Immune gene variation in the group living lizard *Egernia stokesii*

Sarah Kim Pearson

Bachelor of Science in Biodiversity and Conservation Honours

(Flinders University of South Australia)

A thesis submitted in fulfilment of the

requirements for the degree of

Doctor of Philosophy

School of Biological Sciences

Faculty of Science and Engineering

Flinders University of South Australia, Australia

July 2016

Table of contents

Summary..**5**

Declaration..**7**

Acknowledgements...**8**

Statements of co-authorship...**10**

Chapter 1: Introduction..**16**

Chapter 2: A method for deriving lizard DNA from their faeces

Pearson, S. K., Tobe, S. S., Fusco, D. A., Bull, C. M., and Gardner, M. G. (2015). Piles of scats for piles of DNA: deriving DNA of lizards from their faeces. Australian Journal of Zoology, 62, 507-514...**29**

Chapter 3: MHC genotyping from next-generation sequencing

Pearson, S.K., Bradford, T.M., Ansari, T.H., Bull, C.M., and Gardner, M.G. (resubmitted). MHC genotyping from next-generation sequencing: detailed methodology for the gidgee skink, *Egernia stokesii*, Transactions of the Royal Society of South Australia..**58**

Supplemental information..**93**

Chapter 4: Selection on skink MHC

Pearson, S.K., Bull, C.M., and Gardner, M.G. (resubmitted). *Egernia stokesii* (gidgee skink) MHC I positively selected sites lack concordance with HLA peptide binding regions, Immunogenetics……………………………………………..**108**

Chapter 5: Larger lizards live longer in a group living lizard

Pearson, S.K., Godfrey, S.S., Bull, C.M., and Gardner, M.G. (accepted). Larger lizards live longer in a the group living lizard *Egernia stokesii*, Australian Journal of Zoology..**151**

Supplementary material..**191**

Chapter 6: Genes and group predict mates in a lizard

Pearson, S.K., Godfrey, S. S., Schwensow, N., Bull, C.M., and Gardner, M.G. (submitted). Genes and group membership together predict reproductive pairs in the family living lizard, *Egernia stokesii*, Journal of Heredity...**194**

Supplementary material..**229**

Chapter 7: Selection on the MHC at a fine spatial scale

Pearson, S.K., Bull, C.M., and Gardner, M.G. (in preparation). Selection outweighs drift at a fine scale: lack of MHC divergence among geographically close populations of a family living lizard...**244**

Chapter 8: Discussion..**278**

Haikus..**284**

Appendix

Appendix 1: Introduction to *Egernia stokesii*..**286**

Appendix 2: Full version as published of Chapter 2..**293**

Appendix 3: *Egernia stokesii* MHC nucleotide sequences...**301**

Appendix 4: *Egernia stokesii* MHC allele evolutionary relationships.........................**311**

Appendix 5: *Egernia stokesii* MHC genotypes...**314**

Appendix 6: *Egernia stokesii* longevity..**322**

Appendix 7: *Egernia stokesii* MHC allele clustering analysis..**323**

Appendix 8: *Egernia stokesii* microsatellite genotypes..**325**

Appendix 9: Other publications...**340**

Summary

Immune gene variation has consequences for individual fitness and species persistence. Because of their role in disease resistance, mate choice, and kin recognition, the genes of the major histocompatibility complex (MHC) are important for species adaptability. Yet lizard MHC structure is unknown and few lizards have had their MHC characterised. Skinks are a basal lizard lineage. Their inclusion in MHC studies should provide valuable insights into the evolutionary ecology of the MHC.

The first aim of this thesis was to characterise MHC variation in one skink species, *E. stokesii,* among three discrete sampling sites. This work derived a reliable method of obtaining lizard DNA from their scats (Chapter 2) and comprehensive methodology for deriving MHC alleles and genotypes from next-generation DNA sequencing (Chapter 3). This represents the most comprehensive characterisation of skink MHC to date, facilitating reliable population level MHC variation inferences and comparisons.

The second aim was to analyse processes and mechanisms that generated the observed *E. stokesii* MHC variation. Although selection is purported to explain MHC variation, selection on skink MHC has been untested. Non-mammalian MHC amino acid sites of selection are usually inferred from human MHC despite often lacking concordance. This work represents the first analysis of selection on skink MHC and demonstrates positively selected skink MHC amino acid sites do not correspond to those in human MHC (Chapter 4), strengthening calls for investigations of reptilian MHC structure.

Genetic drift and gene flow influence MHC variation, not selection alone. These processes influence MHC variation on small spatial scales, yet are usually investigated at broad scales. This work provides evidence that selection on the MHC is stronger than genetic drift and constrained gene flow at a fine spatial scale (Chapter 5). Thus, adaptive and neutral genetic

variation do not always align. Both should be accounted for in species genetic diversity assessments.

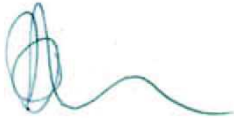
Social structure can influence MHC variation, yet studies of social structure effects on MHC variation are rare. Space use is an important aspect of social structure. In this work, almost 50 lizards sampled from three discrete sites had been captured nearly a decade before and 65% of lizards were recaptured in the same space they previously occupied (Chapter 6). Site fidelity was more likely in larger lizards.

Parasite mediated balancing selection and MHC based mate choice drive MHC variation. Comparative studies of mate choice across taxa representing a diversity of social structures should improve our currently limited understanding of social structure effects on MHC variation. Results from this study indicated mate choice was predicted by both adaptive and neutral genetic variation of potential mates. Group membership together with genetic variables predicted *E. stokesii* reproductive pairings (Chapter 7), suggesting social structure influences genetic variation in this group living species.

Future work will benefit from the characterisation of MHC variation and selection in additional skinks and other lizard taxa. The members of the *Egernia* group of lizards, which represent a diversity of social structures, are suitable candidates for future investigations of the influence of group living on the MHC.

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.



Sarah Kim Pearson

24 July 2016

Acknowledgements

So here it is finally in ink,

My four year study of a social skink!

Thanks to Dr Michael Gardner's molecular insight,

And Professor Michael Bull's ecological might,

I got up close and personal with molecular ecology,

Not to mention some fundamental biology!

Nothing was possible without dollars[1] to pay for things,

But once that was sorted, the research had wings!

First came new friends in the field[2], catching lizards and collecting poo.

Then new friends in the lab[3], helping out with all things new!

There were researchers, academics, conservationists, and citizen scientists too,

Thank you all for helping make this thesis come true!

Last but not least, thanks to Pete.

Good news - it's finally complete!

Footnotes

[1] Holsworth Wildlife Research Endowment, Field Naturalist Society of South Australia Lirabenda Endowment Fund Research Grant, Sir Mark Mitchell Research Foundation, Nature Foundation South Australia Scholarship Grant, Royal Society of SA Small Research Grants Scheme, Biology Society of SA Field Research Grant, Nature Conservation Society of South Australia Conservation Biology Grant, and the Linnean Society of NSW Joyce Vickery Scientific Research Fund are gratefully acknowledged for their financial support, and

members of the Research Services Office of Flinders University of SA are thanked for grant administration.

[2] Aaron Fenner, Jess Hacking, Sarah Mantel, David Frahm, Rhiannan Walker, Sherise Flipsons, Peter Matejcic, Jess Skewes, Thao Doan Dinh, Erin Murray, Carol Dennis, Tessa Roberts, Brian Matthews, Chloe Warden-Flood, Tony Dingwall, Jade Currie and Mel Dodd, Nature Foundation SA for access to Witchelina, John and Julie Rowe & the people of Hawker, members of Prof. Mike Bull research lab, Dominic Reppucci, and Leslie Morrison.

[3] Mina Ansari, Alison Fitch, Tessa Bradford, Amy Slender, Rebecca Dew, Steffi Graff, Olivia Davies, Nahid Shokri Bousjein, Carmel Mahr, members of School of Biology IT, Kathy Saint, and students and staff working in the SA Regional Facility for Molecular Ecology and Evolution, SA Museum and University of Adelaide

Statement of co-authorship

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

Sarah K. Pearson (Candidate), School of Biological Sciences, Flinders University, Adelaide, Australia.

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

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

Shanan S. Tobe (Author 4), Department of Chemistry and Physics, Arcadia University, Glenside, Philadelphia, USA

Diana A. Fusco (Author 5), School of Biological Sciences, Flinders University, Adelaide, Australia.

Stephanie S. Godfrey (Author 6), School of Veterinary and Life Sciences, Murdoch University, Murdoch, Australia

Tessa M. Bradford (Author 7), School of Biological Sciences, Flinders University, Adelaide, Australia.

Talat H. Ansari (Author 8), School of Biological Sciences, Flinders University, Adelaide, Australia.

Nina Schwensow (Author 9), School of Biological Sciences, University of Adelaide, Australia 5005, Australia; Institute of Evolutionary Ecology and Conservation Genomics, University of Ulm, Germany

Chapter details and relative author contributions (proportion of work undertaken) are:

<u>Chapter 2 is published as:</u>

Pearson, S. K., Tobe, S. S., Fusco, D. A., Bull, C. M., and Gardner, M. G. (2015).

Piles of scats for piles of DNA: deriving DNA of lizards from their faeces. *Australian Journal of Zoology*, 62, 507-514

The candidate was the primary author. The candidate and author 5 sampled lizard scats. The candidate conducted all laboratory works with assistance of author 5. Author 4 provided methodology advice. The candidate and authors 2 and 3 contributed to the development of main ideas and approach. The candidate and authors 2, 3, 4 and 5 contributed to refining the text.

SP 65%, ST 5%, DF 10%, CMB 10%, MG 10%

<u>Chapter 3 has been resubmitted as:</u>

Pearson, S.K., Bradford, T.M., Ansari, T.H., Bull, C.M., and Gardner, M.G

MHC genotyping from next-generation sequencing: detailed methodology for the gidgee skink, *Egernia stokesii, Transactions of the Royal Society of South Australia*, resubmitted July 2016

The candidate was the primary author. The candidate conducted field work during 2012-13 and utilised data derived from earlier field surveys undertaken by members of the research group of author 3. The candidate conducted all laboratory work with assistance and advice from authors 7 and 8. The candidate conducted all bioinformatics with assistance from author 7. The candidate as well as authors 2 and 3 contributed to developing the main ideas and approach. The candidate as well as authors 2, 3, 7 and 8 contributed to refining the text.

SP 75%, TB 5%, TA 5%, CMB 5%, MG 10%

Chapter 4 has been resubmitted as:

Pearson, S.K., Bull, C.M., and Gardner, M.G.

Egernia stokesii (gidgee skink) MHC I positively selected sites lack concordance with HLA peptide binding regions, *Immunogenetics*, resubmitted July 2016

The candidate was the primary author and conducted all analyses. The candidate conducted field work during 2012-13 and utilised data derived from earlier field surveys undertaken by members of the research group of author 3. The candidate as well as author 2 and 3 developed the main ideas and approach, and assisted with refining the text.

SP 80%, CMB 5%, MG 15%

Chapter 5 is accepted as:

Pearson, S.K., Godfrey, S.S., Bull, C.M., and Gardner, M.G.

Larger lizards live longer in the group living *Egernia stokesii*, *Australian Journal of Zoology*, accepted July 2016

The candidate was the primary author. The candidate conducted field work during 2012-13 and utilised data derived from earlier field surveys undertaken by members of the research group of author 3. The candidate conducted all spatial analyses. Author 6 conducted the generalized linear mixed model analysis using data, hypotheses and model frameworks provided by the candidate. The candidate as well as authors 2, 3 and 6 contributed to developing the main ideas and approach as well as refining the text.

SP 70%, SG 10%, CMB 10%, MG 10%

Chapter 6 is submitted as:

Pearson, S.K., Godfrey, S.S., Schwensow, N., Bull, C.M., and Gardner, M.G.

Genes and group membership predict gidgee skink (*Egernia stokesii*) reproductive pairs, *Journal of Heredity*, submitted June 2016

The candidate was the primary author and conducted all of the laboratory work. The candidate utilised data derived from earlier field surveys undertaken by members of the research group of author 3 and the PhD of author 2. Author 6 conducted the generalized linear mixed model analysis using data, hypotheses and model frameworks provided by the candidate. Author 9 assisted with methodological approaches for calculating MHC variability. The candidate as well as authors 2 and 3 developed the main ideas and approach. The candidate and authors 2, 3, 6 and 9 assisted with refining the text.

SP 75%, SG 10%, NS 5%, CMB 5%, MG 5%

Chapter 7 is in preparation as:

Pearson, S.K., Bull, C.M., and Gardner, M.G.

Selection outweighs drift at a fine scale: lack of MHC differentiation within a family living lizard across geographically close but disconnected rocky outcrops

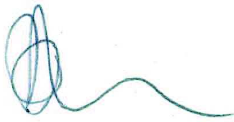
The candidate was the primary author. The candidate conducted field work during 2012-13. The candidate as well as authors 2 and 3 developed the main ideas and approach, and assisted with refining the text.

SP 80%, CMB 10%, MG 10%

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

Signed:

Candidate: Sarah Kim Pearson

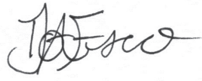


Michael G. Gardner

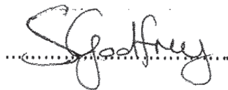
C. Michael Bull

Shanan S. Tobe

Diana A. Fusco

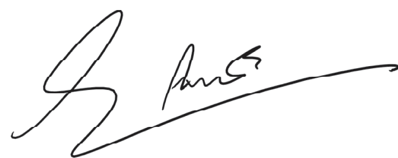


Stephanie S. Godfrey



Tessa M. Bradford

Talat H. Ansari



Nina Schwensow

Preface

This thesis contains six data chapters (Chapters 2-7) that communicate research undertaken as part of a doctoral program. Statements connecting the work of previous and subsequent chapters are provided between chapters. Each chapter was written to stand alone, therefore references are provided at the end of each chapter.

One chapter has been published (Chapter 2), one has been accepted for publication (Chapter 5), two have been revised and resubmitted for publication (Chapters 3 and 4), one has been submitted (Chapters 6), and one will be prepared for publication following thesis submission (Chapter 7). A published version of Chapter 2 is provided in the thesis Appendices.

Chapters are presented in a format according to the journal in which they have been published, submitted, or resubmitted. Within text references to other chapters of this thesis are in the format e.g. Pearson et al. Chapter 3.

Although I conducted the majority of the work, chapters have been written as manuscripts therefore the plural "we" is generally used instead of the singular “I” due to the contribution of co-authors. The Statement of co-authorship details the contribution of each author to each chapter.

The following chapter (Chapter 1) presents the structure of the thesis contained within a brief introduction that places the work of the thesis in a broad evolutionary ecology context and provides the aims and objectives of the work. Because each chapter was written as a stand alone manuscript, each contains its own introduction. Therefore, extensive detail is not provided in Chapter 1 as the reader will find these in subsequent chapters.

Chapter 1: Introduction

This thesis contains six data chapters (Chapters 2-7), plus this brief introduction and a discussion (Chapter 8), documenting aspects of adaptive genetic variation in a scincid lizard. Adaptive genetic variation has fitness consequences with implications for species persistence (Holderegger et al., 2006). Because of their role in disease resistance, mate choice, and kin recognition, the genes of the major histocompatibility complex are commonly associated with the importance of genetic variation for species adaptability and therefore survival (Sommer, 2005, Milinski, 2006, Spurgin and Richardson, 2010, Piertney and Oliver, 2006). The MHC is a multigene family that encodes molecules involved in self-nonself recognition and immune response (Piertney and Oliver, 2006, Neefjes et al., 2011).

High variability is a widely upheld feature of the MHC (Klein et al., 1993), which has been attributed to the gene family's importance for parasite resistance and individual identification. However, evidence from mammals and birds indicates that high variation is not always a feature of the MHC (Schut et al., 2011, Babik et al., 2012, Sonsthagen et al., 2014). In addition, the observation that MHC is highly variable is derived from studies that lack taxonomic breadth. Notably, MHC variation in reptiles is relatively unknown compared to that in mammals and birds (Elbers and Taylor, 2016). For example, despite representing the largest number of species of reptiles, little is known about lizard MHC structure and the MHC of very few lizards has been characterised (Elbers and Taylor, 2016, Miller et al., 2006). Thus, the assumption of high MHC variability remains relatively untested amongst lizards and limits our understanding of the processes and mechanisms that generate, and constrain, MHC variation. With more than 1600 species (Zheng and Wiens, 2016), the Scincidae is one of the most diverse families of lizards yet MHC variation has only been characterised in ten species and in all cases sample sizes were small (< 10; Elbers and Taylor,

2016). Because skinks are a basal lizard lineage, their inclusion as a study group should yield valuable insights into the evolution of the MHC.

Thus, the ***first aim*** of this thesis was to characterise MHC variation in three populations of a scincid lizard. The study species was *Egernia stokesii*, a member of the *Egernia* group of scincid lizards. An introduction to *E. stokesii* is provided in Appendix 1. Sampling the study species is the first step in characterising MHC variation. Non-invasive sampling is increasingly being used as an alternative to traditional, invasive sampling methods (Taberlet and Luikart, 1999, Beja-Pereira et al., 2009). Non-invasive sampling also provides an opportunity to increase sample sizes generated via traditional methods. Faecal pellets (scats) are widely used for non-invasive genetic sampling. Yet, the use of scats as a DNA source for reptiles was restricted to a single application in snakes and none in lizards (Jones et al., 2008). Therefore, the ***first objective*** of the thesis was to determine the suitability of using DNA derived from lizard scat as a means of increasing samples sizes arising from traditional methods. **Chapter 2** describes the development of a method to derive lizard DNA from their scats. If it could be established that reliable DNA could be derived from *E. stokesii* scats collected during field surveys, sample sizes could be increased.

MHC variation is characterised using allele and genotype frequ(Sommer et al., 2013)encies derived from DNA samples. Therefore, the ***second objective*** of this thesis was to distinguish *E. stokesii* MHC alleles and derive *E. stokesii* MHC genotypes, and then use these to describe *E. stokesii* MHC variation. *Egernia stokesii* MHC alleles and genotypes were derived from next-generation sequencing data, which although widely adopted, is lacking detailed methodology in the literature (but see Sommer et al., 2013, Lighten et al., 2014). **Chapter 3** documents the method used to identify *E. stokesii* MHC alleles and derive MHC genotypes from next-generation sequencing. Alleles and genotypes for two classes of MHC (I and II) were derived. Data derived employing the documented methodology was used in work

documented in Chapters 4, 6 and 7. By providing a level of detail that is not usually available, this work should be useful for the adoption of the approach by other researchers. In addition, this work represents the first account of the pattern of skink MHC variation in multiple individuals from multiple populations. Previous reports of MHC variation in skinks have been based on small sample sizes (Elbers and Taylor, 2016), which are not conducive to drawing conclusions about population wide variation. Furthermore, this work represents the first known published documentation of MHC II variation in skinks and the second for any lizard (reviewed in Elbers and Taylor, 2016). Documenting population level MHC I and II variation in a skink provides a foundation for comparative studies and for exploring the evolutionary and ecological mechanisms that have generated that variation.

Following the identification of patterns of MHC variation, the ***secondary aim*** of this thesis was to critically analyse the processes and mechanisms likely to have generated the observed patterns. Although MHC variation is influenced by selection, genetic drift and gene flow, selection is proposed to be the key driver of polymorphism (Bernatchez and Landry, 2003, Apanius et al., 1997). There are two key modes of selection on the MHC: 1) parasite mediated balancing selection and 2) sexual selection via MHC based mate choice (Doherty and Zinkernagel, 1975, Apanius et al., 1997, Spurgin and Richardson, 2010, Edwards and Hedrick, 1998, Penn, 2002). Evidence for selection on the MHC is purported to be strong yet we very little about selection on the MHC of lizards and selection on the MHC remains untested within the scincid lineage. As a result, it is unknown if selection on scincid MHC is similar to that at mammalian and bird MHC. Therefore, before the relative roles of gene flow, genetic drift, and selection on scincid MHC variation can be assessed, it is necessary to identify if selection is acting on the MHC.

Selection, if it acts on MHC alleles, is strongest at codons contained within the peptide binding regions in which the cellular immune response occurs (Edwards and Hedrick, 1998,

Hughes and Yeager, 1998). Human MHC is commonly used to infer these regions in non-mammalian species (Strandh et al., 2011, Wang et al., 2010, Wegner, 2008). However, positively selected codons identified in non-mammalian taxa commonly lack concordance with the human MHC (Glaberman and Caccone, 2008, Wegner, 2008). Therefore, an adaptive role can incorrectly be assigned to a codon, with implications for the reliability of analyses (Wegner, 2008). Further, MHC sites under selection are yet to be identified in skinks. Consequently it is unknown if those sites in skinks correspond with sites under selection in human MHC. Therefore, the ***third objective*** of this thesis was to test for evidence of selection on *E. stokesii* MHC alleles. Then, if evidence for selection was found, to identify the particular codons under selection and compare them to the corresponding regions in non-lizard MHC. **Chapter 4** details tests for selection on *E. stokesii* MHC and a comparison of *E. stokesii* MHC codons under selection with those in humans and other taxa.

Social structure is an important component of a species social system that has implications for MHC variation (Sommer et al., 2002, Hambuch and Lacey, 2002). Yet studies of the influence of social structure on MHC variation is limited and earlier studies of effects of social structure on the MHC show contrasting results and a lack of clear patterns (Hambuch and Lacey, 2002, Califf et al., 2013). Comparative studies of the effect of social structure on MHC are necessary to increase our understanding of the mechanisms that generate MHC variation. A recent investigation of aggregations within squamates reveals a wide range of social structures present within squamates (Gardner et al., 2015). The social structure diversity present within the *Egernia* group of scincid lizards make it a suitable group for comparative work. The study species of this thesis, *E. stokesii*, form stable family group and long-term pair bonds, characteristics present in other taxa such as primates, making it useful for future comparisons.

Space use is an important aspect of social structure, with implications for genetic variation (Campbell et al., 2008, Rossiter et al., 2012). *Egernia stokesii* site fidelity has previously been documented in one population over six consecutive years (Duffield and Bull, 2002). During field surveys conducted for this work, some individuals were captured that been sampled in previous surveys which provided an opportunity to test whether this behaviour was consistent across three isolated populations over a longer period of nearly a decade. In addition, longevity estimates were able to be assessed using recapture data. Previous assessments of site fidelity in lizards have usually been for one population over a short-medium time period and little is known regarding factors influencing space use (Effenberger and Mouton, 2007, Chapple and Keogh, 2006, but see Kerr and Bull, 2006). Therefore, the ***fourth objective*** of this thesis was to assess lizard site fidelity over a period of almost a decade and test factors that may influence recapture likelihood and space use (**Chapter 5**).

Sexual selection via MHC based mate choice is a key driver of MHC variation (Milinski, 2006, Winternitz et al., 2013). Social structure is one of the contexts on which MHC based mate choice depends, yet our understanding of the influence of social structure on MHC variation is limited. Comparative studies of mate choice across species representing a range of social structures are required and will benefit from inclusion of groups across a broad taxonomic range (Huchard and Pechouskova, 2013, Winternitz, 2015). *Egernia stokesii* exhibit high levels of social group, mate, and site fidelity (Duffield and Bull, 2002, Gardner et al., 2007, Gardner et al., 2002), which are likely to constrain mate choice. As such, a choice of mate for genetic benefits may be particularly important. Therefore, the ***fifth objective*** of this thesis was to determine if there is a genetic basis for *E. stokesii* reproductive pairings. **Chapter 6** tests the relative importance of potential predictors of *E. stokesii* reproductive pairings. Non-genetic variables are rarely incorporated in studies of the genetic basis of mate choice. However, because of the strong social group structure of *E. stokesii*, the

role of group membership was investigated as a potential predictor of pairings, as well as adaptive and neutral genetic variables. In addition, although most previous studies of MHC based choice only used one class of MHC (i.e. either I or II), these regions have different functions therefore both were included in this work.

MHC variation is not influenced by selection alone. Genetic drift and gene flow also have a role (Bernatchez and Landry, 2003, Dionne et al., 2008, Sutton et al., 2011). To further our understanding of the relative influence of these processes, it is useful to compare variation at the MHC variation with that at neutral genetic loci (for example microsatellites) (Bernatchez and Landry, 2003). Although gene flow, genetic drift, and selection can influence MHC variation on a small spatial scale, investigations of the influence of these processes on MHC variation are usually undertaken at a broad scale (Bichet et al., 2015, Strand et al., 2012, Zeisset and Beebee, 2014). In addition, differences exists in the relative influence of each of the aforementioned processes on MHC variation (Boyce et al., 1997, Loiseau et al., 2009, Rico et al., 2015). Therefore, the ***sixth objective*** of this thesis was to determine if the effects of selection outweigh the effects of genetic drift and limited gene flow on *E. stokesii* MHC variation among populations at a fine spatial scale. In **Chapter 7**, *E. stokesii* MHC and microsatellite variation is compared within and among populations as a means of assessing the relative influence of these three processes on MHC variation. This work is novel in its quantification of MHC differentiation within and among geographically close but demographically separated populations.

Chapter 8 presents a short discussion and conclusion summarising the contribution of the work contained in Chapters 2 – 7 to the field and proposed areas for further research.

References

APANIUS, V., PENN, D., SLEV, P. R., RUFF, L. R. & POTTS, W. K. 1997. The nature of selection on the major histocompatibility complex. *Critical Reviews in Immunology,* 17**,** 179-224.

BABIK, W., KAWAŁKO, A., WÓJCIK, J. M. & RADWAN, J. 2012. Low Major Histocompatibility Complex Class I (MHC I) Variation in the European Bison (*Bison bonasus*). *Journal of Heredity,* 103**,** 349-359.

BEJA-PEREIRA, A., OLIVEIRA, R., ALVES, P. C., SCHWARZ, M. K. & LUIKART, G. 2009. Advancing ecological understanding through technological transformations in noninvasive genetics. *Molecular Ecology Resources,* 9**,** 1279-1301.

BERNATCHEZ, L. & LANDRY, C. 2003. MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? *Journal of Evolutionary Biology,* 16**,** 363-377.

BICHET, C., MOODLEY, Y., PENN, D. J., SORCI, G. & GARNIER, S. 2015. Genetic structure in insular and mainland populations of house sparrows (*Passer domesticus*) and their hemosporidian parasites. *Ecology and Evolution,* 5**,** 1639-1652.

BOYCE, W. M., HEDRICK, P. W., MUGGLI-COCKETT, N. E., KALINOWSKI, S., PENEDO, M. C. T. & RAMEY, R. R. 1997. Genetic variation of major histocompatibility complex and microsatellite loci: a comparison in bighorn sheep. *Genetics,* 145**,** 421-433.

CALIFF, K. J., RATZLOFF, E. K., WAGNER, A. P., HOLEKAMP, K. E. & WILLIAMS, B. L. 2013. Forces shaping major histocompatibility complex evolution in two hyena species. *Journal of Mammalogy,* 94.

CAMPBELL, R., GALES, N., LENTO, G. & BAKER, C. 2008. Islands in the sea: extreme female natal site fidelity in the Australian sea lion, *Neophoca cinerea*. *Biology letters,* 4**,** 139-142.

CHAPPLE, D. G. & KEOGH, S. J. 2006. Group structure and stability in social aggregations of White's skink, *Egernia whitii*. *Ethology,* 112**,** 247-257.

DIONNE, M., CARON, F., DODSON, J. J. & BERNATCHEZ, L. 2008. Landscape genetics and hierarchical genetic structure in Atlantic salmon: the interaction of gene flow and local adaptation. *Molecular Ecology,* 17**,** 2382-2396.

DOHERTY, P., C. & ZINKERNAGEL, R., M. 1975. Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. *Nature,* 256**,** 50-52.

DUFFIELD, G. A. & BULL, C. M. 2002. Stable aggregations in an Australian lizard, *Egernia stokesii*. *Naturwissenschaften,* 89**,** 424-427.

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

EFFENBERGER, E. & MOUTON, P. L. F. N. 2007. Space use in a multi-male group of the group-living lizard. *Journal of Zoology,* 272**,** 202-208.

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

GARDNER, M. G., BULL, C. M. & COOPER, S. J. B. 2002. High levels of genetic monogamy in the group-living Australian lizard *Egernia stokesii*. *Molecular Ecology,* 11**,** 1787-1794.

GARDNER, M. G., BULL, C. M., FENNER, A., MURRAY, K. & DONNELLAN, S. C. 2007. Consistent social structure within aggregations of the Australian lizard, *Egernia stokesii* across seven disconnected rocky outcrops. *Journal of Ethology,* 25**,** 263-270.

GARDNER, M. G., PEARSON, S. K., JOHNSTON, G. R. & SCHWARZ, M. P. 2015. Group living in squamate reptiles: a review for stable aggregations. *Biological Reviews*.

GLABERMAN, S. & CACCONE, A. 2008. Species-specific evolution of class I MHC genes in iguanas (Order: Squamata; Subfamily: Iguaninae). *Immunogenetics,* 60**,** 371-382.

HAMBUCH, T. M. & LACEY, E. A. 2002. Enhanced selection for MHC diversity in social tuco-tucos. *Evolution,* 56**,** 841-845.

HOLDEREGGER, R., KAMM, U. & GUGERLI, F. 2006. Adaptive vs. neutral genetic diversity: implications for landscape genetics. *Landscape Ecology,* 21**,** 797-807.

HUCHARD, E. & PECHOUSKOVA, E. 2013. The major histocompatibility complex and primate behavioral ecology: new tools and future questions. *International Journal for Parasitology,* 35**,** 11-31.

HUGHES, A. L. & YEAGER, M. 1998. Natural selection at major histocompatibilty complex loci of vertebrates. *Annual Review of Genetics,* 321**,** 415-435.

JONES, R., CABLE, J. & BRUFORD, M. W. 2008. An evaluation of non-invasive sampling for genetic analysis in northern European reptiles. *Herpetological Journal,* 18**,** 32-39.

KERR, G. D. & BULL, C. M. 2006. Exclusive core areas in overlapping ranges of the sleepy lizard, *Tiliqua rugosa*. *Behavioral Ecology,* 17**,** 380-391.

KLEIN, J., SATTA, Y. & O'HUIGIN, C. 1993. The molecular descent of the major histocompatibility complex. *Annual Review of Immunology,* 11**,** 269-295.

LIGHTEN, J., OOSTERHOUT, C., PATERSON, I. G., MCMULLAN, M. & BENTZEN, P. 2014. 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,* 14**,** 753-767.

LOISEAU, C., RICHARD, M., GARNIER, S., CHASTEL, O., JULLIARD, R., ZOOROB, R. & SORCI, G. 2009. Diversifying selection on MHC class I in the house sparrow (*Passer domesticus*). *Molecular Ecology,* 18**,** 1331-1340.

MILINSKI, M. 2006. The major histocompatibility complex, sexual selection, and mate choice. *Annual Review of Ecology, Evolution and Systematics,* 37**,** 159-186.

MILLER, H. C., BELOV, K. & DAUGHERTY, C. H. 2006. MHC class I genes in the tuatara (Sphenodon spp.): evolution of the MHC in an ancient reptilian order. *Molecular biology and evolution,* 23**,** 949-956.

NEEFJES, J., JONGSMA, M. L., 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.

PENN, D. J. 2002. The scent of genetic compatibility: sexual selection and the major histocompatibility complex. *Ethology,* 108**,** 1-21.

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

RICO, Y., MORRIS-POCOCK, J., ZIGOURIS, J., NOCERA, J. J. & KYLE, C. J. 2015. Lack of Spatial Immunogenetic Structure among Wolverine (*Gulo gulo*) Populations Suggestive of Broad Scale Balancing Selection. *PloS one,* 10**,** e0140170.

ROSSITER, S. J., ZUBAID, A., MOHD-ADNAN, A., STRUEBIG, M. J., KUNZ, T. H., GOPAL, S., PETIT, E. J. & KINGSTON, T. 2012. Social organization and genetic structure: insights from codistributed bat populations. *Molecular Ecology,* 21**,** 647-661.

SCHUT, E., AGUILAR, J., MERINO, S., MAGRATH, M., KOMDEUR, J. & WESTERDAHL, H. 2011. Characterization of MHC-I in the blue tit (*Cyanistes caeruleus*) reveals low levels of genetic diversity and trans-population evolution across European populations. *Immunogenetics,* 63**,** 531-542.

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

SOMMER, S., COURTIOL, A. & MAZZONI, C. J. 2013. MHC genotyping of non-model organisms using next-generation sequencing: a new methodology to deal with artefacts and allelic dropout. *BMC Genomics,* 14.

SOMMER, S., SCHWAB, D. & GANZHORN, J. 2002. MHC diversity of endemic Malagasy rodents in relation to geographic range and social system. *Behavioral Ecology and Sociobiology,* 51**,** 214-221.

SONSTHAGEN, S., FALES, K., JAY, C., SAGE, G. & TALBOT, S. 2014. Spatial variation and low diversity in the major histocompatibility complex in walrus (*Odobenus rosmarus*). *Polar Biology,* 37**,** 497-506.

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**,** 979-988.

STRAND, T. M., SEGELBACHER, G., QUINTELA, M., XIAO, L., AXELSSON, T. & HÖGLUND, J. 2012. Can balancing selection on MHC loci counteract genetic drift in small fragmented populations of black grouse? *Ecology and Evolution,* 2**,** 341-353.

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**,** 653-666.

SUTTON, J. T., NAKAGAWA, S., ROBERTSON, B. C. & JAMIESON, I. G. 2011. Disentangling the roles of natural selection and genetic drift in shaping variation at MHC immunity genes. *Molecular Ecology,* 20**,** 4408-4420.

TABERLET, P. & LUIKART, G. 1999. Non-invasive genetic sampling and individual identification. *Biological Journal of the Linnean Society,* 68**,** 41-55.

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.

WEGNER, K. M. 2008. Historical and contemporary selection of teleost MHC genes: did we leave the past behind? *Journal of Fish Biology,* 73**,** 2110-2132.

WINTERNITZ, J. C. 2015. Examining the evidence for major histocompatibility complex-dependent mate selection in humans and nonhuman primates. *Research and Reports in Biology,* 6**,** 73-88.

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 of London, Series B: Biological Sciences,* 280**,** 20131605.

ZEISSET, I. & BEEBEE, T. J. 2014. Drift rather than selection dominates MHC class II allelic diversity patterns at the biogeographical range scale in natterjack toads *Bufo calamita*. *PloS one,* 9**,** e100176.

ZHENG, Y. & WIENS, J. J. 2016. Combining phylogenomic and supermatrix approaches, and a time-calibrated phylogeny for squamate reptiles (lizards and snakes) based on 52 genes and 4162 species. *Molecular phylogenetics and evolution,* 94**,** 537-547.

Chapter 2: A method for deriving lizard DNA from their faeces

Published as:

Pearson, S. K., Tobe, S. S., Fusco, D. A., Bull, C. M., and Gardner, M. G. (2015). Piles of scats for piles of DNA: deriving DNA of lizards from their faeces. *Australian Journal of Zoology*, 62, 507-514 (**see Appendix 2 for the published version**)

Sampling of the study species is the first step in characterising MHC variation. Non-invasive methods are increasingly used as an alternative to, or to complement, traditional, invasive sampling methods. Although widely used in other animal groups such as mammals, the use of faecal pellets (scats) as a DNA source for genetic investigations of reptiles was constrained to a single published study in snakes and none in lizards.

Chapter 2 describes the development of a reliable method of deriving lizard DNA from their scats. The method is cost intensive due to the low quality and quantity of DNA typically present in scats. The method was not applied in this work because funding was not obtained. Therefore, the information here is not used in subsequent chapters.

Piles of scats for piles of DNA: Deriving DNA of lizards from their faeces, *Australian Journal of Zoology*, 62(6) 507-514

Pearson, S.K.[A,B], Tobe, S.S.[A], Fusco, D.A.[A], Bull, C.M.[A], Gardner, M.G.[A,C]

[A]School of Biological Sciences, Flinders University of South Australia, Bedford Park 5042, Australia

[C]Evolutionary Biology Unit, South Australian Museum, North Terrace, Adelaide 5000, Australia.

[B]Corresponding author. Email: sarah.pearson@flinders.edu.au

Short summary

Non-invasive genetic sampling using scats has a well established role in conservation biology; but has rarely been applied to reptiles. We evaluated various storage and DNA extraction methods and identified a reliable method of deriving genotypes and sequences from gidgee skink, *Egernia stokesii,* scats. Results highlight the opportunity for using scat-derived-DNA in lizard studies, particularly for species that deposit scats in piles.

Running header A method for deriving lizard DNA from their faeces

Abstract

Non-invasive genetic sampling using scats has a well established role in conservation biology, but has rarely been applied to reptiles. Using scats from captive and wild *Egernia stokesii* (Squamata, Scincidae) we evaluated two storage and six DNA extraction methods and the reliability of subsequent genotype and sequence data. Accurate genotype and sequence data were obtained from frozen and dried captive lizard scat DNA extracted using a QIAamp ® DNA Stool Mini Kit and a modified Gentra ®Puregene ® method, but success rates were reduced for wild lizard scats. Wild *E. stokesii* eat more plants than their captive counterparts; possibly resulting in scat DNA extracts containing plant compounds that inhibit PCR-amplifications. Notably, reliable genotypes and sequences were obtained from wild *E. stokesii* scat DNA extracted using a Qiagen DNeasy ® Plant Mini Kit, a method designed to remove plant inhibitory compounds. Results highlight the opportunity for using scat derived DNA in lizard studies, particularly for species that deposit scats in piles.

Additional keywords: PCR inhibition; lizard scat; non-invasive sampling; DNA extraction; faecal DNA; scat piles; microsatellites; *Egernia* group.

Introduction

Faecal pellets (scats) are a widely used source of non-invasive genetic sampling of animals (Taberlet *et al.* 1999; Beja-Pereira *et al.* 2009) providing information on species identification and distribution (Harris *et al.* 2010), individual identity (Brinkman *et al.* 2010), and relatedness between individuals within a population (Steinglein *et al.* 2011). At the population level, genetic data derived from scats have shed light on population history, structure, and genetic diversity (Frantz *et al.* 2003; Iyengar *et al.* 2005). Scat collection is less intrusive than most traditional methods used to extract DNA from collected tissue. It causes less stress and less disruption of normal behaviour to the study individuals (Taberlet and Luikart 1999), and may be less demanding in terms of field time, collection effort, equipment and costs (Solberg *et al.* 2006; Vynne *et al.* 2012). For secretive species that are hard to locate or catch, non-invasive genetic sampling using scats may be the only viable option (e.g. Alacs *et al.* 2003). For threatened species, the use of scats can provide a means to overcome collecting permit restrictions.

However, there are specific problems in deriving donor DNA from scats. as they may contain many other components including exotic DNA from food remains and gut parasites (Morin *et al.* 2001; Broquet *et al.* 2007; Marrero *et al.* 2009). Extraction and amplification of the DNA of the scatting individual may be inhibited by this accompanying material (Marrero *et al.* 2009; Panasci *et al.* 2011). Additionally, sample age and environmental conditions since the time of scat deposition can result in DNA degradation (Piggott 2004; Murphy *et al.* 2007; Panasci *et al.* 2011). Despite the challenges posed by the low quantity and quality of DNA, the use of scats for genetic data now has a well established place in ecological studies.

In studies of reptiles, scats have been widely used to derive non-genetic data on diet (Barrows 2006; Germano *et al.* 2007; Pavey *et al.* 2010), species distribution and abundance (Turner and Medica 1982), recognition and communication (Bull *et al.* 1999; Wilgers and Horne

2009), parasite infections (Fenner and Bull 2008; Smith *et al.* 2009), and territoriality (Wilgers and Horne 2009). However, despite its wide application in studies of mammals, the use of scats as a DNA source for genetic studies of reptiles is limited to only a single published study in snakes (Jones *et al.* 2008), with none in lizards. One explanation may be that lizard scats may contain fewer cells from the scatting individual and lower DNA yields than mammal scats. In mammal scats, donor DNA is found in a mucous layer of colorectal epithelial cells that have collected on the surface of the scat as it moves through the digestive tract (Waits and Paetkau 2005; Ball *et al.* 2007; Herbert *et al.* 2011). Amplification of target DNA is more successful using the outer coating of scats than material from inside the scats (Wehausen *et al.* 2004), and the scat coating is regularly targeted for DNA extraction in mammal studies (Piggott and Taylor 2003; Ball *et al.* 2007; Herbert *et al.* 2011). Lizard scats appear to have a reduced mucosal coating. Despite the potential challenges, the scarcity of studies using DNA derived from lizard scats highlights an opportunity to develop this non-invasive genetic sampling method for this animal group. An important component of this process is to determine the best methods for storing scat samples, and extracting DNA from them, to maximise the yield and quality of DNA for genotyping and sequencing analyses.

Published studies in which DNA has been extracted from scats reveal a range of methods for scat storage. These include freezing (Nagy 2010), drying (Nsubuga *et al.* 2004), and storage in a buffer (Frantz *et al.* 2003). Methods to extract DNA from scats also vary. For example there are off-the-shelf scat DNA extraction kits (Steinglein *et al.* 2011; Watts *et al.* 2011), or scats can be treated using blood or tissue DNA extraction kits (Brinkman *et al.* 2010; Harris *et al.* 2010). It appears that no one method of both storage and extraction suits all species, and Valiere *et al.* (2007) and Renan *et al.* (2012) recommended a pilot study be undertaken to identify optimal methods for each new study species. We investigated alternative methods for storing scats and deriving DNA from them for an Australian scincid lizard, *Egernia stokesii*.

Our objectives were to: 1) identify optimal *E. stokesii* scat storage and DNA extraction methods; and 2) assess the reliability of DNA genotypes and sequences derived from *E. stokesii* scats using these methods. Once developed, these methods could complement traditional invasive sampling methods in this and other lizard species.

Materials and methods

Study species

Egernia stokesii (gidgee skink, J.E. Gray, 1845) is a large (180mm snout-vent length, Cogger (1983)), long-living, viviparous skink (Duffield and Bull 2002) widely distributed across eastern and central areas of semi-arid Australia. *Egernia stokesii* individuals live in stable family groups (Gardner *et al.* 2001a; Duffield and Bull 2002); have high levels of genetic monogamy (Gardner *et al.* 2002) and limited dispersal (Gardner *et al.* 2001b). They produce scats upon rock platforms immediately outside of the rocky crevices in which they reside, resulting in distinctive scat piles or deposits (Duffield and Bull 1998). Using olfaction, individuals can discriminate between scats from familiar group and non-group members, suggesting scat piles play an important role in social group cohesion in this species (Bull *et al.* 1999). The use of scat derived DNA in this and similarly scat piling lizard species could provide quick access to the DNA of most group members, without the time and effort required to capture the lizards for tissue samples. Additionally, collecting scat may provide a more complete genetic sampling of social groups as some individuals may not be caught.

Scat sampling

Scats were sampled from two sources. First we used captive *E. stokesii* housed at the Flinders University of South Australia. These included individuals originally captured near Hawker (31°54′S; 138°25′E) in the southern Flinders Ranges, South Australia, during the summers of 1993-1998 and their progeny (Main and Bull 1996; Lanham and Bull 2004; Arida and Bull

2008). We randomly selected nine *E. stokesii* individuals and kept them in nine separate cages so scats could be confidently assigned to an individual. Each individual was housed in a cage (40cm high x 40cm wide x 50cm deep) in a room with a temperature of 25ºC (± 2ºC), with ceiling lights on for 12 hours a day, and heat lamps on for 6 hours a day. Scats were collected twice weekly for four weeks (total 128 scats, average 14.22 scats per lizard, SE ± 2.13). Second, scats were collected within an estimated four hours of defecation during field surveys of three *E. stokesii* populations near Hawker conducted between September 2012 and March 2013 (409 scats). Scat freshness was assessed based on colour, moisture, compaction, by the presence of a uric acid spot, and by comparison with scats of known age from the captive colony. In addition, some lizards captured during surveys defecated during handling, ensuring complete freshness of the scat samples.

In each case, scats were collected using forceps that had been cleaned in 90% ethanol between each collection, and were stored using alternative methods as described below. The diet of captive and wild *E. stokesii* differed. Captives of all ages were fed a mix of boiled eggs, fruits and vegetables, and reptile supplement, while adult wild *E. stokesii* feed largely on plant material (Duffield and Bull 1998). Based on their size, we deduced that the wild scats used in this study were from adult *E. stokesii* (Duffield and Bull 1998); this was confirmed by visual inspection which showed a largely plant derived content of wild scats.

Positive controls

To confirm that DNA derived from scats accurately represented the DNA of the scatting individual we collected blood samples from some individuals as an alternative source for DNA characterisation. Blood (up to 0.5 mL) was taken from the caudal vein of the nine isolated captive individuals and from 29 wild individuals that produced a scat while captured. Blood was stored on Whatman FTA ® Elute for later DNA extraction. We used established

methods for deriving mitochondrial DNA (mtDNA) sequences and microsatellite DNA genotypes from FTA stored *E. stokesii* blood (Gardner *et al.* 2007).

Scat storage

We compared two methods of storage for the captive lizard scats. Scats were either frozen at -20°C (Frantz *et al.* 2003) (27 scats; all from the lab colony), or dried (72 scats; 54 lab colony, 18 field). Scats to be dried were sprayed with 90% ethanol and then stored on silica beads (hereafter termed dried; modified from Roeder *et al.* 2004) and kept at room temperature until DNA extraction. Samples from the field were all stored dried as this method was considered more practical for sampling in extreme conditions and away from amenities.

DNA extraction of captive lizard scats

We trialled six DNA extraction methods using 81 scats (27 frozen, 54 dried) from captive lizards (Table 1): 1) QIAamp ® DNA Stool Mini Kit (QIAGEN ®, Catalogue 51504); 2) ISOLATE Fecal DNA Kit (Bioline ®, Catalogue BIO-52037); 3) a standard Chelex ® 100 extraction; 4) Chelex ® 100 without boiling (adopted from Casquet *et al.* 2011); 5) a modified Gentra ®Puregene ® (Gentra Systems) method; and 6) a direct PCR method. The first two methods were off the shelf kits specifically derived for scat samples. The next three were standard kit methods used for tissue or blood samples. The last method involved amplification without first extracting or purifying the DNA and allows for maximum recovery of sample, but can suffer from inhibitors that are normally removed during extraction. The two kits (QIAamp and ISOLATE) were used according to manufacturer guidelines. The modified Gentra ® Puregene ® method involved immersion of the whole scat in SLP buffer (500mM Tris-HCl pH 8.0, 50mM EDTA, 10mM NaCl, (modified from Deuter *et al.* 1995) followed by agitation on a rotor wheel for one hour, protein precipitation with

Proteinase K, DNA precipitation with ammonium acetate and isopropanol, ethanol wash, and DNA hydration in TLE buffer.

For the kit extractions, scats were selected based on recommended weight ranges where possible; *E. stokesii* scats ranged in weight from about 10 – 900mg (average 199.60mg, SE ± 16.63), therefore total weight may have been outside the recommended range (180 – 220mg QIAamp ® DNA Stool Mini Kit, up to 150mg ISOLATE Fecal DNA Kit). Where a scat was large enough, a surface scrape of the scat was used in kit extractions as this is where most of the donor individual's DNA is expected to be found. Alternatively a segment of the scat, or the entire scat was used, depending upon the protocol. In all methods, filtered pipette tips were used to minimise contamination and negative extraction controls (scat material was not added to the extraction) were used to assess contamination. Separate laboratories were used for extraction, amplification preparation and reaction. Replicate scat extractions are sometimes recommended (Taberlet *et al.* 1999) but this was not possible as a single extraction often required the whole scat to be used.

DNA amplification captive lizards

Initially, mtDNA was targeted in DNA amplification trials because cells contain more mtDNA than nuclear DNA (nuDNA), suggesting that if mtDNA could not be amplified then targeting nuDNA was likely to be futile (Taberlet *et al.* 1999). However, amplification success may be increased for smaller DNA markers (Broquet *et al.* 2007). Because only larger mtDNA genetic markers (~800bp) were currently available for *E. stokesii*, and because scat DNA may be of low quantity and quality (Navidi *et al.* 1992; Taberlet *et al.* 1996), we developed genetic markers to amplify ~200 bp of the mtDNA ND4 gene. Three primer pairs were designed in Geneious 5.6 (Biomatters Ltd 2012) based on a consensus sequence derived from 159 existing *E. stokesii* mtDNA sequences. These primer pairs were trialled in DNA derived from *E. stokesii* blood; forward primer M1544 (5'-

TATGAACGCACCCATAGCCG-3') and reverse primer M1545 (5'-GCTGCTGTTAGAAGAGTGCC-3') were selected for this study.

For mtDNA only 1:5 and 1:50 dilutions were trialled. A dilution of 1:5 has previously been successful for DNA from blood in this species, but we considered that overcoming inhibitors in scat DNA may require increased dilution (Monteiro *et al.* 1997; Ball *et al.* 2007; Arandjelovic *et al.* 2009). Polymerase chain reaction (PCR) amplifications were conducted at a total volume of 25-µL consisting of 1 x PCR Gold Buffer (Applied Biosystems), 0.20 µM of each primer, 0.80 mM dNTPs, 2 mM MgCl2, 0.5 U AmpliTaq ® Gold DNA polymerase (Applied Biosystems), 2 µL of extracted DNA, and PCR grade water. The cycling conditions were nine minutes at 95 ºC, 34 cycles of 45 seconds at 94 ºC, 45 seconds at 60 ºC, one minute at 72 ºC, and a final elongation step of ten minutes at 72 ºC followed by 30 seconds at 25 ºC. To ensure that non-amplification was due to the test procedure rather than a failure of the PCR, and that positive results were not the result of contamination, one PCR positive (DNA extracted from blood and known to amplify) and two PCR negatives (TLE buffer and the negative DNA extraction) controls were used in each PCR. Neat DNA, from which mtDNA PCR-amplifications were successful, was quantified using Qubit ® 2.0 Fluorometer (Life Technologies Corporation, CA); although we acknowledge that the DNA measured may have included both target and non-target DNA.

Where mtDNA amplification was successful, as determined by the presence of a band on an agarose gel, amplification trials continued using a previously developed species specific microsatellite genetic marker (Est 1, Gardner *et al.* 1999). For nuDNA, a QIAGEN ® Multiplex PCR Kit (QIAGEN ®, Catalogue 206143) was used. Each 10-µL uniplex reaction mix contained 0.10 x QIAGEN Multiplex PCR Master Mix, 0.25 µM of each primer, 0.50 x Q-solution, 2 µL of extracted DNA, and RNAse Free Water. The cycling conditions were 15 minutes at 95 ºC, 35 cycles of 30 seconds at 94 ºC, 90 seconds at 57 ºC, one minute at 72 ºC,

and a final elongation step of 30 minutes at 60 ºC followed by 30 seconds at 25 ºC. Amplification success of the Est 1 locus was determined by the presence of a band on an agarose gel. A dilution of 1:50 has previously been successful for nuDNA from blood in this species. Because increased dilutions may be required to reduce the effect of inhibitors, if amplification was not successful for 1:50 dilutions, a range of DNA dilutions (neat, 1:5, 1:10, 1:100, 1:500, 1:1000) were then trialled. If Est 1 failed to amplify for any dilution, the extraction method was deemed unsuccessful for nuDNA.

DNA extraction and amplification in scats from wild lizards

Although the QIAamp ® DNA Stool Mini Kit and modified Gentra ®Puregene ® method were successful in preliminary trials using captive lizard scats (see results), the modified Gentra ®Puregene ® method had a lower per sample cost, therefore we chose that method for validation using six wild scats; positive and negative controls were used as described for captive scats above. None of the six wild scat DNA extractions amplified for mtDNA (results not shown). We considered diet differences between captive and wild *E. stokesii* may explain differences in amplification success rates. Earlier studies have suggested that diet derived inhibitors in scats may reduce both DNA extraction yields and amplification success (Kohn and Wayne 1997; Herbert *et al.* 2011; Panasci *et al.* 2011; Monroe *et al.* 2013). This problem could be particularly relevant for omnivorous or herbivorous lizards due to the presence of polysaccharides and polyphenols found in plants (Marrero *et al.* 2009; Panasci *et al.* 2011).

To investigate the possible presence of PCR inhibitors, DNA was extracted (using the modified Gentra ®Puregene ® method) from a further six wild lizard scats. To test for PCR inhibition, one of the scat extractions was replicated in the PCR, once with only DNA extracted from the scat, and once with the scat DNA plus 2 µL of a positive control. We could infer inhibitors were likely to be preventing amplification if both reactions failed. In an effort to reduce the impact of potential inhibitors, a subset of extracted DNA from each of the

six wild scats was purified using Microcon Ultracel YM-100 filters. Purified extractions were then assessed for mtDNA amplification success using the reaction mix and conditions outlined above. One sample was replicated in this PCR, with one replicate spiked with control DNA to directly assess the effect of the inhibitor clean-up process (i.e. the same sample was used as in the earlier PCR).

Given the low success rates of the modified Gentra ®Puregene ® method on wild lizard scat (see results), and the additional cost and effort associated with clean-up, a Qiagen DNeasy ® Plant Mini Kit was trialled for removing inhibitors. DNA was extracted from a further six wild scats according to the manufacturer instructions except initial scat sample disruption and homogenisation was avoided. Instead, the scat was left intact and, where required due to the size of the scat, additional Buffer AP1 and RNAase A stock solution (100 mg/ml) were used to ensure scats were fully immersed prior to incubation.

Validation via genotyping and sequencing

For captive samples, where an extraction method was successful, both scat and blood samples from a subset of two lizards were sequenced for the mtDNA and genotyped for seven previously described polymorphic microsatellite loci (Est 1, Est 4, Est 8, Est 13, (Gardner *et al.* 1999); TrL 28, TrL 29, TrL 35, (Gardner *et al.* 2008)) in PCR-amplifications according to the reaction mix and conditions described above except that reactions were performed in two multiplex reactions rather than uniplex (multiplex 1: Est 1, Est 4, Est 8, Est 13; multiplex 2: TrL 28, TrL 29, TrL 35). For wild samples, scat DNA of six lizards that defaecated during handling was extracted using the Qiagen DNeasy ® Plant Mini Kit, and blood DNA from the same lizards, extracted using the Whatman FTA ® Elute, were similarly genotyped and sequenced. Prior to sequencing, mtDNA PCR products were purified using multiscreen PCR filter plates (Millipore Billerica, MA) to remove unincorporated primers and dNTPs. Sequence reactions were prepared using a BigDYE Terminator Cycle Sequencing Kit v3.1

(Applied Biosystems) following manufacturer recommendations, using the same primers as those used in PCR amplification. The cycling conditions were three minutes at 96 ºC, 30 cycles of 30 seconds at 96 ºC, 15 seconds at 50 ºC, four minutes at 60 ºC, and a final elongation step of three minutes at 25 ºC followed by 30 seconds at 25 ºC. Sequence products were purified using multiscreen PCR filter plates (Millipore Billerica, MA) prior to submission of DNA to the Australian Genome Research Facility (AGRF) for capillary separation on an ABI Prism 3730*xl* 96-capillary sequencer. The resulting sequences were compared against data on GenBank, to confirm species identification, using the Basic Local Alignment Search Tool (BLAST) available at http://blast.ncbi.nlm.nih.gov/Blast.cgi. For nuDNA, we compared genotypes derived from both blood and scat samples from the same individuals. PCR products were analysed by capillary separation on an AB3730 DNA analyser (Australian Genome Research Facility) and resulting fragments were scored using GeneMapper ® (Applied Biosystems). Although recommended (Taberlet *et al.* 1999; Valiere *et al.* 2007), we did not perform replicate PCRs for mtDNA or nuDNA in initial trials; instead we used blood DNA samples from the same individuals as a positive control. At this stage we were interested in determining if a sequence and genotype could be derived from scat DNA and, if they matched those derived from blood derived DNA.

Assessment of genotyping reliability

Once we identified a method for deriving genotypes from wild *E. stokesii* scats (see results), we assessed genotyping reliability using three independent PCRs (adopted from Panasci *et al.* 2011; Stenglein *et al.* 2011). DNA amplification and genotyping were undertaken as uniplexes (see "validation via genotyping and sequencing" above).

Results

We used 81captive *E. stokesii* scats in preliminary trials (27 frozen, 54 dried, Table 1). Both frozen and dried samples amplified for both mtDNA and nuDNA, and the amplification success rates for each method are given in Table 1. Of the six extraction methods trialled using captive lizard scats, the QIAamp ® DNA Stool Mini Kit, ISOLATE Fecal DNA Kit, and modified Gentra ®Puregene ® methods were successful for mtDNA (Table 1). The Chelex ® 100 and direct PCR methods failed to amplify mtDNA, and therefore were not trialled for nuDNA. Both the 1:5 and 1:50 DNA dilutions were successful for mtDNA, while for nuDNA, neat DNA was the most successful (results not shown). The QIAamp ® DNA Stool Mini Kit and modified Gentra ®Puregene ® method were further tested for reliability of sequencing and genotyping using captive *E. stokesii* scats. Of twenty captive lizard scat DNA sequences derived using the QIAamp ® DNA Stool Mini Kit (n= 17) and modified Gentra ®Puregene ® method (n = 3) assessed in BLAST, 80% (n= 16) were identified as *E. stokesii*, 15% (n=3) as *Egernia sp,* and one sequence (5%) was too short to provide meaningful results. For nuDNA, all seven microsatellite loci were successfully derived from captive lizard scat DNA extracted using both the QIAamp ® DNA Stool Mini Kit and the modified Gentra ®Puregene ® method, and all scat derived genotypes matched those derived from blood.

Based on trials using captive lizard scats, the modified Gentra ®Puregene ® method was initially chosen for use with six wild lizard scat samples; mtDNA amplification was unsuccessful. In subsequent trials using a further six wild lizard scat samples, the PCR positive control DNA on its own was successfully amplified, but the wild lizard scat sample spiked with the positive control DNA failed to amplify, suggesting the presence of inhibitors. Following application of a purification method, an additional three wild scat DNA samples amplified for mtDNA. Notably, the elute DNA of six of the 12 wild lizard scat extractions

using the modified Gentra ®Puregene ® method ranged from a light tea colour to muddy brown; whereas the elute DNA of all captive lizard scat extractions was clear.

Six wild lizard scat samples extracted using the Qiagen DNeasy ® Plant Mini Kit were successfully sequenced and genotyped. All six samples were identified as *E. stokesii* using BLAST analyses. All seven microsatellite DNA loci could be scored and the resulting genotypes matched those derived from blood. The elute DNA was clear for all wild lizard scats extracted using the Qiagen DNeasy ® Plant Mini Kit. Successful mtDNA PCR-amplification quantifications are available as Supplementary Material on the Journal website. The reliability of genotypes derived from DNA extracted using the Qiagen DNeasy ® Plant Mini Kit was further assessed for seven loci in five wild *E. stokesii* scat samples. All loci amplified in all replicates for all samples, except TrL 35 which failed in all replicates for one sample. Matching heterozygotes were observed in all replicates for most samples, with three exceptions. Firstly, all replicates for Est 1 and TrL 28 in one sample showed matching homozygotes. Secondly, allelic dropout was evident for Est 1 in one sample which showed two matching heterozygotes and one homozygote. Lastly, for TrL 28 in one sample, two replicates showed matching homozygotes, while one showed a heterozygote, suggesting either allelic dropout in two replicates or a false allele in one replicate. The allelic dropout and false allele incidences represent an overall genotyping error rate of 2%.

Discussion

We have identified a reliable method for deriving DNA sequences and genotypes from wild *E. stokesii* scat samples. Genotypes and sequences were successfully derived from DNA extracted from field collected scats using a Qiagen DNeasy ® Plant Mini Kit. The overall reliability of genotypes derived using this method was supported by a low genotyping error rate. Adoption of this method would complement traditional capture-mark-recapture methods for estimating local abundance of *E. stokesii* and other lizard species, and for estimating

genetic structure and diversity, particularly for those species that create easily sampled scat piles. Co-located scats provide greater confidence of matching of scat to lizard location, making this a potentially useful tool for assessing social structures and relatedness among social group members. In addition, this method provides an alternative, non-invasive technique for threatened or secretive lizards.

Two DNA extraction methods (QIAamp ® DNA Stool Mini Kit and modified Gentra ®Puregene ® method) were successful for captive scats although success rates decreased when applied to wild scats. On the other hand, the Qiagen DNeasy ® Plant Mini Kit successfully extracted DNA from wild lizard scats; suggesting plant inhibitors present in the scats of herbivorous lizards may often prevent amplification of DNA unless they are filtered out. Although both mtDNA and nuDNA were successfully amplified from frozen and dried scat samples, the drying method will be more suitable when sampling in semi-arid to arid locations away from electricity supplies. As false alleles and allelic dropout may arise in scat samples with low quality and quantity of DNA (Taberlet *et al.* 1999; Broquet and Petit 2004; Valiere *et al.* 2007) error checking protocols should normally be adopted. DNA amplification replicates and assessment using a consensus approach have previously been suggested (Navidi *et al.* 1992; Taberlet *et al.* 1999; Broquet and Petit 2004) and an assessment of power such as probability of identity is recommended (Valiere 2002). We also recommend the use of replicate PCRs to assess the reliability of genotypes derived from DNA extracted using the Qiagen DNeasy ® Plant Mini Kit. In addition, as methods are not necessarily transferable between species (Taberlet *et al.* 1999) preliminary trials, incorporating genotyping error rates, are recommended prior to their use with other lizard species.

There are other potential problems when deriving lizard DNA from their scats. We found reptile scales on the surface of some captive and wild scats (Pearson, *pers. obs*.). Captive lizards were isolated so it could be assumed that in those cases the scales belonged to the

lizard from which scats were collected. The same assumption cannot be made for field collected scats as lizards may eat the sloughed skin of other individuals, or even conspecific neonates (Lanham and Bull 2004), potentially contaminating the sample with other conspecific DNA. Further, the co-location of scats may result in cross contamination between scats that are in contact but from different individuals. Also DNA extraction and amplification success is likely to decline with scat age as the DNA deteriorates (Demay *et al.* 2013). Wild lizard samples used in this study were fresh; we therefore recommend that future studies consider temporal sampling thresholds.

A further potential complication concerns the identification of scats from the target species. In this study, few other lizard species were sighted during the sampling of wild *E. stokesii* and the size and location of *E. stokesii* scats in piles immediately outside occupied crevice entrances facilitated identification. However, geckos were present and gecko scats may be confused with sub-adult *E. stokesii* scats, although species identification may be verified via sequencing. In addition, the field sites used in this study consisted of rocky outcrops with sparse vegetation where scats were easily found. Scats may be harder to locate and identify in an area with denser vegetation or higher lizard diversity. Knowledge of the behaviour of the target species and an awareness of other resident and transient species would be essential in such cases. However, this study indicates that more confidence may be applied to the identification of species from their scat in scat piling species, making non-invasive genetic sampling particularly applicable for such species.

