. (1993) The influence of mating system and overlapping generations on effective population size. *Evolution* **47**, 1329+. [In English]

Olsson, M., Healey, M., Wapstra, E., Schwartz, T., Lebas, N., and Uller, T. (2007) Mating system variation and morph fluctuations in a polymorphic lizard. *Molecular Ecology* **16**.

Olsson, M., Schwartz, T., Uller, T., and Healey, M. (2007) Sons are made from old stores: sperm storage effects on sex ratio in a lizard. *Biology Letters* **3**(5), 491.

Olsson, M., Schwartz, T., Uller, T., and Healey, M. (2009) Effects of sperm storage and male colour on probability of paternity in a polychromatic lizard. *Animal Behaviour* **77**.

Pemberton, J.M., Slate, J., Bancroft, D.R., and Barrett, J.A. (1995) Nonamplifying alleles at microsatellite loci: a caution for parentage and population studies. *Molecular Ecology* **4**(2), 249-252.

Phillips, K.P., Jorgensen, T.H., Jolliffe, K.G., Jolliffe, S.-M., Henwood, J., and Richardson, D.S. (2013) Reconstructing paternal genotypes to infer patterns of sperm storage and sexual selection in the hawksbill turtle. *Molecular Ecology* **22**(8), 2301-2312.

Rankin, K.J., McLean, C.A., Kemp, D.J., and Stuart-Fox, D. (2016) The genetic basis of discrete and quantitative colour variation in the polymorphic lizard, *Ctenophorus decresii*. *BMC Evolutionary Biology* **16**(1), 179.

Rousset, F. (2008) genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources* **8**(1), 103–106.

Schwartz, T., Warner, D.B., LB, and Olsson, M. (2007) Microsatellite loci for Australian agamid lizards. *Molecular Ecology Notes* **7**(3), 528-531.

Smith, L., and Burgoyne, L. (2004) Collecting, archiving and processing DNA from wildlife samples using FTA® databasing paper. *BMC Ecology* **4**(1), 4.

Stow, A.J., and Sunnucks, P. (2004) High mate and site fidelity in Cunningham's skinks (*Egernia cunninghami*) in natural and fragmented habitat. *Molecular Ecology* **13**(2), 419-430.

Todd, E.V., Blair, D., Limpus, C.J., Limpus, D.J., and Jerry, D.R. (2013) High incidence of multiple paternity in an Australian snapping turtle (*Elseya albagula*). *Australian Journal of Zoology* **60**(6), 412-418.

Tolley, K.A., Chauke, L.F., Jackson, J.C., and Feldheim, K.A. (2014) Multiple paternity and sperm storage in the Cape Dwarf Chameleon (*Bradypodion pumilum*). *African Journal of Herpetology* **63**(1), 47-56.

Uller, T., and Olsson, M. (2008) Multiple paternity in reptiles: patterns and processes. *Molecular Ecology* **17**(11), 2566-80. [In eng]

Uller, T., Schwartz, T., Koglin, T., and Olsson, M. (2013) Sperm storage and sperm competition across ovarian cycles in the dragon lizard, *Ctenophorus fordi*. *J Exp Zool A Ecol Genet Physiol* **319**.

Uller, T., Stuart-Fox, D., and Olsson, M. (2010) Evolution of Primary Sexual Characters in Reptiles. In 'The evolution of primary sexual characters in animals.' (Eds. J Leonard and A Corodoba-Aguilar). (Oxford University Press: Oxford and New York)

van Schaik, J., and Kerth, G. (2017) Host social organization and mating system shape parasite transmission opportunities in three European bat species. *Parasitology Research* **116**(2), 589-599.

Wang, J. (2004) Sibship Reconstruction From Genetic Data With Typing Errors. *Genetics* **166**(4), 1963-1979.

Wang, J. (2007) Triadic IBD coefficients and applications to estimating pairwise relatedness. *Genetic Resources* **89**(3), 135-53. [In eng]

Wang, J. (2017) User's guide for software COLONY. In '.' 2.0.6.4 edn. (Zoological Society of London: London)

White, J., Richard, M., Massot, M., and Meylan, S. (2011) Cloacal Bacterial Diversity Increases with Multiple Mates: Evidence of Sexual Transmission in Female Common Lizards. *PLOS ONE* **6**(7), e22339.

Whitlock, M.C., and Bürger, R. (2004) 'Fixation of new mutations in small populations.' (Cambridge Univ. Press: Cambridge)

Wusterbarth, T.L., King, R.B., Duvall, M.R., Grayburn, S., and Burghardt, G.M. (2010) Phylogenetically widespread multiple paternity in new world natricine snakes. *Herpetological Conservation and Biology* **5**(1), 86-93.

Yewers, M.S.C. (2016) The function and evolution of colour polymorphism in the tawny dragon lizard. PhD Thesis, The University of Melbourne, Melbourne

York, J.R., and Baird, T.A. (2015) Testing the adaptive significance of sex-specific mating tactics in collared lizards (*Crotaphytus collaris*). *Biological Journal of the Linnean Society* **115**(2), 423-436.

Tables and Figures

Table 1. Microsatellite characteristics

k, number of alleles; N, number of individuals typed; H_o, observed heterozygosity; H_e, expected heterozygosity; PIC, polymorphic information content, Exclusion prob., mean probability of excluding an incorrect parent, given one parent is known; Error rate, estimated based on known parent-offspring mismatches (null allele rate is provided in brackets); HWE p-value, Hardy-Weinberg equilibrium exact test p-value. *Statistically significant at 0.05. ^HWE calculations are based on a subset of unrelated individuals (r <0.25, n=23).

Locus	k	N	H_o	H_e	PIC	Exclusion prob.	Error rate	HWE p-value^
Ctde03	20	103	0.699	0.890	0.875	0.773	0.025 (0.024)	0.045*
Ctde05	19	111	0.793	0.928	0.919	0.846	0.000 (0.000)	0.025*
Ctde08	15	111	0.775	0.899	0.886	0.789	0.033 (0.033)	0.517
Ctde12	18	112	0.893	0.907	0.896	0.806	0.000 (0.000)	0.374
Ctde21	13	105	0.571	0.866	0.847	0.723	0.063 (0.063)	0.043*
Ctde45	12	106	0.708	0.865	0.847	0.727	0.025 (0.025)	0.675
CP10	21	107	0.720	0.905	0.892	0.801	0.029 (0.000)	0.601
CP11	4	112	0.598	0.555	0.500	0.310	0.023 (0.000)	0.261

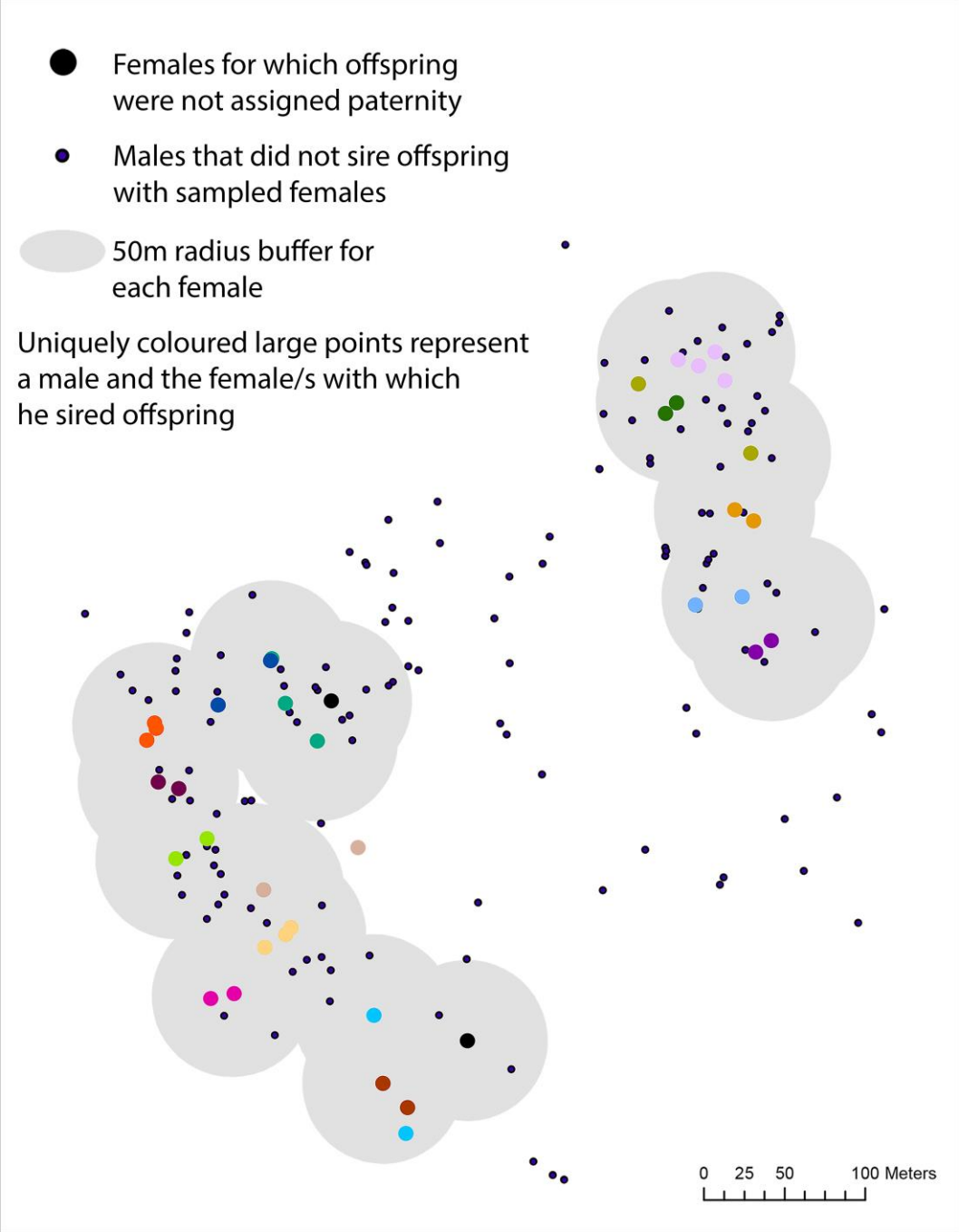


Figure 1. Spatial position of females (mothers) and all adult males sampled. One quarter of the males that sired offspring did so with more than one sampled female (range 1-3) and most clutches (96%) were sired by a single male.

Supplementary Information

Very low rate of multiple paternity detected in clutches of a wild agamid lizard

Hacking, J., Stuart-Fox, D., Gardner, M.

Australian Journal of Zoology, In Press, (2018).

Text S1. Details of captive hatching methodology and results

Captive hatching methods

The captive hatching schedule is provided in figure S1. After capture, gravid females were housed at the field station in Hawker for a maximum of three days before they were transported to Flinders University, Adelaide (five hour drive). During housing at the field station and transport to Adelaide, gravid females were secured in individual calico bags and placed in a portable electric fridge to keep them at approximately 15 °C, thereby reducing actively levels. The lid on the fridge was ajar for airflow. Once at Flinders University, gravid females were placed in individual enclosures.

Enclosures were made of large opaque plastic storage tubs (60 cm x 45 cm x 30 cm) with wire mesh lids and included UV lighting, basking lamps, warm and cool hides and nesting substrate (fig. S2). For female enclosures, nesting substrate was piled approximately 20cm high and 20cm wide up the side of one of the cool corners of the enclosure and consisted of a 70% sand and 30% peat moss mix. To ensure the nesting pile remained firm enough for females to dig safely, it was kept damp by spraying with water every second day. Three to six live small to medium sized crickets were provided every second day. Adult *C. decresii* were observed eating foliage at the Hawker field site, therefore shredded greens (bok choy, kale or cos lettuce) were provided once per week. Water was provided every second day by spraying both tiles and the sides of the container until pools of water formed. Gravid females were monitored twice daily and behaviour was recorded as either hiding, active (e.g. foraging or digging), basking or abnormal (e.g. running circles around enclosure, limping). Feeding records were kept, including the number of crickets provided, whether females were observed eating or foraging, and whether crickets remained from the previous feed.

Once females had laid their eggs they were removed from the enclosure temporarily while eggs were removed. Females were retained for approximately seven days after laying to increase body condition before release at their capture location. During transport females were secured in calico bags inside a portable electric fridge at 15 °C, with the lid ajar for airflow. Thirty minutes before release the

temperature of the portable fridge was increased to 25 °C so that females were active when released. Opportunistic sightings of females in the weeks following release were recorded to determine whether females remained close to the original capture location and whether they became gravid again later in the season.

Eggs were removed from enclosures and the number of eggs laid (clutch size), individual egg weight and clutch weight were recorded before they were placed in separate containers for incubation (fig. S3). Due to regular monitoring, eggs were placed in the incubator within 12 hours of laying. Incubating substrate, consisting of a damp vermiculite mixture of 1 parts water and 3 parts vermiculite, was placed in incubating containers. Eggs were incubated in an Exo Terra Reptile Incubator at 30°C, which is likely to produce an even sex ratio (Harlow 2000; pers. comm. Devi Stuart-Fox).

Once a hatchling was fully emerged from the egg standard morphometric measurements were taken and it was then housed in an individual enclosure. Enclosures consisted of plastic storage containers (30 cm x 20 cm x 11 cm) with half of the lid replaced with wire mesh. Enclosures included UV lighting, basking lamps, and warm and cool hides (fig. S4). Hatchlings were fed approximately 10 'baby' or 'pinhead' sized crickets every second day. Water was provided daily by spraying tiles and the sides of the enclosures. When weather permitted, hatchling enclosures were taken outside so that hatchlings could bask in sunlight. Monitoring was undertaken twice daily to record hatchling behaviour. Feeding records were also kept to ensure hatchlings were eating. Any hatchlings that did not eat after four days were hand fed for up to a week (3-4 feeds).

At hatching, six weeks and 12 weeks standard morphometric measurements were taken. Structural abnormalities (e.g. kinked tail) were recorded at hatching. The overall percentage of eggs that hatched and the percentage of hatchlings that survived to 12 weeks of age was calculated, along with mean growth rate (g/day) from hatching to 12 weeks of age and relative clutch weight (RCW). RCW is a measure of maternal investment as is calculated as total clutch weight divided by post-laying weight of the mother (Vitt and Congdon 1978). At approximately 12 weeks of age hatchlings were transported to Hawker and released into vacant rock crevices (checked with a torch) near their mothers capture location. During transport hatchlings were secured in calico bags inside a portable electric fridge at 15 °C, with the lid ajar to allow airflow. The temperature was increased to 25 °C 30 minutes prior to release so that hatchlings were active when released.

For both female and hatching enclosures, native foliage was provided to create structural heterogeneity and fine filtered sand was used as a substrate. Enclosures had a warm end and a cool end, each with a tile hide. At the warm end of the enclosure a basking spot was created during daylight hours (8am-5pm) with heat lamps providing 35°C, and UV lights (URS Outback Max 10% UVA/UVB)

providing adequate levels of UVA and UVB light. Overhead lighting was provided during daylight hours (7am-7pm). The cool end of the enclosure remained at approximately 28°C during daylight hours. Room temperature was set to 28°C during the day, and 23°C at night.

Strict quarantine protocols were followed to ensure that gravid females, hatchlings and other animals within the Flinders University Animal House were not exposed to pathogens. This allowed females and hatchlings to be released into the wild, hence reducing the impact on the population. Quarantine protocols included separate rooms for *C. decresii* and other animals within the Flinders University Animal House, disinfecting rooms before and after use, disinfecting enclosures before and after use, disinfecting calico bags used for temporarily holding lizards before and after use, disinfecting hands before and after handling, feeding and watering lizards and other animals, handling lizards only when necessary, and limiting handling of food items before feeding.

Captive hatching results

On average, females were held at the Hawker field station for one night before being transported to Flinders University. Female were then housed, on average, for 20 days before release back into the wild. Of the 27 females captured, four were likely not gravid and were released after 2 – 3 weeks. All gravid females (n = 23) buried their eggs within the enclosure and eggs were incubated within 12 hours of laying. After release, 12 females were opportunistically sighted (individual ID number written on back with non-toxic paint pen) throughout the field season on 1-3 different occasions. Of those females, two were observed to be heavily gravid again based on the presence of visible egg lumps. All resighted females were within 31 m of the original capture location (average, 10 m). Although released females were not actively targeted for resightings, these findings suggest that at least some females remained close to the original capture location and were reproductively active for the remainder of the breeding season.

Eggs were incubated for 49-65 days and average egg weight was 1.13 g. Average weight at hatching was 1.19 g and average SVL was 31 mm. This increased to 2.18 g and 36 mm, respectively, by 12 weeks of age. The mean growth rate among hatchlings was 0.012 g/day. Overall, hatching survival was high, with a mean of 95 % hatchling surviving to 12 weeks across clutches. Mean relative clutch weight (RCW) was 45 % and varied among clutches, ranging from 32 % to 67 % (table 1). Overall, 17 hatchlings (12%) presented structural abnormalities, with 10 clutches affected. Structural abnormalities included kinked tails and slightly concaved spines. Mean growth rate was similar between hatchlings with (0.010 g/day) and without (0.012 g/day) structural abnormalities, indicating that hatchling growth was not impacted.

One hatchling displayed a possible case of exomphalos and was euthanised immediately after hatching. Hatchlings were housed for a maximum of 110 days before release.

All females ate the crickets provided, with few crickets remaining after each feed. When water was provided some females were observed drinking from the surface of tiles although none were observed drinking from the sides of the enclosures. During captivity (excluding the first 4 days) 12 females predominantly (>50% of daily observations) displayed 'basking' behaviour and 14 females predominantly displayed 'hiding' behaviour. One female predominantly displayed 'active' behaviour (e.g. digging, foraging, regulating body temperature).

Most hatchlings ate crickets provided by the second or third day after hatching and drank water from the surface of tiles. Most hatchlings were observed displaying normal behaviour (e.g. basking, foraging and hiding/retreating). A small portion of the hatchlings (n=5) did not start eating crickets and remained inactive after the third day. Therefore, these individuals were hand-fed 3-5 pin-head sized crickets every second day for a week, after which they started eating for themselves. Despite efforts, these hatchlings didn't thrive and died at 4-6 weeks of age.

Table S1. Captive hatching statistics and hatchling morphometrics for *Ctenophorus decresii.* Relative clutch weight; total clutch weight divided by post-laying weight of mother.

	Mean	Minimum	Maximum
Number of eggs within a clutch	6	4	8
Incubation time (days) at 30°C	57	49	65
Percentage of clutch hatched	83	0	100
Egg weight (g)	1.13	0.93	1.51
Relative clutch weight (%)	45	32	67
Weight (g) at hatching	1.19	0.95	1.47
Weight (g) at 12 weeks	2.18	1.42	3.02
Snout-to-vent length (mm) at hatching	31	29	32
Snout-to-vent length (mm) at 12 weeks	36	30	40
Average hatchling growth rate (g/day)	0.012	0.002	0.020
Clutch hatchling survival to 12 weeks (%)	95	50	100

	2014				2015			
	September	October	November	December	January	February	March	April
Sampling male population Hawker, 1 week trips								
Gravid females collected Hawker, 1 week trips								
Egg incubation Flinders Uni Animal House								
Eggs hatching Flinders Uni Animal House								
Hatchling care Flinders Uni Animal House								
Female releasing Hawker, 1 week trips (Oct-Nov) Hawker, day/overnight trips (Dec)								
Hatchling releasing Hawker, day/overnight trips								

Figure S1. Captive hatching schedule used to obtain family group data for *Ctenophorus decresii*

Figure S2. Example of an enclosure used to house gravid female *Ctenophorus decresii*

Figure S3. A *Ctenophorus decresii* egg prepared for incubation

Figure S4. Example of *Ctenophorus decresii* hatchling enclosures

Table S2. Review of the rate of male and female multiple mating in squamate reptiles determined using molecular methods. Rate; percentage of individuals that mated multiply within a single mating season. Sample size is given in brackets. In all cases female multiple mating was assessed by the contribution of >1 sire to a single clutch or litter. Cases in which multiple paternity within clutches or litters was not observed are indicated with bolded text. Only studies undertaken on wild or free-ranging captive animals are included.

Species	Male multiple mating detected?	Rate (percentage of males)	Female multiple mating detected?	Rate (percentage of females)	Reference
		Scincidae			
Egernia whitii, White's skink	Yes	—	Yes	17% (NA)	While, Uller *et al.* (2009)
Egernia whitii, White's skink	Yes	—	Yes	12% (5/43)	Chapple and Keogh (2005)
Egernia stokesii, Gidgee skink	Yes	9% (2/23)	Yes	25% (4/16)	Gardner, Bull *et al.* (2002)
Egernia cunninghami, Cunningham's skink	Yes	12% (4/33)	Yes	3% (1/38)	Stow and Sunnucks (2004)*
Tiliqua rugosa, Sleepy lizard	Yes	18% (7/39)	Yes	19% (4/21)	Bull, Cooper *et al.* (1998)
Tiliqua adelaidensis, Pigmy blue-tongue	Yes	—	Yes	75% (18/24)	Schofield, Gardner *et al.* (2014)
Eulamprus heatwolei, Southern water skink	Yes	—	Yes	42% (5/12)	Keogh, Noble *et al.* (2012)
Eulamprus heatwolei, Southern water skink	Yes	—	Yes	82% (14/17)	Morrison, Scott Keogh *et al.* (2002)
Eulamprus quoyii, Eastern water skink	—	—	Yes	65% (41/63)	Noble, Keogh *et al.* (2013)^
Eulamprus leuraensis, Blue Mountains water skink	—	—	Yes	27% (11/41)	Dubey, Chevalley *et al.* (2011)
Oligosoma grande, Grand skink	Yes	70% (7/10)	Yes	47% (7/15)	Berry (2006)
Pseudomoia eurecateuixii, Mountain log skink	—	—	Yes	27% (3/11)	Stapley and Keogh (2006)^

Pseudomoia eurecateuixii, Mountain log skink	—	—	Yes	57% (9/17)	Stapley, Hayes *et al.* (2003)
Plestiodon fasciatus, Common Five-lined Skink	—	—	Yes	65% (13/20)	Bateson, Krenz *et al.* (2011)
Liopholis kintorei, **Great desert skink**	**Yes**	**40% (NA)**	**No**	**0% (NA)**	**McAlpin, Duckett *et al.* (2011)***
		Lacertidae			
Iberolacerta cyreni, Spanish rock lizard	Yes	—	Yes	48.5% (16/33)	Salvador, Díaz *et al.* (2008)
Lacerta vivipara, Common lizard	Yes	51% (22/43)	Yes	65% (17/26)	Hofmann and Henle (2006)
Lacerta agilis, Sand lizard	—	—	Yes	80% (4/5)	Gullberg, Olsson *et al.* (1997)
Podarcis muralis, Common wall lizard	—	—	Yes	87% (27/31)	Oppliger, Degen *et al.* (2007)
Gallotia bravoana, Giant lizard of La Gomera	Yes	60% (3/5)	Yes	27% (3/11)	Gonzalez, Cerón-Souza *et al.* (2014)
		Agamidae			
Amphibolurus muricatus, Jacky lizard	—	—	Yes	30% (20/67)	Warner, Woo *et al.* (2010)^
Ctenophorus pictus, Painted dragon	—	—	Yes	18% (9/51)	Uller and Olsson (2008)
Ctenophorus ornatus, Ornate dragon	Yes	—	Yes	25% (5/20)	Lebas (2001)
Intellagama lesueurii, Water dragon	—	—	Yes	86% (19/22)	Frère, Chandrasoma *et al.* (2015)
		Iguanidae			
Ctenosaura pectinate, Black spiny-tailed iguana	—	—	Yes	11% (1/9)	Faria, Zarza *et al.* (2010)
Crotaphytus collaris, Collard lizard	Yes	—	Yes	—	York and Baird (2015)
		Chamaeleonidae			
Bradypodion pumilum, Cape Dwarf Chameleon	—	—	Yes	100% (6/6)	Tolley, Chauke *et al.* (2014)
		Phrynosomatidae			

Sceloporus virgatus, Striped plateau lizard	Yes	—	Yes	61.5% (8/13)	Abell (1997)
Uta stansburiana, Side-blotched lizard	—	—	Yes	73% (71/97)	Zamudio and Sinervo (2000)[¥]
		Teiidae			
Ameiva exsul, Teiid lizard	Yes	—	Yes	—	Lewis, Tirado *et al.* (2000)^
		Diplodactylidae			
Oedura reticulate, Reticulated velvet gecko	Yes	97% (29/30)	Yes	100% (0/26)	Lange, Gruber *et al.* (2013)
		Dactyloidae			
Anolis cristatellus, Puerto Rican crested anole	—	—	Yes	30% (7/23)	Eales, Thorpe *et al.* (2010)
		Colubridae			
Nerodia sipedon, Northern water snake	—	—	Yes	62% (NA)	Kissner, Weatherhead *et al.* (2005)^
Nerodia sipedon, Northern water snake	Yes	—	Yes	58% (46/81)	Prosser, Weatherhead *et al.* (2002)^
Nerodia sipedon, Northern water snake	—	—	Yes	86% (12/14)	Barry, Weatherhead *et al.* (1992)
Nerodia rhombifer, Diamondback water snake	—	—	Yes	100% (1/1)	Wusterbarth, King *et al.* (2010)
Elaphe obsolete, Black rat snake	Yes	13% (5/39)	Yes	88% (30/34)	Blouin-Demers, Gibbs *et al.* (2005)
Natrix natrix, Grass snake	—	—	Yes	90% (10/11)	Meister, Ursenbacher *et al.* (2012)
Stegonotus cucullatus, Slatey-grey snake	Yes	—	Yes	—	Dubey, Brown *et al.* (2009)
Thamnophis elegans, Western terrestrial garter snake	—	—	Yes	50% (3/6)	Garner and Larsen (2005)

Thamnophis sirtalis parietalis, Red-sided garter snake	—	—	Yes	85% (50/59)	Friesen, Mason *et al.* (2014)
Thamnophis sirtalis, Garter snake	—	—	Yes	75% (6/8)	McCracken, Burghardt *et al.* (1999)
Thamnophis sirtalis, Garter snake	—	—	Yes	37.5% (6/16)	Garner, Gregory *et al.* (2002)
Thamnophis sirtalis, Garter snake	—	—	Yes	100% (4/4)	King, Milstead *et al.* (2001)
Thamnophis sirtalis, Garter snake	—	—	Yes	100% (1/1)	Wusterbarth, King *et al.* (2010)
Thamnophis melanogaster, **Blackbelly garter snake**	**—**	**—**	**No**	**0% (0/1)**	**Wusterbarth, King *et al.* (2010)**
Thamnophis sauritus, Ribbon snake	—	—	Yes	100% (1/1)	Wusterbarth, King *et al.* (2010)
Storeria dekayi, Brown snake	—	—	Yes	100% (1/1)	Wusterbarth, King *et al.* (2010)
Storeria occipitomaculata, Redbelly snake	—	—	Yes	100% (1/1)	Wusterbarth, King *et al.* (2010)
		Elapidae			
Enhydris enhydris, Rainbow water snake	**—**	**—**	Yes	100% (2/2)	Voris, Karns *et al.* (2007)
Enhydris subtaeniata, Mekong mud snake	**—**	**—**	Yes	100% (4/4)	
Hydrophis elegans, **Elegant sea snake**	**—**	**—**	**No**	**0% (0/2)**	**Lukoschek and Avise (2011)**
Hydrophis kingii, **King's sea snake**	**—**	**—**	**No**	**0% (0/1)**	
Hydrophis mcdowelli, **Small-headed sea snake**	**—**	**—**	**No**	**0% (0/3)**	
Hydrophis ocellatus, Spotted sea snake	**—**	**—**	**No**	**0% (0/1)**	
Hydrophis pacificus, Pacific sea snake	**—**	**—**	**No**	**0% (0/1)**	

***Lapemis curtus*, Shaw's sea snake**	—	—	No	0% (0/4)	
		Viperidae			
Vipera berus, Adder	Yes	29% (6/21)	Yes	69% (9/13)	Ursenbacher, Erny *et al.* (2009)
Crotalus atrox, Western diamond-backed rattlesnake	Yes	—	Yes	50% (12/24)	Clark, Schuett *et al.* (2014)
Crotalus horridus, Timber Rattlesnake	Yes[#]	—	Yes	43% (3/7)	Lind, Flack *et al.* (2016)
Gloydius halys, Halys pit viper	—	—	Yes	100% (1/1)	Simonov and Wink (2011)
		Pythonidae			
Liasis fuscus, Water python	—	—	Yes	14% (2/14)	Madsen, Ujvari *et al.* (2005)

^Study undertaken on a free-ranging captive population.

*Assessed from juveniles in family groups, not hatchlings

[#]Based on field observations

[¥]Based on KINSHIP program. Using the COLONY program multiple paternity was estimated at 68% (66/97)

Table S3. *Ctenophorus decresii* microsatellite genotypes. Paternity assignment probability is provided for each father in brackets and is based on the polygyny Colony model. The number of paternal alleles were inferred at each locus for each clutch. Maternal alleles are coloured green and paternal alleles are coloured red. Missing data or alleles affected by missing data are coloured purple and mismatches are coloured yellow. Evidence for multiple paternity was found for a single clutch (mother F_958_F).

		CP11		Ctde03		Ctde05		Ctde12		Ctde21		Ctde45		CP10		Ctde08	
Mother	F_1016_F	144	144	0	0	203	230	245	281	0	0	0	0	132	132	213	213
Father (1.000)	M_655_F	144	146	295	295	242	242	277	302	325	401	157	157	120	120	156	156
	H_1016-2_F	144	146	295	316	203	242	281	302	329	401	157	157	120	132	156	156
	H_1016-3_F	144	144	295	316	230	242	245	277	325	329	157	157	120	132	156	156
	H_1016-4_F	144	144	295	355	230	242	277	281	325	329	157	157	120	134	156	156
Number inferred paternal alleles:		1		NA		1		2		NA		NA		1 - 2		1	
Mother	F_1022_F	144	150	316	359	195	242	237	253	291	308	211	211	120	134	213	213
Father (0.949)	M_848_F	144	146	277	277	242	242	241	253	287	287	157	176	0	0	158	177
	H_1022-1_F	144	150	316	359	195	242	241	253	287	308	157	211	120	171	158	213
	H_1022-2_F	144	150	316	359	242	242	237	253	287	308	176	211	120	171	177	213
	H_1022-3_F	144	144	316	359	242	242	241	253	291	291	176	176	120	188	177	177
	H_1022-4_F	144	150	277	316	195	242	237	241	308	308	157	157	134	171	158	158
	H_1022-5_F	144	144	316	359	242	242	237	241	291	291	157	157	134	134	158	158
Number inferred paternal alleles:		1		2		1		2		2		2		2 - 3		2	
Mother	F_903_F	142	144	299	316	167	230	265	273	405	405	161	176	120	149	160	177
Father (1.000)	M_910_F	144	146	291	295	222	230	241	281	325	325	169	207	120	130	168	209
	H_903-2_F	0	0	291	316	167	222	265	281	401	405	0	0	120	120	177	209
	H_903-3_F	144	146	295	299	222	230	241	265	401	405	161	207	130	149	160	209
	H_903-4_F	142	144	291	316	167	230	273	281	325	325	161	169	120	149	160	168
	H_903-5_F	144	144	291	316	167	230	241	273	401	405	161	169	0	0	160	168
	H_903-6_F	142	146	295	316	222	230	265	281	325	325	176	207	130	149	177	209
Number inferred paternal alleles:		2		2		2		2		2		2		2		2	
Mother	F_906_F	144	150	295	295	210	246	281	297	291	325	161	161	134	140	160	160
Father (1.000)	M_609_F	144	144	0	0	167	222	237	285	291	291	145	207	120	138	144	209
	H_906-1_F	144	150	316	316	210	222	237	281	291	325	145	161	120	140	144	160
	H_906-2_F	144	144	316	316	210	222	237	297	291	325	161	207	120	140	160	209
	H_906-3_F	144	144	0	0	210	222	237	297	291	291	145	161	120	134	144	160
	H_906-4_F	144	150	0	0	210	222	281	285	291	291	161	207	134	138	160	209
	H_906-5_F	144	144	295	295	167	210	237	297	291	291	161	207	138	140	160	209
	H_906-6_F	144	150	295	295	167	210	285	297	291	291	145	161	138	140	144	160
Number inferred paternal alleles:		1		2		2		2		1		2		2		2	

		CP11		Ctde03		Ctde05		Ctde12		Ctde21		Ctde45		CP10		Ctde08	
Mother	F_913_F	144	144	332	367	222	239	281	285	291	291	0	0	130	146	197	201
Father (1.000)	M_668_F	144	144	328	336	167	167	273	289	325	325	157	211	0	0	156	213
	H_913-1_F	144	144	332	336	222	230	281	289	291	325	196	211	125	144	197	213
	H_913-2_F	144	144	332	336	167	222	273	281	291	325	200	211	120	120	201	213
	H_913-3_F	144	144	332	336	167	239	273	281	291	325	157	196	120	130	156	197
	H_913-4_F	144	144	328	332	222	230	273	281	291	325	157	196	146	168	156	197
	H_913-5_F	144	144	336	367	167	222	281	289	291	325	200	211	120	146	201	213
	H_913-6_F	144	144	336	367	167	239	285	289	291	325	196	211	130	168	197	213
	H_913-7_F	144	144	336	367	230	239	273	285	291	325	157	196	120	130	156	197
Number inferred paternal alleles:		1		2		2		2		1		NA		2		2	
Mother	F_915_F	144	144	295	316	239	239	277	281	287	295	157	207	134	152	158	209
Father (1.000)	M_609_F	144	144	0	0	167	222	237	285	291	291	145	207	120	138	144	209
	H_915-1_F	144	144	316	316	222	239	281	285	287	291	145	157	120	134	144	158
	H_915-2_F	144	144	295	316	222	239	281	285	287	287	157	207	120	152	158	209
	H_915-3_F	144	144	295	295	167	239	277	285	295	295	145	207	120	152	144	209
	H_915-5_T	144	144	316	316	167	239	277	285	287	291	207	207	120	152	209	209
	H_915-6_F	144	144	295	316	222	239	237	277	291	295	207	207	138	152	209	209
	H_915-7_F	144	144	295	295	167	239	281	285	287	287	145	157	120	134	144	158
Number inferred paternal alleles:		1		2		2		2		2		2		2		2	
Mother	F_916_F	144	144	299	320	167	230	253	273	325	325	157	161	120	149	158	160
Father (1.000)	M_648_F	142	144	316	355	230	239	285	289	0	0	211	211	125	130	213	213
	H_916-1_F	142	144	299	316	167	239	253	289	333	333	157	211	120	130	158	213
	H_916-2_F	144	144	316	320	167	230	273	285	325	325	157	211	125	149	158	213
	H_916-3_F	144	144	299	316	230	239	253	289	333	333	161	211	130	149	160	213
	H_916-4_F	144	144	0	0	167	239	0	0	325	325	0	0	120	125	160	213
	H_916-5_F	144	144	316	320	167	239	273	289	333	333	161	211	120	125	160	213
	H_916-6_F	144	144	299	316	167	239	273	285	333	333	157	211	130	149	158	213
	H_916-7_F	142	144	299	316	230	239	253	289	333	333	157	211	130	149	158	213
	H_916-8_F	144	144	316	320	167	230	273	285	333	333	161	211	120	130	160	213
Number inferred paternal alleles:		2		2		2		2		1		1		2		1	
Mother	F_922_F	144	144	295	305	218	230	237	241	291	295	0	0	130	144	209	213
Father (1.000)	M_631_F	144	144	291	295	211	259	265	289	287	287	176	211	134	142	177	213
	H_922-1_F	144	144	295	295	218	259	237	265	295	295	176	211	134	144	177	213
	H_922-2_F	144	144	291	305	230	259	237	265	287	295	207	211	130	142	209	213
	H_922-3_F	144	144	295	305	230	259	237	265	291	291	207	211	142	144	209	213
	H_922-4_F	144	144	295	295	211	230	237	289	287	295	176	207	130	142	177	209
	H_922-5_F	144	144	295	295	230	259	237	265	287	291	176	211	142	144	177	213
	H_922-6_F	144	144	291	305	218	259	241	289	287	291	176	211	142	144	177	213
	H_922-7_F	144	144	295	305	230	259	241	289	295	295	176	207	130	142	177	209
	H_922-8_F	144	144	295	295	230	259	241	289	295	295	207	211	142	144	209	213
Number inferred paternal alleles:		1		2		2		2		1		NA		2		2	

		CP11		Ctde03		Ctde05		Ctde12		Ctde21		Ctde45		CP10		Ctde08	
Mother	F_932_F	146	150	340	340	207	211	241	277	287	308	157	200	134	135	156	201
Father (1.000)	M_908_F	144	144	316	320	222	230	273	285	329	329	161	176	120	179	160	177
	H_932-1_F	144	146	316	340	211	222	277	285	0	0	176	200	135	135	177	201
	H_932-2_F	144	150	316	340	211	230	277	285	308	329	157	161	135	179	156	160
	H_932-3_F	144	146	316	340	211	222	241	285	308	308	157	176	120	134	156	177
	H_932-4_F	144	146	320	320	211	230	273	277	308	308	157	161	120	134	156	160
Number inferred paternal alleles:		1		2		2		2		2		2		2		2	
Mother	F_934_F	144	144	277	316	214	222	265	281	287	405	176	211	134	149	177	213
Father (1.000)	M_805_F	144	150	316	316	230	230	237	265	291	329	176	211	138	138	177	213
	H_934-1_F	144	144	316	316	222	230	265	265	329	405	176	211	138	149	177	213
	H_934-2_F	144	150	277	316	214	230	265	281	287	329	211	211	134	138	213	213
	H_934-3_F	144	150	277	316	222	230	265	281	287	329	176	176	149	149	177	177
	H_934-4_F	144	144	277	316	222	230	265	265	287	329	176	176	134	179	177	177
	H_934-5_F	144	150	277	316	222	230	265	281	287	329	176	211	149	179	177	213
	H_934-6_T	144	144	277	316	214	230	265	281	287	291	176	211	149	149	177	213
	H_934-7_F	144	150	316	316	214	230	265	265	329	405	211	211	149	179	213	213
	H_934-8_F	144	144	316	316	214	230	237	281	287	329	211	211	149	179	213	213
Number inferred paternal alleles:		2		1		1		2		2		2		2		2	
Mother	F_943_F	144	146	259	277	222	222	241	277	295	321	157	211	120	130	156	213
Father (0.999)	M_825_F	144	144	311	332	167	222	273	281	291	299	161	161	142	142	160	164
	H_943-1_F	144	144	277	311	167	222	241	273	299	321	157	161	120	142	156	160
	H_943-2_F	144	144	259	311	222	222	273	277	299	321	161	211	130	142	160	213
	H_943-3_F	144	146	277	311	167	222	241	273	295	299	161	211	130	142	160	213
	H_943-4_F	144	144	259	311	222	222	273	277	295	299	157	161	130	142	156	160
	H_943-5_F	144	146	259	311	222	222	241	273	299	321	211	211	120	142	164	213
	H_943-6_F	144	144	259	311	222	222	273	277	299	321	157	161	120	142	156	160
	H_943-7_F	144	144	277	311	167	222	277	281	299	321	161	211	130	130	160	213
	H_943-8_F	144	144	259	311	222	222	273	277	291	321	211	211	130	130	164	213
Number inferred paternal alleles:		1		1		2		1		2		1		1		2	
Mother	F_944_F	144	146	363	367	222	242	253	277	291	291	0	0	130	149	158	197
Father (1.000)	M_833_F	0	0	311	311	214	239	270	277	325	325	149	169	0	0	149	168
	H_944-1_F	144	144	311	363	222	239	277	277	291	325	149	196	130	146	149	197
	H_944-2_F	144	144	367	367	214	222	253	277	291	401	149	196	130	134	149	197
	H_944-3_F	144	146	363	363	214	242	270	277	291	401	169	196	130	146	168	197
	H_944-4_F	144	144	311	363	222	239	277	277	325	325	169	196	134	134	168	197
	H_944-5_F	144	144	363	363	239	242	277	277	291	325	157	169	146	149	158	168
	H_944-6_F	144	146	363	363	214	222	253	270	291	325	149	196	130	134	149	197
	H_944-7_F	144	146	311	363	214	222	270	277	291	401	157	169	130	134	158	168
Number inferred paternal alleles:		1		1		2		2		2		NA		2		2	

		CP11		Ctde03		Ctde05		Ctde12		Ctde21		Ctde45		CP10		Ctde08	
Mother	F_955_F	144	146	295	295	167	242	281	285	291	291	0	0	120	134	149	213
Father (0.998)	M_825_F	144	144	311	332	167	222	273	281	291	299	161	161	142	142	160	164
	H_955-1_F	144	144	295	332	222	242	273	285	291	299	149	161	134	142	149	160
	H_955-2_F	144	146	295	332	167	222	273	285	291	299	211	211	124	134	164	213
	H_955-3_F	144	144	332	355	167	242	273	281	291	291	149	161	134	134	149	160
	H_955-4_F	144	146	295	332	167	167	281	281	291	291	161	211	120	120	160	213
	H_955-5_F	144	146	295	332	167	242	281	281	291	299	161	211	134	142	160	213
	H_955-6_F	144	146	295	332	167	222	281	281	291	299	211	211	134	134	164	213
	H_955-7_F	144	144	295	332	167	222	281	285	291	291	211	211	134	134	164	213
Number inferred paternal alleles:		1		1		2		2		2		NA		2		2	
Mother	F_957_F	144	150	316	328	218	218	277	314	287	287	157	200	130	146	158	201
Father (1.000)	M_919_F	144	144	295	328	218	227	241	289	308	308	157	196	130	134	156	197
	H_957-1_F	144	144	295	316	218	218	241	277	333	333	157	196	130	134	158	197
	H_957-2_F	144	144	316	328	218	227	241	314	287	287	157	200	130	134	156	201
	H_957-3_F	144	144	316	328	218	227	289	314	287	308	157	157	130	146	156	158
	H_957-5_F	144	144	295	328	218	218	241	277	287	308	157	157	130	146	156	158
	H_957-6_F	144	144	295	316	218	227	241	314	333	333	157	200	134	146	156	201
	H_957-7_F	144	150	316	328	218	218	277	289	287	287	196	200	130	130	197	201
	H_957-8_F	144	144	295	316	218	218	277	289	308	308	157	200	130	146	156	201
Number inferred paternal alleles:		1		2		2		2		2		2		2		2	
Mother	F_960_F	144	144	305	336	227	230	281	281	325	325	157	196	134	144	156	197
Father (1.000)	M_668_F	144	144	328	336	167	167	273	289	325	325	157	211	0	0	156	213
	H_960-1_F	144	144	305	328	230	230	0	0	325	325	157	157	144	168	156	156
	H_960-2_T	144	144	305	336	167	230	273	281	325	325	157	157	120	144	156	156
	H_960-3_F	144	144	305	328	167	230	273	281	325	325	157	157	120	144	156	156
	H_960-4_F	144	144	305	328	230	230	273	281	325	325	196	211	120	134	197	213
	H_960-5_F	144	144	328	336	167	227	273	281	325	325	196	211	134	168	197	213
Number inferred paternal alleles:		1		2		2		1		1		2		2		2	
Mother	F_974_F	146	150	316	332	227	230	265	285	287	299	157	169	112	144	156	168
Father (1.000)	M_657_F	144	150	311	328	203	227	237	285	287	325	176	211	120	124	177	213
	H_974-1_F	144	150	311	316	203	227	0	0	299	325	169	176	120	144	168	177
	H_974-2_F	142	146	311	316	227	227	237	265	287	287	169	176	124	144	168	177
	H_974-3_F	144	146	316	328	203	230	285	285	299	325	157	211	120	144	156	213
	H_974-4_F	150	150	328	332	203	230	285	285	287	299	157	211	120	144	156	213
Number inferred paternal alleles:		2		2		2		2		2		2		2		2	

		CP11		Ctde03		Ctde05		Ctde12		Ctde21		Ctde45		CP10		Ctde08	
Mother	F_987_F	142	150	277	277	207	218	281	285	321	321	200	211	142	146	201	213
Father (0.834)	M_700_F	150	150	0	0	203	214	237	237	0	0	157	157	0	0	0	0
	H_987-1_F	142	150	277	316	214	218	237	285	321	325	157	200	120	146	158	201
	H_987-2_F	150	150	277	316	207	214	237	281	287	287	157	200	146	179	158	201
	H_987-4_F	0	0	277	320	207	214	237	281	321	325	157	200	120	146	158	201
	H_987-6_F	142	150	277	320	214	218	237	281	321	325	157	200	142	179	158	201
Number inferred paternal alleles:		1		2		1		1		1 - 2		1		2		1	
Mother	F_990_F	144	150	299	311	218	242	245	285	325	329	196	200	130	144	197	201
Father (1.000)	M_655_F	144	146	295	295	242	242	277	302	325	401	157	157	120	120	156	156
	H_990-1_F	144	150	295	299	218	242	245	277	329	401	157	196	120	144	156	197
	H_990-2_F	144	146	295	311	242	242	245	302	325	401	157	196	120	130	156	197
	H_990-3_T	146	150	295	299	242	242	285	302	325	329	157	196	120	130	156	197
	H_990-4_F	144	150	295	311	218	242	285	302	325	401	157	200	120	130	156	201
	H_990-5_F	146	150	295	311	218	242	277	285	329	401	157	200	120	130	156	201
	H_990-6_F	144	146	295	299	242	242	277	285	325	329	157	200	120	130	156	201
	H_990-7_F	144	144	295	311	218	242	285	302	325	401	157	200	120	130	156	201
Number inferred paternal alleles:		2		1		1		2		2		1		1		1	
Mother	F_995_F	144	144	277	299	211	222	249	270	299	325	161	211	134	149	160	213
Father (0.958)	M_857_F	144	144	316	316	183	222	237	281	291	308	157	157	122	122	158	158
	H_995-1_F	144	144	299	316	211	222	270	281	308	325	157	211	122	122	158	213
	H_995-2_F	144	144	299	316	183	222	270	281	299	308	211	211	134	134	164	213
	H_995-3_F	144	144	277	316	183	211	270	281	308	325	161	161	146	149	160	160
	H_995-4_F	144	144	277	316	183	222	249	281	299	308	161	161	146	149	160	160
	H_995-5_F	144	144	277	316	183	222	237	270	291	325	211	211	134	134	164	213
	H_995-6_F	144	144	299	316	183	211	270	281	299	308	157	211	149	149	158	213
Number inferred paternal alleles:		1		1		2		2		2		1		2		2	
Mother	F_997_F	144	150	0	0	222	222	277	285	287	287	157	161	124	125	158	160
Father (1.000)	M_655_F	144	146	295	295	242	242	277	302	325	401	157	157	120	120	156	156
	H_997-1_F	144	150	295	328	222	242	277	285	325	325	157	157	120	124	156	158
	H_997-2_F	144	146	295	328	222	242	285	302	325	325	157	161	120	124	156	160
	H_997-3_F	144	146	295	328	222	242	277	277	287	401	157	161	120	125	156	160
	H_997-4_T	146	150	295	328	222	242	285	302	287	401	157	161	120	124	156	160
	H_997-5_F	144	146	295	328	222	242	277	302	325	325	0	0	120	125	156	160
	H_997-6_F	144	150	295	311	222	242	277	302	325	401	157	161	120	124	156	160
	H_997-7_F	144	150	295	311	222	242	277	285	287	401	157	161	120	124	156	160
Number inferred paternal alleles:		2		NA		1		2		2		1		1		1	

		CP11		Ctde03		Ctde05		Ctde12		Ctde21		Ctde45		CP10		Ctde08	
Mother	F_998_F	142	150	295	311	207	210	277	281	308	321	157	176	134	142	156	177
Father (0.998)	M_841_F	144	144	311	328	203	227	237	285	325	325	196	211	120	124	197	213
	H_998-1_F	144	150	311	311	207	227	281	285	321	325	176	211	120	142	207	227
	H_998-2_F	142	144	311	328	210	227	237	281	321	401	157	196	120	142	156	197
	H_998-3_F	144	144	295	311	203	207	281	285	308	401	157	196	124	142	156	197
	H_998-4_F	144	150	295	311	203	210	277	285	308	401	176	196	120	142	177	197
	H_998-5_F	142	144	311	328	203	207	281	285	308	401	176	211	124	134	177	213
	H_998-6_F	142	144	295	328	210	227	237	281	308	325	157	196	124	134	156	197
Number inferred paternal alleles:		1		2		2		2		2		2		2		2	
		MULTIPLE PATERNITY															
Mother (father/s unknown)	F_958_F	144	144	336	363	167	259	249	273	325	325	157	207	120	134	156	209
	H_958-1_F	144	150	277	336	167	259	273	285	325	325	149	157	120	130	156	156
	H_958-2_F	144	144	277	363	167	203	249	285	321	325	149	207	120	120	149	209
	H_958-3_F	144	150	316	363	207	259	257	273	325	325	207	211	120	144	209	213
	H_958-4_F	144	150	316	363	214	259	249	257	325	325	196	207	134	149	197	209
	H_958-5_F	144	150	336	336	167	167	249	285	325	325	149	207	120	120	149	209
Number inferred paternal alleles:		1		2		3		2		1		3		3		3	
Mother (father/s unknown)	F_996_F	144	144	277	316	214	230	245	277	287	405	165	211	132	138	164	213
	H_996-1_F	144	144	299	316	230	242	277	281	287	287	196	211	130	138	197	213
	H_996-2_F	144	144	277	299	230	242	245	285	291	405	165	196	138	138	164	197
	H_996-3_F	144	144	316	332	214	242	277	281	287	291	196	211	138	168	197	213
	H_996-4_F	144	144	277	332	167	214	277	285	291	405	165	196	130	138	164	197
	H_996-5_F	144	144	277	332	214	242	277	281	287	291	157	165	138	168	156	164
	H_996-6_F	0	0	277	332	230	242	245	285	287	405	157	211	130	138	156	213
Number inferred paternal alleles:		1		2		2		2		1		2		2		2	

References

Abell, A.J. (1997) Estimating paternity with spatial behaviour and DNA fingerprinting in the striped plateau lizard, *Sceloporus virgatus* (Phrynosomatidae). *Behavioral Ecology and Sociobiology* **41**(4), 217-226.