Acknowledgements

We thank Leslie Morrison and staff of the Flinders University of South Australia animal house for assistance with captive lizards. This work was part funded by the Royal Society of South Australia, the Biology Society of South Australia, and the Nature Conservation Society of South Australia. Captive and wild lizards were managed in accordance with standards and

procedures in compliance with the Australian Code of Practice for the use of animals for scientific purposes. We also thank two anonymous reviewers and Steve Cooper for their helpful comments on earlier versions of the manuscript.

Figures and tables:

Table 1. Results of captive *E. stokesii* scat trials

Showing the number of scats used in trial for each storage and DNA extraction method, and numbers (%) successfully amplified for mt- and nuDNA for each storage and DNA extraction method.

Extraction method	Storage method		mtDNA amplified		nuDNA amplified	
	Frozen*	Dried*	Frozen	Dried	Frozen	Dried
QIAamp ® DNA Stool Mini Kit	11	20	10 (91)	20 (100)	9 (82)	17 (85)
ISOLATE Fecal DNA Kit	4	16	3 (75)	11 (69)	1 (25)	4 (25)
Direct	3	3	0	0	n/a	n/a
Chelex with boiling	3	3	0	0	n/a	n/a
Chelex without boiling	3	3	0	0	n/a	n/a
Modified Gentra ® Puregene ® method	3	9	3 (100)	9 (100)	3 (100)	9 (100)
Total	27	54	16	40	13	30

* Frozen: -20°C, Dried: sprayed with 90% ethanol then stored on silica beads at room temperature

References

Alacs, E., Alpers, D., de Tores, P.J., Dillon, M., and Spencer, P.B.S. (2003). Identifying the presence of quokkas (*Setonix brachyurus*) and other macropods using cytochrome b analyses from faeces. *Wildlife Research* **30**, 41-47.

Arandjelovic, M., Guschanski, K., Schubert, G., Harris, R., Thalmann, O., Siedel, H., and Vigilant, L. (2009). Two-step multiplex polymerase chain reaction improves the speed and accuracy of genotyping using DNA from noninvasive and museum samples. *Molecular Ecology Resources* **9**, 28-36.

Arida, E.A., and Bull, C.M. (2008). Optimising the design of artificial refuges for the Australian skink, *Egernia stokesii*. *Applied Herpetology* **5**(2), 161-172.

Ball, M.C., Pither, R., Manseau, M., Clark, J., Petersen, S.D., Kingston, S., Morrill, N., and Wilson, P. (2007). Characterisation of target nuclear DNA from faeces reduces technical issues associated with the assumptions of low-quality and quantity template. *Conservation Genetics* **8**, 577-586.

Barrows, C.W. (2006). Population dynamics of a threatened sand dune lizard. *The Southwestern Naturalist* **51**(4), 514-523.

Beja-Pereira, A., Oliveira, R., Alves, P.C., Schwarz, M.K., and Luikart, G. (2009). Advancing ecological understanding through technological transformations in noninvasive genetics. *Molecular Ecology Resources* **9**, 1279-1301.

Brinkman, T.J., Person, D.K., Schwartz, M.K., Pilgrim, K.L., Colson, K.E., and Hundertmark, K.J. (2010). Individual identification of Sitka black-tailed deer (*Odocoileus hemionus sitkensis*) using DNA from fecal pellets. *Conservation Genetics Resources* **2**, 115-118.

Broquet, T., Menard, N., and Petit, E. (2007). Noninvasive population genetics: a review of sample source, diet, fragment length and microsatellite motif effects on amplification success and genotyping error rates. *Conservation Genetics* **8**, 249-260.

Broquet, T., and Petit, E. (2004). Quantifying genotyping errors in noninvasive population genetics. *Molecular Ecology* **13**, 3601-3608.

Bull, C.M., Griffith, S.C., and Perkins, M.V. (1999). Some properties of a pheromone allowing individual recognition, from the scats of an Australian lizard, *Egernia striolata*. *Acta Ethologica* **2**, 35-42.

Casquet, J., Thebaud, C., and Gillespie, R.G. (2011). Chelex without boiling, a rapid and easy technique to obtain stable amplifiable DNA from small amounts of ethanol-stored spiders. *Molecular Ecology Resources* **2**(1), 136-141.

Cogger, H.G. (1983) 'Reptiles and amphibians of Australia.' (AH & AW Reed Pty Ltd, New South Wales)

Demay, S.M., Becker, P.A., Eidson, C.A., Rachlow, J.L., Johnson, T.R., and Waits, L.P. (2013). Evaluating DNA degradation rates in faecal pellets of the endangered pygmy rabbit. *Molecular Ecology Resources* **13**, 654-662.

Deuter, R., Pietsch, S., Hertel, S., and Muller, O. (1995). A method of preparation of fecal DNA suitable for PCR. *Nucleic Acids Research* **23**(18), 3800-3801.

Duffield, G.A., and Bull, C.M. (1998). Seasonal and ontogenetic changes in the diet of the Australian skink *Egernia stokesii*. *Herpetologica* **54**(3), 414-419.

Duffield, G.A., and Bull, C.M. (2002). Stable aggregations in an Australian lizard, *Egernia stokesii*. *Naturwissenschaften* **89**, 424-427.

Fenner, A.L., and Bull, C.M. (2008). The impact of nematode parasites on the behaviour of an Australian lizard, the gidgee skink *Egernia stokesii*. *Ecological Research* **23**, 897-903.

Frantz, A.C., Pope, L.C., Carpenter, P.J., Roper, T.J., Wilson, G.J., Delahay, R.J., and Burke, T. (2003). Reliable microsatellite genotyping of the Eurasian badger (*Meles meles*) using faecal DNA. *Molecular Ecology* **12**, 1649-1661.

Gardner, M.G., Bull, C.M., A., F., Murray, K., and Donnellan, S.C. (2007). Consistent social structure within aggregations of the Australian lizard, *Egernia stokesii* across seven disconnected rocky outcrops. *Journal of Ethology* **25**, 263-270.

Gardner, M.G., Bull, C.M., Cooper, J.B., and Duffied, G.A. (2001a). Genetic evidence for a family structure in stable social aggregations of the Australian lizard *Egernia stokesii*. *Molecular Ecology* **10**, 175-183.

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**, 1787-1794.

Gardner, M.G., Bull, C.M., Cooper, S.J.B., and Duffield, G.A. (2001b). Genetic evidence for a family structure in stable social aggregations of the Australian lizard *Egernia stokesii*. *Molecular Ecology* **10**, 175-183.

Gardner, M.G., Cooper, S.J.B., Bull, C.M., and Grant, W.N. (1999). Isolation of microsatellite loci from a social lizard, *Egernia stokesii*, using a modified enrichment procedure. *The Journal of Heredity* **90**(2), 301-304.

Gardner, M.G., Sanchez, J.J., Dudaniec, R.Y., Rheinberger, L., Smith, A.L., and Saint, K.S. (2008). *Tiliqua rugosa* microsatellites: isolation via enrichment and characterisation of loci

for multiplex PCR in *T. rugosa* and the endangered *T. adelaidensis*. *Conservation Genetics* **9**, 233-237.

Germano, D.J., Smith, P.T., and Tabor, S.P. (2007). Food habits of the blunt-nosed leopard lizard (*Gambelia sila*). *The Southwestern Naturalist* **52**(2), 318-323.

Harris, R.B., Winnie, J., J., Amish, S.J., Beja-Pereira, A., Godinho, R., Costa, V., and Luikart, G. (2010). Argali abundance in the Afghan Pamir using capture-recapture modeling from fecal DNA. *Journal of Wildlife Management* **74**(4), 668-677.

Herbert, L., Darden, S.K., Pedersen, B.V., and Dabelsteen, T. (2011). Increased DNA amplification success of non-invasive genetic samples by successful removal of inhibitors from faecal samples collected in the field. *Conservation Genetics Resources* **3**, 41-43.

Iyengar, A., Babu, V.N., Hedges, S., Venkataraman, B., Maclean, N., and Morin, P.A. (2005). Phylogeography, genetic structure, and diversity in the dhole (*Cuon alpinus*). *Molecular Ecology* **14**, 2281-2297.

Jones, R., Cable, J., and Bruford, M.W. (2008). An evaluation of non-invasive sampling for genetic analysis in northern European reptiles. *Herpetological Journal* **18**, 32-39.

Kohn, M.H., and Wayne, R.K. (1997). Facts from feces revisited. *TRENDS in Ecology and Evolution* **12**(6), 223-227.

Lanham, E.J., and Bull, C.M. (2004). Enhanced vigilance in groups in *Egernia stokesii*, a lizard with stable social aggregations. *Journal of Zoology, London* **263**, 95-99.

Main, A.R., and Bull, C.M. (1996). Mother-offspring recognition in two Australian lizards, *Tiliqua rugosa* and *Egernia stokesii*. *Animal Behavior* **52**, 193-200.

Marrero, P., Fregel, R., Cabrera, V.M., and Nogales, M. (2009). Extraction of high-quality host DNA from feces and regurgitated seeds: a useful tool for vertebrate ecological studies. *Biological Researcg* **42**, 147-151.

Monroe, C., Grier, C., and Kemp, B.M. (2013). Evaluating the efficiency of various thermo-stable polymerases against co-extracted PCR inhibitors in ancient DNA samples. *Forensic Science International* **228**, 142-153.

Monteiro, L., Bonnemaison, D., Vekris, A., Petry, K.G., Bonnet, J., Vidal, R., Cabrita, J., and Mégraud, F. (1997). Complex polysaccharides as PCR inhibitors in feces: *Helicobacter pylori* model. *Journal of Clinical Microbiology* **35**(4), 995-998.

Morin, P.A., Chambers, K.E., Boesch, C., and Vigilant, L. (2001). Quantitative polymerase chain reaction analysis of DNA from noninvasive samples for accurate microsatellite genotyping of wild chimpanzees (*Pan troglodytes verus*). *Molecular Ecology* **10**, 1835-1844.

Murphy, M.A., Kendall, K.C., Robinson, A., and Waits, L.P. (2007). The impact of time and field conditions on brown bear (*Ursus arctos*) faecal DNA amplification. *Conservation Genetics* **8**, 1219-1224.

Nagy, Z.T. (2010). A hands-on overview of tissue preservation methods for molecular genetic analyses. *Organisms Diversity & Evolution* **10**, 91-105.

Navidi, W., Arnheim, N., and Waterman, M.S. (1992). A multiple-tubes approach for accurate genotyping of very small DNA samples by using PCR: statistical considerations. *American Journal of Human Genetics* **50**(2), 347-359.

Nsubuga, A.M., Robbins, M.M., Roeder, A.D., Morin, A., Boesch, C., and Vigilant, L. (2004). Factors affecting the amount of genomic DNA extracted from ape faeces and the identification of an improved sample storage method. *Molecular Ecology* **13**, 2089-2094.

Panasci, M., Ballard, W.B., Breck, S., Rodriguez, D., Densmore III, L.D., Wester, D.B., and Baker, R.J. (2011). Evaluation of fecal DNA preservation techniques and effects of sample age and diet on genotyping success. *The Journal of Wildlife Management* **75**(7), 1616-1624.

Pavey, C.R., Burwell, C.J., and Nano, C.E.M. (2010). Foraging ecology and habitat use of Slater's skink (*Egernia slateri*): an endangered Australian desert lizard. *Journal of Herpetology* **44**(4), 563-571.

Piggott, M.P. (2004). Effect of sample age and season of collection on the reliability of microsatellite genotyping of faecal DNA. *Wildlife Research* **31**, 485-493.

Piggott, M.P., and Taylor, A.C. (2003). Extensive evaluation of faecal preservation and DNA extraction methods in Australian native and introduced species. *Australian journal of Zoology* **51**, 341-355.

Renan, S., Speyer, E., Shahar, N., Gueta, T., Templeton, A.R., and Bar-David, S. (2012). A factorial design experiment as a pilot study for noninvasive genetic sampling. *Molecular Ecology Resources* **12**, 1040-1047.

Roeder, A.D., Archer, F.I., Poinar, H.N., and Morin, P.A. (2004). A novel method for collection and preservation of faeces for genetic studies. *Molecular Ecology* **4**, 761-764.

Smith, A.L., Fenner, A.L., Bull, C.M., and Gardner, M.G. (2009). Genotypes and nematode infestations in an endangered lizard, *Tiliqua adelaidensis*. *Applied Herpetology* **6**, 300-305.

Solberg, K.H., Bellemain, E., Drageset, O.-M., Taberlet, P., and Swenson, J.E. (2006). An evaluation of field and non-invasive genetic methods to estimate brown bera (*Ursus arctos*) population size. *Biological Conservation* **128**(2), 158-168.

Steinglein, J.L., Waits, L.P., Ausband, D.E., Zager, P., and Mack, C.M. (2011). Estimating gray wolf pack size and family relationships using noninvasive genetic sampling at rendevous sites. *Journal of Mammalogy* **92**(4), 784-795.

Stenglein, J.L., Waits, L.P., Ausband, D.E., Zager, P., and Mack, C.M. (2011). Estimating gray wolf pack size and family relationships using noninvasive genetic sampling at rendevous sites. *Journal of Mammalogy* **92**(4), 784-795.

Taberlet, P., Griffin, S., Goossens, B., Questiau, S., Manceau, V., Escaravage, N., Waits, L.P., and Bouvet, J. (1996). Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research* **24**(16), 3189-3194.

Taberlet, P., and Luikart, G. (1999). Non-invasive genetic sampling and individual identification. *Biological Journal of the Linnean Society* **68**, 41-55.

Taberlet, P., Waits, L.P., and Luikart, G. (1999). Noninvasive genetic sampling: look before you leap. *TRENDS in Ecology and Evolution* **14**(8), 323-327.

Turner, F.B., and Medica, P.A. (1982). The distribution and abundance of the flat-tailed horned lizard (*Phrynosoma mcallii*). *Copeia* **1982**(4), 815-823.

Valiere, N. (2002). GIMLET: a computer program for analysing genetic individual identification data. *Molecular Ecology Notes* **2**, 377-379.

Valiere, N., Bonenfant, C., Toigo, C., Luikart, G., Gaillard, J.-M., and Klein, F. (2007). Importance of a pilot study for non-invasive genetic sampling: genotyping errors and population size estimation in red deer. *Conservation Genetics* **8**, 69-78.

Vynne, C., Baker, M.R., Breuer, Z.K., and Wasser, S.K. (2012). Factors influencing degradation of DNA and hormones in maned wolf scat. *Animal Conservation* **15**, 184-194.

Waits, L.P., and Paetkau, D. (2005). Noninvasive genetic sampling tools for wildlife biologists: a review of applications and recommendations for accurate data collection. *Journal of Wildlife Management* **69**(4), 1419-1433.

Watts, H.E., Scribner, K.T., Garcia, H.A., and Holekamp, K.E. (2011). Genetic diversity and structure in two spotted hyena populations reflects social organization and male dispersal. *Journal of Zoology* **285**, 281-291.

Wehausen, J.D., Ramey II, R.R., and Epps, C.W. (2004). Experiments in DNA extraction and PCR amplification from bighorn sheet feces: the importance of DNA extraction method. *Journal of Heredity* **95**(6), 503-509.

Wilgers, D.J., and Horne, E.A. (2009). Discrimination of chemical stimuli in conspecific fecal pellets by a visually adept iguanid lizard, *Crotaphytus collaris*. *Journal of Ethology* **27**, 157-163.

Chapter 3: MHC genotyping from next-generation sequencing

Accepted pending changes that have been addressed:

Pearson, S.K., Bradford, T.M., Ansari, T.H., Bull, C.M., and Gardner, M.G., MHC genotyping from next-generation sequencing: detailed methodology for the gidgee skink, *Egernia stokesii*, *Transactions of the Royal Society of South Australia*, resubmitted July 2016.

There is an increasing application of next-generation sequencing techniques in studies of the MHC, yet concise and comprehensive protocols are generally lacking. Next-generation sequencing data was used to identify *E. stokesii* MHC alleles and derive MHC genotypes for sampled lizards. In the course of identifying *E stokesii* MHC data, the following work provides detailed methodology relating to deriving alleles and genotypes from next-generation sequencing data.

MHC genotyping from next-generation sequencing: detailed methodology for the gidgee skink, *Egernia stokesii*

Pearson, S.K.[A,B], Bradford, T.M.[A], Ansari, T.H.[A], Bull, C.M.[A], Gardner, M.G.[A,C]

[A]School of Biological Sciences, Flinders University of South Australia, Bedford Park 5042, Australia, +61 8 8201 2315

[B]Corresponding author.

[C]Evolutionary Biology Unit, South Australian Museum, North Terrace, Adelaide 5000, Australia.

Pearson, S.K. +61 8 8201 2315 sarah.pearson@flinders.edu.au

Bradford, T.M. +61 8 8201 2315 tessa.bradford@flinders.edu.au

Ansari, T.H. +61 8 8201 2315 mina.ansari@flinders.edu.au

Bull, C.M. +61 8 8201 2263 michael.bull@flinders.edu.au

Gardner, M.G. +61 8 8201 2315 michael.gardner@flinders.edu.au

Resubmitted to *Transactions of the Royal Society of South Australia*, July 2016

Abstract

Next-generation sequencing has revolutionised molecular ecology. Its key advantages are a more accurate representation of genetic variation made possible by the generation of large volumes of data, more quickly and at a lower price per sequence than traditional sequencing methods. Yet these benefits come with a cost. For example, next-generation sequencing is error prone and requires increased quality control compared with traditional methods. Problems associated with next-generation sequencing may be exacerbated when sequencing gene complexes such as the major histocompatibility complex (MHC). Although not eliminated, significant progress has been made in addressing some of those problems and there is an increasing literature utilising this technology for studies of the MHC. However, what is generally lacking is detailed documentation of the methods used, and clear reasoning, for each step. Here we document detailed methodology, using an Australian lizard, *Egernia stokesii*, as a case study, with explanations, for MHC amplification, sequencing, and allele identification. This work provides molecular ecologists with a comprehensive guide to follow, particularly when first employing next-generation sequencing techniques similar to those used here. In addition, the *E. stokesii* MHC genotypes derived from this work provide foundation data for future investigations of the influence of social structure on the MHC.

Key words: Next-generation sequencing, MHC, genotyping, methodology, *Egernia* group, lizard

Introduction

Next-generation sequencing has revolutionised molecular ecology. The Sanger method, first described in 1977 (Sanger, Nicklen, & Coulson, 1977) represented the 'first-generation' of DNA sequencing. Non-Sanger methods, collectively termed 'next-generation' sequencing (Metzker, 2010), appeared in the sequencing market about a decade ago (Schuster, 2007). Because millions of sequences can be simultaneously processed in parallel, compared to 96 at a time with Sanger sequencing (Schuster, 2007), next-generation sequencing is sometimes called massive parallel sequencing. Outputs from next-generation sequencing contain high numbers of sequencing reads (strings of nucleotide bases) representing all variants (sequencing reads that differ in at least one base pair) in a sample of multiple individuals, even those occurring at low or rare frequencies (Thomas et al., 2006). The key advantages of next-generation sequencing are more accurate representation of genetic variation within the sample population, made possible by the generation of large volumes of data, more quickly and at a lower price per sequence than traditional methods (Metzker, 2010).

Yet these benefits come with a cost. Next-generation sequencing is error prone which can result in incorrectly identified nucleotide bases, with implications for genotyping and downstream analysis (McElroy, Thomas, & Luciani, 2014). Error profiles differ across the various next-generation sequencing platforms (e.g. 454 GS Junior, Miseq, and Ion Torrent PGM) (McElroy et al., 2014), necessitating individually tailored approaches to error minimisation in all steps of the sequencing pipeline and downstream analysis. Another next-generation sequencing problem is that the method has a hugely increased demand for computational power, data storage capabilities, processing and analysis requirements, and quality control compared to traditional methods (Depristo et al., 2011; Metzker, 2010). Notwithstanding these problems, researchers have been keen to apply next-generation sequencing in ecological and evolutionary studies of the highly polymorphic major

histocompatibility complex (MHC) (Babik, Taberlet, Ejsmond, & Radwan, 2009; Galan, Guivier, Caraux, Charbonnel, & Cosson, 2010; Zagalska-Neubauer et al., 2010).

The MHC consists of a cluster of genes that play a key role in parasite resistance, mate choice and kin recognition (Milinski, 2006; Piertney & Oliver, 2006; Sommer, 2005). Originally, the study domain of immunologists (Bjorkman & Parham, 1990; Klein, 1979; Simpson, 1988), the MHC is now incorporated in ecological and biological research (Potts & Wakeland, 1990; Schmid-Hempel, 2003; Sommer, 2005). The MHC can be divided into four classes: I, II, III, and IV (Acevedo-Whitehouse & Cunningham, 2006). Due to their role in an adaptive immune response, MHC class I and II are of major interest to ecologists and evolutionary biologists (Bernatchez & Landry, 2003; Milinski, 2006; Piertney & Oliver, 2006; Sommer, 2005). Problems associated with next-generation sequencing techniques are exacerbated when sequencing gene complexes such as the MHC. The presence of multiple loci, high allelic diversity within those loci, gene duplications, chimeras (sequences that contain sequence motifs originating from two different sequences) and pseudogenes (non-functional sequences) complicates the accurate identification of true MHC alleles (Babik, 2010; Babik et al., 2009).

Numerous pre- and post-sequencing approaches have been and continue to be developed in an effort to deal with the complexities of next-generation sequencing of the MHC. These include incorporating protocols aimed at minimising polymerase chain reaction (PCR) artefacts (sequences that are not true alleles) during the amplification of the MHC (Burri, Promerová, Goebel, & Fumagalli, 2014; Lenz & Becker, 2008), including technical DNA amplification and sequencing replicates (Robasky, Lewis, & Church, 2014; Sommer, Courtiol, & Mazzoni, 2013), and utilising programs and scripts specifically designed to deal with the intricacies of the MHC (e.g. jMHC, Sommer et al., 2013; Stuglik, Radwan, & Babik, 2011; Stutz & Bolnick, 2014).

In summary, although not eliminated, significant progress has now been made in addressing problems relating to interpreting the outputs of next-generation sequencing. This has allowed its rapid adoption in addressing ecological and evolutionary questions. The result is an increasing literature on next-generation sequencing derived studies of the MHC. But often those studies lack full and detailed descriptions of the methods used, and clear reasoning for each step (but see Lighten, Oosterhout, Paterson, McMullan, & Bentzen, 2014; Sommer et al., 2013). Here we document a methodology we used for genotyping the MHC of an Australian lizard, *Egernia stokesii*, a member of the *Egernia* group of scincid lizards. The wide range of life history characteristics, mating systems, and social structures (from solitary to family living) exhibited by related species in the *Egernia* group of lizards (Gardner, Pearson, Johnston, & Schwarz, 2015) offers comparative taxa to investigate the evolution and ecology of the MHC. For example, living in groups increases the risk of pathogen transmission, making *Egernia* an emerging model group for studying the influence of group living on the MHC (Gardner et al., 2015; While, Chapple, Gardner, Uller, & Whiting, 2015). Increased parasite risk predicts stronger selective pressure on the Class I and Class II MHC genes, to combat disease in group living compared to solitary species (Alexander, 1974; Moller, Merino, Brown, & Robertson, 2001). Group living has evolved independently in the *Egernia* lineage (Gardner et al., 2015) which allows comparison to studies of MHC in other group living animals.

We document a detailed methodology for characterising MHC with next-generation sequencing, using the group living *E. stokesii* as a case study. We explain the methods we have used for successful MHC amplification, sequencing, and allele identification. Details are provided of some procedural steps not readily found elsewhere. For other steps, we consolidate detail obtained from published work. We thus provide a comprehensive guide to follow. We also provide the first account of population level MHC allelic diversity for a

skink, as a foundation for exploring the evolutionary and ecological processes that have generated or maintained that diversity. The broad application of the technique documented here, particularly in the *Egernia* group of lizards, will create opportunities for many interesting studies related to the potential correlations between sociality and MHC.

Method

Study species

Egernia stokesii (J.E. Gray, 1845) is a large, long lived (Duffield & Bull, 1996), viviparous lizard widely distributed across semi-arid Australia (Cogger, 1983). It resides within crevices among rocky outcrops, forms stable family groups (Duffield & Bull, 2002; Gardner, Bull, Cooper, & Duffied, 2001), and is genetically and socially monogamous (Gardner, Bull, & Cooper, 2002). Although there is limited dispersal (Duffield & Bull, 2002) there is no evidence of loss of genetic diversity through inbreeding (Gardner, Godfrey, Fenner, Donnellan, & Bull, 2012). Individuals of *Egernia stokesii* host a range of parasites (Godfrey, Bull, Murray, & Gardner, 2006) and more socially connected lizards have higher parasite loads than less connected lizards (Godfrey, Bull, James, & Murray, 2009). The high level of sociality, and the correspondingly higher risk of parasite infection resulting from frequent social proximity in this species suggests an evolutionary challenge to the genetic components of disease resistance, including class I and class II of the MHC. This paper reports on the next-generation sequencing methods used to characterise MHC genes in this species, and the background pattern of MHC allelic diversity derived from those methods.

Method overview

A visual representation of all steps used (and described below) is shown in Figure 1. We adopted an approach that utilised free software for the purpose of demonstrating a method that may be employed under conditions of limited funding.

Sample collection & DNA extraction

Individuals of *Egernia stokesii* were sampled during field surveys undertaken between 1994-1998 at Camel Hill (CAM, hereafter termed the 1994-98 dataset) and between 2012-2013 at CAM, Castle Rock (CAS), and Castle Rock Ridge (CRR) (hereafter termed the 2012-13 dataset) near Hawker (31°54′S; 138°25′E) in the southern Flinders Ranges, South Australia. Sites and sampling methodology have already been described in detail elsewhere (Gardner et al., 2001; Gardner, Bull, Fenner, Murray, & Donnellan, 2007; Godfrey et al., 2006). Blood (up to 0.5 mL) was taken from the caudal vein of each lizard and stored in ethanol/saline solution (50%:0.85% v/v) during 1994-98 surveys or on Whatman FTA ® Elute during 2012-13 surveys. DNA was extracted from blood stored in ethanol/saline solution using a Gentra ® Puregene ® (Gentra Systems) method and from blood stored on Whatman FTA ® Elute using manufacturer guidelines.

DNA amplification

We targeted the peptide binding regions both of MHC I exon 2 (corresponding to the α-1 domain) and of MHC II exon 2 (corresponding to the β-1 domain) (Bjorkman et al., 1987a, 1987b; Brown et al., 1993). Peptide binding regions are sites at which recognition of self and non-self peptides occurs, ensuring an adaptive immune response is enacted to eliminate pathogens and maintain self-tolerance. Although the peptide binding region of MHC II crosses both α-1 and β-1, α-1 was not included because peptide contact generally occurs at β-1 (Brown et al., 1993; Madden, 1995).

We used an Illumina Miseq next-generation sequencing platform (see *MHC sequencing & data pre-processing* below), for which samples from different individuals are pooled for sequencing. Therefore, we needed a method of amplifying MHC by which data pertaining to each individual sampled could be extracted from the sequencing results. This was achieved in

a two-stage process (see Appendix 1, Supplemental Information for detailed schematic). First, we used locus specific primers to amplify the MHC. We initially trialled a number of degenerate and non-degenerate forward and reverse locus specific primers previously developed for *Tilqua rugosa*, another member of the *Egernia* group of lizards (Ansari, 2016; Ansari, Bertozzi, Miller, & Gardner, 2015). Degenerate primers include some base positions at which a number of bases are possible. As a result, degenerate primers are less specific and therefore more flexible in amplifying DNA, which is useful when the DNA sequence is unknown (but see Babik, 2010). Because the MHC may be expected to be more conserved among more closely related species, utilising primers that amplified *T. rugosa* MHC aimed to increase the likelihood of amplifying *E. stokesii* MHC compared with primers developed for less related species. Following primer trial and design (Appendix 1 and 2, Supplemental Information) we amplified a 216 base pair region of MHC I exon 2 (α-1) using the forward locus specific primer E2F1 and the reverse locus specific primer E2I2R1. Similarly, we amplified a 105 base pair region of MHC II exon 2 (β-1) using the forward locus specific primer TrIIβ1F1-tagF and the reverse locus specific primer ESB1R1. Details are provided in Appendix 1 and 3, Supplemental Information. Second, the cleaned up product from the first PCR was used in a second PCR using only primers that were essential for 1) individual identification, and 2) Illumina Miseq sequencing methodology (hereafter termed outer primers) (see Appendix 1, Supplemental Information for details). Both locus specific primers and outer primers contained a common sequence (hereafter termed an adapter) for the purpose of annealing locus specific primers and outer primers during the second PCR (see Appendix 1, Supplemental Information for details). Individual identification was possible through the inclusion of two eight base pair index sequences, also called barcodes or multiplex identifier (MID) tags (Meyer & Kircher, 2010), in the outer primers used in the second PCR. Every sample was assigned a unique combination of forward and reverse index

sequences. The same unique combination was used during the second PCR for both MHC I and MHC II for a given sample because we determined the two regions could be separated via their sequence identity post sequencing. Where samples are pooled for sequencing there is a risk of falsely assigning a sequence to a sample. The use of index sequences of eight base pair length meant index sequences were a minimum distance apart of three base pairs. This minimised the index false assignment rate (Meyer, Stenzel, & Hofreiter, 2008) i.e. three sequence errors would need to occur in the index before it resulted in the same index as that used for another sample. The reaction mix and cycling conditions for the first and second PCRs are provided in Appendix 1, Supplemental Information. As in the case of the first PCR, product from the second PCR was subsequently cleaned up for the purpose of removing non-target PCR product.

Clean product from the second PCR was then pooled in equimolar amounts for each of MHC class I and II based on a combination of molarity and concentration for a subset of samples, and fragment size for every sample. The molarity of each pool (MHC I and II) was determined because equimolar concentrations were required for an equal distribution of reads during sequencing. We determined concentration because a final pooled sample of at least 10 µl at 4 nM was required for sequencing on the Miseq platform. Molarity and concentration were quantified using an Agilent 2200 TapeStation © (Agilent Technologies). Fragment size was determined using gel electrophoresis. Only those samples with a band of expected size on the gel were included in the pool to minimise sequencing of non-target regions. Following pooling, the molarity and concentration of each pooled sample was again assessed using the TapeStation. The two pools were then combined into a single pool, as required for Illumina Miseq sequencing, based on equimolar concentrations. As a last step to remove non-target product, we performed another clean-up of the single pooled sample using Agencourt AMPure XP © beads (Beckman Coulter, Inc; product number A63880) before final

quantification using the TapeStation. We included replicate PCRs as recommended by Robasky et al. (2014) to account for variation in next-generation sequencing outputs (Metzker, 2010; Schirmer et al., 2015). We processed 10 % of samples in independent PCRs as technical duplicates. To do this we included independently amplified repeats of the same sample. Each technical duplicate was assigned a unique index sequence combination, different to that used for the first amplification of that sample.

MHC sequencing & data pre-processing

Samples were sequenced as 300 base pair paired end reads (read 1 and read 2, hereafter named R1 and R2), spiked with 10% PhiX (a viral genome). PhiX has a balanced nucleotide representation (i.e. approximate equal proportions of nucleotide bases A, T, G, and C) which can contribute to sequencing quality control particularly if sample sequences have low diversity or unbalanced nucleotide representation. This is particularly critical in the first few cycles of the Illumina sequencing when the sequencing clusters are being detected as our amplicons all have the same starting bases. Post cluster determination, Illumina sequencer algorithms are optimised with a balanced nucleotide representation. We therefore used 10% PhiX as a conservative measure as the diversity and nucleotide representation of *E. stokesii* samples was not known.

Sequencing was undertaken on an Illumina Miseq platform at the Australian Genome Research Facility (AGRF, Adelaide). Paired end sequencing generates sequence reads from both ends of a DNA fragment template. The 300 base pair sequence length permitted an overlap of paired end reads. The subsequently assembled consensus of this (see below) was a longer sequence than either of the single reads, thereby increasing coverage of the target regions for downstream analysis. Preliminary post-sequencing processing was automated in the MiSeq platform at the AGRF during which reads were de-multiplexed according to their unique index sequences, outer primers were removed as they were not required in

downstream analysis, and R1 and R2 were batched into one file per sample. Using the R1 and R2 files provided by the AGRF, we then used FastQC V 0.11.2 (Babraham Institute bioinformatics group, http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ accessed November 2014) to explore the quality of reads (e.g. number of reads per MHC region per sample, sequence length distribution, quality score thresholds) prior to assembling R1 and R2 for each sample with PEAR V 0.9.5 (Zhang, Kobert, Flouri, & Stamatakis, 2014) (Appendix 4, Supplemental Information). During assemblage, R1 and R2 were merged, based on defined quality parameters, to a minimum length of 50 base pairs (as an initial filter of non-target sequences), and a Phred quality score threshold of 20 which corresponds to a 1% base calling error rate (Nielsen, Paul, Albrechtsen, & Song, 2011). Assembled R1/R2 reads were converted from fastq to fasta format (as required in downstream analysis) using the FASTQ-to-FASTA converter in the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/ (Appendix 4, Supplemental Information).

Allele identification

Allele identification was undertaken separately for datasets derived from 1994-98 and 2012-13. Using the assembled fasta files, unique sequences that contained complete inner primers (forward and reverse adapters, forward and reverse locus specific primers) and that were without ambiguities in the target region (i.e. no bases labelled as "N"), were extracted using jMHC version 1.6.1624 https://code.google.com/p/jmhc) (Stuglik et al., 2011). Only sequences of expected length were selected for further analysis. In addition, unique sequences were only retained where they were detected in a minimum of three reads in a minimum of two samples. We deduced there was a low probability of the same artefact being observed three times (Galan et al., 2010). The two sample criterion was based on the two independent PCR rule where only alleles derived from two independent PCRs are considered true alleles (Babik, 2010; but see Radwan, Kuduk, Levy, Lebas, & Babik, 2014). Although we filtered on

expected sequence length, we also undertook an exploratory analysis of sequences that were three base pairs less or more than the expected length, in order to identify potential frameshift mutations, which would have resulted in a change in the translated amino acid sequence due to an insertion or deletion.

MHC genes evolve rapidly and are highly variable but this variability may be over-estimated because MHC sequences are known to occur as pseudogenes and amplification of MHC regions can result in PCR artefacts such as chimeras (Babik, 2010). Because of this, we checked all of the unique sequences that we identified using jMHC, for artefacts. We checked codon start positions before reviewing sequences for artefacts because an incorrect start position can result in erroneous stop codons within an amino acid sequence. For each unique sequence, we calculated the within sample frequency (i.e. the percentage of reads pertaining to each unique sequence within each sample), and we determined the maximum per sample frequency of that sequence across the entire dataset. We then sorted all of the sequences based on their frequencies (the maximum per amplicon frequency (MPAF) approach of Radwan et al., 2012). We chose a threshold of 2% based on the results of a review (in the method as detailed below for the subset of sequences above 2 % MPAF) of all sequences of expected length occurring in a minimum of three reads in a minimum of two samples for MHC I from the 1994-98 dataset (MPAF range 0.08 – 100%, results not shown). This cut-off, which assumes any sequences with an MPAF less than 2 % is an artefact, is within the range of MPAF thresholds either arbitrarily adopted or identified by others (Kuduk, Babik, et al., 2012; Nadachowska-brzyska, Zieliński, Radwan, & Babik, 2012; Radwan et al., 2012). Starting at an MPAF cut-off of 2 % and working up to higher frequencies, we selected three samples (where available) and reviewed sequences in Mega6 (Tamura, Stecher, Peterson, Filipski, & Kumar, 2013) for the presence of pseudogenes, chimeras, and single nucleotide base substitutions. Pseudogenes were detected by the presence of a stop codon in the amino

acid sequence. Chimeras were detected by eye by comparing the less frequent nucleotide sequence to putative parental sequences (sequences occurring in higher frequencies) based on the assumption that chimeras are likely to occur at lower frequencies (Radwan et al., 2012; Sommer et al., 2013). Unlike sequences derived from 454-sequencing in which the main errors are indels associated with homopolymers (McElroy et al., 2014), artefactual nucleotide base substitutions are a feature of sequences derived from Illumina MiSeq sequencing (Jünemann et al., 2013). A nucleotide base substitution error occurs when the sequencer substitutes a different nucleotide base than the actual base in a sequence being sequenced, resulting in a misrepresentation of that base. In our screening process, where a nucleotide sequence differed by one base pair from a putative parental sequence, it was considered an artefact. We acknowledge this exclusion is a conservative measure that may underestimate real MHC diversity. Remaining sequences, considered putative alleles, were then validated using two approaches: 1) comparison with previously published MHC sequences available on GenBank (international DNA sequence database) using the standard nucleotide Basic Local Alignment Search Tool (Altschul, Gish, Miller, Myers, & Lipman, 1990) (BLASTn) available at http://blast.ncbi.nlm.nih.gov/Blast.cgi and 2) amino acid alignment. Although the MHC is highly diverse, it contains conserved regions across divergent taxa, which are useful for allele validation using these two approaches. The nucleotide sequences of different putative alleles were first aligned using Clustal W in Mega6 then translated into amino acids. The amino acid sequences were then aligned in Mega6 against corresponding regions of the human leukocyte antigen system (HLA, the human version of MHC) (Bondinas, Moustakas, & Papadopoulos, 2007; Chelvanayagam, 1996; Reche & Reinherz, 2003) and the MHC of other species (Appendix 5, Supplemental Information). Because we were interested in identifying functional sequences, putative alleles that translated into an identical amino acid sequence (i.e. nucleotide base differences were synonymous) were subsequently treated as

the same putative allele. Again, this is a conservative approach to estimating actual allelic diversity. An alignment of putative alleles was then performed as a final check for one base pair differences between nucleotide sequences, the presence of chimeric sequences, and the presence of stop codons within amino acid sequences. Due to the multi-locus nature of the MHC, it is difficult to assign putative alleles to particular MHC loci. As the primers we used were not specific to single loci we considered the identified alleles to be a mixture of sequences from multiple loci. Alleles were named according to the MHC nomenclature of Klein et al. (1990). As our allele identification approach did not consider rare alleles we subsequently reviewed the entire dataset for cases of sequences of expected length occurring in only one individual with a filter of per amplicon frequency greater than 2 % and a minimum of three reads in an individual.

MHC genotyping

Individuals were genotyped for both MHC regions sequenced unless amplification or sequence failure occurred. Only putative alleles validated by the steps above were used in subsequent genotyping of individuals and downstream analysis. The coverage (total number of reads, hereafter called T1) required for reliable genotyping was determined using the approach and R scripts of Sommer et al. (2013) (doi:10.1186/1471-2164-14-542). Assays of multi-locus regions including the MHC can be affected by inefficiencies in the amplification of loci and alleles due to factors such as changes in the bases where the primers bind, and the stochastic nature of PCR. To investigate whether our determination of genotypes might have been affected by these problems we first used maximum likelihood methods to calculate allele amplification efficiencies (following Sommer et al., 2013). These efficiency values are relative. They are dependent upon parameter start points and the optimisation process of the algorithm used to calculate efficiencies (Sommer et al., 2013). Therefore, we then estimated the standardised amplification efficiency of each allele based on a reference allele. In this

method, values are not dependent on algorithm start point and optimisation process (Sommer et al., 2013). The reference allele was selected as the one that was most similar to the reference sequence used in primer design, identified from a neighbour-joining phylogenetic tree constructed using 1000 bootstraps in Mega6. We determined the T1 necessary to determine 99.9% of an individual's complete genotype. The T1 value was calculated using the R function 'T1.min.efficiency.replicated' of Sommer et al. (2013), following Galan et al. (2010) which is based on a given number of alleles (here, the maximum number of alleles found in an individual in the dataset) and which takes into account minimum amplification efficiencies. We allowed for minimum amplification efficiencies because they are reliable when used in predicting coverage thresholds (Sommer et al., 2013). To be conservative, we used the maximum, rather than median or minimum, T1 value. Samples with reads lower than the maximum T1 value determined by this method were excluded from downstream analysis. To further assess coverage thresholds, we used the Spearman rank correlation coefficient to test for any relationship between the number of alleles and total number of reads. A significant positive correlation would suggest some alleles may have been missed if a sample had a low number of reads. Within each sample retained after applying coverage thresholds, unique alleles were identified where the reads for that allele were equal to or greater than three. Genotyping error was assessed by comparing genotypes derived for technical (PCR) duplicates. In addition, the genotypes of individuals from family groups (Gardner, 2000) were used to review the inheritance of MHC alleles and assess the reliability of genotyping. Six parent/single offspring groups were assessed for reliability of MHC I genotyping. For MHC II, five parent/offspring groups (three groups with a single offspring, two with two offspring) were assessed.

Results

Sample collection and DNA extraction

DNA was extracted from 302 lizards; 67 from 1994-98 and 235 from 2012-13 (Table 1).

DNA amplification

Of the 302 unique samples, 296 (98%) amplified a product of the appropriate size for MHC I and 239 (79%) for MHC II (Table 1). The concentration of the final pooled, cleaned up MHC I and II product for submission to the AGRF was 3.95ng/ µl.

MHC sequencing and data pre-processing

Pooled PCR product from 302 unique lizards (Table 1) plus replicate product samples from 33 of those lizards were submitted for Illumina MiSeq sequencing (Australian Genomics Research Facility, Adelaide Australia). For all samples combined, 12,236,513 reads were obtained after de-multiplexing, of which 10,356,001 (84.6%) were retained after R1/R2 assemblage. The MHC I sequence was aligned with the open reading frame, therefore a 216 base pair region was used in downstream analysis. However, two nucleotide bases had to be trimmed from the beginning and one base from the end of MHC II exon 2 (β-1) sequences to align with the open reading frame. Therefore, a 102 base pair region of MHC II was used in downstream analyses.

Allele identification

After filtering based on sequence length and MPAF greater than 2 %, the number of unique sequences retained for artefact review were 39 for MHC I and 20 for MHC II in the 1994-98 dataset and 74 for MHC I and 47 for MHC II in the 2012-13 dataset (Appendix 6, Supplemental Information). The number of sequences removed due to a one base difference to a more frequent allele were seven 7 for MHC I and 12 for MHC II in the 1994-98 dataset and 36 for MHC I and 39 for MHC II in the 2012-13 dataset (Appendix 6, Supplemental Information). A putative pseudogene was identified in both MHC I datasets, with an MPAF of 9.88% in the 1994-98 dataset and 16.47% in the 2012-13 dataset (Appendix 6,

Supplemental Information). The nucleotide sequence of the putative pseudogene was identical in both MHC I datasets. One putative pseudogene, with an MPAF of 11.55% was identified in the MHC II 2012-13 dataset (Appendix 6, Supplemental Information). The lowest MPAF for a sequence subsequently validated as a true allele was 2.48% and 2.55% for MHC I and 7.57% and 5.78% for MHC II for the 1994-98 and 2012-13 datasets respectively. Chimeras were not detected in any dataset. Two sequences were removed from MHC I in the 2012-13 dataset and MHC II in both the 1994-98 and 2012-13 datasets due to translation to identical amino acid sequences as parental putative alleles (Appendix 6, Supplemental Information). No duplicate amino acid sequences were found among MHC I sequences following artefact removal in the 1994-98 dataset (Appendix 6, Supplemental Information). No rare alleles (as defined above) were detected nor were any true alleles found three base pairs different from the expected length for either region. Following filtering, artefact removal and allele validation, 39 MHC I and five MHC II alleles were identified (Figure 2, 3; Appendix 4, 5; Pearson Appendix 3, 4). Nine *E. stokesii* MHC I codons were conserved with other lizard species, and four of those were conserved across reptilian and non-reptilian taxa (Figure 2). For MHC II, eleven *E. stokesii* codons were conserved with other lizard species, and three of those across a wider range of taxa (Figure 3).

MHC genotyping

There was a significant positive correlation between the number of alleles and total reads ($P < 0.05$) for MHC I in both the 1994-98 and 2012-13 datasets (Appendix 7, Supplemental Information). Coverage thresholds for each region and dataset are shown in Appendix 7 of the Supplemental Information. Eight 1994-98 and twenty 2012-13 MHC I samples were removed after applying coverage thresholds calculated incorporating unequal and lowest allele amplification efficiency. Any potential bias in identifying different alleles for MHC I was addressed during coverage threshold filtering. The correlation between the number of

alleles and total reads for MHC II was not significant ($P > 0.05$). After applying coverage thresholds, we genotyped 63 (94%) 1994-98 samples and 201 (86%) 2012-13 samples (Table 1) corresponding to 55 unique samples for MHC I and 55 unique samples for MHC II for 1994-98, and 198 unique samples for MHC I and 89 unique samples for MHC II for 2012-13 (Table 1) (Pearson Appendix 5). Based on our sample of 33 technical duplicates, genotyping error was estimated at 5.14% for MHC I (n=14) and 4.84% for MHC II (n=18). For MHC I, in four parent offspring groups, all alleles in offspring were found in their parents, whereas the offspring in the other two groups each had one allele (out of a total of four alleles in one sample and seven in the other sample) with three or more reads, which occurred with only one or two reads in only one of their parents. The total number of MHC I alleles in offspring genotypes was 43, of which two were absent in parents, which represented an error rate of 4.65%. For MHC II, in all groups, all of the alleles identified in the offspring were also found in one or both parents.

Discussion

The highly polymorphic nature of the MHC demands individually tailored next-generation sequencing approaches. Despite the early adoption of next-generation sequencing to studies of the MHC, publication of detailed methodology is rare. Here we have provided detailed methodology for the amplification, sequencing, and identification of *E. stokesii* MHC alleles, including processes for screening unique sequences that might not represent unique alleles. This guide should provide a useful tool for others employing approaches similar to those used here in this and other taxa.

We derived genotypes for two MHC regions in just over 300 samples from *Egernia stokesii*. Although the genotyping error rates in this study are comparable to those reported elsewhere for next-generation data (Herdegen, Babik, & Radwan, 2014; Nadachowska-brzyska et al., 2012; Sepil, Moghadam, Huchard, & Sheldon, 2012), we suggest that increasing sequencing

coverage should improve the reliability of genotypes (Lighten et al., 2014; Oomen, Gillett, & Kyle, 2013). One of the challenges in undertaking this work was the range of programs and platforms required during post-sequencing processing, highlighting an increasing demand for bioinformatics skills by molecular ecologists.

The MHC genotypes for *E. stokesii* derived from this work will provide a foundation for future investigations of the influence of social structure on the MHC. Technical (e.g. allele identification methodology) differences make comparison of patterns of MHC allelic diversity among other taxa problematic. Nevertheless, the 39 MHC I alleles we identified for *E. stokesii,* appears to be much less diverse than the 226 MHC I alleles previously identified for the territorial Australian agamid lizard *Ctenophorus ornatus* (Radwan et al., 2014). On the other hand, the 39 MHC I alleles for *E. stokesii*, substantially exceed the eight recorded for another member of the *Egernia* group, *Tiliqua rugosa* (sleepy lizard; Ansari et al., 2015). Studies comparing allelic diversity in regions of both MHCI and II within a species have produced conflicting results. As in *E. stokesii*, there was greater allelic diversity in MHC I than MHC II in colony living sparrows (Bonneaud et al., 2004) and family living marmots (Kuduk, Johanet, Allainé, Cohas, & Radwan, 2012), but the opposite trend has been reported territorial common yellowthroats (Dunn, Bollmer, Freeman-Gallant, & Whittingham, 2013) and schooling lake whitefish (Binz, Largiader, Müller, & Wedekind, 2001). Although this study increases the number of comparable cases, more useful insights may be gained if datasets are generated using multiple methods and outputs compared. Additionally, the mandatory inclusion of scripts in published methodologies and the identification of standards may facilitate comparative studies. Future characterisation of allelic diversity in both MHC classes and in additional members of the *Egernia* group will help in investigations of the influence of social structure on MHC variability.

Acknowledgements

We thank field survey volunteers; Kathy Saint and Terry Bertozzi for assistance with laboratory work. We also thank the anonymous reviewer for their helpful comments and suggestions, which have improved the overall quality of the manuscript. Field survey methods were approved by the Flinders University Animal Welfare Committee, using standards and procedures in compliance with the Australian Code of Practice for the use of animals for scientific purposes. Laboratory work was undertaken at the South Australian Research Facility for Molecular Ecology and Evolution, Adelaide.

This work was supported by the Field Naturalists Society of South Australia, Sir Mark Mitchell Research Foundation, and Holsworth Wildlife Research Endowment.

References

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

Alexander, R. D. (1974). The evolution of social behavior. *Annual Review of Ecology and Systematics, 5*, 325-383.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology, 215*, 403-410. doi:10.1016/S0022-2836(05)80360-2

Ansari, T. H. (2016). *The distribution of genetic variation in sleepy lizards.* (Doctoral dissertation), Flinders University of South Australia.

Ansari, T. H., Bertozzi, T., Miller, R. D., & Gardner, M. G. (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.

Babik, W. (2010). Methods for MHC genotyping in non-model vertebrates. *Molecular Ecology Resources, 10*, 237-251.

Babik, W., Taberlet, P., Ejsmond, M. J., & Radwan, J. (2009). New generation sequencers as a tool for genotyping of highly polymorphic multilocus MHC system. *Molecular Ecology Resources, 9*, 713-719. doi:10.1111/j.1755-0998.2009.02622.x

Bernatchez, L., & Landry, C. (2003). MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? *Journal of Evolutionary Biology, 16*, 363-377.

Binz, T., Largiader, C., Müller, R., & Wedekind, C. (2001). Sequence diversity of Mhc genes in lake whitefish. *Journal of Fish Biology, 58*, 359-373.

Bjorkman, P. J., & Parham, P. (1990). Structure, function, and diversity of class I major histocompatibility complex molecules. *Annual Review of Biochemistry, 59*, 253-288.

Bjorkman, P. J., Saper, M. A., Samraoui, B., Bennett, W. S., Strominger, J. L., & Wiley, D. C. (1987a). The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. *Nature, 329*, 512-518. doi:10.1038/329512a0

Bjorkman, P. J., Saper, M. A., Samraoui, B., Bennett, W. S., Strominger, J. L., & Wiley, D. C. (1987b). Structure of the human class I histocompatibility antigen, HLA-A2. *Nature, 329*, 506-512. doi:10.1038/329506a0

Bondinas, G., Moustakas, A., & Papadopoulos, G. (2007). The spectrum of HLA-DQ and HLA-DR alleles, 2006: a listing correlating sequence and structure with function. *Immunogenetics, 59*, 539-553. doi:10.1007/s00251-007-0224-8

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*, 855-865. doi:10.1007/s00251-004-0648-3

Brown, J. H., Jardetzky, T., S., Gorga, J., C., Stern, L., J. , Urban, R., G., Strominger, J., L., & Wiley, D., C. (1993). Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature, 364*, 33-39. doi:10.1038/364033a0

Burri, R., Promerová, M., Goebel, J., & Fumagalli, L. (2014). PCR -based isolation of multigene families: lessons from the avian MHC class IIB. *Molecular Ecology Resources, 14*, 778-788. doi:10.1111/1755-0998.12234

Chelvanayagam, G. (1996). A roadmap for HLA-A, HLA-B, and HLA-C peptide binding specificities. *Immunogenetics, 45*, 15-26.

Cogger, H. G. (1983). *Reptiles and amphibians of Australia*: AH & AW Reed Pty Ltd.

Depristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., . . . Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics, 43*, 491-498. doi:10.1038/ng.806

Duffield, G. A., & Bull, C. M. (1996). Characteristics of the litter of the gidgee skink, *Egernia stokesii*. *Wildlife Research, 23*, 337-342.

Duffield, G. A., & Bull, C. M. (2002). Stable aggregations in an Australian lizard, *Egernia stokesii*. *Naturwissenschaften, 89*, 424-427.

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*, 679-687. doi:10.1111/j.1558-5646.2012.01799.x

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

Gardner, M. G. (2000). *A genetic investigation of sociality in the Australian group living lizard Egernia stokesii.* (Doctoral dissertation), Flinders University of South Australia, Adelaide.

Gardner, M. G., Bull, C. M., Cooper, J. B., & Duffied, G. A. (2001). Genetic evidence for a family structure in stable social aggregations of the Australian lizard *Egernia stokesii*. *Molecular Ecology, 10*, 175-183.

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

Gardner, M. G., Bull, C. M., Fenner, A., Murray, K., & Donnellan, S. C. (2007). Consistent social structure within aggregations of the Australian lizard, *Egernia stokesii* across seven disconnected rocky outcrops. *Journal of Ethology, 25*, 263-270.

Gardner, M. G., Godfrey, S. S., Fenner, A. L., Donnellan, S. C., & Bull, C. M. (2012). Fine-scale spatial structuring as an inbreeding avoidance mechanism in the social skink *Egernia stokesii*. *Australian Journal of Zoology, 60*, 272-277.

Gardner, M. G., Pearson, S. K., Johnston, G. R., & Schwarz, M. P. (2015). Group living in squamate reptiles: a review for stable aggregations. *Biological Reviews*. doi:10.1111/brv.12201

Godfrey, S. S., Bull, C. M., James, R., & Murray, K. (2009). Network structure and parasite transmission in a group living lizard, the gidgee skink, *Egernia stokesii*. *Behavioral Ecology and Sociobiology, 63*, 1045-1056.

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

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*, 2347-2359.

Jünemann, S., Sedlazeck, F. J., Prior, K., Albersmeier, A., John, U., Kalinowski, J., . . . Harmsen, D. (2013). Updating benchtop sequencing performance comparison. *Nature Biotechnology, 31*, 294-296. doi:10.1038/nbt.2522

Klein, J. (1979). The major histocompatibility complex of the mouse. *Science, 203*, 516-521.

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

Kuduk, K., Babik, W., Bojarska, K., Sliwińska, E. B., Kindberg, J., Taberlet, P., . . . Radwan, J. (2012). Evolution of major histocompatibility complex class I and class II genes in the brown bear. *BMC Evolutionary Biology, 12*, 197. doi:10.1186/1471-2148-12-197

Kuduk, K., Johanet, A., Allainé, D., Cohas, A., & Radwan, J. (2012). Contrasting patterns of selection acting on MHC class I and class II DRB genes in the Alpine marmot (*Marmota marmota*). *Journal of Evolutionary Biology, 25*, 1686-1693. doi:10.1111/j.1420-9101.2012.02537.x

Lenz, T. L., & 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*, 117-123. doi:10.1016/j.gene.2008.09.013

Lighten, J., Oosterhout, C., Paterson, I. G., McMullan, M., & Bentzen, P. (2014). 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, 14*, 753-767.

Madden, D. R. (1995). The three-dimensional structure of peptide-MHC complexes. *Annual Review of Immunology, 13*(1), 587-622.

McElroy, K., Thomas, T., & Luciani, F. (2014). Deep sequencing of evolving pathogen populations: applications, errors, and bioinformatic solutions. *Microbial Informatics and Experimentation, 4*(1), 1. doi:10.1186/2042-5783-4-1

Metzker, M., L. (2010). Sequencing technologies — the next generation. *Nature Reviews: Genetics, 11*, 31-46. doi:10.1038/nrg2626

Meyer, M., & Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols, 2010*(6), t5448.

Meyer, M., Stenzel, U., & Hofreiter, M. (2008). Parallel tagged sequencing on the 454 platform. *Nature Protocols, 3*, 267-278. doi:10.1038/nprot.2007.520

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

Moller, A. P., Merino, S., Brown, C. R., & Robertson, R. J. (2001). Immune defense and host sociality: A comparative study of swallows and martins. *The American Naturalist, 158*, 136-145.

Nadachowska-brzyska, K., Zieliński, P., Radwan, J., & Babik, W. (2012). Interspecific hybridization increases MHC class II diversity in two sister species of newts. *Molecular Ecology, 21*, 887-906. doi:10.1111/j.1365-294X.2011.05347.x

Nielsen, R., Paul, J., S., Albrechtsen, A., & Song, Y., S. (2011). Genotype and SNP calling from next-generation sequencing data. *Nature Reviews: Genetics, 12*, 443-451. doi:10.1038/nrg2986

Oomen, R. A., Gillett, R. M., & Kyle, C. J. (2013). Comparison of 454 pyrosequencing methods for characterizing the major histocompatibility complex of nonmodel species and the advantages of ultra deep coverage. *Molecular Ecology Resources, 13*, 103-116.

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

Potts, W. K., & Wakeland, E. K. (1990). Evolution of diversity at the major histocompatibility complex. *Trends in Ecology & Evolution, 5*, 181-187.

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. doi:10.1111/mec.12984

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*, 2469-2479.

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), 623-641. doi:10.1016/s0022-2836(03)00750-2

Robasky, K., Lewis, N., E., & Church, G., M. (2014). The role of replicates for error mitigation in next-generation sequencing. *Nature Reviews: Genetics, 15*, 56-62. doi:10.1038/nrg3655

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences, 74*, 5463-5497.

Schirmer, M., Ijaz, U. Z., D'Amore, R., Hall, N., Sloan, W. T., & Quince, C. (2015). Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Research, 43*, e37. doi:10.1093/nar/gku1341

Schmid-Hempel, P. (2003). Variation in immune defence as a question of evolutionary ecology. *Proceedings of the Royal Society of London B: Biological Sciences, 270*, 357-366.

Schuster, S. C. (2007). Next-generation sequencing transforms today's biology. *Nature Methods, 5*, 16-18. doi:10.1038/nmeth1156

Sepil, I., Moghadam, H. K., Huchard, E., & Sheldon, B. C. (2012). Characterization and 454 pyrosequencing of Major Histocompatibility Complex class I genes in the great tit reveal complexity in a passerine system. *BMC Evolutionary Biology, 12*, 1.

Simpson, E. (1988). Function of the MHC. *Immunology, 64*, 27.

Sommer, S. (2005). The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Frontiers in Zoology, 2*, 16. doi:doi:10.1186/1742-9994-2-16

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

Stuglik, M. T., Radwan, J., & Babik, W. (2011). jMHC: software assistant for multilocus genotyping of gene families using next-generation amplicon sequencing. *Molecular Ecology Resources, 11*, 739-742. doi:10.1111/j.1755-0998.2011.02997.x

Stutz, W. E., & Bolnick, D. I. (2014). Stepwise threshold clustering: a new method for genotyping MHC loci using next-generation sequencing technology. *PloS one, 9*, e100587. doi:10.1371/journal.pone.0100587

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution, 30*, 2725-2729. doi:10.1093/molbev/mst197

Thomas, R. K., Nickerson, E., Simons, J. F., Jänne, P. A., Tengs, T., Yuza, Y., . . . Meyerson, M. (2006). Sensitive mutation detection in heterogeneous cancer specimens by massively parallel picoliter reactor sequencing. *Nature Medicine, 12*, 852-855.

While, G. M., Chapple, D. G., Gardner, M. G., Uller, T., & Whiting, M. J. (2015). *Egernia* lizards. *Current Biology, 25*, R593-R595. doi:10.1016/j.cub.2015.02.070

Zagalska-Neubauer, M., Babik, W., Stuglik, M., Gustafsson, L., Cichoń, M., & Radwan, J. (2010). 454 sequencing reveals extreme complexity of the class II Major Histocompatibility Complex in the collared flycatcher. *BMC Evolutionary Biology, 10*, 395-395. doi:10.1186/1471-2148-10-395

Zhang, J., Kobert, K., Flouri, T., & Stamatakis, A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics, 30*, 614-620. doi:10.1093/bioinformatics/btt593

Tables

Table 1: Number of unique *Egernia stokesii* from which DNA was extracted, submitted for sequencing and genotyped for the major histocompatibility complex. Sites surveyed were Camel Hill (CAM), Castle Rock (CAS), and Castle Rock Ridge (CRR) during the Austral summers of 1994-98 and 2012-13.

Site/Year	DNA extracted	Submitted for sequencing		Genotyped	
		MHC I	*MHC II*	*MHC I*	*MHC II*
CAM 94-98	67	65	56	55	55
CAM 12-13	79	78	62	64	35
CAS 12-13	83	81	63	70	26
CRR 12-13	73	72	58	64	28
Total	302	296	239	253	144

Figures

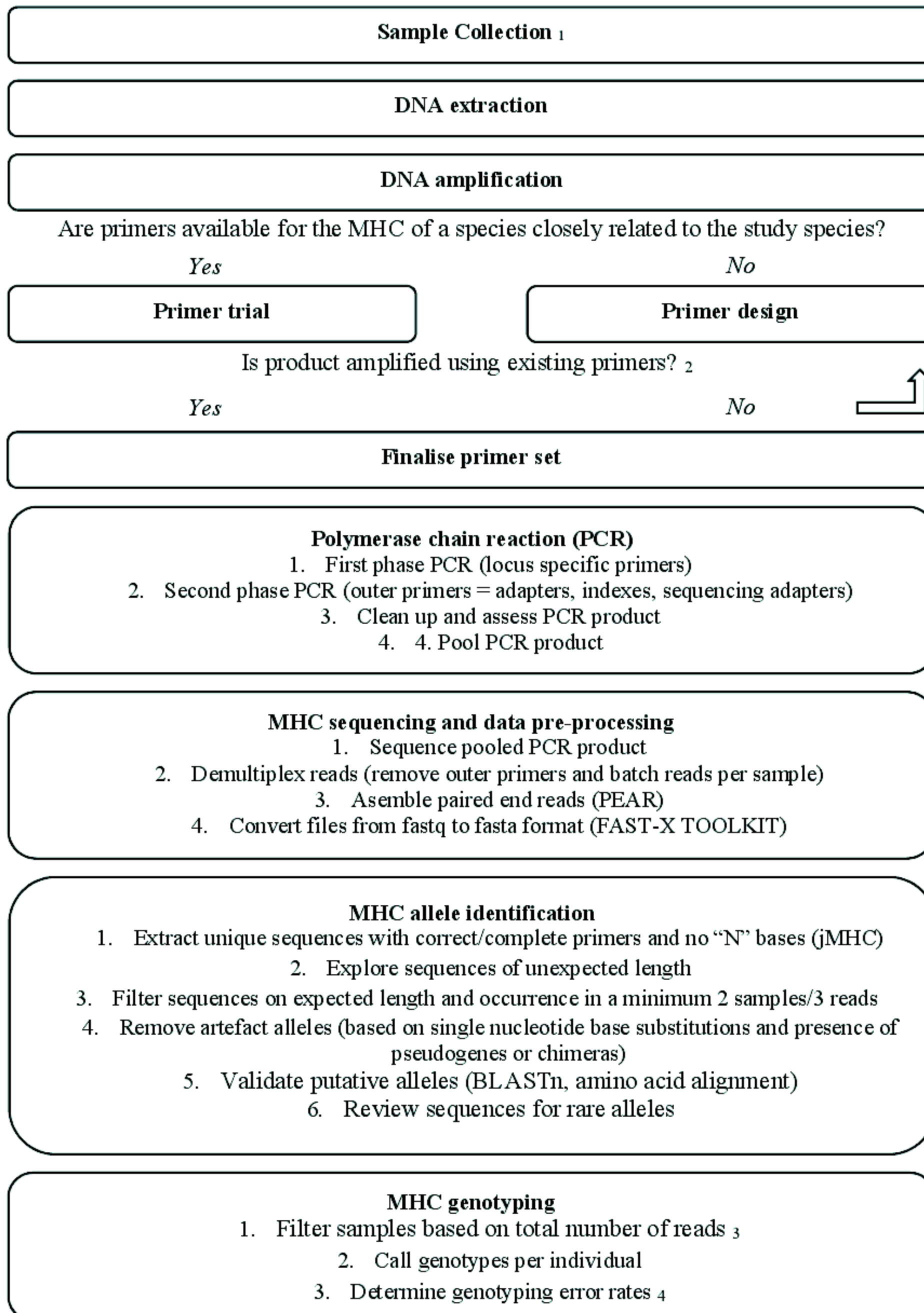


Figure 1: Summary of steps used to derive *Egernia stokesii* MHC alleles and genotypes.

Site number	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47		
			#		#							#																
Igig-UB*0101	E	V	P	Q	F	F	S	V	G	Y	V	D	D	Q	Q	F	V	S	Y	D	T	N			T	R		
Igig-UB*02	.	.	.	.	.	.	A	.	A	.	.	.	.	.	P	.	A	.	.	.	A	.			.	.		
Amcr-UB*01	.	.	.	R	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	P	.			A	.		
Amcr-UB*02	.	.	.	.	.	.	L	.	.	.	.	.	.	.	.	.	.	.	.	.	P	.			.	.		
Amcr-UB*03	.	.	.	R	.	.	Y	.	.	.	.	.	N	.	.	.	.	.	.	.	P	.			A	.		
Amcr-UB*0401	.	.	.	R	.	.	L	.	.	.	.	.	.	.	.	.	.	.	.	.	P	.			.	.		
Cosu-UB*0101	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	P	D			.	.		
Cosu-UB*02	.	.	.	.	.	.	Y	.	.	.	.	.	.	.	.	.	.	.	.	.	P	.			.	.		
Cosu-UB*03	.	.	.	.	.	.	I	.	.	.	M	.	.	.	R	.	.	Y	.	.	A	.			S	G		
Tiru-UA*0301	G	L	.	.	.	T	V	.	A	.	A	.	.	.	L	I	G	Y	.	.	S	.			R	.		
Tiru-UA*0402	.	L	.	W	.	L	I	.	.	.	.	.	G	.	R	.	F	H	.	.	S	D			G	.		
Hosa-HLA-A	G	E	.	R	.	I	A	.	.	.	.	.	.	T	.	.	.	R	F	.	S	D	A	A	S	Q		
Sppu-U*01	G	L	.	E	.	M	A	.	.	.	.	.	E	.	.	L	D	W	F	.	S	E			R	.		
Gaga-B-F10	G	Q	.	W	.	V	T	.	.	.	.	.	G	E	L	.	.	H	.	N	S	T			A	R		
Omny-UBA	G	F	.	.	.	V	A	.	.	I	.	.	G	M	.	I	D	Y	F	.	S	V			S	E		
Maru-UB*01	G	E	.	R	.	V	.	.	.	.	.	.	.	.	P	.	M	.	F	.	.	D			S	P	G	Q
Egst-UA*04	-	-	-	-	-	-	-	-	A	.	.	.	.	.	L	I	G	Y	.	.	S	S			R	.		
#Egst-UB*06	G	L	.	L	.	T	A	M	A	.	L	.	.	.	F	I	G	F	.	.	S	S			R	.		
#Egst-UB*07	G	L	.	L	.	T	A	M	A	.	L	.	.	.	F	I	G	F	.	.	S	S			W	G		
#Egst-UB*08	G	L	.	L	.	T	A	M	A	.	L	.	.	.	F	I	G	F	.	.	S	S			W	G		
#Egst-UB*09	G	L	.	L	.	T	A	M	A	.	L	.	.	.	L	I	G	F	.	.	S	S			R	.		
*Egst-UB*10	.	L	.	W	.	L	I	.	.	.	.	.	G	.	R	.	F	H	.	.	S	A			G	.		
Egst-UB*11	.	L	.	W	.	L	I	.	.	.	.	.	G	.	R	.	F	H	.	.	S	A			G	.		
Egst-UB*12	.	L	.	W	.	L	I	.	.	.	.	.	G	.	R	.	F	H	.	.	S	A			G	.		
Egst-UB*13	A	Q	.	R	.	I	K	E	.	.	L	.	G	.	C	I	M	Y	.	.	N	.			R	N		
*Egst-UB*14	A	Q	.	R	.	I	M	E	.	.	L	.	G	.	C	I	M	Y	.	.	N	H			R	N		
*Egst-UB*15	A	Q	.	W	.	I	M	E	.	.	L	.	G	.	C	I	M	Y	.	.	N	H			R	N		
Egst-UB*16	G	L	.	.	.	M	A	L	A	.	T	.	.	.	L	M	G	V	.	.	S	.			R	.		
Egst-UB*17	G	L	.	L	.	T	A	T	A	.	L	.	.	.	L	I	G	F	.	.	S	S			G	G		
Egst-UB*18	G	L	.	L	.	T	A	R	A	.	L	.	.	.	L	I	G	F	.	.	S	S			G	G		
*Egst-UB*19	G	L	.	L	.	T	A	T	A	.	L	.	.	.	L	I	G	F	.	.	S	S			R	G		
Egst-UB*20	G	L	.	L	.	T	A	T	A	.	L	.	.	.	L	I	G	F	.	.	S	S			R	G		

Site number	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95
																																														#		
Igig-UB*0101	R	R	L	P	T	V	P	W	I	K	K	V	D	E	D	D	P	Q	Y	W	E	R	N	S	Q	R	V	Q	G	T	E	E	T	F	Q	V	N	L	L	T	A	L	K	Y	Y	N	Q	S
Igig-UB*02	.	V	.	.	K	S	S	.	.	R	.	.	.	K	.	.	.	.	.	.	D	K	.	.	.	Y	.	S	D	S	.	Q	V	.	R	G	.	.	V	.	.	R	.	.	.	.	.	.
Amcr-UB*01	.	.	F	.	.	.	.	.	M	T	E	.	.	K	.	.	.	.	.	.	.	.	.	.	.	A	.	R	N	A	.	Q	V	.	R	.	G	.	V	.	L	R	.	.	.	.	.	.
Amcr-UB*02	.	H	F	.	.	.	.	.	M	T	E	.	.	K	.	.	.	.	.	.	.	W	.	.	.	A	A	R	N	A	.	Q	.	.	R	.	S	.	V	.	L	R	.	.	.	.	.	.
Amcr-UB*03	.	S	.	.	K	.	.	.	M	T	E	.	.	K	.	.	.	.	.	.	.	.	.	.	D	N	M	K	N	A	.	Q	A	.	R	.	G	.	V	N	V	R	.	.	.	.	.	.
Amcr-UB*0401	.	H	F	.	.	.	.	.	M	T	E	.	.	K	.	.	.	.	.	.	.	W	.	.	.	T	A	R	N	A	.	Q	A	.	R	.	G	.	V	.	.	R	.	.	.	.	.	.
Cosu-UB*0101	.	H	F	.	.	.	.	.	M	T	E	.	.	K	.	.	.	.	.	.	.	.	T	.	.	V	I	W	N	A	.	Q	V	.	R	.	D	.	V	N	V	P	.	.	.	.	.	.
Cosu-UB*02	.	S	.	.	K	.	.	.	.	T	.	.	.	K	.	.	.	.	.	.	.	W	.	.	D	G	M	K	N	A	.	Q	V	.	R	G	.	.	V	.	.	R	.	.	.	.	.	.
Cosu-UB*03	.	V	R	.	A	.	.	.	M	T	E	.	.	K	.	.	.	.	.	.	.	W	.	.	.	A	A	.	N	A	.	Q	V	.	R	G	.	.	V	.	.	R	.	.	.	.	.	.
Tiru-UA*0301	.	I	V	S	P	.	.	.	V	Q	.	I	E	K	E	.	.	.	L	.	D	M	H	T	.	.	A	.	N	E	.	L	Y	L	G	W	H	.	A	D	V	Q	.	L	H	.	.	.
Tiru-UA*0402	.	A	M	.	A	G	.	.	M	D	Q	.	C	Q	E	.	.	H	.	.	D	Q	E	T	.	S	A	.	.	E	.	A	.	L	R	S	Y	.	E	I	.	R	Q	R	.	.	.	.
Hosa-HLA-A	K	M	E	.	R	A	.	.	.	-	-	-	E	Q	E	G	.	E	.	.	D	Q	E	T	R	N	M	K	A	H	S	Q	.	D	R	A	.	.	G	.	L	R	G	.	.	.	.	.
Sppu-U*01	Q	S	E	.	R	A	E	.	M	A	-	-	Q	N	E	G	.	.	Y	.	.	.	Q	T	.	N	L	.	.	W	K	A	W	.	H	.	.	.	P	.	L	R	Q	R	.	.	.	.
Gaga-B-F10	.	Y	V	.	R	T	E	.	.	A	-	-	A	K	A	.	Q	.	.	.	D	G	Q	T	.	I	G	.	.	N	.	Q	I	D	R	E	.	.	G	I	L	Q	R	R	.	.	.	T
Omny-UBA	K	N	V	L	K	Q	S	.	M	E	-	-	-	-	-	G	V	R	D	E	K	M	I	T	N	T	R	K	.	N	Q	Q	.	.	K	A	.	V	E	I	G	M	Q	R	F	.	.	T
Maru-UB*01	.	E	E	.	R	G	.	.	M	D	S	M	K	H	E	E	.	E	F	.	.	.	Q	T	W	I	H	R	A	.	A	Q	N	Y	R	.	G	.	E	N	L	R	G	.	F	.	.	S
Egst-UA*04	.	I	V	S	S	.	.	.	M	Q	.	T	E	K	E	.	.	H	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
#Egst-UB*06	.	I	V	T	S	.	A	.	M	Q	.	T	E	K	E	.	.	.	L	.	A	M	H	T	K	M	A	.	R	E	.	L	F	L	R	R	D	.	V	.	V	Q	N	V	H	.	.	.
#Egst-UB*07	.	I	.	S	P	.	T	.	G	Q	.	T	E	K	E	.	.	.	L	.	A	T	H	T	.	L	A	.	N	N	.	L	F	R	R	K	D	.	A	.	V	Q	N	L	H	.	.	.
#Egst-UB*08	.	I	.	S	P	.	T	.	G	Q	.	T	E	K	E	.	.	.	L	.	A	M	H	T	K	M	A	.	R	E	.	L	F	L	R	R	D	.	V	.	V	Q	N	V	H	.	.	.
#Egst-UB*09	.	I	V	T	S	.	A	.	M	Q	.	T	E	K	E	.	.	.	L	.	A	T	H	T	.	L	A	.	N	N	.	L	F	L	R	R	D	.	A	.	V	Q	N	L	H	.	.	.
*Egst-UB*10	.	A	R	.	A	A	.	.	M	D	.	.	C	Q	E	.	.	H	.	.	D	Q	E	T	.	S	A	R	.	E	.	A	.	L	R	S	Y	.	E	I	.	R	Q	R	.	.	.	.
Egst-UB*11	.	A	M	.	A	A	.	.	M	D	.	.	C	Q	E	.	.	H	.	.	D	Q	E	T	.	S	A	R	.	E	.	A	.	L	R	S	Y	.	E	I	.	R	Q	R	.	.	.	.
Egst-UB*12	.	A	M	.	A	.	.	.	M	D	.	.	C	Q	E	.	.	H	.	.	D	Q	E	T	.	S	A	.	.	E	.	A	.	L	R	S	Y	.	E	I	.	R	Q	R	.	.	.	.
Egst-UB*13	.	V	M	H	R	.	.	.	M	E	.	A	.	K	E	T	S	P	S	.	D	G	Q	P	.	S	E	M	.	.	.	A	I	.	R	A	.	.	E	I	L	Q	.	R	.	.	.	.
*Egst-UB*14	.	V	M	H	R	.	.	.	M	E	.	.	.	K	K	T	S	L	S	.	D	V	Q	T	.	S	A	F	.	S	.	A	I	.	T	.	.	.	E	I	L	Q	.	R	H	.	H	.
*Egst-UB*15	.	V	M	H	R	.	.	.	M	E	.	A	.	K	K	T	S	L	S	.	D	V	Q	T	.	S	A	F	.	S	.	A	I	.	T	.	.	.	E	I	L	Q	.	R	H	.	.	.
Egst-UB*16	.	I	V	.	Y	L	.	.	V	.	.	I	E	K	E	.	.	.	F	.	A	M	P	T	.	F	M	.	N	Q	K	L	Y	L	G	.	D	.	D	.	V	Q	E	H	.	.	.	.
Egst-UB*17	.	I	V	S	P	.	A	.	G	Q	.	T	E	K	E	.	.	E	L	.	A	M	H	T	R	M	A	.	N	E	.	L	S	L	R	.	H	.	A	.	V	Q	N	L	H	.	.	.
Egst-UB*18	.	I	V	S	P	.	A	.	G	Q	.	T	E	K	E	.	.	E	L	.	A	M	H	T	R	M	A	.	N	E	.	L	S	L	R	.	H	.	A	.	L	Q	N	L	H	.	.	.
*Egst-UB*19	.	I	V	S	P	.	A	L	G	Q	.	T	E	K	E	.	.	H	L	.	V	.	H	T	.	M	A	.	N	N	.	L	F	L	R	R	D	.	A	.	V	Q	N	L	H	.	.	.
Egst-UB*20	.	I	V	S	P	.	A	L	G	Q	.	T	E	K	E	.	.	H	L	.	V	.	H	T	.	M	A	.	N	N	.	L	F	L	R	R	D	.	V	.	V	Q	N	V	H	.	.	.

(continued next page)

(continued from previous page)

Site number	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47
			#		#							#														
*Egst-UB*21*	G	L	.	L	.	T	A	T	A	.	L	.	.	.	L	I	G	F	.	.	S	S			W	G
*Egst-UB*22*	G	L	.	L	.	T	A	M	A	.	L	.	.	.	L	I	G	F	.	.	S	S			R	.
**Egst-UB*23*	G	L	.	L	.	T	A	M	A	.	L	.	.	.	L	I	G	F	.	.	S	S			R	G
*Egst-UB*24*	G	L	.	L	.	T	A	M	A	.	L	.	.	.	L	I	G	F	.	.	S	S			R	.
**Egst-UB*25*	G	L	.	L	.	T	A	M	A	.	L	.	.	.	L	I	G	F	.	.	S	S			G	G
*Egst-UB*26*	G	L	.	L	.	T	A	M	A	.	L	.	.	.	L	I	G	F	.	.	S	S			W	.
*Egst-UB*27*	G	L	.	L	.	T	A	M	A	.	L	.	.	.	L	I	G	F	.	.	S	S			W	E
*Egst-UB*28*	G	L	.	L	.	T	A	M	A	.	L	.	.	.	L	I	G	F	.	.	S	S			W	G
*Egst-UB*29*	G	L	.	L	.	T	A	M	A	.	L	.	.	.	L	I	G	F	.	.	S	S			W	G
*Egst-UB*30*	G	L	.	L	.	T	A	M	A	.	L	.	.	.	L	I	G	F	.	.	S	S			W	G
*Egst-UB*31*	G	L	.	L	.	T	A	M	A	.	L	.	.	.	L	I	G	F	C	.	S	S			G	G
*Egst-UB*32*	G	L	.	L	.	T	A	M	A	.	L	.	.	.	F	I	G	F	.	.	S	S			W	G
**Egst-UB*33*	G	L	.	L	.	T	A	M	A	.	L	.	.	.	F	I	G	F	.	.	S	S			W	G
*Egst-UB*34*	G	L	.	L	.	T	A	M	A	.	L	.	.	.	F	I	G	F	.	.	S	S			W	G
*Egst-UB*35*	G	L	.	L	.	T	A	T	A	.	L	.	.	.	L	I	G	F	.	.	S	S			R	G
*Egst-UB*36*	G	L	.	L	.	T	A	M	A	.	L	.	.	.	L	I	G	F	.	.	S	S			W	G
*Egst-UB*37*	G	L	.	L	.	T	A	M	A	.	L	.	.	.	L	I	G	F	.	.	S	S			W	G
*Egst-UB*38*	G	L	.	L	.	T	A	M	A	.	L	.	.	.	L	I	G	F	.	.	S	S			R	.
*Egst-UB*39*	G	L	.	L	.	T	A	M	A	.	L	.	.	.	L	I	G	F	.	.	S	S			G	G
*Egst-UB*40*	G	L	.	L	.	T	A	M	A	.	L	.	.	.	L	I	G	F	.	.	S	S			W	G
**Egst-UB*41*	G	L	.	L	.	T	A	M	A	.	L	.	.	.	L	I	G	F	.	.	S	S			G	G
*Egst-UB*42*	G	L	.	L	.	T	A	M	A	.	L	.	.	.	L	I	G	F	.	.	S	S			W	G
*Egst-UB*43*	G	L	.	L	.	T	A	M	A	.	L	.	.	.	L	I	G	F	.	.	S	S			W	G
*Egst-UB*44*	G	L	.	L	.	T	A	M	A	.	L	.	.	.	L	I	G	F	.	.	S	S			W	G

Site number	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71
*Egst-UB*21*	.	I	.	S	P	.	A	.	G	Q	.	T	E	K	E	.	.	.	L	.	D	M	H	M
*Egst-UB*22*	.	I	V	T	S	.	A	.	M	Q	.	T	E	K	E	.	.	.	L	.	D	M	H	T
**Egst-UB*23*	.	I	V	S	P	.	A	L	G	Q	.	T	E	K	E	.	.	H	L	.	V	S	H	T
*Egst-UB*24*	.	I	M	T	S	.	A	.	M	Q	.	T	E	K	E	.	.	.	L	.	D	M	H	T
**Egst-UB*25*	.	I	V	S	P	.	A	.	G	Q	.	T	E	K	E	.	.	E	L	.	A	M	H	T
*Egst-UB*26*	S	I	V	S	P	.	A	.	V	Q	N	T	E	K	E	.	.	E	L	.	A	M	H	T
*Egst-UB*27*	.	I	V	S	P	.	A	.	V	Q	.	A	E	K	E	.	.	E	L	.	A	M	H	T
*Egst-UB*28*	.	I	.	.	P	.	A	.	G	Q	.	T	E	K	E	.	.	.	L	.	A	M	H	T
*Egst-UB*29*	.	I	V	S	P	.	A	.	V	Q	.	T	E	K	E	.	.	E	L	.	A	M	H	T
*Egst-UB*30*	.	I	V	S	P	.	A	.	V	Q	.	A	E	K	E	.	.	E	L	.	A	M	H	T
*Egst-UB*31*	.	I	V	S	P	.	A	.	V	Q	.	T	E	R	K	E	Y	.	L	.	A	M	H	T
*Egst-UB*32*	.	I	.	S	P	.	T	.	G	Q	.	T	E	K	E	.	.	.	L	.	A	T	H	T
**Egst-UB*33*	.	I	.	S	P	.	T	.	G	Q	.	T	E	K	E	.	.	H	L	.	V	.	H	T
*Egst-UB*34*	.	I	V	S	A	.	A	.	M	Q	.	T	E	K	E	.	.	.	L	.	D	M	H	T
*Egst-UB*35*	.	I	V	S	P	.	A	L	G	Q	.	T	E	K	E	.	.	H	L	.	V	S	H	T
*Egst-UB*36*	.	I	V	S	P	.	A	.	V	Q	.	A	E	K	E	.	.	E	L	.	A	M	H	A
*Egst-UB*37*	.	I	V	S	P	.	A	.	V	Q	.	A	E	K	E	.	.	E	L	.	A	M	H	V
*Egst-UB*38*	.	I	V	T	S	.	A	.	M	Q	.	T	E	K	E	.	.	.	L	.	A	M	H	T
*Egst-UB*39*	.	I	V	S	P	.	A	.	V	Q	.	A	E	K	E	.	.	E	L	.	A	M	H	T
*Egst-UB*40*	.	I	V	S	P	.	A	.	V	Q	.	A	E	K	E	.	.	E	L	.	A	M	H	T
**Egst-UB*41*	.	I	V	S	P	.	A	.	V	Q	N	T	E	K	E	.	.	E	L	.	A	M	H	T
*Egst-UB*42*	.	I	M	S	P	.	A	.	V	Q	N	T	E	K	E	.	.	E	L	.	A	M	H	T
*Egst-UB*43*	.	I	M	S	P	.	A	.	V	Q	N	T	E	K	E	.	.	E	L	.	A	M	H	T
*Egst-UB*44*	.	I	M	S	P	.	A	.	V	Q	.	A	E	K	E	.	.	E	L	.	A	M	H	T

Site number	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95
																						#		
*Egst-UB*21*	.	M	A	.	R	E	.	L	V	L	R	R	H	.	A	.	V	Q	N	L	H	.	.	.
*Egst-UB*22*	K	M	A	.	R	E	.	L	F	L	R	R	D	.	V	.	V	Q	N	V	H	.	.	.
**Egst-UB*23*	.	M	A	.	N	N	.	L	F	L	R	R	D	.	V	.	V	Q	N	V	H	.	.	.
*Egst-UB*24*	K	M	A	.	R	E	.	L	F	L	R	R	D	.	V	.	V	Q	N	V	H	.	.	.
**Egst-UB*25*	R	M	A	.	N	E	.	L	S	L	R	.	H	.	A	.	L	Q	N	L	H	.	.	.
*Egst-UB*26*	E	M	A	.	N	E	.	L	S	L	R	R	H	.	A	.	V	Q	N	L	H	.	.	.
*Egst-UB*27*	R	M	A	.	N	.	.	L	F	R	R	K	D	.	A	.	V	Q	N	L	H	.	.	.
*Egst-UB*28*	E	M	A	.	N	E	.	L	S	L	R	R	H	V	A	.	V	Q	N	V	H	.	.	.
*Egst-UB*29*	.	.	A	K	N	N	.	L	F	L	R	G	H	.	A	.	V	Q	N	L	H	.	.	.
*Egst-UB*30*	R	M	A	.	N	.	.	L	F	R	R	K	D	.	A	.	V	Q	N	L	H	.	.	.
*Egst-UB*31*	.	V	A	.	K	E	.	L	S	L	R	R	D	.	A	.	V	Q	N	L	H	.	.	.
*Egst-UB*32*	.	L	A	.	N	N	.	L	F	L	R	R	D	.	A	.	V	Q	N	L	H	.	.	.
**Egst-UB*33*	.	M	A	.	N	N	.	L	F	L	R	R	D	.	V	.	V	Q	N	V	H	.	.	.
*Egst-UB*34*	K	M	A	.	N	N	.	L	F	R	R	R	D	.	A	.	V	Q	N	L	H	.	.	.
*Egst-UB*35*	.	M	A	.	N	N	.	L	F	L	R	R	D	.	V	.	V	Q	N	V	H	.	.	.
*Egst-UB*36*	R	L	A	.	N	N	.	L	F	R	R	K	D	.	A	.	V	Q	N	L	H	.	.	.
*Egst-UB*37*	R	L	A	.	N	N	.	L	F	R	R	K	D	.	A	.	V	Q	N	L	H	.	.	.
*Egst-UB*38*	K	M	A	.	R	E	.	L	F	L	R	R	D	.	V	.	V	Q	N	V	H	.	.	.
*Egst-UB*39*	R	M	A	.	N	.	.	L	F	R	R	K	D	.	A	.	V	Q	N	L	H	.	.	.
*Egst-UB*40*	R	M	A	.	N	N	.	L	F	R	R	K	D	.	A	.	V	Q	N	L	H	.	.	.
**Egst-UB*41*	.	M	A	.	R	E	.	L	V	L	R	R	H	.	A	.	V	Q	N	L	H	.	.	.
*Egst-UB*42*	.	M	A	.	N	N	.	L	F	L	R	R	H	.	A	.	V	Q	N	L	H	.	.	.
*Egst-UB*43*	.	M	A	.	R	E	.	L	V	L	R	R	H	.	A	.	V	Q	N	L	H	.	.	.
*Egst-UB*44*	R	M	A	.	N	N	.	L	S	L	R	R	H	.	A	.	V	Q	N	L	H	.	.	.

Figure 2: Amino acid sequences of 39 MHC I exon 2 (α-1) alleles of *Egernia stokesii* (shown in italics), aligned with the corresponding region of other species and the human leucocyte antigen system (HLA, human version of MHC). *Egernia stokesii* alleles prefixed with # were found only in the 1994-98 dataset, alleles prefixed with * were found only in the 2012-13 dataset, all other alleles were common to both datasets. Sites conserved across all lizard species included in the alignment are shaded in grey, sites conserved across all species in the alignment are indicated with '#'. Other sequences have been trimmed to match *E. stokesii* open reading frame. Site number corresponds to Igig-UB*0101. Species names, common names, and Genbank allele and accession numbers are detailed in Appendix 5, Supplemental Information.

		#														#				#														
	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79
Amcr-DAB1*01	A	R	G	E	F	E	A	V	A	A	L	G	E	P	D	A	R	Y	W	N	G	Q	K	E	V	L	E	Q	N	R	A	A	A	D
Amcr-DAB2*01	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	H	.	.	.	.	.	.	.	.	.	N	A	.	.	.	.	.
Amcr-DAB2*02	.	.	.	.	Y	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	D	R	R	.	.	E	V	.
Amcr-DAB3*01	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	N	A	.	.	.	.	.
Amcr-DAB3*02	.	.	.	.	Y	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	R	.	.	.	.	.
Amcr-DAB4*01	.	.	.	.	Y	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	D	R	R	.	.	E	V	.
Amcr-DAB4*02	.	.	.	.	Y	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	D	R	R	.	.	E	V	.
*Egst-DRB*01*	R	.	.	S	.	.	.	.	T	E	.	.	.	L	E	.	.	V	.	.	S	.	T	.	F	.	.	A	M	.	T	D	V	N
*Egst-DRB*02*	R	.	.	S	.	.	.	I	T	E	.	.	.	.	E	.	.	V	.	.	S	.	T	.	F	.	.	A	M	.	T	D	V	.
*Egst-DRB*03*	R	.	.	T	Y	.	.	.	T	.	.	.	.	.	S	.	.	.	.	.	S	L	T	D	Y	M	.	R	K	.	T	E	V	.
*Egst-DRB*04*	R	.	.	S	.	.	.	.	T	E	.	.	K	.	E	.	.	V	.	.	S	.	T	.	F	.	.	A	M	.	T	D	V	.
*Egst-DRB*05*	R	.	.	S	.	.	.	I	T	E	.	.	.	.	E	.	.	I	.	.	S	.	T	.	F	.	.	A	M	.	T	D	V	.
Hosa-DRB1	D	V	.	.	Y	R	.	.	E	E	.	.	R	.	.	.	E	.	.	.	S	.	.	D	L	.	.	.	K	.	G	Q	V	.
Sppu-DAB1*01	D	.	.	L	.	.	P	.	T	E	.	.	R	.	.	.	E	S	.	.	R	.	T	.	I	.	Q	D	R	.	T	.	V	E
Taac-DZB*01	D	V	.	V	F	V	S	.	T	E	.	D	S	K	.	.	Q	.	.	.	S	.	.	D	I	M	D	.	K	.	.	E	V	.
Gaga-BLB1	D	V	.	K	Y	V	.	D	T	P	.	.	.	P	Q	.	E	Y	.	.	S	N	A	.	F	.	.	N	R	M	N	E	V	.
Brre-DAB1	T	V	.	K	Y	V	G	Y	T	E	Q	.	V	I	F	.	.	N	F	.	K	N	Q	A	Y	.	Q	.	R	K	.	E	V	E

Figure 3: Amino acid sequences of five MHC II exon 2 (β-1) alleles of *Egernia stokesii* (shown in italics), aligned with the corresponding region of other species and the human leucocyte antigen system (HLA, human version of MHC). Sites conserved across all lizard species included in the alignment are shaded in grey, sites conserved across all species in the alignment are indicated with '#'. Other lizard sequences have been trimmed to match *E. stokesii* open reading frame. Site number corresponds to Amcr-DAB*01. Species names, common names, and Genbank allele and accession numbers are detailed in Appendix 5, Supplemental Information.

MHC genotyping from next-generation sequencing: detailed methodology for the gidgee skink, *Egernia stokesii*

Supplemental Information

Pearson, S.K.[A,B], Bradford, T.M.[A], Ansari, T.H.[A], Bull, C.M.[A], Gardner, M.G.[A,C]

[A]School of Biological Sciences, Flinders University of South Australia, Bedford Park 5042, Australia, +61 8 8201 2315

[B]Corresponding author.

[C]Evolutionary Biology Unit, South Australian Museum, North Terrace, Adelaide 5000, Australia.

Pearson, S.K. +61 8 8201 2315 sarah.pearson@flinders.edu.au

Bradford, T.M. +61 8 8201 2315 tessa.bradford@flinders.edu.au

Ansari, T.H. +61 8 8201 2315 mina.ansari@flinders.edu.au

Bull, C.M. +61 8 8201 2263 michael.bull@flinders.edu.au

Gardner, M.G. +61 8 8201 2315 michael.gardner@flinders.edu.au

Appendix 1: Primer trials

We initially trialled a number of degenerate and non-degenerate forward and reverse locus specific primers previously developed for amplifying the MHC of *Tiliqua rugosa* (Ansari, Bertozzi, Miller, & Gardner, 2015; Ansari, unpublished) (details of a subset of primers trialled are provided in Table 1). A member of the *Egernia* group of lizards, *T. rugosa* is more related to *E. stokesii* than other lizards e.g. iguanids for which primers had already been developed. Trialling primers that amplify MHC in a closely related species is a useful first step prior to primer design because the MHC is likely to be more conserved between the study species and those that are more closely related which can increase the likelihood of amplifying study species MHC using existing primers. Degenerate primers include bases at which alternate bases are possible, which is particularly useful for the highly polymorphic MHC and when the DNA sequence is not known (but see Babik, 2010). Fifteen *T. rugosa* primers were used in trials for MHC I, of which 12 were degenerate primers and 11 *T. rugosa* primers were used in trials for MHC II, of which six were degenerate primers. Primer combinations were trialled in one or both of two PCR protocols: 1) Taq ® Gold and 2) Multiplex Ready Technology (MRT; Hayden, Nguyen, Waterman, & Chalmers, 2008) (reaction mix and cycling conditions are provided in Table 2). The enzyme used in amplifying the MHC can impact errors within sequencing reads (reviewed in Brandariz-Fontes et al., 2015) and although a high fidelity polymerase is sometimes recommended (but see Babik, 2010), we did not use one in this work. Based on earlier trials, all PCRs used DNA dilutions of 1:50 for MHC I, 1:5 for MHC II.

Where *T. rugosa* primer pairs resulted in a visible band on 1.5% agarose gel, Sanger sequencing was used to further assess primer pair utility for generating *E. stokesii* MHC sequences. PCR products were purified prior to sequencing; unincorporated primers and dNTPs were removed using multiscreen PCR filter plates (Millipore Billerica, MA). Where

multiple bands were present on the agarose gel, DNA was isolated from the gel then purified using either a MinElute® Gel Extraction Kit (Qiagen) or UltraClean® GelSpin® DNA Purification Kit (Mo Bio Laboratories). Sequence reactions were prepared using a BigDYE® Terminator Cycle Sequencing Kit v3.1 (Applied Biosystems) according to manufacturer guidelines, using the same primers as those used in PCR amplification. Sequence reaction cycling conditions were: 3 minutes at 96 ºC; 31 cycles of 30 seconds at 96 ºC, 15 seconds at 50 ºC, 4 minutes at 60 ºC; final elongation for 3 minutes at 25 ºC, 30 seconds at 25 ºC. Sequence reaction products were purified using multiscreen PCR filter plates (Millipore Billerica, MA) prior to submission to the Australian Genome Research Facility (AGRF, Adelaide, Australia) for capillary separation on an ABI Prism 3730xl capillary sequencer. In some cases, 10µl of purified PCR was submitted to the AGRF for big dye terminator sequencing and capillary separation.

Following Sanger sequencing, ambiguous bases were manually corrected before assessment of an alignment of *E. stokesii* nucleotide sequences against MHC sequences of *T. rugosa* (Ansari et al., 2015; Ansari, unpublished) and other lizards available from GenBank (Appendix 5). Only primer pairs from which we derived *E. stokesii* MHC amino acid sequences and that aligned with conserved MHC regions of other lizards were selected for further work i.e. primer design and/or DNA amplification (see Appendix 2 and 3).

Table 1: Details of primers used in primer trials and subsequent amplification of *Egernia stokesii* MHC. Several locus specific primer pairs were trialled, only those for which *E. stokesii* MHC sequences were derived following Sanger sequencing are listed. The relationship between locus specific primers, adapters, and outer primer components is shown in Figure 1.

Primer name	Sequence (from 5' to 3')	Application
Locus specific primers		
MHC1α 1 F1	GGY KCC TCB TCC CAC TCK GYG MRG TA	*T. rugosa* forward degenerate primer used in primer trials and from which *E. stokesii* MHC I sequences were derived
Trα1α2r	GTC CAG GTG AGG GTC TCC TT	*T. rugosa* reverse primer used in primer trials and from which *E. stokesii* MHC I sequences were derived
TrIIβ1F1-tagF	GAG YWC GYG CGC TTC GAC A	*T. rugosa* forward degenerate primer used in primer trials and from which *E. stokesii* MHC II sequences were derived.
TrIIβ1R1_tagR	WGA TKC CRT AGT TGT RCC GGC AG	*T. rugosa* reverse degenerate primer used in primer trials and from which *E. stokesii* MHC II sequences were derived.
TrIIβ1R1B_tagR	CGA TGC CGT AGT TGT GCC GGC AG	*T. rugosa* reverse primer used in primer trials and from which *E. stokesii* MHC II sequences were derived.
E2F1	GTG TCG GAG CCT GGC CAG	Forward primer designed following primer trials and used to amplify *E. stokesii* MHC I prior to next-generation sequencing
E2I2R1	CCA CCT CTC CAC TCA CCT CC	Reverse primer designed following primer trials and used to amplify *E. stokesii* MHC I prior to next-generation sequencing

TrIIβ1F1-tagF	GAG YWC GYG CGC TTC GAC A	*T. rugosa* forward degenerate primer used to amplify *E. stokesii* MHC II prior to next-generation sequencing
ESB1R1	GGT TCT GCC GGT ACA ACT ATG G	Reverse primer designed following primer trials and used to amplify *E. stokesii* MHC II prior to next-generation sequencing
Outer primer components		
P5	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC	Sequencing adapter, required for Illumina MiSeq sequencing
P7	CAA GCA GAA GAC GGC ATA CGA GAT	Sequencing adapter, required for Illumina MiSeq sequencing
Index 2 (i5)	*for example Index_F_1:* TCTCTGTG	Unique forward index sequence used to assign reads to a single sample during post-sequencing processing
Index 1 (i7)	*for example Index_R_1:* ATCGTCTG	Unique reverse index sequence used to assign reads to a single sample during post-sequencing processing
PE Read 1 Sequencing Primer	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT	Sequencing primer which anneals to the template strand
Multiplexing Read 2 Sequencing Primer	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T	Sequencing primer which anneals to the complementary strand
Adapters		
Forward adapter	ACG ACG TTG TAA AA	Anneals the locus specific primer and outer primers
Reverse adapter	CAT TAA GTT CCC ATT A	Anneals the locus specific primer and outer primers

Table 2: Reaction mix and cycling condition of Polymerase Chain Reactions (PCR) used in primer trials and subsequent amplification of *Egernia stokesii* MHC

PCR protocol	Reaction mix	Cycling conditions	Application
Taq ® Gold	Total 25 µl containing 1 X Taq Gold ® Buffer (Applied Biosystems), 2 mM $MgCl_2$, 0.8 mM dNTPs, 0.2µM primer, 0.5 U AmpliTaq ® Gold DNA polymerase (Applied Biosystems), and 2 µl of extracted DNA	Initial denaturation: 10 min at 94 ºC Followed by: 34 cycles of 30 s at 94 ºC, 30 s at 60 ºC, 90 s at 72 ºC Final extension: 10 min at 72 ºC, 30 s at 25 ºC	1. Primer trials
Multiplex-Ready Technology (MRT) (Hayden et al., 2008)	Total 12 µl containing 0.8 mM dNTP, 0.05 mg/ml BSA, 1 X Immolase buffer, 0.3 U Immolase (Bioline), 0.075 µM indexed forward primer, 0.075 µM indexed reverse primer, locus specific primer pair (0.02µM for MHC I, 0.06µM for MHC II), and 2 µl DNA	Initial denaturation: 95°C for 10 minutes Followed by: two PCR phases: First: 5 cycles of 92°C for 60 seconds, 50°C for 90 seconds, and 72°C for 60 seconds, followed by 20 cycles of 92°C for 30 seconds, 63°C for 90 seconds, and 72°C for 60 seconds. Second: 40 cycles of 92°C for 15 seconds, 54°C for 60 seconds, and 72°C for 60 seconds Final extension: 72°C for 10 minutes	1. Primer trials 2. First phase DNA amplification
Immolase	20 µl containing 0.8 mM dNTP, 0.05 mg/ml BSA, 1 X Immolase buffer, 0.5 U Immolase, 0.2 µM indexed forward primer, and 0.2 µM indexed reverse primer	Initial denaturation: 95°C for 10 minutes Followed by: 5 cycles of 94°C for 45 seconds, 54°C for 45 seconds, and 72°C for 60 seconds Final extension: 72°C for 6 minutes	1. Second phase DNA amplification

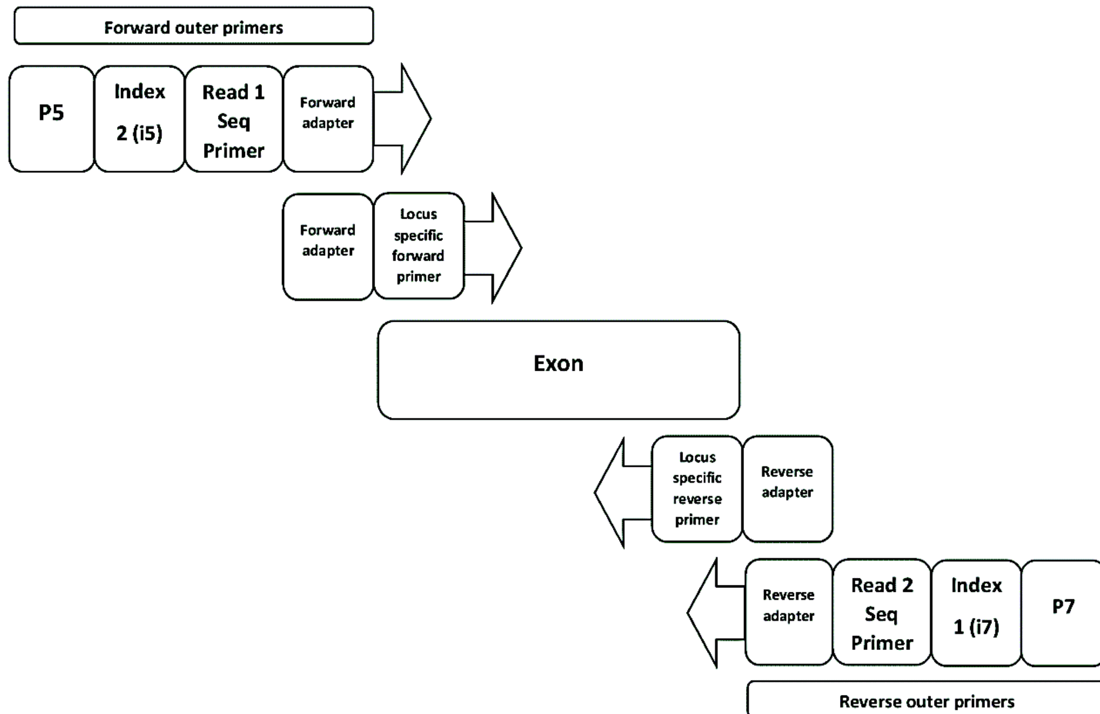


Figure 1: Schematic of primers used to amplify *Egernia stokesii* MHC. Locus specific primers were used during a first phase amplification of DNA. Forward and reverse adapters were incorporated into forward and reverse locus specific primers and outer primers for the purpose of annealing locus specific primers and outer primers during a second phase of DNA amplification. P5 and P7 were sequencing adapters used in Illumina Miseq sequencing. Index 1 and 2 were unique combinations of forward and reverse sequences used to assign sequenced reads to a single sample during post-processing sequencing. Read 1 and Read 2 sequencing primers were used to anneal primers during sequencing.

Appendix 2: Primer design

Although some MHC product was amplified using *T. rugosa* primers in primer trials, amplification success rates were low and inconsistent across samples. Therefore we used *E. stokesii* gDNA sequences which aligned with amino acid sequences of other lizards as described above to design *E. stokesii* locus specific primers.

We used sequence data derived using *Tiliqua rugosa* forward locus specific primer MHC1α 1 F1 (Ansari et al., 2015) and *T. rugosa* reverse locus specific primer Trα1α2r (Ansari, unpublished) to develop locus specific primers for *E. stokesii* exon 2 (α-1) of MHC I (see Table 1, Appendix 1 for primer details).

Sequence data derived using *T. rugosa* forward locus specific primer TrIIβ1F1-tagF and *T. rugosa* reverse locus specific primers TrIIβ1R1_tagR (CAT and TrIIβ1R1B_tagR (Ansari, unpublished) were used to develop locus specific primers for domain β-1 (DRB, exon 2) of MHC II (see Table 1, Appendix 1 for primer details).

For primer design, we initially used primer design tools available in GENEIOUS® 7.1.4 (created by Biomatters, available at www.geneious.com; Kearse et al., 2012). Where primers could not be generated in Geneious due to limited sequence data, we manually designed primers using Oligo Analyzer © (Integrated DNA Technologies, Inc) available online at http://sg.idtdna.com/analyzer/Applications/OligoAnalyzer/ (accessed May 2014). Subsequent primer pairs were trialled in PCRs using a Taq ® Gold protocol (Table 2, Appendix 1) with the same cycling conditions as initial primer trial PCRs except that the PCR annealing temperature was modified to suit the melting temperatures of the primer pair. Amplified PCR products were subsequently cleaned, sequenced and assessed as described in Appendix 1.

Appendix 3: DNA amplification

Following *T. rugosa* locus specific primer pair trials and subsequent *E. stokesii* locus specific primer design (Appendix 1 and 2), we amplified a 216 base pair region of MHC I exon 2 (α-1) using forward locus specific primer E2F1 and reverse locus specific primer E2I2R1 and a 105 base pair region of MHC II exon 2 (β-1) using forward locus specific primer TrIIβ1F1-tagF and reverse locus specific primer ESB1R1 (see Table 1, Appendix 1 for primer details). Because *E. stokesii* MHC has not been fully characterised, it is unknown what proportion of the each region was amplified. However, if *E. stokesii* MHC I is similar to that in *T. rugosa*, the region we amplified may represent 89% of exon 2 (242bp; Ansari et al., 2015). As required for sequencing on the Illumina Miseq platform, locus specific primers were used in combination with forward and reverse adapters and forward and reverse outer primers (see Table 1 and Figure 1, Appendix 1, for primer details and a schematic of the primer concept). Optimal primer concentration was determined by trialling 10, 20, 40, 60, 80 and 100 µM of 0.4 µM locus specific primers.

We initially amplified the MHC using MRT PCRs (Table 2, Appendix 1), later referred to as first phase DNA amplification. However, gel electrophoresis showed the presence of a locus specific band sometimes with and sometimes without the outer primers attached. Therefore we cleaned up the product from the first PCR and amplified the MHC a second time in a second PCR using outer primers only (Immolase protocol; see Table 2, Appendix 1, for reaction mix and cycling conditions), later referred to as second phase DNA amplification. This second phase aimed to anneal the outer primers (required for MiSeq sequencing and for assignment of index sequences to an individual) to the locus specific primers used in the first PCR. Subsequent assessment of product from the second PCR using gel electrophoresis indicated that the outer primers had, in most cases, annealed to the locus specific primers. Only samples for which the outer primers were incorporated were submitted for sequencing.

Appendix 4: Scripts used in bioinformatics relating to identification of *E. stokesii* MHC alleles

4a) PEAR script for assembling paired end reads

```
#unzip the files
for i in */; do gunzip -r "${i%/}.gz" "$i"; done
#move the files from their sub-folders to a new directory
for i in */; do cp $i*.fastq /…/fastq1/; done
#Merge paired-end reads in PEAR version 0.9.5
#Zhang et al (2014) Bioinformatics 30(5): 614-620 | doi:10.1093/#bioinformatics/btt593
#!/bin/bash
PREFIX=$(ls *.fastq | cut -d_ -f1 | uniq)

for p in ${PREFIX}
do
        f1=${p}*R1_001.fastq
        f2=${p}*R2_001.fastq
        pear -f ${f1} -r ${f2} -q 20 -o ${p}
done
```

4b) Script for converting files from fastq to fasta format

```
#FASTX-Toolkit
#http://hannonlab.cshl.edu/fastx_toolkit/index.html
#https://github.com/agordon/fastx_toolkit
#FASTQ-to-FASTA
for i in *.fastq; do fastq_to_fasta -n -i $i -o $i.converted.fasta; done
```

Appendix 5: Species name, common name, and GenBank allele and accession number for species used in amino acid alignment of *Egernia stokesii* MHC I and II. GenBank accession numbers used in the alignment of MHC II alignment are marked with #, all others were used for MHC I.

Species	Common name	GenBank accession number
Amblyrhynchus cristatus	Galápagos marine iguana	Amcr-UB*01: EU604308
		Amcr-UB*02: EU604309
		Amcr-UB*03: EU604310
		Amcr-UB*0401: EU604311
		Amcr-DAB1*01: FJ623746#
		Amcr-DAB2*01: FJ623747#
		Amcr-DAB2*02: FJ623748#
		Amcr-DAB3*01: FJ623749#
		Amcr-DAB3*02: FJ623750#
		Amcr-DAB4*01: FJ623751#
		Amcr-DAB4*02: FJ623752#
Brachydanio rerio	Zebrafish	Brre-DAB1 L04805 #
Conolophus subcristatus	Galápagos land iguana	Cosu-UB*0101: EU604313
		Cosu-UB*02: EU604315
		Cosu-UB*03: EU604316
Egernia stokesii	Gidgee skink	Egst-UA*04 KM515947
Gallus gallus	Chicken	Gaga-B-F10 X12780
		Gaga-BLB1 AL023516 #
Homo sapiens	Human	Hosa-HLA-A U07161
		Hosa-DRB1 M11161 #

Iguana iguana	Green iguana	Igig-UB*0101:EU604317
		Igig-UB*02: EU604319
Macropus rufogriseus	Red-necked wallaby	Maru-UB*01 L04952
Oncorhynchus mykiss	Rainbow trout	Omny-UBA AF287487
Sphenodon punctatus	Tuatara	Sppu-U*01 DQ145788.1
		Sppu-DAB*01 DQ124231 #
Tachyglossus aculeatus	Short beaked echidna	Taac-DZB*01 AY288075 #
Tiliqua rugosa	Sleepy lizard	Tiru-UB*01: KM515952
		Tiru-UB*02: KM515953

Appendix 6: Allele identification

Table 1: Number of unique *Egernia stokesii* MHC I and II sequences retained following filtering, artefact removal, and allele validation

Dataset	MHC region	Number of nucleotide sequences with MPAF > 2 %	Number of amino acid sequences with stop codons (pseudo-genes)	Number of nucleotide sequences with 1 base pair substitution	Number of nucleotide sequences translating to a duplicate amino acid sequence	Total number of nucleotide sequences removed	Total number of nucleotide sequences retained for analysis
1994-98	MHC I	39	1	7	0	8	31
1994-98	MHC II	20	0	13	2	15	5
2012-13	MHC I	74	1	36	2	39	35
2012-13	MHC II	47	1	39	2	42	5

Appendix 7: Results of tests for correlation between number of MHC alleles and total reads, and read coverage thresholds

Table 1: *P*-values from tests for correlation between number of MHC alleles and total reads (Spearman's rho, two-tailed, significant values in bold) and values relating to identification of read coverage thresholds using R script provided by Sommer et al (2013) (doi:10.1186/1471-2164-14-542) used to identify read coverage threshold for 99.9% genotyping accuracy

	MHC I						MHC II					
Dataset	*P*-value	TI_{MEDIAN}	TI_{MIN}	$T1_{MAX}$	$Alelles_{MAX}$	$AmpEff_{MIN}$	*P*-value	TI_{MEDIAN}	TI_{MIN}	$T1_{MAX}$	$Alelles_{MAX}$	$AmpEff_{MIN}$
CAM 1994-98	**< 0.001**	329.50	323	340	16	0.50	0.396	813	798	831	4	0.04
Hawker 2012-13	**< 0.001**	5228.5	5106	5328	21	0.04	0.096	2117	2076	2157	5	0.02

Median T1 value = TI_{MEDIAN}, Minimum T1 value = TI_{MIN}, Maximum T1 value = $T1_{MAX}$, Maximum number of alleles in any sample in the dataset = $Alelles_{MAX}$, Minimum allele amplification efficiency identified using *efficiency.standardised* script = $AmpEff_{MIN}$

References

Ansari, T. H., Bertozzi, T., Miller, R. D., & Gardner, M. G. (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.

Babik, W. (2010). Methods for MHC genotyping in non-model vertebrates. *Molecular Ecology Resources, 10*, 237-251.

Brandariz-Fontes, C., Camacho-Sanchez, M., Vilà, C., Vega-Pla, J. L., Rico, C., & Leonard, J. A. (2015). Effect of the enzyme and PCR conditions on the quality of high-throughput DNA sequencing results. *Scientific reports, 5*.

Hayden, M. J., Nguyen, T. M., Waterman, A., & Chalmers, K. J. (2008). Multiplex-Ready PCR: A new method for multiplexed SSR and SNP genotyping. *BMC Genomics, 9*, 1-12. doi:10.1186/1471-2164-9-80

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*, 1647-1649.

Chapter 4: Selection on skink MHC

Accepted pending changes that have been addressed:

Pearson, S.K., Bull, C.M., and Gardner, M.G., *Egernia stokesii* (gidgee skink) MHC I positively selected sites lack concordance with HLA peptide binding regions, *Immunogenetics*, resubmitted July 2016.

Once the MHC of a target species has been characterised (Chapter 3), the next step in investigating genetic drift, gene flow and selection, and their relative influence on MHC variation, is to identify if selection is acting on the MHC. Further, knowledge of specific amino acid sites under stronger selection is required for reliable downstream analysis. Such sites in human MHC are commonly used to infer those in a study species, yet previous studies demonstrate these sites do not always align across mammalian and non-mammalian species. No previous studies have compared peptide binding regions and sites under selection between human MHC and skink MHC. In the following chapter, the *E. stokesii* alleles and genotypes derived using the methodology described in Chapter 3 are used in tests of selection on the MHC. Then, if *E. stokesii* MHC alleles are under selection, specific amino acid sites under selection are identified and compared to those in humans and other taxa.

***Egernia stokesii* (gidgee skink) MHC I positively selected sites lack concordance with HLA peptide binding regions**

Pearson, Sarah K.[A, B], Bull, C. Michael[A], Gardner, Michael G.[A, C]

[A]School of Biological Sciences, Flinders University of South Australia, Bedford Park 5042, Australia

[C]Evolutionary Biology Unit, South Australian Museum, North Terrace, Adelaide 5000, Australia.

[B]Corresponding author. Email: sarah.pearson@flinders.edu.au; Phone: +618 82012315; Fax: +61 8 8201 3015

Key words: MHC, peptide binding region, positively selected sites, Squamata, Scincidae, *Egernia stokesii*

***Immunogenetics*, resubmitted July 2016**

Abstract

Genes of the major histocompatibility complex (MHC) play an important role in vertebrate disease resistance, kin recognition and mate choice. Mammalian MHC is the most widely characterised of all vertebrates and attention is often given to the peptide binding regions of the MHC, because they are presumed to be under stronger selection than non-peptide binding regions. For vertebrates where the MHC is less well understood, researchers commonly use the amino acid positions of the peptide binding regions of the human leucocyte antigen (HLA) to infer the peptide binding regions within the MHC sequences of their taxon of interest. However, positively selected sites within MHC have been reported to lack correspondence with the HLA in fish, frogs, birds, and reptiles including squamates. Despite squamate diversity, the MHC has been characterised in few snakes and lizards. The *Egernia* group of scincid lizards is appropriate for investigating mechanisms generating MHC variation, as their inclusion will add a new lineage (i.e. Scincidae) to studies of selection on the MHC. We aimed to identify positively selected sites within the MHC of *Egernia stokesii* and then determine if these sites corresponded with the peptide binding regions of the HLA. Six positively selected sites were identified within *E. stokesii* MHC I, only two were homologous with the HLA. *Egernia stokesii* positively selected sites corresponded more closely to non-lizard than other lizard taxa. The characterisation of the MHC of more intermediate taxa within the squamate order is necessary to understand the evolution of the MHC across all vertebrates.

Introduction

Genes of the highly polymorphic major histocompatibility complex (MHC) play an important role in vertebrate disease resistance, kin recognition and mate choice (reviewed in Milinski 2006; Piertney and Oliver 2006; Sommer 2005). The MHC comprises a cluster of genes which encode cell surface molecules, binding pathogen peptide fragments to elicit an immune response (reviewed in Neefjes et al. 2011). These interactions occur at peptide binding regions, recognition sites for presentation to T cells (Knapp 2005). Selection, which acts on MHC alleles, accounts for the remarkable diversity of the MHC, with selection being strongest at the peptide binding regions (Edwards and Hedrick 1998; Hughes and Yeager 1998). The two central hypotheses to explain the observed high levels of MHC variation within populations are parasite mediated selection and MHC dependent sexual selection (Apanius et al. 1997; Milinski 2006; Piertney and Oliver 2006; Spurgin and Richardson 2010). Based on tests of these hypotheses, evidence of selection on the MHC is widespread (Apanius et al. 1997; Bernatchez and Landry 2003; Klein et al. 1993) yet debate persists about the relative roles of the mechanisms of selection, and factors that influence them, in generating MHC variation (Bernatchez and Landry 2003; Spurgin and Richardson 2010; Winternitz et al. 2013). Characterising MHC variation and identifying signatures of selection within that genetic complex, across a range of taxa, is a fundamental step in gaining greater clarity regarding the mechanisms that generate MHC variation and the broad aim of this paper is to contribute to that step in a reptile system.

The mammalian MHC, which is the most widely characterised of all vertebrates to date, can be divided into four classes: I, II, III, and IV (Acevedo-Whitehouse and Cunningham 2006). Studies commonly focus on MHC I and II because they are highly variable compared to III and IV (Acevedo-Whitehouse and Cunningham 2006). Further, attention is often given to the peptide binding regions of the MHC of class I and II because they are presumed to be under

stronger selection than non-peptide binding regions as argued above. The MHC in other vertebrates is less well known. Therefore, although differences to mammalian MHC have been found in birds, fish, and amphibians (reviewed in Kelley et al. 2005), mammalian MHC is usually the reference point in the absence of other data. For some mammals, and for non-mammals where the MHC structure and function is less well known, researchers commonly use the amino acid positions of the peptide binding regions of the human leucocyte antigen (HLA, the human MHC) to infer the sites involved in peptide binding within the MHC sequences of their taxon of interest (Strandh et al. 2011; Wang et al. 2010; Wegner 2008). However, the corresponding codons in other taxa are not always found to be under selection. Notably, codons within non-mammalian MHC identified to be under selection (i.e. positively selected sites within the non-mammalian MHC amino acid sequence), are not always concordant with the peptide binding regions of the HLA (Glaberman and Caccone 2008; Wegner 2008). If codons under selection differ from those inferred from the HLA, testing hypotheses about the mechanisms of selection becomes difficult. In addition, inferring peptide binding sites in non-mammalian MHC from the HLA can falsely assign a functional role to a codon (Wegner 2008), with implications for the reliability of analyses based on such codons.

Positively selected sites within MHC have been reported to lack correspondence with the HLA in fish (*Acipenser sinensis*; Wang et al. 2010), frogs (*Agalychnis callidryas*; Kiemnec-Tyburczy et al. 2012), birds (*Halobaena caerulea*; Strandh et al. 2011), and reptiles including sphenodontids (the tuatara *Spenodon punctatus*; Miller et al. 2007) and squamates (*Iguana iguana*; Glaberman and Caccone 2008). Although there are more than 9500 squamate species (Uetz 2015), the MHC has been characterised in remarkably few snakes and lizards and knowledge of squamate MHC structure is limited (Elbers and Taylor 2016; Miller et al. 2006). Squamates for which the MHC has been partly characterised include representatives

of Pythonidae (*Liasis fuscus*; Madsen and Ujvari 2006), Polychrotidae (*Anolis sagrei*; Hung 2013), Iguanidae (e.g. *Iguana iguana*; Glaberman and Caccone 2008), Agamidae (*Ctenophorus ornatus*; Radwan et al. 2014) and, more recently, Scincidae (*Egernia stokesii*, Pearson et. al. Chapter 3; *Tiliqua rugosa*, Ansari et al. 2015).

Both *E. stokesii* and *T. rugosa* are members of the relatively well-studied *Egernia* group of scincid lizards. The *Egernia* group is appropriate for investigating mechanisms generating MHC variation as their inclusion will add a new lineage (i.e. Scincidae) to studies of selection on the MHC and should contribute to an improved understanding of the evolution of peptide binding regions and MHC structure across squamates. Using *Egernia stokesii* as the study species, the specific aim of this study was to identify positively selected amino acid sites within the MHC of *E. stokesii* and then determine if the sites identified as being under selection corresponded with the peptide binding regions of the HLA. Based on previous findings for lizards, (Glaberman and Caccone 2008; Glaberman et al. 2009; Radwan et al. 2014), we predicted that positively selected codons in *E. stokesii*, if detected, would lack concordance with the HLA. Our results will contribute information on MHC regions under selection from another lineage of squamate reptiles, the skinks, and add to current understanding of the codons potentially involved in targeting pathogens. In addition, this work should strengthen the impetus for investigations into non-mammalian MHC structure and function.

Method

Study species

Egernia stokesii is a large (180 mm snout-vent length), long lived, viviparous lizard, widely distributed across semi-arid Australia (Cogger 1983). Populations in southern Australia are confined to rocky outcrops where the lizards occupy crevice refuges for shelter. Both the

number of lizards and the number of social groups within a population are constrained by the availability of rocky crevices and resident lizards in each outcrop population exhibit limited dispersal (Gardner et al. 2001). This has been confirmed by analyses of microsatellite DNA data that show geographically adjacent populations are genetically isolated (Pearson unpublished data). Within populations, *Egernia stokesii* form stable family groups (Duffield and Bull 2002; Gardner et al. 2001) with high levels of social and genetic monogamy (Gardner et al. 2002), delayed maturity and limited between-group dispersal (Duffield and Bull 2002). Individuals harbour a diverse array of parasite species (Duffield and Bull 1996; Hallas et al. 2005; Keirans et al. 1996; Stein 1999; Stein and Dyce 2002; Telford and Stein 2000), with varying prevalence among populations (Stein 1999). At least one group of parasites, the gut nematodes, affect *E. stokesii* basking time and movement, suggesting infection related fitness consequences (Fenner and Bull 2008). Using social network approaches, Godfrey et al. (2009) found more socially connected *E. stokesii* had higher parasite loads than less connected lizards. *Egernia stokesii* are capable of recognising kin and group and non-group members, possibly via the use of chemical cues which may also play a role in mate choice and social group cohesion in this species (Bull et al. 2000; Main and Bull 1996).

Sample collection

Egernia stokesii were sampled during field surveys undertaken between 1994-1998 at one population (Camel Hill, CAM) and between 2012-2013 at three populations (CAM, Castle Rock (CAS), and Castle Rock Ridge (CRR)) (hereafter termed 1994-98 and 2012-13 datasets respectively), all near Hawker (31°54′S; 138°25′E) in the southern Flinders Ranges, South Australia. Each population was located on a rocky outcrop with crevices for *E. stokesii* refuges and separated from each other population by 300 m – 1.5 km of unsuitable *E. stokesii* habitat (Gardner et al. 2007). Sample collection has been described in detail elsewhere

(Gardner et al. 2007; Gardner et al. 2001; Gardner et al. 2002; Godfrey et al. 2006; Pearson et al. Chapter 5). A total of 411 unique lizards were captured consisting of 151 from CAM during 1994-98 field surveys (Duffield and Bull 2002; Gardner et al. 2001) and 260 individuals from CAM, CAS, and CRR during 2012-13 field surveys (Pearson et al., Chapter 5).

MHC alleles

The methodology used to identify MHC alleles (defined here as unique nucleotide sequences that translated to unique amino acid sequences) and derive MHC genotypes is detailed in Pearson et al. (Chapter 3). Therefore, we only provide brief details here. We derived genotypes for a 216 base pair region of MHC I exon 2 (corresponding to the α-1 domain) and for a 102 base pair region of MHC II exon 2 (corresponding to the β-1 domain) using Illumina Miseq next-generation sequencing. Because this technology necessitated samples from different individuals to be pooled, we needed a method of extracting data for each individual from the sequencing results. We achieved this in a two-stage polymerase chain reaction (PCR) process. PCR cycling conditions are provided in Appendix 1, Pearson et al. Chapter 3 Supplemental Information.

First, locus specific primers were used to amplify the MHC. We initially trialled primers developed for *Tiliqua rugosa*, another member of the *Egernia* group of lizards (Ansari 2016; Ansari et al. 2015). Using *T. rugosa* primers, some sequences were derived which we used to design species-specific primers. Subsequently, for the first stage PCR for MHC I, we used forward locus specific primer E2F1 (5' - GTG TCG GAG CCT GGC CAG - 3') and reverse locus specific primer E2I2R1 (5' - CCA CCT CTC CAC TCA CCT CC - 3'). For MHC II, we used *T. rugosa* forward primer TrIIβ1F1-tagF (5' - GAG YWC GYG CGC TTC GAC A - 3') and reverse locus specific primer ESB1R1 (5' - GGT TCT GCC GGT ACA ACT ATG G - 3'). In the second stage, we used the cleaned up product from the first PCR in a second PCR

using only primers required for 1) individual identification, and 2) Illumina Miseq sequencing (hereafter called outer primers, Pearson et al., Chapter 3). We included multiplex identifier (MID) tags (Meyer and Kircher 2010) in the outer primers for the purpose of identifying unique sample sequencing data. Every sample was assigned a unique combination of MID tags.

Samples were sequenced as 300 base pair paired end reads. The regions sequenced corresponded to 82% and 34% of the HLA for MHC I and MHC II respectively. We identified codons corresponding to the peptide binding regions based on amino acid alignment of the HLA (Bondinas et al. 2007; Chelvanayagam 1996; Reche and Reinherz 2003) with *E. stokesii* MHC alleles. The regions sequenced included thirteen and seven putative peptide binding regions in MHC I and II respectively (Fig. 1 and 2).