Barry, F.E., Weatherhead, P.J., and Philipp, D.P. (1992) Multiple paternity in a wild population of northern water snakes, *Nerodia sipedon*. *Behavioral Ecology and Sociobiology* **30**(3), 193-199.

Bateson, Z.W., Krenz, J.D., and Sorensen, R.E. (2011) Multiple paternity in the common five-lined skink (*Plestiodon fasciatus*). *Journal of Herpetology* **45**(4), 504-510.

Berry, O.F. (2006) Inbreeding and promiscuity in the endangered grand skink. *Conservation Genetics* **7**(3), 427-437.

Blouin-Demers, G., Gibbs, H.L., and Weatherhead, P.J. (2005) Genetic evidence for sexual selection in black ratsnakes, *Elaphe obsoleta*. *Animal Behaviour* **69**(1), 225-234.

Bull, C.M., Cooper, S.J.B., and Baghurst, B.C. (1998) Social monogamy and extra-pair fertilization in an Australian lizard, *Tiliqua rugosa*. *Behavioral Ecology and Sociobiology* **44**(1), 63-72.

Chapple, D.G., and Keogh, J.S. (2005) Complex mating system and dispersal patterns in a social lizard, *Egernia whitii*. *Molecular Ecology* **14**(4), 1215-1227.

Clark, R.W., Schuett, G.W., Repp, R.A., Amarello, M., Smith, C.F., and Herrmann, H.-W. (2014) Mating systems, reproductive success, and sexual selection in secretive species: A case study of the western diamond-backed rattlesnake, *Crotalus atrox*. *PLOS ONE* **9**(3), e90616.

Dubey, S., Brown, G.P., Madsen, T., and Shine, R. (2009) Sexual selection favours large body size in males of a tropical snake (*Stegonotus cucullatus*, Colubridae). *Animal Behaviour* **77**(1), 177-182.

Dubey, S., Chevalley, M., and Shine, R. (2011) Sexual dimorphism and sexual selection in a montane scincid lizard (*Eulamprus leuraensis*). *Austral Ecology* **36**(1), 68-75.

Eales, J., Thorpe, R.S., and Malhotra, A. (2010) Colonization history and genetic diversity: adaptive potential in early stage invasions. *Molecular Ecology* **19**(14), 2858-2869.

Faria, C.M.A., Zarza, E., Reynoso, V.H., and Emerson, B.C. (2010) Predominance of single paternity in the black spiny-tailed iguana: conservation genetic concerns for female-biased hunting. *Conservation Genetics* **11**(5), 1645-1652.

Frère, C.H., Chandrasoma, D., and Whiting, M.J. (2015) Polyandry in dragon lizards: inbred paternal genotypes sire fewer offspring. *Ecology and Evolution* **5**(8), 1686-1692.

Friesen, C.R., Mason, R.T., Arnold, S.J., and Estes, S. (2014) Patterns of sperm use in two populations of Red-sided Garter Snake (*Thamnophis sirtalis parietalis*) with long-term female sperm storage. *Canadian Journal of Zoology* **92**, 33+. [In English]

Gardner, M.G., Bull, C.M., and Cooper, S.J.B. (2002) High levels of genetic monogamy in the group-living Australian lizard *Egernia stokesii*. *Molecular Ecology* **11**(9), 1787-1794.

Garner, T.W.J., Gregory, P.T., McCracken, G.F., Burghardt, G.M., Koop, B.F., McLain, S.E., Nelson, R.J., and McEachran, J.D. (2002) Geographic variation of multiple paternity in the common garter snake (*Thamnophis sirtalis*). *Copeia* **2002**(1), 15-23.

Garner, T.W.J., and Larsen, K.W. (2005) Multiple paternity in the western terrestrial garter snake, *Thamnophis elegans*. *Canadian Journal of Zoology* **83**(5), 656-663. [In English]

Gonzalez, E.G., Cerón-Souza, I., Mateo, J.A., and Zardoya, R. (2014) Island survivors: population genetic structure and demography of the critically endangered giant lizard of La Gomera, *Gallotia bravoana*. *BMC Genetics* **15**(1), 121.

Gullberg, A., Olsson, M., and Tegelström, H. (1997) Male mating success, reproductive success and multiple paternity in a natural population of sand lizards: behavioural and molecular genetics data. *Molecular Ecology* **6**(2), 105-112.

Harlow, P.S. (2000) Incubation Temperature Determines Hatchling Sex in Australian Rock Dragons (Agamidae: Genus *Ctenophorus*). *Copeia* **2000**(4), 958-964.

Hofmann, S., and Henle, K. (2006) Male reproductive success and intrasexual selection in the common lizard determined by DNA-microsatellites. *Journal of Herpetology* **40**(1), 1-6.

Keogh, J.S., Noble, D.W.A., Wilson, E.E., and Whiting, M.J. (2012) Activity predicts male reproductive success in a polygynous lizard. *PLOS ONE* **7**(7), e38856.

King, R.B., Milstead, W.B., Gibbs, H.L., Prosser, M.R., and et al. (2001) Application of microsatellite DNA markers to discriminate between maternal and genetic effects on scalation and behavior in multiply-sired garter snake litters. *Canadian Journal of Zoology* **79**(1), 121-128. [In English]

Kissner, K.J., Weatherhead, P.J., and Gibbs, H.L. (2005) Experimental assessment of ecological and phenotypic factors affecting male mating success and polyandry in northern watersnakes, *Nerodia sipedon*. *Behavioral Ecology and Sociobiology* **59**(2), 207.

Lange, R., Gruber, B., Henle, K., Sarre, S.D., and Hoehn, M. (2013) Mating system and intrapatch mobility delay inbreeding in fragmented populations of a gecko. *Behavioral Ecology* **24**(5), 1260-1270.

Lebas, N.R. (2001) Microsatellite determination of male reproductive success in a natural population of the territorial ornate dragon lizard, *Ctenophorus ornatus*. *Molecular Ecology* **10**(1), 193-203. [In eng]

Lewis, A.R., Tirado, G., and Sepulveda, J. (2000) Body size and paternity in a Teiid lizard (*Ameiva exsul*). *Journal of Herpetology* **34**(1), 110-120.

Lind, C.M., Flack, B., Rhoads, D.D., and Beaupre, S.J. (2016) The mating system and reproductive life history of female timber rattlesnakes in northwestern Arkansas. *Copeia* **104**(2), 518-528.

Lukoschek, V., and Avise, J.C. (2011) Genetic monandry in 6 viviparous species of true sea snakes. *Journal of Heredity* 102(3), 347-351.

Madsen, T., Ujvari, B., Olsson, M., and Shine, R. (2005) Paternal alleles enhance female reproductive success in tropical pythons. *Molecular Ecology* 14(6), 1783-1787.

McAlpin, S., Duckett, P., and Stow, A. (2011) Lizards cooperatively tunnel to construct a long-term home for family members. *PLOS ONE* 6(5), e19041.

McCracken, G.F., Burghardt, G.M., and Houts, S.E. (1999) Microsatellite markers and multiple paternity in the garter snake *Thamnophis sirtalis*. *Molecular Ecology* 8(9), 1475-9. [In eng]

Meister, B., Ursenbacher, S., and Baur, B. (2012) Frequency of multiple paternity in the grass snake (*Natrix natrix*). *Amphibia-Reptilia* 33(2), 308-312.

Morrison, S.F., Scott Keogh, J., and Scott, I.A.W. (2002) Molecular determination of paternity in a natural population of the multiply mating polygynous lizard *Eulamprus heatwolei*. *Molecular Ecology* 11(3), 535-545.

Noble, D.W.A., Keogh, J.S., and Whiting, M.J. (2013) Multiple mating in a lizard increases fecundity but provides no evidence for genetic benefits. *Behavioral Ecology* 24(5), 1128-1137.

Oppliger, A., Degen, L., Bouteiller-Reuter, C., and John-Alder, H.-B. (2007) Promiscuity and high level of multiple paternity in common wall lizards (*Podarcis muralis*): data from microsatellite markers. *Amphibia-Reptilia* 28(2), 301-303.

Prosser, M.R., Weatherhead, P.J., Gibbs, H.L., and Brown, G.P. (2002) Genetic analysis of the mating system and opportunity for sexual selection in northern water snakes (*Nerodia sipedon*). *Behavioral Ecology* **13**(6), 800-807.

Salvador, A., Díaz, J.A., Veiga, J.P., Bloor, P., and Brown, R.P. (2008) Correlates of reproductive success in male lizards of the alpine species Iberolacerta cyreni. *Behavioral Ecology* **19**(1), 169-176.

Schofield, J.A., Gardner, M.G., Fenner, A.L., and Michael Bull, C. (2014) Promiscuous mating in the endangered Australian lizard *Tiliqua adelaidensis*: a potential windfall for its conservation. *Conservation Genetics* **15**(1), 177-185.

Simonov, E., and Wink, M. (2011) Cross-amplification of microsatellite loci reveals multiple paternity in Halys pit viper (*Gloydius halys*). *Acta Herpetologica* **6**(2).

Stapley, J., Hayes, C.M., and Scott Keogh, J. (2003) Population genetic differentiation and multiple paternity determined by novel microsatellite markers from the Mountain Log Skink (*Pseudemoia entrecasteauxii*). *Molecular Ecology Notes* **3**(2), 291-293.

Stapley, J., and Keogh, J.S. (2006) Experimental and molecular evidence that body size and ventral colour interact to influence male reproductive success in a lizard. *Ethology Ecology & Evolution* **18**(4), 275-288.

Stow, A.J., and Sunnucks, P. (2004) High mate and site fidelity in Cunningham's skinks (*Egernia cunninghami*) in natural and fragmented habitat. *Molecular Ecology* **13**(2), 419-430.

Tolley, K.A., Chauke, L.F., Jackson, J.C., and Feldheim, K.A. (2014) Multiple paternity and sperm storage in the Cape Dwarf Chameleon (*Bradypodion pumilum*). *African Journal of Herpetology* 63(1), 47-56.

Uller, T., and Olsson, M. (2008) Multiple paternity in reptiles: patterns and processes. *Molecular Ecology* 17(11), 2566-80. [In eng]

Ursenbacher, S., Erny, C., and Fumagalli, L. (2009) Male reproductive success and multiple paternity in wild, low-density populations of the adder (*Vipera berus*). *Journal of Heredity* 100(3), 365-370.

Vitt, L.J., and Congdon, J.D. (1978) Body Shape, Reproductive Effort, and Relative Clutch Mass in Lizards: Resolution of a Paradox. *The American Naturalist* 112(985), 595-608.

Voris, H.K., Karns, D.R., Feldheim, K.A., Kechavarvi, B., and Rinehart, M. (2007) Multiple paternity in the oriental-australian rear-fanged watersnakes (homalopsidae). *Herpetological Conservation and Biology* 3(1), 88-102.

Warner, D.A., Woo, K.L., Van Dyk, D.A., Evans, C.S., and Shine, R. (2010) Egg incubation temperature affects male reproductive success but not display behaviors in lizards. *Behavioral Ecology and Sociobiology* 64(5), 803-813.

While, G.M., Uller, T., and Wapstra, E. (2009) Within-population variation in social strategies characterize the social and mating system of an Australian lizard, *Egernia whitii*. *Austral Ecology* 34(8), 938-949.

Wusterbarth, T.L., King, R.B., Duvall, M.R., Grayburn, S., and Burghardt, G.M. (2010) Phylogenetically widespread multiple paternity in new world natricine snakes. *Herpetological Conservation and Biology* **5**(1), 86-93.

York, J.R., and Baird, T.A. (2015) Testing the adaptive significance of sex-specific mating tactics in collared lizards (*Crotaphytus collaris*). *Biological Journal of the Linnean Society* **115**(2), 423-436.

Zamudio, K.R., and Sinervo, B. (2000) Polygyny, mate-guarding, and posthumous fertilization as alternative male mating strategies. *Proceedings of the National Academy of Sciences* **97**(26), 14427-14432.

CHAPTER 4

De novo genotyping of the MHC in an Australian dragon lizard aided by family group data

Note to examiners

This chapter presents the methodology used to genotype *Ctenophorus decresii* individuals at the MHC class I region and describes the diversity present. Family group data produced in chapter 3 aided in genotyping and was used to investigate the segregation of MHC alleles. This chapter has been submitted to *Transactions of the Royal Society of South Australia* and is therefore formatted according to the specifications for that journal. Please refer to Appendix 1 for additional details regarding transcriptome assembly and primer design. Appendix 1 is a thesis-only appendix and is therefore not present in the publication associated with this chapter.

Hacking, J., Bradford, T., Pierce, K., Gardner, M., (2018). *De novo* genotyping of the MHC in an Australian dragon lizard aided by family group data. *Manuscript submitted for publication.*

De novo genotyping of the major histocompatibility complex in an Australian dragon lizard aided by family group data

Jessica Hacking[a] (corresponding author)
jessica.hacking@flinders.edu.au

Tessa Bradford[a,b,c]
tessa.bradford@samuseum.sa.gov.au

Kelly Pierce[a]
kellyannepierce@gmail.com

Michael Gardner[a,b]
michael.gardner@flinders.edu.au

[a]College of Science and Engineering, Flinders University, Bedford Park SA 5042, Australia

[b]Evolutionary Biology Unit, South Australian Museum, Adelaide SA 5000, Australia

[c]School of Biological Sciences, University of Adelaide, Adelaide, SA 5005, Australia

Abstract

The major histocompatibility complex (MHC) is a hypervariable gene family that plays an essential role in the recognition of pathogens and immune response. Research on the reptilian MHC has lagged behind other vertebrate groups. Here, we genotyped individuals of an Australian agamid lizard species at MHC class I loci using a recently developed clustering method and family group data. Our method allowed identification of low amplification efficiency alleles and estimation of both type I and II genotyping error rates. The number of MHC class I alleles per individual varied within populations and together with allele segregation patterns, suggests either natural copy number variation or allele dropout. Genotypes from individuals across five populations revealed shared alleles among populations and low allelic diversity in an island population. Finally, we identified sites under selection and designated putative peptide binding sites. Our results provide a foundation for future work on the MHC class I region of agamid lizards.

Keywords: Agamidae; Next generation amplicon sequencing; Allele segregation; AmpliSAT; Peptide binding site; Copy number variation

Short title: *De novo* MHC genotyping for an Australian agamid lizard

Introduction

The major histocompatibility complex (MHC) is a hypervariable gene region that plays a central role in the adaptive immune system of all jawed vertebrates (Piertney and Oliver, 2006). Genes of the MHC encode cell surface glycoproteins that bind to self- and pathogen-derived peptides, presenting them to circulating T-lymphocyte cells. Diversity at the MHC is manifested at multiple levels: it is polygenic, often possesses many duplicated loci, and is usually extremely polymorphic within loci (Janeway, Travers, Walport, & Shlomchik, 2005). Furthermore, the total number of MHC loci may differ among individuals within and among populations; a phenomenon termed copy number variation (CNV; De Groot, Blokhuis, Otting, Doxiadis, & Bontrop, 2015; Lighten, Oosterhout, & Bentzen, 2014; Schrider and Hahn, 2010). Accurate MHC genotyping of non-model species is challenging due to this complex genomic architecture. However, individuals can be genotyped at the MHC with a high degree of confidence using high-throughput next-generation sequencing (NGS) technologies along with thorough post-sequencing quality control (Lighten, et al., 2014; Sebastian, Herdegen, Migalska, & Radwan, 2016; Stutz and Bolnick, 2014).

Next-generation amplicon sequencing has become the method of choice when quantifying diversity at the MHC, particularly in the case of non-model organisms (reviewed in Lighten, et al., 2014). Next generation sequencing allows hundreds to thousands of amplicons (multiplexing) to be screened hundreds to thousands of times each (deep sequencing). However, with an increase in complexity and read depth comes an increase in errors, with artefact sequences generated both during PCR amplification and NGS (Lighten, et al., 2014; Sommer, Courtiol, & Mazzoni, 2013). Artefacts produced during PCR are predominantly due to random base-mismatch errors and chimera formation (Lenz and Becker, 2008), whereas artefact sequences produced during NGS are usually due to substitution-type miscalls (Liu et al., 2012; Loman et al., 2012; Ross et al., 2013). An additional source of error is cross-sample contamination during NGS library preparation (Biedrzycka, Sebastian, Migalska, Westerdahl, & Radwan, 2016) and tag-swapping during sequencing (Schnell, Bohmann, & Gilbert, 2015). As a result, thorough post-sequencing error correction and the inclusion of sample replicates is required to distinguish true alleles from artefactual sequences when using next-generation amplicon sequencing technologies to genotype individuals. In relation to the MHC (and other complex gene families), family group data can be used to verify genotypes and investigate MHC allele segregation patterns (Gaigher et al., 2016; Pearson, Bradford, Ansari, Bull, & Gardner, 2016). Another challenge associated with MHC genotyping, which is not necessarily limited to NGS methodologies, is assigning MHC alleles to their respective loci. The complex structure of the MHC region (e.g. CNV), the potential for allele sharing among loci, and the use of primers that amplify across multiple loci, inhibits the ability to assign MHC alleles to loci in non-model species (Lighten, et al., 2014).

Four inter-related issues need to be considered when distinguishing true alleles from artefacts within next-generation amplicon sequencing datasets. These are uneven amplification of true alleles, high read depth for artefact sequences, highly similar true alleles and the presence of a true allele in only some amplicons (summarised in table 1). Several post-sequencing error correction protocols have been developed to genotype individuals at the MHC (reviewed in Lighten, et al., 2014). Many of the earlier protocols do not take several of the aforementioned issues into account (table 1, Babik, Taberlet, Ejsmond, & Radwan, 2009; Galan, Guivier, Caraux, Charbonnel, & Cosson, 2010; Zagalska-Neubauer et al., 2010), or require every amplicon to be independently replicated (Sommer, et al., 2013); a costly process that reduces the number of amplicons that can be sequenced for a given amount of research funding. A recently developed clustering method (Sebastian, et al., 2016; Stutz and Bolnick, 2014), however, significantly contributes towards overcoming the issues associated with accurately and efficiently genotyping individuals at the MHC. This clustering method is based on the stepwise threshold clustering method recently developed by Stutz and Bolnick (2014) and was implemented as an online tool (AmpliSAS) as part of the AmpliSAT suite by Sebastian, et al. (2016).

AmpliSAS has been found to outperform other genotyping methods (Biedrzycka, et al., 2016; Sebastian, et al., 2016) and its flexible framework is suited to *de novo* genotyping of multigene families in non-model organisms. During AmpliSAS clustering genotypes are resolved on an amplicon-by-amplicon basis; taking into account unequal amplification among samples and among alleles, rather than applying a global read depth frequency threshold (first and fourth issue, table 1). Variants are ordered by read depth and each variant either forms a cluster and becomes a dominant sequence or is deemed an artefact sequence based on clustering parameters. Clustering parameters include user-defined substitution and indel error rates, sequence length parameters, and a dominance frequency threshold at which a sequence within a cluster is allowed to form its own cluster, becoming another dominant sequence. Dominant sequences that then pass user-defined filtering parameters, such as excluding very low frequency variants and chimeras, are considered true alleles. When a variant is classified as an artefact its read depth is added to that of the dominant sequence from which it likely originated. This helps to separate true alleles that have lower read depths and artefact sequences that occur at higher frequencies (second issue, table 1). Finally, the dominance frequency threshold applied within clusters allows the discrimination of true alleles that are relatively similar (third issue, table 1). In addition, the AmpliCHECK pre-genotyping tool, which is also part of the AmpliSAT suite (Sebastian, et al., 2016), is extremely useful for *de novo* genotyping in non-model systems as it provides preliminary information on amplicon read depth, variant frequencies and estimated error rates.

While the MHC is well characterised in many mammal model species (e.g. humans and other primates, Wang and He 2006, Bontrop et al. 1999), the chicken (Kaufman, 2015), *Xenopus* frogs (Ohta,

Goetz, Hossain, Nonaka, & Flajnik, 2006) and model fish species (Dirscherl, McConnell, Yoder, & Jong, 2014; Grimholt et al., 2015), little is known regarding the genomic structure of and diversity present at the reptilian MHC (Elbers and Taylor, 2016). We previously characterised eight MHC class I (henceforth MHC I) transcripts in an Australian dragon lizard, *Ctenophorus decresii* (Hacking, et al., 2018). It is likely that six of these transcripts (*Ctde-UA*001 – Ctde-UA*006*) undertake classical MHC I functions within the immune system, although confirmation of high polymorphism (allelic diversity) and sites under positive selection is required to further validate classical functionality. First, we genotyped individuals across five geographically distinct locations using a recently developed clustering method (Sebastian, et al., 2016), which was aided by the analysis of family group data. Second, we used family group data to examine allele segregation patterns. Finally, we briefly described diversity among and within populations and identified sites that are under selection, allowing the identification of putative peptide binding sites. This work furthers knowledge on the squamate MHC I region and provides a foundation for future work investigating the factors shaping MHC diversity in *C. decresii*.

Materials and Methods

Sample collection and DNA extraction

The tawny dragon is a small agamid lizard endemic to rocky ranges within South Australia (McLean, Moussalli, & Stuart-Fox, 2013). During spring and summer between 2013 and 2015 we undertook field surveys at five populations, spanning most of the current distribution of the tawny dragon: Hawker (31°57’17.5”S, 138°22’26.4”E), Mount Remarkable (32°50’32.1”S, 138°03’27.7”E), Barossa (34°33’26.5”S, 139°12’12.7”E), Morialta (34°54’14.9”S, 138°42’13.3”E), and Kangaroo Island (35°57’46.2”S, 136°39’15.9”E). Hawker and Mount Remarkable represent the ‘northern’ lineage, and Morialta and Kangaroo Island represent the ‘southern’ lineage, with Barossa at the contact zone (Fig. 1, McLean, Stuart-Fox, & Moussalli, 2014). Blood (≤100µL) was collected from captured individuals via venipuncture of the *vena angularis*, accessed from the corner of the mouth before release at the capture location. Whole blood was preserved on Whatman® Classic FTA cards (GE Healthcare, Buckinghamshire, UK) and stored at room temperature until DNA extraction, which was completed according to Smith and Burgoyne (2004). Family group data was collected using a captive hatching method, as described in Hacking, Stuart-Fox, & Gardner (In Press). A total of 21 complete family groups with three to eight (average of six) hatchlings each were obtained and a predominantly polygynous mating system was described (Hacking, et al., In Press).

Determining genotyping thresholds

Illumina MiSeq library preparation and pre-genotyping bioinformatics methodology is provided as supplementary material. AmpliCHECK and AmpliSAS, which are part of AmpliSAT, a suite of next generation amplicon sequencing analysis tools, were used to genotype individuals (Sebastian, et al., 2016). The genotyping process is summarised in figure S3. First, AmpliCHECK analysis and preliminary AmpliSAS analysis was undertaken in order to determine optimal clustering and filtering thresholds for final AmpliSAS genotyping analysis, which is described below (genotyping section). AmpliCHECK provides an initial analysis of amplicon sequencing results. Reads are de-multiplexed and collapsed into unique variants per amplicon. For each variant, the depth (number of contributing reads), frequency (percentage of the total reads assigned to the variant) and coverage (number of samples in which the variant is present) is calculated. The number of variants present within each amplicon, the total number of reads assigned to each amplicon, and the frequency and depth of each variant within each amplicon (PAF; per amplicon frequency) are also calculated. The mean, minimum and maximum PAF across amplicons is provided for each variant. Furthermore, AmpliCHECK categorises each variant as either 'good', 'suspicious' or 'probably artefact' based on estimated error rates and comparisons among amplicons. A variant is classified as 'probably artefact' if it can be explained as a base mismatch or a chimera, or a chimera with mismatches (putative accumulated errors) in all of the amplicons in which it is present. A variant is identified as chimeric only if parent sequences making up the chimera are present in the same amplicon. Variants are classified as 'suspicious' if they can be explained as an artefact in only some amplicons and 'good' variants aren't identified as artefacts in any amplicons.

AmpliCHECK analysis was performed for both α1 and α2 with a substitution error rate of 1% and an indel error rate of 0.001% (AmpliCHECK default for Illumina data), and a minimum PAF cutoff of 1%. Substitution and indel error rates are specific to the Illumina platform and are based on findings of previous NGS studies (Sebastian, et al., 2016). A preliminary AmpliSAS analysis was also undertaken for comparison with AmpliCHECK results. The maximum number of alleles per amplicon was set to 10 (based on AmpliCHECK results) and the minimum amplicon depth was set to 100. Any amplicons with fewer than 100 reads were excluded. Default parameters were used for clustering (a dominance frequency threshold of 25%) and a 4% PAF threshold was used for filtering. Chimeras were identified and discarded by the program. Read depth was capped at 5000 for all amplicons for both AmpliCHECK and AmpliSAS analysis (maximum allowed by AmpliCHECK and AmpliSAS). As most amplicons had a read depth lower than 5000 (see results), this was not an issue. Within the AmpliCHECK and preliminary AmpliSAS datasets, family groups and technical replicates were examined in order to determine optimal clustering and filtering thresholds for the final AmpliSAS analyses for α1 and α2.

Analysing family group data provided the opportunity to confirm artefact sequences identified by AmpliCHECK and identify additional artefact sequences. Family group data also allowed the identification of putative low amplification efficiency alleles (LAEA) within the dataset. Prior to analysing family groups and technical replicates, all variants with very low frequencies (< 0.03 PAF) were assumed to be artefacts and were excluded from the dataset (Herdegen, Babik, & Radwan, 2014). An artefact sequence within a family group was defined as a variant that only occurred in one or more offspring but not parents. A LAEA was defined as a variant that was present in one or both parents and offspring but had a low frequency (< 0.05 PAF) in >50% of family members. Furthermore, these variants must have had a mean PAF lower than the overall mean PAF for all putative true alleles, based on preliminary AmpliSAS analysis (i.e. LAEA did not reach high PAF in any amplicon). These putative LAEA could be a result of low primer annealing efficiency or could represent artefact sequences that contain several accumulated errors. In cases where a variant occurred in only parents and could not be explained as an error or a LAEA, it was assumed to be a true allele that was not passed to offspring. For each family group, the lowest PAF of a true allele, and the highest PAF of an artefact sequence were recorded.

To determine the optimal PAF filtering threshold for final AmpliSAS analyses, the rates of type I (incorrect rejection of a true allele) and type II (incorrect acceptance of an artefact sequence) error were calculated using family group data under different PAF filtering threshold values, ranging from 3% to 6%. For each PAF filtering threshold (3% to 6%) the type I genotyping error rate was calculated as the average percentage of individuals within family groups with putative true alleles with lower frequencies than the given PAF filtering threshold. The type II genotyping error rate was calculated as the average percentage of offspring within a family group possessing artefact sequences with a higher frequency than the given PAF filtering threshold. LAEA and artefact sequences containing putative accumulated errors (identified by AmpliCHECK and family group analysis) were excluded when calculating genotyping error rates as these variants were removed during the genotyping process (see below). A PAF filtering threshold that resulted in the lowest combined type I and type II genotyping error rates without unacceptably inflating either type I or II error was used for final AmpliSAS analyses. Technical replicates were used to estimate the overall genotyping error rate.

Genotyping

For the final AmpliSAS analysis, the maximum number of alleles at each amplicon was set to 10 (based on AmpliCHECK and preliminary AmpliSAS analyses). The minimum number of reads required for an amplicon was 100 and the maximum (cutoff) was set to 5000. For clustering, a dominance frequency threshold of 25% was used with a substitution error rate of 1% and an indel error rate of 0.001% (AmpliSAS default for Illumina data). A PAF threshold of 5% was used for $\alpha 1$ and a PAF threshold of 4% was used for $\alpha 2$ during the filtering step (see results). A dominance frequency threshold of 25% allowed

any variant (that isn't identified as an artefact) within a cluster that had at least 25% of the reads of the dominant variant within that cluster, to become a dominant variant in its own cluster. All reads within a cluster were assigned to the dominant variant. This process was repeated within each new cluster. After clustering, any variant with a PAF lower than 5% for α1 or 4% for α2 was removed. Post-genotyping filtering was undertaken to remove artefact sequences identified during family group analysis that were not removed during genotyping, pseudoalleles, and LAEA (Fig. S3). AmpliSAS genotyping analysis does not exclude variants that are chimeras with mismatches (potential accumulated errors). During AmpliSAS clustering, base mismatches (based on the substitution error rate provided) are identified and then chimeras are excluded during the filtering step. Therefore, variants that are chimeras with mismatches may not have the mismatches identified during clustering, and are not identified as chimeras during filtering. This is likely because such variants could be natural recombinants. However, almost all such variants with potential accumulated errors that were identified in the AmpliCHECK analysis as 'probably artefacts' were confirmed as artefacts during analysis of family group data (see results). Consequently, all other variants considered 'probably artefacts' by AmpliCHECK due to accumulated errors that were not represented in family groups, were conservatively removed from the final dataset. All putative alleles were aligned and translated in Geneious and any alleles with premature stop codons were classified as pseudoalleles and removed from the dataset. All LAEA identified during family group analysis were conservatively removed from the final dataset due to the uncertainty surrounding these variants.

To determine whether the sequencing depth was sufficient for genotyping, the relationship between the number of alleles and read depth within amplicons for both α1 and α2 was examined using a simple linear regression in R ver. 3.3.1 (R Core Team, 2016).

Segregation analysis

Segregation patterns were examined for MHC I α1 and α2 alleles for each family group, taking LAEA into account. Mendelian inheritance was assumed and where alleles did not follow simple segregation patterns or where parents possessed differing numbers of alleles, a model of CNV and resulting hemizygosity was assessed. Where possible, alleles were assigned to the same locus across family groups.

Genetic diversity and tests for selection

Percent pairwise nucleotide and amino acid identity was calculated using Geneious for all alleles within the dataset and within each population. Amino acid variability (Wu-Kabat variability coefficient) was calculated using the Protein Variability Server (Garcia-Boronat, Diez-Rivero, Reinherz, & Reche, 2008) across all α1 and α2 alleles. The HyPhy Datamonkey web server (Delport, Poon, Frost, & Kosakovsky

Pond, 2010) was used to test for sites under positive and negative selection across all α1 and α2 alleles, using multiple models. Firstly, a model of codon evolution was determined using Codon Model Selection (CMS) and a test for recombination was performed (SBP analysis, Kosakovsky Pond, Posada, Gravenor, Woelk, & Frost, 2006). Recombination was taken into account when testing for positive and negative selection using the SLAC, REL, FEL (Kosakovsky Pond and Frost, 2005) and MEME (Murrell et al., 2012) models. The REL model was not used for α2 alleles due to an alignment size restriction. Each of these models differ in the way they test for selection, and only MEME accounts for episodic diversifying selection. Consensus between at least two models was required for a site to be considered under selection.

Two Bayesian phylogenetic trees were created based on aligned α1 and α2 nucleotide sequences using MrBayes ver. 3.2.6 (Ronquist et al., 2011). Each tree included α1 or α2 alleles identified in this study and the α1 or α2 domain of eight previously described transcripts (Hacking, et al., 2018). The α1 or α2 domain of marine iguana (*Amblyrhynchus cristatus*) MHC I sequence (*Amcr-UB*01*, EU604308.1) was used as an outgroup in each tree. PartitionFinder2 (Guindon and Gascuel, 2003; Lanfear, Calcott, Ho, & Guindon, 2012; Lanfear, Frandsen, Wright, Senfeld, & Calcott, 2017) was employed to determine optimal partitioning and the best model of evolution. Accordingly, a GTR+I+G model was used for Bayesian analysis with no partitioning. Each Bayesian analysis included two independent runs, with four Markov chains per run. Each chain was run for 20 million generations, with a sample frequency of 1000 and a 25% burn-in period. Convergence diagnostics, including the standard deviation of split frequencies between runs, the potential scale reduction factor (PSRF) and the average effective sample size (ESS) were examined to confirm run convergence. FigTree ver. 1.4.2 (Rambaut, 2012) was used to annotate trees produced by MrBayes. In addition, to investigate phylogenetic relationships among true alleles, artefact sequences, LAEA and pseudoalleles Bayesian trees were created including all variants produced by AmpliSAS genotyping for both α1 and α2 datasets. The same parameters as described above were used for Bayesian analyses.

Results

Determining genotyping thresholds

Illumina MiSeq and pre-genotyping bioinformatics results are provided as supplementary material. AmpliCHECK analysis for α1 produced 697 variants (excluding those with a PAF lower than 1%; table 2). A total of 478 variants were considered 'probably artefacts' due to either 1-2bp mismatches, chimeras, or a combination of both (accumulated error), and 52 variants were classified as 'suspicious'. The remaining variants (n = 167) were considered 'good' and could not be explained by 1-2bp mismatches

and/or chimeras in any amplicons. AmpliCHECK analysis for α2 produced 582 variants (table 2), 188 of which were considered 'good' variants. 'Suspicious' variants numbered 57 and 337 variants were considered 'probably artefacts'. For both α1 and α2 several of the 'suspicious' variants were identified as potential artefacts within amplicons due to putative accumulated errors (chimeras with mismatches) and could represent natural recombination. Variants that were identified as 'probably artefacts' generally had low PAF, with an average PAF of 2.4% for α1 and 1.8% for α2 with 70% present in only one sample for α1 and 60% for α2.

Of the 21 family groups included in the MiSeq library, 16 complete family groups (both parents) were obtained for α1, and 20 were obtained for α2. Analysis of family group data revealed that for α1 the mean *lowest* PAF for a true allele within family groups was 9.3% (range 3.0% - 29.8%) and the mean *highest* PAF for an artefact sequence within families was 7.8% (range 3.5% - 31.6%). For α2 the mean *lowest* PAF for a true allele within family groups was 3.5% (range 3.0% - 4.5%) and the mean *highest* PAF for an artefact sequence within families was 6.8% (range 3.3% - 16.0%). Therefore, there was a PAF 'grey zone' for both α1 and α2 but was most severe for α2. Within family groups, all variants that were considered 'probably artefacts' by AmpliCHECK due to accumulated errors (chimeras and mismatches) were either only present in offspring (not parents) or only present in a parent at a low PAF. Hence it is likely that these variants are the results of accumulated errors and are not natural recombinants. LAEA were identified in both α1 (n = 11) and α2 (n= 14) datasets. The mean PAF of LAEA was 10.12% for α1 and 10.08% for α2. Some additional LAEA may exist within the dataset but were unidentified because they didn't occur in family groups. Additional artefact sequences that were not attributed to mismatches, chimeras or chimeras with mismatches (accumulated error) were identified during family group analysis (present in offspring but not parents). These putative artefacts could be true alleles that contaminated other samples during library preparation (e.g. pipetting error) or sequencing (tag-swapping) (Biedrzycka, et al., 2016; Schnell, et al., 2015).

Based on type I and II error rates across different PAF filtering thresholds (3% - 6%), a PAF filtering threshold of 5% was considered appropriate for α1 and a PAF filtering threshold of 4% was considered appropriate for α2 (fig. 2). Given a PAF of 5%, the type I error rate for α1 was 4% and the type II error rate was 5%. For α2 both type I and II error rates were higher, at 11% and 14%, given a PAF filtering threshold of 4%. Increasing or decreasing the PAF filtering threshold for α1 or α2 resulted in a large increase in either type I or type II error (fig. 2). For example, decreasing the PAF filtering threshold to 4% for α1 resulted in a type II error rate of 13% (compared to a 5% error rate at a 5% PAF filtering threshold). None of the nine technical replicates had discrepancies for α1 and five out of 11 replicates had discrepancies for α2.

Genotyping

Final AmpliSAS analysis (including filtering of artefact sequences, LAEA and pseudoalleles) resulted in 63 alleles for α1 and 95 alleles for α2, which translated into 61 and 75 amino acid sequences, respectively (table 2). A total of 443 samples (91.3%) were genotyped for α1 and 468 (96.5%) were retained for α2, including replicates. The most common α1 allele was present in 89 individuals and there were 11 rare alleles that were each present in a single individual. For α2, the most common allele was present in 294 individuals, with 21 rare alleles present in single individuals. For α1 the number of alleles per individual ranged from 1 to 4, and from 1 to 7 for α2, indicating that at least four and up to seven different MHC I loci exist in *C. decresii*. Individuals within each population had varying numbers of alleles. In regards to post-genotyping filtering, all variants that were identified as 'probably artefact' by AmpliCHECK due to accumulated errors (many of which were confirmed as artefacts in family group analysis) were removed from both α1 (n = 13) and α2 (n = 5) datasets. Alignment and translation of all alleles uncovered one pseudoallele in both the α1 and α2 dataset, which were subsequently removed. All LAEA identified for α1 (n = 11) and α2 (n= 14) were conservatively removed from the dataset.

No correlation was detected between the number of alleles per individual *after post-genotyping filtering* and read depth across individuals for both α1 ($F_{1, 452}$=0.013, p=0.9082) and α2 ($F_{1, 476}$=0.370, p=0.5433). However, there was a weak but significant correlation between read depth and the number of alleles per individual *before post-genotyping filtering* for α2 ($F_{1, 476}$=3.935, p=0.0479), but not α1 ($F_{1, 452}$=1.750, p=0.1865).

Segregation analysis

Examination of segregation patterns of alleles amongst hatchlings for α1 and α2 revealed that all family groups possessed loci at which a single allele was present (putative hemizygosity) in parents and/or offspring. Putative hemizygosity in parents was indicated by the presence of an allele in only some offspring and putative hemizygosity in offspring was indicated by differences in the number of loci between parents. In most family groups (α1 = 81%, α2 = 65%) parents had differing numbers of observed loci. The maximum number of loci suggested by family groups was four for α1 and six for α2. Examples of family group allele segregation patterns are provided in figure 3. Family groups for α1 did not show evidence of allele sharing among loci; certain alleles could be consistently placed at the same hypothetical loci for each family group. These allele groupings into hypothetical loci did not correspond to allele relationships displayed in the Bayesian gene tree (results not shown). For α2, we were unable to place alleles at the same hypothetical locus in each family group. However, this may be a result of the higher number of alleles and more complex patterns rather than allele sharing.

Genetic diversity and tests for selection

Percent identity among alleles, for both nucleotide and amino acid sequences, was similar among populations. Only Kangaroo Island displayed slightly lower identity among alleles and amino acids for α1 (table 3). The mean number of alleles per individual differed among populations and the total number of alleles per population also differed (table 4).

Recombination was detected for both α1 and α2 and was taken into account in tests for selection, which revealed a total of nine sites under positive selection for α1 and ten sites under positive selection for α2. Two sites in both α1 and α2 were identified as being under positive selection by MEME analysis only. A total of 13 conserved sites were identified for α1, and 10 for α2. Figure 5 displays this information along with amino acid variability (Wu-Kubat variability coefficient) and normalised dN-dS (ratio of non-synonymous to synonymous substitutions, from SLAC analysis), which were plotted for each amino acid position for α1 and α2. Six of the positively selected sites (PSS) for α1 (66%) and α2 (60%) had a conserved site (site under negative selection) within 2 amino acid positions. Three (33%) α1 PSS were homologous with PSS in other vertebrates and known human PBS. Half (n = 5) of the α2 PSS agreed with at least one other vertebrate PSS and three (30%) were homologous with known human PBS. High amino acid variability did not always correspond to a high dN-dS ratio for both domains.

Bayesian phylogenetic analysis of α1 and α2 alleles revealed a lack of geographical structure, with all populations represented in all parts of each tree (Fig. 4). Most alleles were shared among two or more populations. Transcripts described in Hacking, et al. (2018) were included in the Bayesian trees to determine with which transcripts alleles clustered. Alleles at the α1 region were similar to five of the transcripts, with no alleles associated with *Ctde-UA*005*, *Ctde-UA*006* or *Ctde-UA*008*. In the case of α2, all transcripts except *Ctde-UA*001* clustered with alleles. Bayesian trees including artefact sequences, LAEA and pseudoalleles for α1 and α2 showed that LAEA did not form clusters but were dispersed throughout the tree. Artefact sequences were also scattered throughout each tree and the pseudoalleles were not divergent but located amongst putatively functional alleles (Fig. S4).

Discussion

Our MHC genotyping protocol employed a recently developed suite of tools for next-generation amplicon sequencing (Sebastian, et al., 2016) and family group data to identify both type I and II genotyping errors. This allowed us to optimise genotyping parameters for a non-model reptile species and provides direction for MHC studies on other non-model organisms. Our study determined genotypes for over 400 tawny dragon lizards at the α1 and α2 domains of MHC I and confirmed

polymorphism for previously characterised MHC I transcripts (Hacking, et al., 2018), providing additional evidence for classical MHC I functionality for those loci. We then investigated MHC I polymorphism within and among populations. This revealed variation in the number of alleles per individual within populations, differences in the mean number of alleles per individual among populations, and variation in the total number of alleles among populations. The Kangaroo Island population had a lower number of MHC alleles compared to other populations in which a similar number of individuals were sampled. Bayesian phylogenetic analyses uncovered a high degree of allele sharing among populations and no geographic clustering. Tests for selection revealed evidence for putative PBS within α1 and α2, further supporting classical functionality predicted by Hacking, et al. (2018). Family group data also allowed us to explore allele segregation patterns, revealing a pattern expected under copy number variation (CNV) for both α1 and α2. However, we are unable to ascertain whether the pattern observed represents actual CNV, allele dropout due to the absence of primer annealing or low annealing efficiency, or a combination of both CNV and allele dropout. Below, we discuss these findings and suggest avenues for future research.