We used jMHC version 1.6.1624 https://code.google.com/p/jmhc (Stuglik et al. 2011) to extract unique sequences from Miseq output data. We filtered sequences based on expected base pair length and occurrence in a minimum of three reads in a minimum of two samples (Babik 2010; Galan et al. 2010). We then adopted the approach of (Radwan et al. 2012) to further filter sequences based on per sequence maximum per amplicon frequency (2 %) before reviewing sequences for artefacts (single nucleotide base substitutions, pseudogenes, and chimeras). Then, we validated sequences retained after artefact filtering using the standard nucleotide Basic Local Alignment Search Tool (Altschul et al. 1990) and amino acid alignment against the HLA and other taxa. For genotyping, we adopted the approach and R scripts of (Sommer et al. 2013) (doi:10.1186/1471-2164-14-542) to determine the minimum number of reads per sample for reliable genotyping. Genotyping error rates, determined by the use of technical (PCR) duplicates, were 5.14% for MHC I (n=14) and 4.84% for MHC II (n=18).

Thirty-nine MHC I alleles and five MHC II alleles were identified (Fig. 1 and 2) and used to derive genotypes for 253 lizards for MHC I and 144 lizards for MHC II. The 39 MHC I alleles consisted of 27 alleles common to the 1994-98 and 2012-13 datasets, eight unique to the 2012-13 dataset (occurring in 2 - 21 individuals) and four unique to the 1994-98 dataset (occurring in 5 - 13 individuals) (Fig. 1). Because of the differences in the alleles identified in each sampling period, the MHC I 1994-98 and 2012-13 datasets were analysed separately. However, because all five MHC II alleles were common to both sampling periods we pooled the samples for a single MHC II allele dataset. Across the combined 1994-98 and 2012-13 datasets there was one to 21 (mean 8.61, SE ± 0.20) MHC I alleles per individual and one to five (mean 2.14, SE ± 0.07) MHC II alleles per individual (Table 1). Because the primers used to amplify the MHC were not locus-specific, we did not assign alleles to MHC loci. We considered signatures of selection and patterns of variation that we detected to be relevant to the function of the whole gene region, in line with the approach of other researchers (Herdegen et al. 2014; Radwan et al. 2014).

Identification of positively selected sites

We identified positively selected sites within *E. stokesii* MHC alleles in three steps. First, because overall allelic variability may be indicative of selection and amino acid variability may indicate specific amino acid sites subject to selection, we undertook a preliminary assessment of *E. stokesii* MHC variability using two approaches: 1) variability at each amino acid site and 2) sequence variation among alleles. The Wu-Kabat variability coefficient (Wu and Kabat 1970) was used to identify amino acid sites with high variability. The Wu-Kabat variability coefficient is calculated by multiplying the number of sequences in the alignment (N) and the number of different amino acids at a given position (k) and dividing N*k by the number of times that the most common amino acid at that position is present (n). Based on Wu-Kabat variability coefficient values, amino acid sites may be monomorphic (= 1),

polymorphic (≥ 2), or highly polymorphic (sites with more than twice the mean Wu-Kabat variability coefficient) (Eimes et al. 2015; Sun et al. 2014).Wu-Kabat plots were generated in the Protein Variability Server available online at http://imed.med.ucm.es/PVS/ (Díez-Rivero and Reche 2009; Garcia-Boronat et al. 2008).

We then investigated sequence variation among alleles by comparing arithmetical mean rates of non-synonymous and synonymous amino acid substitutions calculated using the Nei–Gojobori method (Nei and Gojobori 1986) with the Jukes & Cantor correction for multiple substitution in Mega 6 (Tamura et al. 2013). We calculated standard errors using 1000 bootstrap replicates. Sites with high variability and elevated rates of non-synonymous substitutions are indicative of selection (Bernatchez and Landry 2003; Hughes and Nei 1988; Hughes and Yeager 1998).

Second, after assessing MHC variability, we determined if the regions of *E. stokesii* MHC I and II that we sequenced had been under historical (i.e. over evolutionary time) positive or negative selection. Negative (or purifying) selection would suggest that the associated gene fragment plays a structural or functional role that is independent of pathogen diversity, whilst positive selection may suggest adaptive responses to pathogen diversity (Bernatchez and Landry 2003; Hughes 2007). We used a common method to detect historical selection on the MHC, comparing ratios of non-synonymous to synonymous nucleotide substitutions, with a higher ratio suggesting stronger positive selection (Bernatchez and Landry 2003). Amino acid change arising from nonsynonymous mutations are likely to be the effect of selection (Bernatchez and Landry 2003). Ratios of non-synonymous/synonymous substitutions (dN/dS) >1 (also represented as $\omega > 1$) indicate positive selection and dN/dS < 1 indicate negative selection (Garrigan and Hedrick 2003; Hughes and Yeager 1998; Kryazhimskiy and Plotkin 2008).

We used two methods to calculate ratios of non-synonymous to synonymous substitutions. First, models of codon evolution were compared using the package CodeML implemented in PAML 4.8 (Xu and Yang 2013; Yang 2007). Three models were tested 1) M0: a single ω (dN/dS ratio) for all codons (the null model), 2) M7: nearly neutral ($\omega \leq 1$), with ω variation approximated by β-distribution, and 3) M8: positive selection (proportion of codons with $\omega > 1$), with ω variation approximated by β-distribution. The best supported model was selected based on the lowest value of the Akaike Information Criterion (AIC) (Posada and Buckley 2004), corrected for small samples sizes (AICc) (Hurvich and Tsai 1989). If the M8 model of codon evolution was determined to have the greatest support, we inferred there had been positive selection. Then, if selection was detected, we attempted to identify from within the nucleotide sequences of *E. stokesii* MHC alleles, the specific codons under positive selection using a Bayes empirical Bayes procedure (Yang et al. 2005). We retained positively selected sites with a posterior probability > 0.95 % for further analysis. We subsequently used a second more conservative measure than that used in the CodeML package for detecting selection (Huchard et al. 2012; Pechouskova et al. 2015): the Z-test of selection with the modified Nei-Gojobori/Jukes-Cantor method with 1000 bootstraps in Mega6. This technique is based on an evolutionary pathways method (Nei and Gojobori 1986) that applies the Jukes & Cantor correction for multiple substitutions. Z-test significance values were adjusted for multiple tests using a false discovery rate test (threshold of 0.1, Benjamini and Hochberg 1995).

Comparison of E. stokesii *positively selected sites to HLA peptide binding regions and other taxa*

Once we identified positively selected sites within *E. stokesii* MHC (see results), we compared them to the corresponding regions in other taxa. Firstly, for exploratory purposes, we identified conserved sites among *E. stokesii* MHC I and II amino acid sequences and

amino acid sequences from the corresponding regions of other taxa (Appendix 5, Pearson et al. Chapter 3 Supplemental Information). The identification of conserved sites was undertaken because the presence of conserved sites across taxa assists investigations of evolutionary relationships (Glaberman and Caccone 2008). MHC nucleotide sequences were aligned, then translated into amino acid sequences in Mega6. We identified sites conserved across 1) all taxa and 2) all lizards in the alignment.

Secondly, we compared *E. stokesii* positively selected MHC sites to the peptide binding regions of the HLA and other taxa, for MHC I only. Analogous comparisons were not performed for MHC II because evidence for historical selection was not detected (see results). We included amino acid sequences of taxa for which positively selected sites had been identified in earlier studies, including fish *(A. sinensis*; Wang et al. 2010), birds (*H. caerulea*; Strandh et al. 2011), frogs (*A. callidryas*; Kiemnec-Tyburczy et al. 2012), lizards (*I. iguana*; Glaberman and Caccone 2008) and tuatara (*S. punctatus*; Miller et al. 2007), to extend the comparison beyond *E. stokesii* and the HLA (Appendix 5, Pearson et al. Chapter 3 Supplemental Information).

Results

Identification of positively selected sites

Amino acid variability, plotted against the peptide binding regions of the HLA, is shown in Wu-Kabat plots (Fig. 3). Sixty three (88 %) MHC I sites were polymorphic (Wu-Kabat variability coefficient ≥ 2) of which seven were highly polymorphic (Wu-Kabat variability coefficient ≥ 8.78). Three of the highly polymorphic MHC I sites corresponded with, and three were adjacent to, HLA peptide binding regions. For MHC II, 16 (47%) sites were polymorphic of which one was highly polymorphic (Wu-Kabat variability coefficient ≥ 3.61) and adjacent to a HLA peptide binding region.

Because positively selected sites were detected in MHC I but not MHC II (see below), we report amino acid variability, synonymous and nonsynonymous substitution rates, and results of Z-tests for selection among complete sequences (i.e. 216 bp) and positively and non-positively selected sites for MHC I. For MHC II, we report results for complete sequences (i.e. 102 bp) and putative peptide binding and non-peptide binding regions. Both the proportion of variable amino acid sites and the average rate of non-synonymous substitutions were higher in positively selected sites than non-positively selected sites for MHC I and in putative peptide binding regions than non-peptide binding regions for MHC II (Table 2, Fig. 4). Both synonymous and non-synonymous substitutions were higher in positively selected sites than non-positively selected sites in MHC I (Table 2, Fig. 4). For MHC II, non-synonymous substitutions were higher in putative peptide binding regions than non-peptide binding regions whereas synonymous substitutions were higher in putative non- peptide binding regions than putative peptide binding regions (Table 2, Fig. 4).

Prior to identifying positively selected sites, we tested for historical selection on the MHC using ratios of synonymous to non-synonymous amino acid substitutions. For MHC I, the model of codon evolution allowing for a fraction of codons under positive selection (M8) had more support than other models (M0, M7), whereas the model of one synonymous/non-synonymous ratio for all codons (M0) had the greatest support for MHC II (Table 3). The Bayes empirical Bayes procedure identified six codons evolving under positive selection in MHC I. For the 1994-98 dataset, four positively selected sites were identified: 53 and 60 with posterior probabilities (*PP)* > 0.95 % and sites 54 and 57 with *PP* > 0.99 % (Fig. 1 and 3a). For the 2012-13 dataset, we identified five positively selected sites: sites 27, 42, and 57 with *PP* > 0.95 % and sites 54 and 60 with *PP* > 0.99 % (Fig. 1 and 3a). Three MHC I sites (54, 57, and 60) were identified as positively selected in both the 1994-98 and 2012-13 datasets. Based on Wu-Kabat variability coefficient values, all MHC I positively selected sites (27, 42,

53, 54, 57, and 60) were polymorphic, of which four (42, 54, 57, and 60) were highly polymorphic (Fig. 3a). We did not apply the Bayes empirical Bayes procedure to MHC II data, as M8 was not the best-supported model.

Z-tests for selection indicated positive selection on MHC I for the entire region and positively selected sites for 1994-98 ($P = 0.010$ and $P < 0.001$ respectively) and 2012-13 ($P = 0.014$ and $P < 0.001$ respectively) datasets () (Table 2). Further, Z-tests indicated positive selection on MHC I non-positively selected sites for the 1994-98 dataset ($P = 0.030$) but not the 2012-13 dataset ($P = 0.070$). For MHC II, Z-tests for selection were not significant (all $P > 0.05$) (Table 2).

Comparison of E. stokesii *positively selected sites to HLA peptide binding regions and other taxa*

Four MHC I and three MHC II *E. stokesii* amino acid sites were conserved among all taxa included in the amino acid alignment and nine MHC I and eleven MHC II *E. stokesii* amino acid sites were conserved with other lizard amino acid sequences (Fig. 1 and 2). Two *E. stokesii* MHC I positively selected sites (54 and 57) occurred at homologous positions with HLA peptide binding regions and four sites (27, 42, 53 and 60) occurred at sites adjacent to an HLA peptide binding region (Fig. 1). MHC I positively selected site 54, in *E. stokesii,* was also under selection in fish, frogs, and tuatara, in addition to being homologous to HLA peptide binding regions. Similarly, *E. stokesii* MHC I site 57 matched those under selection in fish, frogs, tuatara and iguanids (Fig. 1). *Egernia stokesii* positively selected sites adjacent to HLA peptide binding regions matched positively selected sites in fish and frogs (site 27) and frogs and birds (site 53) (Fig. 1). Two *E stokesii* positively selected sites (sites 42 and 60) lacked correspondence with HLA peptide binding regions and positively selected sites in other taxa (Fig. 1). In summary, of the six *E. stokesii* MHC I positively selected sites

identified, four matched sites under selection in frogs, three in fish, two in tuatara, and one in both iguanids and birds, with two being specific to *E. stokesii* (Fig. 1).

Discussion

Human MHC (HLA) is commonly used to infer peptide binding regions, i.e. codons at which selection is predicted to occur, in other taxa. We detected positive selection in *E. stokesii* MHC I but not MHC II and identified six positively selected *E. stokesii* MHC I sites, of which two were homologous with, and four adjacent to, HLA peptide binding regions. No positively selected sites were common among all taxa included in our alignment. A greater proportion of the six *E. stokesii* MHC I positively selected sites were most closely aligned to sites under selection in frogs, not lizards. Only one site was identified as being under positive selection in both an iguanid and *E. stokesii*. As Scincidae is an older lineage than Iguanidae (Pyron et al. 2013), the closer correspondence of positively selected sites between *E. stokesii* and frogs rather than *E. stokesii* and the iguanid in the aligned taxa, suggests that changes in the putative peptide binding regions in the iguanid occurred since the skink and iguanid lineages split and that these sites in skinks have retained ancestral function. Our findings support earlier work demonstrating a lack of concordance between non-mammalian positively selected sites and human peptide binding regions in fish (reviewed in Wegner 2008), birds (Alcaide et al. 2013; Radwan et al. 2012; Sutton et al. 2013), frogs (Wang et al 2010), iguanid lizards (Glaberman and Caccone 2008) and tuatara (Miller et al. 2007). Thus, our results add a new lineage, the skinks, to comparisons of non-mammalian and mammalian MHC, strengthen the findings from earlier comparative studies, and emphasise calls for further research into the structure and evolutionary history of non-avian reptile MHC (Jaratlerdsiri et al. 2014; Miller et al. 2015).

We found a strong signal historical selection on *E. stokesii* MHC I but positive selection was not detected for MHC II. Although the region of MHC II used in this study included seven

putative peptide binding regions, it covered only 34% of the corresponding region in humans. Therefore the detection of selection on MHC II may have been difficult and possibly the sites under selection were not sequenced in our study. In addition, the low diversity as exhibited in *E. stokesii* MHC II can be problematic for detecting selection. Although selection was not evident, non-synonymous substitutions were higher in putative peptide binding regions than non-peptide binding regions in MHC II, which suggests that the location of the peptide binding regions in *E. stokesii* MHC II might match those in humans. Where both MHC I and II have been investigated within a single species, selection is commonly detected in both regions (Bonneaud et al. 2004; Pokorny et al. 2010). However, this is not always the case. A study of MHC I and II in marmots found evidence of selection in MHC II not I (Kuduk et al. 2012). This suggests the strength of selection from parasites and mate choice may be different for each region. Incorporating both regions in future studies will enable comparisons of selection on MHC I and II and should improve our understanding of the relative effect of parasite-mediated selection and MHC based mate choice on MHC variation.

Although high levels of polymorphism is a feature of the MHC, levels of variability can differ between MHC classes. *Egernia stokesii* MHC I had higher allelic variability than MHC II did. Although selection is associated with variability, *E. stokesii* MHC I positively selected sites did not always occur at positions with the highest amino acid variation. *Egernia stokesii* MHC II sites with the highest amino acid variation corresponded more closely to HLA peptide binding regions than in MHC I, which suggests selection may be occurring at these sites in MHC II but we were not able to detect it. Differences in variability between *E. stokesii* MHC I and II alleles demonstrated in this study are consistent with contrasting patterns of variability between MHC I and II previously identified in fish e.g. *Salmo salar* (Consuegra et al. 2011; Consuegra et al. 2005); mammals e.g. *Marmota marmot* (Kuduk et al. 2012), and birds e.g. *Halobaena caerulea* (Strandh et al. 2011). Several reasons have been

proposed to explain contrasting variability between MHC classes. For example, different peptide binding modes may differentially influence selection processes (Consuegra et al. 2005). Further, intra- and extra-cellular parasites may impose different selective pressures on MHC I and II respectively (Kuduk et al. 2012) and MHC I and II may be differentially involved in mate choice for example via the use of odour cues (Strandh et al. 2012; Wedekind and Penn 2000). The contrasting variability between MHC regions demonstrated by earlier studies emphasise the complexity of the mechanisms and factors generating MHC variation.

This study, using a member of the *Egernia* group of lizards, *E. stokesii*, extends the data available for comparison of non-mammalian positively selected sites and HLA peptide binding regions, and adds the scincid lineage to those squamates for which selection on the MHC has been investigated. We found high allelic variability and evidence of selection for *E. stokesii* MHC I, but not MHC II. Six positively selected sites were identified within *E. stokesii* MHC I alleles, of which only two were homologous with the HLA. The sites at which positive selection was detected in *E. stokesii* corresponded more closely to non-lizard than lizard taxa. The characterisation of the MHC of more intermediate taxa within the squamates is necessary to understand the evolution of the MHC within this group and subsequently across all vertebrates.

Acknowledgements

We thank volunteers for assistance with field surveys; Kathy Saint and Terry Bertozzi for assistance with laboratory work; and Katina Krasnec for assistance with bioinformatics. We undertook laboratory components of this work at the South Australian Research Facility for Molecular Ecology and Evolution, Adelaide. Funding from The Field Naturalists Society of South Australia and the Sir Mark Mitchell Research Foundation supported this work.

Ethical approval: All applicable international, national, and/or institutional guidelines for the care and use of animals were followed. All procedures performed in studies involving animals were in accordance with the ethical standards of the institution at which the studies were conducted.

Conflict of Interest: The authors declare that they have no conflict of interest.

Table 1: Summary of *Egernia stokesii* individuals and MHC I & MHC II alleles derived from sequencing a 216 base pair region of MHC I exon 2 (α-1) and a 102 base pair region of MHC II exon 2 (β-1).

MHC I	CAM[a]	CAM[b]	CAS	CRR	Total 12-13	Total All
Number of lizards genotyped for MHC	55	64	70	64	198	253
Total number of MHC alleles	31	32	31	31	35	39
Mean MHC alleles per individual	8.96 (SE±0.39)	8.22 (SE±0.36)	9.09 (SE±0.39)	8.16 (SE±0.41)	8.51 (SE±0.23)	8.61 (SE±0.20)
Minimum MHC alleles per individual	2	1	2	2	1	1
Maximum MHC alleles per individual	16	16	17	21	21	21
MHC II						
Number of lizards with MHC genotypes	55	35	26	28	89	144
Total number of MHC alleles	5	5	3	5	5	5

Mean MHC alleles per individual	2.16 (SE±0.12)	2.09 (SE±0.12)	1.77 (SE±0.10)	2.50 (SE±0.17)	2.12 (SE±0.08)	2.14 (SE±0.07)
Minimum MHC alleles per individual	1	1	1	1	1	1
Maximum MHC alleles per individual	4	4	3	5	5	5

Sites sampled were CAM[a]: Camel Hill 1994-98, CAM[b]: Camel Hill 2012-13, CAS: Castle Rock and CRR: Castle Rock Ridge both 2012-13. Total 12-13 is the total for CAM[b], CAS, CRR. Total All is the total for CAM[a], CAM[b], CAS, CRR.

Table 2: Rates of synonymous and non-synonymous substitutions calculated for *Egernia stokesii* MHC I and II alleles. Synonymous and non-synonymous substitutions were calculated in Mega6 separately then together in Z-tests for selection. Calculations were performed separately for MHCI and MHCII. There were some differences in MHC I alleles in the 1994-98 and 2012-13 datasets therefore the two datasets were analysed separately whereas MHC II alleles were identical in the 1994-98 and 2012-13 datasets therefore analysed only once. Positively selected sites were identified using the Bayes empirical Bayes procedure in PAML for MHC I. For MHC II, peptide binding regions were based on human leucocyte antigen (HLA). Metrics were calculated using 31 MHC I alleles for 1994-98, 35 MHC I alleles for 2012-13 and five alleles for MHC II (identical alleles for 1994-98 and 2012-13).

Dataset	Codons (n)	Variable sites (n)	dN (±SE)	dS (±SE)	*P*	Z
MHC I - 1994-98[a]						
All	72	60	0.186 (0.023)	0.188 (0.028)	**0.010***	2.361
PSS	4	4	0.733 (0.129)	0.349 (0.098)	**0.000***	3.900
Non-PSS	68	45	0.162 (0.019)	0.179 (0.028)	**0.030***	1.895
MHC I - 2012-13[b]						
All	72	63	0.243 (0.029)	0.259 (0.043)	**0.014***	2.231

PSS	5	5	0.751 (0.074)	0.272 (0.081)	**0.000***	6.366
Non-PSS	67	58	0.216 (0.026)	0.255 (0.044)	**0.070**	1.487
MHCII - 1994-98 & 2012-13						
All	34	16	0.121 (0.030)	0.201 (0.081)	1.000	-0.424
PBR	7	6	0.287 (0.162)	0.136 (0.191)	0.195	0.861
Non-PBR	27	10	0.090 (0.026)	0.209 (0.091)	1.000	-0.882

dN and dS (±SE): non-synonymous and synonymous rates and standard error calculated in Mega6 using the modified Nei & Gojobori (1986) method with the Jukes-Cantor (1969) correction for multiple substitutions

P & Z: *Z*-test statistic value and *P* value of the *Z*-test of the alternate hypothesis of positive selection (HA: dN > dS), calculated in Mega6 using Nei & Gojobori (1986) method with the Jukes-Cantor (1969) correction for multiple substitutions, * = significant at P < 0.05. Bold values indicate significant after applying false discovery rate (threshold 0.1) correction for multiple tests.

[a] Samples from Camel Hill 1994-98 field surveys

[b] Samples from Camel Hill, Castle Rock, Castle Rock Ridge 2012-13 field surveys

Table 3: Evaluation of the goodness of fit for different models of codon evolution and estimated parameter values for *Egernia stokesii* MHC I (1994-98 = 31 alleles, 2012-13 = =35 alleles) and MHC II (1994-98 and 2012-13 = 5 alleles). There were some differences in MHC I alleles in the 1994-98 and 2012-13 datasets therefore the two datasets were analysed separately whereas MHC II alleles were identical in the 1994-98 and 2012-13 datasets therefore analysed only once. The model of positive selection (M8) had the greatest support for the MHC I whereas the model of one ω ratio had the greatest support for MHC II.

Model	ln L	AIC_c	ΔAIC_c	Parameters
MHC I - 1994-98[a]				
M8	-1694.029	3389.596	0 - best	$p0 =$ 0.937, $p1 =$ 0.064, $\omega^2 =$ 6.231
M7	-1717.728	3435.885	46.289	
M0	-1735.267	3470.672	81.076	$\omega^1 = 1.147$
MHC I - 2012-13[b]				
M8	-1793.423	3588.179	0 - best	$p0 =$ 0.703, $p1 =$ 0.297, $\omega^2 =$ 2.246
M7	-1800.225	3600.824	12.645	

M0	-1814.213	3628.547	40.368	$\omega^1 = 1.05123$
MHC II				
M0	-270.329	541.992	0 - best	$\omega = 0.541$
M7	-270.289	546.577	4.585	
M8	-269.994	579.987	37.995	$p0 = 0.939$, $p1 = 0.061$, $\omega^2 = 2.862$

M0 (one ω ratio), M7 (nearly neutral with beta), M8 (positive selection with beta ($\omega0 \leq 1$, $\omega1 > 1$)); AICc (AIC with bias adjustment for small sample sizes); ΔAICc Difference between the value of the AICc of a model and the best model; $\omega1$ = dN/dS; $\omega2$ = Estimated ω for sites under positive selection; p0 = Proportion of sites with $\omega \leq 1$; p1 = Proportion of positively selected sites ($\omega > 1$); [a] 1994-98 - Camel Hill sampled 1994-98; [b] 2012-13 - Camel Hill, Castle Rock, Castle Rock Ridge sampled 2012-13.

Site number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22			23	24			25	26	27	28	29	30	31	32	33	34	35	36	37	38		39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72
Hosa-HLA-A	G	E	P	R	F	I	A	V	G	Y	V	D	D	T	Q	F	V	R	F	D	S	D	A	A	S	Q			K	M	E	P	R	A	P	W	I	-	-	-	E	Q	-	E	G	P	E	Y	W	D	Q	E	T	R	N	M	K	A	H	S	Q	T	D	R	A	N	L	G	T	L	R	G	Y	Y	N	Q	S
Igig-UB*0101	E	V	.	Q	.	F	S	.	.	.	.	.	.	Q	.	.	.	S	Y	.	T	N	-	-	T	R			R	R	L	.	T	V	.	.	.	K	K	V	D	E	-	D	D	.	Q	.	.	E	R	N	S	Q	R	V	Q	G	T	E	E	.	F	Q	V	.	.	L	.	A	L	K	.	.	.	.	.
Igig-UB*02	E	V	.	Q	.	F	.	.	A	.	.	.	.	Q	P	.	A	S	Y	.	A	N			T	R			R	V	L	.	K	S	S	.	.	R	K	V	D	K	-	D	D	.	Q	.	.	.	K	N	S	Q	Y	V	S	D	S	E	.	V	F	.	G	.	.	V	.	A	.	K	.	.	.	.	.
Amcr-UB*01	E	V	.	.	.	F	S	.	.	.	.	.	.	Q	.	.	.	S	Y	.	P	N			A	R			R	R	F	.	T	V	.	.	M	T	E	V	D	K	-	D	D	.	Q	.	.	E	R	N	S	Q	A	V	R	N	A	E	.	V	F	.	V	G	.	V	.	.	.	K	.	.	.	.	.
Amcr-UB*02	E	V	.	Q	.	F	L	.	.	.	.	.	.	Q	.	.	.	S	Y	.	P	N			T	R			R	H	F	.	T	V	.	.	M	T	E	V	D	K	-	D	D	.	Q	.	.	E	W	N	S	Q	A	A	R	N	A	E	.	.	F	.	V	S	.	V	.	.	.	K	.	.	.	.	.
Amcr-UB*03	E	V	.	.	.	F	Y	.	.	.	.	.	N	Q	.	.	.	S	Y	.	P	N			A	R			R	S	L	.	K	V	.	.	M	T	E	V	D	K	-	D	D	.	Q	.	.	E	R	N	S	D	.	.	.	N	A	E	.	A	F	.	V	G	.	V	N	V	.	K	.	.	.	.	.
Amcr-UB*0401	E	V	.	.	.	F	L	.	.	.	.	.	.	Q	.	.	.	S	Y	.	P	N			T	R			R	H	F	.	T	V	.	.	M	T	E	V	D	K	-	D	D	.	Q	.	.	E	W	N	S	Q	T	A	R	N	A	E	.	A	F	.	V	G	.	V	.	A	.	K	.	.	.	.	.
Cosu-UB*0101	E	V	.	Q	.	F	S	.	.	.	.	.	.	Q	.	.	.	S	Y	.	P	.			T	R			R	H	F	.	T	V	.	.	M	T	E	V	D	K	-	D	D	.	Q	.	.	E	R	T	S	Q	V	I	W	N	A	E	.	V	F	.	V	D	.	V	N	V	P	K	.	.	.	.	.
Cosu-UB*02	E	V	.	Q	.	F	Y	.	.	.	.	.	.	Q	.	.	.	S	Y	.	P	N			T	R			R	S	L	.	K	V	.	.	.	T	K	V	D	K	-	D	D	.	Q	.	.	E	W	N	S	D	G	.	.	N	A	E	.	V	F	.	G	.	.	V	.	A	.	K	.	.	.	.	.
Cosu-UB*03	E	V	.	Q	.	F	I	.	.	.	M	.	.	Q	R	.	.	Y	Y	.	A	N			.	G			R	V	R	.	A	V	.	.	M	T	E	V	D	K	-	D	D	.	Q	.	.	E	W	N	S	Q	A	A	Q	N	A	E	.	V	F	.	G	.	.	V	.	A	.	K	.	.	.	.	.
Tiru-UA*0301	.	L	.	Q	.	T	V	.	A	.	A	.	.	Q	L	I	G	Y	Y	.	.	N			R	R			R	I	V	S	P	V	.	.	V	Q	K	I	.	K	-	.	D	.	Q	L	.	.	M	H	.	Q	R	A	Q	N	E	E	L	Y	L	G	W	H	.	A	D	V	Q	K	L	H	.	.	.
Tiru-UA*0402	E	L	.	W	.	L	I	.	.	.	.	.	G	Q	R	.	F	H	Y	.	.	.			G	R			R	A	M	.	A	G	.	.	M	D	Q	V	C	.	-	.	D	.	H	.	.	.	.	.	.	Q	S	A	Q	G	E	E	A	.	L	.	S	Y	.	E	I	A	.	Q	R	.	.	.	.
Sppu-U*01	.	L	.	E	.	M	.	.	.	.	.	.	E	Q	.	L	D	W	.	.	.	E			R	R			Q	S	.	.	.	.	E	.	M	A	-	-	Q	N	-	.	.	.	Q	Y	.	E	R	Q	.	Q	.	L	Q	G	W	K	A	W	F	H	V	.	.	P	.	.	.	Q	R	.	.	.	.
Gaga-B-F10	.	Q	.	W	.	V	T	.	.	.	.	.	G	E	L	.	.	H	Y	N	.	T			A	R			R	Y	V	.	.	T	E	.	.	A	-	-	A	K	-	A	D	Q	Q	.	.	.	G	Q	.	Q	I	G	Q	G	N	E	.	I	.	.	E	.	.	.	I	.	Q	R	R	.	.	.	T
Haca-UA*33L	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			-	-	-	-	-	-	-	-	A	-	-	-	A	N		L	D	Q	Q	.	.	.	T	.	.	Q	R	A	.	N	T	Q	H	I	H	.	L	D	.	D	I	R	Q	R	R	.	.	.	.
Omny-UBA	.	F	.	Q	.	V	.	.	.	I	.	.	G	M	.	I	D	Y	.	.	.	V			.	E			.	N	V	L	K	Q	S	.	M	E	-	-	-	-	-	-	.	V	R	D	E	K	M	I	.	N	T	R	.	G	N	Q	.	.	F	K	.	.	V	E	I	G	M	Q	R	F	.	.	T
Maru-UB*01	.	.	.	.	.	V	S	.	.	.	.	.	.	Q	P	.	M	S	.	.	T	.			.	P	G	Q	R	E	.	.	.	G	.	.	M	D	S	M	K	H	-	.	E	.	.	F	.	E	R	Q	.	W	I	H	R	.	T	A	.	N	Y	.	V	G	.	E	N	.	.	.	.	F	.	.	S
Acsi-UBA*01	E	F	.	E	.	T	I	.	.	L	.	.	.	V	.	H	M	H	Y	N	.	-	-	V	T	K			.	V	R	.	T	Q	D	.	M	-	-	-	.	R	S	.	.	.	.	.	.	K	R	.	.	L	T	A	V	G	E	Q	.	N	F	K	V	.	I	.	I	A	M	Q	R	F	.	.	T
Agca-UA*02	.	L	.	E	.	S	.	.	.	.	.	.	.	Q	E	I	T	N	Y	N	.	.	D	G	-	-			E	L	I	.	K	V	D	.	M	-	-	-	K	K	V	.	-	.	.	.	.	G	R	N	K	Q	V	F	.	G	T	E	A	A	Y	.	H	.	V	K	I	A	M	S	R	F	.	.	T
Egst-UA*04	-	-	-	-	-	-	-	-	A	.	.	.	.	Q	L	I	G	Y	Y	.	.	S			R	R			R	I	V	S	S	V	.	.	M	Q	K	T	.	K	-	.	D	.	H	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
*#Egst-UB*06*	.	L	.	L	.	T	.	M	A	.	L	.	.	Q	F	I	G	F	Y	.	.	S			R	R			R	I	V	T	S	V	A	.	M	Q	K	T	.	K	-	.	D	.	Q	L	.	A	M	H	.	K	M	A	Q	R	E	E	L	F	L	.	R	D	.	V	.	V	Q	N	V	H	.	.	.
*#Egst-UB*07*	.	L	.	L	.	T	.	M	A	.	L	.	.	Q	F	I	G	F	Y	.	.	S			W	G			R	I	L	S	P	V	T	.	G	Q	K	T	.	K	-	.	D	.	Q	L	.	A	T	H	.	Q	L	A	Q	N	N	E	L	F	R	.	K	D	.	A	.	V	Q	N	L	H	.	.	.
*#Egst-UB*08*	.	L	.	L	.	T	.	M	A	.	L	.	.	Q	F	I	G	F	Y	.	.	S			W	G			R	I	L	S	P	V	T	.	G	Q	K	T	.	K	-	.	D	.	Q	L	.	A	M	H	.	K	M	A	Q	R	E	E	L	F	L	.	R	D	.	V	.	V	Q	N	V	H	.	.	.
*#Egst-UB*09*	.	L	.	L	.	T	.	M	A	.	L	.	.	Q	L	I	G	F	Y	.	.	S			R	R			R	I	V	T	S	V	A	.	M	Q	K	T	.	K	-	.	D	.	Q	L	.	A	T	H	.	Q	L	A	Q	N	N	E	L	F	L	.	R	D	.	A	.	V	Q	N	L	H	.	.	.
**Egst-UB*10*	E	L	.	W	.	L	I	.	.	.	.	.	G	Q	R	.	F	H	Y	.	.	A			G	R			R	A	R	.	A	.	.	.	M	D	K	V	C	.	-	.	D	.	H	.	.	.	.	.	.	Q	S	A	R	G	E	E	A	.	L	.	S	Y	.	E	I	A	.	Q	R	.	.	.	.
*Egst-UB*11*	E	L	.	W	.	L	I	.	.	.	.	.	G	Q	R	.	F	H	Y	.	.	A			G	R			R	A	M	.	A	.	.	.	M	D	K	V	C	.	-	.	D	.	H	.	.	.	.	.	.	Q	S	A	R	G	E	E	A	.	L	.	S	Y	.	E	I	A	.	Q	R	.	.	.	.
*Egst-UB*12*	E	L	.	W	.	L	I	.	.	.	.	.	G	Q	R	.	F	H	Y	.	.	A			G	R			R	A	M	.	A	V	.	.	M	D	K	V	C	.	-	.	D	.	H	.	.	.	.	.	.	Q	S	A	Q	G	E	E	A	.	L	.	S	Y	.	E	I	A	.	Q	R	.	.	.	.
*Egst-UB*13*	A	Q	.	.	.	.	K	E	.	.	L	.	G	Q	C	I	M	Y	Y	.	N	N			R	N			R	V	M	H	.	V	.	.	M	E	K	A	D	K	-	.	T	S	P	S	.	.	G	Q	P	Q	S	E	M	G	T	E	A	I	F	.	.	.	.	E	I	.	Q	K	R	.	.	.	.
**Egst-UB*14*	A	Q	.	.	.	.	M	E	.	.	L	.	G	Q	C	I	M	Y	Y	.	N	H			R	N			R	V	M	H	.	V	.	.	M	E	K	V	D	K	-	K	T	S	L	S	.	.	V	Q	.	Q	S	A	F	G	S	E	A	I	F	T	V	.	.	E	I	.	Q	K	R	H	.	H	.
Conserved all taxa			#		#							#																																																																#		
Conserved all lizards			*		*					*		*		*						*																												*																										*		*
Human peptide binding regions																a														a																		a				a			a	a			a			a	a			a			a	a			a				
Fish PSS (Acsi-UBA*-1)							b								b	b							b						b	b																					b				b	b	b		b			b								b							
Frog PSS (Agca-UA*02)							c									c												c	c											c								c	c				c		c	c			c	c						c							
Bird PSS (Haca-UA*33L)																																																							d	d		d					d						d							
Tuatara PSS (Sppu-U*01)																																																								e			e			e															
Iguanid PSS (Igig-UB*0101)							f																						f																				f			f	f	f	f				f								f							
E. stokesii PSS																															g															g										g	g			g			g													

(continued next page)

(continued from previous page)

Site number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22		23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38		39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72
*Egst-UB*15	A	Q	.	W	.	.	M	E	.	.	L	.	G	Q	C	I	M	Y	Y	.	N	H		R	N	R	V	M	H	.	V	.	.	M	E	K	A	D	K	-	K	T	S	L	S	.	.	V	Q	.	Q	S	A	F	G	S	E	A	I	F	T	V	.	.	E	I	.	Q	K	R	H	.	.	.
Egst-UB*16	.	L	.	Q	.	M	.	L	A	.	T	.	.	Q	L	M	G	V	Y	.	.	N		R	R	R	I	V	.	Y	L	.	.	V	K	K	I	.	K	-	.	D	.	Q	F	.	A	M	P	.	Q	F	.	Q	N	Q	K	L	Y	L	G	V	D	.	D	.	V	Q	E	H	.	.	.	.
Egst-UB*17	.	L	.	L	.	T	A	T	A	.	L	.	.	Q	L	I	G	F	Y	.	.	S		G	G	R	I	V	S	P	V	A	.	G	Q	K	T	.	K	-	.	D	.	.	L	.	A	M	H	.	.	M	A	Q	N	E	E	L	S	L	.	V	H	.	A	.	V	Q	N	L	H	.	.	.
Egst-UB*18	.	L	.	L	.	T	A	R	A	.	L	.	.	Q	L	I	G	F	Y	.	.	S		G	G	R	I	V	S	P	V	A	.	G	Q	K	T	.	K	-	.	D	.	.	L	.	A	M	H	.	.	M	A	Q	N	E	E	L	S	L	.	V	H	.	A	.	.	Q	N	L	H	.	.	.
*Egst-UB*19	.	L	.	L	.	T	A	T	A	.	L	.	.	Q	L	I	G	F	Y	.	.	S		R	G	R	I	V	S	P	V	A	L	G	Q	K	T	.	K	-	.	D	.	H	L	.	V	R	H	.	Q	M	A	Q	N	N	E	L	F	L	.	R	D	.	A	.	V	Q	N	L	H	.	.	.
Egst-UB*20	.	L	.	L	.	T	A	T	A	.	L	.	.	Q	L	I	G	F	Y	.	.	S		R	G	R	I	V	S	P	V	A	L	G	Q	K	T	.	K	-	.	D	.	H	L	.	V	R	H	.	Q	M	A	Q	N	N	E	L	F	L	.	R	D	.	V	.	V	Q	N	V	H	.	.	.
Egst-UB*21	.	L	.	L	.	T	A	T	A	.	L	.	.	Q	L	I	G	F	Y	.	.	S		W	G	R	I	L	S	P	V	A	.	G	Q	K	T	.	K	-	.	D	.	Q	L	.	.	M	H	M	Q	M	A	Q	R	E	E	L	V	L	.	R	H	.	A	.	V	Q	N	L	H	.	.	.
Egst-UB*22	.	L	.	L	.	T	A	M	A	.	L	.	.	Q	L	I	G	F	Y	.	.	S		R	R	R	I	V	T	S	V	A	.	M	Q	K	T	.	K	-	.	D	.	Q	L	.	.	M	H	.	K	M	A	Q	R	E	E	L	F	L	.	R	D	.	V	.	V	Q	N	V	H	.	.	.
*Egst-UB*23	.	L	.	L	.	T	A	M	A	.	L	.	.	Q	L	I	G	F	Y	.	.	S		R	G	R	I	V	S	P	V	A	L	G	Q	K	T	.	K	-	.	D	.	H	L	.	V	S	H	.	Q	M	A	Q	N	N	E	L	F	L	.	R	D	.	V	.	V	Q	N	V	H	.	.	.
Egst-UB*24	.	L	.	L	.	T	A	M	A	.	L	.	.	Q	L	I	G	F	Y	.	.	S		R	R	R	I	M	T	S	V	A	.	M	Q	K	T	.	K	-	.	D	.	Q	L	.	.	M	H	.	K	M	A	Q	R	E	E	L	F	L	.	R	D	.	V	.	V	Q	N	V	H	.	.	.
*Egst-UB*25	.	L	.	L	.	T	A	M	A	.	L	.	.	Q	L	I	G	F	Y	.	.	S		G	G	R	I	V	S	P	V	A	.	G	Q	K	T	.	K	-	.	D	.	.	L	.	A	M	H	.	.	M	A	Q	N	E	E	L	S	L	.	V	H	.	A	.	.	Q	N	L	H	.	.	.
Egst-UB*26	.	L	.	L	.	T	A	M	A	.	L	.	.	Q	L	I	G	F	Y	.	.	S		W	R	S	I	V	S	P	V	A	.	V	Q	N	T	.	K	-	.	D	.	.	L	.	A	M	H	.	E	M	A	Q	N	E	E	L	S	L	.	R	H	.	A	.	V	Q	N	L	H	.	.	.
Egst-UB*27	.	L	.	L	.	T	A	M	A	.	L	.	.	Q	L	I	G	F	Y	.	.	S		W	E	R	I	V	S	P	V	A	.	V	Q	K	A	.	K	-	.	D	.	.	L	.	A	M	H	.	.	M	A	Q	N	T	E	L	F	R	.	K	D	.	A	.	V	Q	N	L	H	.	.	.
Egst-UB*28	.	L	.	L	.	T	A	M	A	.	L	.	.	Q	L	I	G	F	Y	.	.	S		W	G	R	I	L	.	P	V	A	.	G	Q	K	T	.	K	-	.	D	.	Q	L	.	A	M	H	.	E	M	A	Q	N	E	E	L	S	L	.	R	H	V	A	.	V	Q	N	V	H	.	.	.
Egst-UB*29	.	L	.	L	.	T	A	M	A	.	L	.	.	Q	L	I	G	F	Y	.	.	S		W	G	R	I	V	S	P	V	A	.	V	Q	K	T	.	K	-	.	D	.	.	L	.	A	M	H	.	Q	R	A	.	N	N	E	L	F	L	.	G	H	.	A	.	V	Q	N	L	H	.	.	.
Egst-UB*30	.	L	.	L	.	T	A	M	A	.	L	.	.	Q	L	I	G	F	Y	.	.	S		W	G	R	I	V	S	P	V	A	.	V	Q	K	A	.	K	-	.	D	.	.	L	.	A	M	H	.	.	M	A	Q	N	T	E	L	F	R	.	K	D	.	A	.	V	Q	N	L	H	.	.	.
Egst-UB*31	.	L	.	L	.	T	A	M	A	.	L	.	.	Q	L	I	G	F	C	.	.	S		G	G	R	I	V	S	P	V	A	.	V	Q	K	T	.	R	-	K	E	Y	Q	L	.	A	M	H	.	Q	V	A	Q	K	E	E	L	S	L	.	R	D	.	A	.	V	Q	N	L	H	.	.	.
Egst-UB*32	.	L	.	L	.	T	A	M	A	.	L	.	.	Q	F	I	G	F	Y	.	.	S		W	G	R	I	L	S	P	V	T	.	G	Q	K	T	.	K	-	.	D	.	Q	L	.	A	T	H	.	Q	L	A	Q	N	N	E	L	F	L	.	R	D	.	A	.	V	Q	N	L	H	.	.	.
*Egst-UB*33	.	L	.	L	.	T	A	M	A	.	L	.	.	Q	F	I	G	F	Y	.	.	S		W	G	R	I	L	S	P	V	T	.	G	Q	K	T	.	K	-	.	D	.	H	L	.	V	R	H	.	Q	M	A	Q	N	N	E	L	F	L	.	R	D	.	V	.	V	Q	N	V	H	.	.	.
Egst-UB*34	.	L	.	L	.	T	A	M	A	.	L	.	.	Q	F	I	G	F	Y	.	.	S		W	G	R	I	V	S	A	V	A	.	M	Q	K	T	.	K	-	.	D	.	Q	L	.	.	M	H	.	K	M	A	Q	N	N	E	L	F	R	.	R	D	.	A	.	V	Q	N	L	H	.	.	.
Egst-UB*35	.	L	.	L	.	T	A	T	A	.	L	.	.	Q	L	I	G	F	Y	.	.	S		R	G	R	I	V	S	P	V	A	L	G	Q	K	T	.	K	-	.	D	.	H	L	.	V	S	H	.	Q	M	A	Q	N	N	E	L	F	L	.	R	D	.	V	.	V	Q	N	V	H	.	.	.
Egst-UB*36	.	L	.	L	.	T	A	M	A	.	L	.	.	Q	L	I	G	F	Y	.	.	S		W	G	R	I	V	S	P	V	A	.	V	Q	K	A	.	K	-	.	D	.	.	L	.	A	M	H	A	.	L	A	Q	N	N	E	L	F	R	.	K	D	.	A	.	V	Q	N	L	H	.	.	.
Egst-UB*37	.	L	.	L	.	T	A	M	A	.	L	.	.	Q	L	I	G	F	Y	.	.	S		W	G	R	I	V	S	P	V	A	.	V	Q	K	A	.	K	-	.	D	.	.	L	.	A	M	H	V	.	L	A	Q	N	N	E	L	F	R	.	K	D	.	A	.	V	Q	N	L	H	.	.	.
Egst-UB*38	.	L	.	L	.	T	A	M	A	.	L	.	.	Q	L	I	G	F	Y	.	.	S		R	R	R	I	V	T	S	V	A	.	M	Q	K	T	.	K	-	.	D	.	Q	L	.	A	M	H	.	K	M	A	Q	R	E	E	L	F	L	.	R	D	.	V	.	V	Q	N	V	H	.	.	.
Egst-UB*39	.	L	.	L	.	T	A	M	A	.	L	.	.	Q	L	I	G	F	Y	.	.	S		G	G	R	I	V	S	P	V	A	.	V	Q	K	A	.	K	-	.	D	.	.	L	.	A	M	H	.	.	M	A	Q	N	T	E	L	F	R	.	K	D	.	A	.	V	Q	N	L	H	.	.	.
Egst-UB*40	.	L	.	L	.	T	A	M	A	.	L	.	.	Q	L	I	G	F	Y	.	.	S		W	G	R	I	V	S	P	V	A	.	V	Q	K	A	.	K	-	.	D	.	.	L	.	A	M	H	.	.	M	A	Q	N	N	E	L	F	R	.	K	D	.	A	.	V	Q	N	L	H	.	.	.
*Egst-UB*41	.	L	.	L	.	T	A	M	A	.	L	.	.	Q	L	I	G	F	Y	.	.	S		G	G	R	I	V	S	P	V	A	.	V	Q	N	T	.	K	-	.	D	.	.	L	.	A	M	H	.	Q	M	A	Q	R	E	E	L	V	L	.	R	H	.	A	.	V	Q	N	L	H	.	.	.
Egst-UB*42	.	L	.	L	.	T	A	M	A	.	L	.	.	Q	L	I	G	F	Y	.	.	S		W	G	R	I	M	S	P	V	A	.	V	Q	N	T	.	K	-	.	D	.	.	L	.	A	M	H	.	Q	M	A	Q	N	N	E	L	F	L	.	R	H	.	A	.	V	Q	N	L	H	.	.	.
Egst-UB*43	.	L	.	L	.	T	A	M	A	.	L	.	.	Q	L	I	G	F	Y	.	.	S		W	G	R	I	M	S	P	V	A	.	V	Q	N	T	.	K	-	.	D	.	.	L	.	A	M	H	.	Q	M	A	Q	R	E	E	L	V	L	.	R	H	.	A	.	V	Q	N	L	H	.	.	.
Egst-UB*44	.	L	.	L	.	T	A	M	A	.	L	.	.	Q	L	I	G	F	Y	.	.	S		W	G	R	I	M	S	P	V	A	.	V	Q	K	A	.	K	-	.	D	.	.	L	.	A	M	H	.	.	M	A	Q	N	N	E	L	S	L	.	R	H	.	A	.	V	Q	N	L	H	.	.	.
Conserved all taxa			#		#							#																																																												#		
Conserved all lizards			*		*					*		*		*						*																										*																										*		*
Human peptide binding regions																a											a																		a				a			a	a			a			a	a			a			a	a			a				
Fish PSS (Acsi-UBA*-1)							b									b	b						b				b	b																				b				b	b	b		b			b								b							
Frog PSS (Agca-UA*02)							c										c										c	c												c								c	c				c		c	c			c	c							c							
Bird PSS (Haca-UA*33L)																																																				d	d		d					d							d							
Tuatara PSS (Sppu-U*01)																																																					e			e			e															
Iguanid PSS (Igig-UB*0101)							f																				f																					f			f	f	f	f					f								f							
E. stokesii PSS																												g																g											g	g			g			g												

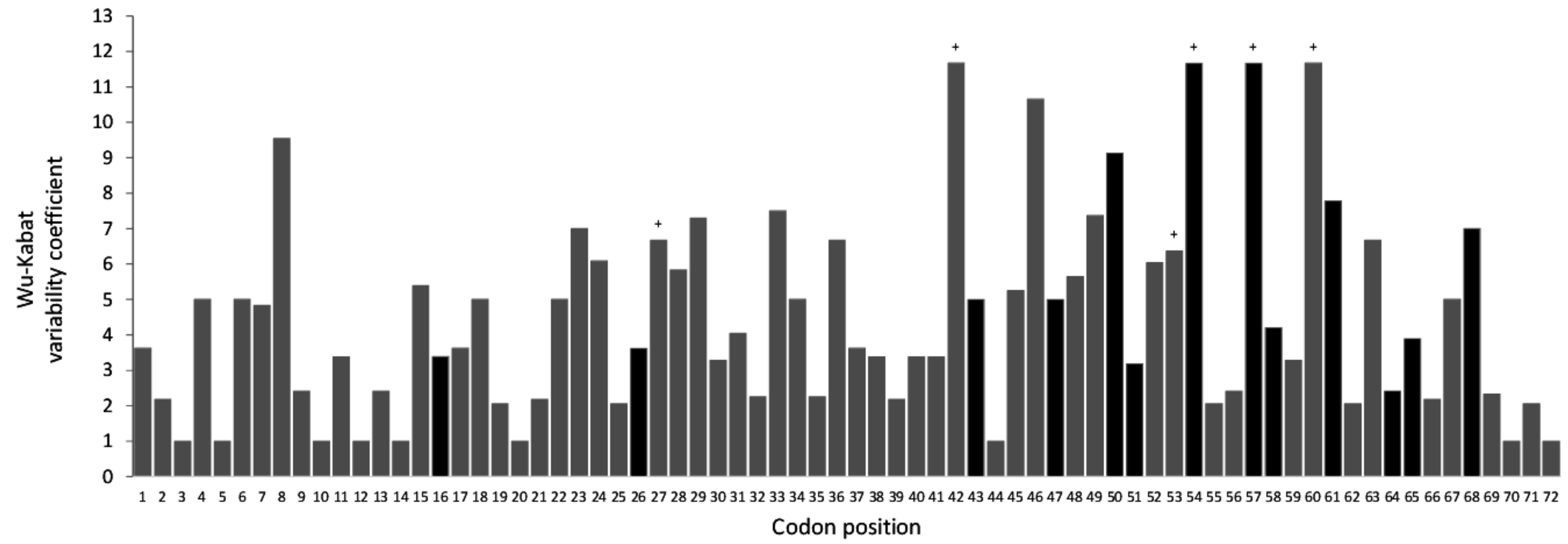
Fig. 1 Amino acid sequences of 39 *Egernia stokesii* MHC I exon 2 (α-1) alleles, aligned with the corresponding region of the human leucocyte antigen (HLA). Sites representing the peptide binding regions of the HLA are shaded in grey. Sites conserved across all lizard species included in the alignment are indicated with '*' and sites conserved across all taxa are indicated with '#'. Alleles were derived for two datasets: Camel Hill sampled 1994-1998, and Hawker (Camel Hill, Castle Rock, and Castle Rock Ridge) sampled 2012-13. Amino acids under positive selection in *E. stokesii* are indicated with '+' where sites 27 and 42 relate to the Hawker 2012-13 dataset, site 53 relates to the Camel Hill 1994-98 datasets, and sites 54, 57, and 60 were common to both datasets. *Egernia stokesii* alleles prefixed with # were found only in Camel Hill 1994-98, alleles prefixed with were found only in Hawker 2012-13, all other alleles were common to both datasets. Other sequences were trimmed to match *E. stokesii* open reading frame. Site number corresponds to the 216 bp *E. stokesii* MHC I region sequenced in this study. Below the alignment, human peptide binding regions (a) are compared to positively identified sites (PSS) identified in non-mammalian taxa including *E. stokesii* (b – g). Species are: Igig (*Iguana iguana*, green iguana), Amcr (*Amblyrhynchus cristatus*, Galapagos marine iguana), Cosu (*Conolophus subcristatus*, Galapagos land iguana), Tiru (*Tiliqua rugosa*, sleepy lizard), Hosa (*Homo sapien*), Sppu (*Sphenodon punctatus*, tuatara), Gaga (*Gallus gallus*, chicken), Haca (*Halobaena caerulea*, blue petrel), Omny (*Oncorhynchus mykiss*, rainbow trout), Maru (*Macropus rufogriseus*, red-necked wallaby), Acsi (*Acipenser sinensis*, Chinese sturgeon), Agca (*Agalychnis callidryas*, red eyed tree frog), Egst (*Egernia stokesii*, gidgee skink). Species included in the alignment and for which positively selected sites have been determined are identified in bold text in the alignment.

Site number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
Hosa-DRB1	D	V	G	E	Y	R	A	V	E	E	L	G	R	P	D	A	E	Y	W	N	S	Q	K	D	L	L	E	Q	K	R	G	Q	V	D
Sppu-DAB1*01	.	R	.	L	F	E	P	.	T	.	.	.	.	.	.	.	.	S	.	.	R	.	T	E	I	.	Q	D	R	.	T	A	.	E
Taac-DZB*01	.	.	.	V	F	V	S	.	T	.	.	D	S	K	.	.	Q	.	.	.	.	.	.	.	I	M	D	.	.	.	A	E	.	.
Gaga-BLB1	.	.	.	K	.	V	.	D	T	P	.	.	E	P	Q	.	.	.	.	.	.	N	A	E	F	.	.	N	R	M	N	E	.	.
Brre-DAB1	T	.	.	K	.	V	G	Y	T	.	Q	.	V	I	F	.	R	N	F	.	K	N	Q	A	Y	.	Q	.	R	K	A	E	.	E
Amcr-DAB1*01	A	R	.	.	F	E	.	.	A	A	.	.	E	.	.	.	R	.	.	.	G	.	.	E	V	.	.	.	N	.	A	A	A	.
Amcr-DAB2*01	A	R	.	.	F	E	.	.	A	A	.	.	E	.	.	.	R	H	.	.	G	.	.	E	V	.	.	N	A	.	A	A	A	.
Amcr-DAB2*02	A	R	.	.	.	E	.	.	A	A	.	.	E	.	.	.	R	.	.	.	G	.	.	E	V	.	D	R	R	.	A	E	.	.
Amcr-DAB3*01	A	R	.	.	F	E	.	.	A	A	.	.	E	.	.	.	R	.	.	.	G	.	.	E	V	.	.	N	A	.	A	A	A	.
Amcr-DAB3*02	A	R	.	.	.	E	.	.	A	A	.	.	E	.	.	.	R	.	.	.	G	.	.	E	V	.	.	.	R	.	A	A	A	.
Amcr-DAB4*01	A	R	.	.	.	E	.	.	A	A	.	.	E	.	.	.	R	.	.	.	G	.	.	E	V	.	D	R	R	.	A	E	.	.
Amcr-DAB4*02	A	R	.	.	.	E	.	.	A	A	.	.	E	.	.	.	R	.	.	.	G	.	.	E	V	.	D	R	R	.	A	E	.	.
*Egst-DRB*01*	R	R	.	S	F	E	.	.	T	.	.	.	E	L	E	.	R	V	.	.	.	.	T	E	F	.	.	A	M	.	T	D	.	N
*Egst-DRB*02*	R	R	.	S	F	E	.	I	T	.	.	.	E	.	E	.	R	V	.	.	.	.	T	E	F	.	.	A	M	.	T	D	.	.
*Egst-DRB*03*	R	R	.	T	.	E	.	.	T	A	.	.	E	.	S	.	R	.	.	.	.	L	T	.	Y	M	.	R	.	.	T	E	.	.
*Egst-DRB*04*	R	R	.	S	F	E	.	.	T	.	.	.	K	.	E	.	R	V	.	.	.	.	T	E	F	.	.	A	M	.	T	D	.	.
*Egst-DRB*05*	R	R	.	S	F	E	.	I	T	.	.	.	E	.	E	.	R	I	.	.	.	.	T	E	F	.	.	A	M	.	T	D	.	.
Conserved all taxa			#													#				#														
Conserved all lizards		*	*			*	*				*	*				*	*		*	*										*				

Fig. 2 Amino acid sequences of five *Egernia stokesii* MHC II exon 2 (β-1) alleles, aligned with the corresponding region of the human leucocyte antigen (HLA). Sites representing the peptide binding regions of the HLA are shaded in grey. Sites conserved across all lizard species included in the alignment are indicated with '*' and sites conserved across all taxa are indicated with '#'. No evidence was found for positive selection at any of the *E. stokesii* sites. Other sequences were trimmed to match *E. stokesii* open reading frame. Site number corresponds to the 102 bp *E. stokesii* MHC I region sequenced in this study. Species are: Amcr (*Amblyrhynchus cristatus*, Galapagos marine iguana), Hosa (*Homo sapien*),

Sppu (*Sphenodon punctatus*, tuatara), Taac (*Tachyglossus aculeatus*, short beaked echidna), Gaga (*Gallus gallus*, chicken), Brre (*Brachydanio rerio*, Zebrafish).

3a)



3b)

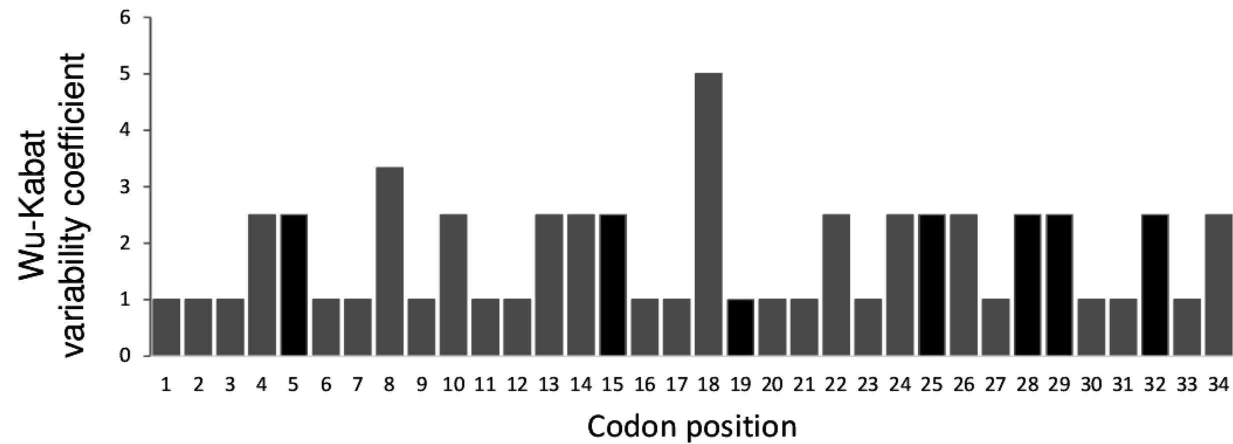


Fig. 3 Plots showing amino acid variability based on the Wu-Kabat variability coefficient for **a)** 39 *E. stokesii* MHC I alleles and **b)** five *E. stokesii* MHC II alleles. Higher coefficient values correlate with higher amino acid variability. The '+' indicates a positively selected site (*E. stokesii* MHC I only) and black bars indicate peptide binding sites of the human leucocyte antigen.

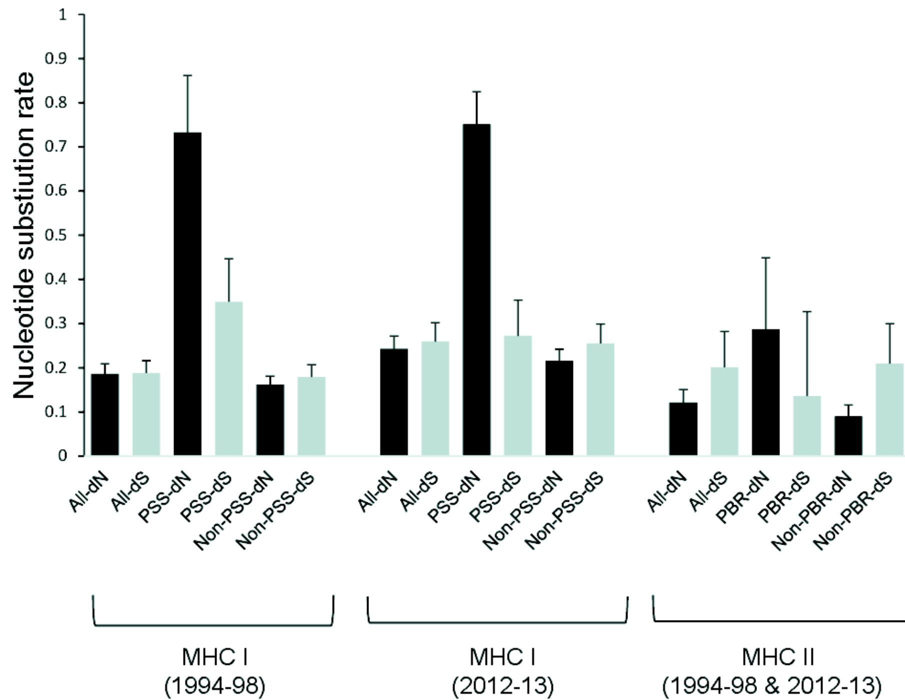


Fig. 4 Rates of non-synonymous (dN, black bars) and synonymous (dS, grey bars) substitutions in *Egernia stokesii* MHC I and II. There were some differences in MHC I alleles in the 1994-98 and 2012-13 datasets therefore the two datasets were analysed separately whereas MHC II alleles were identical in the 1994-98 and 2012-13 datasets therefore analysed only once. Error bars indicate standard errors. All = entire sequence (216 bp MHC I, 102 bp MHC II). PSS = rates for positively selected sites only (shown for MHC I only as none identified in MHC II). PBR = rates for putative peptide binding regions (based on HLA) only (shown for MHC II only as PSS identified in MHC I likely to be the PBRs for *E. stokesii*)

References

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

Alcaide M, Liu M, Edwards SV (2013) Major histocompatibility complex class I evolution in songbirds: universal primers, rapid evolution and base compositional shifts in exon 3. PeerJ 1:e86.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403-410. doi:10.1016/S0022-2836(05)80360-2

Ansari TH (2016) The distribution of genetic variation in sleepy lizards. Flinders University of South Australia

Ansari TH, Bertozzi T, Miller RD, Gardner MG (2015) MHC in a monogamous lizard - characterization of class I MHC genes in the Australian skink *Tiliqua rugosa*. Dev Comp Immunol 53:320-327.