De novo genotyping of MHC I

We employed a recently developed clustering method (Sebastian, et al., 2016; Stutz and Bolnick, 2014) to genotype individuals at the MHC with the aid of family group data. Family group data is rarely used as part of MHC genotyping protocols despite its benefits, likely because complete family group data is difficult to obtain for many species, especially for wild populations. Studies such as Gaigher, et al. (2016) and Ferrandiz-Rovira, Bigot, Allaine, Callait-Cardinal, & Cohas (2015) have shown that family group data can be used to genotype individuals at the MHC, investigate linkage between MHC loci, and investigate CNV within the MHC. Using a combination of the AmpliSAT tools, AmpliCHECK and AmpliSAS, (Sebastian, et al., 2016) and family group data, we were able to identify artefact sequences and gain insight into the composition of artefact sequences. Furthermore, we were able to identify LAEA, allowing us to estimate type I and II error rates under different genotyping thresholds and find a balance between type I and II errors for final genotyping.

All of the artefact sequences that consisted of mismatches (based on the error rate provided) or chimeras were identified and removed during AmpliSAS analysis. However, variants with putative accumulated errors (chimeras with mismatches) identified by AmpliCHECK and verified as artefacts in family groups, were not removed during AmpliSAS analysis. These artefact sequences consisting of accumulated errors are not likely to be natural recombinants because the parent variants making up the chimera were always present within the same amplicon, as per AmpliCHECK methodology (Biedrzycka, et al., 2016; Sebastian, et al., 2016). Consequently, all such artefact sequences with accumulated errors were manually removed from final genotypes. Variants with accumulated errors

could conceivably occur if an artefact sequence was created early on during PCR when more time is available for further errors to occur. AmpliCHECK identifies these sequences as chimera with mismatches, however, some of these variants could also be a result of three-way chimeras, as was reported by Lenz and Becker (2008). Chimeras can be quite numerous within next-generation amplicon sequencing datasets (Biedrzycka, et al., 2016) and if left undetected can cause issues in downstream analyses (Lenz and Becker, 2008; Sommer, et al., 2013). We are confident that all artefact sequences resulting from chimeras were removed from our dataset.

Using family group data, we were able to identify variants that likely represented alleles that amplified at a low efficiency. These putative alleles were present in several family members but at a low frequency, and had an overall low frequency across all amplicons. The presence of LAEA is not unusual for next-generation MHC amplicon sequencing datasets; Stutz and Bolnick (2014) identified several such alleles and attributed them to a single locus with differences at primer annealing sites. However, unlike Stutz and Bolnick (2014), our LAEA did not cluster together in a neighbour-joining tree and hence are unlikely to belong to a single locus. Thus, it is possible that alleles across several loci differ sufficiently at primer annealing sites to cause lower rates of amplification compared to other alleles. The presence of LAEA raises the question of whether other alleles may have been dropped completely (allele dropout) (Sommer, et al., 2013). We tested for correlation between read depth and the number of alleles per individual and found a significant correlation for α2 only when LAEA were included. This suggests that biases within the dataset due to LAEA were alleviated with the removal of LAEA. However, a lack of correlation between read depth and the number of alleles per individual does not provide evidence for complete amplification of all alleles at loci if certain alleles are dropped in all amplicons.

We were able to obtain an acceptable genotyping error rate for α1 alleles, with the type I and II error estimated at 4% and 5%, respectively. This PAF filtering threshold is similar to PAF filtering thresholds employed by other studies (Gaigher, et al., 2016; Sebastian, et al., 2016). However, the genotyping error rate for α2 was unacceptably high, at 11% for type I and 14% for type II. We suggest that these error rate make the use of α2 genotypes unreliable for future work. These high error rates may be associated with cross-sample contamination due to a laboratory mistake or tag-swapping. However, if tag-swapping were the cause it would be expected that higher error rates would also be observed for α1 genotypes.

MHC I diversity and comparison with other vertebrates

Genotyping revealed a total of 63 alleles for α1 and 95 for α2. The number of alleles per individual suggests at least four and up to seven loci at the *C. decresii* MHC I. However, family group analysis

suggested six loci at most were amplified. This level of diversity is comparable to that reported for the MHC I of other non-avian reptiles (Elbers and Taylor, 2016; Pearson, Bradford, et al., 2016; Radwan, Kuduk, Levy, LeBas, & Babik, 2014), although comparison is difficult as different MHC I loci may be amplified among species. Comparison among populations revealed differences in the mean number of alleles per individual, which may represent differences in the number of loci among populations, or perhaps that different populations were differentially impacted by allelic dropout. There was also variation in the total number of alleles among populations, which may represent variation in the impact of drift (e.g. Kangaroo Island) or selective pressures such as parasite load, or again, allele dropout. Alleles were shared among populations and showed no signs of geographic clustering, although allele frequencies may differ among populations. Further work investigating the genetic structure of *C. decresii* populations using MHC markers is required.

Recombination sites were detected among both α1 and α2 alleles, providing further evidence that gene conversion has played a role in the evolution of the MHC I in *C. decresii*, as suggested by Hacking, et al. (2018). Several PSS were identified for α1 and α2 domains, some of which aligned with known human PBS and are likely involved in peptide binding in *C. decresii*. The position of PBS and PSS within the α1 and α2 domains is variable among jawed vertebrates (Kaufman, Volk, & Wallny, 1995; Pearson, Bull, et al., 2016). The PSS identified for *C. decresii* that were not homologous with PSS in other vertebrates or known human PBS may also be involved in peptide binding or may undertake other important functions such as interaction with T-cell receptors (Janeway, et al., 2005). We discovered that a third of *C. decresii* α1 PSS were the same as at least two other vertebrate groups (reptiles, birds, mammals, amphibians and fish), two of those included PSS described for iguanids. For α2, half of the *C. decresii* PSS were the same as at least one other vertebrate group, three of which included iguanid lizards. The PSS identified here allow future investigation into the functional clustering of alleles, providing insight into the functional diversity present at MHC I in *C. decresii*. Conserved sites (negative selection) were often located near PSS and some likely undertake important regulatory functions (Kaufman, Salomonsen, & Flajnik, 1994). Finally, the fact that two sites were identified by only the MEME method suggests that episodic selection plays a dominant role at these sites, rather than positive directional selection (Grueber, Wallis, & Jamieson, 2014; Murrell, et al., 2012).

The number of MHC I alleles among *C. decresii* individuals varied within populations and allele segregation analysis revealed patterns that would be expected under CNV. These results could be interpreted as CNV. However, at this stage we are unable to exclude the possibility of extensive allele dropout. The existence of LAEA across multiple loci lends evidence towards allele dropout rather than CNV explaining the patterns observed. However, it is possible that a combination of allele dropout and CNV, or CNV alone are producing these patterns. Thus, it is impossible to make inferences about CNV

based on our data because single exons were amplified (potentially underestimate CNV) and there was a high likelihood of allele dropout (potentially overestimate CNV). Genomic mapping and the analysis of complete haplotypes across several individuals would provide insight into the extent of MHC CNV in *C. decresii* (Llaurens, McMullan, & van Oosterhout, 2012).

Conclusion and future directions

We successfully genotyped *C. decresii* individuals across five populations at the α1 exon of the MHC I region. However, high error rates limited our ability to resolve genotypes for the α2 exon. Our α1 data likely represent a subsample of the diversity present at the MHC I of *C. decresii* as some loci were not encompassed by the primers and it is likely that allele dropout occurred across loci. We consider the α1 genotypes we have presented here useful for future research into the factors shaping diversity at the MHC in *C. decresii* as long as the restrictions of the data are acknowledged. We successfully identified positively selected sites and putative peptide binding sites, and described MHC I diversity within and among populations, providing a base for future work on the MHC in *C. decresii*. Complete haplotype data, perhaps using long read sequencing, primer re-design and improvements to allele amplification protocols (Marmesat, Soriano, Mazzoni, Sommer, & Godoy, 2016), or transcriptome-based exon capture (Bi et al., 2012) are required to gain a more thorough survey of the diversity at the *C. decresii* MHC I region. Our study demonstrates that both type I and type II genotyping error rates can be determined when genotyping is supplemented by the analysis of family group data, allowing informed decisions to be made when setting genotyping thresholds. Future directions include investigation into the existence of CNV at the *C. decresii* MHC class I region and identification of the mechanisms maintaining MHC class I diversity in *C. decresii*.

Acknowledgements

All methods were approved by the Flinders University Animal Welfare Committee (E379) and the Department of Environment, Water and Natural Resources (U26225). Fieldwork at the Yourambulla Caves Aboriginal Heritage site in Hawker was approved by the Adnyamathanha Traditional Lands Association. We thank volunteers and colleagues that helped with fieldwork. We also thank Terry Bertozzi for helpful discussions. This work was funded by the Holsworth Wildlife Research endowment, Nature Foundation of South Australia and American Society for the Study of Amphibians and Reptiles, awarded to JH.

References

Babik, W., Taberlet, P., Ejsmond, M., & Radwan, J. (2009). New generation sequencers as a tool for genotyping of highly polymorphic multilocus MHC system. *Molecular Ecology Resources, 9*, pp. 713-719.

Bi, K., Vanderpool, D., Singhal, S., Linderoth, T., Moritz, C., & Good, J. M. (2012). Transcriptome-based exon capture enables highly cost-effective comparative genomic data collection at moderate evolutionary scales. [journal article]. *BMC Genomics, 13*(1), p 403. doi:10.1186/1471-2164-13-403 Retrieved from https://doi.org/10.1186/1471-2164-13-403

Biedrzycka, A., Sebastian, A., Migalska, M., Westerdahl, H., & Radwan, J. (2016). Testing genotyping strategies for ultra-deep sequencing of a co-amplifying gene family: MHC class I in a passerine bird. *Molecular Ecology Resources*doi:10.1111/1755-0998.12612

Bontrop, R. E., Otting, N., de Groot, N. G., & Doxiadis, G. G. M. (1999). Major histocompatibility complex class II polymorphisms in primates. *Immunological Reviews, 167*(1), pp. 339-350. doi:10.1111/j.1600-065X.1999.tb01403.x Retrieved from http://dx.doi.org/10.1111/j.1600-065X.1999.tb01403.x

De Groot, N., Blokhuis, J. H., Otting, N., Doxiadis, G. G., & Bontrop, R. E. (2015). Co-evolution of the MHC class I and KIR gene families in rhesus macaques: ancestry and plasticity. *Immunological Reviews, 267*(2015), pp. 228–245.

Delport, W., Poon, A. F., Frost, S. D., & Kosakovsky Pond, S. L. (2010). Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics, 26*(19), pp. 2455-2457. doi:10.1093/bioinformatics/btq429

Dirscherl, H., McConnell, S., Yoder, J., & Jong, J. (2014). The MHC class I genes of zebrafish. *Developmental & Comparative Immunology, 46*(1), pp. 11-23.

Elbers, J., & Taylor, S. (2016). Major histocompatibility complex polymorphism in reptile conservation. *Herpetological Conservation and Biology, 11*(1), pp. 1-12.

Ferrandiz-Rovira, M., Bigot, T., Allaine, D., Callait-Cardinal, M. P., & Cohas, A. (2015). Large-scale genotyping of highly polymorphic loci by next-generation sequencing: how to overcome the

challenges to reliably genotype individuals. [Original Article]. *Heredity, 114*(5), pp. 485-493. doi:10.1038/hdy.2015.13 Retrieved from http://dx.doi.org/10.1038/hdy.2015.13

Gaigher, A., Burri, R., Gharib, W., Taberlet, P., Roulin, A., & Fumagalli, L. (2016). Family-assisted inference of the genetic architecture of major histocompatibility complex variation. *Molecular Ecology Resources*

Galan, M., Guivier, E., Caraux, G., Charbonnel, N., & Cosson, J. (2010). A 454 multiplex sequencing method for rapid and reliable genotyping of highly polymorphic genes in large-scale studies. *BMC Genomics, 11*(296)

Garcia-Boronat, M., Diez-Rivero, C. M., Reinherz, E. L., & Reche, P. A. (2008). PVS: a web server for protein sequence variability analysis tuned to facilitate conserved epitope discovery. *Nucleic Acids Research, 36*(Web Server issue), pp. W35-41. doi:10.1093/nar/gkn211

Glaberman, S., & Caccone, A. (2008). Species-specific evolution of class I MHC genes in iguanas (order: Squamata; subfamily: Iguaninae). *Immunogenetics, 60*(7), pp. 371-382. doi:10.1007/s00251-008-0298-y Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/18488213

Glaberman, S., Du Pasquier, L., & Caccone, A. (2009). Characterization of a Nonclassical Class I MHC Gene in a Reptile, the Galapagos Marine Iguana (*Amblyrhynchus cristatus*). *PLOS ONE, 3*(8), p e2859. doi:10.1371/journal.pone.0002859.t001

Grimholt, U., Tsukamoto, K., Azuma, T., Leong, J., Koop, B. F., & Dijkstra, J. M. (2015). A comprehensive analysis of teleost MHC class I sequences. *BMC evolutionary biology, 15*, p 32. doi:10.1186/s12862-015-0309-1

Grueber, C. E., Wallis, G. P., & Jamieson, I. G. (2014). Episodic Positive Selection in the Evolution of Avian Toll-Like Receptor Innate Immunity Genes. *PLOS ONE, 9*(3), p e89632. doi:10.1371/journal.pone.0089632 Retrieved from https://doi.org/10.1371/journal.pone.0089632

Guindon, S., & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol, 52*doi:10.1080/10635150390235520 Retrieved from http://dx.doi.org/10.1080/10635150390235520

Hacking, J., Bertozzi, T., Moussalli, A., Bradford, T., & Gardner, M. (2018). Characterisation of major histocompatibility complex class I transcripts in an Australian dragon lizard. *Developmental & Comparative Immunology, 84*, pp. 164-171. doi:https://doi.org/10.1016/j.dci.2018.02.012 Retrieved from https://www.sciencedirect.com/science/article/pii/S0145305X17306729

Hacking, J. D., Stuart-Fox, D. M., & Gardner, M. G. (In Press). Very low rate of multiple paternity detected in clutches of a wild agamid lizard. *Australian Journal of Zoology*

Herdegen, M., Babik, W., & Radwan, J. (2014). Selective pressures on MHC class II genes in the guppy (*Poecilia reticulata*) as inferred by hierarchical analysis of population structure. *Journal of Evolutionary Biology, 27*(11), pp. 2347-2359. doi:10.1111/jeb.12476 Retrieved from http://dx.doi.org/10.1111/jeb.12476

Janeway, C. A., Travers, P., Walport, M., & Shlomchik, M. J. (2005). *Immunobiology: the immune system in health and disease* London: Garland Sience.

Kaufman, J. (2015). What chickens would tell you about the evolution of antigen processing and presentation. *Current Opinion in Immunology, 34*, pp. 35-42. doi:10.1016/j.coi.2015.01.001

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

Kaufman, J., Volk, H., & Wallny, H. J. (1995). A "minimal essential Mhc" and an "unrecognized Mhc": two extremes in selection for polymorphism. *Immunological Reviews, 143*doi:10.1111/j.1600-065X.1995.tb00670.x Retrieved from http://dx.doi.org/10.1111/j.1600-065X.1995.tb00670.x

Kiemnec-Tyburczy, K., Richmond, J., Savage, A., Lips, K., & Zamudio, K. (2012). Genetic diversity of MHC class I loci in six non-model frogs is shaped by positive selection and gene duplication. *Heredity, 109*(3), pp. 146-155. doi:10.1038/hdy.2012.22 Retrieved from http://dx.doi.org/10.1038/hdy.2012.22

Kosakovsky Pond, S. L., & Frost, S. D. (2005). Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Molecular Biology and Evolution, 22*(5), pp. 1208-1222. doi:10.1093/molbev/msi105

Kosakovsky Pond, S. L., Posada, D., Gravenor, M. B., Woelk, C. H., & Frost, S. D. W. (2006). GARD: a genetic algorithm for recombination detection. *Bioinformatics, 22*(24), pp. 3096-3098. doi:10.1093/bioinformatics/btl474 Retrieved from http://dx.doi.org/10.1093/bioinformatics/btl474

Lanfear, R., Calcott, B., Ho, S. Y., & Guindon, S. (2012). Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution, 29*(6), pp. 1695-1701. doi:10.1093/molbev/mss020

Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., & Calcott, B. (2017). Partitionfinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution, 34*(3), pp. 772-773. doi:10.1093/molbev/msw260 Retrieved from http://dx.doi.org/10.1093/molbev/msw260

Lenz, T., & Becker, S. (2008). Simple approach to reduce PCR artefact formation leads to reliable genotyping of MHC and other highly polymorphic loci--implications for evolutionary analysis. *Gene, 427*(1-2), pp. 117-123. doi:10.1016/j.gene.2008.09.013 Retrieved from http://dx.doi.org/10.1016/j.gene.2008.09.013

Lighten, J., Oosterhout, C., & Bentzen, P. (2014). Critical review of NGS analyses for *de novo* genotyping multigene families. *Molecular Ecology, 2014*(23), pp. 3957–3972.

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., . . . Law, M. (2012). Comparison of next-generation sequencing systems. *Journal of biomedicine & biotechnology, 2012*, p 251364. doi:10.1155/2012/251364 Retrieved from http://dx.doi.org/10.1155/2012/251364

Llaurens, V., McMullan, M., & van Oosterhout, C. (2012). Cryptic MHC polymorphism revealed but not explained by selection on the class iib peptide-binding region. *Molecular Biology and Evolution, 29*(6), pp. 1631-1644. doi:10.1093/molbev/mss012 Retrieved from http://dx.doi.org/10.1093/molbev/mss012

Loman, N., Misra, R., Dallman, T., Constantinidou, C., Gharbia, S., Wain, J., & Pallen, M. (2012). Performance comparison of benchtop high-throughput sequencing platforms. *Nature Biotechnology, 30*(2012), pp. 434-439.

Marmesat, E., Soriano, L., Mazzoni, C. J., Sommer, S., & Godoy, J. A. (2016). PCR strategies for complete allele calling in multigene families using high-throughput sequencing approaches. *PLOS ONE, 11*(6), p e0157402. doi:10.1371/journal.pone.0157402 Retrieved from https://doi.org/10.1371/journal.pone.0157402

McLean, C. A., Moussalli, A., & Stuart-Fox, D. (2013). Taxonomic Assessment of the *Ctenophorus decresii* Complex (Reptilia: Agamidae) Reveals a New Species of Dragon Lizard from Western New South Wales. *Records of the Australian Museum, 65*(3), pp. 51–63.

McLean, C. A., Stuart-Fox, D., & Moussalli, A. (2014). Phylogeographic structure, demographic history and morph composition in a colour polymorphic lizard. *Journal of Evolutionary Biology, 27*(10), pp. 2123-2137.

Miller, H. C., Andrews-Cookson, M., & Daugherty, C. H. (2007). Two patterns of variation among MHC class I loci in Tuatara (*Sphenodon punctatus*). *Journal of Heredity, 98*(7), pp. 666-677. doi:10.1093/jhered/esm095 Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/18032462

Murrell, B., Wertheim, J. O., Moola, S., Weighill, T., Scheffler, K., & Pond, S. K. (2012). Detecting individual sites subject to episodic diversifying selection. *PLOS Genetics, 8*doi:10.1371/journal.pgen.1002764 Retrieved from http://dx.doi.org/10.1371/journal.pgen.1002764

Ohta, Y., Goetz, W., Hossain, M. Z., Nonaka, M., & Flajnik, M. F. (2006). Ancestral organization of the MHC revealed in the amphibian Xenopus. *Journal of Immunology, 176*doi:10.4049/jimmunol.176.6.3674 Retrieved from http://dx.doi.org/10.4049/jimmunol.176.6.3674

Pearson, S., Bradford, T., Ansari, T., Bull, M., & Gardner, M. (2016). MHC genotyping from next-generation sequencing: detailed methodology for the gidgee skink, *Egernia stokesii*. *Transactions of the Royal Society of South Australia, 140*(2), pp. 244-262.

Pearson, S. K., Bull, C., & Gardner, M. (2016). *Egernia stokesii* (gidgee skink) MHC I positively selected sites lack concordance with HLA peptide binding regions. *Immunogenetics, 69*(1), pp. 49-61.

Piertney, S., & Oliver, M. (2006). The evolutionary ecology of the major histocompatibility complex. *Heredity, 96*(1), pp. 7-21. Retrieved from http://dx.doi.org/10.1038/sj.hdy.6800724

R Core Team. (2016). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Radwan, J., Kuduk, K., Levy, E., LeBas, N., & Babik, W. (2014). Parasite load and MHC diversity in undisturbed and agriculturally modified habitats of the ornate dragon lizard. *Molecular Ecology, 23*doi:10.1111/mec.12984 Retrieved from https://doi.org/10.1111/mec.12984

Rambaut, A. (2012). FigTree (Version 1.4.2). Institute of Evolutionary Biology, University of Edinburgh. Retrieved from http://tree.bio.ed.ac.uk/

Reche, P. A., & Reinherz, E. L. (2003). Sequence variability analysis of human class I and class II MHC molecules: functional and structural correlates of amino acid polymorphisms. *Journal of Molecular Biology, 331*(3), pp. 623-641.

Ronquist, F., Teslenko, P., Ayres, D., Darling, S., Hohna, S., Larget, B., . . . Suchard, M. H., JP. (2011). MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology, 6*(3), pp. 539-542.

Ross, M. G., Russ, C., Costello, M., Hollinger, A., Lennon, N. J., Hegarty, R., . . . Jaffe, D. B. (2013). Characterizing and measuring bias in sequence data. [journal article]. *Genome Biology, 14*(5), pp. 1-20. doi:10.1186/gb-2013-14-5-r51 Retrieved from http://dx.doi.org/10.1186/gb-2013-14-5-r51

Schnell, I., Bohmann, K., & Gilbert, M. (2015). Tag jumps illuminated – reducing sequence-to-sample misidentifications in metabarcoding studies. *Molecular Ecology Resources, 15*(6), pp. 1289-1303.

Schrider, D., & Hahn, M. (2010). Gene copy-number polymorphism in nature. *Proceedings of the Royal Society B: Biological Sciences, 2010*(277), pp. 3213-3221.

Sebastian, A., Herdegen, M., Migalska, M., & Radwan, J. (2016). AMPLISAS: a web server for multilocus genotyping using next-generation amplicon sequencing data. *Molecular Ecology, 16*(2), pp. 498–510.

Smith, L., & Burgoyne, L. (2004). Collecting, archiving and processing DNA from wildlife samples using FTA® databasing paper. [journal article]. *BMC Ecology, 4*(1), p 4. doi:10.1186/1472-6785-4-4 Retrieved from https://doi.org/10.1186/1472-6785-4-4

Sommer, S., Courtiol, A., & Mazzoni, C. (2013). MHC genotyping of non-model organisms using next-generation sequencing: a new methodology to deal with artefacts and allelic dropout. *BMC Genomics, 14*, p 542. doi:10.1186/1471-2164-14-542 Retrieved from http://dx.doi.org/10.1186/1471-2164-14-542

Strandh, M., Lannefors, M., Bonadonna, F., & Westerdahl, H. (2011). Characterization of MHC class I and II genes in a subantarctic seabird, the blue petrel, Halobaena caerulea (Procellariiformes). *Immunogenetics, 63*(10), pp. 653-666. doi:10.1007/s00251-011-0534-8 Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/21607694

Stutz, W., & Bolnick, D. (2014). Stepwise threshold clustering: A new method for genotyping MHC loci using next-generation sequencing technology. *PLOS ONE, 9*(7), p e100587.

Wang, D., Zhong, L., Wei, Q., Gan, X., & He, S. (2010). Evolution of MHC class I genes in two ancient fish, paddlefish (*Polyodon spathula*) and Chinese sturgeon (*Acipenser sinensis*). *FEBS letters, 584*(15), pp. 3331-3339. doi:10.1016/j.febslet.2010.05.065 Retrieved from http://dx.doi.org/10.1016/j.febslet.2010.05.065

Wang, X. X., & He, X. X. (2006). Summary of the evolution studies on primate major histocompatibility complex class I. *Yi Chuan, 28*(5), pp. 611-616.

Zagalska-Neubauer, M., Babik, W., Stuglik, M., Gustafsson, L., Cichon, M., & Radwan, J. (2010). 454 sequencing reveals extreme complexity of the class II Major Histocompatibility Complex in the collared flycatcher. *BMC Evolutionary Biology, 10*, p 395. doi:10.1186/1471-2148-10-395 Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/21194449

Tables and Figures

Table 1. Issues associated with genotyping multi-gene families using next-generation amplicon sequencing.

Issue	Causes	Consequences
Issue 1: Uneven amplification of true alleles	Uneven sampling from the library pool or differences in primer annealing efficiencies.	Type I error. Low read depth or complete dropout for some true alleles.
Issue 2: High read depth for artefact sequences	Errors can be amplified or repeated during PCR and repeated during sequencing so that artefact sequences can attain high coverage.	Type II error. Artefact sequences may be mistaken for true alleles due to high read depth.
Issue 3: Highly similar true alleles	A true allele may share more similarities with another true allele than with an artefact sequence originating from that true allele due to multiple errors.	Type I and II error. A true allele may be assumed to be an artefact due to its similarity to another true allele or an artefact sequence may be assumed to be a true allele due to its relative dissimilarity to true alleles.
Issue 4: True allele in only some amplicons	High sequence similarity among true alleles means that an artefact sequence could, by chance, become identical to a true allele in a different amplicon. Cross-sample contamination and tag-swapping can also cause a true allele to be erroneously included in an amplicon in which it does not occur.	Type II error. An artefact may be assumed to be a true allele in one amplicon because it is a true allele in other amplicons.

References: Biedrzycka, et al. (2016), Lighten, et al. (2014), Sebastian, et al. (2016), Sommer, et al. (2013), Stutz and Bolnick (2014)

Table 2. Summary of Illumina MiSeq run and genotyping results for *Ctenophorus decresii* MHC class I data. The average number of reads per amplicon attributed to true alleles is shown in brackets.

	Post assembly*		AmpliCHECK^		AmpliSAS€	
	α1	*α2*	*α1*	*α2*	*α1*	*α2*
Number of reads	1 245 773	1 903 037	681 819	875 906	735 472 (1660)	938 559 (1963)
Number of variants/alleles	-	-	697	582	63	98
Number of amino acid sequences	-	-	-	-	61	78

*Reads ≥50bp and ≥20 quality score; ^Variants of correct length and ≥1% PAF; €Alleles; variants that passed AmpliSAS clustering and filtering, and post-genotyping filtering.

Table 3. Percent nucleotide and amino acid identity for all *Ctenophorus decresii* MHC class I alleles across all populations and per population.

	Mean percent nucleotide identity		Mean percent amino acid identity	
	α1	*α2*	*α1*	*α2*
All populations	86.9	83.8	80.5	73.4
Hawker	86.2	84.0	79.7	73.7
Mount Remarkable	87.7	84.0	81.9	73.4
Barossa	86.1	83.1	79.1	73.3
Morialta	86.5	83.9	79.8	73.6
Kangaroo Island	82.6	82.8	76.0	72.6

Table 4. Number of alleles per population and average number of alleles per individual. Number of individuals included is given in brackets.

	Total number of alleles		Average number of alleles per individual	
	α1	*α2*	*α1*	*α2*
All populations	63 (443)	95 (468)	1.8	3.8
Hawker*	28 (300)	40 (322)	1.8	3.7
Mount Remarkable	24 (40)	41 (40)	1.8	3.9
Barossa	12 (26)	24 (26)	1.6	4.2
Morialta	20 (36)	34 (39)	1.7	4.0
Kangaroo Island	5 (41)	6 (41)	2.2	3.5

*Includes family groups

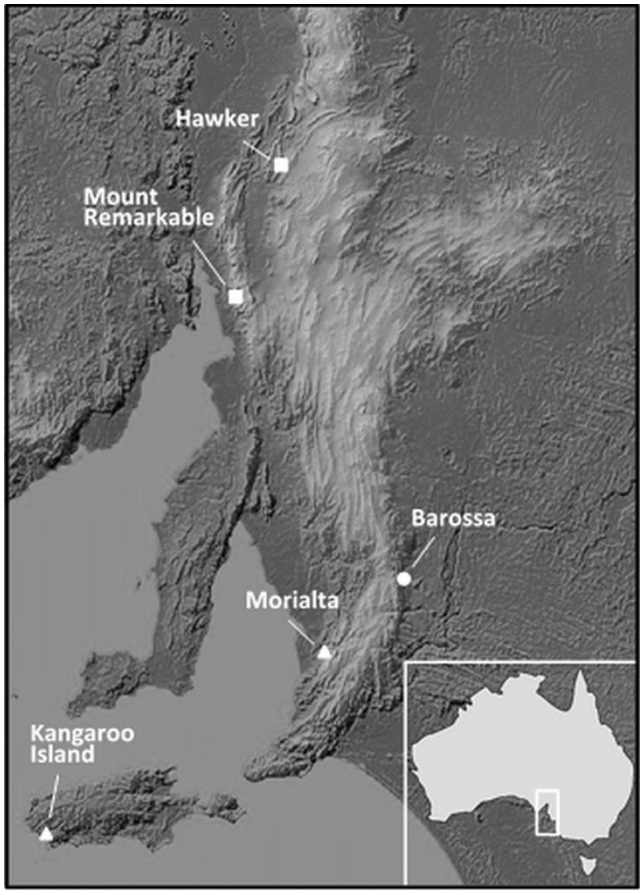


Fig. 1 Map of *Ctenophorus decresii* populations sampled. Squares indicate the northern lineage, triangles indicate the southern lineage and Barossa (circle) lies within the contact zone, as per McLean, et al. (2014)

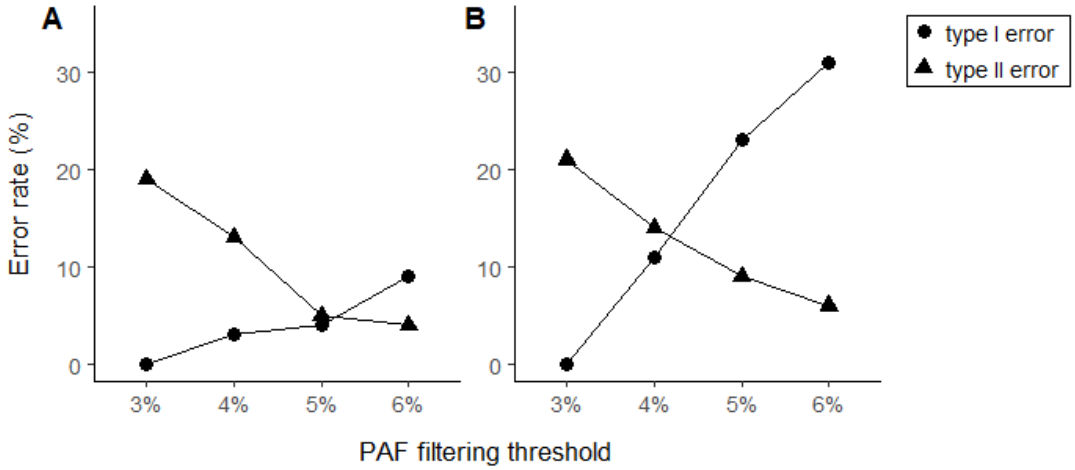


Fig. 2. Visual representation of differences in type I (incorrect rejection of true allele) and type II (incorrect acceptance of artefact sequence) genotyping error rates under different per amplicon frequency (PAF) filtering thresholds used during preliminary AmpliSAS genotyping of the MHC in *C. decresii* for α1 (A) and α2 (B) domains.

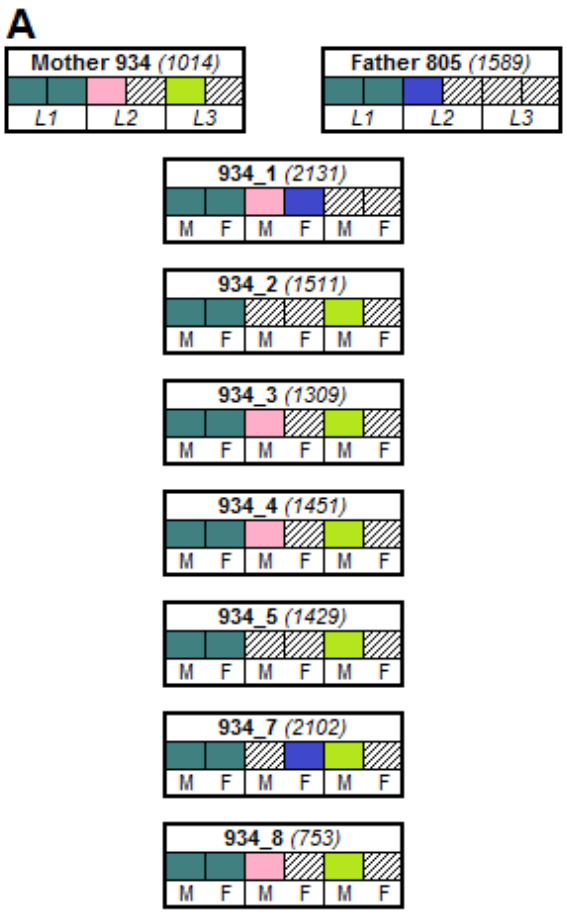

A
Mother 934 (1014)
L1
L2
L3
Father 805 (1589)
L1
L2
L3
934_1 (2131)
M F M F M F
934_2 (1511)
M F M F M F
934_3 (1309)
M F M F M F
934_4 (1451)
M F M F M F
934_5 (1429)
M F M F M F
934_7 (2102)
M F M F M F
934_8 (753)
M F M F M F
B
Mother 990 (2285)
L1
L2
L3
Father 655 (3883)
L1
L2
L3
990_1 (1629)
M F M F M F
990_2 (1424)
M F M F M F
990_4 (1685)
M F M F M F
990_5 (735)
M F M F M F
990_6 (999)
M F M F M F
990_7 (675)
M F M F M F

Fig. 3 Examples of segregation patterns observed for *Ctenophorus decresii* MHC I α1 (A) and α2 (B). Hatchling genotypes are displayed underneath the mother and father genotypes. Locus numbers are indicted for the mother and father. For offspring genotypes, 'M' indicates alleles originating from the mother and 'F' indicates alleles originating from the father. Each colour represents a different allele. The absence of an allele is indicated with slanted lines. The read depth (number of reads) is provided beside each individual identification number in brackets

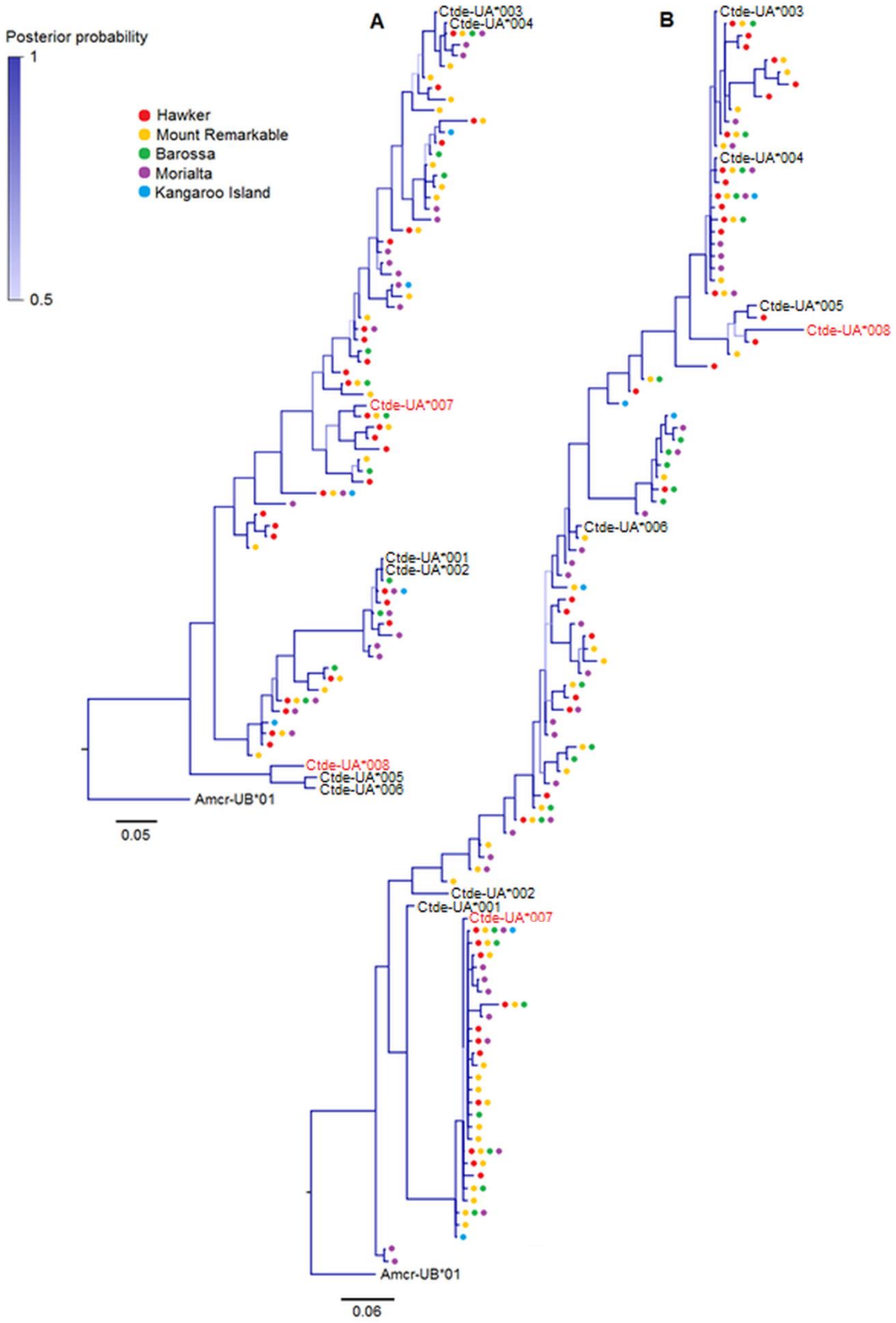

Posterior probability
1
0.5
Hawker
Mount Remarkable
Barossa
Morialta
Kangaroo Island
A
Ctde-UA*003
Ctde-UA*004
Ctde-UA*007
Ctde-UA*001
Ctde-UA*002
Ctde-UA*008
Ctde-UA*005
Ctde-UA*006
Amcr-UB*01
0.05
B
Ctde-UA*003
Ctde-UA*004
Ctde-UA*005
Ctde-UA*008
Ctde-UA*006
Ctde-UA*002
Ctde-UA*001
Ctde-UA*007
Amcr-UB*01
0.06

Fig. 4. Bayesian trees of *Ctenophorus decresii* α1 (A) and α2 (B) alleles, including the α1 and α2 domains (respectively) of eight *C. decresii* transcripts described in Hacking, Bertozzi, Moussalli, Bradford, & Gardner (2018) (*Ctde-UA*001 – Ctde-UA*008*). Transcripts in red text (*Ctde-UA*007* and *Ctde-UA*008*) are truncated, likely due to incomplete transcriptome assembly. These transcripts are missing parts of the α3 domain and the Tm/Cyt domains. Coloured circles represent populations in which each allele is present (refer to the embedded legend). Branch colour corresponds to posterior probability (refer to posterior probability scale). The scale bar indicates the expected changes per site. Each tree is rooted using the α1 (A) or α2 (B) domain of iguanid MHC1 sequence (*Amblyrhynchus cristatus*; Amcr-UB*01, Glaberman, Du Pasquier, & Caccone (2009)).

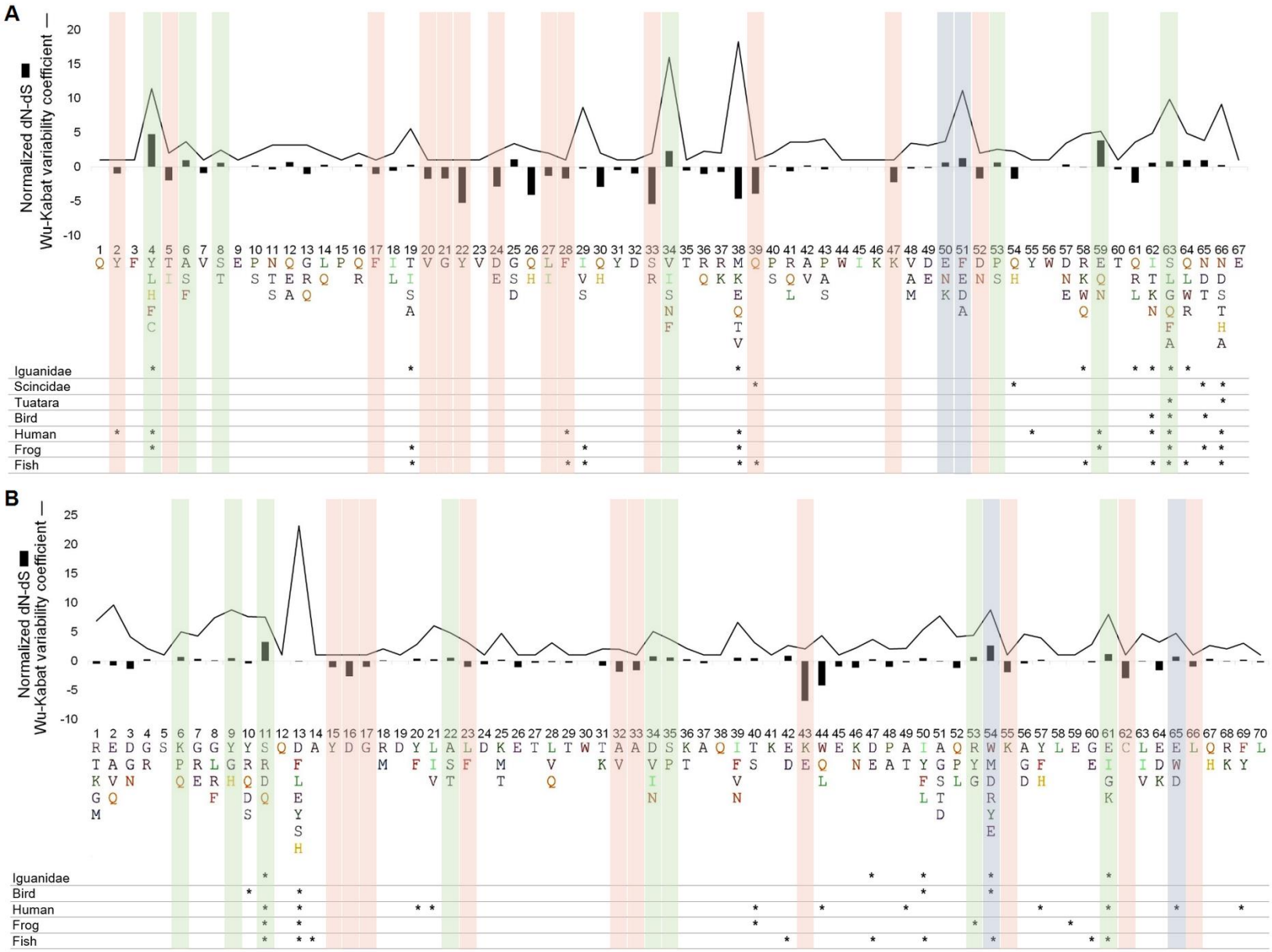

A
Normalized dN-dS
Wu-Kabat variability coefficient
Iguanidae
Scincidae
Tuatara
Bird
Human
Frog
Fish
B
Normalized dN-dS
Wu-Kabat variability coefficient
Iguanidae
Bird
Human
Frog
Fish

Fig. 5. Amino acid variability and sites under selection for α1 (A) and α2 (B) domains of *Ctenophorus decresii* class I MHC alleles. Amino acid variation (Wu-Kabat coefficient) is displayed as a line graph and the dN-dS ratio (according to FEL analysis) is displayed as a bar graph. Site numbers correspond to *C. decresii* α1 and α2 amplicons and every amino acid variant is displayed for each site. Green shaded sites indicate sites under positive selection. Blue shaded sites indicate sites under positive selection according to MEME analysis only. Red shaded sites indicate negatively selected sites. Positively selected sites in other vertebrates and known peptide binding sites in humans are indicated with an asterisks. Iguanid *spp.* (Glaberman and Caccone, 2008), Scincidae; *Egernia stokesii* (Pearson, Bull, & Gardner, 2016), tuatara; *Sphenodon punctatus* (Miller, Andrews-Cookson, & Daugherty, 2007), bird; *Halobaena caerulea* (Strandh, Lannefors, Bonadonna, & Westerdahl, 2011), human; *Homo sapiens* (Reche and Reinherz, 2003), frog *spp.* (Kiemnec-Tyburczy, Richmond, Savage, Lips, & Zamudio, 2012) and fish; *Acipenser sinensis* (Wang, Zhong, Wei, Gan, & He, 2010).

Supplementary Information

De novo genotyping of the MHC in an Australian dragon lizard aided by family group data

Hacking, J., Bradford, T., Pierce, K., Gardner, M.

Illumina MiSeq library preparation methods

A two-step PCR protocol was used to prepare a MiSeq sequencing library following Pearson, Bradford, Ansari, Bull, & Gardner (2016). MHC-targeting primers were used in the first PCR, followed by a second PCR to add Illumina sequencing adapters and sample indexes. MHC-targeting primers were designed for α1 and α2 based on *C. decresii* transcriptome data (table S2, Hacking, Bertozzi, Moussalli, Bradford, & Gardner, 2017) using Primer 3 ver. 2.3.4 (Untergasser et al., 2012) in Geneious ver. 8.1.7 (Kearse et al., 2012). The MHC-targeting primers and a MRT (multiplex ready technology, Hayden, Nguyen, Waterman, & Chalmers, 2008) linker sequence constitute the inner primers. The outer primers consisted of indexes unique to each individual and Illumina-specific adapters with an MRT linker sequence. The inner and outer primers are connected via the MRT primer sequence. A summary of the whole library preparation process, primer design details and all inner and outer primer sequences are provided (figure S1, table S1).