Apanius V, Penn D, Slev PR, Ruff LR, Potts WK (1997) The nature of selection on the major histocompatibility complex. Crit Rev Immunol 17:179-224.

Babik W (2010) Methods for MHC genotyping in non-model vertebrates. Mol Ecol Resour 10:237-251.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B (Methodological) 57:289-300.

Bernatchez L, Landry C (2003) MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? J Evol Biol 16:363-377.

Bondinas G, Moustakas A, Papadopoulos G (2007) The spectrum of HLA-DQ and HLA-DR alleles, 2006: a listing correlating sequence and structure with function. Immunogenetics 59:539-553. doi:10.1007/s00251-007-0224-8

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 55:855-865. doi:10.1007/s00251-004-0648-3

Bull CM, Griffith SC, Lanham EJ, Johnston GR (2000) Recognition of pheromones from group members in a gregarious lizard, *Egernia stokesii*. J Herpetol 34:92-99.

Chelvanayagam G (1996) A roadmap for HLA-A, HLA-B, and HLA-C peptide binding specificities. Immunogenetics 45:15-26.

Cogger HG (1983) Reptiles and amphibians of Australia. AH & AW Reed Pty Ltd,

Consuegra S, de Eyto E, McGinnity P, Stet RJM, Jordan WC (2011) Contrasting responses to selection in class I and class IIα major histocompatibility-linked markers in salmon. Heredity 107:143-154.

Consuegra S, Megens HJ, Leon K, Stet RJM, Jordan WC (2005) Patterns of variability at the major histocompatability class II alpha locus in Atlantic salmon constrast with those at the class I locus. Immunogenetics 57:16-24.

Díez-Rivero CM, Reche P (2009) Discovery of conserved epitopes through sequence variability analyses. In: Bioinformatics for immunomics. Springer, pp 95-101

Duffield GA, Bull CM (1996) Host location by larvae of the reptile tick *Amblyomma vikirri* (Acari: Ixodidae). Experimental & Applied Acarology 20:575-582.

Duffield GA, Bull CM (2002) Stable aggregations in an Australian lizard, *Egernia stokesii*. Naturwissenschaften 89:424-427.

Edwards SV, Hedrick PW (1998) Evolution and ecology of MHC molecules: from genomics to sexual selection. Trends Ecol Evol 13:305-311. doi:10.1016/s0169-5347(98)01416-5

Eimes JA, Townsend AK, Sepil I, Nishiumi I, Satta Y (2015) Patterns of evolution of MHC class II genes of crows (Corvus) suggest trans-species polymorphism. PeerJ 3:e853.

Elbers JP, Taylor SS (2016) Major histocompatibility complex polymorphism in reptile conservation. Herpetological Conservation and Biology 11:1-12.

Fenner AL, Bull CM (2008) The impact of nematode parasites on the behaviour of an Australian lizard, the gidgee skink *Egernia stokesii*. Ecol Res 23:897-903.

Galan M, Guivier E, Caraux G, Charbonnel N, Cosson J-F (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 CM, Reinherz EL, Reche PA (2008) PVS: a web server for protein sequence variability analysis tuned to facilitate conserved epitope discovery. Nucleic Acids Res 36:W35-W41.

Gardner MG, Bull CM, A. F, Murray K, Donnellan SC (2007) Consistent social structure within aggregations of the Australian lizard, *Egernia stokesii* across seven disconnected rocky outcrops. Journal of Ethology 25:263-270.

Gardner MG, Bull CM, Cooper JB, Duffied GA (2001) Genetic evidence for a family structure in stable social aggregations of the Australian lizard *Egernia stokesii*. Mol Ecol 10:175-183.

Gardner MG, Bull CM, Cooper SJB (2002) High levels of genetic monogamy in the group-living Australian lizard *Egernia stokesii*. Mol Ecol 11:1787-1794.

Garrigan D, Hedrick PW (2003) Perspective: detecting adaptive molecular polymorphism: lessons from the MHC. Evolution 57:1707-1722. doi:10.1111/j.0014-3820.2003.tb00580.x

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, Moreno MA, Caccone A (2009) Characterization and evolution of MHC class II B genes in Galapagos marine iguanas (*Amblyrhynchus cristatus*). Dev Comp Immunol 33:939-947.

Godfrey SS, Bull CM, Gardner MG (2006) Associations between blood parasite infection and a mircosatellite DNA allele in an Australian scincid lizard (*Egernia stokesii*). Parasitol Res 100:109-109.

Godfrey SS, Bull CM, James R, Murray K (2009) Network structure and parasite transmission in a group living lizard, the gidgee skink, *Egernia stokesii*. Behav Ecol Sociobiol 63:1045-1056.

Hallas G, Bull CM, Bursey CR (2005) *Pharyngodon tiliquae* and *Thelandros trachysauri* (Nematoda: Pharyngodonidae), new parasite records for *Egernia stokesii* (Scincidae) from Australia. Comp Parasitol 72:119-120.

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. J Evol Biol 27:2347-2359.

Huchard E, Albrecht C, Schliehe-Diecks S, Baniel A, Roos C, Peter PMK, Brameier M (2012) Large-scale MHC class II genotyping of a wild lemur population by next generation sequencing. Immunogenetics 64:895-913.

Hughes AL (2007) Looking for Darwin in all the wrong places: the misguided quest for positive selection at the nucleotide sequence level. Heredity 99:364-373.

Hughes AL, Nei M (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature 335:167-170.

Hughes AL, Yeager M (1998) Natural selection at major histocompatibilty complex loci of vertebrates. Annu Rev Genet 321:415-435.

Hung A (2013) MHC and Mate Choice in Anolis sagrei. Dissertation, The University of New Mexico

Hurvich CM, Tsai C-L (1989) Regression and time series model selection in small samples. Biometrika 76:297-307. doi:10.2307/2336663

Jaratlerdsiri W, Isberg SR, Higgins DP, Ho SY, Salomonsen J, Skjodt K, Miles LG, Gongora J (2014) Evolution of MHC class I in the Order Crocodylia. Immunogenetics 66:53-65.

Keirans JE, Bull CM, Duffied GA (1996) *Amblyomma vikkiri* n. sp. (Acari: Ixodida: Ixodidae), a parasite of the gidgee skink *Egernia stokesii* (Reptilia: Scincidae) from South Australia. Syst Parasitol 34:1-9.

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

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 109:146-155.

Klein J, Satta Y, O'hUigin C (1993) The molecular descent of the major histocompatibility complex. Annu Rev Immunol 11:269-295.

Knapp LA (2005) The abcs of mhc. Evolutionary Anthropology: Issues, News, and Reviews 14:28-37.

Kryazhimskiy S, Plotkin JB (2008) The population genetics of dN/dS. PLoS Genet 4:e1000304.

Kuduk K, Johanet A, Allainé D, Cohas A, Radwan J (2012) Contrasting patterns of selection acting on MHC class I and class II DRB genes in the Alpine marmot

(*Marmota marmota*). J Evol Biol 25:1686-1693. doi:10.1111/j.1420-9101.2012.02537.x

Madsen T, Ujvari B (2006) MHC class I variation associates with parasite resistance and longevity in tropical pythons. J Evol Biol 19:1973-1978.

Main AR, Bull CM (1996) Mother-offspring recognition in two Australian lizards, *Tiliqua rugosa* and *Egernia stokesii*. Anim Behav 52:193-200.

Meyer M, Kircher M (2010) Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb Protoc 2010:t5448.

Milinski M (2006) The major histocompatibility complex, sexual selection, and mate choice. Annu Rev Ecol, Evol Syst 37:159-186.

Miller HC, Andrews-Cookson M, Daugherty CH (2007) Two patterns of variation among MHC class I loci in tuatara (*Sphenodon punctatus*). J Hered 98:666-677.

Miller HC, Belov K, Daugherty CH (2006) MHC class I genes in the tuatara (Sphenodon spp.): evolution of the MHC in an ancient reptilian order. Mol Biol Evol 23:949-956.

Miller HC, O'Meally DO, Ezaz T, Amemiya C, Marshall-Graves JA, Edwards S (2015) Major Histocompatibility Complex genes map to two chromosomes in an evolutionarily ancient reptile, the tuatara *Sphenodon punctatus*. G3: Genes, Genomes, Genetics doi:doi:10.1534/g3.115.017467

Neefjes J, Jongsma ML, 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, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418-426.

Pechouskova E, Dammhahn M, Brameier M, Fichtel C, Kappeler PM, Huchard E (2015) MHC class II variation in a rare and ecological specialist mouse lemur reveals lower allelic richness and contrasting selection patterns compared to a generalist and widespread sympatric congener. Immunogenetics 67:229-245.

Piertney SB, Oliver MK (2006) The evolutionary ecology of the major histocompatibility complex. Heredity 96:7-21.

Pokorny I, Sharma R, Goyal S, Mishra S, Tiedemann R (2010) MHC class I and MHC class II DRB gene variability in wild and captive Bengal tigers (*Panthera tigris tigris*). Immunogenetics 62:667-679. doi:10.1007/s00251-010-0475-7

Posada D, Buckley T (2004) Model Selection and Model Averaging in Phylogenetics: Advantages of Akaike Information Criterion and Bayesian Approaches Over Likelihood Ratio Tests. Syst Biol 53:793-808. doi:10.1080/10635150490522304

Pyron RA, Burbrink FT, Wiens JJ (2013) A phylogeny and revised classification of Squamata, including 4161 species of lizards and snakes. BMC Evol Biol 13:1.

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. Mol Ecol 23:5966-5978. doi:10.1111/mec.12984

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. Mol Ecol 21:2469-2479.

Reche PA, Reinherz EL (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. doi:10.1016/s0022-2836(03)00750-2

Sommer S (2005) The importance of immune gene variability (MHC) in evolutionary ecology and conservation. Front Zool 2:16. doi:doi:10.1186/1742-9994-2-16

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

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.

Stein J (1999) An ecological study of the blood parasites of *E. stokesii*. Dissertation, Flinders Universtiy of South Australia

Stein J, Dyce AL (2002) Field observations on the phlebotomine sand fly *Australophlebotomus mackerrasi* Lewis and Dyce feeding on the Gidgee skink *Egernia stokesii* Gray. Parasitol Res 88:278-279. doi:10.1007/s00436-001-0541-z

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 63:653-666.

Strandh M, Westerdahl H, Pontarp M, Canback B, Dubois M-P, Miquel C, Taberlet P, Bonadonna F (2012) Major histocompatibility complex class II compatibility, but not class I, predicts mate choice in a bird with highly developed olfaction. Proc R Soc Lond, Ser B: Biol Sci 279:4457-4463.

Stuglik MT, Radwan J, Babik W (2011) jMHC: software assistant for multilocus genotyping of gene families using next-generation amplicon sequencing. Mol Ecol Resour 11:739-742. doi:10.1111/j.1755-0998.2011.02997.x

Sun Y, Xi D, Li G, Hao T, Chen Y, Yang Y (2014) Genetic characterization of MHC class II DQB exon 2 variants in gayal (*Bos frontalis*). Biotechnology & Biotechnological Equipment 28:827-833.

Sutton J, Robertson B, Grueber C, Stanton J-A, Jamieson I (2013) Characterization of MHC class II B polymorphism in bottlenecked New Zealand saddlebacks reveals low levels of genetic diversity. Immunogenetics 65:619-633. doi:10.1007/s00251-013-0708-7

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol Biol Evol 30:2725-2729. doi:10.1093/molbev/mst197

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

The Reptile Database (2015) http://www.reptile-database.org. Accessed June 3 2016

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.

Wedekind C, Penn D (2000) MHC genes, body odours, and odour preferences. Nephrology Dialysis Transplantation 15:1269-1271.

Wegner KM (2008) Historical and contemporary selection of teleost MHC genes: did we leave the past behind? J Fish Biol 73:2110-2132.

Winternitz JC, Minchey SG, Garamszegi LZ, Huang S, Stephens PR, Altizer S (2013) Sexual selection explains more functional variation in the mammalian major histocompatibility complex than parasitism. Proc R Soc Lond, Ser B: Biol Sci 280:20131605. doi:10.1098/rspb.2013.1605

Wu TT, Kabat EA (1970) An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. The Journal of Experimental Medicine 132:211-250.

Xu B, Yang Z (2013) PAMLX: A graphical user interface for PAML. Mol Biol Evol 30:2723-2724. doi:10.1093/molbev/mst179

Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586-1591. doi:10.1093/molbev/msm088

Yang Z, Wong WSW, Nielsen R (2005) Bayes Empirical Bayes Inference of Amino Acid Sites Under Positive Selection. Mol Biol Evol 22:1107-1118. doi:10.1093/molbev/msi097

Chapter 5: Larger lizards live longer in a group living lizard

Accepted for publication as:

Pearson, S.K., Godfrey, S.S., Bull, C.M., and Gardner, M.G., Larger lizards live longer in the group living *Egernia stokesii*, *Australian Journal of Zoology*, accepted July 2016.

Chapter 4 provides evidence of selection on *E. stokesii* MHC. The remainder of this thesis documents some preliminary analyses of the mechanisms generating and maintaining *E stokesii* MHC variation. Space use is an important aspect of animal behaviour with implications for MHC variation. Although short-term space use in group-living lizards has received some attention, investigations of long-term space use and factors influencing space use have been limited. Previous work provided evidence of stability of space and social bonds in one *E. stokesii* population over a period of six years. While conducting field surveys during the summer of 2012-13, a number of lizards were captured that had previously been captured in 2003-04. The following work investigated if this behaviour was consistent across three isolated populations over nearly a decade and considered for the first time a number of individual characteristics that may be associated with lizard spatial and social stability.

Larger lizards live longer in the group living *Egernia stokesii*

Pearson, S.K.[A,B], Godfrey, S.S. [C], Bull, C.M.[A], Gardner, M.G.[A,D]

[A]School of Biological Sciences, Flinders University of South Australia, Bedford Park 5042, Australia

[C]School of Veterinary and Life Sciences, Murdoch University, Murdoch, Australia.

[D]Evolutionary Biology Unit, South Australian Museum, North Terrace, Adelaide 5000, Australia.

[B]Corresponding author. Email: sarah.pearson@flinders.edu.au

Running header

Larger lizards live longer in group living lizard

Short summary

Space use is an important aspect of animal behaviour. This investigation of gidgee skink site fidelity found lizards that were larger in initial surveys were more likely to be recaptured, and a high proportion of those recaptured were close to their original capture sites. Why some lizards changed space while others didn't is yet to be fully understood.

Australian Journal of Zoology, accepted July 2016

Abstract

Animal space use has implications for gene flow, disease dynamics, mating systems and the evolution of sociality. Given recent attention to sociality in reptiles, lizards are an important group to expand our understanding of animal space use. Lizard space use is commonly investigated within one population over a short period and limited attention has been given to potential predictors of site fidelity. This study evaluated site fidelity in three populations of group living *Egernia stokesii* (gidgee skink) between two field surveys separated by almost a decade. Of forty-three recaptured lizards, twenty-eight (65%) occupied their original space, and fifteen (36%) of those shared their space with the same other lizard or lizards in both surveys. This confirmed long-term site and social bond fidelity in *E. stokesii*. We found that larger lizards were more likely to be recaptured. Neither body size, individual genetic heterozygosity, nor the availability of refuges strongly predicted whether lizards were recaptured in the same or a different place. The reasons why some lizards stayed in the same space while others moved are yet to be resolved.

Key words

Egernia stokesii, site fidelity, space use, lizard, group living

Introduction

Space use is an important aspect of animal behaviour with implications for gene flow (Stow and Sunnucks 2004; Wolf and Trillmich 2008), disease transmission (Wey *et al.* 2008; Boyer *et al.* 2010; Paull *et al.* 2011), mating systems (Greenwood 1980), and the evolution and maintenance of group living (Emlen 1982; Kerth 2008). Home range patterns are commonly used to describe space use (Börger *et al.* 2008), which is influenced by multiple factors including the availability of food (Godfrey 2013; Fernandez-Duque 2015), mates (Emlen and Oring 1977; Cézilly *et al.* 2000; Kitchen *et al.* 2005), and sleeping and refuge sites (Kerr and Bull 2006; Kerth 2008; Zhou *et al.* 2011). Space use will differ depending on whether resources are limited yet permanent, or widely available yet ephemeral (Switzer 1993; Spiegel *et al.* 2015). Limited, permanent resources may promote sedentary behaviour with high site fidelity (Gardner *et al.* 2007; Wartmann *et al.* 2014).

Site fidelity refers to low variability of the area an animal occupies over time (Ramos-Fernandez *et al.* 2013). Where the benefits of site fidelity outweigh the costs of dispersal over longer time frames, group living may evolve (Emlen 1982; Emlen 1995). There are two models commonly used to investigate the evolution of group living. The ecological constraints model predicts dispersal may be delayed where breeding vacancies are limited by resources (Emlen 1982; Emlen 1995). In turn, site fidelity may arise where dispersal is constrained by resource availability. The benefits of philopatry model predicts benefits are derived from remaining in the natal space with relatives (Emlen 1995). Thus, indirect benefits of improved kin fitness associated with philopatry may promote site fidelity (Meise *et al.* 2013). Other benefits, including advantages of enhanced survival derived from living with relatives, and opportunities to inherit the natal space (Stacey and Ligon 1991; Emlen 1995), may further support site fidelity associated with philopatry.

Site fidelity by group living animals has consequences for social group structure (Moyer *et al.* 2006), the nature of interactions between individuals (Godfrey 2013; Godfrey *et al.* 2014), and the evolution of behaviour such as cooperation and kin selection (Shorey *et al.* 2000). Space use can also influence mating systems. For example, site fidelity may promote pair stability and long-term pair bonds (Shields 1984; Cézilly *et al.* 2000; Gardner *et al.* 2002) whereas wider ranging behaviour may be more associated with promiscuity. Given recent attention to group living in squamate reptiles (Doody *et al.* 2013; Gardner *et al.* 2015; While *et al.* 2015), lizards are an important group to expand our understanding of the dynamics of space use by social animals. Although animals exhibit site fidelity over varying temporal scales (Meise *et al.* 2013; Ramos-Fernandez *et al.* 2013; Wartmann *et al.* 2014; Ebrahimi *et al.* 2015), space use in group living lizards has usually only been assessed over short periods such as within a season or over a few seasons or years.

Examples of evidence for short-medium term site fidelity include armadillo lizards (*Ouroborus cataphractus*; three consecutive months; Effenberger and Mouton 2007), desert night lizards (*Xantusia vigilis*; five consecutive years; Davis *et al.* 2011), White's skinks (*Egernia whitii*; two - three consecutive years; Chapple and Keogh 2006; While *et al.* 2009), gidgee skinks (*Egernia stokesii*; six consecutive years; Duffield and Bull 2002b), sleepy lizards (*Tiliqua rugosa*; two - five consecutive years; Bull and Freake 1999; Kerr and Bull 2006; Spiegel *et al.* 2015), and black rock skinks (*E. saxatilis*; Oct-Mar over four consecutive years; O'Connor and Shine 2003). Longer term site fidelity has been investigated in only a few group living lizard species such as the ten year study of the great desert skink (*Liopholis kintorei*; McAlpin *et al.* 2011) and little is known regarding the consistency of space use behaviour across different populations of the same species. In addition, few studies have investigated factors that may influence lizard space use (but see Clobert *et al.* 1994; Kerr and Bull 2006).

While conducting field surveys of three *E. stokesii* populations during the austral summer of 2012-13 we recaptured 48 individuals originally sampled in a previous study during 2003-04 (Godfrey *et al.* 2006; Gardner *et al.* 2007). This presented an opportunity to extend the temporal scale of earlier studies of site fidelity in this species. These recaptures also permitted an investigation of variable characteristics of individual *E. stokesii* that may predict their long-term persistence in one place. *Egernia stokesii* is a large, long-lived (25 years; Duffield and Bull 1996) lizard found across semi-arid regions of Australia (Cogger 1983). In South Australia, populations occupy isolated rocky outcrops and individuals spend most of their time inside rock crevice refuges, or basking on the rocks beside their crevice entrances retreating into the crevice when disturbed. Within populations, many individuals form stable family groups that can include adult lizards and several consecutive annual cohorts of their offspring that share multiple crevice refuges within a group home range (Gardner *et al.* 2001; Duffield and Bull 2002b). Some other individuals move among multiple groups or are solitary (Gardner *et al.* 2001; Duffield and Bull 2002b; Godfrey *et al.* 2006). *Egernia stokesii* exhibit delayed maturity, with sub-adults remaining in their parental group for at least five years (Duffield and Bull 2002b), and breeding adults have high levels of social and genetic monogamy (Gardner *et al.* 2002). Juvenile mortality is high (40%) but mortality is generally low after two years of age (Duffield and Bull 2002b). Group stability is exhibited by individuals from all age classes (Duffield and Bull 2002b). For example, where juveniles were sampled over more than one season during a five year study, only one of 31 juveniles changed to a different social group (Duffield and Bull 2002b). When lizards disperse, movement to another outcrop is very rare, instead individuals usually move to an adjacent social group within the same population, crossing only one or two groups from their natal crevice (Gardner *et al.* 2001, Pearson unpublished data). Neither inbreeding, nor sex-biased dispersal across multiple populations has been detected by genetic analyses (Gardner *et al.*

2012). Limited crevice availability probably determines the spatial structure of *E. stokesii* social groups (Gardner *et al.* 2007), contributing to fragmented social networks (Godfrey *et al.* 2009) and to fine scale genetic structure (Pearson, unpublished data).

In a previous study, Duffield and Bull (2002b) reported *E. stokesii* site fidelity in one population over six consecutive years. In the current study, we asked whether this behaviour was consistent across three isolated populations over a longer period of nearly a decade. Populations are limited by crevice availability and exhibit fine scale social and genetic structure, therefore we predicted this species would display consistent long-term site and social bond fidelity. We also considered for the first time a number of individual characteristics that may be associated with whether an individual lizard remained in the population and its social group. We tested these predictors, which included body size, individual genetic diversity, population crevice density, and social status, as a set of alternative hypotheses relating to site fidelity (Table 1).

Material and methods

Study sites

The study was conducted over two spring-summer periods separated by nine years, at three geographically separate *E. stokesii* populations within two kilometres of Hawker (31°54′S; 138°25′E) in the southern Flinders Ranges of South Australia. The populations were at Camel Hill (CAM), Castle Rock (CAS), and Castle Rock Ridge (CRR) (Figure 1, Appendix 1; Supplementary Material available on the Journal website). Each population was located on a rocky outcrop with refuge crevices, and was separated from each other population by 300 m to 1.5 km of non-rocky habitat (Gardner *et al.* 2007). The matrix between populations was sparsely vegetated with small shrubs and low grasses and lacked appropriate refuges for *E. stokesii*. Other isolated populations previously surveyed around Hawker (Gardner *et al.*

2007), were 500 m to 35 km beyond those sampled in both periods. Dispersal of lizards to other populations is rare and it is extremely unlikely individuals moved beyond the populations sampled in both periods. There has been one previous record of a human-assisted dispersal event (Duffield and Bull 2002a). During the spring and summer of the first sampling period, August 2003 to March 2004, the average maximum daily temperature was 30.5^{O}C (SE ± 2.04) and the average monthly rainfall was 16.89 mm (SE ± 7.38). In the second sampling period, September 2012 to March 2013, the average maximum daily temperature was 31.09 ^{O}C (SE ± 1.57) and the average monthly rainfall was 16.94 mm (SE ± 4.23).

Field surveys, genetic and social analyses 2003-04

The first sampling period occurred in April 2003 and between August 2003 and March 2004. *Egernia stokesii* were visibly located within crevices or detected by the presence of fresh scats outside crevices. We assigned unique labels to occupied crevices and recorded their locations using a GPS (UTM). GPS accuracy data was not recorded therefore we estimated it to be 4 m. We captured lizards by hand or in unbaited Elliot traps placed outside marked crevices. Most lizards were marked with a microchip, inserted sub-dermally. We used toe clipping to mark juvenile lizards with a snout-vent length (SVL) less than 150 mm because they were too small to be micro-chipped. For within season identification, we also marked lizards with a unique paint pattern with acrylic non-toxic paint. On first capture, lizards were measured (SVL, mm), weighed (gm), and sexed by inverting the hemipenes of males. We inferred lizard age class (adult, sub-adult or juvenile) from SVL (Gardner *et al.* 2007). A sample of up to 0.5 mL of blood from the caudal vein of each lizard and stored on Whatman FTA ® Classic paper. DNA was extracted following (Smith and Burgoyne 2004) and twelve polymorphic microsatellite loci (Est1, Est3, Est4, Est8, Tr3.2, Tr4.11, Tr5.20, Tr5.21, Ecu1, Ecu2, Ecu4, Ecu5) were amplified following Cooper *et al.* (1997); Gardner *et al.* (1999), and

Stow (2002). We retained all loci for use in downstream analyses following tests for linkage disequilibrium, conformance to Hardy-Weinberg equilibrium, and the incidence of null alleles and large allele dropout (Gardner 2000, Pearson unpublished data). Summaries of the genetic structure and social organisation of the three populations during the first sampling period are reported in Godfrey *et al.* (2006) and Gardner *et al.* (2007).

Within this period, we identified previously captured lizards within crevices without capture, either with a modified microchip reader inserted into the crevice, or by their paint pattern. Any new locations were recorded to determine individual and group home ranges, and social group membership of lizards sampled during 2003-04 (Godfrey *et al.* 2006). We derived individual home ranges from recapture locations, and defined social groups based on home range overlap among group members (Godfrey *et al.* 2006). Where lizards were not assigned to a social group, they were designated as a floater (a lizard occupying marginal, less frequently used crevices, within the home ranges of two or more social group) (Duffield and Bull 2002b), or as an isolate (a solitary lizard occupying crevices outside the home ranges of any social groups) (Godfrey *et al.* 2006).

Field surveys 2012-13

The second sampling took place at the same three sites (CAM, CAS, and CRR) in the spring and summer of September 2012 to March 2013 using identical survey methods. Average GPS accuracy was 4 m.

Site fidelity

We assessed site fidelity by comparing capture locations in 2003-04 and 2012-13. We used ArcGIS (ESRI ® ArcMap™ 10.0) to overlay capture locations for each individual captured in both survey periods. First, we considered those individuals recaptured in 2012-13 who were captured three or more times in 2003-04. We considered the lizard occupied the same

individual space if any 2012-13 capture locations fell within the polygon of 2003-04 capture locations, or less than 8 m (allowing for 4 m GPS error in each survey period) outside the polygon edge. Similarly, if any 2012-13 capture locations fell within or less than 8 m outside the polygon of its 2003 -04 social group, we considered the lizard occupied the same group space. If all 2012-13 capture locations of an individual were further than 8 m outside both the previous individual and social group spaces, we deduced it had shifted space use, and we visually assessed if it now occupied the group space of a different 2003-04 social group. We could not derive polygons for lizards only captured at one or two locations in 2003-04. In these cases, we assessed site fidelity based on the distance between capture points in the two survey periods. Distances were calculated using the geographic distance option in GenAlEx v6.41 (Peakall and Smouse 2006). If we recaptured a lizard within 8 m of at least one of its 2003-04 capture locations, we considered it to be in the same individual space. Additional information on the long-term stability of social bonds came both from considering cases of two or more lizards occupying the same shared space over both surveys, and from assessing any change in status of recaptured lizards that were isolates in the first survey.

Predictors of recapture and site fidelity

First, we compared lizards from the first survey that we recaptured in the second survey to lizards that we did not recapture in the second survey. We considered recapture to be an indicator of survival in the population, although we acknowledge that some lizards may have survived but not been recaptured in the second survey period. Then, among the recaptured lizards, we compared those still occupying the same space with those recaptured in a different space. We assessed five potential predictors of *E. stokesii* recapture and site fidelity (see below) using an AIC model selection approach. For both analyses we used Generalized Linear Mixed Models (*glmer*) implemented in R (version 3.1.3) using the lme4 (Bates *et al.* 2014) and AIC_cmodavg (Mazerolle 2015) packages. In analyses of recapture likelihood in

2013-14, we included whether the lizard was recaptured as a binomial response variable (0 = not recaptured, 1 = recaptured), and population (CAM, CAS and CRR) as a random effect. In analyses of site fidelity (recapture location), we included as a binomial response variable whether the lizard was in a different (= 0) or the same (= 1) space, and population as a random effect. In each model set, we initially included five predictor variables, a fixed effect of body size (as at 2003-04 capture, indicated by SVL), individual genetic diversity, population crevice density (2003-04 crevices per hectare as detailed in Gardner *et al.* 2007), sex, and 2003-04 social status (group member, floater or isolate). Individual genetic diversity was measured from the genotypes across the 12 microsatellite loci for each sampled individual from the 2003-04 survey as homozygosity by locus (HL, Aparicio *et al.* 2006), calculated in R as implemented in Genhet (Coulon 2010). HL ranges from 0 when all loci are heterozygous to 1 when all loci are homozygous i.e. lower HL values indicate higher genetic diversity (Aparicio *et al.* 2006). After preliminary modelling (results not shown), we excluded sex from both analyses due to the extent of missing data resulting mainly from juveniles of unknown sex. In addition, we excluded social status from site fidelity models due to low numbers of recaptured lizards that had not previously belonged to a social group.

We incorporated the remaining variables (and combinations of variables) within several alternative hypotheses about factors that influenced: 1) the likelihood of recapturing a lizard and 2) among recaptured lizards, the likelihood of recapture within the same or a different space (Table 1). Each hypothesis corresponds to a model in our model set. Model selection was performed using an information theoretic approach (Burnham 2002) where the goodness of fit of models was measured with the Akaike Information Criterion (AIC) (Akaike 1973), corrected for small sample size (AIC_c) (Hurvich and Tsai 1989). Within each model set, models were ranked using AIC_c, and ΔAIC_c (the difference between the top model and all subsequent models) was used to calculate Akaike weights (ω) and identify a candidate set of

models. Models with $\Delta AIC_c < 2$ were deemed to have substantial support (Burnham 2002) although all models with $\Delta AIC_c \leq 7$ were considered as possible candidate models (Burnham *et al.* 2011). We used model averaging to derive coefficient values and 95% confidence intervals (CIs) for each of the variables included in more than one model in a model set.

Because in the analysis output individual genetic diversity was included in the model set predicting recapture (see results), we also investigated whether any specific alleles were associated with lizard persistence. Alleles more common in recaptured lizards than those not recaptured may indicate alleles associated with genetic fitness. We considered all alleles with a frequency of 20% or more in a population in the first sampling period. Fisher exact tests were used to detect any significant change in frequency of an allele among recaptured lizards compared to lizards not recaptured in the second sampling period. Fisher exact tests were first performed only for adult lizards in 2003-04 (to remove potential biases from sampling families), and then for all lizards captured and genotyped in 2003-04. Significance levels were corrected using a false discovery rate test (threshold of 0.1, Benjamini and Hochberg 1995).

Results

Field surveys, genetic and social analyses 2003-04

A total of 182 lizards were located across CAM, CAS and CRR during 66 sampling days in 2003 -04 (Table 2), made up of 159 lizards in 33 social groups, plus nine floaters, 11 isolates and three lizards with an ambiguous social status (it was unclear if they were group members or isolates) (Gardner *et al.* 2001; Godfrey *et al.* 2006). Lizards were captured a mean of 3.85 (SE 0.183) times (range one – 17) (Godfrey *et al.* 2006). Site characteristics (taken from Gardner *et al.* 2007) are shown in Table 2.

Field surveys 2012-13

We sampled 260 lizards during 56 sampling days in 2012-13 (Table 2). Lizards were captured a mean of 3.90 (SE 0.281) times (range one – 30) (Pearson, unpublished data). Forty-eight (26.5%) captured individuals from 2003-04 were recaptured in 2012-13 (19 males and 29 females). Genotypes derived in 2003-04 were available for 41 of the 48 recaptured lizards. Among the recaptures, 35 (73%) were adults and 13 (27%) were sub-adults in 2003-04. We did not recapture any of the 40 juveniles captured in 2003-04. See Pearson Appendix 6 for an estimate of *E. stokesii* longevity based on recapture data. Of the 48 recaptures, five had inadequate location data in 2003-04 to assess social group membership. Among the other 43 lizards, 34 (79%) had previously been assigned to a social group, five (11%) had been described as floaters, and two (5%) as isolates. Another two (5%) previously had ambiguous social group membership (it was unclear if they were group members or isolates).

Site fidelity

We included 43 lizards with capture location data from both surveys in our spatial analyses, of which 28 (65%) were recaptured in the same individual or group space and 15 (35%) in a different space (Table 2, Fig. 1). Of the 34 social group lizards from the first survey, 24 (71%) lizards were recaptured in the same individual or group space, fifteen of those (63%) were in the same individual and group space, and nine (37%) were in the same group space, but a different individual space. We recaptured four out of five floaters (80%) in the same individual space and one in a different space. The two lizards for which it was previously unclear if they were isolates or belonged to a group, were each recaptured in a different individual space. We also recaptured the two isolates in a different individual space. Of these four lizards, three were in the same population site, but one had moved about 300 m between population sites (captured in CAS in 2003-04 and CRR in 2012-13). An additional two lizards not previously sampled in 2003-04 were observed to move approximately 350 m between populations within the 2012-13 field survey season (Pearson, unpublished data).

Although five of the nine recaptured, non-social group members (56%) had shifted space, compared to ten of 34 previous social group members (29%), the difference in spatial stability between previous social group members and non-members was not statistically significant (chi-squared = 2.14; df = 1; $P = 0.14$).

There were six cases of two and one case of three lizards from the same social group in 2003-04 being recaptured in the same shared space in 2012-13 (Table 2). These 15 individuals represented 35% of lizards included in the spatial analysis. Additionally, we recaptured both of the 2003-04 isolates in crevices occupied, although not at the same time, by other lizards during 2012-13. Thus, we deduced they had now joined a social group. One lizard with an ambiguous social status in 2003-04, also appeared to have joined a social group because it was recaptured sharing a crevice with another lizard in 2012-13. The other ambiguous social status lizard in 2003-04 was captured only once in 2012-13 and was the sole user of a crevice at the time of capture.

Predictors of recapture and site fidelity

We compared eleven models for their ability to predict recapture likelihood and seven models for recapture location (Table 3). Model outputs are available in detail in Appendix 1, Supplementary Material. From the recapture likelihood model set, *Body size* was the only model with $\Delta AIC_c \leq 2$, and had the highest ω (0.687) and lowest AIC_c (169.14) (Table 3). We were more likely to recapture larger lizards than smaller lizards (Fig. 2). Three further models, each including *Body size* (in combination with individual genetic diversity, crevice density, and social status respectively) had $\Delta AIC_c \leq 7$ (Table 3). In combination with larger body size of recaptured lizards, trends indicated: 1) recaptured lizards had higher heterozygosity (i.e. lower HL) than lizards not recaptured, 2) recapture was more likely in populations with higher rather than lower crevice density, and 3) floaters were more likely to be recaptured (55%) than social group lizards (21%) or isolates (18%). The 95% CIs of the

model-averaged coefficients for *Body size* did not overlap zero whereas the CIs of all other variables included zero (Appendix 1).

From the recapture location model set, two models had $\Delta AIC_c \leq 2$ (Table 3). *Body size* had the highest ω (0.429) and the lowest AIC_c (55.86) while the *Null* model ranked second (ω = 0.248, AIC_c = 56.96) (Table 3). We were more likely to recapture larger lizards in a different rather than the same space (Fig 2). Five further models had $\Delta AIC_c \leq 7$, two of which incorporated *Body size* (Table 3). We identified two trends. First, we were more likely to recapture lizards with lower heterozygosity (i.e. higher HL) in a different rather than the same space. Second, lizards occupying sites with higher crevice density were more likely to be recaptured in a different, rather than the same, space. The 95% CIs of the model-averaged coefficients of all variables included zero (Appendix 1).

In addition to incorporating genetic diversity in models, we compared allele frequencies between recaptured and not recaptured lizards. For this analysis, we used nineteen alleles, representing nine microsatellite loci, that occurred with a frequency of 20% or more in genotypes derived from the first survey. For adults only, one of those 19 alleles had a significantly higher frequency in recaptured lizards in one population (n = 21, allele 171, locus Est4, CAS; P=0.016). This result was not significant after applying a false discovery rate. Allele 171 was absent in CAM, and had similar frequencies in recaptured and not recaptured lizards in CRR (P>0.05, results not shown). Using all lizards (adults and juveniles; n=183), five alleles differed in frequency between captured and not recaptured lizards in individual populations, with two retaining significant differences after applying a false discovery rate (allele 150, locus Ecu2, CAS; $P < 0.001$ and allele 115, locus Ecu5, CAM, $P = 0.002$). Allele 150 was present in all populations but only significantly different in CAS, whereas allele 115 was only present in CAM.

Discussion

This study evaluated site fidelity in three populations of group living *Egernia stokesii* (gidgee skink) between two field surveys separated by almost a decade. We found 48 (26.5%) lizards from an earlier survey still present in our surveys nine years later. Body size was a key predictor of recapture likelihood: we were more likely to recapture larger lizards than smaller lizards. After an interval of nine years, 65% of *E. stokesii* recaptured across three populations occupied the same space they were first found in. In addition, we found evidence of long-term social bonds, with 36% of recaptured lizards sharing the same space with at least one other identical lizard in both surveys.

The most likely explanation of the fate of lizards we did not recapture is that they have not survived. We are confident we resurveyed most surviving lizards because dispersal has rarely been recorded, even between our very close study populations, and we could monitor every crevice refuge at each population site. Each population was surrounded and separated by habitat that lacked either suitable rocky refuges or alternate refuges. Therefore, it is unlikely that any lizards resided outside of the sampled areas. Thus, we considered recapture as a good proxy for survival. In this study, we were more likely to recapture larger *E. stokesii* from the first survey than smaller lizards, suggesting both a relatively long life, and a survival benefit derived from having achieved a large body size. Body size may also influence space use because there was also a trend for larger recaptured lizards to be in a different space. However, confidence intervals overlapped zero, indicating weak support for this finding. Both the population size and the number of social groups within a population are limited by the availability of crevices (Gardner *et al.* 2007). Being larger may allow lizards the competitive ability to move into higher quality crevices and thus survive longer.

Within the context of the ecological constraints models (Emlen 1982; Emlen 1995), a habitat with limited refuges may promote site fidelity and stable social bonds. Although *E. stokesii* occupy stable rocky habitats, the local density and location of crevices within a rock outcrop

is variable among population sites, and the number of crevices predicts both the number of *E. stokesii* individuals and the number of social groups within an outcrop (Gardner *et al.* 2007). Where alternate crevices are available, the heterogeneity of food resources, particularly plants which are the main component of adult *E. stokesii* diet (Duffield and Bull 1998), may influence *E. stokesii* site fidelity. Habitat availability influences space use within other group living lizard species. Examples include the availability of rocks and shrubs for White's skink (*E. whitii*; Chapple and Keogh 2006) and logs and rocks for desert night lizards (*X. vigilis*; Zweifel and Lowe 1966). Patterns of *E. striolata* aggregations have been shown to be influenced by rocky habitat heterogeneity, complexity and quality (Michael *et al.* 2010). Habitat quality can also influence space use. Common lizards (*Zootoca vivipara*) have been shown to disperse less in poorer quality habitats (Clobert *et al.* 1994) and an investigation of the impact of habitat fragmentation on *E. cunninghami* dispersal found lizards dispersed less in deforested habitat compared to habitat with natural vegetation (Stow *et al.* 2001). Trends indicated higher crevice density increased *E. stokesii* recapture likelihood and recapture within a different space. However, crevice density was not a major factor in recapture or recapture location models. Thus, it is unlikely that habitat alone influences site fidelity in this species. One explanation for this finding is that all crevices, and subsequently social groups, are saturated. Another possibility is that finer scale ecological factors, such as vegetation for food and shade, are more important than crevice density. Comprehensive investigations of the influence of ecological constraints on site fidelity will require sampling across more populations with a diversity of crevice densities and incorporating more ecological factors.

Benefits-of-philopatry models (Stacey and Ligon 1991; Emlen 1995) offer an alternative but not mutually exclusive explanation for *E. stokesii* site fidelity and the persistence of individuals within a shared space across a time gap of nearly a decade. Lizards may gain fitness benefits from staying in their social group. The benefits might arise from reliable

access to shelter or food resources and from the uncertainty of finding similar resources if they disperse. *Egernia stokesii* exhibit natal philopatry (Gardner *et al.* 2001), a behaviour reported in only a few other lizard species, including the great desert skink (*L. kintorei*; McAlpin *et al.* 2011) and desert night lizards (*X. vigilis*; Davis *et al.* 2011). In these cases there is likely to be tolerance of kin, also reported in other group living lizards, and the potential for indirect parental care (through allowing offspring access to adult refuges). Protection from infanticide and access to thermal and food resources have been proposed to explain the tolerance of adult black rock skinks (*E. saxatilis*) to their co-habiting juveniles (O'Connor and Shine 2004). Experiments have shown that *E. stokesii* voluntarily aggregate (Lanham 2001), lending support to models of natal philopatry as a basis for site fidelity. *Egernia stokesii* can recognise kin from non-kin (Main and Bull 1996) and group from non-group members (Bull *et al.* 2000), behaviour that may serve to reinforce the stability of groups once established within a cluster of adjacent available crevices (Main and Bull 1996). The benefits of philopatry may be reinforced by the long life span and viviparous reproductive mode to promote site fidelity in this and other group living lizard species (Gardner *et al.* 2015).

Although individual genetic diversity was not included in the top models to predict either the likelihood of recapture or the recapture location, it was included in other feasible models. We were more likely to recapture larger lizards, with higher heterozygosity than smaller, less heterozygous lizards. This trend suggests a fitness advantage associated with higher genetic diversity. However, confidence intervals only supported body size. We used twelve microsatellite loci to measure genetic diversity. A larger sample size (lizards x loci) may yield greater insights into the influence of genetic diversity on site fidelity. Although genetic diversity did not strongly predict recapture probability or space use, one microsatellite allele within one population, allele 171 at locus Est4, was more frequent among recaptured lizards

than those not recaptured. A previous study (Godfrey *et al.* 2006) reported a significantly different prevalence of infection of a blood parasite (*Hemolivia*) between *E. stokesii* lizards with or without another allele (159) at the same locus. While our result was not significant after applying a false discovery rate, the coincidental association of alleles at this locus with differential survival and differential parasite prevalence in two independent studies suggests that the genetic region close to this locus may have some influence on fitness. That warrants further investigation.

In summary, we found long-term site and social bond fidelity across three *E. stokesii* populations within a time span of almost a decade. When considered in conjunction with earlier findings of site fidelity in one population within and across six consecutive seasons (Gardner *et al.* 2001; Duffield and Bull 2002b), this study demonstrates that *E. stokesii* exhibit site fidelity across multiple temporal scales. Despite intensive monitoring, records of dispersal among populations have been rare, supporting our assumption of lizards not recaptured as not surviving across sampling periods. One lizard (out of 188) had moved between the surveyed populations during the period between 2003-04 and 2012-13. An additional two lizards were observed to move approximately 350 m between populations during 2012-13 (Pearson, unpublished data). Larger lizards were more likely to persist in populations i.e. body size predicted recapture likelihood. Two fundamental questions remain: 1) why do some members of these group living lizards remain in their space whereas others move? and 2) what is different about those that do move? An improved understanding of space use of group living lizards is warranted.

Acknowledgements

We thank the volunteers involved in the 2012-13 field surveys, and Aaron Fenner and Glen Duffield for earlier work. We also thank the two anonymous reviewers for their helpful comments and suggestions, which have improved the overall quality of the manuscript. Field

survey methods were approved by the Flinders University Animal Welfare Committee, using standards and procedures in compliance with the Australian Code of Practice for the use of animals for scientific purposes. This work was supported by The Field Naturalists Society of South Australia, Nature Foundation of SA, and the Holsworth Wildlife Research Endowment.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In 'Second International Symposium on Information Theory.' (Eds. B.N. Petrov and F. Caski) pp. 267-281. (Akademinai Kiado)

Aparicio, J. M., Ortego, J., and Cordero, P. J. (2006). What should we weigh to estimate heterozygosity, alleles or loci? *Molecular Ecology* **15**, 4659-4665.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2014). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* **67,** 1-48.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 289-300.

Börger, L., Dalziel, B. D., and Fryxell, J. M. (2008). Are there general mechanisms of animal home range behaviour? A review and prospects for future research. *Ecology letters* **11**, 637-650.

Boyer, N., Réale, D., Marmet, J., Pisanu, B., and Chapuis, J. l. (2010). Personality, space use and tick load in an introduced population of Siberian chipmunks *Tamias sibiricus*. *Journal of Animal Ecology* **79**, 538-547.

Bull, C. M., and Freake, M. J. (1999). Home-range fidelity in the Australian sleepy lizard, *Tiliqua rugosa*. *Australian Journal of Zoology* **47**, 125-132.

Bull, C. M., Griffith, S. C., Lanham, E. J., and Johnston, G. R. (2000). Recognition of pheromones from group members in a gregarious lizard, *Egernia stokesii*. *Journal of Herpetology* **34**, 92-99.

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

Burnham, K. P. (2002). 'Model selection and multi-model inference: A practical information-theoretic approach.' 2nd ed. edn. (Secaucus : Springer-Verlag New York, Incorporated: Secaucus)

Cézilly, F., Dubois, F., and Pagel, M. (2000). Is mate fidelity related to site fidelity? A comparative analysis in Ciconiiforms. *Animal Behaviour* **59**, 1143-1152.

Chapple, D. G., and Keogh, S. J. (2006). Group structure and stability in social aggregations of White's skink, *Egernia whitii*. *Ethology* **112**, 247-257.

Clobert, J., Massot, M., Lecomte, J., Sorci, G., de Fraipont, M., and Barbault, R. (1994). Determinants of dispersal behavior: the common lizard as a case study. In 'Lizard ecology:

historical and experimental perspectives'. (Eds L. J. Vitt and E. R. Pianka) pp. 183-206. (Princeton University Press: Princeton .)

Cogger, H. G. (1983). 'Reptiles and amphibians of Australia.' (AH & AW Reed Pty Ltd)

Cooper, S. J. B., Bull, C. M., and Gardner, M. G. (1997). Characterization of microsatellite loci from the socially monogamous lizard *Tiliqua rugosa* using a PCR-based isolation technique. *Molecular Ecology* **6**, 793-795.

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

Davis, A. R., Corl, A., Surget-Groba, Y., and Sinervo, B. (2011). Convergent evolution of kin-based sociality in a lizard. *Proceedings of the Royal Society of London, B* **278**, 1507-1514.

Doody, S. J., Burghardt, G. M., and Dinets, V. (2013). Breaking the social-non-social dichotomy: A role for reptiles in vertebrate social behavior research? *Ethology* **119**, 1-9.

Duffield, G. A., and Bull, C. M. (1996). Characteristics of the litter of the gidgee skink, *Egernia stokesii*. *Wildlife Research* **23**, 337-342.

Duffield, G. A., and Bull, C. M. (1998). Seasonal and ontogenetic changes in the diet of the Australian skink *Egernia stokesii*. *Herpetologica* **54**, 414-419.

Duffield, G. A., and Bull, C. M. (2002a). *Egernia stokesii* (gidgee skink) opportunistic dispersal. *Herpetological Review* **33**, 204-205.

Duffield, G. A., and Bull, C. M. (2002b). Stable aggregations in an Australian lizard, *Egernia stokesii*. *Naturwissenschaften* **89**, 424-427.

Ebrahimi, M., Fenner, A. L., Bull, C. M., and Godfrey, S. S. (2015). Long and short term residence in refuge burrows by endangered pygmy bluetongue lizards. *Amphibia-Reptilia* **36**, 119-124.

Effenberger, E., and Mouton, P. L. F. N. (2007). Space use in a multi-male group of the group-living lizard. *Journal of Zoology* **272**, 202-208.

Emlen, S. T. (1982). The evolution of helping. I. An ecological constraints model. *American Naturalist* **119**, 29-39.

Emlen, S. T. (1995). An evolutionary theory of the family. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 8092-8099.

Emlen, S. T., and Oring, L. W. (1977). Ecology, sexual selection, and the evolution of mating systems. *Science* **197**, 215-223.

Fernandez-Duque, E. (2015). Social monogamy in wild owl monkeys (*Aotus azarae*) of Argentina: the potential influences of resource distribution and ranging patterns. *American Journal of Primatology* **78**, 355-371.

Gardner, M. G. (2000). A genetic investigation of sociality in the Australian group living lizard *Egernia stokesii*. Doctoral dissertation. Flinders University of South Australia, Adelaide

Gardner, M. G., Bull, C. M., Cooper, J. B., and Duffied, G. A. (2001). Genetic evidence for a family structure in stable social aggregations of the Australian lizard *Egernia stokesii*. *Molecular Ecology* **10**, 175-183.

Gardner, M. G., 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**, 1787-1794.

Gardner, M. G., Bull, C. M., Fenner, A., Murray, K., and Donnellan, S. C. (2007). Consistent social structure within aggregations of the Australian lizard, *Egernia stokesii* across seven disconnected rocky outcrops. *Journal of Ethology* **25**, 263-270.

Gardner, M. G., Cooper, S. J. B., Bull, C. M., and Grant, W. N. (1999). Isolation of microsatellite loci from a social lizard, *Egernia stokesii*, using a modified enrichment procedure. *Journal of Heredity* **90**, 301-304.

Gardner, M. G., Godfrey, S. S., Fenner, A. L., Donnellan, S. C., and Bull, C. M. (2012). Fine-scale spatial structuring as an inbreeding avoidance mechanism in the social skink *Egernia stokesii*. *Australian Journal of Zoology* **60**, 272-277.

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

Godfrey, S. S. (2013). Networks and the ecology of parasite transmission: A framework for wildlife parasitology. *International Journal for Parasitology: Parasites and Wildlife* **2**, 235-245.

Godfrey, S. S., Ansari, T. H., Gardner, M. G., Farine, D. R., and Bull, C. M. (2014). A contact-based social network of lizards is defined by low genetic relatedness among strongly connected individuals. *Animal Behaviour* **97**, 35-43.

Godfrey, S. S., Bull, C. M., James, R., and Murray, K. (2009). Network structure and parasite transmission in a group living lizard, the gidgee skink, *Egernia stokesii*. *Behavioral Ecology and Sociobiology* **63**, 1045-1056.

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

Greenwood, P. J. (1980). Mating systems, philopatry and dispersal in birds and mammals. *Animal Behavior* **28**, 1140-1162.

Hurvich, C. M., and Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika* **76**, 297-307.

Kerr, G. D., and Bull, C. M. (2006). Exclusive core areas in overlapping ranges of the sleepy lizard, *Tiliqua rugosa*. *Behavioral Ecology* **17**, 380-391.

Kerth, G. (2008). Causes and consequences of sociality in bats. *BioScience* **58**, 737-746.

Kitchen, A. M., Gese, E. M., Karki, S. M., and Schauster, E. R. (2005). Spatial ecology of swift fox social groups: from group formation to mate loss. *Journal of Mammalogy* **86**, 547-554.

Lanham, E. J. (2001). Group-living in the Australian skink, *Egernia stokesii*. Doctoral dissertation. Flinders University of South Australia, Adelaide

Main, A. R., and Bull, C. M. (1996). Mother-offspring recognition in two Australian lizards, *Tiliqua rugosa* and *Egernia stokesii*. *Animal Behaviour* **52**, 193-200.

Mazerolle, M. J. (2015). Model selection and multimodel inference based on (Q)AIC(c). In '.' 1.21 edn. (R Foundation for Statistical Computing, Vienna)

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

Meise, K., Krüger, O., Piedrahita, P., and Trillmich, F. (2013). Site fidelity of male Galápagos sea lions: a lifetime perspective. *Behavioral Ecology and Sociobiology* **67**, 1001-1011.

Michael, D. R., Cunningham, R. B., and Lindemayer, D. B. (2010). The social elite: Habitat heterogeneity, complexity and quality in granite inselbergs influence patterns of aggregation in *Egernia striolata* (Lygosominae: Scincidae). *Austral Ecology* **35**, 862-870.

Moyer, M. A., McCown, J. W., Eason, T. H., and Oli, M. K. (2006). Does genetic relatedness influence space use pattern? a test on Florida black bears. *Journal of Mammalogy* **87**, 255-261.

O'Connor, D., and Shine, R. (2003). Lizards in 'nuclear families': a novel reptilian social system in *Egernia saxatilis* (Scincidae). *Molecular Ecology* **12**, 743-752.

O'Connor, D. E., and Shine, R. (2004). Parental care protects against infanticide in the lizard *Egernia saxatilis* (Scincidae). *Animal Behavior* **68**, 1361-1369.

Paull, S. H., Song, S., McClure, K. M., Sackett, L. C., Kilpatrick, A. M., and Johnson, P. T. (2011). From superspreaders to disease hotspots: linking transmission across hosts and space. *Frontiers in Ecology and the Environment* **10**, 75-82.

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

Ramos-Fernandez, G., Smith Aguilar, S. E., Schaffner, C. M., Vick, L. G., and Aureli, F. (2013). Site fidelity in space use by spider monkeys (*Ateles geoffroyi*) in the Yucatan peninsula, Mexico. *PLoS ONE* **8**, e62813.

Shields, W. M. (1984). Factors affecting nest and site fidelity in Adirondack barn swallows (*Hirundo rustica*). *The Auk* **101**, 780-789.

Shorey, L., Piertney, S., Stone, J., and Höglund, J. (2000). Fine-scale genetic structuring on *Manacus manacus* leks. *Nature* **408**, 352-353.

Smith, L. M., and Burgoyne, L. A. (2004). Collecting, archiving and processing DNA from wildlife samples using FTA databasing paper. *BMC Ecology* **doi:10.1186/1472-6785-4-4**.

Spiegel, O., Leu, S. T., Sih, A., Godfrey, S. S., and Bull, C. M. (2015). When the going gets tough: behavioural type-dependent space use in the sleepy lizard changes as the season dries. *Proceedings of the Royal Society B, Biological Sciences* **282**, 20151768.

Stacey, P. B., and Ligon, J. D. (1991). The benefits-of-philopatry hypothesis for the evolution of cooperative breeding: variation in territory quality and group size effects. *The American Naturalist* **137**, 831-846.

Stow, A. J. (2002). Microsatellite loci from the Cunningham's skink (*Egernia cunninghami*). *Molecular Ecology Notes* **2**, 256-257.

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**, 419-430.

Stow, A. J., Sunnucks, P., Briscoe, D. A., and Gardner, M. G. (2001). The impact of habitat fragmentation on dispersal of Cunningham's skink (*Egernia cunninghami*): evidence from allelic and genotypic analyses of microsatellites. *Molecular Ecology* **10**, 867-878.

Switzer, P. V. (1993). Site fidelity in predictable and unpredictable habitats. *Evolutionary Ecology* **7**, 533-555.

Wartmann, F., Juárez, C., and Fernandez-Duque, E. (2014). Size, site fidelity, and overlap of home ranges and core areas in the socially monogamous owl monkey (*Aotus azarae*) of Northern Argentina. *The Official Journal of the International Primatological Society* **35**, 919-939.

Wey, T., Blumstein, D. T., Shen, W., and Jordan, F. (2008). Social network analysis of animal behaviour: a promising tool for the study of sociality. *Animal Behaviour* **75**, 333-344.

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

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**, 938-949.

Wolf, J. B. W., and Trillmich, F. (2008). Kin in space: social viscosity in a spatially and genetically substructured network. *Proceedings of the Royal Society B, Biological Sciences* **275**, 2063–2069.

Zhou, Q., Tang, X., Huang, H., and Huang, C. (2011). Factors affecting the ranging behavior of white-headed langurs (*Trachypithecus leucocephalus*). *International Journal of Primatology* **32**, 511-523.

Zweifel, R. G., and Lowe, C. H. (1966). The ecology of a population of *Xantusia vigilis*, the desert night lizard. *American Museum novitates* **2247**, 1-57.

Tables and figures:

Table 1. Generalized Linear Mixed Models, shown with corresponding hypotheses, and predictions, used to assess *Egernia stokesii* 1) recapture likelihood, and 2) recapture location (i.e. recaptured in the same or different space).

Model name	Hypothesis	Prediction
Body size		
Recapture likelihood	Mortality rates are higher for juvenile lizards	Larger (~older) lizards are more likely to have survived and be recaptured than smaller lizards
Recapture location	Larger (~older) lizards are more likely to have established positions in their social groups due to a competitive advantage	Larger (~older) lizards are more likely to be recaptured in the same location than smaller lizards
Body size x Crevice density		
Recapture likelihood and recapture location	Crevice availability influences the likelihood of lizards dispersing and body size predicts competitive ability	Larger (~older) lizards with an established position in a social group are more likely to be competitive in retaining resources than smaller lizards when crevice availability is limited, therefore larger (~older) lizards are more likely to be recaptured, and recaptured in the same space
Body size x Genetic diversity		
Recapture likelihood and recapture location	Genetic diversity promotes fitness and survival	More genetically diverse and larger (~older) lizards are more likely to survive and maintain their group membership and therefore more likely to be be recaptured and recaptured in the same space than smaller lizards with lower genetic diversity.
Body size x Social status		
Recapture likelihood	Larger lizards are more likely to be competitive at gaining and maintaining a position in a social group and therefore less pressured to disperse	Larger lizards are more likely to be recaptured than smaller lizards that have not established a position in a group.

Crevice density		
Recapture likelihood and recapture location	Higher crevice density (i.e. more crevices available per hectare) reduces pressure on lizards to disperse	Lizards in higher density populations are more likely to be recaptured, and recaptured in the same space
Crevice density x Genetic diversity		
Recapture likelihood and recapture location	Higher genetic diversity may enhance a lizards capacity to cope with a resource shortage (i.e. lower crevice density)	Where crevice density is lower, lizards with higher genetic diversity are more likely to be recapatured, and maintain their position and therefore be recaptured in the same space than lizards than lizards with lower genetic diversity
Crevice density x Social status		
Recapture likelihood	Crevices limit the number of social groups and social group membership reduces pressure to disperse	Lizards belonging to a social group are more likely to be recaptured
Genetic diversity		
Recapture likelihood and recapture location	Genetic diversity promotes fitness and survival and larger (~older) lizards are likely to have a competitive advantage	Lizards with higher genetic diversity are more likely to be recaptured, and more likely to maintain group membership and therefore be recaptured in the same space, than lizards with lower genetic diversity
Social status		
Recapture likelihood	Lizards that are members of a social group are less likely to disperse as they have access to resources	Lizards belonging to a social group are more likely to be recaptured than lizards that do not belong to a social group
Social status x Genetic diversity		
Recapture likelihood	Lizards with higher genetic diversity (~ higher fitness) are more likely to gain membership to a social group and its associated resources	Lizards with higher genetic diversity are more likely to be recaptured than lizards with lower genetic diversity
Null model	Recapture likelihood and recapture location is random	Recapture likelihood and recapture location is random

Not all models were applied to recapture location due to a low number of data points.

Table 2. Site characteristics and number of *Egernia stokesii* captured during two field surveys, the number of lizards used in the analysis of long-term space use across three populations, and the results of the spatial analysis.

Site	Site area (ha)	Crevices (/ha)	2003-04	2012-13	Recaps	Spatial analyses	Same space	Different space	Shared space
CAM	1	25 (25)	55	82	17	15	10	5	2
CAS	6.1	21 (3.4)	49	92	15	14	8	6	1
CRR	6.4	42 (6.5)	78	86	16	14	10	4	4
Total	-	-	182	260	48	43	28	15	7

Sites: CAM (Camel Hill), CAS (Castle Rock), CRR (Castle Rock Ridge); site area and crevices taken from Gardner *et al. 2007*, 2003-04: lizards captured during 2003-04 surveys; 2012-13: lizards captured during 2012-13 surveys; Recaps: number of lizards originally captured in 2003-04 and recaptured in 2012-13; Spatial analyses: number of recaptured lizards included in spatial analyses; Same space: number of recaptured lizards found in the same space; Different space: number of recaptured lizards found in a different space; Shared space: number of instances of ≥ two recaptured lizards found in the same shared space.

Table 3. Results of Generalized Linear Mixed Models investigating *Egernia stokesii* recapture likelihood and recapture location (i.e. whether the recaptured lizard was in the same or different space).

Rank	Model	*k*	AIC_c	ΔAIC_c	ω
	Recapture likelihood				
1	**Body size**	**3**	**169.14**	**0.00**	**0.687**
2	Body size x Genetic diversity	5	172.43	3.29	0.133
3	Body size x Crevice density	5	172.89	3.75	0.105
4	Body size x Social status	7	173.57	4.43	0.075
5	Genetic diversity	3	191.48	22.35	0.000
6	Null model	2	191.67	22.54	0.000
7	Crevice density	3	193.39	24.25	0.000
8	Social status	4	193.42	24.29	0.000
9	Social status x Genetic diversity	7	194.00	24.87	0.000
10	Crevice density x Genetic diversity	5	195.27	26.13	0.000
11	Crevice density x Social status	7	196.78	27.64	0.000
	Recapture location				
1	**Body size**	**3**	**55.86**	**0.00**	**0.429**
2	**Null model**	**2**	**56.96**	**1.10**	**0.248**
3	Genetic diversity	3	59.01	3.15	0.089
4	Crevice density	3	59.25	3.39	0.079

5	Body size x Genetic diversity	5	59.75	3.89	0.061
6	Crevice density x Genetic diversity	5	59.99	4.13	0.055
7	Body size x Crevice density	5	60.64	4.78	0.039

k: number of parameters; ΔAIC_c: increase in AIC_c compared with the top model; ω:model weight. Models shown in bold were considered based on $\Delta AIC_c \leq 2$.

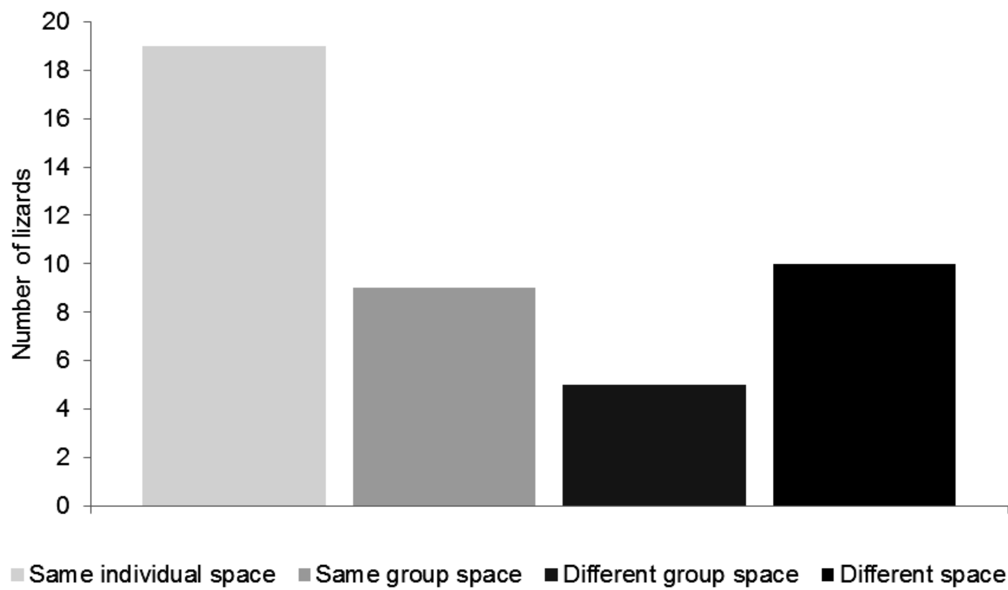


Fig. 1. Results of analysis of *Egernia stokesii* space use change over two field surveys nine years apart. The number of lizards recaptured in the same space (same individual space, same group space) or different space (different group space, or > 8 metres of original capture location/s) to that occupied when captured during 2003-04 is shown. Total of 43 recaptured lizards; 28 (65%) in the same space, 15 (35%) in a different space.

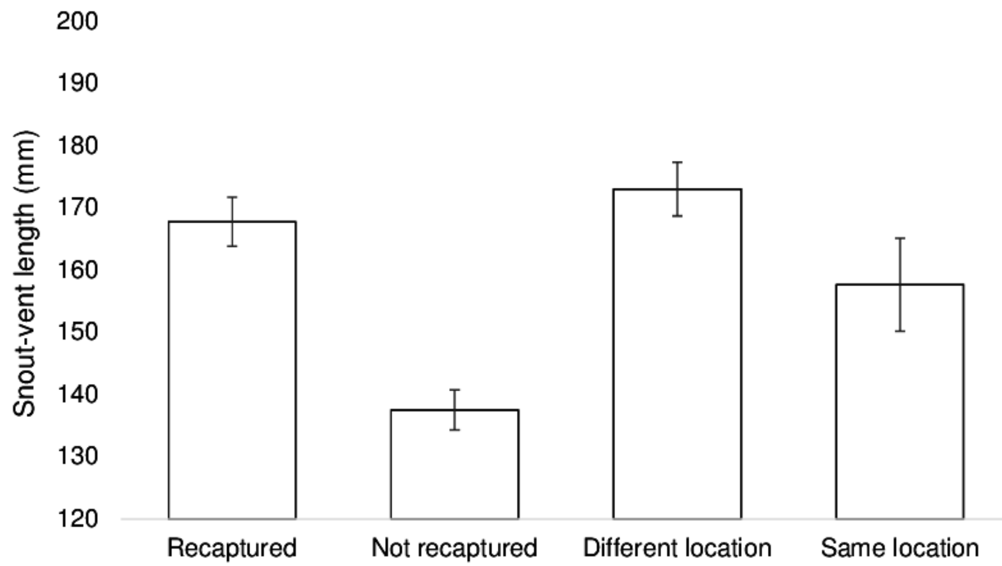


Fig. 2. Mean and standard error of *Egernia stokesii* body size (mm), comparing 1) lizards recaptured and not recaptured between 2003-04 and 2012-13 and 2) lizards found in the same and different space.

Larger lizards live longer in the group living *Egernia stokesii*

Pearson, S.K.[A,B], Godfrey, S.S. [C], Bull, C.M.[A], Gardner, M.G.[A,D]

Appendix 1, Supplementary Material

Table 1: *Egernia stokesii* recapture likelihood GLMM model estimates, standard errors and lower and upper confidence intervals (CIs) shown for variables used in models.

	Estimate	Std. Error	Lower, Upper 95% CI
Recapture likelihood			
Intercept	-1.271	0.201	-1.664, -0.878
Crevice density	0.013	0.019	-0.025, 0.051
Homozygosity by locus	-1.771	1.188	-4.099, 0.557
SVL	**0.288**	**0.067**	**0.156, 0.419**
Social Status Group Member	-1.113	0.705	-2.496, 0.270
Social Status Isolate	-1.277	1.036	-3.307, 0.753
Recapture location			
Intercept	0.657	0.329	-0.059, 0.073
Crevice Density	0.007	0.034	-0.059, 0.073
Homozygosity by locus	1.156	2.173	-3.103, 5.415
SVL	0.245	0.136	-0.022, 0.512

Variables for which confidence intervals did not overlap are shown in bold text. Three categories were utilised for Social Status (group member, floater, isolate) for which floater was the reference category in modelling).

Figure 1: Google Earth (2016) image of three populations of *Egernia stokesii* surveyed during 2003-04 and 012-13. Populations are 1) Camel Hill, 2) Castle Rock, and 3) Castle Rock Ridge.

Chapter 6: Genes and group predict mates in a lizard

Submitted as:

Pearson, S.K., Godfrey, S. S., Schwensow, N., Bull, C.M., and Gardner, M.G, Genes and group membership together predict reproductive pairs in the family living lizard *Egernia stokesii*, *Journal of Heredity*, June 2016.

MHC variation is hypothesised to be due to parasite mediated selection or MHC based mate choice. Chapter 4 demonstrates that selection is acting on *E. stokesii* MHC alleles and Chapter 5 demonstrates a high level of *E. stokesii* site and social fidelity. Longevity, high site fidelity, and long term social bonds may constrain mate choice and subsequently increase pressure on MHC based mate choice. The following chapter utilises MHC alleles identified in Chapter 3 and sites of selection identified in Chapter 4 to investigate factors predicting *E. stokesii* mate choice.

Genes and group membership predict gidgee skink (*Egernia stokesii*) reproductive pairs

Pearson, Sarah K.[A,B], Godfrey, Stephanie S.[C], Schwensow, N.[D,E], Bull, C. Michael[A], Gardner, Michael G.[A,F]

[A]School of Biological Sciences, Flinders University of South Australia

[C]School of Veterinary and Life Sciences, Murdoch University, Australia

[D] School of Biological Sciences, University of Adelaide, Australia

[E]Institute of Evolutionary Ecology and Conservation Genomics, University of Ulm, Germany

[F]Evolutionary Biology Unit, South Australian Museum

[B]Corresponding author. School of Biological Sciences, Flinders University of South Australia, Bedford Park 5042, Australia. Phone: +61 08 8201 2315, Fax: +61 08 8201 3015, Email: sarah.pearson@flinders.edu.au

Running title: Genes and group predict mates in a lizard

Submitted to *Journal of Heredity*, June 2016

Abstract

Due to their role in mate choice, disease resistance and kin recognition, genes of the major histocompatibility complex (MHC) are good candidates for investigating the genetic basis of mate choice. MHC based mate choice is context dependent and influenced by many factors including social structure. A diversity of different social structures makes the *Egernia* group of lizards a suitable group for comparative studies of MHC based mate choice. We investigated mate choice in the gidgee skink (*Egernia stokesii*), a lizard that exhibits high levels of social group and spatial stability. Group membership was incorporated into tests of the *good genes as heterozygosity* and *compatible genes* hypotheses for adaptive (MHC) and neutral (microsatellite) genetic diversity (n = 47). We found females were more likely to pair with a male with higher MHC diversity and lower pairwise relatedness, and males were more likely to pair with a female with higher microsatellite heterozygosity and a lower proportion of shared MHC alleles. Also, lizards were more likely to mate with an individual from within, rather than from outside, their social group, which confirmed earlier findings for this species and indicates that mate choice had already largely occurred. Thus, a combination of genes and group membership, rather than group membership alone, predicted mate choice in this species. By providing a foundation for comparisons among other members of the *Egernia* group this work will contribute to an enhanced understanding of group formation within squamates and a deeper understanding of the evolution of sociality within all vertebrates.

Key words

MHC, mate choice, social structure, group living, lizard

Introduction

Why an individual chooses one mate over another remains a fundamental question of sexual selection. Individuals may choose a mate based on direct, non-genetic benefits for themselves or their offspring, such as access to resources and food or other forms of parental care, or based on indirect, genetic benefits via improved offspring fitness (Kokko *et al.*, 2003; Andersson & Simmons, 2006; Kempenaers, 2007). Mate choice for genetic benefits has subsequent consequences for genetic variation and genetic structure within populations, and for species boundaries (Bonneaud *et al.*, 2006; Gillingham *et al.*, 2009; Tung *et al.*, 2012; Huchard *et al.*, 2013). Two main hypotheses in relation to mate choice for genetic benefits are: 1) the *good genes as heterozygosity* hypothesis (Brown, 1997; Landry *et al.*, 2001), and 2) the *compatible genes* hypothesis (reviewed in Penn & Potts, 1999; Tregenza & Wedell, 2000; Neff & Pitcher, 2005). Although both hypotheses use the term 'genes', they usually refer to alleles. However, as that term is almost never used (Kempenaers, 2007) we use 'genes' here for consistency. The *good genes as heterozygosity* hypothesis refers to the genes of the chosen partner only and predicts a preference for genetic diversity associated with fitness traits (Brown, 1997; Landry *et al.*, 2001). The *compatible genes* hypothesis refers to a process of matching the genotypes of both partners and commonly predicts preference for a partner with a dissimilar genotype to maximise the genetic variability among the offspring (Neff & Pitcher, 2005) and to avoid inbreeding (Tregenza & Wedell, 2000). Good genes, on their own, confer additive benefits, whereas compatible genes, in combination with alleles from the other partner, provide non-additive benefits to genetic variation in offspring fitness (Neff & Pitcher, 2005; Puurtinen *et al.*, 2009; Gohli *et al.*, 2013). Both strategies confer genetic benefits to offspring.

Genes of the highly polymorphic major histocompatibility complex (MHC) have important roles in mate choice, disease resistance and kin recognition (Brown & Eklund, 1994;

Sommer, 2005a; Milinski, 2006; Piertney & Oliver, 2006), and thus are good candidates for investigating the genetic basis of mate choice (Huchard & Pechouskova, 2013; Winternitz *et al.*, 2015). A number of studies have reported observations of mate choice that support the *good genes as heterozygosity* hypothesis with regard to MHC genes, in that females prefer males with higher MHC diversity. For example, in the solitary rodent *Ctenomys talarum* (Talas tuco-tuco) females chose more MHC heterozygous males (Cutrera *et al.*, 2012) and in the obligate pair-living *Cheirogaleus medius* (fat-tailed dwarf lemur) females paired with males with a higher number MHC alleles than random (Schwensow *et al.*, 2008b). Similarly, a study of extra-pair paternity in the socially monogamous *Carpodacus erythrinus* (scarlet rosefinch) found females were less likely to cheat on social male pairs with higher MHC heterozygosity (Promerova *et al.*, 2011), and females of the cooperatively breeding *Acrocephalus sechellensis* (Seychelles warbler) were more likely to mate with extra-pair males that had higher MHC diversity than their male social partners (Richardson *et al.*, 2005).

Other studies have observed MHC based mate choice patterns that support the *compatible genes* hypothesis, commonly by the choice of partners with dissimilar genes, and by dis-assortative mating. For example, females prefer MHC dissimilar males in the solitary, promiscuous *Microcebus murinus* (grey mouse lemur; Schwensow *et al.*, 2008a; Huchard *et al.*, 2013) and the colonial living, monogamous *Halobaena caerulea* (blue petrel; Strandh *et al.*, 2012). Other studies have suggested MHC compatibility when mates are chosen with similar MHC genes (Sommer, 2005b; Bonneaud *et al.*, 2006; Sin *et al.*, 2015), potentially to avoid either disrupting co-adapted gene complexes or reduced levels of immunocompetence that may result from the presence of too many alleles (Nowak *et al.*, 1992; Neff, 2004).

However, some earlier studies did not detect MHC based mate choice (e.g. primate, Huchard *et al.*, 2010; e.g. canid, Galaverni *et al.*, 2015). In those cases, neither the *good genes as heterozygosity* nor the *compatible genes* hypotheses were supported (Kuduk *et al.*, 2014; Sepil

et al., 2015). This lack of support may indicate other mechanisms facilitate inbreeding avoidance, or individuals are unable to discriminate MHC profiles of potential mates (Sepil *et al.*, 2015). Alternatively, where males coerce females to mate, female pre-copulatory choice may be reduced or absent (Clutton-Brock & Parker, 1995; Bisazza *et al.*, 2001). Thus, although MHC may have a general role in mate choice, there is no clear trend among different species, probably because MHC mate choice is context dependent (Setchell & Huchard, 2010). One of the contexts in which mate choice occurs is population social structure which can constrain mate choice strategies. Investigating mate choice across species representing a range of social structures may improve our understanding of MHC dependent mate choice (Setchell & Huchard, 2010; Huchard & Pechouskova, 2013; Winternitz, 2015).

Comparative studies of MHC based mate choice in wild animal populations will benefit from consideration of a broader taxonomic range, particularly from vertebrate groups representing more basal lineages such as squamates, for example lizards. To date, few studies have investigated MHC based mate choice in lizards, let alone other reptiles (but see Olsson *et al.*, 2003; Miller *et al.*, 2009). The *Egernia* group of Australian scincid lizards includes a wide range both of social structures, from solitary, to seasonal yet stable breeding pairs, and to year round stable family groups, and of mating systems, from genetic monogamy, to polyandry and polygamy (Gardner *et al.*, 2015). One well-studied member of this group is *Egernia stokesii* (gidgee skink, J.E. Gray, 1845).

Egernia stokesii is a large (180 mm snout-vent length; Cogger, 1983), long – living (25 years; Duffield & Bull, 1996), viviparous lizard that occupies rocky outcrops across semi-arid Australia (Cogger, 1983). It lives in stable family groups, with high levels of genetic monogamy and low dispersal (Gardner *et al.*, 2001; Duffield & Bull, 2002; Gardner *et al.*, 2002). The number of *E. stokesii* social groups within a rocky outcrop is constrained by the availability of the rocky crevices they shelter in (Gardner *et al.*, 2007). There is genetic

clustering of this species within both social groups and populations, and this may reflect the spatial patterns of social groups and the isolated habitat availability for populations (Gardner *et al.*, 2007). Reports based on microsatellite DNA genotypes have shown that the parents of litters are less related to each-other than non-paired individuals in the population (Gardner *et al.*, 2001) and there is no evidence of significant inbreeding (Gardner *et al.*, 2012). Behavioural mate choice, rather than sex-biased dispersal, appears to be the means of avoiding inbreeding (Gardner *et al.*, 2012). Individuals can discriminate between conspecific kin and non-kin, and between group and non-group members based on both body odour cues and scats (Main & Bull, 1996; Bull *et al.*, 2000). These chemical cues may facilitate recognition and contribute to *E. stokesii* social group cohesion (Bull *et al.*, 2000).

In the current study, we extended previous studies by investigating evidence for *E. stokesii* discrimination among potential alternative mating partners based on several factors. Because *E. stokesii* are group living, we assessed the importance of belonging to the same social group in mate choice decisions. In addition, we were interested in evidence for, and the relative role of 'good' and 'compatible' genes, including both MHC and microsatellite markers, on *E. stokesii* mate choice. Here, we use the term 'mate choice' in the broadest sense of any process leading to a non-random pairing between individuals of the opposite sex (see Edward, 2015 for a detailed review of mate choice terminology). The choice of mating partner in *E. stokesii* is likely to be limited by fine scale social structure, limited dispersal and high levels of previously established monogamous partnerships. The *group membership* hypothesis suggests that living in a group confers a benefit of collaborative access to refuge crevices, and vigilance against approaching predators (Lanham & Bull, 2004). This predicts that lizards would prefer to mate with another individual from within their social group, to retain group membership and its benefits. Based on the assumption that higher genetic diversity represents higher fitness, the *good genes as heterozygosity* hypothesis predicts that lizards would be

more likely to choose a mate with higher genetic diversity in order to optimise genetic outcomes from mating events. The prediction of the *compatible genes* hypothesis is based on observations that group members reside in stable, close proximity to each other within rocky crevices, which is likely to increase the risk of parasite transmission (Godfrey *et al.*, 2006; Godfrey *et al.*, 2009). As a result, lizards should be more likely to mate with genetically dissimilar mates to increase the number of different MHC alleles in offspring for parasite and pathogen resistance. Because these three hypotheses are not necessarily mutually exclusive, our aim in this study was to determine the relative influence of each on mate choice in this lizard species. This work will provide a foundation for comparisons among other members of the *Egernia* group thereby enhancing our understanding of the formation of groups within squamates, a lineage in which group formation is rare (Gardner *et al.*, 2015), and facilitating a deeper understanding of the evolution of sociality within all vertebrates.

Methods

Study sample

We used DNA samples and social, microsatellite, and breeding pair data derived from earlier field surveys of *E. stokesii* at Camel Hill, a rocky outcrop located near Hawker in the southern Flinders Ranges of South Australia (31°54S; 138°25E). Field survey, laboratory, and bioinformatic methodology has largely been described elsewhere (Gardner, 2000; Gardner *et al.*, 2001; Duffield & Bull, 2002; Gardner *et al.*, 2007; Pearson *et al.* Chapter 3) therefore, for brevity, are summarised in the Supplementary Material (S1) rather than included here. MHC genotypes were derived for a 216 base pair region of MHC I exon 2 (corresponding to the α-1 domain) and a 102 base pair region of MHC II exon 2 (corresponding to the β-1 domain). Microsatellite genotypes were previously derived for six microsatellite loci and used in parentage analysis in Cervus (Marshall *et al.*, 1998) to determine both parents of juveniles and sub-adults in the population and hence to determine breeding pairs (Gardner, 2000; Gardner *et*

al., 2002). Although pairs were inferred from parentage analyses, we use the term 'mate', where relevant, for simplicity.

Group membership hypothesis

We predicted *E. stokesii* choice for a mate from within the same social group if the *group membership* hypothesis represented the process behind *E. stokesii* mate choice. The method for determination of social group membership is described in the S1 of the Supplementary Material and social group membership data are summarised in Table 1.

Measures used to test the good genes as heterozygosity *hypothesis*

We predicted *E. stokesii* choice for a mate with higher genetic diversity if the *good genes as heterozygosity* hypothesis represented the process behind mate choice in *E. stokesii*. Four within individual measures were used to assess genetic diversity: 1) number of MHC alleles, 2) mean within individual MHC genotypic distance (Schwensow *et al.*, 2008b), 3) within individual MHC functional distance (Radwan *et al.*, 2012), and 4) individual microsatellite homozygosity by locus (Aparicio *et al.*, 2006). These measures are described in detail in the S1 of the Supplementary Material.

Measures used to test the compatible genes *hypothesis*

Three pairwise measures were used to assess the prediction that *E. stokesii* mate choice for *compatible genes* would be demonstrated by a preference for a genetically dissimilar mate: 1) proportion of shared MHC alleles (Wetton *et al.*, 1987), 2) pairwise MHC genotypic distance (following Landry *et al.*, 2001), and 3) pairwise microsatellite relatedness (Richardson *et al.*, 2005). In addition, we predicted that the choosing individual's genetic diversity (as defined by the four measures described in *Measures used to test the* good genes as heterozygosity *hypothesis* above) would influence their mate choice (Reusch *et al.*, 2001; Aeschlimann *et al.*, 2003). We predicted that the strength of choice for genetic diversity would be higher for less

heterozygous individuals. All measures are described in detail in S1 of the Supplementary Material.

Mate choice test datasets

Our three hypotheses were tested across eight datasets of candidate partnerships. These datasets are briefly described here and detail is provided in S1 of the Supplementary Material. First, all tests were undertaken for both female choice and male choice because it is not known which is the 'choosy' sex in *E. stokesii*, or if mate choice is mutual. For simplicity, we just describe the methodology for female choice, but the methods for analysis of male choice were analogous. Second, we conducted analyses within the whole population, and then just within social groups. Third, we ran separate models for each of two datasets: one for individuals genotyped for both MHC I and microsatellites, another for individuals genotyped for both MHC II and microsatellites (hereafter called MHC I and MHC II datasets respectively). Thus, to explore the alternative hypotheses, eight separate datasets were analysed: female-choice and male-choice, each at the whole population and social group level, and for each of MHC I and MHC II. The numbers of pair-wise comparisons included in each test are provided in Table 2.

Hypothesis testing

To compare the relative importance of alternative hypotheses regarding mate choice in *E. stokesii*, for each data set, we employed an Akaike Information Criterion (AIC) (Akaike, 1973) model selection approach, using Generalized Linear Mixed Models (GLMM) (*glmer*) implemented in R (version 3.1.3) using the *lme4* (Bates *et al.*, 2014) and *AICcmodavg* (Mazerolle, 2015) packages. We considered models corresponding to the *good genes as heterozygosity* and *compatible genes* hypotheses, with *group membership* included as an additional possible explanatory factor only in the population level analyses. Model selection

was performed using an information theoretic approach (Burnham & Anderson, 2001), measuring the goodness of fit of models with the AIC, corrected for small sample size (AIC_c) (Hurvich & Tsai, 1989). Detailed GLMM and model selection methodology is provided in S1 of the Supplementary Material. We used model averaging to derive coefficient values and 95% confidence intervals (CIs) for each of the variables included in more than one model in a model set.

Data Availability

In accordance with the Journal of Heredity data archiving policy (Baker, 2013), upon acceptance of this manuscript we are prepared to deposit the primary data underlying these analyses as follows:

- MHC and microsatellite genotypes: Dryad

- MHC sequences: Dryad

- Pair data: Dryad

Results

Study sample

We identified 31 MHC I and five MHC II alleles in the samples assayed in this study. Social data were available for 67 adults, including 27 breeding pairs (Table 1). Following sample filtering (Pearson *et al.* Chapter 3), we retained MHC I genotypes for 47 adults (25 males, 22 females) and MHC II genotypes for 46 adults (25 males, 21 females) (Table 1). Retained samples represented 17 breeding pairs for MHC I (including eleven within group pairs and six extra group pairs) and 13 for MHC II (including nine within group pairs and four extra group pairs) (Table 1). There was no significant difference in the mean number of alleles per individual between males and females for either MHCI or MHC II ($P > 0.05$, analysis results

not shown). Next, we provide a summary of the outputs from GLMM in relation to the *group membership, good genes as heterozygosity*, and *compatible genes* hypotheses.

Alternate hypotheses

Among all the models tested at the whole of population level, seven models had $\Delta AIC_c \leq 2$, all of which included group membership (Table 2). Four of these incorporated *group membership* and *good genes as heterozygosity* measures, two incorporated *group membership* and *compatible gene* measures, and one was *group membership* alone (Table 2). In most cases, models that incorporated both genes and group membership ranked higher than group membership alone (Table 2). At the social group level, two non-null models had $\Delta AIC_c \leq 2$, one representing the *good genes as heterozygosity hypothesis*, the other representing the *compatible genes* hypothesis (Table 3). Next we report results pertaining to each dataset. Because *group membership* had overall strong support, we focus our reporting of results on the genetic hypotheses. In addition, models in the 95% confidence set were generally equivocal therefore we focus on reporting models with $\Delta AIC_c \leq 2$. Details for all models within 95% confidence sets are presented in Table 2 and 3 and model averaged coefficients with 95% confidence intervals are reported in S2 of the Supplementary Material.

MHC I-associated female choice

For female-choice in the whole of population MHC I dataset, evidence was strongest for the *good genes as heterozygosity* hypothesis, with male MHC I functional distance in addition to the influence of group membership the most highly ranked model. Females were more likely to pair with a male from within the same social group with a higher functional distance of amino acids at MHC I positively selected sites (Fig. 1). Although the support for this model was relatively weak ($\omega = 0.261$, Table 2), the model averaged coefficient confidence intervals did not overlap zero (S2, Supplementary Material). This variable also ranked highest in the

social group level analyses, and the weighting was slightly stronger (ω = 0.353, Table 3) and model averaged coefficient confidence intervals did not overlap zero (S2, Supplementary Material). Other models with $\Delta AIC_c \leq 2$ included another *good genes as heterozygosity* variable that ranked higher than group membership alone, the number of MHC I alleles, but this was very weakly supported (ω = 0.189), and the *compatible genes* variable pairwise microsatellite relatedness, but this was ranked below a model containing group membership alone (Table 3).

MHC II-associated female choice

For female choice in the whole of population MHC II dataset, evidence was strongest for the *compatible genes* hypothesis with pairwise microsatellite relatedness in addition to the influence of group membership the most highly ranked model (Table 2). Females were more likely to pair with a male from within the same social group with lower pairwise microsatellite relatedness (ω = 0.478). For female choice in the social group MHC II dataset, the null model was the only model with $\Delta AIC_c \leq 2$ (ω = 0.825, Table 3). The 95% CIs of the model averaged coefficients for female choice for group membership in the whole population analysis and female choice for male HL and proportion of shared MHC II alleles in the social group analysis did not overlap zero whereas the CIs of all other variables included zero (S2, Supplementary Material).

MHC I-associated male choice

For male choice in the whole of population MHC I dataset, evidence was strongest for the *good genes as heterozygosity* hypothesis with female HL in addition to the influence of group membership the most highly ranked model (Table 2). Males were more likely to pair with a female from within the same social group with higher microsatellite heterozygosity (ω = 0.409, Fig. 1). For male choice in the social group level MHC I dataset, one non-null model

had $\Delta AICc \leq 2$: MHC I alleles shared ($\omega = 0.354$, Table 3). Males were more likely to pair with a female with a lower proportion of shared MHC I alleles (i.e. less similar). The 95% CIs of model averaged coefficients for group membership and pairwise microsatellite relatedness in the population level analysis did not overlap zero whereas the CIs of all other variables included zero (S2, Supplementary Material).

MHC II-associated male choice

As with MHC I, for male choice in the whole of population MHC II dataset, evidence was strongest for the *good genes as heterozygosity* hypothesis with female HL in addition to the influence of group membership the most highly ranked model ($\omega = 0.407$, Table 2). For male choice in the social group MHC II dataset, only the *Null* model had a $\Delta AIC_c \leq 2$ ($\omega = 0.830$, Table 3). The 95% CIs of model averaged coefficients for male choice for group membership in the population level analysis did not overlap zero whereas the CIs of all other variables included zero (S2, Supplementary Material).

Discussion

In this study we found that *E. stokesii* pairings were predicted by both genetic factors and social group membership, with a combination of group membership and genes having more influence than group membership alone. Thus, this work develops conceptual understanding beyond previous work that found *E. stokesii* usually pair with unrelated individuals within their social group (Gardner *et al.*, 2001) and provides new insights into factors that might influence choice of partners within *E. stokesii* social groups. Group membership was a key predictor of *E. stokesii* pairs which indicates that when lizards were sampled, mate choice had already largely occurred. We found evidence to support both the *good genes as heterozygosity* (MHC functional distance, number of MHC alleles, microsatellite heterozygosity) and the *compatible genes* (pairwise microsatellite relatedness, shared MHC alleles) hypotheses to

explain mate choice preferences. Results indicate both male and female *E. stokesii* genetic makeup is a factor in *E. stokesii* mate choice. Overall, results support more than one hypotheses, suggesting that multiple factors are involved in *E. stokesii* mate choice.

Group membership was incorporated in all models within our 95% confidence set, indicating that social group membership had a major influence on *E. stokesii* mate-choice. Individuals were more likely to pair if they were in the same group, although there was still a substantial number of extra group pairings within the whole population level data sets. Overall, however, models incorporating both group membership and genetic factors ranked more highly than group membership alone. Although there was more support for the *good genes as heterozygosity* than *compatible genes* hypothesis, evidence for both highlights the complexities of examining mate choice where more than one hypothesis gains support. While results indicated the genetic makeup of both males and females are involved in *E. stokesii* mate choice, there is no clear trend which suggests a lack of sex-bias in *E. stokesii* mate choice. Although previous work found breeding pairs had lower pairwise microsatellite relatedness than random (Gardner *et al.*, 2001), this factor only ranked highly in our analyses in models of female mate choice at the population level for the MHC II dataset, revealing that factors other than microsatellite relatedness are better predictors of *E. stokesii* pairings. MHC-associated variables generally ranked the highest in female choice models whereas microsatellite-associated models generally ranked the highest in male choice models. Although sex-specific differences in mate choice are beginning to be identified (Bahr *et al.*, 2012), how contrasting strategies evolve is not yet clear (Edward & Chapman, 2011). Results suggest that both adaptive and neutral genes influence *E. stokesii* mate choice. Although MHC-associated models in the MHC I datasets were generally equivocal, results indicate a role for MHC based mate choice. For MHC II data sets, the role of MHC in partner choice was less clear. Although this implies MHC II genotypes have less influence on mate choice,

low variation in this region may have reduced the power to detect a pattern. Alternatively, sequencing a greater region of MHC II may yield greater insights into the role of MHC II in *E. stokesii* mate choice.

We found evidence for a preference for both 'good' and 'compatible' genes in *E. stokesii* pairings. *Egernia stokesii* preference for partners with high genetic diversity (the *good genes as heterozygosity* hypothesis) was evident in female choice for males with higher MHC I functional distance and higher number of MHC I alleles, and male choice for females with higher microsatellite heterozygosity. Our findings of a preference for a genetically diverse mate are consistent with the findings for other species including *Passer domesticus* (house sparrow; Bonneaud *et al.*, 2006), *C. talarum* (Talas tuco-tuco; Cutrera *et al.*, 2012), and *C. medius* (fat-tailed dwarf lemur; Schwensow *et al.*, 2008b). *Egernia stokesii* preference for genetically dissimilar individuals (the *compatible genes* hypothesis) was evident in female choice of partners with lower pairwise microsatellite relatedness and male choice of partners with a lower proportion of shared MHC I alleles. This choice for dissimilarity is consistent with findings in other species including *M. murinus* (grey mouse lemur; Huchard *et al.*, 2013) and *H. caerulea* (blue petrel; Strandh *et al.*, 2012). On the other hand, it contrasts with results found for *P. domesticus* (house sparrow; Bonneaud *et al.*, 2006; Bichet *et al.*, 2014) and *Meles meles* (badger; Sin *et al.*, 2015) in which similar mates were preferred. The aforementioned species represent a range of social structures including colonial, solitary, obligate pairs, and social groups. Although both *M. meles* and *E. stokesii* form stable social groups, the species contrast in their preferences for compatible genes.

Adaptation to local parasites can favour mating with similar individuals (Bonneaud *et al.*, 2006; Sin *et al.*, 2015) yet our results indicate *E. stokesii* prefer to pair with dissimilar individuals. Living a long life and forming stable pair bonds within a social group may strengthen *E. stokesii* preference for dissimilar individuals. *Egernia stokesii* preference for

more genetically diverse and dissimilar social group mates will generate higher diversity among offspring, thereby increasing offspring potential to resist pathogens. This may be particularly important given *E. stokesii* longevity and stability of social groups and pair bonds which make mate decisions critical. *Egernia stokesii* dependence on rocky crevices within limited rocky outcrops, which promotes high relatedness among individuals within social groups (Gardner *et al.*, 2007), would further strengthen the importance of an unrelated mate. Although we found evidence for both a 'good' and 'compatible' gene basis for *E. stokesii* pairings, it is not clear how these patterns arise. However, given the importance of group membership, mate choice dynamics may be closely linked to social group dynamics.

Our findings suggest that *E. stokesii* social group formation and maintenance is a critical factor in mate choice. Although our findings indicate a genetic basis for *E. stokesii* pairings within social groups, it is not known how *E. stokesii* groups are formed and when such preferences may be realised i.e. before or after group formation. One possibility is that individuals find a partner with 'good' (e.g. higher MHC functional distance) and/or 'compatible' (e.g. lower proportion of shared MHC alleles) genes and then establish a new group if vacant, suitable habitat is available. However this mechanism is constrained by the limited number of rocky crevices available and low dispersal capacity. A second possibility is that dispersers differentially settle in groups of individuals with higher than average genetic diversity, or that individuals within a group exhibit choice by allowing or rejecting potential new occupants of groups before they have settled. Third, there may be differential dispersal out of groups of individuals with lower genetic diversity, leaving adults with higher than random genetic diversity within social groups. Fourth, unpaired individuals within an existing group may pair up based on good and/or compatible genes. This last mechanism is a way in which a natal crevice may be inherited which may be of importance for *E. stokesii* given lizard and social group numbers are limited by habitat (Gardner *et al.*, 2007).

Egernia stokesii are long lived and social groups are stable therefore, if mate choice predominantly occurs after groups have formed, individuals are likely to have considerable time to become familiar with each-other. Familiarity is likely to facilitate the ability to observe phenotypes and potentially detect genetic quality, for example via odour cues. On the other hand, *E. stokesii* individuals may not be making a choice of partner. Copulations have not been observed in the field and it is unknown if males monopolise or defend females. However, females need to lift their tail for copulation which suggests some level of 'choice', as they could retreat and wedge themselves in crevices which would prohibit mating. There is no clear male or female biased dispersal in this species (Gardner *et al.*, 2012) so both males and females may move and attempt to establish in existing groups if mating opportunities in the natal group are limited. Despite uncertainty regarding *E. stokesii* group formation, our results highlight the importance of incorporating non-genetic and genetic variables when investigating mate choice yet this is rarely done (but see Sin *et al.*, 2015). Because group formation is a fundamental precursor in the evolution of stable social aggregations, a greater understanding of *E. stokesii* group establishment and maintenance, and how mate choice influences group dynamics, warrants further investigation.

The evidence for *E. stokesii* mate choice based on genetic factors as well as group membership gives rise to numerous questions for further study. First, how are *E. stokesii* groups formed and maintained and at what stage are mate decisions made? Second, if *E. stokesii* choose a mate based on 'good' or 'compatible' genes, how is this differentiated? What are the relative roles of phenotypic and odour cues? Future work would benefit from greater sample sizes for increased power to detect mate choice signals and to facilitate investigations of decisions behind extra pair matings. Group living is rare among squamates. This study now allows comparisons among members of the *Egernia* group and related clades

that represent a range of social structures, thereby enhancing our understanding of the dynamics of group formation and the evolution of sociality within all vertebrates.

Funding

This work was supported by The Field Naturalists Society of South Australia and the Sir Mark Mitchell Research Foundation.

Acknowledgements

We thank Glen Duffield, Tessa Bradford and Mina Ansari. Field survey methods were approved by the Flinders University Animal Welfare Committee (project number E324), using standards and procedures in compliance with the Australian Code of Practice for the use of animals for scientific purposes. Laboratory work was undertaken at the South Australian Research Facility for Molecular Ecology and Evolution, Adelaide.

Figures and tables

Table 1: Summary of data used in *Egernia stokesii* mate choice analysis. Two datasets were used: the MHC I dataset included individuals genotyped for MHC I and six microsatellite loci, the MHC II dataset included individuals genotyped for MHC II and six microsatellite loci.

Dataset	MHC I dataset	MHC II dataset
Sequence length - nucleotides	216	102
Sequence length - codons	72	34
Number of variable amino acid sites (% total sites)	60 (83%)	16 (47%)
Number of peptide binding regions	13	7
Number of positively selected sites	4	0
Number of alleles	31	5
Mean alleles per individual	9.10 (SE±0.43)	2.24 (SE±0.13)
Minimum alleles per individual	2	1
Maximum alleles per individual	16	4
Mean alleles per female	9.09 (SE±0.79)	2.43 (SE±0.16)
Mean alleles per male	9.12 (SE±0.43)	2.08 (SE±0.19)
Total number of adults sampled	67	67
Total number of adults genotyped (% total adults)	47 (70%)	46 (69%)
Total number of adult males genotyped	25	25
Total number of adult females genotyped	22	21
Total number of breeding pairs identified	27	27
Number of breeding pairs included in analysis (% total breeding pairs)	17 (63%)	13 (48%)

Number of within group breeding pairs	11	9
Number of extra group breeding pairs	6	4
Female choice number of pairwise comparisons - population analysis	350	286
Male choice number of pairwise comparisons - population analysis	300	273
Female choice number of pairwise comparisons - social group analysis	22	18
Male choice number of pairwise comparisons - social group analysis	14	12

Table 2: Population level results of Generalized Linear Mixed Models investigating *good genes as heterozygosity* (GG), *compatible genes* (CG) and *group membership* hypotheses in relation to the likelihood of *Egernia stokesii* pairings. Only models within a 95% confidence set (a cumulative Akaike weight of ≥ 95%) are shown.

Rank	Model	k	AIC_c	ΔAIC_c	ω
	Female choice dataset for individuals genotyped for MHC I				
1	**GG: Group Membership + Male MHC I Functional distance**	**7**	**95.77**	**0.00**	**0.261**
2	**GG: Group Membership + Male MHC I Alleles**	**7**	**96.41**	**0.64**	**0.189**
3	**Group Membership**	**6**	**97.11**	**1.34**	**0.133**
4	**CG - Group Membership + Pairwise relatedness**	**7**	**97.28**	**1.52**	**0.122**
5	GG: Group Membership + Male MHC I Amino acid distance	7	98.82	3.05	0.057
6	CG: Group Membership + MHC I Alleles shared	7	99.18	3.42	0.047
7	GG: Group Membership + Male HL	7	99.19	3.42	0.047
8	CG: Group Membership + MHC I Amino acid pairwise distance	7	99.19	3.42	0.047

9	CG: Group Membership + Male MHC I Functional Distance x Female MHC I Functional Distance	9	99.91	4.14	0.033
10	CG: Group Membership + Male HL x Female HL	9	100.36	4.60	0.026
	Female choice dataset for individuals genotyped for MHC II				
1	**CG: Group Membership + Pairwise relatedness**	**7**	**80.50**	**0.00**	**0.478**
2	Group Membership	6	83.13	2.63	0.128
3	CG: Group Membership + MHC II Alleles shared	7	83.81	3.30	0.092
4	GG: Group Membership + Male HL	7	84.28	3.78	0.072
5	GG: Group Membership + Male MHC II Alleles	7	84.58	4.08	0.062
6	GG: Group Membership + Male MHC II Amino acid distance	7	84.88	4.37	0.054
7	CG: Group Membership + MHC II Amino acid pairwise distance	7	85.19	4.69	0.046
8	CG: Group Membership + Male MHC II Alleles x Female MHC II Alleles	9	85.85	5.34	0.033
	Male choice dataset for individuals genotyped for MHC I				
1	**GG: Group Membership + Female HL**	**7**	**89.94**	**0.00**	**0.409**
2	Group Membership	6	92.12	2.18	0.137
3	GG: Group Membership + Female MHC I Amino acid distance	7	93.11	3.17	0.084

4	CG: Group Membership + Female HL x Male HL	9	93.25	3.31	0.078
5	CG: Group Membership + MHC I Amino acid pairwise distance	7	93.79	3.85	0.060
6	CG: Group Membership + MHC I Alleles shared	7	93.90	3.96	0.056
7	GG : Group Membership + Female MHC I Functional Distance	7	94.01	4.07	0.053
8	GG: Group Membership + Female MHC I Alleles	7	94.22	4.28	0.048
	Male choice dataset for individuals genotyped for MHC II				
1	**GG: GroupMembership + Female HL**	**7**	**73.86**	**0.00**	**0.407**
2	CG: Group Membership + Pairwise relatedness	7	75.89	2.03	0.148
3	Group Membership	6	76.29	2.44	0.120
4	CG: Group Membership + Female HL x Male HL	9	76.67	2.82	0.100
5	GG: Group Membership + Female MHC II Amino acid distance	7	77.25	3.39	0.075
6	CG: Group Membership + MHC II Alleles shared	7	78.21	4.35	0.046
7	CG: Group Membership + MHC II Amino acid pairwise distance	7	78.25	4.40	0.045
8	GG: Group Membership + Female MHC II Alleles	7	78.30	4.44	0.044

GG: *good genes as heterozygosity* measures; CG: *compatible genes* measures; HL: microsatellite homozygosity by locus; *k:* number of parameters; ΔAICc: increase in AICc compared with the top model; ω: model weight. Models shown in bold are the most parsimonious models based on ΔAICc ≤ 2

Table 3: Social group level results of Generalized Linear Mixed Models investigating *good genes as heterozygosity* (GG) and *compatible genes* (CG) in relation to the likelihood of *Egernia stokesii* pairings. Only models within a 95% confidence set (a cumulative Akaike weight of ≥ 95%) are shown.

Rank	Model	*k*	AIC_c	ΔAIC_c	ω
	Female choice dataset for individuals genotyped for MHC I				
1	**GG: Male MHC1 Functional Distance**	**6**	**37.25**	**0.00**	**0.353**
2	**Null model**	**5**	**37.65**	**0.40**	**0.289**
3	CG: Pairwise relatedness	6	39.39	2.13	0.122
4	CG: MHC1 Alleles shared	6	41.17	3.92	0.050
5	GG: Male MHC1 Alleles	6	41.40	4.15	0.044
6	GG: Male MHC1 Amino acid distance	6	41.48	4.22	0.043
7	CG: MHC1 Amino acid pairwise distance	6	41.49	4.24	0.042
8	GG: Male HL	6	41.86	4.60	0.035
	Female choice dataset for individuals genotyped for MHC II				

1	**Null model**	**5**	**28.87**	**0.00**	**0.825**
2	GG: Male HL	6	33.88	5.01	0.067
3	CG: MHC2 Alleles shared	6	34.47	5.60	0.050
4	CG: Pairwise relatedness	6	34.77	5.90	0.043
	Male choice dataset for individuals genotyped for MHC I				
1	**CG: MHC1 Alleles shared**	**6**	**37.61**	**0.00**	**0.354**
2	**Null model**	**5**	**37.89**	**0.29**	**0.307**
3	GG: Female HL	6	40.18	2.57	0.098
4	GG: Female MHC1 Amino acid distance	6	40.22	2.61	0.096
5	CG: MHC1 Amino acid pairwise distance	6	41.40	3.79	0.053
6	GG: Female MHC1 Functional Distance	6	42.45	4.84	0.031
7	GG: Female MHC1 Alleles	6	42.52	4.91	0.030
	Male choice dataset for individuals genotyped for MHC II				
1	**Null model**	**5**	**30.81**	**0.00**	**0.830**
2	GG: Female HL	6	35.02	4.21	0.101
3	CG: Pairwise relatedness	6	37.09	6.27	0.036

GG: *good genes as heterozygosity* measures; CG: *compatible genes* measures; HL: homozygosity by locus; k: number of parameters; ΔAICc: increase in AICc compared with the top model; ω: model weight. Models shown in bold are the most parsimonious models based on ΔAICc ≤ 2

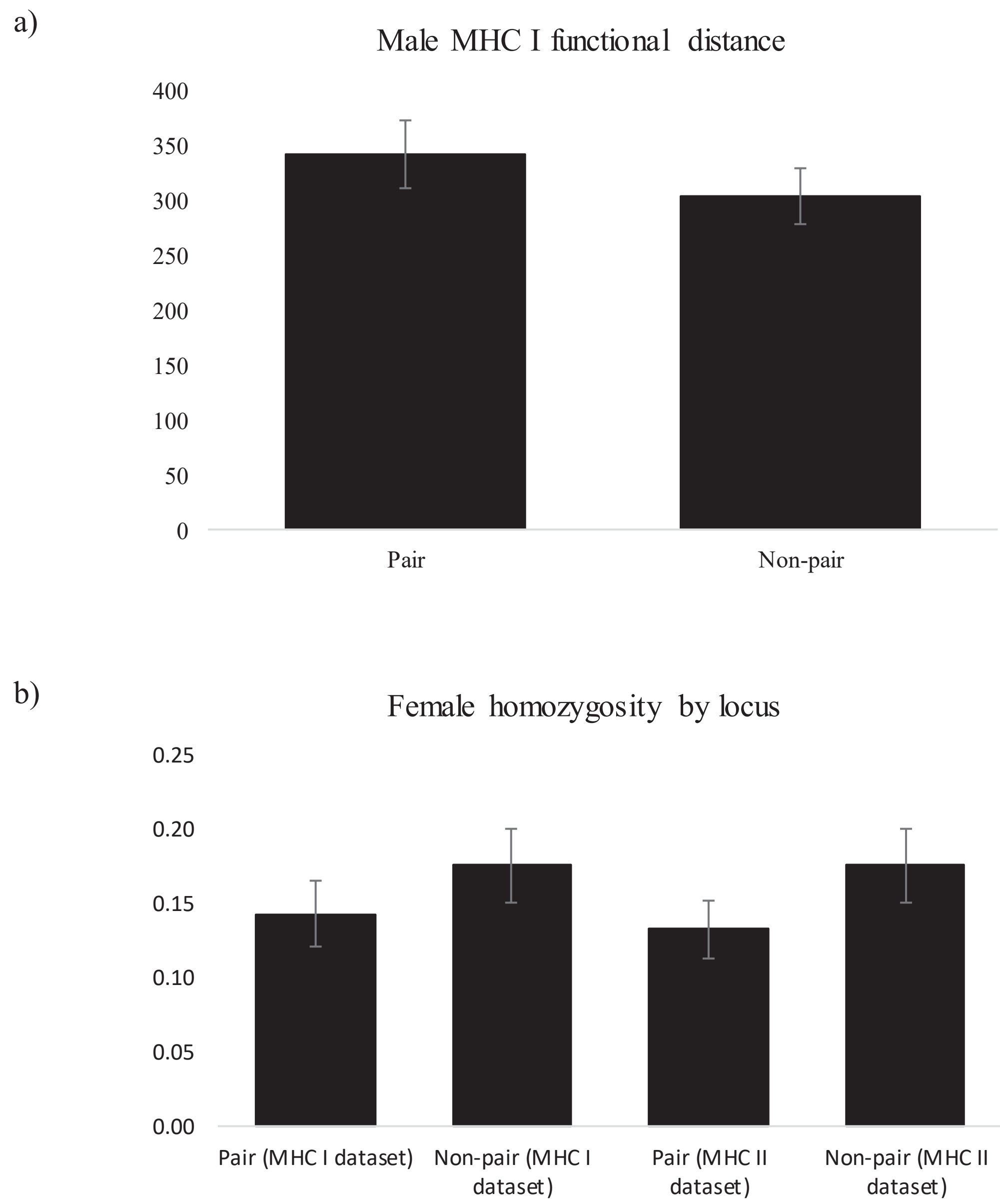


Figure 1: Bar graphs showing mean values for a) MHC I functional distance of positively selected codons for paired and non-paired males; and b) microsatellite homozygosity by locus for paired and non-paired females, identified by Generalized Linear Mixed Models as potential predictors of *Egernia stokesii* pairings. Bars represent standard errors.

References

Aeschlimann, P., Häberli, M., Reusch, T., Boehm, T. & Milinski, M. 2003. Female sticklebacks *Gasterosteus aculeatus* use self-reference to optimize MHC allele number during mate selection. *Behav. Ecol. Sociobiol.* **54:** 119-126.

Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. *Second International Symposium on Information Theory*. B. N. Petrov & F. Caski, Akademinai Kiado**:** 267-281.

Andersson, M. & Simmons, L.W. 2006. Sexual selection and mate choice. *Trends Ecol. Evol.* **21:** 296-302.

Aparicio, J.M., Ortego, J. & Cordero, P.J. 2006. What should we weigh to estimate heterozygosity, alleles or loci? *Mol. Ecol.* **15:** 4659-4665.

Bahr, A., Sommer, S., Mattle, B. & Wilson, A.B. 2012. Mutual mate choice in the potbellied seahorse (*Hippocampus abdominalis*). *Behav. Ecol.***:** ars045.

Baker, C.S. 2013. Journal of heredity adopts joint data archiving policy. *J. Hered.* **104:** 1-1.

Bates, D., Mächler, M., Bolker, B. & Walker, S. 2014. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**.

Bichet, C., Penn, D.J., Moodley, Y., Dunoyer, L., Cellier-Holzem, E., Belvalette, M.*, et al.* 2014. Females tend to prefer genetically similar mates in an island population of house sparrows. *BMC Evol. Biol.* **14:** 1.

Bisazza, A., Vaccari, G. & Pilastro, A. 2001. Female mate choice in a mating system dominated by male sexual coercion. *Behav. Ecol.* **12:** 59-64.

Bonneaud, C., Pérez-Tris, J., Federici, P., Chastel, O. & Sorci, G. 2006. Major histocompatibility alleles associated with local resistance to malaria in a passerine. *Evolution* **60:** 383-389.

Brown, J.L. & Eklund, A. 1994. Kin recognition and the major histocompatibility complex: an integrative review. *Am. Nat.***:** 435-461.

Brown, J.L. 1997. A theory of mate choice based on heterozygosity. *Behav. Ecol.* **8:** 60-65.

Bull, C.M., Griffith, S.C., Lanham, E.J. & Johnston, G.R. 2000. Recognition of pheromones from group members in a gregarious lizard, *Egernia stokesii*. *J. Herpetol.* **34:** 92-99.

Burnham, K.P. & Anderson, D.R. 2001. Kullback-Leibler information as a basis for strong inference in ecological studies. *Wildl. Res.* **28:** 111-119.

Clutton-Brock, T.H. & Parker, G.A. 1995. Sexual coercion in animal societies. *Anim. Behav.* **49:** 1345-1365.

Cogger, H.G. 1983. *Reptiles and amphibians of Australia*. AH & AW Reed Pty Ltd.

Cutrera, A.P., Fanjul, M.S. & Zenuto, R.R. 2012. Females prefer good genes: MHC-associated mate choice in wild and captive tuco-tucos. *Anim. Behav.* **83:** 847-856.

Duffield, G.A. & Bull, C.M. 1996. Characteristics of the litter of the gidgee skink, *Egernia stokesii*. *Wildl. Res.* **23:** 337-342.

Duffield, G.A. & Bull, C.M. 2002. Stable aggregations in an Australian lizard, *Egernia stokesii*. *Naturwissenschaften* **89:** 424-427.

Edward, D.A. & Chapman, T. 2011. The evolution and significance of male mate choice. *Trends Ecol. Evol.* **26:** 647-654.

Edward, D.A. 2015. The description of mate choice. *Behav. Ecol.* **26:** 301-310.

Galaverni, M., Caniglia, R., Milanesi, P., Lapalombella, S., Fabbri, E. & Randi, E. 2015. Choosy Wolves? Heterozygote Advantage But No Evidence of MHC-Based Disassortative Mating. *J. Hered.***:** esv090.

Gardner, M.G. 2000. A genetic investigation of sociality in the Australian group living lizard *Egernia stokesii*. *School of Biological Sciences*, Flinders University of South Australia. Adelaide. **Doctoral dissertation**.

Gardner, M.G., Bull, C.M., Cooper, J.B. & Duffied, G.A. 2001. Genetic evidence for a family structure in stable social aggregations of the Australian lizard *Egernia stokesii*. *Mol. Ecol.* **10:** 175-183.

Gardner, M.G., Bull, C.M. & Cooper, S.J.B. 2002. High levels of genetic monogamy in the group-living Australian lizard *Egernia stokesii*. *Mol. Ecol.* **11:** 1787-1794.

Gardner, M.G., Bull, C.M., Fenner, A., Murray, K. & Donnellan, S.C. 2007. Consistent social structure within aggregations of the Australian lizard, *Egernia stokesii* across seven disconnected rocky outcrops. *J. Ethol.* **25:** 263-270.

Gardner, M.G., Godfrey, S.S., Fenner, A.L., Donnellan, S.C. & Bull, C.M. 2012. Fine-scale spatial structuring as an inbreeding avoidance mechanism in the social skink *Egernia stokesii*. *Aust. J. Zool.* **60:** 272-277.

Gardner, M.G., Pearson, S.K., Johnston, G.R. & Schwarz, M.P. 2015. Group living in squamate reptiles: a review for stable aggregations. *Biological Reviews*.

Gillingham, M.A., Richardson, D.S., Løvlie, H., Moynihan, A., Worley, K. & Pizzari, T. 2009. Cryptic preference for MHC-dissimilar females in male red junglefowl, *Gallus gallus*. *Proc. R. Soc. Lond., Ser. B: Biol. Sci.* **276:** 1083-1092.

Godfrey, S.S., Bull, C.M., Murray, K. & Gardner, M.G. 2006. Transmission mode and distribution of parasites among groups of the social lizard *Egernia stokesii*. *Parasitol. Res.* **99:** 223-230.

Godfrey, S.S., Bull, C.M., James, R. & Murray, K. 2009. Network structure and parasite transmission in a group living lizard, the gidgee skink, *Egernia stokesii*. *Behav. Ecol. Sociobiol.* **63:** 1045-1056.

Gohli, J., Anmarkrud, J.A., Johnsen, A., Kleven, O., Borge, T. & Lifjeld, J.T. 2013. Female promiscuity is positively associated with neutral and selected genetic diversity in passerine birds. *Evolution* **67:** 1406-1419.

Huchard, E., Knapp, L.A., Wang, J., Raymond, M. & Cowlishaw, G. 2010. MHC, mate choice and heterozygote advantage in a wild social primate. *Mol. Ecol.* **19:** 2545-2561.

Huchard, E., Baniel, A., Schliehe-diecks, S. & Kappeler, P.M. 2013. MHC -disassortative mate choice and inbreeding avoidance in a solitary primate. *Mol. Ecol.* **22:** 4071-4086.

Huchard, E. & Pechouskova, E. 2013. The major histocompatibility complex and primate behavioral ecology: new tools and future questions. *Int. J. Parasitol.* **35:** 11-31.

Hurvich, C.M. & Tsai, C.-L. 1989. Regression and time series model selection in small samples. *Biometrika* **76:** 297-307.

Kempenaers, B. 2007. Mate choice and genetic quality: a review of the heterozygosity theory. *Adv. Stud. Behav.* **37:** 189-278.

Kokko, H., Brooks, R., Jennions, M.D. & Morley, J. 2003. The evolution of mate choice and mating biases. *Proc. R. Soc. Lond., Ser. B: Biol. Sci.* **270:** 653-664.

Kuduk, K., Babik, W., Bellemain, E., Valentini, A., Zedrosser, A., Taberlet, P*., et al.* 2014. No evidence for the effect of MHC on male mating success in the brown bear. *PloS one* **9:** e113414.

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*). *Proc. R. Soc. Lond., Ser. B: Biol. Sci.* **268:** 1279-1285.

Lanham, E.J. & Bull, C.M. 2004. Enhanced vigilance in groups in *Egernia stokesii*, a lizard with stable social aggregations. *J. Zool. (Lond).* **263:** 95-99.

Main, A.R. & Bull, C.M. 1996. Mother-offspring recognition in two Australian lizards, *Tiliqua rugosa* and *Egernia stokesii*. *Anim. Behav.* **52:** 193-200.

Marshall, T.C., Slate, J., Kruuk, L.E.B. & Pemberton, J.M. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. *Mol. Ecol.* **7:** 639-655.

Mazerolle, M.J. 2015. Model selection and multimodel inference based on (Q)AIC(c), R Foundation for Statistical Computing, Vienna.

Milinski, M. 2006. The major histocompatibility complex, sexual selection, and mate choice. *Annu. Rev. Ecol., Evol. Syst.* **37:** 159-186.

Miller, H.C., Moore, J.A., Nelson, N.J. & Daugherty, C.H. 2009. Influence of major histocompatability complex genotype on mating success in a free-ranging reptile population. *Proceedings of the Royal Society B, Biological Sciences* **276:** 1695-1704.

Neff, B.D. 2004. Stabilizing selection on genomic divergence in a wild fish population. *Proc. Natl. Acad. Sci. USA* **101:** 2381-2385.

Neff, B.D. & Pitcher, T.E. 2005. Genetic quality and sexual selection: an integrated framework for good genes and compatible genes. *Mol. Ecol.* **14:** 19-38.

Nowak, M.A., Tarczy-Hornoch, K. & Austyn, J.M. 1992. The optimal number of major histocompatibility complex molecules in an individual. *Proc. Natl. Acad. Sci. USA* **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* **270:** S254-256.

Penn, D.J. & Potts, W.K. 1999. The evolution of mating preferences and major histocompatibility complex genes. *Am. Nat.* **153:** 145-164.

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

Promerova, M., Vinkler, M., Bryja, J., Polakova, R., Schnitzer, J., Munclinger, P.*, et al.* 2011. Occurrence of extra-pair paternity is connected to social male's MHC-variability in the scarlet rosefinch *Carpodacus erythrinus*. *J. Avian Biol.* **42:** 5-10.

Puurtinen, M., Ketola, T. & Kotiaho, J. 2009. The Good-genes and compatible-genes benefits of mate choice. *Am. Nat.* **174:** 741-752.

Radwan, J., Zagalska-Neubauer, M., Cichon, M., Sendecka, J., Kulma, K., Gustafsson, L.*, et al.* 2012. MHC diversity, malaria and lifetime reproductive success in collared flycatchers. *Mol. Ecol.* **21:** 2469-2479.

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:** 300-302.

Richardson, D.S., Komdeur, J., Burke, T. & Von Schantz, T. 2005. MHC-based patterns of social and extra-pair mate choice in the Seychelles warbler. *Proc. R. Soc. Lond., Ser. B: Biol. Sci.* **272:** 759-767.

Schwensow, N., Eberle, M. & Sommer, S. 2008a. Compatibility counts: MHC-associated mate choice in a wild promiscuous primate. *Proc. R. Soc. Lond., Ser. B: Biol. Sci.* **275:** 555-564.

Schwensow, N., Fietz, J., Dausmann, K. & Sommer, S. 2008b. MHC-associated mating strategies and the importance of overall genetic diversity in an obligate pair-living primate. *Evol. Ecol.* **22:** 617-636.

Sepil, I., Radersma, R., Santure, A.W., De Cauwer, I., Slate, J. & Sheldon, B.C. 2015. No evidence for MHC class I based disassortative mating in a wild population of great tits. *J. Evol. Biol.* **28:** 642-654.

Setchell, J.M. & Huchard, E. 2010. The hidden benefits of sex: Evidence for MHC-associated mate choice in primate societies. *Bioessays* **32:** 940-948.

Sin, Y.W., Annavi, G., Newman, C., Buesching, C., Burke, T., Macdonald, D.W.*, et al.* 2015. MHC class II-assortative mate choice in European badgers (Meles meles). *Mol. Ecol.* **24:** 3138-3150.

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

Sommer, S. 2005b. Major histocompatibility complex and mate choice in a monogamous rodent. *Behav. Ecol. Sociobiol.* **58:** 181-189.

Strandh, M., Westerdahl, H., Pontarp, M., Canback, B., Dubois, M.-P., Miquel, C.*, et al.* 2012. Major histocompatibility complex class II compatibility, but not class I, predicts mate choice in a bird with highly developed olfaction. *Proc. R. Soc. Lond., Ser. B: Biol. Sci.* **279:** 4457-4463.

Tregenza, T. & Wedell, N. 2000. Genetic compatibility, mate choice and patterns of parentage: invited review. *Mol. Ecol.* **9:** 1013-1027.

Tung, J., Charpentier, M.J., Mukherjee, S., Altmann, J. & Alberts, S.C. 2012. Genetic effects on mating success and partner choice in a social mammal. *Am. Nat.* **180:** 113.

Wetton, J.H., Carter, R.E., Parkin, D.T. & Walters, D. 1987. Demographic study of a wild house sparrow population by DNA fingerprinting. *Nature* **327:** 147-149.

Winternitz, J., Promerova, M., Polakova, R., Vinker, M., Schnitzer, J., Munclinger, P.*, et al.* 2015. Effects of heterozygosity and MHC diversity on patterns of extra-pair paternity in the socially monogamous scarlet rosefinch. *Behav. Ecol. Sociobiol.* **69:** 459-469.

Winternitz, J.C. 2015. Examining the evidence for major histocompatibility complex-dependent mate selection in humans and nonhuman primates. *Res. Rep. Biol.* **6:** 73-88.

Supplementary Material

S1: Field survey and laboratory method and results detail

S2: Model names and outputs for Generalized Linear Mixed Models of *Egernia stokesii* mate choice for eight datasets (female and male choice for MHC I/microsatellites and MHC II/microsatellites, population and social group level)

S1: Field survey and laboratory method and results detail

Method

Field surveys

We used data derived from field surveys conducted during lizard activity seasons (September – March) between 1994 and 1998 (Duffield & Bull, 2002). Sampling methods are described in Gardner *et al.* (2001). Briefly, lizards were captured in Elliot traps placed close to refuge crevices, or by hand extraction from a crevice, they were marked by toe clipping, and their capture location and association with other individuals recorded (Duffield & Bull, 2002). Lizard capture location was recorded based on the grid cell within a 10 x 10 m grid across the study site, where the occupied crevice was located. Individuals were assigned to a social group if they shared crevice use with another member of the group on 80% or more of observations (Duffield & Bull, 2002).

Microsatellite data

During surveys, DNA was sampled from 152 lizards. Blood (up to 0.5 mL) was taken from the caudal vein of each lizard, transferred to tubes containing ethanol/saline solution (50%:0.85% v/v) and stored at room temperature. Individuals were previously genotyped (Gardner *et al.*, 2001; Gardner *et al.*, 2002) for five *E. stokesii* specific microsatellite loci (Est 1, 2, 3, 4, and 8; Gardner *et al.*, 1999) and one microsatellite locus previously isolated from *Tiliqua rugosa* (Tr 3.2; Cooper *et al.*, 1997), another member of the *Egernia* group. All loci were retained as they showed no linkage disequilibrium, deviations from Hardy-Weinberg equilibrium, or incidences of null alleles and large allele dropout (Gardner, 2000).

MHC data

Methods of DNA extraction and MHC amplification, sequencing, allele identification, and genotyping are described in Pearson *et al.* (Chapter 3). In *E. stokesii*, the MHC I region is

more variable than MHC II and there is evidence for historical and contemporary positive selection on MHC I but not MHC II alleles (Pearson *et al.* Chapter 4). Four positively selected amino acid sites have been detected in the 216 base pair region of MHC I utilised here (Pearson *et al.* Chapter 4). Although equivalent positive selection on MHC II has not yet been detected in *E. stokesii* this region is also used in this study because MHC II proteins influence individual odour profiles in some species that use odour cues in mate choice (Wedekind *et al.*, 1995; Milinski, 2006; Strandh *et al.*, 2012).

Measures used to test the good genes as heterozygosity *hypothesis*

Four within individual measures were used to assess genetic diversity in the context of the *good genes as heterozygosity* hypothesis: 1) number of MHC alleles, 2) mean within individual MHC genotypic distance (Schwensow *et al.*, 2008), 3) within individual MHC functional distance (Radwan *et al.*, 2012), and 4) individual microsatellite homozygosity by locus (Aparicio *et al.*, 2006). *First* we recorded the number of MHC alleles in each individual. *Second*, we calculated the mean within individual genotypic distance, based on the mean pairwise amino acid distances between all MHC alleles in an individual (Schwensow et al 2008), using the p-distance method, the proportion of nucleotides that are different between two sequences compared, with 1000 bootstraps in MEGA6 (Tamura et al 2013). Mean individual genotypic distance was calculated based on the four positively selected site sequences for MHC I, and for the entire sequence for MHC II. For our *third* measure, we calculated within individual MHC functional distance based on the physiochemical properties of amino acids (Schwensow *et al.*, 2007; Agbali *et al.*, 2010). Due to low variability and lack of positively selected sites in MHC II, functional distance analysis was performed for MHC I only. Each amino acid within the four MHC I positively selected sites was described using five z-descriptors: z_1 (hydrophobicity), z_2 (steric bulk), z_3 (polarity), and z_4 and z_5 (electronic effects) (Sandberg *et al.*, 1998). Alleles could not be grouped into

functional units or 'supertypes' (Doytchinova & Flower, 2005; Schwensow *et al.*, 2007) (Pearson Appendix 7) therefore instead of the supertype approach we followed the method of Radwan *et al.* (2012) and calculated within individual MHC I functional distance using the sum of pairwise Euclidean distances between z-descriptors of all alleles within an individual. For our *fourth* measure, we used individual microsatellite homozygosity by locus (Aparicio *et al.*, 2006), calculated in R as implemented in Genhet (Coulon, 2010), using the six microsatellite loci described above. Values of homozygosity by locus range from 0 in individuals when all loci are heterozygous, to 1 when all loci in an individual lizard are homozygous (Aparicio *et al.*, 2006). The lower the homozygosity by locus value, the higher the individual microsatellite heterozygosity (Aparicio *et al.*, 2006).