MHC-targeting primer pairs MHC1_A1_G234_F/MHC1_A1_G234_R and MHC1_A2_G234_F/MHC1_A2_G12_R (Hacking, et al., 2017) were used to amplify parts of the MHC I α1 (exon 2) and α2 (exon 3) domains in DNA samples for 337 individuals from Hawker, 40 individuals from Mount Remarkable, 27 individuals from the Barossa, 40 individuals from Morialta, and 41 individuals from Kangaroo Island (total of 485). Technical replicates (n = 12) amplified in independent PCR reactions were included in the MiSeq library. A small number of replicates were used as family group data were also available to investigate the accuracy of MHC genotypes. Amplification reactions for the first PCR totalled 12µl and contained 4mM each dNTP, 0.25 mg/ml BSA, 0.4µM forward and reverse primers, 0.5U Immolase enzyme and approximately 20ng of template DNA. Thermal cycling conditions consisted of an initial denaturisation step of 95°C for 10 minutes, 30 cycles of 94°C for 45 seconds, 57°C for 45 seconds and 72°C for 60 seconds, followed by a final extension of 72°C for 6 minutes. After PCR clean-up using MultiScreenHTS 384-well filter plates on a vacuum manifold (Merck Millipore), a second PCR was undertaken to attach the outer primers. Amplification reactions were 12µl and contained 4mM dNTP, 0.25 mg/ml BSA 0.4µM forward and reverse outer primer, 0.5U Immolase enzyme and approximately 20ng of template DNA. Thermal cycling conditions for the second PCR consisted of an initial denaturisation step of 95°C for 10 minutes, 10 cycles of 92°C for 15 seconds, 54°C

for 60 seconds and 72°C for 1 minute and 30 seconds, followed by a final extension of 72°C for 10 minutes.

Gel electrophoresis (1.5% agarose) was performed for every amplicon following the second PCR and the resulting bands were scored for band strength (absent, weak, moderate, and strong). TapeStation analysis (Agilent 2200 TapeStation ©; Agilent Technologies) was undertaken on a subset of amplicons from each gel score category to determine the concentration, molarity and the amount of contaminating adapter dimer present at 125bp. Nine individuals did not have visible amplicons on the gel and were excluded from downstream analyses. Amplicons were combined into a weak pool, and a moderate/strong pool based on gel band score and TapeStation analysis. The weak pool contained a moderate amount of adapter dimer that was successfully removed using an Ampure size-selection clean-up (Agencourt AMPure beads, Beckman Coulter, Inc., Li, Hofreiter, Straube, Corrigan, & Naylor, 2013) prior to sequencing. TapeStation analysis was undertaken on each cleaned pool to determine concentration and molarity. A portion of each of the cleaned pools was combined to make a final pool, with the volume added based on the number of amplicons within each pool, after equalising molarity. The completed MiSeq library was sent to the Australian Genome Research Facility (AGRF) for 300bp paired-end sequencing on the Illumina MiSeq platform with 10% PhiX (a viral genome) added to create additional complexity within the library for more efficient base detection.

Pre-genotyping bioinformatics methods

Several bioinformatics steps were undertaken to prepare data for genotyping (figure S2). MiSeq paired-end data was received from the AGRF de-multiplexed by sample index, with two separate files per sample (read 1; R1 and read 2; R2). The overall quality of the data was determined by performing FastQC ver. 0.11.2 (Andrews, 2010) analyses on all R1 and all R2 files (concatenated across individuals). Paired-end reads for each individual were assembled using PEAR ver. 0.9.5 (Zhang, Kobert, Flouri, & Stamatakis, 2014) with a 50bp minimum overlap of reads and a Q-value cutoff of 20. Improvement in the quality of the reads (henceforth, assembled paired-end reads are referred to as 'reads') was assessed by FastQC analysis of 15 randomly selected samples. Target region sequence lengths were investigated for a subset of samples (n=15) in Geneious ver. 8.1.7 (Kearse, et al., 2012) using jMHC ver. 1.6.1624 (Stuglik, Radwan, & Babik, 2011) output, in which all identical reads are collapsed for each individual for a given locus-specific primer pair. Pseudogenes were identified by the presence of premature stop codons. Only lengths that were not pseudogenes or a subset of a longer length were included in downstream analyses. At the time of analysis, the programs used for genotyping (AmpliCHECK and AmpliSAS) required reads to be indexed and concatenated into a single file.

Therefore, a python script (https://github.com/kellyp2738/genotyping_methods_hacking_etal) was written to re-index the reads, which were then concatenated into a single compressed file. From this file, target lengths were extracted using Geneious to create smaller, more manageable files suitable for AmpliCHECK and AmpliSAS analysis.

Illumina MiSeq and pre-genotyping bioinformatics results

The complete MiSeq run produced 6.5 million paired-end reads, with 3.1 million of those assigned to the amplicons used in this study (one α1 primer pair and one α2 primer pair; table 2). Initial FastQC analysis of raw, un-paired reads indicated high per-base and per-sequence Phred quality scores and negligible per-base N content. Adapter dimers contributed to a small percentage of the reads produced. Post paired-end read assembly FastQC analysis of a subset of samples revealed that the per-base and per-sequence Phred quality scores were slightly improved (per-sequence quality score 39-40) and contaminating adapter had been removed. On average, there were 17,174 reads per sample (including both α1 and α2 amplicons) and 99.7% of raw reads were paired. Investigation of variant lengths revealed that the α1 primers amplified a 206bp and 225bp fragment. All variants of 225bp in length contained premature stop codons and were assumed to be a pseudogene and consequently excluded from further analysis. Variants of 217bp and 214bp in length were produced for α2, with the size difference due to a 3bp insertion or deletion, with both size variants translating. All target lengths were present in all populations and the 214bp and 217bp α2 variants were analysed together.

Tables and figures

Table S1. List of inner and outer primer sequences and indexes used to amplify MHC class I in *C. decresii* and identify individuals.

Primer/index name	Description	Sequence (5' – 3')
MHC1_A1_G234_F	MHC class I α1 forward primer. Part of 'inner' primer.	GGCTCCTCCTCGCACTCCCTG
MHC1_A1_G234_R	MHC class I α1 reverse primer. Part of 'inner' primer.	ACATTCTGCAGGCTCACTCTGAAC
MHC1_A2_G234_F	MHC class I α2 forward primer. Part of 'inner' primer.	TGCAGTTGATGTACGGCTGTGAGC
MHC1_A2_G12_R	MHC class I α2 reverse primer. Part of 'inner' primer.	CCTCAGTAGGCTCTCCCTCCCG
MRT forward tag	Link inner and out primers	ACGACGTTGTAAAA
MRT reverse tag	Link inner and out primers	CATTAAGTTCCCATTA
P5	Part of 'outer' primer. Forward illumina-specific adapter.	AATGATACGGCGACCACCGAGATCTACAC
P7	Part of 'outer' primer. Reverse illumina-specific adapter.	CAAGCAGAAGACGGCATACGAGAT
PE Read 1 Sequencing Primer	Part of 'outer' primer. Forward illumina-specific adapter.	ACACTCTTTCCCTACACGACGCTCTTCCGATCT
Multiplexing Read 2 Sequencing Primer	Part of 'outer' primer. Reverse illumina-specific adapter.	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Index_F_1	i5 forward barcode	TCTCTGTG

Index_F_2	i5 forward barcode	ACTCACTG
Index_F_3	i5 forward barcode	TCTACTCG
Index_F_4	i5 forward barcode	TAGTAGCG
Index_F_5	i5 forward barcode	AGACGACG
Index_F_6	i5 forward barcode	ACTCGTAG
Index_F_8	i5 forward barcode	TGACGCAG
Index_F_10	i5 forward barcode	ACGCTATC
Index_F_11	i5 forward barcode	AGTGCTGC
Index_F_12	i5 forward barcode	TACTACGC
Index_F_14	i5 forward barcode	TCAGCTAC
Index_F_15	i5 forward barcode	AGAGCGAC
Index_F_16	i5 forward barcode	ATGCTCAC
Index_F_17	i5 forward barcode	TAGCACAC
Index_F_18	i5 forward barcode	TGTACGTG
Index_F_19	i5 forward barcode	AGAGTATG
Index_F_20	i5 forward barcode	ATGACTCG
Index_F_21	i5 forward barcode	AGATAGCG
Index_F_22	i5 forward barcode	TGTAGACG
Index_F_23	i5 forward barcode	TGCAGTAG
Index_F_24	i5 forward barcode	AGCTGATG
Index_F_25	i5 forward barcode	ATAGAGAG
Index_F_26	i5 forward barcode	TGTCACGC
Index_F_27	i5 forward barcode	ATACTGCG
Index_F_28	i5 forward barcode	ACTGTGTC
Index_F_29	i5 forward barcode	TACACAGC

Index_F_30	i5 forward barcode	TCGCTACG
Index_F_31	i5 forward barcode	ACGTACTC
Index_F_32	i5 forward barcode	TCGATGAC
Index_F_33	i5 forward barcode	TGCGATGC
Index_F_34	i5 forward barcode	AGCATCAC
Index_F_35	i5 forward barcode	TATCGATG
Index_F_36	i5 forward barcode	TATCAGAG
Index_F_37	i5 forward barcode	AGTCTAGC
Index_F_38	i5 forward barcode	TATATGCG
Index_F_39	i5 forward barcode	TACATGTC
Index_F_40	i5 forward barcode	ATACGTAC
Index_R_1	i7 reverse barcode	ATCGTCTG
Index_R_2	i7 reverse barcode	TGATCTAG
Index_R_3	i7 reverse barcode	ATGCATGC
Index_R_4	i7 reverse barcode	AGATGCAC
Index_R_5	i7 reverse barcode	ATGCGATG
Index_R_6	i7 reverse barcode	ACGCAGAG
Index_R_7	i7 reverse barcode	ATGTGAGC
Index_R_8	i7 reverse barcode	TGCTCGCG
Index_R_9	i7 reverse barcode	ATGACGTC
Index_R_10	i7 reverse barcode	AGTAGTAC
Index_R_11	i7 reverse barcode	TCTGACTC
Index_R_12	i7 reverse barcode	AGAGTCGC
Index_R_13	i7 reverse barcode	TCAGCACG
Index_R_14	i7 reverse barcode	TGCGTACG

Index_R_15	i7 reverse barcode	TCATGTCG
Index_R_16	i7 reverse barcode	TGACGATC
Index_R_17	i7 reverse barcode	TCACAGCG

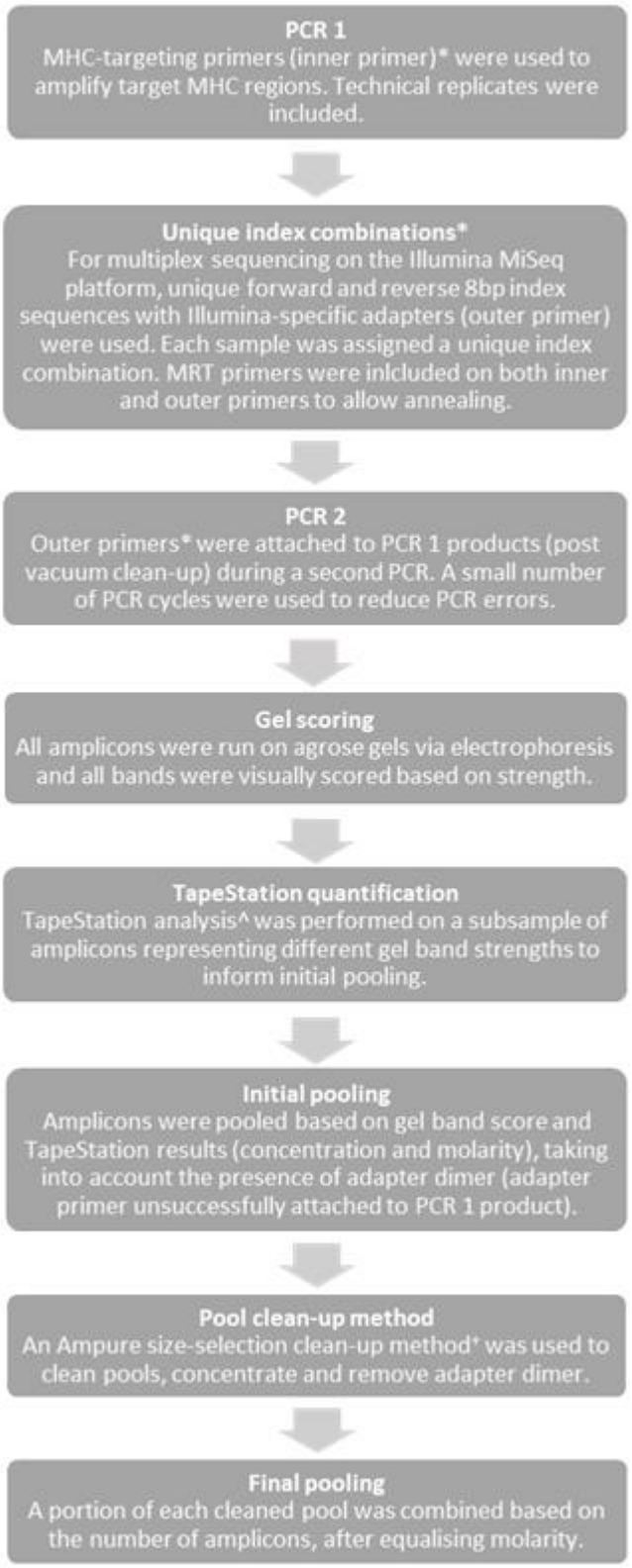


Figure S1. Steps taken during the two-step PCR Illumina MiSeq amplicon library preparation for *Ctenophorus decresii* MHC class I. *Primer concept as per (Pearson, et al., 2016); ^Agilent 2200 TapeStation © (Agilent Technologies); ⁺Agencourt AMPure © beads (Beckman Coulter, Inc.).

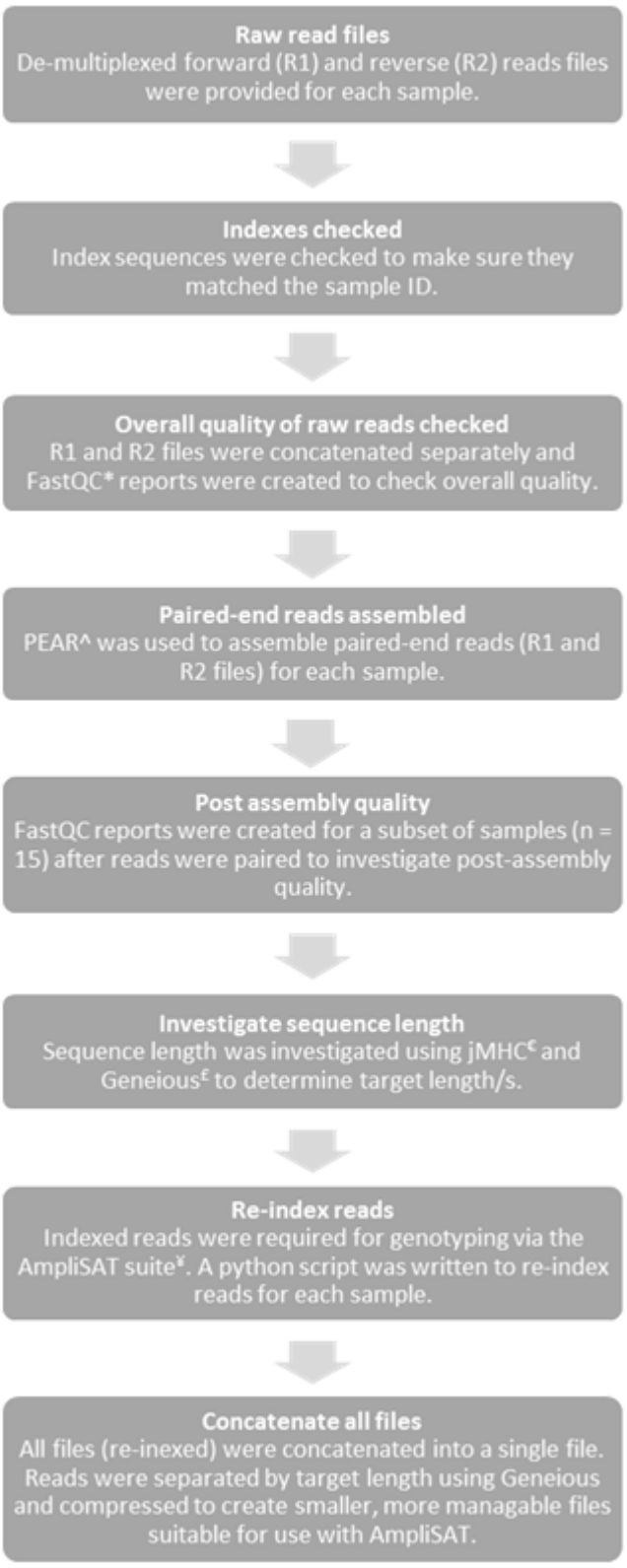


Figure S2. Pre-genotyping bioinformatics steps for *Ctenophorus decresii* class I MHC data. *FastQC ver. 0.11.2 (Andrews, 2010), ^PEAR ver. 0.9.5 (Zhang, et al., 2014), €jMHC ver. 1.6.1624 (Stuglik, et al., 2011), £Geneious ver. 8.1.7 (Kearse, et al., 2012), ^AmpliSAT (Sebastian, Herdegen, Migalska, & Radwan, 2016).

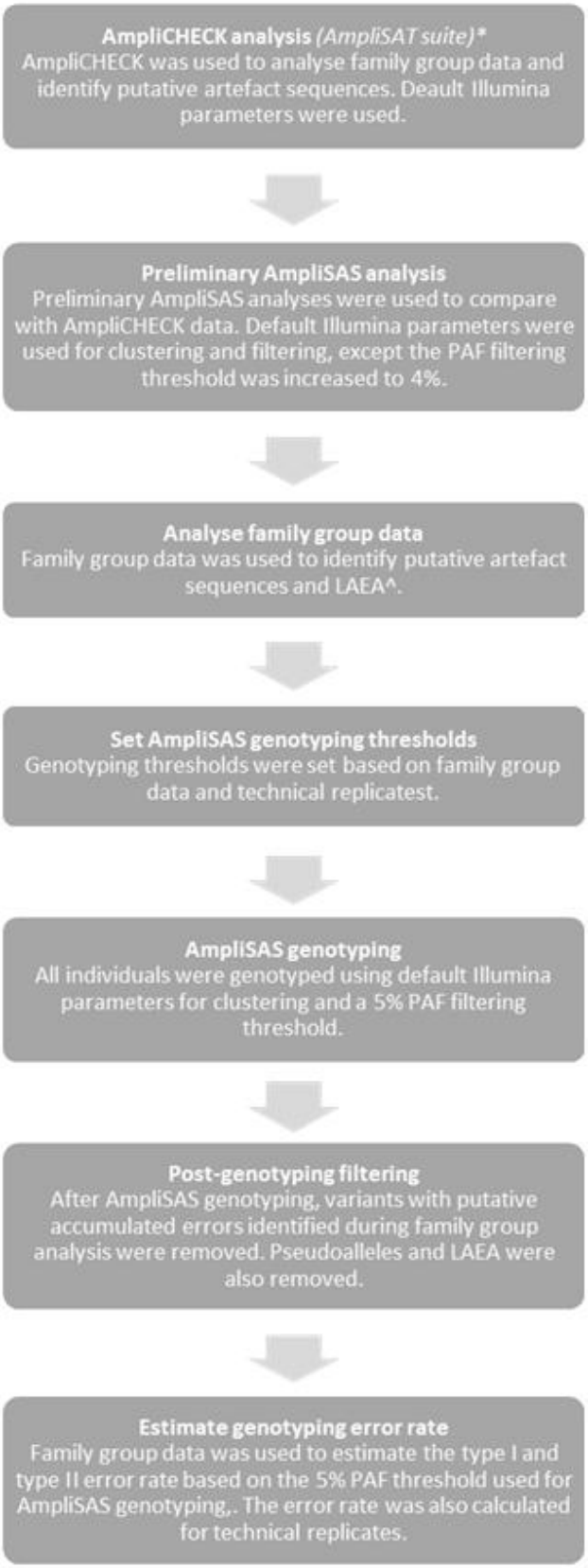


Figure S3. Steps taken during *Ctenophorus decresii* MHC class I genotyping. *AmpliSAT (Sebastian, et al., 2016). ^LAEA: low amplification efficiency alleles.

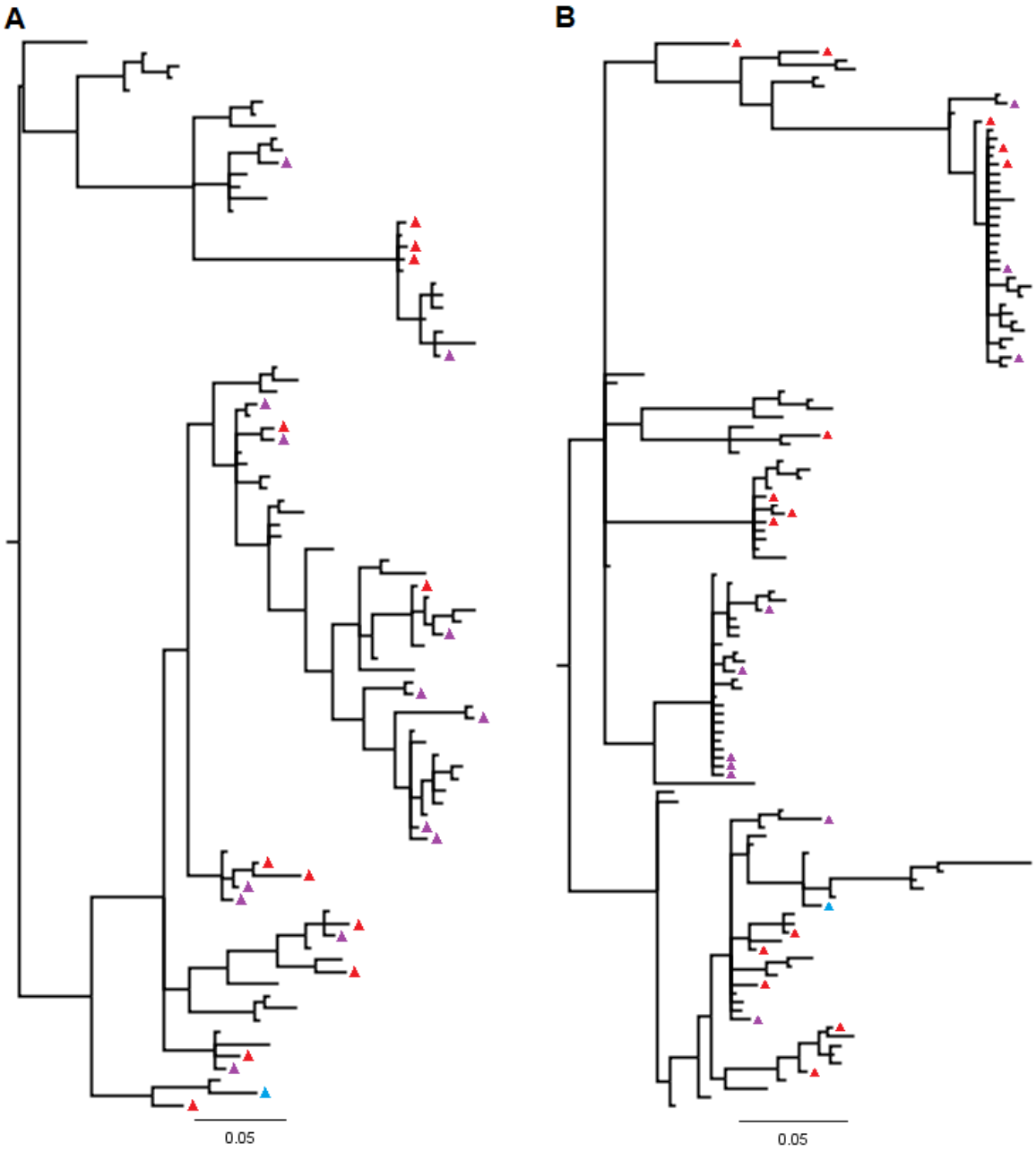


Figure S4. Bayesian trees including all α1 (A) and α2 (B) variants prior to post-genotyping filtering. Low amplification efficiency alleles are indicated with red triangles, purple triangles represent putative artefact sequences with accumulated errors (chimera and mismatch) and pseudogenes are indicated with blue triangles.

References

Andrews, S. (2010). FastQC a Quality Control Tool for High Throughput Sequence Data [Online]. Retrieved.

Hacking, J. D., Bertozzi, T., Moussalli, A., Bradford, T. M., & Gardner, M. G. (2017). Characterization of major histocompatibility complex class I transcripts in an Australian dragon lizard. *Manuscript submitted for publication*

Hayden, M., Nguyen, T., Waterman, A., & Chalmers, K. (2008). Multiplex-Ready PCR: A new method for multiplexed SSR and SNP genotyping. *BMC Genomics, 9*, pp. 1-12.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., . . . Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics, 28*(12), pp. 1647-1649. Retrieved from http://www.geneious.com

Li, C., Hofreiter, M., Straube, N., Corrigan, S., & Naylor, G. (2013). Capturing protein-coding genes across highly divergent species. *BioTechniques, 54*, pp. 321-326.

Pearson, S., Bradford, T., Ansari, T., Bull, M., & Gardner, M. (2016). MHC genotyping from next-generation sequencing: detailed methodology for the gidgee skink, *Egernia stokesii*. *Transactions of the Royal Society of South Australia, 140*(2), pp. 244-262.

Sebastian, A., Herdegen, M., Migalska, M., & Radwan, J. (2016). AMPLISAS: a web server for multilocus genotyping using next-generation amplicon sequencing data. *Molecular Ecology, 16*(2), pp. 498–510.

Stuglik, M., Radwan, J., & Babik, W. (2011). jMHC: Software assistant for multilocus genotyping of gene families using next-generation amplicon sequencing. *Molecular Ecology Resources, 11*, pp. 739-742.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3—new capabilities and interfaces. *Nucleic Acids Research, 40*(15), pp. e115-e115. doi:10.1093/nar/gks596 Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3424584/

Zhang, J., Kobert, K., Flouri, T., & Stamatakis, A. (2014). PEAR: A fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics 30*(5), pp. 614-620.

CHAPTER 5

Specific MHC class I supertype associated with parasite infection and colour morph in a wild lizard population

Note to examiners

This chapter investigates the mechanisms maintaining MHC diversity at the within-population level in *C. decresii*. Specifically, hypotheses regarding sexual selection and parasite-mediated selection were tested. Mated pairs were determined in chapter 3 and MHC genotypes were established in chapter 4. Furthermore, the positively selected sites described in chapter 4 were used in this chapter to cluster MHC alleles into functional supertypes. This chapter has been submitted to *Molecular Ecology* and is therefore formatted according to the specifications of that journal.

Hacking, J. D., Stuart-Fox, D. M., Godfrey, S. S., Gardner, M. G., (2018). Specific MHC class I supertype associated with parasite infection and colour morph in a wild lizard population. *Manuscript submitted for publication.*

Specific MHC class I supertype associated with parasite infection and colour morph in a wild lizard population

Jessica D. Hacking[a] (corresponding author)

jessica.hacking@flinders.edu.au

Devi Stuart-Fox[b]

d.stuart-fox@unimelb.edu.au

Stephanie S. Godfrey[c]

stephanie.godfrey@otago.ac.nz

Michael G. Gardner[a,d]

michael.gardner@flinders.edu.au

[a]College of Science and Engineering, Flinders University, Bedford Park SA 5042, Australia

[b]School of BioSciences, University of Melbourne, Parkville VIC 3010, Australia

[c]Department of Zoology, University of Otago, Dunedin 9016, New Zealand

[d]Evolutionary Biology Unit, South Australian Museum, Adelaide SA 5000, Australia

Abstract

The major histocompatibility complex (MHC) is a large gene family that plays a central role in the immune system of all jawed vertebrates. Non-avian reptiles are under-represented within the MHC literature and little is understood regarding the mechanisms maintaining MHC diversity in this vertebrate group. Here, we examined the relative roles of parasite-mediated selection and sexual selection in maintaining MHC class I diversity of a colour polymorphic lizard. We discovered evidence for parasite-mediated selection acting via rare-allele advantage or fluctuating selection as ectoparasite load was significantly lower in the presence of a specific MHC supertype (functional clustering of alleles); supertype four. Based on comparisons between ectoparasite prevalence and load, and assessment of the impact of ectoparasite load on host fitness, we suggest that supertype four confers quantitative resistance to ticks or an intracellular tick-borne parasite. We also discovered trends towards MHC-associated mating whilst taking the spatial position, relatedness and overall genetic diversity of mates and available males into account. An association was uncovered between supertype four and male throat colour morph. However, it is unlikely that male throat colouration acts as a signal of MHC genotype to conspecifics because we found no evidence to suggest that male throat colouration predicts male mating status. Overall, our results suggest that parasite-mediated selection plays a role in maintaining MHC diversity in this population via rare allele advantage and/or fluctuating selection. Further work is required to determine whether sexual selection also plays a role in maintaining MHC diversity in agamid lizards.

Keywords: Agamidae; major histocompatibility complex; parasite-mediated selection; MHC-associated mating; *Ctenophorus decresii*

Running title: Links between MHC, parasites and colouration

Introduction

Pathogen-host relationships can strongly influence ecological and evolutionary processes within wild populations (Harvell 2004). Understanding the mechanisms shaping host immunity is required for wildlife disease management and to evaluate the evolutionary consequences of disease (Acevedo-Whitehouse & Cunningham 2006). The major histocompatibility complex (MHC) is an extremely diverse gene family that plays a central role within the immune system of all jawed vertebrates (Kulski *et al.* 2002). Both parasite-mediated selection and sexual selection have been found to maintain this diversity (Piertney & Oliver 2006). However, these alternative sources of selection are rarely examined together in the same species, limiting the ability to explore their relative roles (although see Dunn *et al.* 2012; Eizaguirre *et al.* 2010; Sepil *et al.* 2013; Sepil *et al.* 2015).

Evidence for parasite-mediated selection acting on MHC genes via rare allele advantage (Borghans *et al.* 2004; Schwensow *et al.* 2017), fluctuating selection (Jones *et al.* 2015; Osborne *et al.* 2017), heterozygote advantage (Doherty & Zinkernagel 1975; Takahata & Nei 1990) and optimal intermediate diversity advantage (Wegner *et al.* 2003) has been uncovered across a range of taxa. However, most evidence points towards rare allele advantage and fluctuating selection as the dominant mechanisms by which pathogen-mediated selection maintains MHC diversity, with little evidence that heterozygote advantage alone can account for the extreme diversity found at MHC loci (De Boer *et al.* 2004).

Individual MHC alleles or supertypes (functional clustering of alleles) may provide resistance against (Savage & Zamudio 2011; Sepil *et al.* 2013), allow tolerance of (Regoes *et al.* 2014), or cause susceptibility to infection (Carrington *et al.* 1999). Interpreting the nature of such relationships requires information on both parasite prevalence and load, and the impact of infection on host fitness; data which is often difficult to obtain for populations in the wild (Råberg 2014; Råberg *et al.* 2007). Resistance may come in the form of complete (qualitative), or partial (quantitative) protection against parasites (Westerdahl *et al.* 2011). Under qualitative resistance, the host prevents the establishment of infection and completely clears infection. Quantitative resistance, on the other hand, allows the host to suppress parasite load but not completely clear infection. Tolerance may co-occur with quantitative resistance and refers to the ability of the host to withstand high parasite load without impacting fitness (Regoes *et al.* 2014). Tolerance is measured as the gradient of the relationship between Darwinian fitness (or a proxy of fitness) and infection intensity (Råberg 2014). Finally, parasite counter-adaptations to host defences may cause certain MHC alleles or supertypes to increase host susceptibility to infection (Kubinak *et al.* 2012). Understanding the nature of host-parasite relationships is important as different types of relationships have different consequences for epidemiology and the evolutionary dynamics of both host and parasite (Westerdahl *et al.* 2011).

MHC-associated mate choice has been discovered in most vertebrate classes, including bony fish (Evans *et al.* 2012; Reusch *et al.* 2001), amphibians (Bos *et al.* 2009), reptiles (Miller *et al.* 2009; Olsson *et al.* 2003; Pearson *et al.* 2017), birds (Juola & Dearborn 2012; Strandh *et al.* 2012) and mammals (Cutrera *et al.* 2012; Schad *et al.* 2012). Mate choice may be influenced by MHC diversity (high or intermediate), compatibility (high or intermediate diversity in offspring), and/or based on specific alleles or supertypes (Ejsmond *et al.* 2014). Spatial or temporal differences in mate choice for MHC characteristics may also occur (fluctuating selection, Cutrera *et al.* 2014). In some systems sexual selection may play a large role in maintaining MHC diversity. For instance, Winternitz *et al.* (2013) found that sexual selection explains more functional variation than parasite-mediate selection in mammals.

Two non-mutually exclusive hypotheses are used to explain MHC-associated mating; the good genes hypothesis and the complementary genes hypothesis. The good genes hypothesis involves mating that is influenced by MHC diversity, or specific alleles or supertypes, irrespective of the genotype of the choosy sex (absolute criteria, Brown 1997; Hamilton & Zuk 1982). The complementary genes hypothesis predicts that mating is based on MHC genotype compatibility between mates (self-referential criteria, Zeh & Zeh 1996). Hence, the genotype of the choosy sex is considered during mate choice. These hypotheses are used to test for evidence of heterozygote or intermediate diversity advantage, or associations with certain alleles or supertypes, indicating rare-allele advantage or fluctuating selection (Spurgin & Richardson 2010). Both olfactory (Boehm & Zufall 2006; Milinski *et al.* 2005; Setchell *et al.* 2011; Strandh *et al.* 2012) and visual (Dunn *et al.* 2012; Hinz *et al.* 2012; Milinski 2014; Olsson *et al.* 2005) traits have been proposed to signal individual MHC genotypes to conspecifics in mammals, birds and fish. For instance, Dunn *et al.* (2012) found that the male black facial masks of common yellowthroat birds likely act as a signal of MHC diversity to mates, and MHC-dependent peptides in mouse urine may signal MHC genotype to conspecifics (Sturm *et al.* 2013). However, the phenotypic traits used by reptiles and amphibians to signal MHC genotype to conspecifics are largely unknown.

Here, we examined the relative roles of sexual selection and parasite-mediated selection in maintaining MHC diversity within a wild reptile population. The Australian tawny dragon lizard (*Ctenophorus decresii*), for which MHC class I has been characterised (Hacking *et al.* 2018a; Hacking *et al.* 2018b), is host to both ectoparasites and intracellular parasites. Male *C. decresii* exhibit secondary sexual colouration on their throat and chest that is emphasised in displays to conspecifics (Gibbons 1979; Osborne 2005a, b; Osborne *et al.* 2012; Stuart-Fox & Johnston 2005). Furthermore, in some populations four discrete male throat colour morphs co-exist. Hence, *C. decresii* represents an excellent model to investigate patterns of MHC variation, parasites and visual signals. First, we investigated the role that parasite-mediated selection plays in maintaining MHC diversity by testing the hypothesis that

specific MHC supertypes are associated with parasite prevalence and/or load. We then determined whether MHC-parasite relationships were associated with resistance, tolerance or susceptibility. Second, we asked whether sexual selection, via MHC-associated mating, plays a role in maintaining MHC diversity. Specifically, we tested the hypothesis that MHC diversity and/or mate MHC compatibility predicts male mating status whilst accounting for the spatial position of mates, pair relatedness and mate overall genetic diversity. Finally, we investigated visual phenotypic traits that may signal MHC genotype to conspecifics.

Materials and methods

Male mating status

The tawny dragon is a small (<30 g) agamid lizard that is endemic to the rocky ranges of South Australia. Individuals were captured from a site near Hawker in the Flinders Ranges, South Australia (31°57'17.5"S, 138°22'26.4"E) by noosing and were then released at the point of capture. We sampled adult males and females during spring and summer between 2013 and 2015 and captive hatching was undertaken during spring and early summer of 2014, with each mother sampled once within the breeding season. Captive hatching and subsequent paternity analysis presented in Hacking *et al.* (In Press) produced 21 complete family groups, within which there was no evidence for multiple paternity. The tawny dragon employs a mostly polygynous genetic mating system (Hacking *et al.* In Press), with males patrolling territories of 213 m^2, on average (Yewers 2016). Male territories likely contain the home ranges of several females and a females' home range could overlap the territory of more than one male.

When investigating mate choice, it is important to take the spatial position of potential mates into account as females are likely to only come into contact with males that are close-by. Failing to account for the geographic distance between females and potential mates may cause mating preferences to be missed. We recorded the location (±3 m) of each individual at capture using a Garmin© handheld GPS. For each mother, adult males that were captured within a 100 m radius were considered to be potential mates. Given that male territory size ranges from 1 m^2 to 898 m^2 and averages 213 m^2 (Yewers 2016), this 100 m radius is large enough to encompass most males that a female may come into contact with but may also include some males that a female did not come into contact with. The average number of males available to each female was 21 (range 9 - 36). Four males mated with more than one female and most 'mated' males were available to other females during the breeding season. The geographic distance between all pairs (mated and available) was calculated.

Parasite infection

We recorded the number of attached nymph ticks (*Amblyomma limbatum*) for each individual (fig. S1) captured between 2013 and 2015 (166 total). Associations between the MHC and arachnid ectoparasites have been uncovered in both mammals (Kamath *et al.* 2014; Oliver *et al.* 2009; Schad *et al.* 2012) and reptiles (Radwan *et al.* 2014). Immune defence against haematophageous ectoparasites involves class II MHC molecules and potentially also MHC class I molecules via cross-presentation (Andrade *et al.* 2005; Rock *et al.* 2016; Wikel 1996). Furthermore, arachnid ectoparasites (ticks and mites) can transmit vector-borne intracellular parasites such as protists, viruses and bacteria to reptile hosts (Allison & Desser 1981; Bonorris & Ball 1955; Camin 1948; Chaisiri *et al.* 2015; Reynolds *et al.* 2017; Smallridge & Bull 1999). MHC I molecules play a direct role in immune defence against such intracellular parasites and vector-borne parasite infections are influenced by the prevalence and load of vectors (Bennett & Cameron 1974; Godfrey *et al.* 2011; Sol *et al.* 2000).

Male morphometrics and colouration

Snout-to-vent length (SVL) was measured for each male to the nearest 0.01 mm and mass to the nearest 0.25 g. Male body condition was estimated using residuals from a regression of adult male mass and SVL. Male *C. decresii* exhibit polymorphism for throat colour within the species' northern lineage, which includes the focal population (McLean *et al.* 2014; Teasdale *et al.* 2013). Four discrete, heritable morph types exist within these populations; orange surrounded by yellow (orange/yellow), orange, yellow or grey (McLean *et al.* 2014; Teasdale *et al.* 2013, fig. S2). The amount of colour on the throat and throat brightness is highly variable within morph types, which is likely influenced by both genetic and environmental factors (Rankin *et al.* 2016; Teasdale *et al.* 2013). Throat brightness and the amount of grey (associated with melanin pigment) is influenced by stress hormones under laboratory conditions (Lewis *et al.* 2017) and carotenoids can only be obtained from the diet (McGraw 2006). The throat is emphasised during male displays (Gibbons 1979) and it is likely that throat colouration plays an important role in social signalling (Yewers *et al.* 2016). To capture variation in colour within and among throat morphs we calculated the proportion of orange and yellow on the throat of each male. Variation in achromatic throat colouration ('brightness') among males, which is independent of morph type (Teasdale *et al.* 2013), was also calculated for each male. Male *C. decresii* also possess grey to black (melanin) chest patch markings that are exposed during male displays (Gibbons 1979, fig. S2). Chest patch size is an important signal in male-male interactions (Osborne 2005a) and in the congener *C. ornatus* chest patch size is associated with territory size and the number of females within a males' territory (Lebas 2001). We therefore also calculated male relative chest patch size. Refer to supplementary material (text S1) for additional details on the quantification of male throat colouration and chest patch size.

MHC genotyping and supertype analysis

MHC genotypes were obtained using next-generation amplicon sequencing and a thorough genotyping protocol, as described in Hacking *et al.* (2018b). The α1 domain (exon 2), which forms part of the peptide binding cleft of the MHC I molecule was targeted to gain information on the functional diversity present. As discussed in Hacking *et al.* (2018b), the MHC I α1 alleles used here likely represent a subsample of the diversity present at the *C. decresii* MHC I region, as is likely the case for many other MHC studies on non-model organisms (Babik 2010).

It is the characteristics of peptide binding sites and other important sites, such as those involved in T-cell receptor interactions within the domains that form the binding cleft of MHC molecules that determine associations between MHC molecules and parasite-derived peptides. Such sites are likely to be under positive selection. Clustering MHC alleles into functional groups (supertypes) based on the properties of positively selected sites (PSS) and putative peptide binding sites allows the phenotypic effects of MHC alleles to be examined (Naugler & Liwski 2008; Trachtenberg *et al.* 2003). Accordingly, PSS within the α1 domain identified in Hacking *et al.* (2018b) were used to cluster *C. decresii* alleles (including all populations sampled in Hacking *et al.* (2018b)) into supertypes based on five amino acid physicochemical descriptors: hydrophobicity, steric bulk, polarity and two electronic effect variables (Doytchinova & Flower 2005; Sandberg *et al.* 1998). First, amino acid positions under positive selection were extracted from the rest of the sequence and the physicochemical properties (Doytchinova & Flower 2005; Sandberg *et al.* 1998) of each amino acid for each allele were recorded. These data were formatted as a matrix with alleles in rows and the physicochemical properties of each amino acid as columns. Alleles were then clustered into supertypes using K-means clustering, implemented using the *adegenet* package (Jombart 2008) in R ver. 3.4.1 (R Core Team 2016).For K-means clustering all principle components were retained (n = 25) and the optimal number of clusters was determined based on Bayesian Information Criterion (BIC)(Jombart *et al.* 2010).

Pair relatedness and genome-wide diversity

Under MHC-associated mate choice, a lack of correlation between MHC diversity and genome-wide diversity would suggest that mating isn't simply due to mate choice for genetically diverse individuals; while a lack of correlation between relatedness and MHC similarity eliminates mating due to general inbreeding avoidance. Therefore, we tested for correlations between MHC diversity (number of MHC alleles) and genome-wide diversity (microsatellite heterozygosity), and between relatedness and percent genetic (amino acid) distance among MHC I alleles shared between pairs. Microsatellite genotypes presented in Hacking *et al.* (In Press) were used to estimate relatedness between each mother, her mate and all available males, using Coancestry ver. 1.0.1.5 (Wang 2007). Microsatellite

genotypes were also used to estimate individual heterozygosity using the *genhet* package (Coulon 2010) in R (R Core Team 2016) as a measure of neutral genome-wide diversity. Correlation analyses were undertaken using the *glm* function in R (R Core Team 2016).

Hypothesis testing overview

An AICc-based information-theoretic approach was used to test our alternative hypotheses about the factors driving MHC diversity in this system (Burnham & Anderson 2004; Burnham *et al.* 2011; Galipaud *et al.* 2014; Grueber *et al.* 2011; Symonds & Moussalli 2011). We examined evidence for parasite mediated selection in model set 1 and MHC-associated mating in model set 2. We then investigated potential phenotypic signals of MHC genotype in model set 3. Correlations between predictor variables were investigated and variance inflation factors were calculated (VIF, *usdm* R package, Naimi, Hamm, Groen, Skidmore, & Toxopeus, 2014) to avoid multicollinearity within models. A VIF below 3 was considered acceptable (Zuur *et al.* 2010). See table S1 for an outline of all models.

We used generalized linear models (GLM) or generalized linear mixed models (GLMM) to estimate the effect of potential predictors on response variables that were normally distributed. Four response variables; percentage of throat coloured orange, percentage of throat coloured yellow and tick load, were zero-inflated with 'true zeros' or 'structure zeros'. That is, zeros resulting from sub-populations within the dataset rather than from random sampling ('false zeros' or 'sampling zeros', He *et al.* 2014; Martin *et al.* 2005). For example, the percentage of throat coloured orange variable is measure of the percentage of a males' throat coloured orange and is zero inflated because two of the four morph types (grey and yellow) don't include orange colouration. Due to the nature of the zero-inflation, hurdle models were fitted when percentage of throat coloured orange, percentage of throat coloured yellow and tick load were used as response variables. Hurdle models include two components: first, a binomial model determines whether a zero or count (non-zero) outcome occurs (presence/absence). Second, a zero-truncated (excluding zeros) model (e.g. poisson) analyses the count data (Dalrymple *et al.* 2003; Guo *et al.* 2016; Hassrick *et al.* 2016; Naimi *et al.* 2014; Welsh *et al.* 1996; Xu *et al.* 2015). The fit of both a negative binomial and poisson distribution were considered for the zero-truncated (count) part of the hurdle models.

For each model set, a global model was constructed, which was standardised using the *arm* R package (Gelman 2008). Standardisation allowed direct comparison among predictor variables during models selection (Gelman 2008; Grueber *et al.* 2011). The *lme4* R package was used to construct GLMMs (Bates *et al.* 2015), the *stats* R package was used to construct GLMs (R Core Team 2016) and the *pscl* R package was used to construct hurdle models (Jackman 2012; Zeileis *et al.* 2008). Then, the *dredge* function from the *MuMIn* R package (Bartoń 2009) was used to construct all possible models

based on the global model, including the null model. When co-variates were used in models they were included in all models, including the null model, constructed by the *dredge* function. The top models were extracted based on a ΔAICc 95% confidence set (Symonds & Moussalli 2011). Finally, top models were averaged so that parameters were recalculated based on the top model set alone. Model selection and averaging was undertaken using the *MuMIn* R package (Bartoń 2009). Model fit was examined using adjusted R^2 for GLMs (Nagelkerke 1991) and marginal (R^2_m) and conditional (R^2_c) R^2 for GLMMs (Nakagawa & Schielzeth 2013). R^2 calculations were undertaken using the *MuMIn* R package (Bartoń 2009). The fit of hurdle models was visually assessed using rootograms, which were created using the *countreg* R package (Kleiber & Zeileis 2016).

The support for models within model sets was determined based on ΔAICc, the evidence ratio (how much better one model explains the data than the next model) and model fit. The importance of specific predictor variables was based on effect size, accumulative Akaike weights (relative importance) and statistical significance (based on 95% CIs).

Hypothesis testing: parasite-mediated selection

In model set 1 we tested the hypothesis that the presence of specific MHC I supertypes predicts tick prevalence and/or load, using a hurdle model (table S1). Such a relationship would indicate that parasite-mediated selection is occurring via rare-allele advantage or fluctuating selection (Spurgin & Richardson 2010). We could not estimate MHC heterozygosity because we amplified MHC I alleles across multiple loci (up to four alleles per individual, Hacking *et al.* 2018b) and therefore did not test for evidence of parasite-mediated selection acting through heterozygote advantage (Spurgin & Richardson 2010).

Information on the impact of parasite load on host fitness is required when delineating MHC-pathogen relationships (Råberg *et al.* 2007). Body condition was used as a proxy for fitness and was calculated using residuals of a regression of SVL against mass (Jakob *et al.* 1996). We plotted male body condition against tick load and grouped data based on the presence and absence of MHC I supertypes that were identified in model selection. A linear regression line was fitted for each group (absence/presence) for each supertype to assess the relationship between body condition and tick load using the *stats* R package (R Core Team 2016). Only male individuals were used as body condition data was not available for females due to the impact of reproduction on female body condition.

Hypothesis testing: sexual selection

We tested two alternative but potentially non-mutually exclusive hypotheses for MHC-associated mating; i) 'mate choice' for MHC diversity or specific MHC supertypes; the good genes hypothesis, and ii) 'mate choice' for MHC-compatible individuals; the complementary genes hypothesis (table S1,

Eizaguirre *et al.* 2009; Landry *et al.* 2001; Miller *et al.* 2009; Pearson *et al.* 2017; Sepil *et al.* 2015). Mate choice was determined by comparing the male a female mated with to a set of 'available' males (within a 100 m radius). This variable may reflect female mate choice (intersexual selection) and/or male-male competition for access to females (intrasexual selection) but is not a measure of male choice or male reproductive success as all females available to males were not sampled. For model set 2, MHC diversity (good genes hypothesis) was estimated as the number of male MHC I alleles and pair MHC I genetic distance (complementary genes hypotheses) was estimated as the average percent genetic distance (amino acid) between shared MHC I alleles of a male and female pair. Male mass and the spatial proximity (m) of available and mated males were included as covariates in model set 2 because these were strong predictors of male mating status in preliminary analyses, with mated males larger, and geographically closer, than available males. Male and female ID were included as random factors within the model to account for repeated individuals within the dataset (i.e. many males were available to a single female, some males mated with more than one female, and many males were both mated and available to other females). To test whether the probability of possessing a particular supertype is dependent on mate mating status (good genes hypothesis) we performed a Fisher's exact test, implemented in R using the *stats* package (R Core Team 2016) and the *rcompanion* package (Mangiafico 2015).

Following model set 2, we investigated potential phenotypic (colouration) signals of male MHC diversity, including all adult males sampled within the population (n = 108, model set 3a-f, table S1). We included percentage of throat coloured orange, percentage of throat coloured yellow, throat brightness and chest patch as predictors. Male number of alleles (model set 3a) and male number of supertypes (model set 3b) are measures of MHC diversity and were used as response variables. We also tested for associations between potential phenotypic signals and specific MHC I supertypes (model sets 3c – 3f). Each of the potential phenotypic signals were used a response variable and each supertype was included as a separate predictor variable, coded as present or absent for each individual. To further investigate trends uncovered between throat colour and supertype four (model sets 3c and 3d) we performed a Fisher's exact test to determine whether the probability of possessing supertype four is dependent on male throat morph type (yellow/orange-yellow/orange/grey).

To confirm that the phenotypic traits that were found to be associated with specific supertypes also predicted mating status, we performed a GLM with mating status as the response and male percentage throat coloured yellow and male percentage throat coloured yellow as predictors (model set 3g, table S1). As with model set 2, male mass was used as a covariate. Only males were included in the mating status variable, rather than pairs of males and females. We also performed a Fisher's exact test to determine whether mating status is dependent upon male throat morph type.

Results

MHC genotyping and supertype analysis

Most of the mothers (18/21, 90%), fathers (14/16, 88%) and available males (95/108, 88%), and 86% (144/166) of the individuals used to test for parasite-mediated selection were successfully genotyped for the MHC I α1 domain. Individuals had between one and four alleles and pair MHC I genetic distance varied from 0.02 to 0.36. A total of 28 MHC I alleles were uncovered, which translated into 27 unique amino acid sequences (Hacking *et al.* 2018b). The number of alleles per individual, as measured by nucleotide sequences, was the same as when measured by amino acid sequences.

Alleles from all *C. decresii* populations sampled in Hacking *et al.* (2018b) clustered into eight supertypes (fig. S4 and S5) and individuals had between one and four supertypes. Seven of these supertypes (ST2 – ST8) were present within the focal population. Most (89%) individuals with more than one MHC allele had an equal number of alleles and supertypes, indicating high within-individual functional diversity. The frequency of each supertype varied, with supertype three present in 52% of individuals and supertype seven present in only 4% of individuals. Supertype six was also rare (9%) and all other supertypes had an intermediate frequency (18-24%, fig. S6).

Pair relatedness and genome-wide diversity

There was no correlation between number of male MHC I alleles and microsatellite heterozygosity (R^2 = 0.002, p = 0.193) or between MHC I genetic (amino acid) similarity between pairs and pair relatedness (R^2 = 0.001, p = 0.240, fig. S7). This suggests that any MHC-associated mating patterns observed aren't due to mate choice for genetically diverse individuals or mating to avoid inbreeding.

Parasite-mediated selection

We uncovered associations between certain supertypes and tick load. Overall, 87% of individuals were infected with ticks and average tick load was seven. The top model included only supertype four, which was present in all models with a ΔAICc less than two (table 1). Supertypes three and seven were also in models with ΔAICc less than two. These associations were driven by the count component of the hurdle model (fig. S11), suggesting that the presence or absence of these supertypes is associated with the tick load, rather than tick prevalence. Indeed, tick prevalence is similar in the presence and absence of supertype four (fig. 1a). Tick load was significantly lower in the presence of supertype four (95% CIs -1.21, -0.11; odds ratio 0.51; fig. 1b). Tick load was also lower in the presence of supertype seven but was higher in the presence of supertype three. However, supertypes three and seven had lower effect sizes compared to supertype four and relationships were not statistically significant (95% CIs

overlapped zero, table S3, fig. S11). The rootogram confirmed that the hurdle model provided a good fit for the data (fig. S8).

The significant negative relationship between supertype four and tick load, and the neutral relationship between supertype four and tick prevalence, suggests that this supertype confers quantitative resistance to ticks or a tick-borne parasite. To investigate whether supertype four also plays a role in tolerance of ticks (or a tick-borne parasite) we examined the relationship between body condition (a measure of fitness) and tick load in light of the presence and absence of supertype four (Fig. 1c). When only individuals that don't possess supertype four are considered there is a slight non-significant negative relationship (p-value = 0.69, slope = -0.01, Fig. 1c) between body condition and tick load, indicating that high tick loads probably have a small impact on body condition. In contrast, individuals that possess supertype four show a much steeper decline in body condition with increasing tick load, although this relationship is also not statistically significant (p-value = 0.08, slope = -0.18, Fig. 1c). It is therefore likely that supertype four offers only resistance and not tolerance to ticks or a tick-borne parasite.

MHC-associated mating

We found limited evidence for mate choice based on MHC diversity or specific MHC supertypes (good genes hypothesis), or pair MHC I genetic distance (complementary genes hypothesis). Pair MHC I genetic distance was 13% higher in mated pairs than available pairs (fig. 2). Mated males had on average 1.65 alleles, which is lower than the average for available males (1.84), and mated males also possessed two MHC alleles at most, whereas available males possessed up to four alleles (fig. 1). However, these differences were not statistically significant. Three models were within the 95% confidence set for model set 2 (table S4). The model with the lowest ΔAICc included pair MHC I genetic distance only, the second model was the null model and the third was the full model, including both pair MHC I genetic distance and number of male MHC I alleles. The 95% CIs overlapped zero and effect sizes were small for both pair MHC I genetic distance (95% CIs -0.85, 2.35 and odds ratio 3.12, table S5, fig. S12) and number of male MHC I alleles (95% CIs -0.80, 0.68. and odds ratio 0.72, table S5, Fig. S12). All top models had ΔAICc values <2 and evidence ratios were small (1.4 – 1.9), indicating that all models explained the data equally well. In regards to mate choice for specific MHC I supertypes, the Fisher's exact test was non-significant (p-value = 0.11), indicating that the probability of possessing a certain supertype is independent of male mating status.

Potential signals of MHC genotype

We found little evidence to suggest that male colouration signals MHC genotype to conspecifics. Neither the number of alleles, nor the number of supertypes were associated with male colouration

(throat colour, throat brightness and relative chest patch size). There was some evidence for an association between male colouration and the presence of certain supertypes (fig. 3), although these associations were not statistically significant. Refer to supplementary material for details of model set 3a – 3f results (text S2). Model sets 3c and 3d indicate that orange and orange/yellow throat morphs are less likely to possess supertype six, and yellow and orange/yellow throat morphs are less likely to possess supertypes three and five. Furthermore, supertype four is less likely to occur in orange, orange/yellow and yellow morphs. Differences among morphs was most pronounced for supertype four; over 50% of grey morphs possessed supertype four, whereas less than 20% of yellow, orange and orange/yellow morphs possessed this supertype (fig. 3c). Indeed, a Fisher's exact test revealed that the probability of possessing supertype four is dependent upon morph type (p-value = 0.02), with the grey morph more likely to possess supertype four than any other morph type.

Because we uncovered trends suggesting that percentage of throat coloured yellow and percentage of throat coloured orange may be associated with certain MHC I supertypes, we investigated whether these variables predicted mating status (model set 3g). The null model had the lowest ΔAICc and 95% CIs for percentage of throat coloured yellow and percentage of throat coloured orange overlapped zero, suggesting that percentage of throat coloured orange and percentage of throat coloured yellow are not strong predictors of mating status (tables S19 and S20). Furthermore, a Fisher's exact test revealed that male mating status is independent of male throat morph type (p-value = 0.58).

Discussion

We tested hypotheses regarding the roles of parasite-mediated selection and sexual selection in maintaining diversity at the MHC within a wild agamid lizard population. The roles of both parasite-mediated selection and sexual selection are rarely considered together in a single population, limiting the ability to make inferences about the relative importance of these mechanisms. Results supported the hypothesis that specific MHC I supertypes are associated with parasite infection, indicating a role for parasite-mediated rare allele advantage and/or fluctuating selection in maintaining MHC diversity. Parasite load was significantly lower in the presence of MHC I supertype four but this supertype had no effect on parasite prevalence, indicating quantitative resistance. Furthermore, the relationship between parasite load and body condition in the presence of supertype four indicated that this supertype offers only resistance, not tolerance. Tolerance is rarely considered when delineating host-parasite relationships despite important implications for epidemiology and host-parasite co-evolutionary dynamics.

In contrast to parasite-mediated selection, we found limited evidence that sexual selection plays a role in maintaining *C. decresii* MHC I diversity. There was little support for mating based on MHC diversity, specific MHC supertypes, or pair MHC compatibility when taking potential mate spatial proximity, inbreeding avoidance and overall genetic diversity into account. Similar results have recently been reported for great tit birds, in which there is strong evidence for parasite-mediated selection (Sepil *et al.* 2013) but little evidence for sexual selection (Sepil *et al.* 2015) maintaining MHC diversity. In line with the lack of strong evidence for MHC-associated mating, we did not find any evidence to suggest that male colouration acts as a signal of MHC genotype to conspecifics. However, we did find a link between male throat colour morph and supertype four, with grey-throated males over twice as likely to possess this supertype compared to other morphs.

Parasite-mediated selection

Understanding the genetic basis of variation in infections within and among wildlife populations requires identification of host-parasite immunogenetic relationships (i.e. resistance, tolerance and susceptibility) and the evolutionary mechanisms driving such relationships. Recently, researchers have aimed to better understand the relationships between host genetic immunity and parasites; attempting to differentiate qualitative resistance, quantitative resistance and susceptibility. For instance, Sepil *et al.* (2013) discovered that two different MHC supertypes are associated with two different avian malaria (*Plasmodium*) species, but one confers qualitative resistance and the other offers quantitative resistance. However, the role of host tolerance is rarely considered in such studies. One of the few examples of research on the role that MHC molecules play in tolerance is Regoes *et al.* (2014), who found that MHC heterozygosity is associated with greater tolerance of human HIV infections.

Our results suggest that supertype four confers quantitative resistance against but not tolerance of ticks or a tick-borne parasite infecting *C. decresii*. Consistent with resistance, we uncovered a negative relationship between supertype four and parasite load. Associations between infection prevalence and specific MHC alleles or supertypes are likely linked to the ability to completely clear infection, not the ability to prevent infection. This is because the MHC is not directly involved in the initial innate immune response to infection (Chaplin 2010). We uncovered a neutral relationship between supertype four and parasite prevalence, indicating that this supertype does not play a role in clearing infection, as expected under quantitative resistance.

It is likely that supertype four does not offer tolerance alongside quantitative resistance, as revealed by the negative relationship between body condition and parasite load in the presence of this supertype. In the absence of supertype four body condition decreased with an increase in parasite load at a lower rate. Discrepancy between individuals with and without supertype four in regards to the tolerance gradient may be caused by a cost associated with immune response. For instance, immune

response is negatively correlated with reproductive effort in birds (Knowles *et al.* 2009) and with body size and development time in field crickets (Rantala & Roff 2005). However, often multiple measures of fitness are required to gauge the impact of infection on individuals. Furthermore, it is unknown whether body condition, correlates with reproductive success in *C. decresii* or agamid lizards more generally. Recent work in sand lizards (Dudek *et al.* 2015) and anolis lizards (Cox & Calsbeek 2015) suggests that body condition may not be an accurate measure of fitness for lizards. Hence, the decrease in body condition associated with increasing parasite load observed for *C. decresii* may not necessarily indicate an impact of parasite load on lizard health or relative fitness.

When testing the parasite-mediated selection hypothesis we did not take into account risk of infection. Individuals that are unexposed to ticks cannot become infected regardless of their MHC genotype. For instance, sleepy lizards (*Tiliqua rugosa*) that are highly socially connected and use the same refuges as neighbouring lizards have higher tick loads (Leu *et al.* 2010). Therefore, if variation in tick exposure exists within the *C. decresii* population the relationship we observed between specific supertypes and tick prevalence may be biased. Finally, it is worth noting that in this study we only considered the α1 domain from a subset of *C. decresii* MHC I loci. We did not include the α2 domain, which together with α1, completes the peptide binding region, or MHC class II loci. Future studies will likely benefit from including the entire peptide binding region or complete MHC haplotypes.

We identified an association between a specific MHC genotype (supertype four) and parasite load, which is one of the signatures expected under parasite-mediated rare allele advantage and fluctuating selection (Spurgin & Richardson 2010). Further work involving long-term spatiotemporal data is required to assess whether rare allele advantage or fluctuating selection play a dominant role, or if they are acting together in this system. Furthermore, we were unable to test for heterozygote advantage because alleles were amplified across multiple loci. Therefore, it is possible that parasite-mediated heterozygote advantage also plays a role in maintaining MHC diversity in *C. decresii*.

Sexual selection: MHC-associated mating

We uncovered limited evidence for MHC-associated mating in regards to both the good genes and complementary genes hypotheses. This suggests that MHC I loci play a secondary role or are of little importance in mate choice decisions in *C. decresii*. In contrast to our results, the few studies that have examined MHC-associated mate choice in non-avian reptiles have uncovered significant relationships between measures of MHC diversity or dissimilarity and mating (Miller *et al.* 2009; Olsson *et al.* 2005; Pearson *et al.* 2017). Other studies have reported a lack of evidence for MHC-associated mating (Kuduk *et al.* 2014; Whitcomb *et al.* 2014). For instance, Sepil *et al.* (2015) found no evidence for MHC-disassortative mating in wild great tits despite large sample sizes. Moreover, we did not identify male phenotypic signals of MHC genotype, providing further evidence that class I MHC may play little role in

mate choice. However, it is possible that other phenotypic traits that we did not measure, such as femoral pore secretions or other olfactory cues signal MHC I genotype to conspecifics. It is important to note that the sample size associated with the mating status variable may have inhibited our ability to identify small effects related to MHC-associated mating and phenotypic mating signals. Furthermore, we only considered a single exon from a subset of MHC I loci; MHC-associated mate choice may be detected when entire MHC I haplotypes or MHC class II loci are considered.

Supertype four, parasites and male morph type

Supertype four likely provides quantitative resistance to ticks or a tick-borne parasite in *C. decresii*. This supertype was also associated with male throat colour; the grey morph was more likely to possess supertype four compared to other morphs. *Ctenophorus decresii* morphs differ in behaviour and hormone levels with the grey morph being the least bold and least aggressive, with low testosterone levels (Yewers *et al.* 2017; Yewers *et al.* 2016). Despite these characteristics likely reducing the ability of grey morphs to defend territories and potentially mates, the grey morph is present within all polymorphic populations, implying a compensatory selective advantage (McLean *et al.* 2014). Our results suggest that a superior ability to reduce parasite load could provide the grey morph with a selective advantage, allowing it to be maintained within the population. Compared to other morphs, the grey morph has the highest expression levels of several genes associated with melanin synthesis (McLean *et al.* 2017). Melanin colouration has also been linked to the MHC in the common yellowthroat, *Geothlypis trichas* (Dunn *et al.* 2012) and brown trout, *Salmo trutta* (Jacob et al. 2010). Although the genetic and biochemical pathways responsible for this link are unknown, melanins undertake several functions within the immune system (Nosanchuk & Casadevall 2006), suggesting a possible link between MHC molecules and melanin pigmentation.

Conclusion

Both parasite-mediated selection and sexual selection are rarely studied together in a single population when investigating the mechanisms maintaining MHC diversity. Here, we considered both sources of selection and discovered strong evidence for parasite-mediated selection but little evidence for MHC-associated mate choice. When testing MHC-associated mate choice hypotheses we controlled for spatial proximity among individuals, relatedness and overall genetic diversity. Such potential confounding variables are not always considered in studies examining mate choice for MHC characteristics. Our results suggest that parasite-mediated selection may be acting via rare-allele-advantage and/or fluctuating selection to maintain MHC diversity with *C. decresii*, with a specific supertype likely conferring quantitative resistance to ticks or a tick-borne parasite. Overall, this study supports a dominant role for parasite-mediated selection in maintaining MHC diversity and further

demonstrates that the role that sexual selection plays is highly variable within and among vertebrate groups.

Acknowledgements

We are grateful to Tessa Bradford for helpful discussions and to Rachel Popelka-Filcoff for assistance with spectrophotometry. Funding was provided by the Wildlife Preservation Society of Australia and Royal Society of South Australia, awarded to JH and the Flinders University Interfaculty Collaboration Grant, awarded to JH, MG and a third applicant.

References

Acevedo-Whitehouse K, Cunningham AA (2006) Is MHC enough for understanding wildlife immunogenetics? *Trends in Ecology and Evolution* **21**, 433-438.

Allison B, Desser SS (1981) Developmental Stages of Hepatozoon lygosomarum (Doré 1919) comb. n. (Protozoa: Haemogregarinidae), a Parasite of a New Zealand Skink, Leiolopisma nigriplantare. *The Journal of Parasitology* **67**, 852-858.

Andrade BB, Teixeira CR, Barral A, Barral-Netto M (2005) Haematophagous arthropod saliva and host defense system: a tale of tear and blood. *Anais da Academia Brasileira de Ciências* **77**, 665-693.

Babik W (2010) Methods for MHC genotyping in non-model vertebrates. *Molecular Ecology Resources* **10**.

Bartoń K (2009) MuMIn: multi-model inference, p. R package.

Bates D, Mächler M, Bolker B, Walker S (2015) Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* **67**, 48.

Bennett GF, Cameron M (1974) Seasonal prevalence of avian hematozoa in passeriform birds of Atlantic Canada. *Canadian Journal of Zoology* **52**, 1259-1264.

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

Bonorris JS, Ball GH (1955) Schellackia occidentalis n.sp., a Blood-inhabiting Coccidian Found in Lizards in Southern California. *The Journal of Protozoology* **2**, 31-34.

Borghans JAM, Beltman JB, De Boer RJ (2004) MHC polymorphism under host-pathogen coevolution. *Immunogenetics* **55**, 732-739.

Bos DH, Williams RN, Gopurenko D, Bulut Z, DeWoody JA (2009) Condition-dependent mate choice and a reproductive disadvantage for MHC-divergent male tiger salamanders. *Molecular Ecology* **18**, 3307-3315.

Brown JL (1997) A theory of mate choice based on heterozygosity. *Behavioral Ecology* **8**, 60-65.

Burnham KP, Anderson DR (2004) Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods & Research* **33**, 261-304.

Burnham KP, Anderson DR, Huyvaert KP (2011) AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behavioral Ecology and Sociobiology* **65**, 23-35.

Camin JH (1948) Mite Transmission of a Hemorrhagic Septicemia in Snakes. *The Journal of Parasitology* **34**, 345-354.

Carrington M, Nelson GW, Martin MP*, et al.* (1999) HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage. *Science* **283**, 1748-1752.

Chaisiri K, McGarry JW, Morand S, Makepeace BL (2015) Symbiosis in an overlooked microcosm: a systematic review of the bacterial flora of mites. *Parasitology* **142**, 1152-1162.

Chaplin DD (2010) Overview of the Immune Response. *The Journal of allergy and clinical immunology* **125**, S3-23.

Coulon A (2010) genhet: an easy-to-use R function to estimate individual heterozygosity. *Molecular Ecology Resources* **10**, 167-169.

Cox RM, Calsbeek R (2015) Survival of the fattest? Indices of body condition do not predict viability in the brown anole (Anolis sagrei). *Functional Ecology* **29**, 404-413.

Cutrera AP, Fanjul MS, Zenuto RR (2012) Females prefer good genes: MHC-associated mate choice in wild and captive tuco-tucos. *Animal Behaviour* **83**, 847-856.

Cutrera AP, Zenuto RR, Lacey EA (2014) Interpopulation differences in parasite load and variable selective pressures on MHC genes in *Ctenomys talarum*. *Journal of Mammalogy* **95**, 679-695.

Dalrymple ML, Hudson IL, Ford RPK (2003) Finite Mixture, Zero-inflated Poisson and Hurdle models with application to SIDS. *Computational Statistics & Data Analysis* **41**, 491-504.

De Boer R, Borghans J, van Boven M, Kesmir C, Weissing F (2004) Heterozygote advantage fails to explain the high degree of polymorphism of the MHC. *Immunogenetics* **2004**, 725-731.

Doherty PC, Zinkernagel RM (1975) Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. *Nature* **256**, 50-52.

Doytchinova IA, Flower DR (2005) In silico identification of supertypes for class II MHCs. *Journal of Immunology* **174**, 7085-7095.

Dudek K, sajkowska Z, gawałek M, ekner-grzyb A (2015) Using body condition index can be an unreliable indicator of fitness: a case of sand lizard *Lacerta agilis* Linnaeus, 1758 (Sauria: Lacertidae). *Turkish Journal of Zoology* **39**, 182-184.

Dunn P, Bollmer J, Freeeman-Gallant CR, Whittingham L (2012) MHC variation is related to a sexually selected ornament, survival, and parasite resistance in common yellowthroats. *Evolution* **67**, 1-9.

Eizaguirre C, Lenz TL, Sommerfeld RD*, et al.* (2010) Parasite diversity, patterns of MHC II variation and olfactory based mate choice in diverging three-spined stickleback ecotypes. *Evolutionary Ecology* **25**, 605-622.

Eizaguirre C, Yeates SE, Lenz T, Kalbe M, Milinski M (2009) MHC-based mate choice combines good genes and maintenance of MHC polymorphism. *Molecular Ecology* **18**, 3316–3329.

Ejsmond M, Radwan J, Wilson A (2014) Sexual selection and the evolutionary dynamics of the major histocompatibility complex. *Proceedings of the Royal Society B: Biological Sciences* **281**, 1-8.

Evans ML, Dionne M, Miller KM, Bernatchez L (2012) Mate choice for major histocompatibility complex genetic divergence as a bet-hedging strategy in the Atlantic salmon (*Salmo salar*). *Proceedings of the Royal Society B: Biological Sciences* **279**, 379-386.

Galipaud M, Gillingham MAF, David M, Dechaume-Moncharmont F-X (2014) Ecologists overestimate the importance of predictor variables in model averaging: a plea for cautious interpretations. *Methods in Ecology and Evolution* **5**, 983-991.

Gelman A (2008) Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine* **27**, 2865-2873.

Gibbons J (1979) The hind leg pushup display of the *Amphibolurus decresii* species complex (Lacertilia: Agamidae). *Copeia* **1**, 29-40.

Godfrey SS, Nelson NJ, Bull CM (2011) Ecology and dynamics of the blood parasite, Hepatozoon tuatarae (Apicomplexa), in tuatara (Sphenodon punctatus) on Stephens Island, New Zealand. *J Wildl Dis* **47**, 126-139.

Grueber CE, Nakagawa S, Laws RJ, Jamieson IG (2011) Multimodel inference in ecology and evolution: challenges and solutions. *Journal of Evolutionary Biology* **24**, 699-711.

Guo F, Wang G, Innes JL*, et al.* (2016) Comparison of six generalized linear models for occurrence of lightning-induced fires in northern Daxing'an Mountains, China. *Journal of Forestry Research* **27**, 379-388.

Hacking J, Bertozzi T, Moussalli A, Bradford T, Gardner M (2018a) Characterisation of major histocompatibility complex class I transcripts in an Australian dragon lizard. *Developmental & Comparative Immunology* **84**, 164-171.

Hacking JD, Bradford TM, Pierce K, Gardner MG (2018b) *De novo* genotyping of the major histocompatibility complex in an Australian dragon lizard aided by family group data. *Manuscript submitted for publication*.

Hacking JD, Stuart-Fox DM, Gardner MG (In Press) Very low rate of multiple paternity detected in clutches of a wild agamid lizard. *Australian Journal of Zoology*.

Hamilton WD, Zuk M (1982) Heritable true fitness and bright birds: a role for parasites? *Science* **218**, 384-387.

Harvell D (2004) Ecology and Evolution of Host-Pathogen Interactions in Nature. *The American Naturalist* **164**, S1-S5.

Hassrick JL, Henderson MJ, Huff DD*, et al.* (2016) Early ocean distribution of juvenile Chinook salmon in an upwelling ecosystem. *Fisheries Oceanography* **25**, 133-146.

He H, Tang W, Wang W, Crits-Christoph P (2014) Structural zeroes and zero-inflated models. *Shanghai Archives of Psychiatry* **26**, 236-242.

Hinz C, Gebhardt K, Hartmann AK, Sigman L, Gerlach G (2012) Influence of kinship and MHC class II genotype on visual traits in zebrafish larvae (*Danio rerio*). *PLOS ONE* **7**, e51182.

Jackman S (2012) pscl: Classes and Methods for R Developed in the Political Science Computational Laboratory, Stanford University. Department of Political Science, Stanford University, Stanford, California.

Jacob A, Evanno G, Von Siebenthal BA, Grossen C, Wedekind C (2010) Effects of different mating scenarios on embryo viability in brown trout. *Molecular Ecology* **19**, 5296-5307.

Jakob EM, Marshall SD, Uetz GW (1996) Estimating Fitness: A Comparison of Body Condition Indices. *Oikos* **77**, 61-67.

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403-1405.

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics* **11**, 94.

Jones MR, Cheviron ZA, Carling MD (2015) Spatially variable coevolution between a haemosporidian parasite and the MHC of a widely distributed passerine. *Ecology and Evolution* **5**, 1045-1060.

Juola FA, Dearborn DC (2012) Sequence-based evidence for major histocompatibility complex-disassortative mating in a colonial seabird. *Proceedings of the Royal Society B: Biological Sciences* **279**, 153-162.

Kamath PL, Turner WC, Kusters M, Getz WM (2014) Parasite-mediated selection drives an immunogenetic trade-off in plains zebras (Equus quagga). *Proc Biol Sci* **281**, 20140077.

Kleiber C, Zeileis A (2016) Visualizing Count Data Regressions Using Rootograms. *The American Statistician* **70**, 296-303.

Knowles SCL, Nakagawa S, Sheldon BC (2009) Elevated reproductive effort increases blood parasitaemia and decreases immune function in birds: a meta-regression approach. *Functional Ecology* **23**, 405-415.

Kubinak J, Ruff J, Hyzer C, Slev P, Potts W (2012) Experimental viral evolution to specific host MHC genotypes reveals fitness and virulence trade-offs in alternative MHC types. *PNAS* **109**, 3422-3427.

Kuduk K, Babik W, Bellemain E*, et al.* (2014) No Evidence for the Effect of MHC on Male Mating Success in the Brown Bear. *PLOS ONE* **9**, e113414.

Kulski JK, Shiina T, Anzai T, Kohara S, Inoko H (2002) Comparative genomic analysis of the MHC: the evolution of class I duplication blocks, diversity and complexity from shark to man. *Immunological Review* 190.

Landry C, Garant D, Duchesne P, Bernatchez L (2001) 'Good genes as heterozygosity': - the major histocompatibility complex and mate choice in Atlantic salmon (*Salmo salar*). *Proceedings of the Royal Society B: Biological Sciences* 2001, 1279-1285.

Lebas NR (2001) Microsatellite determination of male reproductive success in a natural population of the territorial ornate dragon lizard, *Ctenophorus ornatus*. *Molecular Ecology* 10, 193-203.

Leu ST, Kappeler PM, Bull CM (2010) Refuge sharing network predicts ectoparasite load in a lizard. *Behavioral Ecology and Sociobiology* 64, 1495-1503.

Lewis AC, Rankin KJ, Pask AJ, Stuart-Fox D (2017) Stress-induced changes in color expression mediated by iridophores in a polymorphic lizard. *Ecology and Evolution* 7, 8262-8272.

Mangiafico SS (2015) *An R Companion for the Handbook of Biological Statistics.* http://rcompanion.org/rcompanion/index.html

Martin TG, Wintle BA, Rhodes JR*, et al.* (2005) Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. *Ecology Letters* 8, 1235-1246.

McGraw KJ (2006) Mechanisms of carotenoid-based colouration. In: *Bird colouration: mechanisms and measurements* (eds. Hill GE, McGraw KJ). Harvard University Press, Cambridge (MA).

McLean CA, Lutz A, Rankin KJ, Stuart-Fox D, Moussalli A (2017) Revealing the biochemical and genetic basis of color variation in a polymorphic lizard. *Molecular Biology and Evolution* 34, 1924-1935.

McLean CA, Stuart-Fox D, Moussalli A (2014) Phylogeographic structure, demographic history and morph composition in a colour polymorphic lizard. *Journal of Evolutionary Biology* 27, 2123-2137.

Milinski M (2014) Arms races, ornaments and fragrant genes: The dilemma of mate choice in fishes. *Neuroscience & Biobehavioral Reviews* 46, 567-572.

Milinski M, Griffiths S, Wegner KM*, et al.* (2005) Mate choice decisions of stickleback females predictably modified by MHC peptide ligands. *PNAS* 102, 4414-4418.

Miller HC, Moore JA, Nelson NJ, Daugherty CH (2009) Influence of major histocompatibility complex genotype on mating success in a free-ranging reptile population. *Proceedings of the Royal Society B: Biological Sciences* 276, 1695-1704.

Nagelkerke NJD (1991) A note on a general definition of the coefficient of determination. *Biometrika* 78, 691-692.

Naimi B, Hamm NAS, Groen TA, Skidmore AK, Toxopeus AG (2014) Where is positional uncertainty a problem for species distribution modelling? *Ecography* **37**, 191-203.

Nakagawa S, Schielzeth H (2013) A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution* **4**, 133-142.

Naugler C, Liwski R (2008) An evolutionary approach to major histocompatibility diversity based on allele supertypes. *Medical Hypotheses* **70**, 933-937.

Nosanchuk JD, Casadevall A (2006) Impact of Melanin on Microbial Virulence and Clinical Resistance to Antimicrobial Compounds. *Antimicrobial Agents and Chemotherapy* **50**, 3519-3528.

Oliver MK, Telfer S, Piertney SB (2009) Major histocompatibility complex (MHC) heterozygote superiority to natural multi-parasite infections in the water vole (Arvicola terrestris). *Proceedings of the Royal Society B: Biological Sciences* **276**, 1119-1128.

Olsson M, Madsen T, Nordby J*, et al.* (2003) Major histocompatibility complex and mate choice in sand lizards. *Proceedings of the Royal Society B: Biological Sciences* **2**, S254-256.

Olsson M, Madsen T, Wapstra E*, et al.* (2005) MHC, health, color, and reproductive success in sand lizards. *Behavioral Ecology and Sociobiology* **58**, 289-294.

Osborne L (2005a) Information content of male agonistic displays in the territorial tawny dragon (*Ctenophorus decresii*). *Journal of Ethology* **23**, 189-197.

Osborne L (2005b) Rival recognition in the territorial tawny dragon (*Ctenophorus decresii*). *Acta Ethologica* **8**, 45-50.

Osborne L, Umbers KD, Backwell PR, Keogh JS (2012) Male tawny dragons use throat patterns to recognize rivals. *Naturwissenschaften* **99**, 869-872.

Osborne MJ, Pilger TJ, Lusk JD, Turner TF (2017) Spatio-temporal variation in parasite communities maintains diversity at the major histocompatibility complex class IIβ in the endangered Rio Grande silvery minnow. *Molecular Ecology* **26**, 471-489.

Pearson SK, Godfrey SS, Schwensow N, Bull CM, Gardner MG (2017) Genes and Group Membership Predict Gidgee Skink (*Egernia stokesii*) Reproductive Pairs. *Journal of Heredity* **108**, 369-378.

Piertney S, Oliver M (2006) The evolutionary ecology of the major histocompatibility complex. *Heredity* **96**, 7-21.

R Core Team (2016) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Råberg L (2014) How to Live with the Enemy: Understanding Tolerance to Parasites. *PLoS Biology* **12**, e1001989.

Råberg L, Sim D, Read AF (2007) Disentangling Genetic Variation for Resistance and Tolerance to Infectious Diseases in Animals. *Science* **318**, 812-814.

Radwan J, Kuduk K, Levy E, LeBas N, Babik W (2014) Parasite load and MHC diversity in undisturbed and agriculturally modified habitats of the ornate dragon lizard. *Molecular Ecology* **23**, 5966-5978.

Rankin KJ, McLean CA, Kemp DJ, Stuart-Fox D (2016) The genetic basis of discrete and quantitative colour variation in the polymorphic lizard, *Ctenophorus decresii*. *BMC Evolutionary Biology* **16**, 179.

Rantala MJ, Roff DA (2005) An analysis of trade-offs in immune function, body size and development time in the Mediterranean Field Cricket, Gryllus bimaculatus. *Functional Ecology* **19**, 323-330.

Regoes RR, McLaren PJ, Battegay M*, et al.* (2014) Disentangling Human Tolerance and Resistance Against HIV. *PLoS Biology* **12**, e1001951.

Reusch TBH, Häberli MA, Aeschlimann PB, Milinski M (2001) Female sticklebacks count alleles in a strategy of sexual selection explaining MHC polymorphism. *Nature* **414**.

Reynolds ES, Hart CE, Hermance ME, Brining DL, Thangamani S (2017) An Overview of Animal Models for Arthropod-Borne Viruses. *Comparative Medicine* **67**, 232-241.

Rock KL, Reits E, Neefjes J (2016) Present Yourself! By MHC Class I and MHC Class II Molecules. *Trends in Immunology* **37**, 724-737.

Sandberg M, Eriksson L, Jonsson J, Sjostrom M, Wold S (1998) New chemical descriptors relevant for the design of biologically active peptides. A multivariate characterization of 87 amino acids. *Journal of Medical Chemistry* **41**, 2481-2491.

Savage AE, Zamudio KR (2011) MHC genotypes associate with resistance to a frog-killing fungus. *PNAS* **108**, 16705-16710.

Schad J, Dechmann DK, Voigt CC, Sommer S (2012) Evidence for the 'good genes' model: association of MHC class II DRB alleles with ectoparasitism and reproductive state in the neotropical lesser bulldog bat, *Noctilio albiventris*. *PLOS ONE* **7**, e37101.

Schwensow N, Mazzoni CJ, Marmesat E*, et al.* (2017) High adaptive variability and virus-driven selection on major histocompatibility complex (MHC) genes in invasive wild rabbits in Australia. *Biological Invasions* **19**, 1255-1271.

Sepil I, Lachish S, Hinks AE, Sheldon BC (2013) Mhc supertypes confer both qualitative and quantitative resistance to avian malaria infections in a wild bird population. *Proceedings of the Royal Society B: Biological Sciences* **280**, 20130134.

Sepil I, Radersma R, Santure AW*, et al.* (2015) No evidence for MHC class I-based disassortative mating in a wild population of great tits. *Journal of Evolutionary Biology* **28**, 642-654.

Setchell JM, Vaglio S, Abbott KM*, et al.* (2011) Odour signals major histocompatibility complex genotype in an Old World monkey. *Proceedings of the Royal Society B: Biological Sciences* 278, 274-280.

Smallridge CJ, Bull CM (1999) Transmission of the blood parasite *Hemolivia mariae* between its lizard and tick hosts. *Parasitology Research* 85, 858-863.

Sol D, Jovani R, Torres J (2000) Geographical variation in blood parasites in feral pigeons: the role of vectors. *Ecography* 23, 307-314.

Spurgin LG, Richardson DS (2010) How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proceedings of the Royal Society B: Biological Sciences* 277, 979-988.

Strandh M, Westerdahl H, Pontarp M*, et al.* (2012) Major histocompatibility complex class II compatibility, but not class I, predicts mate choice in a bird with highly developed olfaction. *Proceedings of the Royal Society B: Biological Sciences* 279.

Stuart-Fox DM, Johnston GR (2005) Experience overrides colour in lizard contests. *Behaviour* 142, 329-350.

Sturm T, Leinders-Zufall T, Maček B*, et al.* (2013) Mouse urinary peptides provide a molecular basis for genotype discrimination by nasal sensory neurons. *Nature Communications* 4, 1616.

Symonds MRE, Moussalli A (2011) A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. *Behavioral Ecology and Sociobiology* 65, 13-21.

Takahata N, Nei M (1990) Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. *Genetics* 124.

Teasdale L, Stevens M, Stuart-Fox D (2013) Discrete colour polymorphism in the tawny dragon lizard (*Ctenophorus decresii*) and differences in signal conspicuousness among morphs. *Journal of Evolutionary Biology* 26.

Trachtenberg E, Korber B, Sollars C*, et al.* (2003) Advantage of rare HLA supertype in HIV disease progression. *Nature Medicine* 9, 928.

Wang J (2007) Triadic IBD coefficients and applications to estimating pairwise relatedness. *Genetic Resources* 89, 135-153.

Wegner K, Kalbe M, Kurtz J, Reusch TB, Milinski M (2003) Parasite selection for immunogenetic optimality. *Science* 301.

Welsh AH, Cunningham RB, Donnelly CF, Lindenmayer DB (1996) Modelling the abundance of rare species: statistical models for counts with extra zeros. *Ecological Modelling* 88, 297-308.

Westerdahl H, Asghar M, Hasselquist D, Bensch S (2011) Quantitative disease resistance: to better understand parasite-mediated selection on major histocompatibility complex. *Proceedings of the Royal Society B: Biological Sciences* 279, 577-584.

Whitcomb AC, Banks MA, O'Malley KG (2014) Influence of immune-relevant genes on mate choice and reproductive success in wild-spawning hatchery-reared and wild-born coho salmon (Oncorhynchus kisutch). *Canadian Journal of Fisheries and Aquatic Sciences* 71, 1000-1009.

Wikel SK (1996) Host Immunity to Ticks. *Annual Review of Entomology* 41, 1-22.

Winternitz JC, Minchey SG, Garamszegi LZ*, et al.* (2013) Sexual selection explains more functional variation in the mammalian major histocompatibility complex than parasitism. *Proceedings of the Royal Society B: Biological Sciences* 280.

Xu L, Paterson AD, Turpin W, Xu W (2015) Assessment and Selection of Competing Models for Zero-Inflated Microbiome Data. *PLOS ONE* 10, e0129606.

Yewers MSC (2016) *The function and evolution of colour polymorphism in the tawny dragon lizard* PhD, The University of Melbourne.

Yewers MSC, Jessop TS, Stuart-Fox D (2017) Endocrine differences among colour morphs in a lizard with alternative behavioural strategies. *Hormones and Behavior* 93, 118-127.

Yewers MSC, Pryke S, Stuart-Fox D (2016) Behavioural differences across contexts may indicate morph-specific strategies in the lizard *Ctenophorus decresii*. *Animal Behaviour* 111, 329-339.

Zeh JA, Zeh DW (1996) The Evolution of Polyandry I: Intragenomic Conflict and Genetic Incompatibility. *Proceedings of the Royal Society B: Biological Sciences* 263, 1711-1717.

Zeileis A, Kleiber C, Jackman S (2008) Regression Models for Count Data in R. *Journal of Statistical Software* 27, 25.

Zuur AF, Ieno EN, Elphick CS (2010) A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution* 1, 3-14.

Tables and figures

Table 1. AIC information-theoretic top model selection results for model set 1 (response: tick load, predictors: MHC I supertypes), after model averaging. Only those models with ΔAICc ≤2 are shown due to the large number of models in the 95% confidence set. See table S3 and figure S10 for summary results for predictor variables and figure S7 for overall model fit.

Model	df	AICc	ΔAICc	Weight	ER
ST4	5	860.35	0.00	0.18	
ST4 + ST3	7	861.33	0.98	0.11	1.6
ST4 + ST7	7	861.71	1.36	0.09	1.2

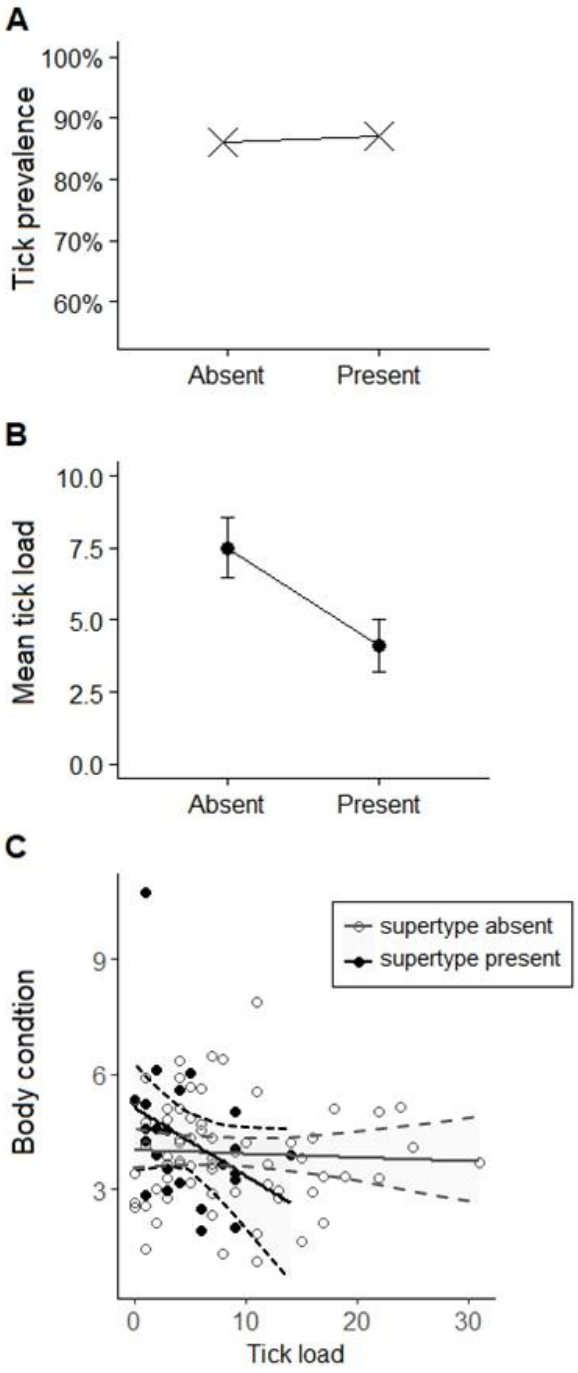


Figure 1. Relationship between the presence and absence of supertype four and tick prevalence (A) and mean (±SE) tick load (B), and the relationship between tick load and body condition (showing 95% confidence intervals with shaded area and dotted lines), comparing individuals with and without supertype four (C).