Measures used to test the compatible genes *hypothesis*

Three pairwise measures were used to assess the prediction that *E. stokesii* mate choice for *compatible genes* would be demonstrated by a preference for a genetically dissimilar mate: 1) proportion of shared MHC alleles (Wetton *et al.*, 1987), 2) pairwise MHC genotypic distance (following Landry *et al.*, 2001), and 3) pairwise microsatellite relatedness (Richardson *et al.*, 2005). *First*, we calculated the proportion of shared MHC alleles for each pair as twice the sum of the number of alleles shared by the pair divided by the total number of alleles present in both individuals in the pair, following Wetton *et al.* (1987) and using GELSTATS v 2.61 (Rogstad & Pelikan, 1996). Therefore, $D = 2F_{AB}/(F_A + F_B)$ where F_{AB} is the number of shared MHC alleles between A and B in a pair and F_A and F_B are the number of MHC alleles in individuals A and B respectively (Bonneaud *et al.*, 2006; Bichet *et al.*, 2014). The proportion of shared alleles ranges from 0 for pairs that share no MHC alleles (most dissimilar) to 1 for pairs that share identical MHC alleles (most similar), with intermediate values between 0 and 1 for pairs sharing at least one MHC allele (Baratti *et al.*, 2012). *Second*, we calculated pairwise genotypic distance based on the average of all mean pairwise amino acid distances

of all MHC alleles carried by the two individuals (following Landry *et al.*, 2001) using MEGA6, calculated only for positively selected site sequences for MHC I, and for the entire sequence for MHC II. Dissimilarity is indicated by higher genotypic distance values. *Third*, we used genotypes of the six microsatellite DNA loci to estimate pairwise relatedness (Wang, 2002), with 10,000 bootstrap replicates used to estimate 95% confidence intervals, in Coancestry v 1.0.1.5 (Wang, 2011). Dissimilarity is indicated by lower pairwise microsatellite relatedness values.

In addition, we predicted that the choosing individual's genetic diversity (as defined by the four measures described in *Measures used to test the* good genes as heterozygosity *hypothesis* above) would influence their mate choice (Reusch *et al.*, 2001; Aeschlimann *et al.*, 2003). For example a focal individual with a lower number of alleles would be more likely to choose a partner with more alleles than a focal individual with a higher number of alleles. We predicted that the strength of choice for genetic diversity would be higher for less heterozygous individuals. Although this prediction utilises genetic diversity measures, it is dependent upon the genotypes of both individuals therefore it is considered here within the context of the *compatible genes* hypotheses.

Mate choice test datasets

Because historically derived field locations were based on a grid system, rather than GPS coordinates, a precise spatial proximity between candidate mates could not be calculated. Therefore we initially took a conservative approach whereby all males in the population were considered as candidate mates, even though this may have included some males that never overlapped the home range of a focal female. We then restricted candidate males to those within the same social group as the focal female. Therefore extra group pairings were excluded from social group analyses, even though they occurred quite often (Table 1). Although we would have been interested in differences between group pair and extra group

pair mates, small sample sizes prevented analysis (Table 1). Although MHC I and II genes may be in linkage disequilibrium (e.g. Sin *et al.*, 2014), the two regions differ in structure, function and expression (Hughes & Yeager, 1998) which may exert different influences on mate choice (Strandh *et al.*, 2012; Sin *et al.*, 2014). Therefore MHC I and MHC II were analysed separately.

Hypothesis testing

For each possible pair-wise combination of females and males within each data set, we considered a subset that contained only females that were identified as belonging to a breeding pair and all their potential male partners. In each model, we included whether each possible pair had or had not mated as a binomial response variable (0 = not mated, 1 = mated), and male and female ID (both nested within group ID) as random effects, to account for the repeated inclusion of both individual males and females as mates and potential mates in these pair-wise comparisons. In each model, we included a fixed effect that represented each of our alternative hypotheses about mate choice in this system. For *compatible genes* models we included an additional set of models to assess interactions between a female's own quality and that of the potential male to incorporate the influence of female 'quality' on their choice of mate. To test our third hypothesis, with whole population level data sets, a model including only group membership (0 = different group; 1 = same group) was included in the model set. During initial model trials, group membership was identified as a strong predictor of mate choice, therefore, subsequently, for each of our models (based on *good genes as heterozygosity* or *compatible genes* hypotheses), group membership was included as an additional covariate to account for the influence of this on mate choice when assessing the influence of the gene-based hypotheses. Our model sets also included a null model, with no predictor variables.

For each model set, models were ranked using AIC_c, and ΔAIC_c (the difference between the top model and all subsequent models) to identify a candidate set of models and calculate Akaike weights (ω). An Akaike weight may be considered as the probability that a model is the best approximating model (but see Richards, 2005). For example if a model has an Akaike weight of 0.305 it means there is a 30.5% chance that it is the best approximating model for the data, relative to all models considered in the model set. Although the model with the smallest AICc may be considered the most plausible (Burnham *et al.*, 2011), models with $\Delta AICc < 2$ can be considered to have substantial support (Burnham, 2002). We utilised a 95 % confidence set which included all models with a cumulative Akaike weight of ≥ 95 % in a review of the best approximating models (Burnham, 2002; Symonds & Moussalli, 2011). To further assess model strength, evidence ratios were calculated (results not shown), where the Akaike weight of the top ranking model (e.g. H1) was compared to the Akaike weight of a second candidate model (e.g. H2) using H1/H2 (Burnham *et al.*, 2011).

S2: Model names and outputs for Generalized Linear Mixed Models of *Egernia stokesii* mate choice for eight datasets (female and male choice for MHC I/microsatellites and MHC II/microsatellites, population and social group level)

Table 1: Results of population level *Egernia stokesii* Generalized Linear Mixed Models assessing mate choice in an MHC I/microsatellite dataset. Showing model coefficients, standard errors (SE), and upper and lower confidence limits (CL) of 95% confidence intervals. Confidence intervals that don't overlap zero are shown in bold.

Variable	Coefficient	SE	Lower.CL	Upper.CL
Female choice				
Intercept	-5.579	4.96	-15.30	4.14
Group membership	**4.348**	**0.68**	**3.01**	**5.69**
Male MHC I Alleles	0.267	0.17	-0.06	0.59
Male MHC I Amino acid distance	3.318	5.51	-7.49	14.13
Male MHC I Functional distance	**0.047**	**0.03**	**0.00**	**0.10**
Male microsatellite homozygosity by locus	-0.055	2.61	-5.17	5.06
MHC I Alleles shared	0.160	1.75	-3.27	3.59
Male MHC I Amino Acid - Pairwise Distance	0.009	5.56	-10.88	10.90
Pairwise microsatellite relatedness	-2.003	1.48	-4.90	0.89
Male choice				
Intercept	-4.143	4.31	-12.58	4.29
Group membership	**4.324**	**0.73**	**2.90**	**5.75**
Female MHC I Alleles	-0.003	0.09	-0.17	0.17
Female MHC I Amino acid distance	4.630	4.51	-4.21	13.47
Female MHC I Functional distance	0.006	0.01	-0.02	0.03
Female microsatellite homozygosity by locus	-6.083	3.15	-12.25	0.09
MHC I Alleles shared	-1.117	1.99	-5.01	2.78
MHC I Amino acid - pairwise distance	4.506	6.91	-9.03	18.04
Pairwise microsatellite relatedness	**-0.028**	**0.01**	**-0.05**	**-0.01**

Table 2: Results of population level *Egernia stokesii* Generalized Linear Mixed Models assessing mate choice in an MHC II/microsatellite dataset. Showing model coefficients, standard errors (SE), and upper and lower confidence limits (CL) of 95% confidence intervals. Confidence intervals that don't overlap zero are shown in bold.

Variable	Coefficient	SE	Lower.CL	Upper.CL
Female choice				
Intercept	-5.042	1.60	-8.18	-1.90
Group membership	**5.032**	**1.16**	**2.77**	**7.30**
Male MHC II Alleles	0.332	0.41	-0.47	1.13
Male MHC II Amino acid distance	2.917	4.68	-6.25	12.09
Male microsatellite homozygosity by locus	3.269	3.40	-3.39	9.93
MHC II Alleles shared	2.729	2.38	-1.93	7.39
MHC II Amino acids - pairwise distance	2.306	10.68	-18.63	23.24
Pairwise microsatellite relatedness	-4.765	2.65	-9.96	0.43
Male choice				
Intercept	-4.181	1.54	-7.20	-1.17
Group membership	**5.227**	**1.11**	**3.05**	**7.40**
Female MHC II Alleles	0.235	0.74	-1.22	1.69
Female MHC II Amino acid distance	17.185	16.48	-15.12	49.49
Female microsatellite homozygosity by locus	-9.255	4.77	-18.61	0.10
MHC II Alleles shared	1.018	2.37	-3.63	5.67
MHC II Amino acids - pairwise distance	3.915	9.79	-15.26	23.09
Pairwise microsatellite relatedness	-4.670	3.28	-11.09	1.75

Table 3: Results of social group level *Egernia stokesii* Generalized Linear Mixed Models assessing mate choice in an MHC I/microsatellite dataset. Showing model coefficients and upper and lower confidence limits (CL) of 95% confidence intervals. Confidence intervals that don't overlap zero are shown in bold.

Variable	Coefficient	LowerCI	UpperCI
Female choice			
Intercept	10.520	2.92	18.12
Male MHC I Alleles	0.333	-1.75	2.42
Male MHC I Amino acid distance	5.360	-51.56	62.28
Male Functional distance	**2.193**	**1.27**	**3.11**
Male microsatellite homozygosity by locus	**-53.323**	**-104.24**	**-2.40**
Male MHC I Alleles shared	**49.262**	**0.13**	**98.40**
Male MHC I Amino acid pairwise distance	-3.835	-70.36	62.69
Pairwise microsatellite relatedness	-55.253	-114.53	4.02
n=22; I8 Males in I2 groups and I2 females in I0 groups			
Male choice			
Intercept	0.751	-0.61	2.11
Female MHC I Alleles	-0.017	-0.29	0.26
Female MHC I Amino acid distance	10.690	-6.91	28.29
Female Functional distance	0.006	-0.04	0.05
Female microsatellite homozygosity by locus	-9.097	-24.95	6.76
Female MHC I Alleles shared	-5.855	-42.19	30.48
Female MHC I Amino acid pairwise distance	12.232	-13.03	37.50
Pairwise microsatellite relatedness	0.357	-6.03	6.74
n=18; 18 females in 12 groups, 10 males in 10 groups			

Table 4: Results of social group level *Egernia stokesii* Generalized Linear Mixed Models assessing mate choice in an MHC II/microsatellite dataset. Showing model coefficients and upper and lower confidence limits (CL) of 95% confidence intervals. Confidence intervals that don't overlap zero are shown in bold. Male choice for female MHC II Alleles is not shown because the term did not converge during modelling.

Variable	Coefficient	LowerCI	UpperCI
Female choice			
Intercept	29.041	12.27	45.81
Male MHC II Alleles	1.009	-1.32	3.34
Male MHC II Amino acid distance	28.307	-52.91	109.52
Male microsatellite homozygosity by locus	**97.670**	**2.34**	**193.00**
Male MHC II Alleles shared	**42.823**	**3.69**	**81.96**
Male MHC II Amino acid pairwise distance	44.020	-49.11	137.15
Pairwise microsatellite relatedness	-8.644	-20.29	3.00
n = 14 (11 Males in 10 groups and 10 females in 9 groups)			
Male choice			
Intercept	1.609	0.09	3.13
Female MHC II Amino	-9.091	-54.97	36.79
Female microsatellite homozygosity by locus	-39.039	-139.74	61.66
Female MHC II Alleles shared	-0.651	-8.89	7.59
Female MHC II Amino acid pairwise distance	12.031	-44.25	68.31
Pairwise microsatellite relatedness	-11.512	-28.28	5.26
n=12; 11 females in 9 groups & 9 males in 9 groups			

References

Aeschlimann, P., Häberli, M., Reusch, T., Boehm, T. & Milinski, M. 2003. Female sticklebacks *Gasterosteus aculeatus* use self-reference to optimize MHC allele number during mate selection. *Behav. Ecol. Sociobiol.* **54:** 119-126.

Agbali, M., Reichard, M., Bryjová, A., Bryja, J. & Smith, C. 2010. Mate choice for nonadditive genetic benefits correlate with MHC dissimilarity in the rose bitterling (*Rhodeus ocellatus*). *Evolution* **64:** 1683-1696.

Aparicio, J.M., Ortego, J. & Cordero, P.J. 2006. What should we weigh to estimate heterozygosity, alleles or loci? *Mol. Ecol.* **15:** 4659-4665.

Baratti, M., Dessi-Fulgheri, F., Ambrosini, R., Bonisoli-Alquati, A., Caprioli, M., Goti, E.*, et al.* 2012. MHC genotype predicts mate choice in the ring-necked pheasant *Phasianus colchicus*. *J. Evol. Biol.* **25:** 1531-1542.

Bichet, C., Penn, D.J., Moodley, Y., Dunoyer, L., Cellier-Holzem, E., Belvalette, M.*, et al.* 2014. Females tend to prefer genetically similar mates in an island population of house sparrows. *BMC Evol. Biol.* **14:** 1.

Bonneaud, C., Pérez-Tris, J., Federici, P., Chastel, O. & Sorci, G. 2006. Major histocompatibility alleles associated with local resistance to malaria in a passerine. *Evolution* **60:** 383-389.

Burnham, K., Anderson, D. & Huyvaert, K. 2011. AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behav. Ecol. Sociobiol.* **65:** 23-35.

Burnham, K.P. 2002. *Model selection and multi-model inference: A practical information-theoretic approach*, 2nd ed. edn. Secaucus : Springer-Verlag New York, Incorporated, Secaucus.

Cooper, S.J.B., Bull, C.M. & Gardner, M.G. 1997. Characterization of microsatellite loci from the socially monogamous lizard *Tiliqua rugosa* using a PCR-based isolation technique. *Mol. Ecol.* **6:** 793-795.

Coulon, A. 2010. genhet: an easy-to-use R function to estimate individual heterozygosity. *Mol. Ecol. Resour.* **10:** 167-169.

Doytchinova, I.A. & Flower, D.R. 2005. In silico identification of supertypes for class II MHCs. *J. Immunol.* **174:** 7085-7095.

Duffield, G.A. & Bull, C.M. 2002. Stable aggregations in an Australian lizard, *Egernia stokesii*. *Naturwissenschaften* **89:** 424-427.

Gardner, M.G., Cooper, S.J.B., Bull, C.M. & Grant, W.N. 1999. Isolation of microsatellite loci from a social lizard, *Egernia stokesii*, using a modified enrichment procedure. *J. Hered.* **90:** 301-304.

Gardner, M.G. 2000. A genetic investigation of sociality in the Australian group living lizard *Egernia stokesii*. *School of Biological Sciences*, Flinders University of South Australia. Adelaide. **Doctoral dissertation**.

Gardner, M.G., Bull, C.M., Cooper, J.B. & Duffied, G.A. 2001. Genetic evidence for a family structure in stable social aggregations of the Australian lizard *Egernia stokesii*. *Mol. Ecol.* **10:** 175-183.

Gardner, M.G., Bull, C.M. & Cooper, S.J.B. 2002. High levels of genetic monogamy in the group-living Australian lizard *Egernia stokesii*. *Mol. Ecol.* **11:** 1787-1794.

Hughes, A.L. & Yeager, M. 1998. Natural selection at major histocompatibilty complex loci of vertebrates. *Annu. Rev. Genet.* **321:** 415-435.

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*). *Proc. R. Soc. Lond., Ser. B: Biol. Sci.* **268:** 1279-1285.

Milinski, M. 2006. The major histocompatibility complex, sexual selection, and mate choice. *Annu. Rev. Ecol., Evol. Syst.* **37:** 159-186.

Radwan, J., Zagalska-Neubauer, M., Cichon, M., Sendecka, J., Kulma, K., Gustafsson, L.*, et al.* 2012. MHC diversity, malaria and lifetime reproductive success in collared flycatchers. *Mol. Ecol.* **21:** 2469-2479.

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:** 300-302.

Richards, S.A. 2005. Testing ecological theory using the information-theoretic approach: examples and cautionary results. *Ecology* **86:** 2805-2814.

Richardson, D.S., Komdeur, J., Burke, T. & Von Schantz, T. 2005. MHC-based patterns of social and extra-pair mate choice in the Seychelles warbler. *Proc. R. Soc. Lond., Ser. B: Biol. Sci.* **272:** 759-767.

Rogstad, S.H. & Pelikan, S. 1996. GELSTATS: a computer program for population genetics analyses using VNTR multilocus probe data. *BioTechniques* **21:** 1128-1131.

Sandberg, M., Eriksson, L., Jonsson, J., Sjöström, M. & Wold, S. 1998. New chemical descriptors relevant for the design of biologically active peptides. A multivariate characterization of 87 amino acids. *J. Med. Chem.* **41:** 2481-2491.

Schwensow, N., Fietz, J., Dausmann, K.H. & Sommer, S. 2007. Neutral versus adaptive genetic variation in parasite resistance: importance of major histocompatibility complex supertypes in a free-ranging primate. *Heredity* **99:** 265-277.

Schwensow, N., Fietz, J., Dausmann, K. & Sommer, S. 2008. MHC-associated mating strategies and the importance of overall genetic diversity in an obligate pair-living primate. *Evol. Ecol.* **22:** 617-636.

Sin, Y.W., Annavi, G., Dugdale, H.L., Newman, C., Burke, T. & Macdonald, D.W. 2014. Pathogen burden, co-infection and major histocompatibility complex variability in the European badger (*Meles meles*). *Mol. Ecol.* **23:** 5072-5088.

Strandh, M., Westerdahl, H., Pontarp, M., Canback, B., Dubois, M.-P., Miquel, C.*, et al.* 2012. Major histocompatibility complex class II compatibility, but not class I, predicts mate choice in a bird with highly developed olfaction. *Proc. R. Soc. Lond., Ser. B: Biol. Sci.* **279:** 4457-4463.

Symonds, M.R. & Moussalli, A. 2011. A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. *Behav. Ecol. Sociobiol.* **65:** 13-21.

Wang, J. 2002. An estimator for pairwise relatedness using molecular markers. *Genetics* **160:** 1203-1215.

Wang, J. 2011. COANCESTRY: a program for simulating, estimating and analysing relatedness and inbreeding coefficients. *Mol. Ecol. Resour.* **11:** 141-145.

Wedekind, C., Seebeck, T., Bettens, F. & Paepke, A.J. 1995. MHC-dependent mate preference in humans. *Proc. R. Soc. Lond., Ser. B: Biol. Sci.* **260:** 245-249.

Wetton, J.H., Carter, R.E., Parkin, D.T. & Walters, D. 1987. Demographic study of a wild house sparrow population by DNA fingerprinting. *Nature* **327:** 147-149.

Chapter 7: Selection on the MHC at a fine spatial scale

In preparation as:

Pearson, S.K., Bull, C.M., and Gardner, M.G., lack of MHC differentiation within a family living lizard across geographically close but disconnected rocky outcrops.

Chapter 4 provides evidence that selection is acting on *E. stokesii* MHC alleles. *Egernia stokesii* site and social fidelity (Chapter 5) and mate choice for genetic diversity and dissimilarity (Chapter 6) may impose selective pressures on the MHC. However, MHC variation is also influenced by genetic drift and gene flow, not selection alone. To date, most studies investigating the relative influence of these processes on MHC variation have been undertaken at broad spatial scales. Chapter 7 investigates whether selection on *E. stokesii* MHC alleles can outweigh the effects of genetic drift and limited gene flow at a fine scale.

Selection outweighs drift at a fine scale: lack of MHC differentiation within a family living lizard across geographically close but disconnected rocky outcrops

Pearson, S.K.[A,B], Bull, C.M.[A], Gardner, M.G.[A,C]

[A]School of Biological Sciences, Flinders University of South Australia, Bedford Park 5042, Australia

[C]Evolutionary Biology Unit, South Australian Museum, North Terrace, Adelaide 5000, Australia.

[B]Corresponding author. Email: sarah.pearson@flinders.edu.au

Key words: MHC, genetic differentiation, selection, fine spatial scale, *Egernia stokesii*

In preparation

Abstract

The highly polymorphic genes of the major histocompatibility complex (MHC) are involved in disease resistance, mate choice, and kin recognition. Therefore, they are among the most widely used markers for investigating adaptive variation. Although selection is the key driver, gene flow and genetic drift also influence adaptive genetic variation, sometimes in opposing ways and with consequences for adaptive potential. Selection drives MHC variation through two key processes: parasite mediated selection and sexual selection, which may act in tandem. To further our understanding of the processes that generate MHC variation, it is helpful to compare variation at the MHC with that at neutral genetic loci. Differences in MHC and neutral genetic variation are useful for inferring the relative influence of selection, gene flow and drift on MHC variation. To date, such investigations have usually been undertaken at a broad spatial scale. Yet, evolutionary and ecological processes can occur at a fine spatial scale, particularly in small or fragmented populations. We investigated spatial patterns of MHC variation among three geographically close, naturally discrete, sampling sites of *Egernia stokesii*, an Australian lizard. The MHC of *E. stokesii* has recently been characterised and there is evidence for historical selection on the MHC. We found *E. stokesii* MHC weakly differentiated among sites compared to microsatellites, suggesting selection, acting similarly at each site, has outweighed any effects of low gene flow or of genetic drift on *E. stokesii* MHC variation. Our findings demonstrate the strength of selection in shaping patterns of MHC variation or consistency at a fine spatial scale and highlight that both adaptive and neutral genetic markers should be used when assessing species genetic variation.

Introduction

Adaptive genetic variation has fitness consequences with implications for the evolutionary potential of populations and species (Hedrick, 2001, Reed and Frankham, 2003, but see Radwan et al., 2010). Because of their role in disease resistance (Doherty and Zinkernagel, 1975, Spurgin and Richardson, 2010), mate choice (Penn, 2002, Milinski, 2006) and kin recognition (Brown and Eklund, 1994), the highly variable genes of the major histocompatibility complex (MHC) have high adaptive potential. Therefore, they are appropriate markers for investigating adaptive variation within and among populations (Bernatchez and Landry, 2003). Although selection is the main driver of adaptive genetic variation, gene flow and genetic drift can also change allele frequencies, often with opposite effects on genetic differentiation within and among populations (Bernatchez and Landry, 2003, Dionne et al., 2008, Taylor et al., 2012). Where selection pressures differ among local populations, genetic differentiation among these populations is predicted to increase at loci subject to those selective forces (Kaltz and Shykoff, 1998, Kawecki and Ebert, 2004). Conversely, if selection pressures are similar among local populations, then we expect little or no genetic differentiation among populations at the selected loci.

Superimposed on the selection processes, if gene flow is limited, genetic drift can reduce genetic variation within populations and increase genetic differentiation among populations (Radwan et al., 2010, Sutton et al., 2011). On the other hand, if gene flow is not restricted, migration and dispersal can reduce genetic differentiation among populations (Slatkin, 1987). Thus, an understanding of the relative influence of selection, gene flow, and genetic drift is important because independently, and in combination, they can affect adaptive potential. Because a loss of MHC variation may reduce resistance to pathogens (Sommer, 2005, but see Radwan et al., 2010, Siddle et al., 2007), understanding the processes that generate and

maintain that variation is a critical component of species conservation. Thus, MHC loci are commonly used in investigations of the relative influence of these processes.

There are two key mechanisms of selection that contribute to maintaining high levels of MHC variability in populations (Apanius et al., 1997, Edwards and Hedrick, 1998). One, parasite mediated selection, infers that, with parasite-host coevolution, greater MHC variation in hosts confers a wider defence against evolving variation in the defence evasion strategies of their parasites (Jeffery and Bangham, 2000, Wegner et al., 2003). The three main types of selection that have been evoked to explain parasite mediated selection are: negative frequency-dependent selection (or rare allele advantage; Clarke and Kirby, 1966, Slade and McCallum, 1992), overdominance (or heterozygote advantage; Doherty and Zinkernagel, 1975, Hughes and Nei, 1988), and diversifying selection (or fluctuating selection; Hill, 1991). Sexual selection is an alternative process put forward to explain how selection favours MHC diversity. This infers that individuals choose mating partners for the diversity or compatibility of their MHC genotypes in order to confer the benefits of MHC variation to their offspring (Edwards and Hedrick, 1998, Reusch et al., 2001, Penn, 2002). The outcomes of these two processes, parasite mediated and sexual selection, may act in tandem and it is difficult to differentiate between them.

However, MHC variation is not influenced by selection alone, and genetic drift and gene flow may also have roles (Bernatchez and Landry, 2003, Sutton et al., 2011, Strand et al., 2012). To understand the relative importance of selection, drift and gene flow in generating adaptive genetic variation in the MHC, it is useful to compare MHC variation with the variation at neutral genetic loci, such as microsatellite DNA, in the same individuals (Bernatchez and Landry, 2003, Holderegger et al., 2006). Whereas gene flow and genetic drift influence both adaptive and neutral variation, only adaptive variation is subject to selection. Therefore, differences in levels of variation between MHC and neutral genetic markers are commonly

used to infer the relative influence of these processes on the maintenance or depletion of genetic variation, the generation of genetic differentiation, and the evolutionary potential of a species.

An understanding of the relative roles of selection, gene flow and genetic drift on MHC variation may be particularly relevant for small or fragmented populations where the isolation of each fragment may restrict gene flow and increase genetic drift, potentially weakening the adaptive potential for selection (Sutton et al., 2011). Although some populations may have become recently isolated because of anthropogenic influences, many others have been naturally fragmented for longer periods. For example, geographic barriers and natural patterns of habitat patchiness can result in isolated populations (Slatkin, 1987). Normal social behaviour, such as the formation of stable social groups within populations, can have equivalent effects, leading to fine scale spatial genetic structure even within apparently continuous populations (Ross, 2001, Rossiter et al., 2012). If spatial and social structure is sufficient to reduce gene flow or increase genetic drift within subsections of populations, those populations may become genetically differentiated at a similar fine scale.

Investigations of MHC variation should be undertaken at a comparable spatial scale to dispersal distance for each species because scale is species specific (Manel et al., 2003, Anderson et al., 2010). Yet investigations are usually undertaken at a broad spatial scale (Rico et al., 2015, Strand et al., 2012, Bichet et al., 2015). Fine scale investigations are more likely to mirror the scale of ecological patterns influenced by natural habitat fragmentation, spatial clustering of kin groups, and variation in parasite communities. This may be particularly relevant for long-lived social species inhabiting permanently disconnected natural habitats. Our study considered the spatial distribution of MHC variation among three geographically close, but naturally fragmented, sampling sites of one such species, *Egernia stokesii*, an Australian scincid lizard.

Egernia stokesii (J.E. Gray, 1845) is a large (180 mm snout-vent length; Cogger, 1983), viviparous lizard widely distributed across semi-arid Australia, residing in crevices within geographically isolated rocky outcrops. *Egernia stokesii* live in stable family groups (Duffield and Bull, 2002, Gardner et al., 2001), limited by the number of available rocky crevices (Gardner et al., 2007). They exhibit high levels of genetic monogamy (Gardner et al., 2002), limited dispersal (Gardner et al., 2001), and strong site fidelity (Pearson et al. Chapter 5). Recent work (Pearson et al. Chapter 6) demonstrated that mate choice in *E. stokesii* was predicted by a combination of their genes and their group membership. Lizards preferred mates from within their social group, but that were genetically diverse and genetically dissimilar (Pearson et al Chapter 6). Avoidance of kin, rather than sex-biased dispersal, appears to be the means by which inbreeding is avoided (Gardner et al., 2012). Individuals of *E. stokesii* are subject to a diversity of parasites across their range (Hallas et al., 2005, Keirans et al., 1996, Duffield and Bull, 1996, Telford and Stein, 2000, Stein, 1999, Stein and Dyce, 2002). Stein (1999) reported that prevalence differed significantly among adjacent outcrops for several blood parasite species, as did the diversity of blood parasite species infecting *E. stokesii.* Tick prevalence and intensity of infection also varied among outcrops (Stein, 1999). However, parasite prevalence was generally consistent within an outcrop across years (Stein, 1999). Although the effect of parasites on *E. stokesii* is poorly understood, at least one parasite, a gut nematode, affected lizard basking time and activity, indicating infection associated behavioural changes with potential fitness consequences (Fenner and Bull, 2008).

Egernia stokesii MHC has recently been characterised (Pearson et al. Chapter 3), and evidence for historical selection on the MHC has been reported (Pearson et al. Chapter 4). In this current study, we assessed if selection has a stronger impact on *E. stokesii* MHC variation than genetic drift or restricted gene flow at a fine spatial scale. To do this, we

compared patterns of variation of MHC and of neutral genetic markers (microsatellites) across three geographically close sampling sites (all within a 2km x 500 m area). We aimed to distinguish the strength of selection compared to gene flow and genetic drift, rather than understanding which form of selection was operating. Three alternate outcomes could arise from a comparison of differentiation among sites at MHC and microsatellite loci. First, we might detect no difference between levels of MHC and microsatellite differentiation, which would indicate a lack of selection. Second, we might detect higher levels of differentiation among microsatellite than among MHC loci, which would indicate some balancing selection on MHC, for instance stabilising similar genetic structures in the face of similar parasite challenges. In this case, selection is stronger than drift and acting in the opposite direction. Third, we might detect the MHC is more strongly differentiated than microsatellite loci, which would imply that diversifying selection is the main force influencing the MHC, but in this case acting in the same direction as drift.

For *E. stokesii*, the limited availability of refuges in rocky outcrops, coupled with a long life span, a stable family group structure and monogamy, is likely to constrain gene flow and increase spatial differentiation of genetic structure through drift. However, we predicted that strong selection on *E. stokesii* MHC is likely to counter these effects, and reduce differentiation of MHC. Previous research has recorded both a significant impact of parasites on individual fitness (Fenner and Bull, 2008) and evidence for MHC based mate choice (Pearson et al. Chapter 6). Although there may be short-term differences among outcrops in the prevalence and infection intensity of some apicomplexan blood parasites in and of ectoparasitic ticks (Stein, 1999), the broad suite of parasites threatening *E. stokesii* will be similar across sites, suggesting selection will not lead to MHC differentiation. Additionally, *E. stokesii* prefer genetically diverse and dissimilar, rather than similar mates (Pearson et al. Chapter 6) which suggests strong selection to increase MHC diversity, but little pressure for

local adaptation (Bonneaud et al., 2006). Therefore, we predicted *E. stokesii* would be less strongly differentiated for the MHC than for microsatellite loci, because stabilising selective forces have outweighed any influence of low gene flow and genetic drift.

Method

Field surveys

In field surveys conducted between September 2012 and March 2013, we sampled *E. stokesii* at three discrete rocky outcrops: Camel Hill (CAM), Castle Rock (CAS), and Castle Rock Ridge (CRR) near Hawker (31°54′S; 138°25′E) in the southern Flinders Ranges, South Australia. Each outcrop was located on a ridge separated from each other outcrop site by 300 m to 1.5 km of non-rocky matrix (Gardner et al., 2007). The matrix between the three outcrops was sparsely vegetated and lacked suitable *E. stokesii* habitat. Although CAS and CRR outcrops were separated by about 300 m at their closest point, we have rarely recorded dispersal between them (Duffield and Bull, 2002; Pearson et al. Chapter 5). Other outcrops previously surveyed around Hawker (Gardner et al., 2007) were located from 500 m to 35 km from the three sampled outcrops (see Fig 1, Appendix 1, Pearson et al. Chapter 5). Study sites and sampling methodology have previously been described (Duffield and Bull, 2002, Godfrey et al., 2006, Gardner et al., 2007). Upon capture by hand or in Elliot traps, we marked lizards by toe clipping or microchip and recorded capture locations. Lizards were weighed (gm), measured (mm), and sexed by inverting the hemipenes of males. We inferred lizard age class from snout vent length (SVL) (Gardner et al., 2007). Blood (up to 0.5 mL) was taken from the lizard caudal vein and stored on Whatman FTA ® Elute.

MHC alleles and genotypes

DNA extraction and MHC amplification, sequencing, allele identification and genotyping are described in detail in Pearson et al. (Chapter 3). Loci developed by Ansari et al. (2015) and

Ansari (2016) were used for MHC amplification (Pearson et al. Chapter 3). We used blood samples collected during field surveys to derive genotypes for a 216 base pair region of MHC I exon 2 (corresponding to the α-1 domain) and a 102 base pair region of MHC II exon 2 (corresponding to the β-1 domain). We previously identified 35 MHC I and five MHC II alleles and genotyped 198 lizards for MHC I and 89 lizards for MHC II (Pearson et al. Chapter 3). Although we sampled 260 *E. stokesii*, not all lizards were successfully sequenced for both MHC I and MHC II.

The sequenced MHC I region contained significantly more variable codons than the sequenced MHC II region (Pearson et al. Chapter 3 and 4). Tests of historical selection on *E. stokesii* MHC indicated strong positive selection on MHC I, in which five positively selected codons have been identified (Pearson et al. Chapter 4). Although similar evidence for selection has not been detected in MHC II (Pearson et al. Chapter 4), we used both MHC I and II in this study because their patterns of structure, function and expression differ (Hughes and Yeager, 1998). Therefore, different selection processes and population patterns may be expected (Wedekind et al., 1995, Milinski, 2006, Strandh et al., 2012).

Microsatellite DNA genotyping

We used blood samples collected during field surveys to derive microsatellite DNA genotypes. A multiplex polymerase chain reaction (PCR) approach was used to amplify 11 previously described polymorphic microsatellite loci (Est 1, Est 3, Est 4, Gardner et al., 1999, Ecu 2, Stow, 2002, Tr 3.2,Tr 5.20, Cooper et al., 1997, TrL 12, TrL 14, TrL 28, TrL 29, TrL 35, Gardner et al., 2008). We performed PCR-amplifications using the QIAGEN ® Multiplex PCR Kit (QIAGEN ®, Catalogue 206143). Each 10 µl uniplex reaction mix contained 0.10 x QIAGEN Multiplex PCR Master Mix, 0.125-0.25 µM of each primer, 2 µl of template DNA, and RNAse Free Water. Cycling conditions were 15 minutes at 95 ºC, 35 cycles of 30 seconds at 94 ºC, 90 seconds at 57 ºC, one minute at 72 ºC, and a final elongation step of 30

minutes at 60 ºC followed by 30 seconds at 25 ºC. PCR product was submitted to the AGRF for capillary separation on an AB3730 DNA analyser.

We assessed the neutrality of loci using the FDIST2 approach of Beaumont and Nichols (1996) implemented in LOSITAN (Antao et al., 2008) and BAYESCAN (Foll and Gaggiotti, 2008). The incidence of null alleles and of large allele dropout was assessed using Micro-Checker (van Oosterhout et al., 2004) and linkage disequilibrium and Hardy-Weinberg equilibrium were tested using Genepop (V4.3; Rousset, 2008). We corrected linkage disequilibrium significance levels using a false discovery rate test (threshold of 0.1; Benjamini and Hochberg, 1995). Genotypes were derived using GeneMapper ® (Applied Biosystems). We amplified a subset of samples in two independent PCRs and these were subsequently both genotyped to assess genotyping reliability.

Datasets used in analyses

The inclusion of multiple samples from close kin can bias assessment of genetic structure within an outcrop. Therefore, in all analyses we used MHC and microsatellite genotypes only from non-full-sibling adults. We identified these individuals by first calculating the average pairwise genetic relatedness (r) of lizards using the Wang (2002) coefficient in Coancestry (Wang, 2011)(http://www.zsl.org/science/software/coancestry) (results not shown). We assumed dyads with $r > 0.5$ were full sibs and then removed one lizard from the pair, usually retaining the individual with most MHC genotype information, or the one that maximised the spatial distribution of samples within a site.

To compare differentiation of *E. stokesii* MHC to that of microsatellites, and to ensure we used the same individuals for each marker, we derived two source datasets: 1) lizards genotyped for both MHC I and microsatellites, 2) lizards genotyped for both MHC II and microsatellites. Then, we performed analyses separately for MHC I and MHC II. For

microsatellites, we used only the microsatellite genotypes of lizards which had also been genotyped for MHC I as this was the larger dataset, and would generate more rigorous genetic estimates.

Genetic diversity

For each marker set, we estimated *E. stokesii* genetic diversity for each outcrop before assessing how the genetic variation was partitioned. MHC diversity was determined by seven commonly used parameters. These were: 1) total number of MHC alleles per population, 2) mean, and minimum and maximum number of alleles per individual within the population, 3) allelic richness (Theta k), 4) number of polymorphic (segregating) sites (S), 5) nucleotide diversity (π, the average proportion of nucleotides that differ between random sequence pairs), 6) nucleotide diversity using the Jukes and Cantor (1969) correction (π(JC), which accounts for sequence evolution by correcting for base mutations that have occurred more than once), and 7) the average number of nucleotide differences between unique alleles (k). We calculated theta (θ) k (the index of allelic richness) in Arlequin v3.5.2 (Excoffier and Lischer, 2010). Theta (θ) k is estimated from the infinite-allele-equilibrium (Ewens, 1972) between the expected number of alleles, the sample size, and θ. We used DnaSP v.5.10.01 (Librado and Rozas, 2009) to calculate the remaining MHC diversity measures.

We assessed microsatellite genetic diversity within each outcrop with the following measures: 1) mean number of alleles (A), 2) mean observed heterozygosity (H_O), and 3) mean expected heterozygosity (H_E) calculated in Arlequin v. 3.5.2.2. Intra-population indices of fixation (F_{IS}) were calculated in Genepop v. 4.3. Allelic richness (AR) was calculated in HP-RARE v. 1.1 (Kalinowski, 2005) using a rarefaction approach, because estimates may be biased by sample size differences (Leberg, 2002). We tested differences in both MHC and microsatellite diversity measures among the three outcrops using an ANOVA with Tukey's HSD post-hoc comparisons in R version 3.1.3 (R Core Team, 2014).

Since genome-wide variability may largely explain MHC variability (Boyce et al., 1997), we also tested for a correlation between MHC allelic richness and microsatellite allelic richness in samples from each outcrop using Spearman's rho calculated in R. Correlations for MHC I and MHC II were calculated separately. A lack of significant correlation between MHC and microsatellite allelic richness would indicate 1) different processes have influenced *E. stokesii* adaptive and neutral genetic variation and 2) sample sizes were sufficient for analysis.

Genetic differentiation among sites

We tested our prediction about the influence of selection, gene flow and genetic drift on MHC differentiation, using two methods. First we calculated pairwise and global F_{ST} in Arlequin, using 10,000 permutations, and pairwise and global *D* (Jost, 2008) for MHC in SPADE v.Feb 2009 (Chao and Shen, 2010) with 10,000 permutations, and for microsatellites using SMOGD v.1.2.5 (Crawford, 2010), available online at http://www.ngcrawford.com/django/jost/ (accessed 4 September 2015) using 1000 bootstrap replicates to derive 95% confidence intervals. Second, we used Analysis of Molecular Variance (AMOVA) (Excoffier et al., 1992) to examine the partitioning of *E. stokesii* genetic variation between two different levels of hierarchical structure: within and among outcrops. We performed an AMOVA for each marker set in Arlequin with 10,000 permutations. PGDSpider (Lischer and Excoffier, 2012) was used to convert files into formats required by the various programs utilised.

Results

Field surveys

We sampled 260 *E. stokesii* individuals during the 2012-13 field survey including 108 adults, 101 sub-adults, and 51 juveniles (Table 1).

Microsatellite and MHC genotyping

For all microsatellite loci, there was no evidence of null alleles, linkage disequilibrium or deviations from Hardy-Weinberg equilibrium. Combined results from two alternative tests of neutrality, indicated no microsatellite loci were under positive selection (results not shown). Therefore, we used all eleven microsatellite loci in genotyping. Three of 255 PCR duplicates did not match, representing a 1.18% genotyping error rate. For MHC population genetic analyses, we retained 66 non-full sibling adults genotyped for MHC I and microsatellites and 34 non-full sibling adults genotyped for MHC II and microsatellites (Table 1). For microsatellite population genetic analyses, we used the same 66 individuals genotyped for MHC I (Table 1).

Genetic diversity

Levels of genetic diversity within sites were not significantly different among outcrops for MHC I or microsatellites ($P > 0.05$, ANOVA, Tukey's HSD test) (Table 2 and 3). However, one outcrop pair (CRR and CAS) differed significantly in one genetic diversity measure for MHC II, the number of alleles ($P = 0.045$, ANOVA, Tukey's HSD test), with CRR having more alleles than CAS (Table 2). There was no significant correlation between allelic richness of either MHC I or MHC II with microsatellite allelic richness ($P > 0.05$), indicating adequate sample sizes.

Genetic differentiation among sites

For MHC I and II there was no significant genetic differentiation among sites, and pairwise and global F_{ST} and D estimates were not significant (Table 4). On the other hand, microsatellite loci were significantly genetically differentiated, with global F_{ST} and some pairwise estimates significant (Table 4). AMOVA analyses showed that all (100%) of the MHC I and II variation was contained within rather than among outcrops ($P > 0.05$ for MHC

I and II) whereas microsatellite variation was explained by variation both within (95.5%) and among (4.5%) outcrops ($P < 0.01$).

Discussion

Our results confirmed the prediction that *E. stokesii* would be less strongly differentiated among sites for the MHC than for microsatellites. We found no evidence for *E. stokesii* MHC differentiation among three rocky outcrops located within two km of each other, while microsatellite loci in the same individuals did diverge. This suggests that selection is the main process generating the observed patterns of *E. stokesii* MHC variation, and that it outweighs any impact of drift. That selection has a key role in *E. stokesii* MHCI variation is supported by evidence for historical selection previously demonstrated (Pearson et al. Chapter 4). In contrast, *E. stokesii* microsatellites were differentiated at a fine spatial scale. Since there is no evidence for selection on the neutral microsatellite loci, the observed microsatellite differentiation suggests that either genetic drift or reduced gene flow have contributed to patterns of *E. stokesii* microsatellite variation.

The patterns of adaptive and neutral genetic variation found for *E. stokesii* are consistent with those found for other species. Genetic differentiation at microsatellites but not at the MHC has also been found in wolverines (Rico et al., 2015), house sparrows (Bichet et al., 2015), black grouse (Strand et al., 2012), and a molly (Tobler et al., 2014). Authors commonly propose overdominance or negative frequency dependent selection on the MHC to explain weaker MHC differentiation compared to microsatellites. In contrast, stronger MHC differentiation compared to microsatellites has been reported in salmon (Miller et al., 2001), great snipe (Ekblom et al., 2007), house sparrows (Loiseau et al., 2009) and racoons (Kyle et al., 2014), which was explained by adaptation to local parasites. In other cases, similar patterns of genetic differentiation for both the MHC and microsatellites are evident in

bighorn sheep (Boyce et al., 1997) and toads (Zeisset and Beebee, 2014) which suggest neutral forces are strongest on both markers.

The aforementioned comparative studies of MHC and microsatellite differentiation sampled multiple populations across broad spatial scales. For example wolverines were sampled across Russia and eight regions of Canada (Rico et al., 2015), black grouse were sampled from eleven locations across Europe (Strand et al., 2012), and house sparrows were sampled across six insular and six mainland populations (Bichet et al., 2015). In contrast, our study provides a novel insight into a lack of MHC differentiation at a much smaller geographic scale. If outcrops are like islands, then reduced gene flow may strengthen the effects of genetic drift, which in turn may weaken selective forces that generate adaptive variation. Microsatellite differentiation among *E. stokesii* populations suggest the presence of barriers to gene flow which are likely to include habitat (rocky outcrop and crevice) availability, low vagility and dispersal, delayed maturity, natal philopatry, high genetic monogamy, and family group structure. Despite the presence of such barriers, *E. stokesii* MHC lacks differentiation among outcrops at the same scale as microsatellites, probably due to the nature of selection.

Parasites may drive the generation of *E. stokesii* MHC variation directly via parasite-host coevolution. Previous work found nematode infection influenced *E. stokesii* basking time and movement (Fenner and Bull, 2008). Infected lizards spent less time basking and moving around which has fitness consequences because basking is important for thermoregulation (Fenner and Bull, 2008). Although parasite mediated selection may be generating *E. stokesii* MHC variation, the lack of MHC differentiation among outcrops suggests a lack of local adaptation. This is further supported by earlier work that found *E. stokesii* prefer diverse and dissimilar mates (Pearson et al. Chapter 6). If parasites were sufficiently different among populations to promote local adaptation, genetically similar mates should be preferred (Bonneaud et al., 2006) and a greater proportion of MHC variation should be partitioned

among, rather than within, outcrops. An alternative explanation is that parasites may be different, but lack significant fitness consequences. We acknowledge that our study did not directly test for associations of MHC alleles and parasites. Further research is required to identify parasite-host dynamics and parasite fitness consequences for *E. stokesii*.

Alternatively, or in addition to the direct effects of parasite-host coevolution, parasites may indirectly influence *E. stokesii* MHC variation if lizards choose a mate based on MHC genotype. Mate choice for an individual with higher MHC diversity or a more compatible MHC genotype provides the benefit of increased offspring MHC variation as a means of parasite resistance (Brown, 1997, Landry et al., 2001). *Egernia stokesii* can recognise kin and evidence exists for the use of chemosensory cues (Bull et al., 2000, Main and Bull, 1996), thus lizards may be able to detect the MHC genotype of potential mates and use that information in mate choice decisions. Evidence exists for a genetic basis of *E. stokesii* mate choice (Pearson et al. Chapter 6). *Egernia stokesii* reproductive pairings were predicted by a combination of membership of the same social group and genetic variables (higher MHC and microsatellite diversity, lower pairwise microsatellite relatedness, and a lower proportion of shared MHC alleles) (Pearson et al. Chapter 6). How lizards form and maintain social groups and how genotypes are detected is yet to be investigated.

We demonstrate that although populations may be ecologically, socially, and genetically clustered, as in the case of *E. stokesii*, selection can outweigh the effects of gene flow and drift at a small scale such that populations avoid divergence and maintain adaptive diversity, with implications for species persistence. Although few studies have been undertaken at a fine scale, results similar to ours have been found for house sparrows. A lack of differentiation among geographical close mainland populations of sparrows has been found despite the sedentary behaviour and limited dispersal exhibited by the species (Bichet et al., 2015). Further, even though island populations of sparrows were more differentiated for the

MHC than mainland populations, island populations were less differentiated for the MHC than microsatellites, indicating that selection can outweigh drift in small, fragmented populations (Bichet et al., 2015). Other studies have found differentiation increased with geographic distance (Loiseau et al., 2009) therefore additional studies of *E. stokesii* populations across varying spatial scales are warranted.

Our investigation of MHC genetic differentiation among *E. stokesii* sampled at three rocky outcrops within two kilometres of each-other provides evidence that selection can outweigh the effects of restricted gene flow and genetic drift for naturally fragmented animal populations at a smaller spatial scale than is usually studied. These findings demonstrate that evidence for neutral genetic differentiation within a species need not imply that adaptive markers are also genetically differentiated. Further, findings lend support to the use of both adaptive and neutral markers when assessing the genetic variation of a species. Indeed, although a loss of adaptive genetic variation does not necessarily increase extinction risk, an investigation of both provides a more comprehensive assessment of species genetic variation. As natural habitat becomes increasingly fragmented, investigations of adaptive versus neutral genetic variation at a fine scale may become progressively more important to understand the relevant influence of selection, gene flow and drift on contemporary populations. This may be particularly relevant for animals that are also fragmented by their socio-sexual system. Once patterns of genetic differentiation have been identified, work is required to understand the processes generating the observed patterns.

Acknowledgements

We thank volunteers for assistance in the field and Mina Ansari and Tessa Bradford for assistance with laboratory work. Field survey methods were approved by the Flinders University Animal Welfare Committee, using standards and procedures in compliance with the Australian Code of Practice for the use of animals for scientific purposes. Laboratory

work was undertaken at the South Australian Research Facility for Molecular Ecology and Evolution, Adelaide. Funding from the Holsworth Wildlife Research Endowment, The Field Naturalists Society of South Australia, Nature Foundation South Australia and the Sir Mark Mitchell Research supported this work.

Figures and tables

Table 1: Summary of *Egernia stokesii* individuals sampled during field surveys, genotyped for MHC I, MHC II, and 11 microsatellite loci, and subsequently retained for population genetic analyses.

Pop	Total	A	SA	J	MHC I_{ALL}	MHC II_{ALL}	$Micsats_{ALL}$	$Adults_1$	$Adults_2$
CAM	82	34	31	17	64	35	78	22	14
CAS	92	42	30	20	70	26	84	26	10
CRR	86	32	40	14	64	28	76	18	10
Total	260	108	101	51	198	89	238	66	34

Pop: CAM (Camel Hill), CAS (Castle Rock), CRR (Castle Rock Ridge); Total: total number of lizards sampled, A=Adult, SA=Subadult, J=Juvenile; MHC I_{ALL}: total number of lizards genotyped for MHC I; MHC II_{ALL}: total number of lizards genotyped for MHC II; $Micsats_{ALL}$: number of lizards genotyped for 11 microsatellite loci; $Adults_1$: number of non-full sibling adults used in MHC I and microsatellite population genetic analyses, $Adults_2$: number of non-full sibling adults used in MHC II population genetic analyses.

Table 2: MHC diversity statistics for three *Egernia stokesii* populations: Camel Hill (CAM), Castle Rock (CAS), and Castle Rock Ridge (CRR) in the southern Flinders Ranges of South Australia. Estimates are derived from genotypes of a 216 base pair region of MHC I exon 2 (α-1 domain) and 102 base pair region of MHC II exon 2 (β-1) for 66 (MHC I) and 34 (MHC II) non-full sibling adults.

	N	N_{ADULTS}	N_{ALL}	N_{MEAN}	N_{MIN}	N_{MAX}	A_{RICH}	S	π	π (JC)	k
MHC I											
CAM	22	31	32	8.41 (SE±0.62)	1	16	10.407	148	0.152	0.195	32.761
CAS	26	28	31	7.75 (SE±0.73)	2	16	8.567	146	0.151	0.194	32.711
CRR	18	24	31	7.33 (SE±0.99)	3	21	8.327	145	0.163	0.211	35.107
MHC II											
CAM	14	4	5	2.07 (SE±0.20)	1	3	1.016	26	0.057	0.064	5.793
CAS	10	2	3	1.60 (SE±0.16)	1	2	0.347	4	0.021	0.021	2.100
CRR	10	5	5	2.40 (SE±0.27)	1	4	1.625	27	0.048	0.053	4.877

Number of non-full sibling adults included in analysis (N), total number of MHC alleles in non-full sibling adults (N_{ADULTS}), total number of MHC alleles in all lizards genotyped (N_{ALL}), mean number of alleles per individual within the population (N_{MEAN}), minimum number of alleles per individual within the population (N_{MIN}), maximum number of alleles per individual within the population (N_{MAX}), allelic richness (A_{RICH}), number of polymorphic (segregating) sites (S), nucleotide diversity (π), nucleotide diversity using Jukes & Cantor (1969) correction (π(JC)), average number of nucleotide differences between unique alleles (k)

Table 3: Microsatellite genetic diversity estimates for three *Egernia stokesii* populations including sample size (N), mean number of alleles (A), allelic richness (AR), mean observed heterozygosity (H_O), mean expected heterozygosity (H_E), and intra-population fixation indices (F_{IS}). Standard errors are shown in brackets. Populations are Camel Hill (CAM), Castle Rock (CAS), and Castle Rock Ridge (CRR). Estimates are derived from genotypes of 11 microsatellite loci for 66 non full sibling adults.

Site	N	*A*	*AR*	H_o	H_e	F_{is}
CAM	22	10.64 (4.68)	6.15	0.80 (0.09)	0.82 (0.13)	0.026
CAS	26	9.18 (3.34)	5.55	0.77 (0.14)	0.77 (0.14)	0.009
CRR	18	10.09 (4.35)	5.92	0.74 (0.18)	0.77 (0.20)	0.034

Table 4: Pairwise and global F_{ST} and *D* for a) MHC I (n=66), b) MHC II (n=34) and c) 11 microsatellite loci (n=66) for *Egernia stokesii* from three populations: Camel Hill (CAM), Castle Rock (CAS), and Castle Rock Ridge (CRR). F_{ST} estimates are below the diagonal and *D* estimates are above the diagonal. *P* values for F_{ST} estimates and confidence intervals for MHC *D* are shown in brackets. Per locus confidence intervals for microsatellite *D* values showed overall D values were significant (results not shown).

a) MHC I	CAM	CAS	CRR
CAM		-0.055 (0.000, 0.015)	-0.061 (0.000, 0.030)
CAS	-0.004 (0.985)		-0.056 (0.000, 0.031)
CRR	-0.005 (0.934)	-0.004 (0.938)	
Global: F_{ST}: -0.004; D: -0.057			
b) MHC II	CAM	CAS	CRR
CAM		0.071 (0.000, 0.349)	-0.080 (0.000, 0.128)
CAS	0.027 (0.243)		0.040 (0.000, 0.317)
CRR	-0.032 (0.937)	0.011 (0.293)	
Global: F_{ST}: -0.006; D: 0.019			
c) Microsatellites	CAM	CAS	CRR
CAM		0.222	0.189
CAS	0.053 (0.00)		0.077
CRR	0.052 (0.00)	0.026 (0.00)	
Global: F_{ST}: 0.045; D: 0.179			

References

ANDERSON, C. D., EPPERSON, B. K., FORTIN, M. J., HOLDEREGGER, R., JAMES, P., ROSENBERG, M. S., SCRIBNER, K. T. & SPEAR, S. 2010. Considering spatial and temporal scale in landscape-genetic studies of gene flow. *Molecular Ecology,* 19**,** 3565-3575.

ANSARI, T. H. 2016. *The distribution of genetic variation in sleepy lizards.* Doctoral dissertation, Flinders University of South Australia.

ANSARI, T. H., BERTOZZI, T., MILLER, R. D. & GARDNER, M. G. 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.

ANTAO, T., LOPES, A., LOPES, R. J., BEJA-PEREIRA, A. & LUIKART, G. 2008. LOSITAN: a workbench to detect molecular adaptation based on a Fst-outlier method. *BMC Bioinformatics,* 9**,** 323.

APANIUS, V., PENN, D., SLEV, P. R., RUFF, L. R. & POTTS, W. K. 1997. The nature of selection on the major histocompatibility complex. *Critical Reviews in Immunology,* 17**,** 179-224.

BEAUMONT, M. A. & NICHOLS, R. A. 1996. Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society B, Biological Sciences,* 263**,** 1619-1626.

BENJAMINI, Y. & HOCHBERG, Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological),* 57**,** 289-300.

BERNATCHEZ, L. & LANDRY, C. 2003. MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? *Journal of Evolutionary Biology,* 16**,** 363-377.

BICHET, C., MOODLEY, Y., PENN, D. J., SORCI, G. & GARNIER, S. 2015. Genetic structure in insular and mainland populations of house sparrows (*Passer domesticus*) and their hemosporidian parasites. *Ecology and Evolution,* 5**,** 1639-1652.

BONNEAUD, C., PÉREZ-TRIS, J., FEDERICI, P., CHASTEL, O. & SORCI, G. 2006. Major histocompatibility alleles associated with local resistance to malaria in a passerine. *Evolution,* 60**,** 383-389.

BOYCE, W. M., HEDRICK, P. W., MUGGLI-COCKETT, N. E., KALINOWSKI, S., PENEDO, M. C. T. & RAMEY, R. R. 1997. Genetic variation of major histocompatibility complex and microsatellite loci: a comparison in bighorn sheep. *Genetics,* 145**,** 421-433.

BROWN, J. L. 1997. A theory of mate choice based on heterozygosity. *Behavioral Ecology,* 8**,** 60-65.

BROWN, J. L. & EKLUND, A. 1994. Kin recognition and the major histocompatibility complex: an integrative review. *American Naturalist***,** 435-461.

BULL, C. M., GRIFFITH, S. C., LANHAM, E. J. & JOHNSTON, G. R. 2000. Recognition of pheromones from group members in a gregarious lizard, *Egernia stokesii*. *Journal of Herpetology,* 34**,** 92-99.

CHAO, A. & SHEN, T.-J. 2010. *Program SPADE (Species Prediction And Diversity Estimation). Program and User's Guide.* [Online]. Available: http://chao.stat.nthu.edu.tw [Accessed 23rd January 2011].

CLARKE, B. & KIRBY, D. 1966. Maintenance of histocompatibility polymorphisms. *Nature* 211**,** 999-1000.

COGGER, H. G. 1983. *Reptiles and amphibians of Australia*, AH & AW Reed Pty Ltd.

COOPER, S. J. B., BULL, C. M. & GARDNER, M. G. 1997. Characterization of microsatellite loci from the socially monogamous lizard *Tiliqua rugosa* using a PCR-based isolation technique. *Molecular Ecology,* 6**,** 793-795.

CRAWFORD, N. G. 2010. SMOGD: software for the measurement of genetic diversity. *Molecular Ecology Resources,* 10**,** 556-557.

DIONNE, M., CARON, F., DODSON, J. J. & BERNATCHEZ, L. 2008. Landscape genetics and hierarchical genetic structure in Atlantic salmon: the interaction of gene flow and local adaptation. *Molecular Ecology,* 17**,** 2382-2396.

DOHERTY, P., C. & ZINKERNAGEL, R., M. 1975. Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. *Nature,* 256**,** 50-52.

DUFFIELD, G. A. & BULL, C. M. 1996. Host location by larvae of the reptile tick *Amblyomma vikirri* (Acari: Ixodidae). *Experimental & Applied Acarology,* 20**,** 575-582.

DUFFIELD, G. A. & BULL, C. M. 2002. Stable aggregations in an Australian lizard, *Egernia stokesii*. *Naturwissenschaften,* 89**,** 424-427.

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

EKBLOM, R., SAETHER, S. A., JACOBSSON, P., FISKE, P., SAHLMAN, T., GRAHN, M., KÅLÅS, J. A. & HÖGLUND, J. 2007. Spatial pattern of MHC class II variation in the great snipe (*Gallinago media*). *Molecular ecology,* 16**,** 1439-1451.

EWENS, W. J. 1972. The sampling theory of selectively neutral alleles. *Theoretical Population Biology,* 3**,** 87-112.

EXCOFFIER, L. & LISCHER, H. E. L. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources,* 10**,** 564-567.

EXCOFFIER, L., SMOUSE, P. E. & QUATTRO, J. M. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics,* 131**,** 479-491.

FENNER, A. L. & BULL, C. M. 2008. The impact of nematode parasites on the behaviour of an Australian lizard, the gidgee skink *Egernia stokesii*. *Ecological Research,* 23**,** 897-903.

FOLL, M. & GAGGIOTTI, O. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics,* 180**,** 977-993.

GARDNER, M. G., BULL, C. M., COOPER, J. B. & DUFFIED, G. A. 2001. Genetic evidence for a family structure in stable social aggregations of the Australian lizard *Egernia stokesii*. *Molecular Ecology,* 10**,** 175-183.

GARDNER, M. G., BULL, C. M. & COOPER, S. J. B. 2002. High levels of genetic monogamy in the group-living Australian lizard *Egernia stokesii*. *Molecular Ecology,* 11**,** 1787-1794.

GARDNER, M. G., BULL, C. M., FENNER, A., MURRAY, K. & DONNELLAN, S. C. 2007. Consistent social structure within aggregations of the Australian lizard, *Egernia stokesii* across seven disconnected rocky outcrops. *Journal of Ethology,* 25**,** 263-270.

GARDNER, M. G., COOPER, S. J. B., BULL, C. M. & GRANT, W. N. 1999. Isolation of microsatellite loci from a social lizard, *Egernia stokesii*, using a modified enrichment procedure. *Journal of Heredity,* 90**,** 301-304.

GARDNER, M. G., GODFREY, S. S., FENNER, A. L., DONNELLAN, S. C. & BULL, C. M. 2012. Fine-scale spatial structuring as an inbreeding avoidance mechanism in the social skink *Egernia stokesii*. *Australian Journal of Zoology,* 60**,** 272-277.

GARDNER, M. G., SANCHEZ, J. J., DUDANIEC, R. Y., RHEINBERGER, L., SMITH, A. L. & SAINT, K. S. 2008. *Tiliqua rugosa* microsatellites: isolation via enrichment and characterisation of loci for multiplex PCR in *T. rugosa* and the endangered *T. adelaidensis*. *Conservation Genetics,* 9**,** 233-237.

GODFREY, S. S., BULL, C. M., MURRAY, K. & GARDNER, M. G. 2006. Transmission mode and distribution of parasites among groups of the social lizard *Egernia stokesii*. *Parasitology Research,* 99**,** 223-230.

HALLAS, G., BULL, C. M. & BURSEY, C. R. 2005. *Pharyngodon tiliquae* and *Thelandros trachysauri* (Nematoda: Pharyngodonidae), new parasite records for *Egernia stokesii* (Scincidae) from Australia. *Comparative Parasitology,* 72**,** 119-120.

HEDRICK, P. W. 2001. Conservation genetics: where are we now? *Trends in Ecology & Evolution,* 16**,** 629-636.

HILL, A. V. 1991. HLA associations with malaria in Africa: some implications for MHC evolution. *Molecular evolution of the major histocompatibility complex.* Springer.

HOLDEREGGER, R., KAMM, U. & GUGERLI, F. 2006. Adaptive vs. neutral genetic diversity: implications for landscape genetics. *Landscape Ecology,* 21**,** 797-807.

HUGHES, A. L. & NEI, M. 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature,* 335**,** 167-170.

HUGHES, A. L. & YEAGER, M. 1998. Natural selection at major histocompatibilty complex loci of vertebrates. *Annual Review of Genetics,* 321**,** 415-435.

JEFFERY, K. J. M. & BANGHAM, C. R. M. 2000. Do infectious diseases drive MHC diversity? *Microbes and Infection,* 2**,** 1335-1341.