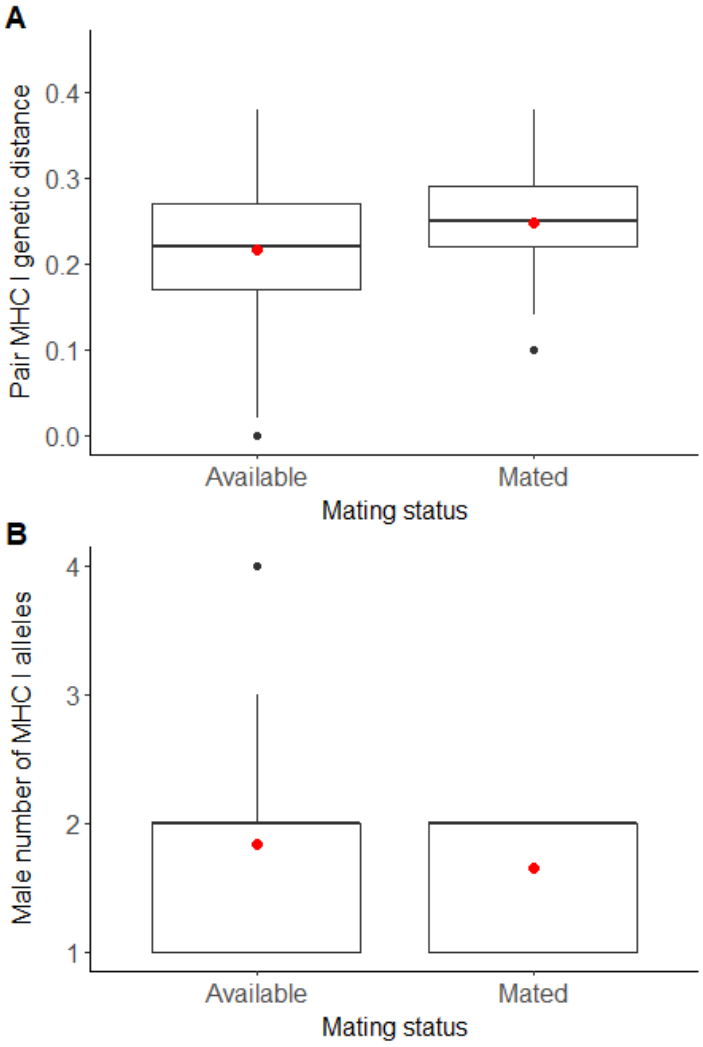


Figure 2. Difference between available and mated *Ctenophorus decresii* males in average pair percent difference among shared MHC I alleles (A) and male number of MHC I alleles (B). Mean values are indicated with a red point.

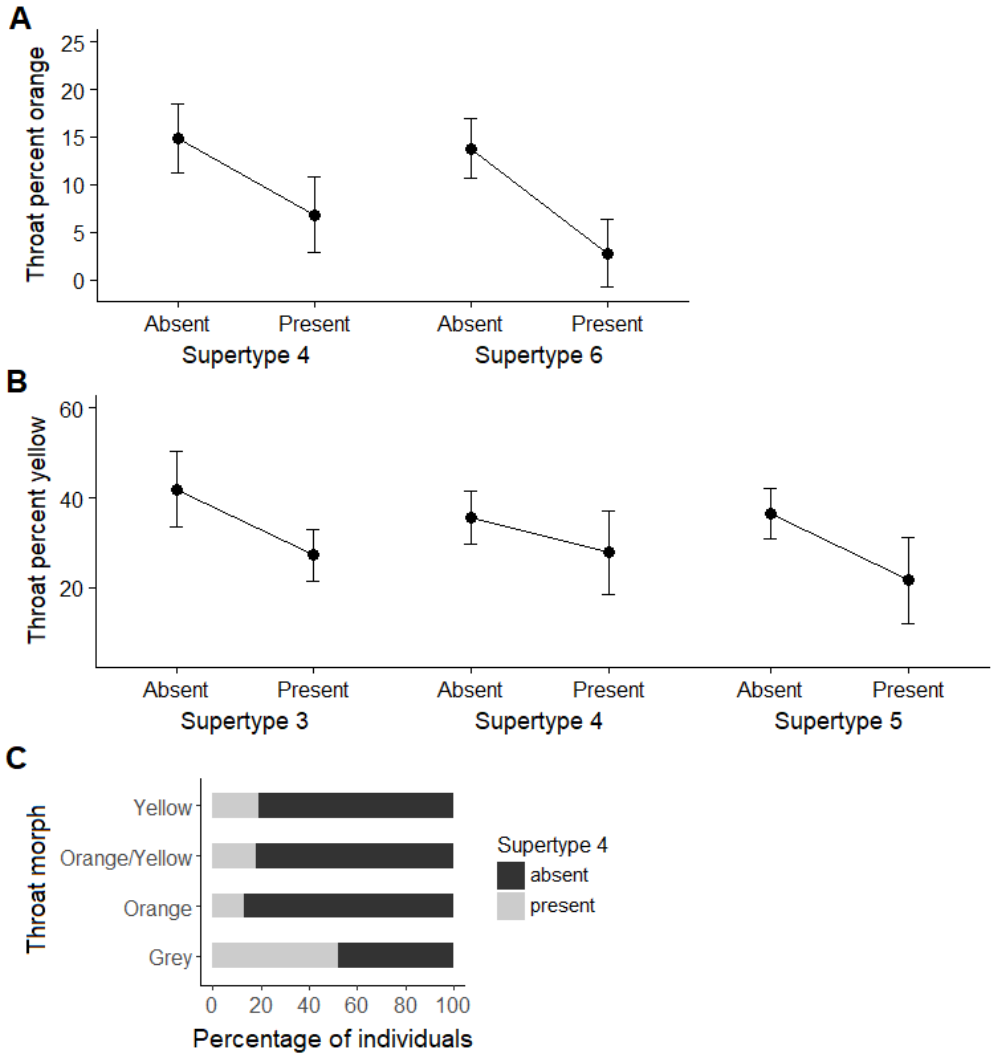


Figure 3. Mean (±SE) percentage of male throat colour (A: orange, B: yellow) as a function of supertype presence (only supertypes that were present in models with ΔAICc ≤2 are displayed) and the percentage of individuals of each male throat morph that possess supertype four (C).

Supplementary Information

Specific MHC class I supertype associated with parasite infection and colour morph in a wild lizard population

Hacking, J., Stuart-Fox, D., Godfrey, S., Gardner, M.

Text S1. Additional methodology for male throat colouration analysis

Measures of colouration are based on the visual system of agamid lizards (i.e. perception by conspecifics) and are independent of the human visual system. The proportion of yellow and orange throat colouration was derived from calibrated photographs by extracting portions of the photograph based on RGB values of each pixel according to set threshold values (Teasdale *et al.* 2013). For each male, photographs were taken at a standard distance and angle with a Canon© PowerShot SX50 HScamera. An X-Rite© ColourChecker Classic Cardcolour card was included in each photograph. Photographs were taken when males were at an active temperature (32°C ± 3°C). Each photograph was calibrated to adjust for differences in illumination among photographs using the RGBGys and reflectance values of the grey scale on the colour card as described in Teasdale *et al.* (2013). After calibration, photographs were cropped to include only the throat and the proportion of orange and yellow were calculated using threshold values of 0.35 and 0.20 for orange and yellow, respectively. Analysis was undertaken in Matlab (The MathWorks, Inc., Natick, MA, USA) using a custom program (Teasdale *et al.* 2013).

A portable Ocean Optics (Dunedin, FL, USA) Jaz spectrometer was used to collect two or three spectral reflectance measurements for each primary and secondary throat colour, holding the probe at a 45° angle to the throat. Measurements were taken every 1 nm within the visible spectrum of agamid lizards (300 to 700 nm) and were expressed relative to a white Spectralon (Labsphere, Inc., North Sutton, NH, USA) 99% diffuse reflectance standard. Males were at an active temperature (33°C ± 3°C) when measurements were taken. Reflectance measures (300 – 700 nm) were smoothed over 5 nm intervals using the Nadaraya–Watson kernel regression estimate implemented in R ver. 3.4.1 (R Core Team 2016) prior to analysis (Fig. S3). After the relative stimulation of *C. decresii* photoreceptors (receptor quantum catches) were estimated (Vorobyev *et al.* 1998), we calculated the chromatic (colour) and achromatic (brightness) contrasts relative to a standard background (reflectance of 1%) under full sun irradiance (fig. S3). Chromatic and achromatic contrast was measured in units of 'just noticeable differences', with a just noticeable difference greater than three generally indicating a noticeable difference (here, from the standard background) from the perspective of the receiver

(Vorobyev & Osorio 1998). Single cone sensitivities were used to calculate chromatic contrast and double cone sensitivity was used to calculate achromatic contrast (Yewers *et al.* 2015). Refer to (Teasdale *et al.* 2013) for further details on the visual models used here. Only the results for achromatic JNDs ('brightness') were used in subsequent statistical analyses. All calculations were undertaken in R, using the *pavo* package (Maia *et al.* 2013; R Core Team 2016).

To quantify chest patch size, photographs (as described above) were set to scale using the ruler included in each photograph and then the area of black/grey chest patch marking was calculated, using ImageJ ver. 1.46r (Schindelin *et al.* 2012). Relative chest patch size was calculated using residuals from a correlation of chest patch size and SVL.

Text S2. Model set 3a – 3f results (male phenotypic signals of MHC genotype)

The 95% confidence set for model set 3a, in which number of MHC I alleles was the response variable, included nine models (tables S10 and S11). The top model (lowest ΔAICc) was the null model. All models had low ΔAICc and evidence ratios. Adjusted R^2 values were very low for all models, indicating that the variables had little predictive power. Model set 3b, in which number of MHC I supertypes was the response variable, contained nine top models (95% confidence set, tables S12 and S13). All models within the 95% confidence set included percentage of male throat coloured yellow (PercY) and the top model included only PercY. PercY was highest in individuals with one supertype and lowest in individuals with three supertypes. All top models had low ΔAICc and ERs, hence there was no clear 'best' model. All models were poor fits, with adjusted R^2 reaching a maximum of 0.18 among the top model set.

Model sets 3c and 3d uncovered an effect of the presence of certain supertypes on percentage of male throat coloured orange (PercO) and percentage of male throat coloured yellow (PercY), although these effects weren't statistically significant. Rootograms showed that hurdle models fit well in both cases, although they suffered from slight underestimation (figs. S8 and S9). The 95% confidence set for model set 3c (PercO as response) included four models that had a ΔAICc less than two (tables S6 and S7, fig. S12). These models included supertypes four and six, with PercO lower in the presence of both supertypes. This association was driven by the zero (binomial) component of the hurdle model, revealing that the presence of supertypes four and six was associated with the absence of orange colour on the throat rather than the amount of orange colour on the throat. Although supertypes four and six were present in the top models and the hurdle model fitted the data well, 95% CIs included zero for both supertypes and effect sizes were small. In regard to model set 3d (PercY as response), the 95% confidence set included five models for which ΔAICc was less than two (tables S8 and S9, fig. S13). These models included supertypes three, four and five. Again, these associations were driven by the zero (binomial) component of the hurdle model, indicating that the presence or absence of yellow on the throat, rather than the amount of yellow on the throat, is associated with certain supertypes. Supertypes three, four and five all displayed a negative relationship with PercY. Supertype five had the largest effect size, although overall effect sizes were small, and 95% CIs overlapped zero for all supertype.

Model sets 3e (throat brightness as response) and 3f (chest patch as response) revealed little evidence that throat brightness and chest patch are associated with certain MHC I supertypes. In model set 3e the null model had the lowest ΔAICc and model fit was poor (tables S14 and S15). Model fit was also poor for model set 3f, although supertypes four and six had lower ΔAICc than the null model (tables S16 and S17).The 95% CIs overlapped zero for all variables in each model set.

Table S1. Summary of statistical tests used to test hypotheses regarding the mechanisms shaping MHC class I diversity within a wild *C. decresii* population

Model set	Response	Predictors	OPV*
		Parasite-mediated selection	
1, hurdle model	Tick load	Supertypes: ST2 – ST8	21
		MHC-associated mating	
2, GLMM	Mating status	Pair MHC I genetic distance, number MHC alleles (male); *co-variates:* pair spatial proximity, male mass; *random factors:* male ID, female ID	4
Fisher's exact test	Mating status	Prevalence of supertypes ST2 – ST8	
		Signals of MHC genotype	
3a, GLM	Number MHC alleles	Percentage of throat coloured orange, percentage of throat coloured yellow, throat brightness, relative chest patch size	14
3b, GLM	Number MHC supertypes		
3c, hurdle model	Percentage of throat coloured orange	Supertypes ST2 – ST8	12
3d, hurdle model	Percentage of throat coloured yellow		12
			10
3e, GLM	Throat brightness		
3f, GLM	Relative chest patch size		12
3g, GLM	Mating status	Percentage of throat coloured orange, percentage of throat coloured yellow; *co-variate:* male mass	5
Fisher's exact test	Supertype ST4	Morph types	
Fisher's exact test	Mating status	Morph types	

*OPV: Observations per variable: The number of observations divided by the number of predictor variables (including co-variates). For binary response variables the number of observations is equal to the smaller of i) the number of events (i.e. mated) or ii) the number non-events (i.e. available).

Table S2. Description of variables used in hypothesis testing

Variable	Abbreviation	Description
Mating status	——	Each male either mated with, or was available to (within 100m radius), a given female
Pair spatial proximity	Geo_dist	Shortest geographic distance (m) between a male and female
Mass	——	Male mass (g)
Body condition	——	
Number MHC alleles	Al	The number of MHC I alleles per individual. 'Male Al' refers to number of male MHC I alleles.
Number MHC supertypes	ST	The number of MHC I supertypes per individual
Supertypes	——	Each supertype (ST2 to ST8) is a separate variable, binary coded for absence or presence
Pair MHC I genetic distance	Al_dist	The average percent genetic (amino acid) distance between shared MHC I alleles of a male and female
Percent orange on throat	PercO	Percentage of male throat coloured orange
Percent yellow on throat	PercY	Percentage of male throat coloured yellow
Throat brightness	TB	Achromatic JNDs of male throat colour against a standard background, a measure of brightness
Chest patch size	CP	Male chest patch size relative to SVL
Tick load	TL	The number of ticks attached to an individual

Table S3. Summary results for predictor variables included in model set 1 top (95% confidence set) models (response: tick load, predictors: MHC I supertypes), after model averaging. See main text for model results. Bolded and italicised CI values do not overlap zero. See table S2 for variable descriptions and abbreviations.

Variable	Estimate	Adjusted SE	Lower CI	Upper CI	Relative importance
Count_ST4	-0.66	0.28	***-1.21***	***-0.11***	0.93
Zero_ST4	0.09	0.59	-1.07	1.25	0.93
Count_ST3	0.10	0.18	-0.25	0.46	0.33
Zero_ST3	0.09	0.31	-0.52	0.69	0.33
Count_ST7	-0.24	0.41	-1.04	0.56	0.34
Zero_ST7	-0.22	0.58	-1.36	0.92	0.34
Count_ST2	-0.06	0.15	-0.35	0.23	0.22
Zero_ST2	-0.08	0.29	-0.65	0.50	0.22
Count_ST5	-0.03	0.11	-0.25	0.19	0.14
Zero_ST5	0.00	0.22	-0.42	0.42	0.14
Count_ST6	-0.04	0.13	-0.30	0.23	0.13
Zero_ST6	-0.02	0.26	-0.52	0.48	0.13
Count_ST8	0.02	0.09	-0.17	0.20	0.10
Zero_ST8	-0.01	0.21	-0.42	0.40	0.10

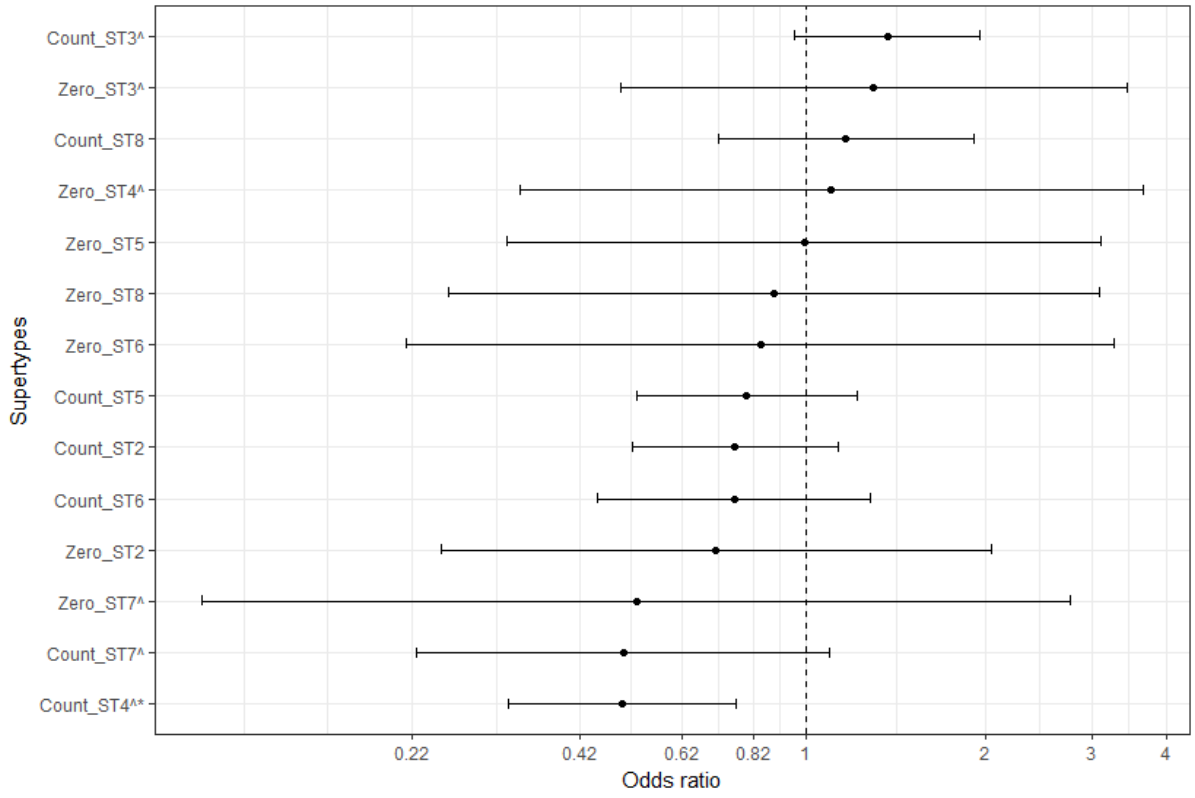


Figure S10. Estimates of effect size (odds ratios with 95% CI) of each *C. decresii* MHC I supertype on tick load (model set 1, top models averaged). Hurdle model with tick load as the response. Zero refers to the binomial (logit link) part of the hurdle model and the count refers to the truncated negative binomial (log link) part of the model. Odds ratios great than 1 indicate a positive association and odds ratios less than 1 indicate a negative association. ^supertype within high ranking models (ΔAICc ≤2). *95% CI does not overlap zero.

Table S4 AIC information-theoretic top (95% confidence set) model selection results for model set 2 (response: male mating status, predictors: MHC I diversity and MHC I genetic distance), after model averaging. The null model includes only co-variables (male mass and pair spatial proximity). See table S2 for variable descriptions and abbreviations.

Model	df	AICc	ΔAICc	Weight	ER	R^2_m	R^2_c
Al_dist	6	110.08	0.00	0.48		0.002	0.025
Null	5	110.74	0.66	0.34	1.4	0.002	0.021
Al_dist + Male_Al	7	111.99	1.91	0.18	1.9	0.488	0.599

Table S5. Summary results for predictor variables included in model set 2 top (95% confidence set) models (response: male mating status, predictors: MHC I diversity and MHC I genetic distance), after model averaging. Effect sizes standardised on two S.D. See main text for model results. See table S2 for variable descriptions and abbreviations.

Variable	Estimate	Adjusted SE	Lower CI	Upper CI	Relative importance
Al_dist	0.75	0.81	-0.85	2.35	0.66
Male_Al	-0.06	0.38	-0.80	0.68	0.18
Male mass (co-variable)	2.32	0.84	0.66	3.97	1.00
Geo_dist (co-variable)	-3.04	0.77	-4.54	-1.54	1.00

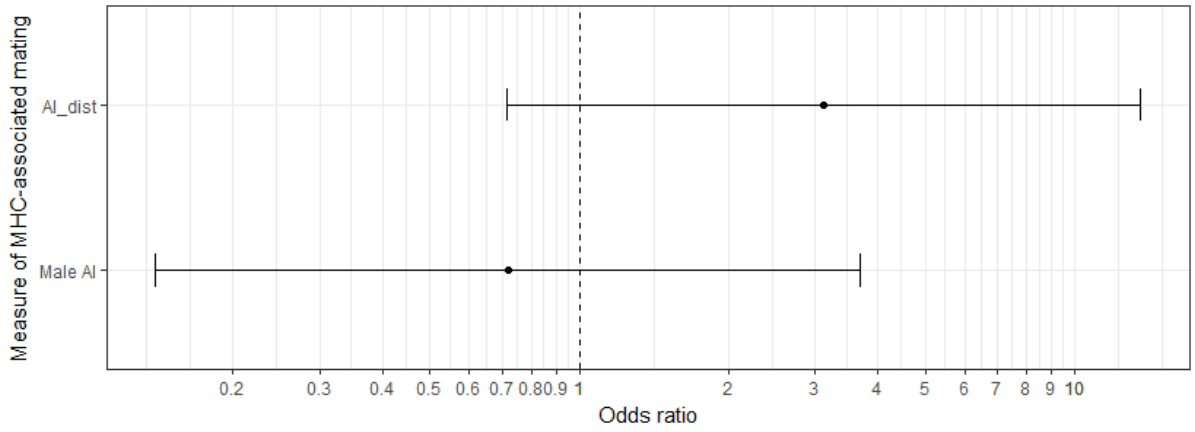


Figure S11. Estimates of effect size (odds ratios with 95% CI) of each *C. decresii* MHC I supertype on mating status (model set 2, top models averaged). Odds ratios great than 1 indicate a positive association and odds ratios less than 1 indicate a negative association. Both variables were within high ranking models (ΔAICc ≤2).

Table S6. AIC information-theoretic top model selection results for model set 3c (response: percentage of male throat coloured orange, predictors: MHC I supertypes), after model averaging. Only those models with ΔAICc ≤2 are shown due to the large number of models in the 95% confidence set. See table S2 for variable descriptions and abbreviations.

Model	df	AICc	ΔAICc	Weight	ER
ST4 + ST6	7	447.02	0.00	0.19	
ST6	5	447.19	0.17	0.17	1.1
Null	3	447.84	0.82	0.12	1.4
ST4	5	448.58	1.57	0.09	1.3

Table S7. Summary results for predictor variables included in model set 3c top (95% confidence set) models (response: male percentage throat coloured orange, predictors: MHC I supertypes), after model averaging. See main text for model results. See table S2 for variable descriptions and abbreviations.

Variable	Estimate	Adjusted SE	Lower CI	Upper CI	Relative importance
Count_ST4	-0.13	0.25	-0.83	0.48	0.46
Zero_ST4	-0.45	0.61	-1.73	0.72	0.46
Count_ST6	-0.15	0.58	-1.80	0.93	0.61
Zero_ST6	-1.28	1.33	-3.08	1.08	0.61
Count_ST5	0.03	0.12	-0.40	0.57	0.12
Zero_ST5	0.02	0.20	-0.62	0.49	0.12
Count_ST2	-0.01	0.09	-0.24	0.23	0.08
Zero_ST2	-0.01	0.16	-0.45	0.38	0.08
Count_ST3	-0.01	0.08	-0.18	0.19	0.08
Zero_ST3	0.01	0.14	-0.31	0.27	0.08
Count_ST7	0.02	0.15	-0.45	0.52	0.07
Zero_ST7	0.02	0.30	-0.60	0.58	0.07
Count_ST8	0.01	0.09	-0.20	0.23	0.09
Zero_ST8	0.01	0.17	-0.30	0.33	0.09

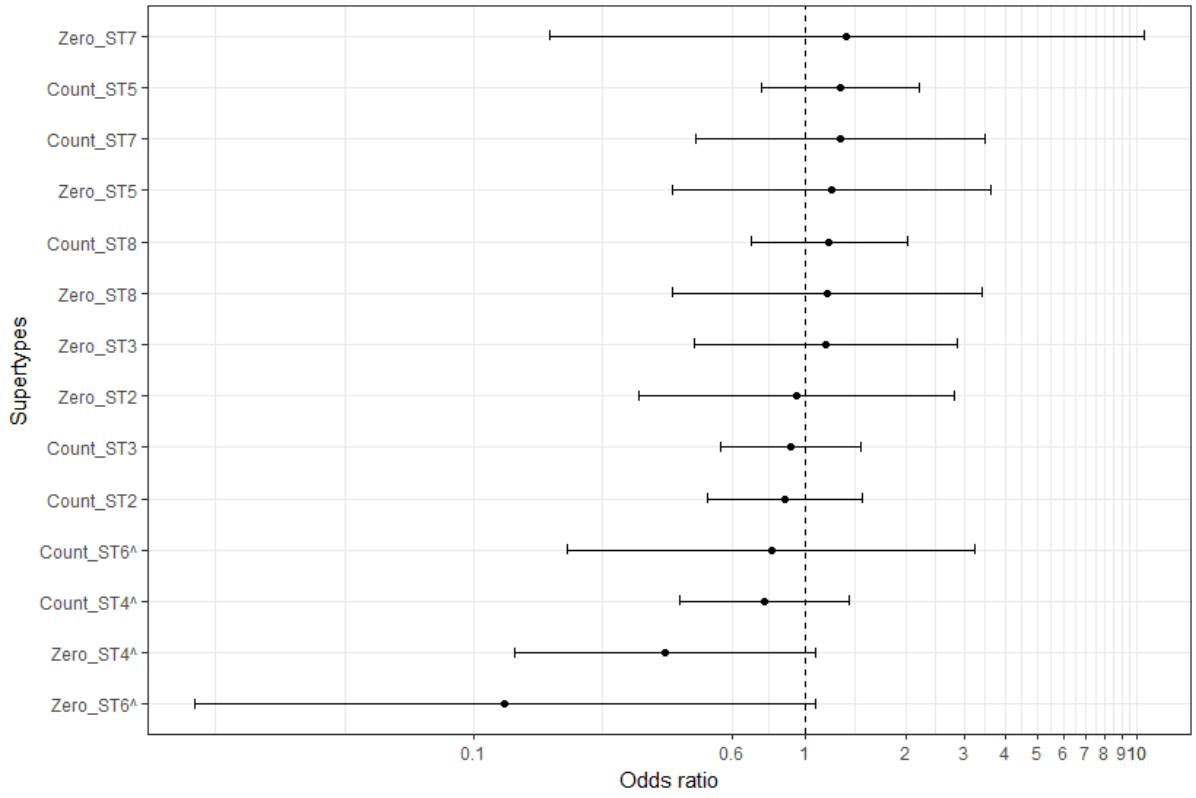


Figure S12. Estimates of effect size (odds ratios with 95% CI) of each *C. decresii* MHC I supertype on percentage of male throat coloured orange (model set 3c, top models averaged). Hurdle model with male percentage of throat coloured orange as the response. Zero refers to the binomial (logit link) part of the hurdle model and the count refers to the truncated negative binomial (log link) part of the model. Odds ratios great than 1 indicate a positive association and odds ratios less than 1 indicate a negative association. ^supertype within high ranking models (ΔAICc ≤2).

Table S8. AIC information-theoretic top model selection results for model set 3d (response: percentage of male throat coloured yellow, predictors: MHC I supertypes), after model averaging. Only those models with ΔAICc ≤2 are shown due to the large number of models in the 95% confidence set. See table S2 for variable descriptions and abbreviations.

Model	df	AICc	ΔAICc	Weight	ER
ST3 + ST5	7	624.56	0.00	0.12	
ST3	5	624.60	0.04	0.12	1.0
Null	3	624.92	0.36	0.10	1.2
ST3 + ST5 + ST4	9	625.42	0.86	0.08	1.3
ST5	5	626.10	1.55	0.06	1.3

Table S9. Summary results for predictor variables included in model set 3d top (95% confidence set) models (response: male percentage throat coloured yellow, predictors: MHC I supertypes), after model averaging. See main text for model results. See table S2 for variable descriptions and abbreviations.

Variable	Estimate	Adjusted SE	Lower CI	Upper CI	Relative importance
Count_ST3	-0.17	0.18	-0.51	0.18	0.60
Zero_ST3	-0.36	0.51	-1.35	0.63	0.60
Count_ST5	-0.06	0.15	-0.36	0.24	0.48
Zero_ST5	-0.59	0.76	-2.08	0.90	0.48
Count_ST4	0.00	0.09	-0.17	0.17	0.26
Zero_ST4	-0.23	0.48	-1.17	0.70	0.26
Count_ST2	-0.02	0.07	-0.16	0.13	0.15
Zero_ST2	0.08	0.34	-0.57	0.74	0.15
Count_ST8	0.01	0.07	-0.13	0.14	0.12
Zero_ST8	0.07	0.29	-0.50	0.64	0.12
Count_ST6	-0.01	0.07	-0.15	0.14	0.10
Zero_ST6	0.07	0.35	-0.61	0.75	0.10
Count_ST7	0.00	0.10	-0.20	0.19	0.07
Zero_ST7	-0.04	0.31	-0.64	0.56	0.07

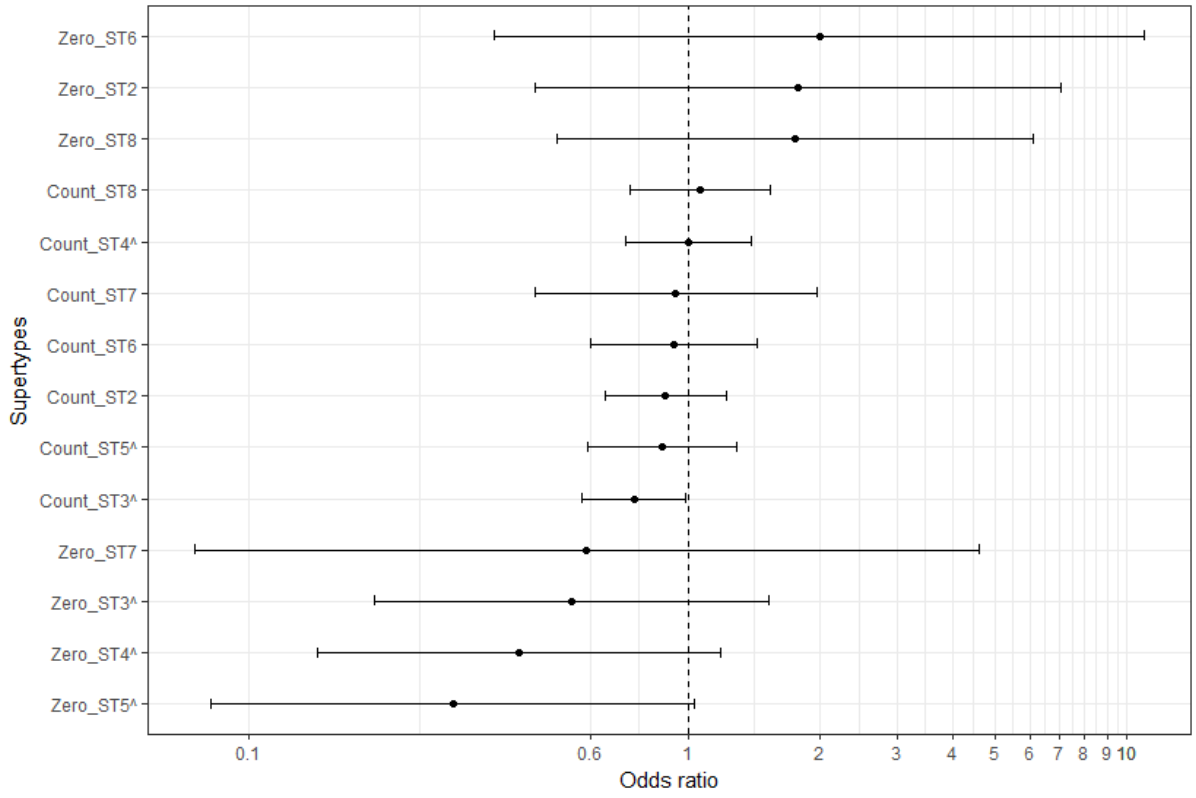


Figure S13. Estimates of effect size (odds ratios with 95% CI) of each *C. decresii* MHC I supertype on percentage of male throat coloured yellow (model set 3d, top models averaged). Hurdle model with male percentage of throat coloured yellow as the response. Zero refers to the binomial (logit link) part of the hurdle model and the count refers to the truncated negative binomial (log link) part of the model. Odds ratios great than 1 indicate a positive association and odds ratios less than 1 indicate a negative association. ^supertype within high ranking models (ΔAICc ≤2).

Table S10. AIC information-theoretic top (95% confidence set) model selection results for model set 3a (response: number MHC I alleles, predictors: potential signals of MHC diversity), after model averaging. See table S2 for variable descriptions and abbreviations.

Model	df	AICc	ΔAICc	Weight	ER	Adj-R^2
Null	2	131.94	0.00	0.23		0.00
PercY	3	132.69	0.76	0.16	1.5	0.03
CP	3	133.63	1.70	0.10	1.6	0.01
TB	3	134.02	2.08	0.08	1.2	0.00
PercO	3	134.15	2.21	0.08	1.1	0.00
PercY + CP	4	134.58	2.65	0.06	1.2	0.04
PercY + TB	4	134.65	2.71	0.06	1.0	0.04
PercY + PercO	4	134.77	2.83	0.06	1.1	0.03
CP + TB	4	135.73	3.80	0.03	1.6	0.01

Table S11. Summary results for predictor variables included in model set 3a top (95% confidence set) models (response: number MHC I alleles, predictors: potential signals of MHC diversity), after model averaging. Effect sizes standardised on two S.D. See table S2 for variable descriptions and abbreviations.

Variable	Estimate	Adjusted SE	Lower CI	Upper CI	Relative importance
PercY	-0.10	0.18	-0.44	0.25	0.39
CP	0.03	0.11	-0.19	0.26	0.23
TB	0.02	0.10	-0.18	0.22	0.20
PercO	-0.01	0.09	-0.18	0.16	0.15

Table S12. AIC information-theoretic top (95% confidence set) model selection results for model set 3b (response: number of MHC I supertypes, predictors: potential signals of MHC diversity), after model averaging. See table S2 for variable descriptions and abbreviations.

Model	df	AICc	ΔAICc	Weight	ER	Adj-R^2
PercY	3	121.25	0.00	0.24		0.12
PercY + TB	4	121.64	0.39	0.19	1.2	0.15
PercY + CP	4	122.63	1.38	0.12	1.6	0.14
PercY + TB + CP	5	122.81	1.56	0.11	1.1	0.17
PercY + PercO	4	122.98	1.73	0.10	1.1	0.13
PercY + TB + PercO	5	123.79	2.54	0.07	1.5	0.16
PercY + CP + PercO	5	124.57	3.33	0.04	1.5	0.14
PercY + TB + CP + PercO	6	125.17	3.92	0.03	1.3	0.18
Null	2	125.32	4.07	0.03	1.1	0.00

Table S13. Summary results for predictor variables included in model set 3b top (95% confidence set) models (response: number of MHC I supertypes, predictors: potential signals of MHC diversity), after model averaging. Effect sizes standardised on two S.D. See table S2 for variable descriptions and abbreviations.

Variable	Estimate	Adjusted SE	Lower CI	Upper CI	Relative importance
PercY	-0.46	0.20	-0.86	-0.06	0.97
TB	0.11	0.17	-0.23	0.45	0.43
CP	0.06	0.14	-0.21	0.32	0.33
PercO	-0.03	0.11	-0.25	0.19	0.26

Table S14. AIC information-theoretic top (95% confidence set) model selection results for model set 3e (response: throat brightness, predictors: supertypes), after model averaging. Only models with a ΔAICc ≤2 are shown due to the high number of models in the 95% confidence set. See table S2 for variable descriptions and abbreviations.

Model	df	AICc	ΔAICc	Weight	ER	Adj-R^2
Null	2	468.64	0.00	0.07		0.00
ST8	3	469.63	0.99	0.04	1.8	0.02
ST2	3	469.90	1.27	0.04	1.0	0.01
ST3	3	469.92	1.28	0.04	1.0	0.01
ST7	3	470.27	1.63	0.03	1.3	0.01
ST5	3	470.40	1.76	0.03	1.0	0.01
ST2 + ST5	4	470.55	1.92	0.03	1.0	0.04
ST4	3	470.60	1.96	0.03	1.0	0.00

Table S15. Summary results for predictor variables included in model set 3e top (95% confidence set) models (response: throat brightness, predictors: supertypes), after model averaging. Effect sizes standardised on two S.D. See table S2 for variable descriptions and abbreviations.

Variable	Estimate	Adjusted SE	Lower CI	Upper CI	Relative importance
ST8	0.92	1.89	-2.78	4.62	0.36
ST2	1.13	1.85	-2.49	4.75	0.44
ST3	-0.43	1.16	-2.70	1.85	0.30
ST7	0.92	2.36	-3.72	5.55	0.30
ST5	-0.76	1.72	-4.14	2.61	0.34
ST4	0.34	1.13	-1.87	2.55	0.27
ST6	0.15	1.42	-2.62	2.93	0.22

Table S16. AIC information-theoretic top (95% confidence set) model selection results for model set 3f (response: chest patch size, predictors: supertypes), after model averaging. Only models with a ΔAICc ≤2 are shown due to the high number of models in the 95% confidence set. See table S2 for variable descriptions and abbreviations.

Model	df	AICc	ΔAICc	Weight	ER	Adj-R^2
ST4	3	197.91	0.00	0.07		0.04
ST4 + ST6	4	198.79	0.87	0.04	1.8	0.06
Null	2	198.88	0.97	0.04	1.0	0.00
ST4 + ST8	4	199.25	1.34	0.03	1.3	0.05
ST4 + ST3	4	199.41	1.50	0.03	1.0	0.05
ST4 + ST2	4	199.46	1.55	0.03	1.0	0.05
ST2	3	199.50	1.58	0.03	1.0	0.02

Table S17. Summary results for predictor variables included in model set 3f top (95% confidence set) models (response: chest patch size, predictors: supertypes), after model averaging. Effect sizes standardised on two S.D. See table S2 for variable descriptions and abbreviations.

Variable	Estimate	Adjusted SE	Lower CI	Upper CI	Relative importance
ST4	0.20	0.23	-0.26	0.66	0.58
ST6	0.10	0.24	-0.36	0.57	0.33
ST8	0.04	0.14	-0.24	0.32	0.27
ST3	-0.05	0.13	-0.31	0.21	0.31
ST2	-0.07	0.16	-0.38	0.24	0.33
ST5	0.00	0.12	-0.24	0.24	0.23
ST7	0.01	0.20	-0.38	0.39	0.22

Table S18. AIC information-theoretic top (95% confidence set) model selection results for model set 3g (response: mating status, predictors: male percentage throat coloured orange and male percentage throat coloured yellow), after model averaging. The null model includes only co-variates (male mass). See table S2 for variable descriptions and abbreviations.

Model	df	AICc	ΔAICc	Weight	ER	Adj-R^2
Null	2	88.85	0.00	0.55		0.12
PercY	3	90.37	1.52	0.26	2.1	0.13
PercO	3	90.96	2.11	0.19	1.4	0.12

Table S19. Summary results for predictor variables included in model set 3g top (95% confidence set) models (response: mating status, predictors: male percentage throat coloured orange and male percentage throat coloured yellow), after model averaging. Effect sizes standardised on two S.D. See table S2 for variable descriptions and abbreviations.

Variable	Estimate	Adjusted SE	Lower CI	Upper CI	Relative importance
PercY	-0.11	0.36	-0.82	0.59	0.26
PercO	0.01	0.26	-0.50	0.51	0.19
Male mass (co-variable)	1.66	0.65	0.40	2.93	1.00

Figure S1. Example of tick (*Amblyomma limbatum*) infection on adult male *Ctenophorus decresii.*

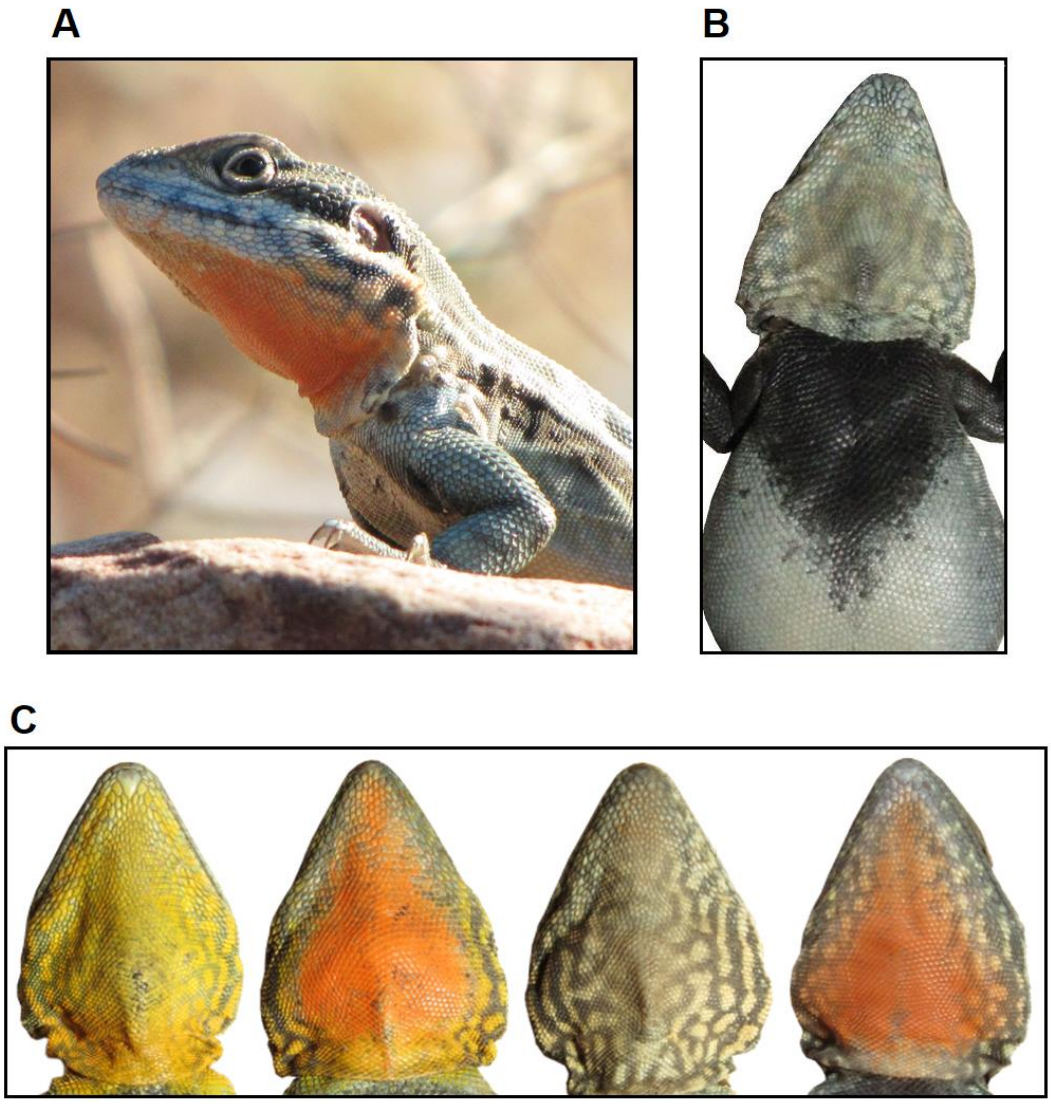


Figure S2. A male tawny dragon lizard (*Ctenophorus decresii*) perched on a rock at the Hawker field site in the Flinders Ranges, South Australia (A). An example of an adult male chest patch marking (B), and the three throat colour morphs present within the Hawker population, ordered left to right from most frequent to least frequent (C).

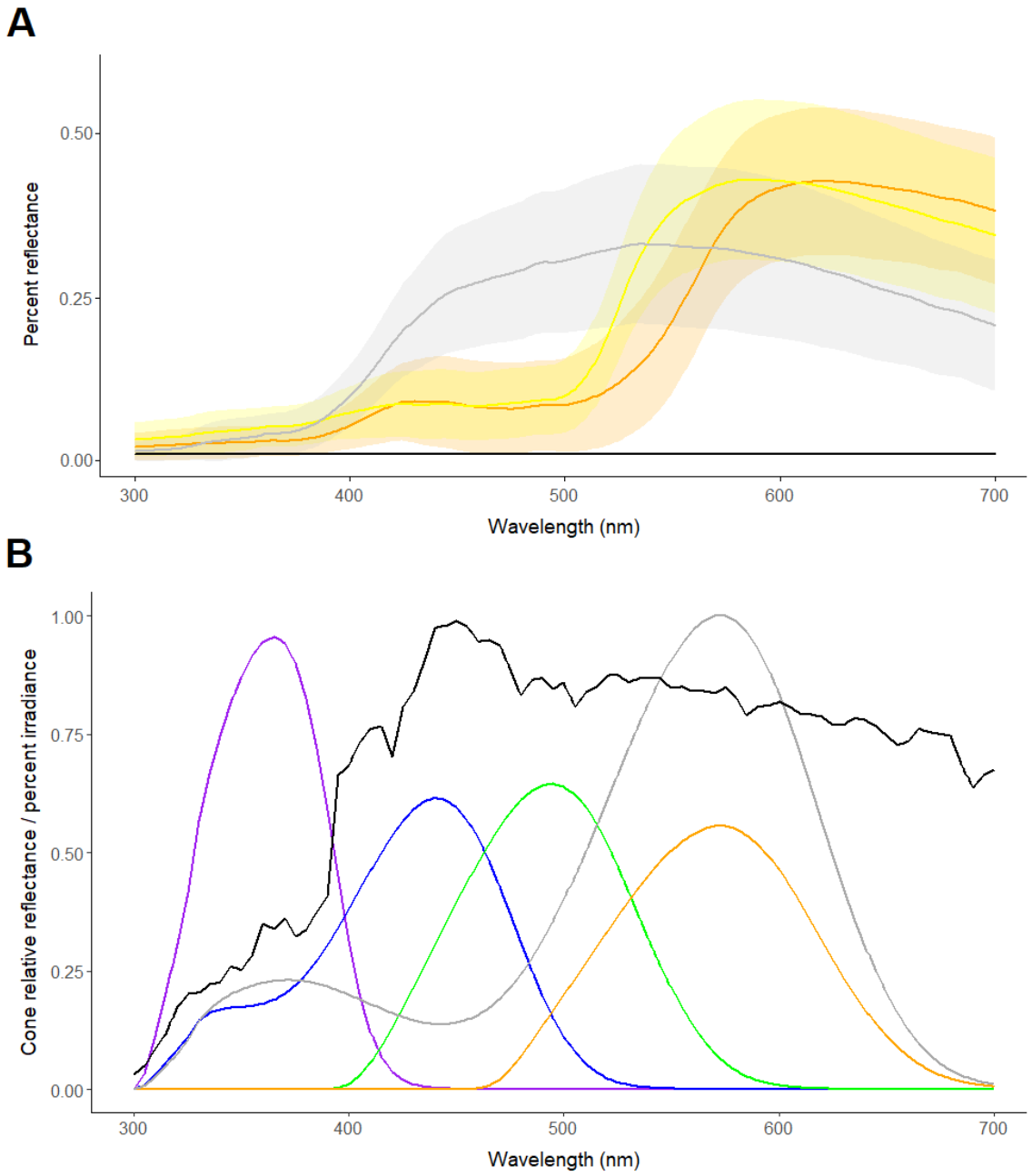


Figure S3. Average (standard deviation) percent spectral reflectance for each male *Ctenophorus decresii* throat colour (yellow, orange and grey) and the standard black background used in analyses of throat colouration (A). Spectral sensitivities of *Ctenophorus decresii* single (UV: purple, short wavelengths: blue, medium wavelengths: green, long wavelengths: orange) and double (grey) cones, and the full sun irradiance data used in analyses of throat colouration (black) (B).

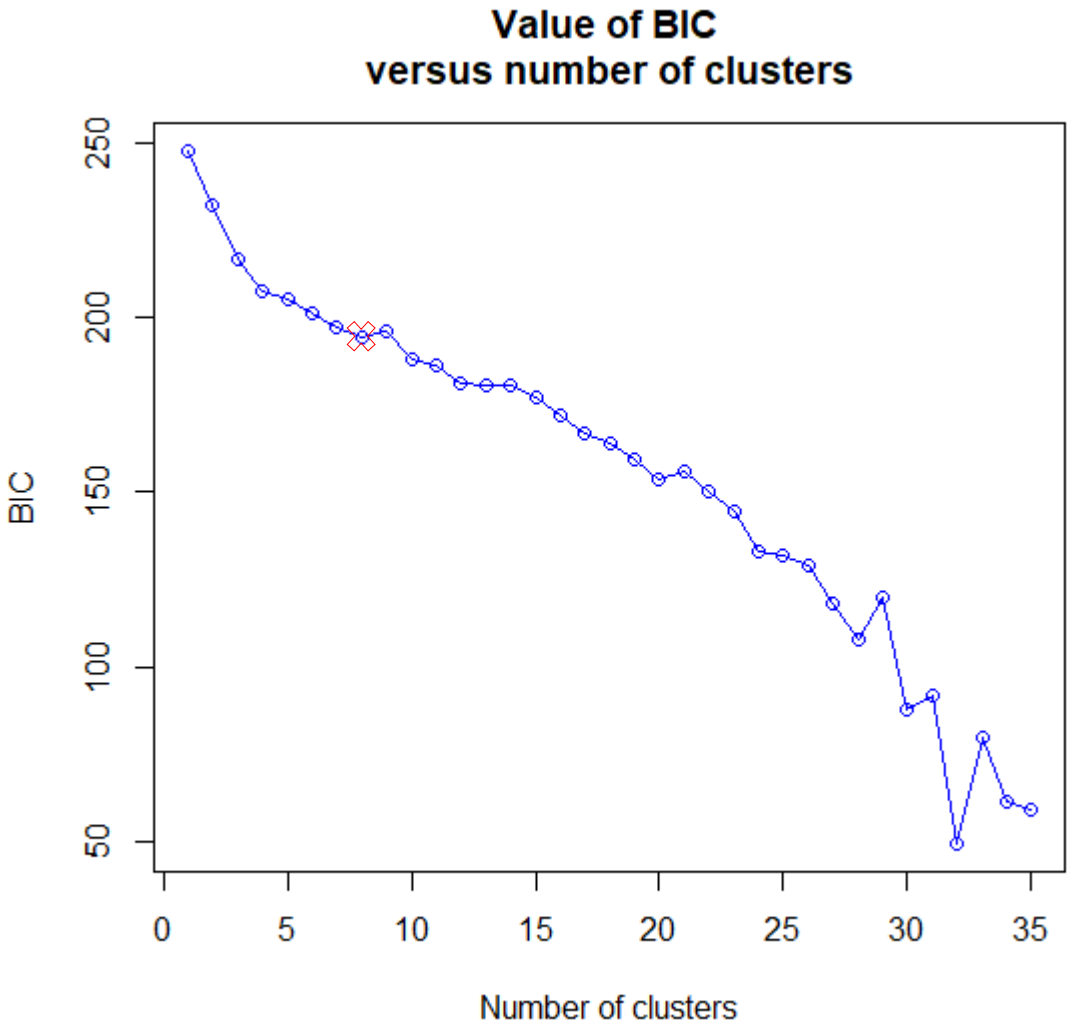


Figure S4. Value of the Bayesian Information Criterion (BIC) for different models (different numbers of clusters, K). The maximum K was set to 35 (above this K there were more cluster centers than distinct data points). A red cross indicates the optimal K (K = 8), which was chosen as after this K the BIC increased (Jombart *et al.* 2010).

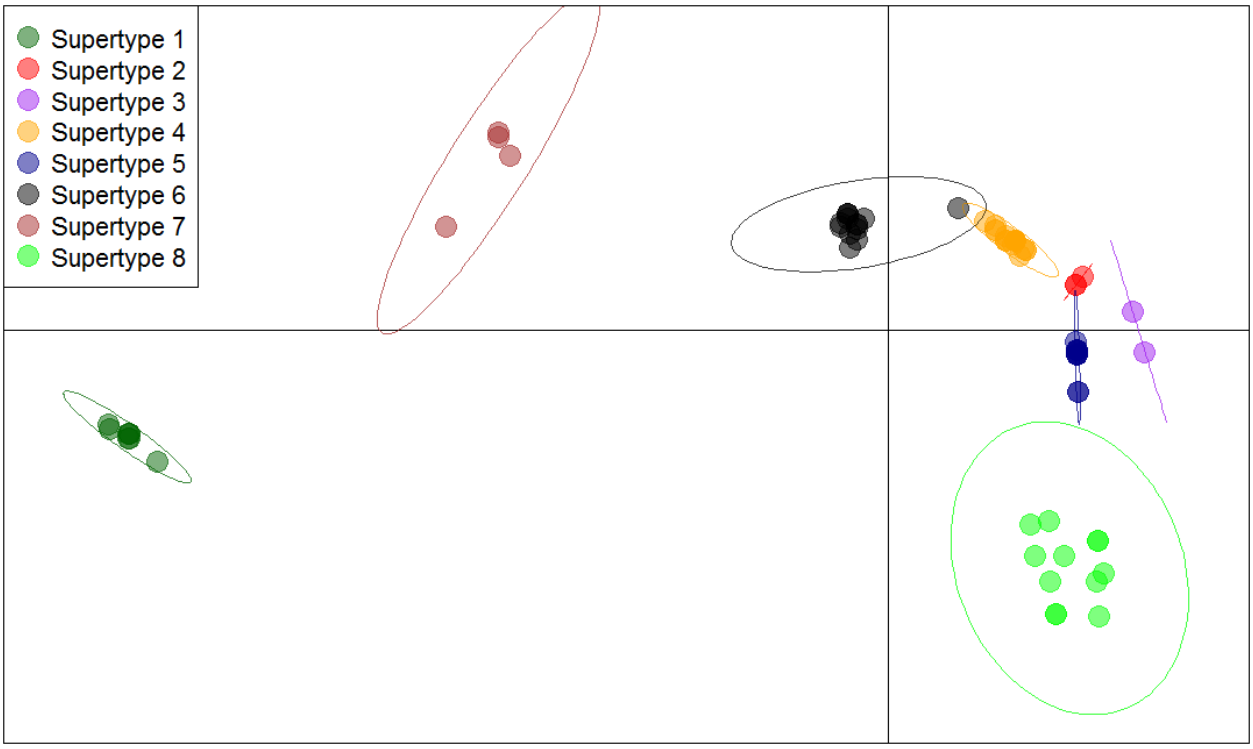


Figure S5. Visualisation of MHC I supertype analysis using k-means clustering. Each point on the scatterplot represents an allele and groups are indicated by inertia ellipses.

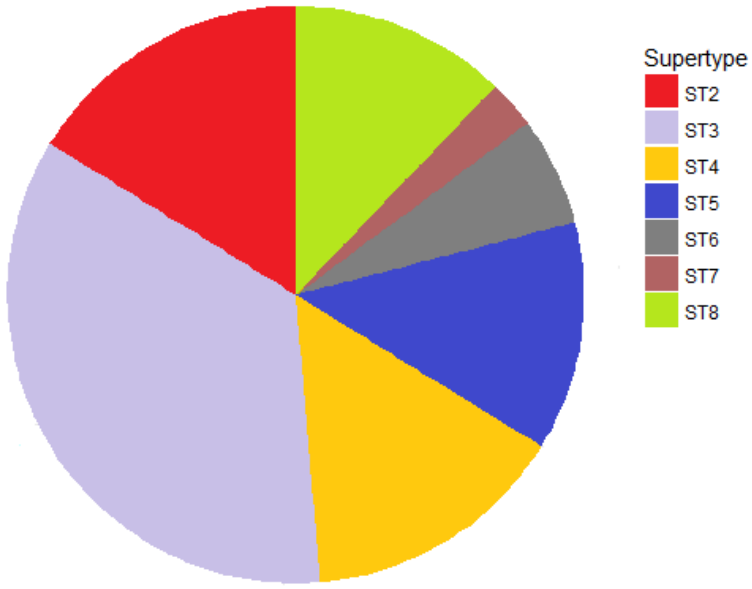


Figure S5. Frequency of each MHC I supertype within the Hawker population

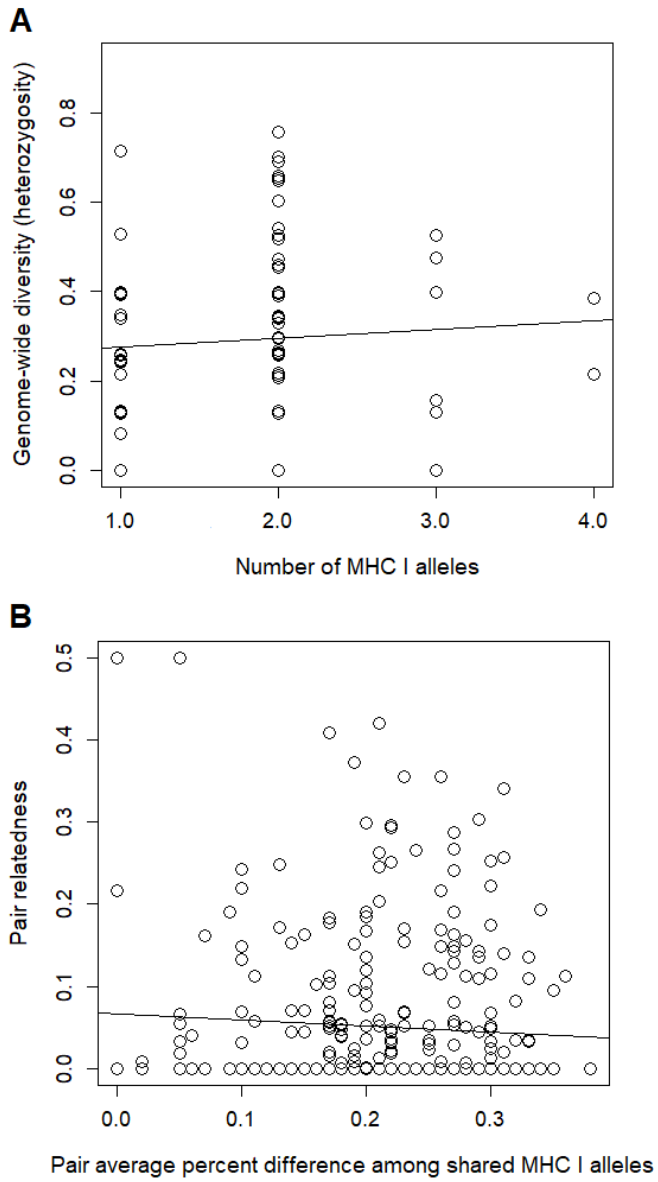


Figure S6. The relationship between genome-wide diversity (microsatellite heterozygosity) and the number of MHC I alleles, Al (A) and the relationship between pair relatedness (estimated from microsatellites) and MHC I allele percent distance between pairs, Al_dist (B).

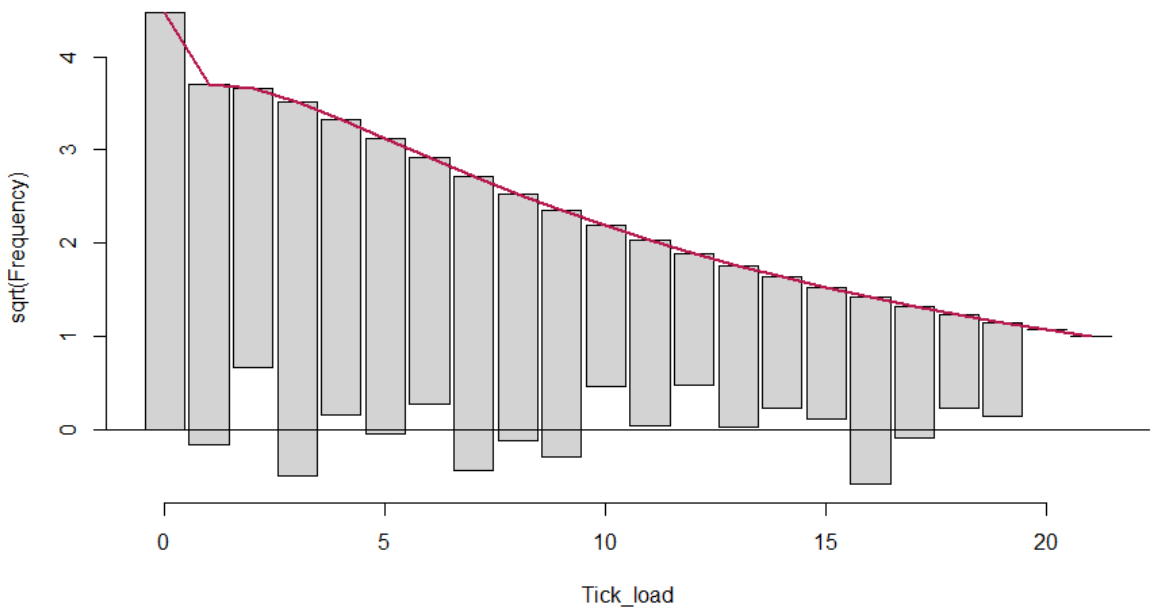


Figure S7. Hanging rootogram for model set 1 used as a visualisation of model fit for hurdle models. The red line indicates expected counts and observed counts are represented by grey bars, which are 'hanging' from the expected counts. The y-axis is the square root of the observed or expected counts, allowing deviations from expectations to be observed at small frequencies. Perfect estimation is indicated by grey bars sitting on the zero reference line. Grey bars above zero represent over-estimation and grey bars below zero represent under-estimation.

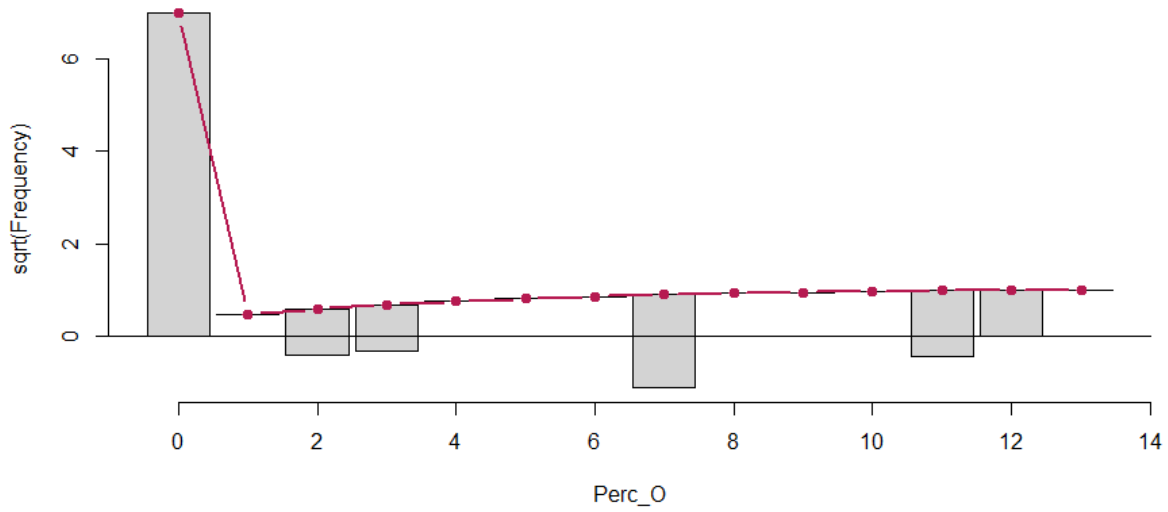


Figure S8. Hanging rootogram for model set 3c used as a visualisation of model fit for hurdle models. The red line indicates expected counts and observed counts are represented by grey bars, which are 'hanging' from the expected counts. The y-axis is the square root of the observed or expected counts, allowing deviations from expectations to be observed at small frequencies. Perfect estimation is indicated by grey bars sitting on the zero reference line. Grey bars above zero represent over-estimation and grey bars below zero represent under-estimation.

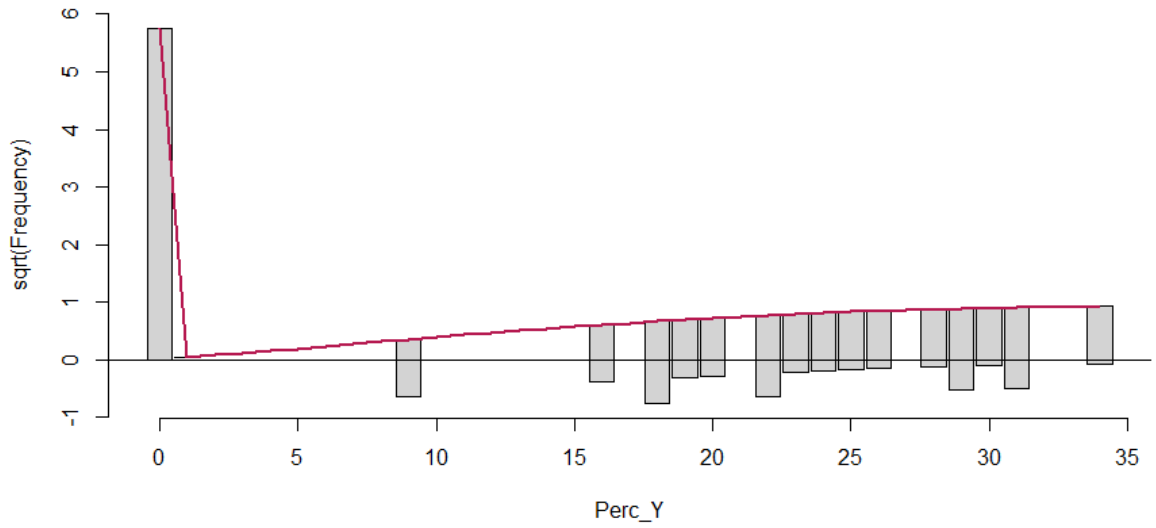


Figure S9. Hanging rootogram for model set 3d used as a visualisation of model fit for hurdle models. The red line indicates expected counts and observed counts are represented by grey bars, which are 'hanging' from the expected counts. The y-axis is the square root of the observed or expected counts, allowing deviations from expectations to be observed at small frequencies. Perfect estimation is indicated by grey bars sitting on the zero reference line. Grey bars above zero represent over-estimation and grey bars below zero represent under-estimation.

References

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics* **11**, 94.

Maia R, Eliason CM, Bitton P-P, Doucet SM, Shawkey MD (2013) pavo: an R package for the analysis, visualization and organization of spectral data. *Methods in Ecology and Evolution* **4**, 906-913.

R Core Team (2016) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Schindelin J, Arganda-Carreras I, Frise E*, et al.* (2012) Fiji: an open-source platform for biological-image analysis. *Nature Methods* **9**, 676-682.

Teasdale L, Stevens M, Stuart-Fox D (2013) Discrete colour polymorphism in the tawny dragon lizard (*Ctenophorus decresii*) and differences in signal conspicuousness among morphs. *Journal of Evolutionary Biology* **26**.

Vorobyev M, Osorio D (1998) Receptor noise as a determinant of colour thresholds. *Proceedings of the Royal Society B: Biological Sciences* **265**, 351-358.

Vorobyev M, Osorio D, Bennett ATD, Marshall NJ, Cuthill IC (1998) Tetrachromacy, oil droplets and bird plumage colours. *Journal of Comparative Physiology A* **183**, 621-633.

Yewers MS, McLean CA, Moussalli A*, et al.* (2015) Spectral sensitivity of cone photoreceptors and opsin expression in two colour-divergent lineages of the lizard *Ctenophorus decresii*. *The Journal of Experimental Biology* **218**, 1556-1563.

CHAPTER 6

General discussion

Characterisation of major histocompatibility complex (MHC) genes in non-avian reptiles has lagged behind other taxonomic groups. The work presented in my thesis contributes to filling this gap by identifying mechanisms generating and maintaining diversity at the MHC in an Australian agamid lizard, the tawny dragon (*Ctenophorus decresii*). Furthermore, when obtaining family group data, I discovered that *C. decresii* employs a predominantly polygynous mating system and has one of the lowest rates of multiple paternity among squamates, contributing to knowledge on the general biology of the species. Here, I discuss the results presented in this thesis and suggest several avenues for future work.

Tawny dragon MHC class I diversity

Transcriptome data was obtained using next-generation sequencing and an iterative assembly strategy successfully identified six full-length MHC class I transcripts (chapter 2). These transcripts possessed all of the exons and conserved sites expected of classical MHC class I genes (chapter 2), were associated with high allelic diversity and possessed several putative peptide binding regions that were under positive and/or episodic selection (chapter 4). Hence, these transcripts are likely to encode classical class I MHC molecules. In addition, two partial transcripts were identified and are likely the result of high similarity among transcripts impeding complete transcriptome assembly. However, it is also possible that these transcripts represent naturally truncated non-classical MHC class I molecules that are secreted and undertake peptide presentation independently from the cell surface (chapter 2). New technologies such as PacBio single-molecule real-time sequencing are likely to eliminate issues surrounding transcriptome assembly by sequencing entire transcripts (Westbrook et al., 2015).

Primers were designed to amplify MHC class I exons involved in peptide binding and were used in next-generation amplicon sequencing of 485 individuals across five populations (chapter 4). A thorough genotyping protocol produced reliable genotypes for the MHC class I α1 domain (chapter 4). This genotyping protocol involved the use of a recently developed clustering method (Sebastian et al., 2016) and was aided by family group data produced in chapter 3, allowing in-depth analysis of both type I and

type II genotyping error rates. The nature of genotyping error (type I or type II) is often not considered in MHC studies despite the fact that type I and II errors have different consequences for down-steam analyses. Family group data also allowed the analysis of MHC allele segregation patterns. Allele segregation patterns and the observed variation in the number of alleles possessed by individuals within populations suggested that either allele dropout occurred or copy number variation (CNV, variation in the number of loci among individuals) exists within the MHC class I region of *C. decresii*. Often, studies examining MHC diversity in non-model species assume CNV despite in most cases the equally likely possibility of partial or complete allele dropout.

MHC genotyping revealed moderate MHC class I allelic diversity within *C. decresii* compared to other reptiles. However, direct comparisons are difficult given that it is difficult to determine the number and type of MHC loci amplified in non-model taxa. MHC class I alleles were clustered by function, revealing eight 'supertypes' (chapter 5). Within the Hawker population most individuals with more than one MHC allele had an equal number of supertypes, highlighting the importance of within-individual functional diversity. A high level of MHC class I allele sharing was observed among populations and alleles did not form clades associated with geographic regions. Future work examining population structure in terms of MHC class I alleles and supertypes will reveal levels of range-wide MHC diversity.

Factors shaping diversity at tawny dragon MHC class I genes

Phylogenetic analysis of Iguania MHC class I sequences in chapter 2 revealed that the birth and death model of evolution likely plays a major role in generating and shaping MHC diversity in agamid lizards, with concerted evolution potentially playing a secondary role. In chapter 5 the mechanisms maintaining diversity at the within-population level were examined. Evidence for parasite-mediated selection was discovered, with ectoparasite load significantly lower in the presence of a specific MHC supertype (supertype four). Such a relationship suggests the MHC diversity may be maintained via parasite-mediated rare allele advantage or fluctuating selection in *C. decresii*. Patterns observed in relation to MHC-associated mating supported both the good genes hypothesis (absolute criteria) and the complementary genes hypothesis (self-referential criteria). Larger sample sizes are required to confirm these small effects. There was little evidence to suggest that male colouration acts as a signal of male MHC genotype. Hence, we were not able to confirm mating based on honest signalling of MHC-based parasite resistance as predicted by the Hamilton-Zuk hypothesis. However, we uncovered a link between MHC I supertype four and morph type, with grey morphs more likely to possess this supertype. The nature of this link is unknown and warrants further investigation. Examining the patterns of MHC

class I diversity among *C. decresii* populations will allow the identification of mechanisms shaping MHC diversity at larger, range-wide scales. It is important to note that only a subset of MHC loci were considered in this thesis. Investigation of other MHC class I and class II loci may provide further insights into the mechanisms shaping MHC diversity in *C. decresii*.

Future directions and conclusions

I was successful in identifying mechanisms generating and maintaining MHC diversity within an agamid lizard. This work broadens knowledge on squamate MHC genes and the evolutionary ecology of the MHC class I. I identified several avenues for future work throughout the thesis. The most intriguing of these are related to copy number variation, ectoparasites or vector-borne parasites associated with a MHC class I supertype, the biochemical and genetic pathways responsible for colour-MHC associations, the mechanisms acting to maintain MHC diversity among populations, and the extremely low rate of multiple paternity detected within the Hawker population.

For some species CNV provides an alternative way of distributing MHC diversity (De Groot et al., 2015). For instance, an Old World monkey, *Macaca mulatta*, has a similar level of within-individual diversity to humans but this diversity is distributed among several genes within individuals with low allelic diversity, rather than a few genes with high allelic diversity, as is the case for humans (De Groot et al., 2015). Does CNV exist within the *C. decresii* MHC class I region? More broadly, what are the factors responsible for some species possessing CNV at the MHC and others not? Is CNV the ancestral or the derived state? Examining CNV within *C. decresii* will require examination of complete MHC I haplotypes across several individuals. Illumina long read sequencing technology would provide such data (Yin et al., 2016).

Work presented in chapter 5 presents the first case of an association between a MHC supertype and parasite load in a non-avian reptile. Is the observed association due to response to the tick (*Amblyomma limbatum*) itself or a tick-borne parasite? It is known that *A. limbatum* is a vector for haemogregarine blood parasites (Smallridge and Bull, 1999), although it is unknown whether this species also transmits other intercellular parasites such as bacteria and viruses. Infection experiments and sequencing technologies such as Duel-Seq, which simultaneously sequences the transcriptome of host and parasite, will help to answer these questions (Avital et al., 2017).

The biochemical pathway responsible for the link between the MHC and odour signalling was suggested 30 years ago (Brown et al., 1987, Singh et al., 1987). Self peptides presented by MHC molecules reflect differences in an individual's MHC repertoire as different MHC molecules bind to different peptides. Later, work in mice showed that it is the self peptides presented by MHC molecules

that are detected via the olfactory vomeronasal gland in rodents (Leinders-Zufall et al., 2004). Given that several studies have reported associations between colouration and the MHC (Dunn et al., 2013, Hinz et al., 2012, Jacob et al., 2010, Setchell et al., 2009), research should move towards an understanding of the underlying mechanisms relating to MHC-colour associations. Completion of the *C. decresii* genome is likely to provide insights into potential links between melanin colouration and the MHC in agamid lizards.

Within this thesis I examined the mechanisms influencing the MHC at the within-population level. However, the factors shaping the MHC on the larger range-wide scale remains to be investigated. Does genetic drift or selection play a dominant role? By contrasting patterns of neutral diversity with MHC diversity, it is possible to separate the effects of selection from drift and make inferences about the type of selection (e.g. balancing or fluctuating) that has taken place (Piertney and Oliver, 2006).

In conclusion, this work significantly furthers knowledge on the mechanisms generating and maintaining MHC diversity in a non-avian reptile and addressed broad evolutionary ecology questions, such as the classic Hamilton-Zuk hypothesis (Hamilton and Zuk, 1982). Furthermore, the research presented here has uncovered several pathways for further investigation, both in relation to *C. decresii* and the MHC in general.

References

AVITAL, G., AVRAHAM, R., FAN, A., HASHIMSHONY, T., HUNG, D. T. & YANAI, I. 2017. scDual-Seq: mapping the gene regulatory program of Salmonella infection by host and pathogen single-cell RNA-sequencing. *Genome Biology,* 18**,** 200.

BROWN, R. E., SINGH, P. B. & ROSER, B. 1987. The Major Histocompatibility Complex and the chemosensory recognition of individuality in rats. *Physiology & Behavior,* 40**,** 65-73.

DE GROOT, N., BLOKHUIS, J. H., OTTING, N., DOXIADIS, G. G. & BONTROP, R. E. 2015. Co-evolution of the MHC class I and KIR gene families in rhesus macaques: ancestry and plasticity. *Immunological Reviews,* 267**,** 228–245.

DUNN, P. O., BOLLMER, J. L., FREEMAN-GALLANT, C. R. & WHITTINGHAM, L. A. 2013. MHC variation is related to a sexually selected ornament, survival, and parasite resistance in common yellowthroats. *Evolution,* 67**,** 679-87.

HAMILTON, W. D. & ZUK, M. 1982. Heritable true fitness and bright birds: a role for parasites? *Science,* 218**,** 384-7.

HINZ, C., GEBHARDT, K., HARTMANN, A. K., SIGMAN, L. & GERLACH, G. 2012. Influence of kinship and MHC class II genotype on visual traits in zebrafish larvae (*Danio rerio*). *PLOS ONE,* 7**,** e51182.

JACOB, A., EVANNO, G., VON SIEBENTHAL, B. A., GROSSEN, C. & WEDEKIND, C. 2010. Effects of different mating scenarios on embryo viability in brown trout. *Molecular Ecology,* 19**,** 5296-5307.

LEINDERS-ZUFALL, T., BRENNAN, P., WIDMAYER, P., S., P. C., MAUL-PAVICIC, A., JÄGER, M., LI, X.-H., BREER, H., ZUFALL, F. & BOEHM, T. 2004. MHC class I peptides as chemosensory signals in the vomeronasal organ. *Science,* 306**,** 1033-1037.

PIERTNEY, S. & OLIVER, M. 2006. The evolutionary ecology of the major histocompatibility complex. *Heredity,* 96**,** 7-21.

SEBASTIAN, A., HERDEGEN, M., MIGALSKA, M. & RADWAN, J. 2016. AMPLISAS: a web server for multilocus genotyping using next-generation amplicon sequencing data. *Molecular Ecology,* 16**,** 498–510.

SETCHELL, J. M., CHARPENTIER, M. J. E., ABBOTT, K. M., WICKINGS, E. J. & KNAPP, L. A. 2009. Is brightest best? Testing the Hamilton-Zuk hypothesis in mandrills. *International Journal of Primatology,* 30**,** 825-844.

SINGH, P. B., BROWN, R. E. & ROSER, B. 1987. MHC antigens in urine as olfactory recognition cues. *Nature,* 327**,** 161.

SMALLRIDGE, C. J. & BULL, C. M. 1999. Transmission of the blood parasite *Hemolivia mariae* between its lizard and tick hosts. *Parasitology Research,* 85**,** 858-63.

WESTBROOK, C. J., KARL, J. A., WISEMAN, R. W., MATE, S., KOROLEVA, G., GARCIA, K., SANCHEZ-LOCKHART, M., O'CONNOR, D. H. & PALACIOS, G. 2015. No assembly required: Full-length MHC class I allele discovery by PacBio circular consensus sequencing. *Human Immunology,* 76**,** 891-6.

YIN, Y., LAN, J. H., NGUYEN, D., VALENZUELA, N., TAKEMURA, P., BOLON, Y. T., SPRINGER, B., SAITO, K., ZHENG, Y., HAGUE, T., PASZTOR, A., HORVATH, G., RIGO, K., REED, E. F. & ZHANG, Q. 2016. Application of high-throughput next-generation sequencing for HLA typing on buccal extracted DNA: Results from over 10,000 donor recruitment samples. *PLOS ONE,* 11.

Appendix 1

This appendix is associated with chapters two and four and is present only in the thesis (not included in the publications associated with chapters two and four). Appendix 1 provides additional information regarding MHC class I transcriptome assembly and primer design.

MHC class I primer design

Methods

Transcriptome assembly

Due to the complex nature of the MHC region, standard transcriptome assembly techniques may not be suitable (method is described in Chapter 2). It was suspected that the original transcriptome assembly was erroneous due to difficulties during primer design, especially in the case of MHC class II (MHC class II characterisation was subsequently abandoned). We therefore employed a method similar to Ansari et al. (2015) to gain a more robust assembly of the MHC class I region (method is described in Chapter 2). This assembly is henceforth referred to as the revised assembly and was completed in 2017.

Primer design

Primers were designed using the original transcriptome sequences (original transcripts) as the revised assembly method was undertaken after primer designed and testing was finalised (next-generation sequencing was already performed). When designing primers, a neighbour-joining tree was constructed using MEGA ver. 5.2 (Tamura et al., 2011) to designate transcript groups, with the thought that similar transcripts likely represent loci, or closely related loci where groups contain more than two transcripts.

Initially, primers were designed to amplify the intron regions surrounding exon 2 (α1) and 3 (α2), so that primers amplifying complete exon sequences from flanking intron regions could be designed. Primers were trialled on genomic DNA obtained from the individual originally used for transcriptome sequencing (SAMR67384). These primers were unsuccessful in amplifying intron regions despite trailing several different PCR enzymes and PCR protocols, including long-range methods. This failure is likely

due to the original assembly not being correct. Another possible cause is that the intron regions between exons were too long for the long-range PCR methods employed (Miller et al., 2015).

Following failed attempts to design primers within intron sequence flanking exons, multiple primers were designed at the 5' and 3' ends of exons 2 (α1) and 3 (α2) using Primer 3 ver. 2.3.4 (Untergasser et al., 2012) in Geneious ver. 8.1.7 (Kearse et al., 2012). In an attempt to target loci separately, unique primers were designed for each group (as per the neighbour-joining tree). In some cases unique primers could not be designed for each group due to similarity at desired priming sites. Primers were therefore designed for multiple groups where necessary.

Results

Transcriptome assembly and primer design

The original assembly produced a total of 14 unique full-length MHC class I transcripts, which clustered into six groups consisting of one to four sequences each (fig. 1). Amino acid alignment of these transcripts with published MHC class I sequences from several vertebrate taxa revealed concordance with conserved regions and regions with predicted functions (Kaufman et al., 1994). These sequences were therefore considered likely to encode MHC class I molecules that probably undertake classical functions. However, as noted above this transcriptome assembly was erroneous.

Some of the original transcripts were similar to the transcripts resulting from the revised assembly for α1 (fig. 2). Sequences from groups 2, 3 and 4 (fig. 1) were all identical at α1 and matched α1 of the revised transcriptome sequences *Ctde-UA*001* and *Ctde-UA*003*. Consequently, the primers designed to amplify these groups (MHC1_A1_G234_F/ MHC1_A1_G234_R) were identical to the priming sites on *Ctde-UA*001* and *Ctde-UA*003*. Sanger sequences produced using the MHC1_A1_G234_F/ MHC1_A1_G234_R primer pair matched the α1 of the revised transcriptome sequences *Ctde-UA*001* and *Ctde-UA*003* but contained many ambiguous sites, suggesting that multiple products (other than *Ctde-UA*001* and *Ctde-UA*003*) were simultaneously amplified.

In regards to α2, only the revised transcript *Ctde-UA*003* was identical to some of the original transcripts at α2 (group 4, fig. 3). As a result, most of the primers designed for α2 were different to the priming sites on the revised transcripts by one to seven base pairs. Despite this, α2 primers amplified product that was similar to revised transcripts. The resulting Sanger sequences had a high number of ambiguous sites, suggesting that α2 primers amplified multiple products simultaneously.

Three unique primer pairs were designed for α1, of which, one pair (1. MHC1_A1_G234_F/ 2. MHC1_A1_G234_R) amplified successfully in all three individuals (fig. 4, tables 1 and 2). This primer

pair encompasses most of the α1 domain (fig. 1), amplifying a 206bp product. Five unique primer pairs were designed for α2, amplifying products ranging from 204bp to 217bp in length (fig. 4, tables 1 and 2).

Tables and figures

Table 1. Primer sequences designed to amplify MHC I exon 2 and 3 in the tawny dragon (*Ctenophorus decresii*). F: forward, R: reverse.

Forward primer ID	Reverse primer ID	Target
1. MHC1_A1_G234_F	2. MHC1_A1_G234_R	α1, exon 2
3. MHC1_A2_G234_F	4. MHC1_A2_G12_R	α2, exon 3
5. MHC1_A2_G16_F	6. MHC1_A2_G3456_R	α2, exon 3
7. MHC1_A2_G5_F	6. MHC1_A2_G3456_R	α2, exon 3
5. MHC1_A2_G16_F	4. MHC1_A2_G12_R	α2, exon 3
3. MHC1_A2_G234_F	6. MHC1_A2_G3456_R	α2, exon 3

Table 2. Sequences of primers designed to amplify exons 2 and 3 of the tawny dragon (*Ctenophorus decresii*) MHC I region. F: forward, R: reverse.

Primer ID	Sequence (5' – 3')
1. MHC1_A1_G234_F	GGCTCCTCCTCGCACTCCCTG
2. MHC1_A1_G234_R	ACATTCTGCAGGCTCACTCTGAAC
3. MHC1_A2_G234_F	TGCAGTTGATGTACGGCTGTGAGC
4. MHC1_A2_G12_R	CCTCAGTAGGCTCTCCCTCCCG
5. MHC1_A2_G16_F	GTACAGCTGTGAGCTGGGGCAG
6. MHC1_A2_G3456_R	GCAGCGTCTCCTTCCCGTACTC
7. MHC1_A2_G5_F	CTGCAGCGGATGTACGGCTGTG

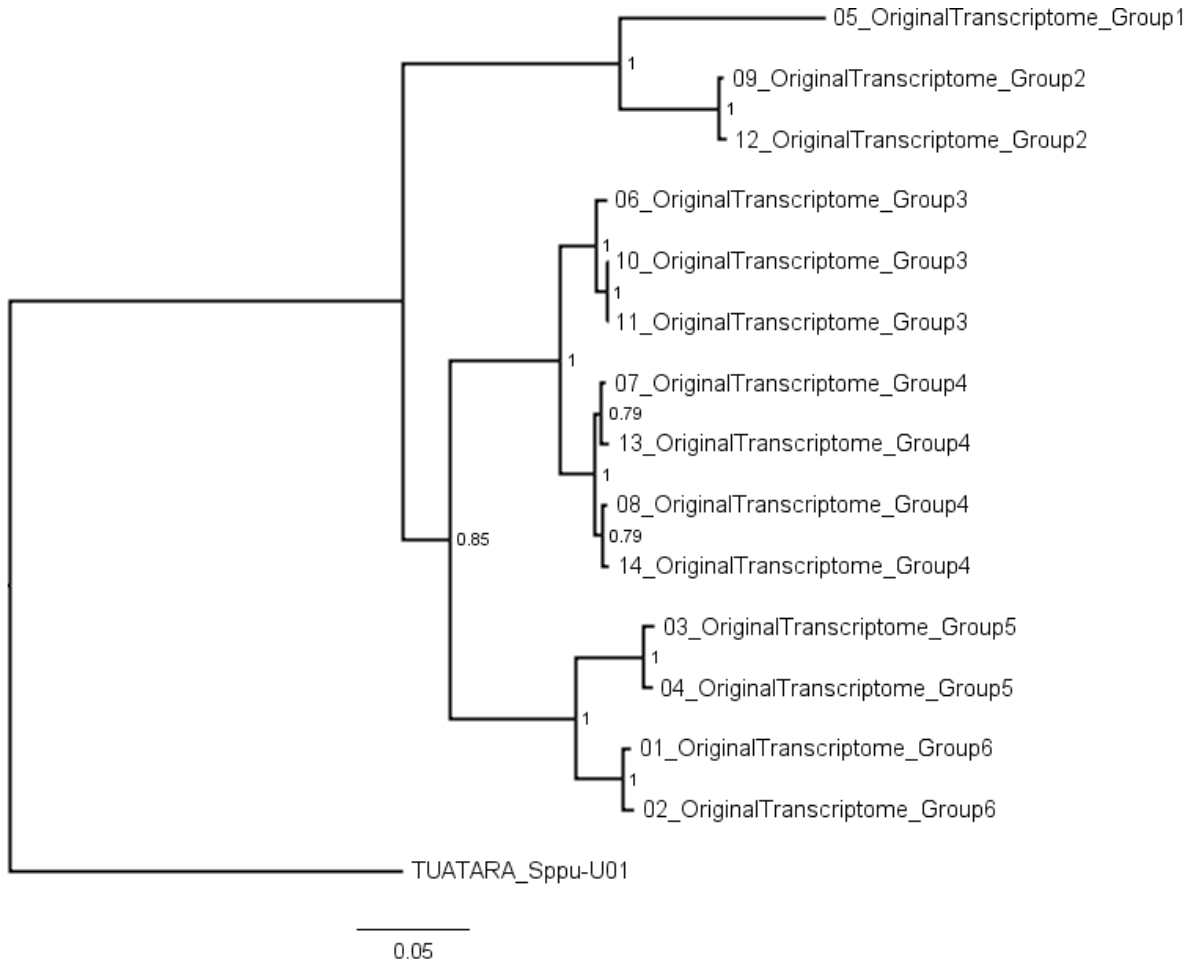


Figure 1. Neighbour-joining tree of original transcripts that were used for primer design. Group designation is displayed on tip labels. Bootstrap support (1000 replicates) is displayed at each node. The tree is rooted using tuatara (*Sphenodon punctatus*) MHC class I sequence.

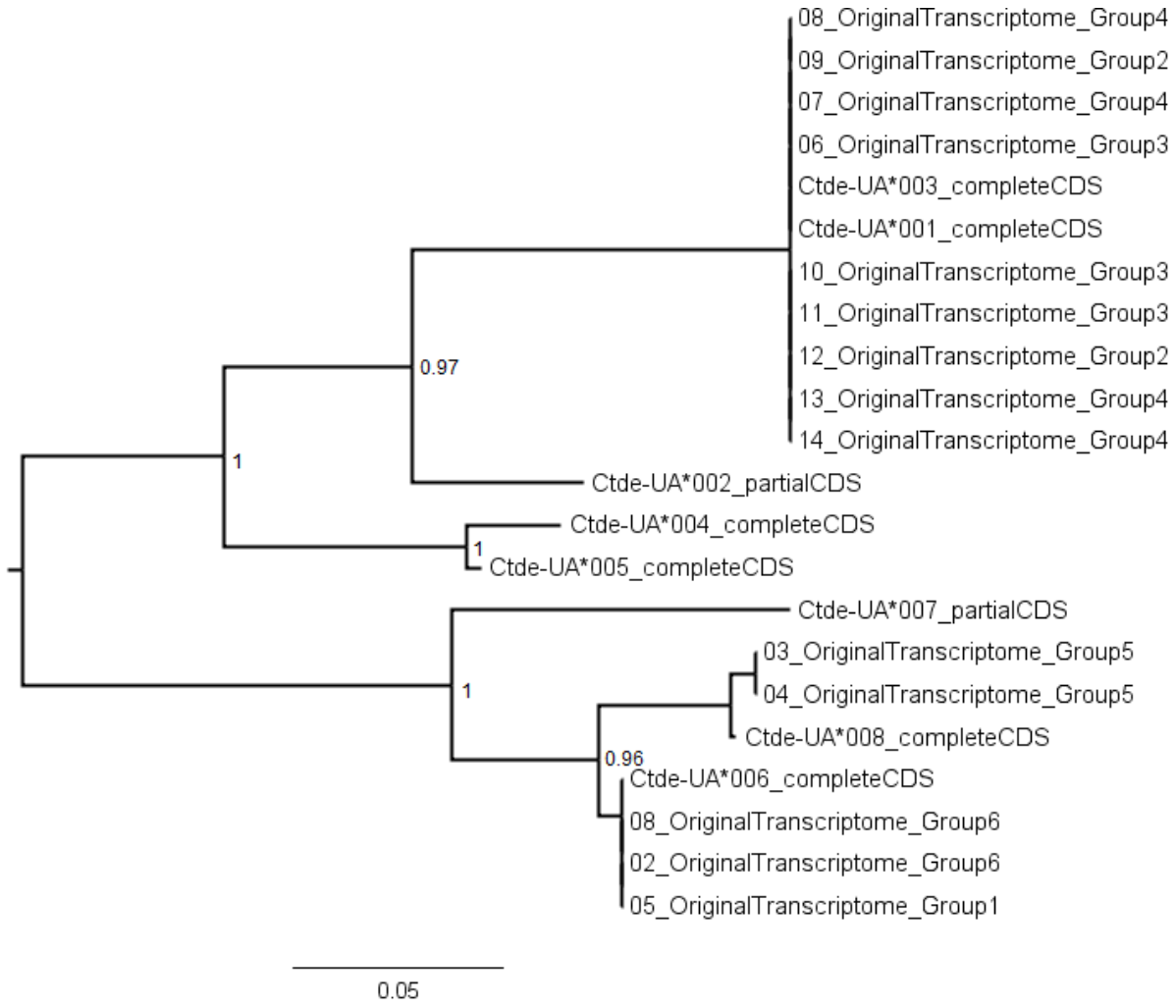


Figure 2. Neighbour-joining tree comparing the α1 domain of original transcripts and revised transcripts. Bootstrap support (1000 replicates) is displayed at each node and the tree is midpoint rooted.

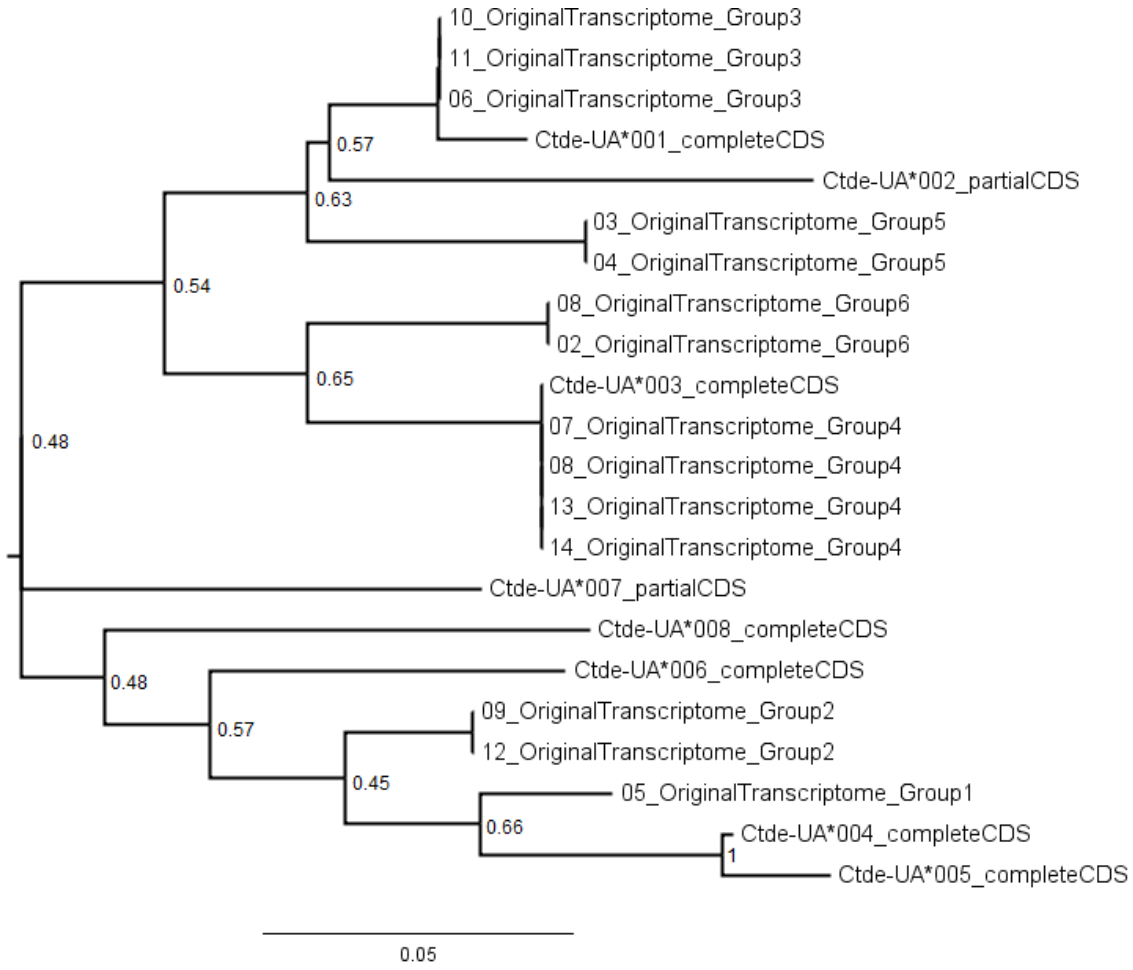


Figure 3. Neighbour-joining tree comparing the α2 domain of original transcripts and revised transcripts. Bootstrap support (1000 replicates) is displayed at each node and the tree is midpoint rooted.

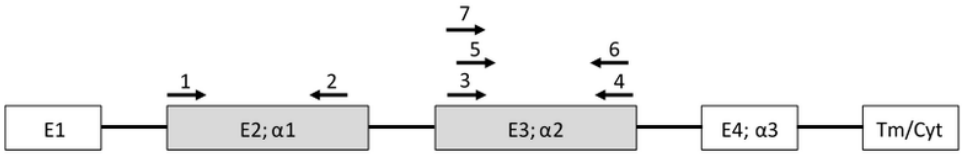


Figure 4. Relative positions of primers designed to amplify exons 2 (α1) and 3 (α2) (shaded grey) of the tawny dragon (*Ctenophorus decresii*) MHC I region. Primer numbers correspond to those listed in tables 4 and 5. Figure not to scale.

References

ANSARI, T., BERTOZZI, T., MILLER, R. & GARDNER, M. 2015. MHC in a monogamous lizard - Characterization of class I MHC genes in the Australian skink *Tiliqua rugosa*. *Developmental and Comparative Immunology,* 53**,** 320-327.

KAUFMAN, J., SALOMONSEN, J. & FLAJNIK, M. 1994. Evolutionary conservation of MHC class I and class II molecules - different yet the same. *Seminars in Immunology,* 6**,** 411-24.

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., MENTJIES, P. & DRUMMOND, A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics,* 28**,** 1647-1649.

MILLER, H. C., O'MEALLY, D., EZAZ, T., AMEMIYA, C., MARSHALL-GRAVES, J. A. & EDWARDS, S. 2015. Major Histocompatibility Complex Genes Map to Two Chromosomes in an Evolutionarily Ancient Reptile, the Tuatara *Sphenodon punctatus*. *G3: Genes|Genomes|Genetics,* 5**,** 1439-1451.

TAMURA, K., PETERSON, D., PETERSOIN, N., STECHER, G., NEI, M. & KUMAR, S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution,* 28**,** 2731-2739.

UNTERGASSER, A., CUTCUTACHE, I., KORESSAAR, T., YE, J., FAIRCLOTH, B. C., REMM, M. & ROZEN, S. G. 2012. Primer3—new capabilities and interfaces. *Nucleic Acids Research,* 40**,** e115-e115.

Appendix 2

This appendix provides the publication associated with chapter two:

Hacking, J., Bertozzi, T., Moussalli, A., Bradford, T., Gardner, M., (2018). Characterisation of major histocompatibility complex class I transcripts in an Australian dragon lizard. Developmental & Comparative Immunology, 84, 164-171.

Developmental and Comparative Immunology 84 (2018) 164–171

Contents lists available at ScienceDirect

Developmental and Comparative Immunology

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

Short communication

Characterisation of major histocompatibility complex class I transcripts in an Australian dragon lizard

Jessica Hacking [a, *], Terry Bertozzi [b, c], Adnan Moussalli [d], Tessa Bradford [a, b, c], Michael Gardner [a, b]

[a] College of Science and Engineering, Flinders University, Bedford Park, SA, 5042, Australia
[b] Evolutionary Biology Unit, South Australian Museum, Adelaide, SA, 5000, Australia
[c] School of Biological Sciences, University of Adelaide, Adelaide, SA, 5005, Australia
[d] Sciences Department, Museum Victoria, Carlton Gardens, VIC, 3053, Australia

ARTICLE INFO

Article history:
Received 15 December 2017
Received in revised form
10 February 2018
Accepted 10 February 2018
Available online 16 February 2018

Keywords:
Transcriptome assembly
Iguania
Agamidae
Ctenophorus decresii
MHC class I evolution

ABSTRACT

Characterisation of squamate major histocompatibility complex (MHC) genes has lagged behind other taxonomic groups. MHC genes encode cell-surface glycoproteins that present self- and pathogen-derived peptides to T cells and play a critical role in pathogen recognition. Here we characterise MHC class I transcripts for an agamid lizard (*Ctenophorus decresii*) and investigate the evolution of MHC class I in Iguanian lizards. An iterative assembly strategy was used to identify six full-length *C. decresii* MHC class I transcripts, which were validated as likely to encode classical class I MHC molecules. Evidence for exon shuffling recombination was uncovered for *C. decresii* transcripts and Bayesian phylogenetic analysis of Iguanian MHC class I sequences revealed a pattern expected under a birth-and-death mode of evolution. This work provides a stepping stone towards further research on the agamid MHC class I region.

1. Introduction

The major histocompatibility complex (MHC) is a multigene family involved in pathogen recognition and immune response, and is one of the most diverse regions of the vertebrate genome (Piertney and Oliver, 2006). MHC genes encode cell surface glycoproteins that present self- and foreign-derived peptides to circulating T-lymphocyte cells (T cells). The evolution of MHC genes is complex and is thought to be governed primarily by the birth-and-death model of evolution in which loci are duplicated or lost, although concerted evolution via inter-locus gene conversion events may also play a role (Edwards and Hedrick, 1998; Nei and Rooney, 2005; Spurgin et al., 2011). These processes can occur over short time scales and it is apparent that MHC genes have undergone numerous independent expansion and diversification events throughout vertebrate evolution (Nei et al., 1997). The MHC is gene rich and is generally extremely polymorphic within loci (Janeway et al., 2001). Pathogen-mediated natural selection and sexual selection (MHC-associated mating) are considered to be the primary mechanisms maintaining these extraordinary levels of diversity (Edwards and Hedrick, 1998; Ejsmond et al., 2014; Milinski, 2006).

The MHC is divided into four classes based primarily on structural and functional differences (Janeway et al., 2001). Genes belonging to classes I and II are further separated into classical or non-classical genes based primarily on function and expression patterns (Alfonso and Karlsson, 2000; Janeway et al., 2001). The structure of classical MHC class I (hereafter MHC I) molecules is conserved among jawed vertebrates and includes a leader peptide, three α domains, and the transmembrane and cytoplasmic (Tm/Cyt) domains, all of which are encoded by a single gene (Kaufman et al., 1994). The α1 and α2 domains form the peptide binding cleft and contain amino acid positions that are directly involved in peptide binding, termed peptide binding regions (PBR) (Janeway et al., 2001). Classical MHC I molecules are anchored to the surface of somatic cells via the Tm/Cyt domains and display self-

Abbreviations: MHC, major histocompatibility complex; PBR, peptide binding region; CNV, copy number variation.

* Corresponding author.
E-mail addresses: jessica.hacking@flinders.edu.au (J. Hacking), terry.bertozzi@samuseum.sa.gov.au (T. Bertozzi), amoussalli@museum.vic.gov.au (A. Moussalli), tessa.bradford@samuseum.sa.gov.au (T. Bradford), michael.gardner@flinders.edu.au (M. Gardner).

https://doi.org/10.1016/j.dci.2018.02.012
0145-305X/

peptides and antigenic peptides derived primarily from intracellular pathogens to cytotoxic T cells. When a particular MHC molecule presents an antigenic peptide and is recognized by a T cell, an immune response is initiated, which usually involves lysis of the infected cell (Neefjes et al., 2011). Non-classical MHC class I genes are distinguished from classical genes by low levels of allelic variation and restricted expression (Janeway et al., 2001). In mammals some non-classical MHC class I genes undertake important roles within the immune system, both at the cell surface and in secreted forms (Adams and Luoma, 2013).

The MHC has been thoroughly characterised in humans and model organisms, primarily due to the critical role that the region plays in organ and tissue transplantation (Garcia et al., 2012). MHC genes are also used in conservation genetic studies as a measure of population genetic health and adaptive potential (Sommer, 2005). However, many vertebrate groups, especially non-avian reptiles, are under-represented within the MHC literature and little is known regarding the mechanisms shaping MHC diversity in these taxa. The tawny dragon (*Ctenophorus decresii*) is a small (<30 g) agamid lizard endemic to South Australia and provides a promising model system in which to investigate the mechanisms shaping MHC diversity. Male *C. decresii* use visual cues during social and sexual interactions (Gibbons, 1979; Osborne, 2005a, b; Yewers et al., 2016) and the species is host to external and intracellular parasites (Hacking et al. unpublished results); providing opportunities to investigate the roles of sexual selection and parasite-mediated selection in maintaining MHC diversity. Here, we characterised MHC I transcripts for *C. decresii* and investigated the evolutionary mechanisms playing a role in the generation of MHC I diversity within Iguanian lizards (Iguanidae, Agamidae, Chamaeleonidae, Dactyloidae and related families, Pyron et al., 2013).

2. Materials and methods

2.1. Sample collection

A single *C. decresii* individual was captured in Burra, South Australia (33°40′57.7″S, 138°56′16.8″E) in October 2012 and taken directly to Adelaide to be euthanized for tissue collection. Burra is located just north of the contact zone between the northern and southern clades of this species (McLean et al., 2014). The thymus and spleen were collected immediately after euthanasia and were stored separately in RNA Later (Qiagen, Venlo, Netherlands) at 4 °C for 48 h and then at −80 °C until required for RNA extraction. The remainder of the specimen was accessioned into the South Australian Museum herpetology collection (SAMAR67384).

2.2. Transcriptome sequencing and MHC class I discovery

Total RNA was extracted using the Qiagen RNeasy mini kit. Sequencing libraries were then prepared using the TruSeq RNA Kit v1 using a polyA purification. These were then multiplexed with two other samples and sequenced (100 bp paired end) on a single lane of the HISEQ 2000. Extractions, library preparation and sequencing were carried out by Georgia Genomic Facility (GGF, University of Georgia, USA). Adaptor sequences and low quality reads were removed or trimmed using Trimmomatic ver. 0.22 (Bolger et al., 2014), with a minimum quality Phred score of 25 per 4bp sliding window and a minimum sequence length of 40bp. Assemblies were then constructed from the trimmed and filtered reads for each sample separately using the program Trinity v1 (r2013-02-25) (Haas et al., 2013) with default settings, followed by an assessment of gene completeness using BUSCO v1.22 (Simao et al., 2015) based on the OrthoDB 'vertebrata' database. Lastly, to identify putative *C. decresii* MHC I transcripts we performed local BLASTX (E-value ≤ 1e-10) searches (Altschul et al., 1997; Camacho et al., 2009) using predicted *Pogona vitticeps* MHC gene models (Georges et al., 2015) as our reference. Putative *C. decresii* MHC I transcripts were aligned manually with published MHC I sequences (Table S1) to confirm expected MHC I structure and the presence of conserved sites (Kaufman et al., 1994).

Due to the high diversity and complex structure of the MHC region, traditional assembly methods may not be sufficient to obtain a robust assembly. To refine the MHC I assemblies, 75bp sequence fragments congruent with the putative antigen binding α2 domain for each unique sequence were iteratively re-assembled using the mirabait utility from MIRA v4.0.2 (Chevreux et al., 1999) as presented in Ansari et al. (2015) but using a kmer length of 31 (size of the search string) and requiring 50 matching kmers (number of matching search strings). Sequences were extended until sequence length stabilized or it was no longer possible to uniquely map reads. The resulting contigs were evaluated by concatenating each separated by 200 Ns and remapping cleaned reads using BWA (Li, 2013) with default settings and visualizing the resultant BAM file in IGV (Robinson et al., 2011). Read pairs spanning sequences were checked for accuracy and suspected chimeric reads removed.

2.3. Validation of MHC I transcripts and comparison with other vertebrates

Putative *C. decresii* MHC class I transcripts were translated and aligned with a subset of published full length MHC I amino acid sequences of other vertebrates (Table S1) using MUSCLE (Edgar, 2004) implemented in MEGA ver. 6.06 (Tamura et al., 2011). The alignment was manually refined to ensure correct alignment of conserved regions. Coding domain boundaries were defined as per Koller and Orr (1985). Aligned *C. decresii* transcripts were validated as likely MHC I sequences by (i) confirming MHC I gene structure (leader peptide, α domains and Tm/Cyt domains), (ii) confirming concordance with known conserved regions and regions with predicted function that are typical of MHC I sequences (Kaufman et al., 1994), and iii) confirming the absence of stop codons within coding regions. Two additional steps needed to further validate transcripts as likely classical MHC I sequences, which were beyond the scope of this study, are confirming polymorphism among individuals, and strong and widespread expression. Pairwise nucleotide and amino acid identity among validated MHC class I transcript sequences was calculated using Geneious ver. 8.1.7 (Kearse et al., 2012). Validated *C. decresii* MHC class I transcripts were named according to Klein et al. (1990); each unique nucleotide sequence was given the species identification prefix (Ctde) followed by U (Uno; class 1) and A (locus group/family designation), and a unique number (e.g. *Ctde-UA*001*). Once full-length MHC I genomic data are available for *C. decresii* locus designations may be defined (i.e. UA1 and UA2).

To investigate relationships among *C. decresii* MHC I transcripts and their position relative to other Iguanian lizards a Bayesian phylogenetic tree was constructed based on an alignment of validated full-length *C. decresii* MHC class I transcripts and all full-length Iguanian MHC I sequences available in GenBank (NCBI Resource Coordinators, 2016) (Table S1). All sequences obtained from GenBank were validated as likely MHC I sequences via confirmation of expected MHC I structure and conserved sites. All squamate MHC I nucleotide sequences were translated before aligning with Muscle, implemented in MEGA ver. 6.06 and then untranslated for phylogenetic analysis. Tuatara (*Sphenodon punctatus*) MHC I sequence was used as an outgroup. Only the three α domains (α1, α2 and α3) of these full-length sequences were used in phylogenetic analysis due to extreme variation at leader and Cyt/

Tm domains inhibiting alignment. To determine optimal partitioning and the best model of evolution for Bayesian phylogenetic analysis, PartitionFinder2 (Guindon and Gascuel, 2003; Lanfear et al., 2012, 2017) was employed. Sequences were split into three data blocks representing codon positions and only models employed by MrBayes were considered, with the best model determined using AICc model selection. Bayesian phylogenetic analyses were undertaken in MrBayes ver. 3.2.6 (Ronquist et al., 2011), with one analysis employing the model of evolution and partitioning identified by PartitionFinder2, and another using mixed models with sequences partitioned by codon position. For both MrBayes analyses, two independent runs were performed, each with four Markov chains run for 20 million generations at a sample frequency of 1000 and a default burn-in period of 25%. Convergence diagnostics, including the standard deviation of split frequencies between runs, the potential scale reduction factor (PSRF) and the average effective sample size (ESS) were examined to confirm run convergence. FigTree ver. 1.4.2 (Rambaut, 2012) was used to annotate trees produced by MrBayes.

3. Results and discussion

3.1. Transcriptome sequencing and MHC class I discovery

RNA extractions had RIN values of 9.1 and 9.8 for the thymus and spleen samples, respectively. Approximately 49 million paired-end reads were obtained for the thymus and approximately 51 million paired-end reads were obtained for the spleen. Of these reads, approximately 86% survived Trimmomatic filtering. Based on the Busco analysis, 73% and 76% of the vertebrata orthologous gene set were recovered and complete for the thymus (C:73% [D:15%],F:7.7%,M:19%,n:3023) and spleen (C:76% [D:13%],F:6.4%,M:16%,n:3023) transcriptomes, respectively. Combing the two transcriptome assemblies resulted in 82% of the vertebrata gene set being complete (C:82% [D:43%],F:5.1%,M:12%,n:3023). In total, eight putative *C. decresii* MHC I transcripts (*Ctde-UA*001 – Ctde-UA*008*) were discovered, based on unique α2 and α3 domains, indicating the presence of at least four different loci. We were unable to confidently assign transcripts to specific loci based on data from a single individual. As a result, when naming transcripts all sequences were designated the letters 'UA' (MHC I locus group 'A') and no specific locus numbers were assigned (i.e. UA1, UA2, UA3 etc).

3.2. Validation of C. decresii MHC I transcripts

Of the eight putative *C. decresii* MHC I transcripts, six were validated as likely MHC I sequences after confirming (i) normal MHC I gene structure, (ii) concordance with known conserved regions and regions with predicted function that are typical of MHC I sequences, and (iii) the absence of stop codons within coding regions (*Ctde-UA*001 – Ctde-UA*006*). Each of these transcripts contained complete leader peptide, α domains and Tm/Cyt domains and did not contain any premature stop codons. An amino acid alignment of putative *C. decresii* MHC I transcripts with published squamate, tuatara (*Sphenodon punctatus*) and human sequences, confirmed concordance with conserved regions and regions with predicted functions that are typical of MHC I molecules (Table S2, Fig. 1, Kaufman et al., 1994). Specifically, nine amino acid positions that bind C- and N-terminal residues of antigenic peptides are highly conserved among classical MHC class I molecules (Kaufman et al., 1994). All nine of these positions were identified within *C. decresii* MHC I transcripts and displayed conserved amino acids or low (≤2 changes) amino acid variability (site numbers: 45, 101, 126, 167, 187, 190, 191, 204 and 219). Similarly, amino acid positions involved in salt bridge formation within α1 and α2 were found to be conserved, with the histidine (H) residues at site 41 and 137, and the aspartic acid (D)/glutamic acid (E) residues at sites 69 and 163. The four cysteine (C) residues involved in intra-domain disulphide bridge formation within α2 and α3 (site numbers: 145, 209, 249 and 305) were conserved across all taxa included in the amino acid alignment. Finally, most vertebrates possess an NQS or NQT nitrogen-linked glycosylation acceptor site near the end of the α1 domain (Kaufman et al., 1994). Nitrogen-linked glycans play an important role in the folding and stability of classical, and likely also non-classical, MHC class I molecules (Ryan and Cobb, 2012, 2015). The amino acid sequence for nitrogen-linked glycosylation site consists of an asparagine (N), any other amino acid except proline, followed by serine (S) or threonine (T) (Ryan and Cobb, 2015). Most of the *C. decresii* transcripts encoded NQS glycosylation sites, with three transcripts (*Ctde-UA*005* and *Ctde-UA*006*) encoding NHS instead. Together, these findings suggest that transcripts *Ctde-UA*001 – Ctde-UA*006* produce molecules that undertake classical MHC I functions at the cell surface.

Classical MHC I gene structure (leader peptide, α domains and Tm/Cyt domains) could not be confirmed for two *C. decresii* putative MHC I transcripts; *Ctde-UA*007* and *Ctde-UA*008*. Both *Ctde-UA*007* and *Ctde-UA*008* ended prematurely; *Ctde-UA*007* was missing the 5′ end of the α3 domain and all of the Tm/Cyt domains and *Ctde-UA*008* was missing both the α3 and the Tm/Cyt domains. Both *Ctde-UA*007* and *Ctde-UA*008* possessed the expected conserved sites associated with peptide binding at the α1 and α2 domains and did not contain any premature stop codons (Fig. 1). These sequences could not be unambiguously resolved during transcriptome assembly, probably due to similarity with the other transcripts. This is not surprising given that MHC loci arise by periodic gene duplication and rearrangement events (Nei and Rooney, 2005). It is also possible that these transcripts are naturally truncated and encode non-classical MHC I molecules that are secreted and undertake peptide presentation away from the cell surface, given the lack of Tm/Cyt domains and the presence of conserved sites associated with peptide presentation (Carlini et al., 2016; Donadi et al., 2011; Glaberman et al., 2009). Given that *Ctde-UA*007* and *Ctde-UA*008* are likely truncated due to incomplete transcriptome assembly they were not assigned to a new locus group and were included in the 'UA' MHC I locus group.

Average nucleotide percent identity among the validated full-length *C. decresii* MHC I transcripts was 79.8%, with this differentiation primarily driven by variation at the α1 and α2 domains (Table S2). Nucleotide identity within α1 and α2 was 81% and 82%, respectively, whereas nucleotide identity within α3 was 88%. Sequences *Ctde-UA*001* and *Ctde-UA*002* were identical only at the α1 domain, suggesting the possibility for exon shuffling recombination in the *C. decresii* MHC I region. Exon shuffling recombination occurs between entire exon regions with breaks within introns and has been found to occur in a range of vertebrates (Holmes and Parham, 1985; Wang et al., 2010; Zhao et al., 2013). Overall percent nucleotide identity between *Ctde-UA*001* and *Ctde-UA*002* is 95.6% and between *Ctde-UA*003* and *Ctde-UA**004 it is 95.1%. Nucleotide identity is slightly lower between *Ctde-UA*005* and *Ctde-UA*006* at 93.7%. The high similarity between these pairs of transcripts indicates that they may be alleles at the same loci, suggesting five loci in total. Bayesian phylogenetic analysis of *C. decresii* transcripts (see section 3.3 for further details) suggest the presence of five or six loci. *Ctde-UA*001* and *Ctde-UA*002,* and *Ctde-UA*003* and *Ctde-UA**004 each clustered together but *Ctde-UA*005* and *Ctde-UA*006* didn't, with *Ctde-UA*006* located between *Ctde-UA*005* and a *P. vitticeps* sequence.

Fig. 1. Amino acid alignment of tawny dragon (*Ctenophorus decresii*) MHC I transcripts with other squamates, tuatara (*Sphenodon punctatus*) and human. Coding domain separations are based on Koller and Orr (1985). Dots indicate identity with *Ctde-UA*001* and dashes indicate alignment gaps. ¥ = partial coding sequence. Residues with expected functions as per Kaufman et al. (1994) are shaded grey; *stars* = conserved peptide-binding residues of antigen N and C termini, *triangles* = salt bridge-forming residues, *circles* = disulphide bridge-forming cysteines, *square* = N-glycosylation site, CD8 = expected CD8 binding site. *Asterisks* represent conserved sites, across *C. decresii* sequences and across all taxa. Gidgee skink (*Egernia stokesii*) positively selected sites (PSS, putative peptide binding regions) (Pearson et al., 2017) and human peptide binding regions (PBR) (Reche and Reinherz, 2003) are indicated with *hashes* and *crosses*, respectively.

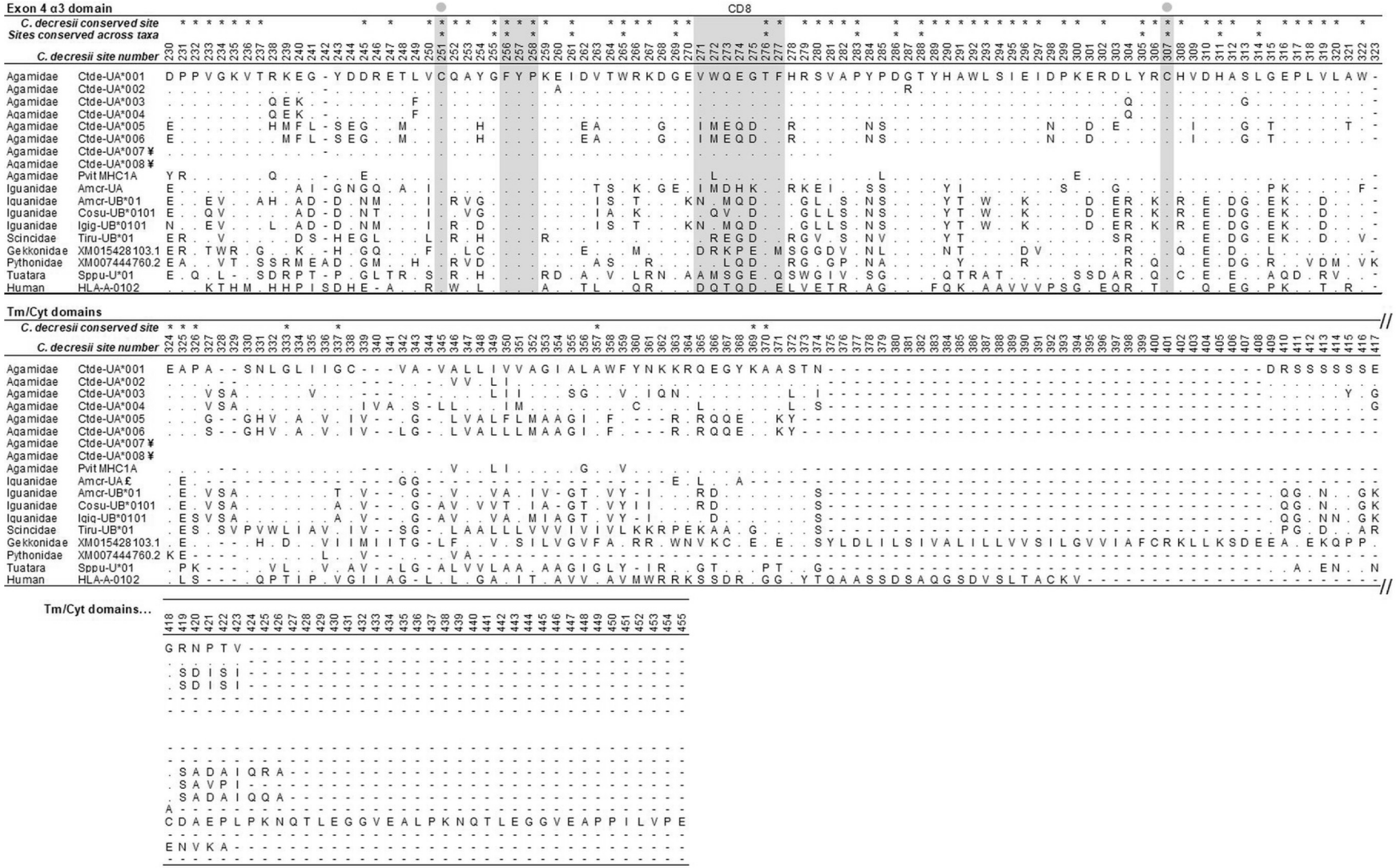


Fig. 1. (*continued*).

3.3. Phylogenetic analysis of Iguanian MHC class I genes

One of the Bayesian phylogenetic analyses of MHC I sequences was conducted with no partitioning of codon positions and a GTR + G model of evolution, according to PartitionFinder2 results. The other analysis was conducted using data partitioned by codon position and allowing mixed models of evolution. For both Bayesian analyses, convergence diagnostics indicated run convergence and high posterior probability values were obtained across most of the tree. The two resulting trees had identical topology but the tree resulting from the PartitionFinder2-informed analysis had slightly better overall posterior probability values and was therefore retained. Posterior probabilities were high (>0.95) for most branch nodes, but low for some outer nodes (Fig. 2).

Full-length MHC I sequences were only available for three Iguania species (*Amblyrhynchus cristatus*, *Conolophus subcristatus* and *Iguana iguana*), one Dactyloidae species (*Anolis carolinensis*) and two Agamidae species (C. *decresii* and *Pogona vitticeps*, Fig. 2, Table S1). No orthologous relationships were observed within the Iguanidae clade, representing three genera that diverged 10–20 million years ago (Rassmann, 1997). Instead, MHC I sequences clustered by species, suggesting loss of ancestral diversity, recent gene duplications and potentially gene conversion events (concerted evolution, Glaberman and Caccone, 2008). In contrast, Agamidae MHC I sequences (*C. decresii* and *P. vitticeps*) displayed orthologous relationships despite similar divergence times to the iguanid species (Hugall et al., 2008). Given divergence of approximately 20 million years between *P. vitticeps* and *C. decresii* (Hugall et al., 2008), the separation of the agamid MHC I sequences into three orthologous clades suggests that at least two gene duplication events occurred greater than 20 million years ago (Fig. 2). The conservation of MHC class I loci over such a time scale is not unusual; some primate MHC I loci have been conserved for at least 46–66 million years (Piontkivska and Nei, 2003). The phylogenetic relationships within the Iguanidae and Agamidae families are consistent with the birth-and-death model of evolution; the continual gain and loss of genes (Edwards and Hedrick, 1998; Nei and Rooney, 2005). Concerted evolution via gene conversion may also play a role in generating diversity within Iguanidae and Agamidae, as evidenced by the patterns observed within Iguanidae and putative exon shuffling recombination in *C. decresii*.

4. Conclusions

A total of eight MHC I transcripts were isolated and characterised for *C. decresii* from a single individual using HiSeq next generation sequencing of thymus and spleen total RNA. Due to the complex nature of the MHC, the initial assembly was refined using an assembly technique capable of distinguishing highly similar transcripts. Six of the putative MHC I transcripts were validated as

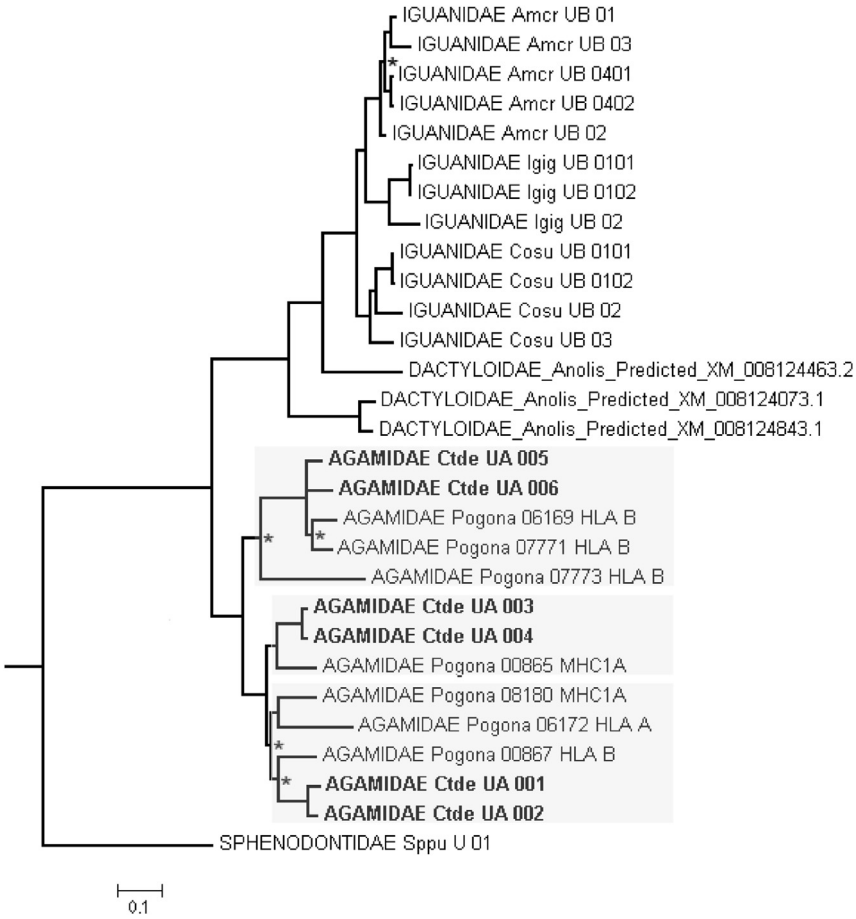


Fig. 2. Bayesian phylogenetic tree of the α1, α2 and α3 domains of available full length Iguania (here, Iguanidae, Agamidae and Dactyloidae) MHC class I nucleotide sequence. Validated full length tawny dragon (*Ctenophorus decresii*) MHC class I transcripts are highlighted by bold text. The tree is rooted using tuatara (*Sphenodon punctatus*) MHC class I sequence. Three orthologous Agamidae clades are shaded grey. Nodes for which the Bayesian posterior probability <0.95 are indicated with an asterisk. The scale bar indicates the number of expected nucleotide substitutions per site. *Anolis* sequences were predicted by NCBI automated computational searches of genomic sequence (NCBI Resource Coordinators, 2016). The *Pogona* sequences were obtained from the Pogona Genome Project (Georges et al., 2015, 2016). Refer to Table S1 for further information on sequences.

likely to encode classical MHC I molecules based on three criteria; (i) normal MHC I gene structure, (ii) concordance with known conserved regions and regions with predicted function, and (iii) the absence of stop codons within coding regions. Two of the putative MHC I transcripts ended prematurely either due to transcriptome assembly restrictions (i.e. due to high sequence similarity among transcripts) or non-classical functionality (naturally truncated). Bayesian phylogenetic analysis indicated that a birth-and-death model of evolution is likely the main mechanism shaping MHC I diversity within Iguanian lizards. MHC I sequences from a wider range of squamates are required to obtain a clearer view of the mechanisms responsible for creating MHC I diversity within Squamata. This work provides a foundation for future work examining the mechanisms shaping diversity at the MHC class I region of agamid lizards.

Acknowledgments

We thank Talat Hojat Ansari and Sarah Pearson for assistance with laboratory work and data analysis, and Aaron Fenner for assistance with fieldwork. This work was funded by the Holsworth Wildlife Research Endowment (no. HOLSW2015-1-F177) and the Nature Foundation of South Australia, awarded to JH. Specimen collection and euthanasia were conducted according to the Flinders University Animal Welfare Committee (approval no. E379) and the South Australian Department of Environment, Water and Natural Resources regulations (permit no. U26225).

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.dci.2018.02.012.

References

Adams, E.J., Luoma, A.M., 2013. The adaptable major histocompatibility complex (MHC) fold: structure and function of nonclassical and MHC class I–like molecules. Annu. Rev. Immunol. 31, 529–561.
Alfonso, C., Karlsson, L., 2000. Nonclassical MHC class II molecules. Annu. Rev. Immunol. 18, 113–142.
Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.
Ansari, T., Bertozzi, T., Miller, R., Gardner, M., 2015. MHC in a monogamous lizard - characterization of class I MHC genes in the Australian skink *Tiliqua rugosa*. Dev. Comp. Immunol. 53, 320–327.
Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120.
Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. BMC Bioinf. 10, 421.
Carlini, F., Ferreira, V., Buhler, S., Tous, A., Eliaou, J.-F., René, C., Chiaroni, J., Picard, C., Di Cristofaro, J., 2016. Association of HLA-a and non-classical HLA class I alleles. PLoS One 11 e0163570.
Chevreux, B., Wetter, T., Suhai, S., 1999. Genome sequence assembly using trace signals and additional sequence information. In: Proceedings of the German Conference on Bioinformatics (GCB), pp. 45–56.
Donadi, E.A., Castelli, E.C., Arnaiz-Villena, A., Roger, M., Rey, D., Moreau, P., 2011. Implications of the polymorphism of HLA-G on its function, regulation, evolution and disease association. Cell. Mol. Life Sci. 68, 369–395.
Edgar, R., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.
Edwards, S., Hedrick, P., 1998. Evolution and ecology of MHC molecules: from genomics to sexual selection. Trends Ecol. Evol. 13, 305–311.
Ejsmond, M., Radwan, J., Wilson, A., 2014. Sexual selection and the evolutionary dynamics of the major histocompatibility complex. Proc. Biol. Sci. 281, 1–8.
Garcia, M., Yebra, B., Flores, A., Guerra, E., 2012. The major histocompatibility complex in transplantation. Journal of Transplantation 2012, 1–7.
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., Matsubara, K., Waters, P., Graves, J.A.M., Sarre, S.D., Zhang, G., 2015. High-coverage sequencing and annotated assembly of the genome of the Australian dragon lizard *Pogona vitticeps*. GigaScience 4, 1–11.
Georges, A., O'Meally, D., Genomics@UC, 2016. The Pogona vitticeps Genome Browser (pvi1.1 Jan 2013). University of Canberra, Canberra. Institute for Applied Ecology.
Gibbons, J., 1979. The hind leg pushup display of the *Amphibolurus decresii* species complex (Lacertilia: Agamidae). Copeia 1, 29–40.
Glaberman, S., Caccone, A., 2008. Species-specific evolution of class I MHC genes in iguanas (order: squamata; subfamily: iguaninae). Immunogenetics 60, 371–382.
Glaberman, S., Du Pasquier, L., Caccone, A., 2009. Characterization of a nonclassical class I MHC gene in a reptile, the galapagos marine Iguana (*Amblyrhynchus cristatus*). PLoS One 3 e2859.
Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52.
Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., MacManes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., LeDuc, R.D., Friedman, N., Regev, A., 2013. *De novo* transcript sequence reconstruction from RNA-Seq: reference generation and analysis with trinity. Nat. Protoc. 8, 1494–1512.
Holmes, N., Parham, P., 1985. Exon shuffling *in vivo* can generate novel HLA class I molecules. EMBO J. 4, 2849–2854.
Hugall, A.F., Foster, R., Hutchinson, M., Lee, M.S.Y., 2008. Phylogeny of Australasian agamid lizards based on nuclear and mitochondrial genes: implications for morphological evolution and biogeography. Biol. J. Linn. Soc. 93.
Janeway, C.J., Travers, P., Walport, M., 2001. Immunobiology: the Immune System in Health and Disease, fifth ed. Garland Science, New York.
Kaufman, J., Salomonsen, J., Flajnik, M., 1994. Evolutionary conservation of MHC class I and class II molecules - different yet the same. Semin. Immunol. 6, 411–424.
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., Mentjies, P., Drummond, A., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649.
Klein, J., Bontrop, R.E., Dawkins, R., Erlich, H.A., Gyllensten, U.B., Heise, E., Jones, P., Parham, P., Wakeland, E., Watkins, D., 1990. Nomenclature for the major histocompatibility complexes of different species: a proposal. Immunogenetics 31, 217–219.
Koller, B.H., Orr, H.T., 1985. Cloning and complete sequence of an HLA-A2 gene: analysis of two HLA-A alleles at the nucleotide level. J. Immunol. 134, 2727–2733 (Baltimore, Md. : 1950).
Lanfear, R., Calcott, B., Ho, S.Y., Guindon, S., 2012. Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. 29, 1695–1701.
Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B., 2017. Partitionfinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol. Biol. Evol. 34, 772–773.
Li, H., 2013. Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM arXiv:1303.3997v2 [q-bio.GN].
McLean, C.A., Stuart-Fox, D., Moussalli, A., 2014. Phylogeographic structure, demographic history and morph composition in a colour polymorphic lizard. J. Evol. Biol. 27, 2123–2137.
Milinski, M., 2006. The major histocompatibility complex, sexual selection, and mate choice. Annu. Rev. Ecol. Evol. Systemat. 37.
NCBI Resource Coordinators, 2016. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 44, D7–D19.
Neefjes, J., Jongsma, M., Paul, P., Bakke, O., 2011. Towards a systems understanding of MHC class I and MHC class II antigen presentation. Nat. Rev. Immunol. 11, 823–836.
Nei, M., Gu, X., Sitnikova, T., 1997. Evolution by the birth-and-death process in multigene families of the vertebrate immune system. Proc. Natl. Acad. Sci. Unit. States Am. 94.
Nei, M., Rooney, A.P., 2005. Concerted and birth-and-death evolution of multigene families. Annu. Rev. Genet. 39, 121–152.
Osborne, L., 2005a. Information content of male agonistic displays in the territorial tawny dragon (*Ctenophorus decresii*). J. Ethol. 23, 189–197.
Osborne, L., 2005b. Rival recognition in the territorial tawny dragon (*Ctenophorus decresii*). Acta Ethol. 8, 45–50.
Pearson, S.K., Bull, C.M., Gardner, M.G., 2017. *Egernia stokesii* (gidgee skink) MHC I positively selected sites lack concordance with HLA peptide binding regions. Immunogenetics 69, 49–61.
Piertney, S., Oliver, M., 2006. The evolutionary ecology of the major histocompatibility complex. Heredity 96, 7–21.
Piontkivska, H., Nei, M., 2003. Birth-and-death evolution in primate MHC class I genes: divergence time estimates. Mol. Biol. Evol. 20.
Pyron, R.A., Burbrink, F.T., Wiens, J.J., 2013. A phylogeny and revised classification of Squamata, including 4161 species of lizards and snakes. BMC Evol. Biol. 13, 93.
Rambaut, A., 2012. FigTree, 1.4.2. University of Edinburgh. Institute of Evolutionary Biology.
Rassmann, K., 1997. Evolutionary age of the galápagos iguanas predates the age of the present galápagos islands. Mol. Phylogenet. Evol. 7, 158–172.
Reche, P.A., Reinherz, E.L., 2003. Sequence variability analysis of human class I and class II MHC molecules: functional and structural correlates of amino acid polymorphisms. J. Mol. Biol. 331, 623–641.
Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., Mesirov, J.P., 2011. Integrative genomics viewer. Nat. Biotechnol. 29, 24–26.
Ronquist, F., Teslenko, P., Ayres, D., Darling, S., Hohna, S., Larget, B., Liu, L.,

Suchard, M.H., JP, 2011. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 6, 539–542.
Ryan, S.O., Cobb, B.A., 2012. Roles for major histocompatibility complex glycosylation in immune function. Semin. Immunopathol. 34, 425–441.
Ryan, S.O., Cobb, B.A., 2015. Major histocompatibility complex: N-Glycosylation form and function. In: Taniguchi, N., Endo, T., Hart, G.W., Seeberger, P.H., Wong, C.-H. (Eds.), Glycoscience: Biology and Medicine. Springer Japan, Tokyo, pp. 643–648.
Simao, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210–3212.
Sommer, S., 2005. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. Front. Zool. 2, 16.
Spurgin, L.G., van Oosterhout, C., Illera, J.C., Bridgett, S., Gharbi, K., Emerson, B.C., Richardson, D.S., 2011. Gene conversion rapidly generates major histocompatibility complex diversity in recently founded bird populations. Mol. Ecol. 20, 5213–5225.
Tamura, K., Peterson, D., Petersoin, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731–2739.
Wang, D., Zhong, L., Wei, Q., Gan, X., He, S., 2010. Evolution of MHC class I genes in two ancient fish, paddlefish (*Polyodon spathula*) and Chinese sturgeon (*Acipenser sinensis*). FEBS Lett. 584, 3331–3339.
Yewers, M.S.C., Pryke, S., Stuart-Fox, D., 2016. Behavioural differences across contexts may indicate morph-specific strategies in the lizard *Ctenophorus decresii*. Anim. Behav. 111, 329–339.
Zhao, M., Wang, Y., Shen, H., Li, C., Chen, C., Luo, Z., Wu, H., 2013. Evolution by selection, recombination, and gene duplication in MHC class I genes of two Rhacophoridae species. BMC Evol. Biol. 13, 113.