JOST, L. 2008. G_{ST} and its relatives do not measure differentiation. *Molecular Ecology,* 17**,** 4015-4026.

JUKES, T. H. & CANTOR, C. R. 1969. *Evolution of protein molecules*, Academic Press.

KALINOWSKI, S. T. 2005. HP-RARE 1.0: a computer program for performing rarefraction on measures of allelic richness. *Molecular Ecology Notes,* 5**,** 187-189.

KALTZ, O. & SHYKOFF, J. A. 1998. Local adaptation in host–parasite systems. *Heredity,* 81**,** 361-370.

KAWECKI, T. J. & EBERT, D. 2004. Conceptual issues in local adaptation. *Ecology letters,* 7**,** 1225-1241.

KEIRANS, J. E., BULL, C. M. & DUFFIELD, G. A. 1996. *Amblyomma vikkiri* n. sp. (Acari: Ixodida: Ixodidae), a parasite of the gidgee skink *Egernia stokesii* (Reptilia: Scincidae) from South Australia. *Systematic Parasitology,* 34**,** 1-9.

KYLE, C. J., RICO, Y., CASTILLO, S., SRITHAYAKUMAR, V., CULLINGHAM, C. I., WHITE, B. N. & POND, B. A. 2014. Spatial patterns of neutral and functional genetic variations reveal patterns of local adaptation in raccoon (*Procyon lotor*) populations exposed to raccoon rabies. *Molecular ecology,* 23**,** 2287-2298.

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 of London, Series B: Biological Sciences,* 268**,** 1279-1285.

LEBERG, P. L. 2002. Estimating allelic richness: Effects of sample size and bottlenecks. *Molecular Ecology,* 11**,** 2445-2449.

LIBRADO, P. & ROZAS, J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics,* 25**,** 1451-1452.

LISCHER, H. E. L. & EXCOFFIER, L. 2012. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics,* 28**,** 298-299.

LOISEAU, C., RICHARD, M., GARNIER, S., CHASTEL, O., JULLIARD, R., ZOOROB, R. & SORCI, G. 2009. Diversifying selection on MHC class I in the house sparrow (*Passer domesticus*). *Molecular Ecology,* 18**,** 1331-1340.

MAIN, A. R. & BULL, C. M. 1996. Mother-offspring recognition in two Australian lizards, *Tiliqua rugosa* and *Egernia stokesii*. *Animal Behaviour,* 52**,** 193-200.

MANEL, S., SCHWARZ, M. K., LUIKART, G. & TABERLET, P. 2003. Landscape genetics: combining landscape ecology and population genetics. *TRENDS in Ecology and Evolution,* 18**,** 189-197.

MILINSKI, M. 2006. The major histocompatibility complex, sexual selection, and mate choice. *Annual Review of Ecology, Evolution and Systematics,* 37**,** 159-186.

MILLER, K. M., KAUKINEN, K. H., BEACHAM, T. D. & WITHLER, R. E. 2001. Geographic heterogeneity in natural selection on an MHC locus in sockeye salmon. *Genetica,* 111**,** 237-257.

PENN, D. J. 2002. The scent of genetic compatibility: sexual selection and the major histocompatibility complex. *Ethology,* 108**,** 1-21.

RADWAN, J., BIEDRZYCKA, A. & BABIK, W. 2010. Does reduced MHC diversity decrease viability of vertebrate populations? *Biological Conservation,* 143**,** 537-544.

R CORE TEAM 2014. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

REED, D. H. & FRANKHAM, R. 2003. Correlation between fitness and genetic diversity. *Conservation Biology,* 17**,** 230-237.

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**,** 300-302.

RICO, Y., MORRIS-POCOCK, J., ZIGOURIS, J., NOCERA, J. J. & KYLE, C. J. 2015. Lack of Spatial Immunogenetic Structure among Wolverine (*Gulo gulo*) Populations Suggestive of Broad Scale Balancing Selection. *PloS one,* 10**,** e0140170.

ROSS, K. G. 2001. Molecular ecology of social behaviour: analyses of breeding systems and genetic structure. *Molecular Ecology,* 10**,** 265-284.

ROSSITER, S. J., ZUBAID, A., MOHD-ADNAN, A., STRUEBIG, M. J., KUNZ, T. H., GOPAL, S., PETIT, E. J. & KINGSTON, T. 2012. Social organization and genetic structure: insights from codistributed bat populations. *Molecular Ecology,* 21**,** 647-661.

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

SIDDLE, H. V., KREISS, A., ELDRIDGE, M. D. B., NOONAN, E., CLARKE, C. J., PYECROFT, S., WOODS, G. M. & BELOV, K. 2007. Transmission of a fatal clonal tumor by biting occurs due to depleted MHC diversity in a threatened carnivorous marsupial. *Proceedings of the National Academy of Sciences,* 104**,** 16221-16226.

SLADE, R. & MCCALLUM, H. 1992. Overdominant vs. frequency-dependent selection at MHC loci. *Genetics,* 132**,** 861.

SLATKIN, M. 1987. Gene flow and the geographic structure of natural populations. *Science,* 236**,** 787-792.

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**,** 979-988.

STEIN, J. 1999. *An ecological study of the blood parasites of E. stokesii.* Doctoral dissertation, Dissertation, Flinders Universtiy of South Australia.

STEIN, J. & DYCE, A. L. 2002. Field observations on the phlebotomine sand fly *Australophlebotomus mackerrasi* Lewis and Dyce feeding on the Gidgee skink *Egernia stokesii* Gray. *Parasitology Research,* 88**,** 278-279.

STOW, A. J. 2002. Microsatellite loci from the Cunningham's skink (*Egernia cunninghami*). *Molecular Ecology Notes,* 2**,** 256-257.

STRAND, T. M., SEGELBACHER, G., QUINTELA, M., XIAO, L., AXELSSON, T. & HÖGLUND, J. 2012. Can balancing selection on MHC loci counteract genetic drift in small fragmented populations of black grouse? *Ecology and Evolution,* 2**,** 341-353.

STRANDH, M., WESTERDAHL, H., PONTARP, M., CANBACK, B., DUBOIS, M.-P., MIQUEL, C., TABERLET, P. & BONADONNA, F. 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 of London, Series B: Biological Sciences,* 279**,** 4457-4463.

SUTTON, J. T., NAKAGAWA, S., ROBERTSON, B. C. & JAMIESON, I. G. 2011. Disentangling the roles of natural selection and genetic drift in shaping variation at MHC immunity genes. *Molecular Ecology,* 20**,** 4408-4420.

TAYLOR, S. S., JENKINS, D. A. & ARCESE, P. 2012. Loss of MHC and neutral variation in Peary caribou: genetic drift is not mitigated by balancing selection or exacerbated by MHC allele distributions. *PloS one,* 7**,** e36748.

TELFORD, S. R. & STEIN, J. 2000. Two malaria parasites (Apicomplexa: Plasmodiidae) of the Australian skink *Egernia stokesii*. *The Journal of Parasitology,* 86**,** 395-406.

TOBLER, M., PLATH, M., RIESCH, R., SCHLUPP, I., GRASSE, A., MUNIMANDA, G. K., SETZER, C., PENN, D. J. & MOODLEY, Y. 2014. Selection from parasites

favours immunogenetic diversity but not divergence among locally adapted host populations. *Journal of Evolutionary Biology,* 27**,** 960-974.

VAN OOSTERHOUT, C., HUTCHINSON, W. F., WILLS, D. P. M. & SHIPLEY, P. 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes,* 2004**,** 535-538.

WANG, J. 2002. An estimator for pairwise relatedness using molecular markers. *Genetics,* 160**,** 1203-1215.

WANG, J. 2011. COANCESTRY: a program for simulating, estimating and analysing relatedness and inbreeding coefficients. *Molecular Ecology Resources,* 11**,** 141-145.

WEDEKIND, C., SEEBECK, T., BETTENS, F. & PAEPKE, A. J. 1995. MHC-dependent mate preference in humans. *Proceedings of the Royal Society of London, Series B: Biological Sciences,* 260**,** 245-249.

WEGNER, K. M., REUSCH, T. B. H. & KALBE, M. 2003. Multiple parasites are driving major histocompatability complex polymorphism in the wild. *Journal of Evolutionary Biology,* 16**,** 224-232.

ZEISSET, I. & BEEBEE, T. J. 2014. Drift rather than selection dominates MHC class II allelic diversity patterns at the biogeographical range scale in natterjack toads *Bufo calamita*. *PloS one,* 9**,** e100176.

Chapter 8: Discussion

Chapter 7 presents evidence that selection on *E. stokesii* MHC can outweigh the effects of genetic drift and limited gene flow at a fine spatial scale. In addition, results demonstrate that adaptive genetic variation may have a different spatial distribution to neutral genetic variation.

Chapters 2-7 present findings in relation to *E. stokesii* MHC variation. A short discussion and conclusion summarising the key findings of this work and recommending areas for further research are presented in the following, final chapter.

Discussion

This thesis documents facets of immune gene variation in the group-living lizard *Egernia stokesii* (gidgee skink). The key outcome of this work was a comprehensive characterisation of MHC variation in a skink and an understanding of some of the processes and mechanisms that may have contributed to the observed variation. By characterising MHC variation in a poorly characterised lineage, the skinks, this work contributes to an advanced understanding of the evolutionary ecology of the MHC.

The research had two aims. First, to characterise *E. stokesii* major histocompatibility complex (MHC) variation among three discrete rocky outcrops. Sampling the study species and deriving alleles and genotypes are fundamental steps in any characterisation of genetic variation. Methods for non-invasive genetic sampling of lizards using their scats were lacking in the literature. Chapter 2 documents a reliable method of deriving lizard DNA from their scats. Within the context of lizard conservation, this method provides a viable option for sampling of threatened lizard species and has since been utilised in an investigation of the population structure of the endangered *Liopholis slateri* (Slater's skink) (Treilibs unpublished). However, application of the methodology has wider applications. For example, a study of *E. stokesii* enteric bacteria (Dodd unpublished) utilised DNA derived from *E. stokesii* scats using the method documented here. Thus, as an alternative, or complement, to traditional sampling methods, the application of the methodology derived here has promise for extending genetic data available for investigations of social structure and parasite transmission.

MHC variation is increasingly being characterised from next-generation DNA sequencing. Yet detailed methodology is generally lacking. Chapter 3 documents a comprehensive methodology for deriving MHC alleles and genotypes from next-generation DNA sequencing

data. This document should prove useful for ecologists adopting the same or similar approaches for other study species. The alleles and genotypes derived using this methodology represent the most comprehensive characterisation of skink MHC to date, thereby providing a previously rare opportunity for reliable population level skink MHC variation inferences and comparisons. However, is *E. stokesii* MHC variation typical of skinks? Characterisation of the MHC and population level MHC variation in other skinks and intermediate lizards is required to extend the molecular data available from this work and permit comparative studies. In addition, future work should include both MHC I and II, and a greater proportion of MHC II.

The second aim of this thesis was to identify processes and mechanisms contributing to *E. stokesii* MHC variation. MHC alleles and genotypes provide the foundation for such investigations. Selection is purported to be the key driver of MHC variation, yet selection on skink MHC had been untested. Chapter 4 details evidence for selection on the MHC in a lineage not previously assessed and augments other studies finding a lack of concordance between non-mammalian positively selected MHC amino acid sites and the peptide binding regions of human MHC. This work provides a foundation for comparing lizards and other taxa for which the characterisation of MHC variation and selection in additional skinks and other lizard taxa is essential. Yet, how does the overall structure of lizard MHC compare to that in mammals? The findings of this work strengthen previous calls for a greater understanding of the structure of reptilian MHC. The use of alleles and genotypes derived from DNA sequences corresponding to high coverage of target MHC regions, once structure is known, is necessary. An improved knowledge of reptilian MHC structure is also likely to enable the assignment of alleles to specific loci, which will improve future analyses.

The characterisation of MHC variation permits the investigation of processes and mechanisms generating the observed variation. Although social systems influence MHC variation, the effect of social structure on MHC variation is poorly understood. An important aspect of social structure is space use. Site fidelity has previously been demonstrated in one *E. stokesii* population over six years. During field surveys undertaken for this thesis, almost 50 lizards sampled from three sites had been captured nearly ten years before. Of these lizards, 65% were recaptured in the same space they previously occupied and 36% were found sharing space with the same other lizard or lizards they originally shared their space with (Chapter 5). Larger lizards were more likely to be recaptured, which suggests a fitness benefit related to body size. This work gives rise to several questions. For example, why have some lizards remained in the same space while others have moved? As an extension, what are the mechanisms of social group formation and maintenance? Is acceptance into a social group dependent upon the genetic diversity of the incoming individual? If so, what is the relative importance of MHC variation and other factors (e.g. body size) in securing and maintaining a position within a social group?

Along with parasite mediated selection, MHC based mate choice is a key driver of MHC variation. Social structure can constrain mate choice. Comparative studies of mate choice across taxa representing a range of social structures would enhance our currently limited understanding of social structure effects on MHC variation. High levels of social group, mate, and site fidelity probably constrain *E. stokesii* mate choice. Chapter 6 documents the finding that group membership together with genetic variables predict *E. stokesii* reproductive pairings, which are likely to influence MHC variation in this species. Group membership and genes were a stronger predictor of pairs than genes alone. Results indicated both male and female directed mate choice and identified preferences for both adaptive and neutral genetic variation. Results highlight the complexity of factors that may predict mate choice and

influence MHC variation. Further to aspects of site fidelity documented in Chapter 5, this work raises additional questions about group formation. For example, are mate decisions made before, or after, group formation? In addition, how is the gene profile of an individual detected, given genes appear to play a role in mate choice?

Genetic drift and gene flow also influence MHC variation, not just selection alone. Although these processes effect MHC variation at small spatial scales, investigations are generally undertaken at broad scales. Ecological processes operate at fine spatial scales. Thus, investigations of the generation of MHC variation should be undertaken at comparable scales. Chapter 7 details evidence that selection has outweighed the effects of genetic drift and limited gene flow on *E. stokesii* MHC variation within three geographically close yet discrete outcrops. This work demonstrates that structuring of neutral genetic variation need not imply that adaptive genetic variation is similarly structured even at a fine spatial scale. Thus, both forms of genetic variation should be accounted for in assessments of species genetic diversity. However, are the observed patterns of *E. stokesii* MHC variation consistent over space and time? Future investigations into the processes and mechanisms generating MHC variation should incorporate a range of spatial and temporal scales. In addition, the inclusion of parasite data are likely to yield useful insights into the mode of selection on the MHC.

This thesis documents facets of MHC variation and selection in a group living skink, *Egernia stokesii*. Parasites are widely considered a key cost of sociality. Living in a group may increase the rate of parasites transmission, giving rise to the general prediction that selection on the MHC is stronger in social species compared to solitary species. Yet, our understanding of the effect of group living on MHC variation is inadequate. Members of the *Egernia* group of lizards represent a diversity of social structures and mating systems. Therefore, the *Egernia* lizards are suitable candidates for future investigations of the influence of group

living on the MHC. The work of this thesis contributes to the foundation required for such investigations.

Haikus

The following presents each chapter of this thesis in the form of a haiku.

Diverse immune genes

Driven by selection and

Desired in a mate.

Chapter 2

Lizard DNA from poo -

a method that can tell you

exactly who's who!

Chapter 3

How to genotype

immune genes: documented

in detail for you!

Chapter 4

Gidgee immune genes -

highly variable and

under selection.

Chapter 4

Skink immune hotspots

compared to mammals: some same

and some different.

Chapter 5

Larger lizards more

likely to be recaptured

a decade later.

Chapter 5

Long term social bonds

and site fidelity too -

stable gidgee skinks!

Chapter 6

Social group and genes,

more than social group alone

makes a gidgee pair!

Chapter 7

Skink immune genes are

similar among sites: it's

selection at work!

Appendix 1: An introduction to *Egernia stokesii*

Egernia stokesii (gidgee skink, J.E. Gray, 1845) is a large (180 mm snout-vent length) , viviparous skink that is widely distributed across eastern and central areas of semi-arid Australia (Cogger, 1983). *Egernia stokesii* are habitat specialists, usually residing in crevices within rocky outcrops (Cogger, 1983). However they are also found within tree hollows and among fallen logs (reviewed in Chapple, 2003). Although geographically widespread, *E. stokesii* fine scale distribution appears constrained by crevice availability (Gardner et al., 2001). However, *E. stokesii* aggregate voluntarily, which suggests that it is not habitat alone that influences group spatial structure (Lanham, 2001).

Egernia stokesii form stable family groups (Gardner et al., 2001) ranging in size from two to 16 (Gardner et al., 2007). Groups comprise of one or more males, one or more females, and multiple age classes (Gardner et al., 2007, Duffield and Bull, 2002). Individuals within groups show higher levels of relatedness than individuals among groups (Gardner et al., 2001). Thermoregulation and predator avoidance have been proposed as mechanisms generating group living in *E. stokesii* (Lanham and Bull, 2004). In turn, crevice sharing may have contributed to the evolution of kin recognition mechanisms (Bull et al., 2000, Main and Bull, 1996). However, not all *E. stokesii* individuals live in a group. Approximately 10% do not belong to a group (Godfrey et al., 2006), the reason for which are currently unknown.

Egernia stokesii form fragmented social networks (Godfrey et al., 2009), which may reflect social structure and habitat availability. *Egernia stokesii* demonstrate kin recognition, presumably based on chemical cues (Main and Bull, 1996). *Egernia stokesii* deposit scats outside their crevices (Duffield and Bull, 1998). Earlier work has shown they are able to discriminate between scats of group and non-group members, suggesting pheromone cues may play a role in social group cohesion (Bull et al., 2000).

Both infanticide and maternal assistance of newborns has been demonstrated in this species (Lanham and Bull, 2000). Tolerance of other individuals is assumed given the presence of numerous *E. stokesii* within a single crevice (Duffield and Bull, 2002). However bite marks have occasionally been observed (Johnston, unpublished, Pearson, pers. obs.). Bite size and shape has been found to be consistent with *E. stokesii* skull morphology (Hutchinson, pers. comm.), suggesting some level of aggression, competition, or lack of tolerance exists within this species.

Egernia stokesii are long lived (25 years, Swan, 1990), reaching maturity at about six years (Duffield and Bull, 2002), after which they exhibit high levels of social and genetic monogamy (males 88.9%, females 63.6%) (Gardner et al., 2001, Gardner et al., 2002). They are active between September and April; mating probably occurs during October-November and birthing during February-March (Duffield and Bull, 1996a). In a laboratory some females failed to produce a litter, suggesting they may not reproduce very year study (Duffield and Bull, 1996a). Kin based behaviour avoidance, rather than sex-biased dispersal, appears to be the mechanism by which inbreeding is avoided (Gardner et al., 2012). Within a rocky outcrop, *E. stokesii* dispersal appears to be low, constrained to within a few social groups from the natal crevice (Gardner et al., 2001; Pearson unpublished). Despite an earlier study suggesting male-biased dispersal within one *E. stokesii* population (Gardner et al., 2001) further work using additional populations failed to find evidence of sex-biased dispersal (Gardner et al., 2012). The results of an earlier five year study of *E. stokesii* identified only one lizard moved between populations (unpublished data).

Egernia stokesii are genetically structured for neutral genetic markers at the population and social group level, probably reflecting both social structure and habitat constraints (Gardner et al., 2007; Pearson unpublished). The genetic structure of a population reflects species dispersal and gene flow in space and time. Using nine microsatellite loci, both Bayesian

analyses (Structure v2.3.3; Pritchard et al., 2000) and Discriminant Analysis of Principal Components (Jombart, 2008) indicated *E. stokesii* within seven populations could be separated into two genetic clusters (Pearson unpublished).

Egernia stokesii harbour numerous ecto- and endo-parasites including nematodes (*Pharyngodon tiliquae* and *Thelandros trachysauri*; Hallas et al., 2005), ticks (*Amblyomma vikirri* and *A. limbatum*; Keirans et al., 1996, Duffield and Bull, 1996b), blood parasites (*Plasmodium circularis*, *P. mackerrasae*, Telford and Stein, 2000, *Hemolivia*, *Schellackia*, and *Hepatozoon*, Stein, 1999), and a sandfly (*Australophlebotomus mackerrasi*, Stein and Dyce, 2002). *Egernia stokesii* parasite ecology was investigated by Stein (1999) who found, in most cases, blood parasite prevalence differed significantly among populations and prevalence was generally consistent within a population across years. The number of blood parasite species infecting *E. stokesii* (i.e. mixed infection) also differed significantly among populations (Stein, 1999). Tick abundance also varied at the population level and ticks were absent altogether in one of seven populations (Stein, 1999). Although parasite prevalence did not differ significantly between males and females, prevalence was positively correlated with weight (Stein, 1999). Within group infections of *Hemolivia* and *Schellackia*, both of which are directly transmitted, have been found to be higher than infections of *Plasmodium* which is indirectly transmitted (Godfrey et al., 2006).

Although numerous *E. stokesii* parasites have been identified, very little is known of their effect on their host. Lizard basking time and movements have been found to be influenced by nematodes, suggesting behaviour changes due to infection with potential fitness consequences (Fenner and Bull, 2008). Using social network approaches, *E. stokesii* parasite transmission networks within four populations have been investigated (Godfrey et al., 2009). Transmission networks were less fragmented than social networks (Godfrey et al., 2009). A relationship between *E. stokesii* sociality and parasite prevalence and infection was found by

Godfrey et al. (2009): lizards with more social connections had higher parasite loads. Godfrey et al. (2009) suggested that the crevice sharing behaviour of *E. stokesii* facilitates parasite transmission, even when crevice sharing is non-synchronous.

To date, comprehensive investigations of *E. stokesii* have been undertaken within seven populations found within a 40km2 area near Hawker within the southern Flinders Ranges of South Australia (31°54′S; 138°25′E) (references herein). A study of ~40 *E. stokesii* has also been undertaken in the northern Flinders Ranges of South Australia (Pearson, unpublished data).

References

BULL, C. M., GRIFFITH, S. C., LANHAM, E. J. & JOHNSTON, G. R. 2000. Recognition of pheromones from group members in a gregarious lizard, *Egernia stokesii*. *Journal of Herpetology,* 34**,** 92-99.

CHAPPLE, D. G. 2003. Ecology, life-history, and behavior in the Australian scincid genus *Egernia*, with comments on the evolution of complex sociality in lizards. *Herpetological Monographs,* 17**,** 145-180.

COGGER, H. G. 1983. *Reptiles and amphibians of Australia*, AH & AW Reed Pty Ltd.

DUFFIELD, G. A. & BULL, C. M. 1996a. Characteristics of the litter of the gidgee skink, *Egernia stokesii*. *Wildlife Research,* 23**,** 337-342.

DUFFIELD, G. A. & BULL, C. M. 1996b. Host location by larvae of the reptile tick *Amblyomma vikirri* (Acari: Ixodidae). *Experimental & Applied Acarology,* 20**,** 575-582.

DUFFIELD, G. A. & BULL, C. M. 1998. Seasonal and ontogenetic changes in the diet of the Australian skink *Egernia stokesii*. *Herpetologica,* 54**,** 414-419.

DUFFIELD, G. A. & BULL, C. M. 2002. Stable aggregations in an Australian lizard, *Egernia stokesii*. *Naturwissenschaften,* 89**,** 424-427.

FENNER, A. L. & BULL, C. M. 2008. The impact of nematode parasites on the behaviour of an Australian lizard, the gidgee skink *Egernia stokesii*. *Ecological Research,* 23**,** 897-903.

GARDNER, M. G., BULL, C. M., COOPER, J. B. & DUFFIED, G. A. 2001. Genetic evidence for a family structure in stable social aggregations of the Australian lizard *Egernia stokesii*. *Molecular Ecology,* 10**,** 175-183.

GARDNER, M. G., BULL, C. M. & COOPER, S. J. B. 2002. High levels of genetic monogamy in the group-living Australian lizard *Egernia stokesii*. *Molecular Ecology,* 11**,** 1787-1794.

GARDNER, M. G., BULL, C. M., FENNER, A., MURRAY, K. & DONNELLAN, S. C. 2007. Consistent social structure within aggregations of the Australian lizard, *Egernia stokesii* across seven disconnected rocky outcrops. *Journal of Ethology,* 25**,** 263-270.

GARDNER, M. G., GODFREY, S. S., FENNER, A. L., DONNELLAN, S. C. & BULL, C. M. 2012. Fine-scale spatial structuring as an inbreeding avoidance mechanism in the social skink *Egernia stokesii*. *Australian Journal of Zoology,* 60**,** 272-277.

GODFREY, S. S., BULL, C. M., JAMES, R. & MURRAY, K. 2009. Network structure and parasite transmission in a group living lizard, the gidgee skink, *Egernia stokesii*. *Behavioral Ecology and Sociobiology,* 63**,** 1045-1056.

GODFREY, S. S., BULL, C. M., MURRAY, K. & GARDNER, M. G. 2006. Transmission mode and distribution of parasites among groups of the social lizard *Egernia stokesii*. *Parasitology Research,* 99**,** 223-230.

HALLAS, G., BULL, C. M. & BURSEY, C. R. 2005. *Pharyngodon tiliquae* and *Thelandros trachysauri* (Nematoda: Pharyngodonidae), new parasite records for *Egernia stokesii* (Scincidae) from Australia. *Comparative Parasitology,* 72**,** 119-120.

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

KEIRANS, J. E., BULL, C. M. & DUFFIED, G. A. 1996. *Amblyomma vikkiri* n. sp. (Acari: Ixodida: Ixodidae), a parasite of the gidgee skink *Egernia stokesii* (Reptilia: Scincidae) from South Australia. *Systematic Parasitology,* 34**,** 1-9.

LANHAM, E. J. 2001. *Group-living in the Australian skink, Egernia stokesii.* Honours thesis, Flinders University of South Australia.

LANHAM, E. J. & BULL, C. M. 2000. Maternal care and infanticide in the Australian skink, *Egernia stokesii*. *Herpetological Review,* 31**,** 151-152.

LANHAM, E. J. & BULL, C. M. 2004. Enhanced vigilance in groups in *Egernia stokesii*, a lizard with stable social aggregations. *Journal of Zoology (London),* 263**,** 95-99.

MAIN, A. R. & BULL, C. M. 1996. Mother-offspring recognition in two Australian lizards, *Tiliqua rugosa* and *Egernia stokesii*. *Animal Behaviour,* 52**,** 193-200.

PRITCHARD, J. K., STEPHENS, M. & DONNELLY, P. 2000. Inference of population structure using multilocus genotype data. *Genetics,* 155**,** 945-959.

STEIN, J. 1999. *An ecological study of the blood parasites of E. stokesii.* PhD, Dissertation, Flinders Universtiy of South Australia.

STEIN, J. & DYCE, A. L. 2002. Field observations on the phlebotomine sand fly *Australophlebotomus mackerrasi* Lewis and Dyce feeding on the Gidgee skink *Egernia stokesii* Gray. *Parasitology Research,* 88**,** 278-279.

SWAN, G. 1990. *A field guide to the snakes and lizards of New South Wales*, Three Sisters Productions.

TELFORD, S. R. & STEIN, J. 2000. Two malaria parasites (Apicomplexa: Plasmodiidae) of the Australian skink *Egernia stokesii*. *The Journal of Parasitology,* 86**,** 395-406.

Appendix 2: Full version as published of Chapter 2

CSIRO PUBLISHING

Australian Journal of Zoology, 2014, **62**, 507–514
http://dx.doi.org/10.1071/ZO14059

Piles of scats for piles of DNA: deriving DNA of lizards from their faeces

S. K. Pearson[A,C], *S. S. Tobe*[A], *D. A. Fusco*[A], *C. M. Bull*[A] *and M. G. Gardner*[A,B]

[A]School of Biological Sciences, Flinders University of South Australia, Bedford Park, SA 5042, Australia.
[B]Evolutionary Biology Unit, South Australian Museum, North Terrace, Adelaide, SA 5000, Australia.
[C]Corresponding author. Email: sarah.pearson@flinders.edu.au

Abstract. Non-invasive genetic sampling using scats has a well established role in conservation biology, but has rarely been applied to reptiles. Using scats from captive and wild *Egernia stokesii* (Squamata, Scincidae) we evaluated two storage and six DNA-extraction methods and the reliability of subsequent genotype and sequence data. Accurate genotype and sequence data were obtained from frozen and dried captive lizard scat DNA extracted using a QIAamp® DNA Stool Mini Kit and a modified Gentra® Puregene® method, but success rates were reduced for wild lizard scats. Wild *E. stokesii* eat more plants than their captive counterparts, possibly resulting in scat DNA extracts containing plant compounds that inhibit PCR-amplifications. Notably, reliable genotypes and sequences were obtained from wild *E. stokesii* scat DNA extracted using a Qiagen DNeasy® Plant Mini Kit, a method designed to remove plant inhibitory compounds. Results highlight the opportunity for using scat-derived DNA in lizard studies, particularly for species that deposit scats in piles.

Additional keywords: DNA extraction, *Egernia* group, faecal DNA, lizard scat, microsatellites, non-invasive sampling, PCR inhibition.

Received 1 August 2014, accepted 9 February 2015, published online 24 February 2015

Introduction

Faecal pellets (scats) are a widely used source of non-invasive genetic sampling of animals (Taberlet *et al*. 1999; Beja-Pereira *et al*. 2009) providing information on species identification and distribution (Harris *et al*. 2010), individual identity (Brinkman *et al*. 2010), and relatedness between individuals within a population (Stenglein *et al*. 2011). At the population level, genetic data derived from scats have shed light on population history, structure, and genetic diversity (Frantz *et al*. 2003; Iyengar *et al*. 2005). Scat collection is less intrusive than most traditional methods used to extract DNA from collected tissue. It causes less stress and less disruption of normal behaviour to the study individuals (Taberlet and Luikart 1999), and may be less demanding in terms of field time, collection effort, equipment and costs (Solberg *et al*. 2006; Vynne *et al*. 2012). For secretive species that are hard to locate or catch, non-invasive genetic sampling using scats may be the only viable option (e.g. Alacs *et al*. 2003). For threatened species, the use of scats can provide a means to overcome collecting permit restrictions.

However, there are specific problems in deriving donor DNA from scats, as they may contain many other components including exotic DNA from food remains and gut parasites (Morin *et al*. 2001; Broquet *et al*. 2007; Marrero *et al*. 2009). Extraction and amplification of the DNA of the scatting individual may be inhibited by this accompanying material (Marrero *et al*. 2009; Panasci *et al*. 2011). Additionally, sample age and environmental conditions since the time of scat deposition can result in degradation of the DNA (Piggott 2004; Murphy *et al*. 2007; Panasci *et al*. 2011). Despite the challenges posed by the low quantity and quality of DNA, the use of scats for genetic data now has a well established place in ecological studies.

In studies of reptiles, scats have been widely used to derive non-genetic data on diet (Barrows 2006; Germano *et al*. 2007; Pavey *et al*. 2010), species distribution and abundance (Turner and Medica 1982), recognition and communication (Bull *et al*. 1999; Wilgers and Horne 2009), parasite infections (Fenner and Bull 2008; Smith *et al*. 2009), and territoriality (Wilgers and Horne 2009). However, despite its wide application in studies of mammals, the use of scats as a DNA source for genetic studies of reptiles is limited to only a single published study in snakes (Jones *et al*. 2008), with none in lizards. One explanation may be that lizard scats may contain fewer cells from the scatting individual and lower DNA yields than mammal scats. In mammal scats, donor DNA is found in a mucous layer of colorectal epithelial cells that have collected on the surface of the scat as it moves through the digestive tract (Waits and Paetkau 2005; Ball *et al*. 2007; Hebert *et al*. 2011). Amplification of target DNA is more successful using the outer coating of scats than material from inside the scats (Wehausen *et al*. 2004), and the scat coating is regularly targeted for DNA extraction in mammal studies (Piggott and Taylor 2003; Ball *et al*. 2007; Hebert *et al*. 2011). Lizard scats appear to have a reduced mucosal coating.

Journal compilation © CSIRO 2014

Despite the potential challenges, the scarcity of studies using DNA derived from lizard scats highlights an opportunity to develop this non-invasive genetic sampling method for this animal group. An important component of this process is to determine the best methods for storing scat samples, and extracting DNA from them, to maximise the yield and quality of DNA for genotyping and sequencing analyses.

Published studies in which DNA has been extracted from scats reveal a range of methods for scat storage. These include freezing (Nagy 2010), drying (Nsubuga *et al*. 2004), and storage in a buffer (Frantz *et al*. 2003). Methods to extract DNA from scats also vary. For example there are off-the-shelf scat DNA extraction kits (Stenglein *et al*. 2011; Watts *et al*. 2011), or scats can be treated using blood or tissue DNA extraction kits (Brinkman *et al*. 2010; Harris *et al*. 2010). It appears that no one method of both storage and extraction suits all species, and Valiere *et al*. (2007) and Renan *et al*. (2012) recommended a pilot study be undertaken to identify optimal methods for each new study species. We investigated alternative methods for storing scats and deriving DNA from them for an Australian scincid lizard, *Egernia stokesii*. Our objectives were to: (1) identify optimal *E. stokesii* scat storage and DNA extraction methods, and (2) assess the reliability of DNA genotypes and sequences derived from *E. stokesii* scats using these methods. Once developed, these methods could complement traditional invasive sampling methods in this and other lizard species.

Materials and methods

Study species

Egernia stokesii (J. E. Gray, 1845) (gidgee skink) is a large (180 mm snout–vent length: Cogger 1983), long-living, viviparous skink (Duffield and Bull 2002) widely distributed across eastern and central areas of semiarid Australia. *Egernia stokesii* individuals live in stable family groups (Gardner *et al*. 2001*a*; Duffield and Bull 2002), have high levels of genetic monogamy (Gardner *et al*. 2002) and limited dispersal (Gardner *et al*. 2001*b*). They produce scats on rock platforms immediately outside of the rocky crevices in which they reside, resulting in distinctive scat piles or deposits (Duffield and Bull 1998). Using olfaction, individuals can discriminate between scats from familiar group and non-group members, suggesting that scat piles play an important role in social group cohesion in this species (Bull *et al*. 1999). The use of scat-derived DNA in this and similarly scat-piling lizard species could provide quick access to the DNA of most group members, without the time and effort required to capture the lizards for tissue samples. Additionally, collecting scat may provide a more complete genetic sampling of social groups as some individuals may not be caught.

Scat sampling

Scats were sampled from two sources. First we used captive *E. stokesii* housed at the Flinders University of South Australia. These included individuals originally captured near Hawker (31°54′S, 138°25′E) in the southern Flinders Ranges, South Australia, during the summers of 1993–98, and their progeny (Main and Bull 1996; Lanham and Bull 2004; Arida and Bull 2008). We randomly selected nine *E. stokesii* individuals and kept them in nine separate cages so scats could be confidently assigned to an individual. Each individual was housed in a cage (40 cm high × 40 cm wide × 50 cm deep) in a room with a temperature of 25°C (±2°C), with ceiling lights on for 12 h per day, and heat lamps on for 6 h per day. Scats were collected twice weekly for four weeks (total 128 scats, average 14.22 scats per lizard, s.e. ± 2.13). Second, scats were collected within an estimated 4 h of defaecation during field surveys of three *E. stokesii* populations near Hawker conducted between September 2012 and March 2013 (409 scats). Scat freshness was assessed on the basis of colour, moisture, compaction, by the presence of a uric acid spot, and by comparison with scats of known age from the captive colony. In addition, some lizards captured during surveys defaecated during handling, ensuring complete freshness of the scat samples.

In each case, scats were collected using forceps that had been cleaned in 90% ethanol between each collection, and were stored using alternative methods as described below. The diet of captive and wild *E. stokesii* differed. Captives of all ages were fed a mix of boiled eggs, fruits and vegetables, and reptile supplement, while adult wild *E. stokesii* feed largely on plant material (Duffield and Bull 1998). On the basis of their size, we deduced that the wild scats used in this study were from adult *E. stokesii* (Duffield and Bull 1998); this was confirmed by visual inspection that showed a largely plant-derived content of wild scats.

Positive controls

To confirm that DNA derived from scats accurately represented the DNA of the scatting individual we collected blood samples from some individuals as an alternative source for DNA characterisation. Blood (up to 0.5 mL) was taken from the caudal vein of the nine isolated captive individuals and from 29 wild individuals that produced a scat while captured. Blood was stored on Whatman FTA® Elute for later DNA extraction. We used established methods for deriving mitochondrial DNA (mtDNA) sequences and microsatellite DNA genotypes from FTA-stored *E. stokesii* blood (Gardner *et al*. 2007).

Scat storage

We compared two methods of storage for the captive lizard scats. Scats were either frozen at –20°C (Frantz *et al*. 2003) (27 scats; all from the laboratory colony), or dried (72 scats; 54 laboratory colony, 18 field). Scats to be dried were sprayed with 90% ethanol and then stored on silica beads (hereafter termed 'dried'; modified from Roeder *et al*. 2004) and kept at room temperature until DNA extraction. Samples from the field were all stored dried as this method was considered more practical for sampling in extreme conditions and away from amenities.

DNA extraction of scats from captive lizards

We trialled six DNA extraction methods using 81 scats (27 frozen, 54 dried) from captive lizards (Table 1): (1) QIAamp® DNA Stool Mini Kit (QIAGEN®, Catalogue 51504); (2) ISOLATE Fecal DNA Kit (Bioline®, Catalogue BIO-52037); (3) a standard Chelex® 100 extraction; (4) Chelex® 100 without boiling (adopted from Casquet *et al*. 2011); (5) a modified Gentra® Puregene® (Gentra Systems) method; and (6) a direct PCR method. The first two methods were off-the-shelf kits specifically

Table 1. Results of scat trials with captive *E. stokesii*
Showing the number of scats used in trial for each storage and DNA extraction method, and numbers (%) successfully amplified for mt- and nuDNA for each storage and DNA extraction method

Extraction method	Storage method		mtDNA amplified		nuDNA amplified	
	Frozen[A]	Dried[B]	Frozen	Dried	Frozen	Dried
QIAamp® DNA Stool Mini Kit	11	20	10 (91)	20 (100)	9 (82)	17 (85)
ISOLATE Fecal DNA Kit	4	16	3 (75)	11 (69)	1 (25)	4 (25)
Direct	3	3	0	0	n.a.	n.a.
Chelex with boiling	3	3	0	0	n.a.	n.a.
Chelex without boiling	3	3	0	0	n.a.	n.a.
Modified Gentra® Puregene® method	3	9	3 (100)	9 (100)	3 (100)	9 (100)
Total	27	54	16	40	13	30

[A]Frozen: −20°C.
[B]Dried: sprayed with 90% ethanol then stored on silica beads at room temperature.

derived for scat samples. The next three were standard kit methods used for tissue or blood samples. The last method involved amplification without first extracting or purifying the DNA and allows for maximum recovery of sample, but can suffer from inhibitors that are normally removed during extraction. The two kits (QIAamp and ISOLATE) were used according to manufacturers' guidelines. The modified Gentra® Puregene® method involved immersion of the whole scat in SLP buffer (500 mM TRIS-HCl pH 8.0, 50 mM EDTA, 10 mM NaCl (modified from Deuter *et al.* 1995)) followed by agitation on a rotor wheel for 1 h, protein precipitation with Proteinase K, DNA precipitation with ammonium acetate and isopropanol, ethanol wash, and DNA hydration in TLE buffer.

For the kit extractions, scats were selected according to recommended weight ranges where possible; *E. stokesii* scats ranged in weight from ~10 to 900 mg (average 199.60 mg, s.e. ±16.63), therefore total weight may have been outside the recommended range (180–220 mg QIAamp® DNA Stool Mini Kit, up to 150 mg ISOLATE Fecal DNA Kit). Where a scat was large enough, a surface scrape of the scat was used in kit extractions as this is where most of the donor individual's DNA is expected to be found. Alternatively, a segment of the scat, or the entire scat, was used, depending upon the protocol. In all methods, filtered pipette tips were used to minimise contamination and negative extraction controls (scat material was not added to the extraction) were used to assess contamination. Separate laboratories were used for extraction, amplification preparation and reaction. Replicate scat extractions are sometimes recommended (Taberlet *et al.* 1999) but this was not possible as a single extraction often required the whole scat to be used.

DNA amplification in captive lizards

Initially, mtDNA was targeted in DNA amplification trials because cells contain more mtDNA than nuclear DNA (nuDNA), suggesting that if mtDNA could not be amplified then targeting nuDNA was likely to be futile (Taberlet *et al.* 1999). However, amplification success may be increased for smaller DNA markers (Broquet *et al.* 2007). Because only larger mtDNA genetic markers (~800 bp) were currently available for *E. stokesii*, and because scat DNA may be of low quantity and quality (Navidi *et al.* 1992; Taberlet *et al.* 1996), we developed genetic markers to amplify ~200 bp of the mtDNA ND4 gene. Three primer pairs were designed in Geneious 5.6 (http://www.geneious.com; Kearse *et al.* 2012) based on a consensus sequence derived from 159 existing *E. stokesii* mtDNA sequences. These primer pairs were trialled in DNA derived from *E. stokesii* blood; forward primer M1544 (5′-TATGAACGCACCCATAGCCG-3′) and reverse primer M1545 (5′-GCTGCTGTTAGAAGAGTGCC-3′) were selected for this study.

For mtDNA only 1 : 5 and 1 : 50 dilutions were trialled. A dilution of 1 : 5 has previously been successful for DNA from blood in this species, but we considered that overcoming inhibitors in scat DNA may require increased dilution (Monteiro *et al.* 1997; Ball *et al.* 2007; Arandjelovic *et al.* 2009). Polymerase chain reaction (PCR) amplifications were conducted at a total volume of 25 µL consisting of 1 × PCR Gold Buffer (Applied Biosystems), 0.20 µM of each primer, 0.80 mM dNTPs, 2 mM $MgCl_2$, 0.5 U AmpliTaq® Gold DNA polymerase (Applied Biosystems), 2 µL of extracted DNA, and PCR-grade water. The cycling conditions were 9 min at 95°C, 34 cycles of 45 s at 94°C, 45 s at 60°C, 1 min at 72°C, and a final elongation step of 10 min at 72°C followed by 30 s at 25°C. To ensure that non-amplification was due to the test procedure rather than a failure of the PCR, and that positive results were not the result of contamination, one PCR-positive (DNA extracted from blood and known to amplify) and two PCR-negative (TLE buffer and the negative DNA extraction) controls were used in each PCR. Neat DNA, from which mtDNA PCR-amplifications were successful, was quantified using Qubit® 2.0 Fluorometer (Life Technologies Corporation, CA), although we acknowledge that the DNA measured may have included both target and non-target DNA.

Where mtDNA amplification was successful, as determined by the presence of a band on an agarose gel, amplification trials continued using a previously developed species-specific microsatellite genetic marker (Est 1: Gardner *et al.* 1999). For nuDNA, a QIAGEN® Multiplex PCR Kit (QIAGEN®, Catalogue 206143) was used. Each 10-µL uniplex reaction mix contained 0.10 × QIAGEN Multiplex PCR Master Mix, 0.25 µM of each primer, 0.50 × Q-solution, 2 µL of extracted DNA, and RNAse Free Water. The cycling conditions were 15 min at 95°C, 35 cycles of 30 s at 94°C, 90 s at 57°C, 1 min at 72°C, and a final elongation step of 30 min at 60°C followed by 30 s at 25°C. Amplification success of the Est 1 locus was determined by the presence of a band on an agarose gel. A dilution of 1 : 50

has previously been successful for nuDNA from blood in this species. Because increased dilutions may be required to reduce the effect of inhibitors, if amplification was not successful for 1 : 50 dilutions, a range of DNA dilutions (neat, 1 : 5, 1 : 10, 1 : 100, 1 : 500, 1 : 1000) were then trialled. If Est 1 failed to amplify for any dilution, the extraction method was deemed unsuccessful for nuDNA.

DNA extraction and amplification in scats from wild lizards

Although the QIAamp® DNA Stool Mini Kit and modified Gentra® Puregene® method were successful in preliminary trials using scats from captive lizards (see Results), the modified Gentra® Puregene® method had a lower per-sample cost so we chose that method for validation using six wild scats; positive and negative controls were used as described for captive scats above. None of the six wild scat DNA extractions amplified for mtDNA (results not shown). We considered diet differences between captive and wild *E. stokesii* may explain differences in amplification success rates. Earlier studies have suggested that diet-derived inhibitors in scats may reduce both DNA extraction yields and amplification success (Kohn and Wayne 1997; Hebert *et al.* 2011; Panasci *et al.* 2011; Monroe *et al.* 2013). This problem could be particularly relevant for omnivorous or herbivorous lizards due to the presence of polysaccharides and polyphenols found in plants (Marrero *et al.* 2009; Panasci *et al.* 2011).

To investigate the possible presence of PCR inhibitors, DNA was extracted (using the modified Gentra® Puregene® method) from a further six scats from wild lizards. To test for PCR inhibition, one of the scat extractions was replicated in the PCR, once with only DNA extracted from the scat, and once with the scat DNA plus 2 µL of a positive control. We could infer that inhibitors were likely to be preventing amplification if both reactions failed. In an effort to reduce the impact of potential inhibitors, a subset of extracted DNA from each of the six wild scats was purified using Microcon Ultracel YM-100 filters. Purified extractions were then assessed for mtDNA amplification success using the reaction mix and conditions outlined above. One sample was replicated in this PCR, with one replicate spiked with control DNA to directly assess the effect of the inhibitor clean-up process (i.e. the same sample was used as in the earlier PCR).

Given the low success rates of the modified Gentra® Puregene® method on scats of wild lizards (see Results), and the additional cost and effort associated with clean-up, a Qiagen DNeasy® Plant Mini Kit was trialled for removing inhibitors. DNA was extracted from a further six scats from wild lizards according to the manufacturer's instructions except initial disruption and homogenisation of the scat sample was avoided. Instead, the scat was left intact and, where required due to the size of the scat, additional Buffer AP1 and RNAase A stock solution (100 mg mL^{-1}) were used to ensure that scats were fully immersed before incubation.

Validation via genotyping and sequencing

For captive samples, where an extraction method was successful, both scat and blood samples from a subset of two lizards were sequenced for the mtDNA and genotyped for seven previously described polymorphic microsatellite loci (Est 1, Est 4, Est 8, Est 13 (Gardner *et al.* 1999); TrL 28, TrL 29, TrL 35 (Gardner *et al.* 2008)) in PCR-amplifications according to the reaction mix and conditions described above except that reactions were performed in two multiplex reactions rather than uniplex (Multiplex 1: Est 1, Est 4, Est 8, Est 13; Multiplex 2: TrL 28, TrL 29, TrL 35). For wild samples, scat DNA of six lizards that defaecated during handling was extracted using the Qiagen DNeasy® Plant Mini Kit, and blood DNA from the same lizards, extracted using the Whatman FTA® Elute, were similarly genotyped and sequenced. Prior to sequencing, mtDNA PCR products were purified using multiscreen PCR filter plates (Millipore Billerica, MA) to remove unincorporated primers and dNTPs. Sequence reactions were prepared using a BigDYE Terminator Cycle Sequencing Kit 3.1 (Applied Biosystems) following manufacturer recommendations, using the same primers as those used in PCR amplification. The cycling conditions were 3 min at 96°C, 30 cycles of 30 s at 96°C, 15 s at 50°C, 4 min at 60°C, and a final elongation step of 3 min at 25°C followed by 30 s at 25°C. Sequence products were purified using multiscreen PCR filter plates (Millipore Billerica, MA) before submission of DNA to the Australian Genome Research Facility for capillary separation on an ABI Prism 3730*xl* 96-capillary sequencer. The resulting sequences were compared against data on GenBank, to confirm species identification, using the Basic Local Alignment Search Tool (BLAST) available at http://blast.ncbi.nlm.nih.gov/Blast.cgi. For nuDNA, we compared genotypes derived from both blood and scat samples from the same individuals. PCR products were analysed by capillary separation on an AB3730 DNA analyser (Australian Genome Research Facility) and resulting fragments were scored using GeneMapper® (Applied Biosystems). Although recommended (Taberlet *et al.* 1999; Valiere *et al.* 2007), we did not perform replicate PCRs for mtDNA or nuDNA in initial trials; instead, we used blood DNA samples from the same individuals as a positive control. At this stage we were interested in determining whether a sequence and genotype could be derived from scat DNA and whether they matched those derived from blood-derived DNA.

Assessment of genotyping reliability

Once we identified a method for deriving genotypes from scats of wild *E. stokesii* (see Results), we assessed genotyping reliability using three independent PCRs (adopted from Panasci *et al.* 2011; Stenglein *et al.* 2011). DNA amplification and genotyping were undertaken as uniplexes (see 'Validation via genotyping and sequencing' above).

Results

We used 81 scats from captive *E. stokesii* in preliminary trials (27 frozen, 54 dried) (Table 1). Both frozen and dried samples amplified for both mtDNA and nuDNA, and the amplification success rates for each method are given in Table 1. Of the six extraction methods trialled using scats from captive lizards, the QIAamp® DNA Stool Mini Kit, ISOLATE Fecal DNA Kit, and modified Gentra® Puregene® methods were successful for mtDNA (Table 1). The Chelex® 100 and direct PCR methods

failed to amplify mtDNA, and therefore were not trialled for nuDNA. Both the 1 : 5 and 1 : 50 DNA dilutions were successful for mtDNA, while for nuDNA, neat DNA was the most successful (results not shown). The QIAamp® DNA Stool Mini Kit and modified Gentra® Puregene® method were further tested for reliability of sequencing and genotyping using scats from captive *E. stokesii*. Of 20 scat DNA sequences from captive lizards derived using the QIAamp® DNA Stool Mini Kit ($n = 17$) and modified Gentra® Puregene® method ($n = 3$) assessed in BLAST, 80% ($n = 16$) were identified as *E. stokesii*, 15% ($n = 3$) as *Egernia* sp., and one sequence (5%) was too short to provide meaningful results. For nuDNA, all seven microsatellite loci were successfully derived from DNA extracted from scats of captive lizards using both the QIAamp® DNA Stool Mini Kit and the modified Gentra® Puregene® method, and all scat-derived genotypes matched those derived from blood.

On the basis of trials using scats from captive lizards, the modified Gentra® Puregene® method was initially chosen for use with six scat samples from wild lizards; mtDNA amplification was unsuccessful. In subsequent trials using a further six scat samples from wild lizards, the PCR-positive control DNA on its own was successfully amplified, but the scat sample from a wild lizard spiked with the positive control DNA failing to amplify, suggesting the presence of inhibitors. Following application of a purification method, an additional three DNA samples from scats of wild lizards were amplified for mtDNA. Notably, the elute DNA of six of the 12 extractions from wild lizard scats using the modified Gentra® Puregene® method ranged from a light tea colour to muddy brown, whereas the elute DNA of all scat extractions from captive lizards was clear.

Six scat samples from wild lizards extracted using the Qiagen DNeasy® Plant Mini Kit were successfully sequenced and genotyped. All six samples were identified as *E. stokesii* using BLAST analyses. All seven microsatellite DNA loci could be scored and the resulting genotypes matched those derived from blood. The elute DNA was clear for all scats from wild lizards extracted using the Qiagen DNeasy® Plant Mini Kit. Successful mtDNA PCR-amplification quantifications are available as Supplementary Material on the Journal website. The reliability of genotypes derived from DNA extracted using the Qiagen DNeasy® Plant Mini Kit was further assessed for seven loci in five scat samples from wild *E. stokesii*. All loci amplified in all replicates for all samples, except TrL 35, which failed in all replicates for one sample. Matching heterozygotes were observed in all replicates for most samples, with three exceptions. First, all replicates for Est 1 and TrL 28 in one sample showed matching homozygotes. Second, allelic dropout was evident for Est 1 in one sample, which showed two matching heterozygotes and one homozygote. Lastly, for TrL 28 in one sample, two replicates showed matching homozygotes, while one showed a heterozygote, suggesting either allelic dropout in two replicates or a false allele in one replicate. The allelic dropout and false allele incidences represent an overall genotyping error rate of 2%.

Discussion

We have identified a reliable method for deriving DNA sequences and genotypes from scat samples of wild *E. stokesii*. Genotypes and sequences were successfully derived from DNA extracted from field-collected scats using a Qiagen DNeasy® Plant Mini Kit. The overall reliability of genotypes derived using this method was supported by a low genotyping error rate. Adoption of this method would complement traditional capture–mark–recapture methods for estimating local abundance of *E. stokesii* and other lizard species, and for estimating genetic structure and diversity, particularly for those species that create easily sampled scat piles. Colocated scats provide greater confidence of matching of scat to lizard location, making this a potentially useful tool for assessing social structures and relatedness among social group members. In addition, this method provides an alternative, non-invasive technique for threatened or secretive lizards.

Two DNA extraction methods (QIAamp® DNA Stool Mini Kit and modified Gentra® Puregene® method) were successful for scats from captive lizards although success rates decreased when applied to scats from wild lizards. On the other hand, the Qiagen DNeasy® Plant Mini Kit successfully extracted DNA from scats of wild lizards; suggesting that plant inhibitors present in the scats of herbivorous lizards may often prevent amplification of DNA unless they are filtered out. Although both mtDNA and nuDNA were successfully amplified from frozen and dried scat samples, the drying method will be more suitable when sampling in semiarid to arid locations away from electricity supplies. As false alleles and allelic dropout may arise in scat samples with low quality and quantity of DNA (Taberlet *et al.* 1999; Broquet and Petit 2004; Valiere *et al.* 2007) error-checking protocols should normally be adopted. DNA amplification replicates and assessment using a consensus approach have previously been suggested (Navidi *et al.* 1992; Taberlet *et al.* 1999; Broquet and Petit 2004) and an assessment of power such as probability of identity is recommended (Valiere 2002). We also recommend the use of replicate PCRs to assess the reliability of genotypes derived from DNA extracted using the Qiagen DNeasy® Plant Mini Kit. In addition, as methods are not necessarily transferable between species (Taberlet *et al.* 1999), preliminary trials incorporating genotyping error rates are recommended before their use with other lizard species.

There are other potential problems when deriving lizard DNA from their scats. We found reptile scales on the surface of some scats from captive and wild lizards (Pearson, pers. obs.). Captive lizards were isolated so it could be assumed that in those cases the scales belonged to the lizard from which scats were collected. The same assumption cannot be made for field-collected scats as lizards may eat the sloughed skin of other individuals, or even conspecific neonates (Lanham and Bull 2004), potentially contaminating the sample with other conspecific DNA. Further, the colocation of scats may result in cross contamination between scats that are in contact but from different individuals. Also, DNA extraction and amplification success is likely to decline with scat age as the DNA deteriorates (DeMay *et al.* 2013). Samples from wild lizards used in this study were fresh; we therefore recommend that future studies consider temporal sampling thresholds.

A further potential complication concerns the identification of scats from the target species. In this study, few other lizard species were sighted during the sampling of wild *E. stokesii* and the size and location of *E. stokesii* scats in piles immediately

outside occupied crevice entrances facilitated identification. However, geckos were present and gecko scats may be confused with scats of subadult *E. stokesii*, although species identification may be verified via sequencing. In addition, the field sites used in this study consisted of rocky outcrops with sparse vegetation where scats were easily found. Scats may be harder to locate and identify in an area with denser vegetation or higher lizard diversity. Knowledge of the behaviour of the target species and an awareness of other resident and transient species would be essential in such cases. However, this study indicates that more confidence may be applied to the identification of species from their scat in scat-piling species, making non-invasive genetic sampling particularly applicable for such species.

Acknowledgements

We thank Leslie Morrison and staff of the Flinders University of South Australia animal house for assistance with captive lizards. This work was part funded by the Royal Society of South Australia, the Biology Society of South Australia, the Nature Conservation Society of South Australia, and the Field Naturalist Society of South Australia. Captive and wild lizards were managed in accordance with standards and procedures in compliance with the Australian Code of Practice for the use of animals for scientific purposes. We also thank two anonymous reviewers and Steve Cooper for their helpful comments on earlier versions of the manuscript.

References

Alacs, E., Alpers, D., de Tores, P. J., Dillon, M., and Spencer, P. B. S. (2003). Identifying the presence of quokkas (*Setonix brachyurus*) and other macropods using cytochrome b analyses from faeces. *Wildlife Research* **30**, 41–47. doi:10.1071/WR01109

Arandjelovic, M., Guschanski, K., Schubert, G., Harris, R., Thalmann, O., Siedel, H., and Vigilant, L. (2009). Two-step multiplex polymerase chain reaction improves the speed and accuracy of genotyping using DNA from noninvasive and museum samples. *Molecular Ecology Resources* **9**, 28–36. doi:10.1111/j.1755-0998.2008.02387.x

Arida, E. A., and Bull, C. M. (2008). Optimising the design of artificial refuges for the Australian skink, *Egernia stokesii. Applied Herpetology* **5**, 161–172. doi:10.1163/157075408784648826

Ball, M. C., Pither, R., Manseau, M., Clark, J., Petersen, S. D., Kingston, S., Morrill, N., and Wilson, P. (2007). Characterisation of target nuclear DNA from faeces reduces technical issues associated with the assumptions of low-quality and quantity template. *Conservation Genetics* **8**, 577–586. doi:10.1007/s10592-006-9193-y

Barrows, C. W. (2006). Population dynamics of a threatened sand dune lizard. *The Southwestern Naturalist* **51**, 514–523. doi:10.1894/0038-4909(2006)51[514:PDOATS]2.0.CO;2

Beja-Pereira, A., Oliveira, R., Alves, P. C., Schwarz, M. K., and Luikart, G. (2009). Advancing ecological understanding through technological transformations in noninvasive genetics. *Molecular Ecology Resources* **9**, 1279–1301. doi:10.1111/j.1755-0998.2009.02699.x

Brinkman, T. J., Person, D. K., Schwartz, M. K., Pilgrim, K. L., Colson, K. E., and Hundertmark, K. J. (2010). Individual identification of Sitka black-tailed deer (*Odocoileus hemionus sitkensis*) using DNA from fecal pellets. *Conservation Genetics Resources* **2**, 115–118. doi:10.1007/s12686-010-9176-7

Broquet, T., and Petit, E. (2004). Quantifying genotyping errors in noninvasive population genetics. *Molecular Ecology* **13**, 3601–3608. doi:10.1111/j.1365-294X.2004.02352.x

Broquet, T., Menard, N., and Petit, E. (2007). Noninvasive population genetics: a review of sample source, diet, fragment length and microsatellite motif effects on amplification success and genotyping error rates. *Conservation Genetics* **8**, 249–260. doi:10.1007/s10592-006-9146-5

Bull, C. M., Griffith, S. C., and Perkins, M. V. (1999). Some properties of a pheromone allowing individual recognition, from the scats of an Australian lizard, *Egernia striolata. Acta Ethologica* **2**, 35–42. doi:10.1007/PL00012230

Casquet, J., Thebaud, C., and Gillespie, R. G. (2011). Chelex without boiling, a rapid and easy technique to obtain stable amplifiable DNA from small amounts of ethanol-stored spiders. *Molecular Ecology Resources* **2**, 136–141.

Cogger, H. G. (1983). 'Reptiles and Amphibians of Australia.' (AH & AW Reed Pty Ltd: Sydney.)

DeMay, S. M., Becker, P. A., Eidson, C. A., Rachlow, J. L., Johnson, T. R., and Waits, L. P. (2013). Evaluating DNA degradation rates in faecal pellets of the endangered pygmy rabbit. *Molecular Ecology Resources* **13**, 654–662. doi:10.1111/1755-0998.12104

Deuter, R., Pietsch, S., Hertel, S., and Muller, O. (1995). A method of preparation of fecal DNA suitable for PCR. *Nucleic Acids Research* **23**, 3800–3801. doi:10.1093/nar/23.18.3800

Duffield, G. A., and Bull, C. M. (1998). Seasonal and ontogenetic changes in the diet of the Australian skink *Egernia stokesii. Herpetologica* **54**, 414–419.

Duffield, G. A., and Bull, C. M. (2002). Stable aggregations in an Australian lizard, *Egernia stokesii. Naturwissenschaften* **89**, 424–427. doi:10.1007/s00114-002-0346-7

Fenner, A. L., and Bull, C. M. (2008). The impact of nematode parasites on the behaviour of an Australian lizard, the gidgee skink *Egernia stokesii. Ecological Research* **23**, 897–903. doi:10.1007/s11284-007-0453-1

Frantz, A. C., Pope, L. C., Carpenter, P. J., Roper, T. J., Wilson, G. J., Delahay, R. J., and Burke, T. (2003). Reliable microsatellite genotyping of the Eurasian badger (*Meles meles*) using faecal DNA. *Molecular Ecology* **12**, 1649–1661. doi:10.1046/j.1365-294X.2003.01848.x

Gardner, M. G., Cooper, S. J. B., Bull, C. M., and Grant, W. N. (1999). Isolation of microsatellite loci from a social lizard, *Egernia stokesii*, using a modified enrichment procedure. *The Journal of Heredity* **90**, 301–304. doi:10.1093/jhered/90.2.301

Gardner, M. G., Bull, C. M., Cooper, J. B., and Duffied, G. A. (2001*a*). Genetic evidence for a family structure in stable social aggregations of the Australian lizard *Egernia stokesii. Molecular Ecology* **10**, 175–183. doi:10.1046/j.1365-294X.2001.01171.x

Gardner, M. G., Bull, C. M., Cooper, S. J. B., and Duffield, G. A. (2001*b*). Genetic evidence for a family structure in stable social aggregations of the Australian lizard *Egernia stokesii. Molecular Ecology* **10**, 175–183. doi:10.1046/j.1365-294X.2001.01171.x

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**, 1787–1794. doi:10.1046/j.1365-294X.2002.01552.x

Gardner, M. G., Bull, C. M., Fenner, A, Murray, K, and Donnellan, S. C. (2007). Consistent social structure within aggregations of the Australian lizard, *Egernia stokesii* across seven disconnected rocky outcrops. *Journal of Ethology* **25**, 263–270. doi:10.1007/s10164-006-0022-z

Gardner, M. G., Sanchez, J. J., Dudaniec, R. Y., Rheinberger, L., Smith, A. L., and Saint, K. S. (2008). *Tiliqua rugosa* microsatellites: isolation via enrichment and characterisation of loci for multiplex PCR in *T. rugosa* and the endangered *T. adelaidensis. Conservation Genetics* **9**, 233–237. doi:10.1007/s10592-007-9316-0

Germano, D. J., Smith, P. T., and Tabor, S. P. (2007). Food habits of the blunt-nosed leopard lizard (*Gambelia sila*). *The Southwestern Naturalist* **52**, 318–323. doi:10.1894/0038-4909(2007)52[318:FHOTBL]2.0.CO;2

Harris, R. B., Winnie, J., Jr, Amish, S. J., Beja-Pereira, A, Godinho, R, Costa, V, and Luikart, G (2010). Argali abundance in the Afghan Pamir

using capture–recapture modeling from fecal DNA. *Journal of Wildlife Management* **74**, 668–677. doi:10.2193/2009-292

Hebert, L., Darden, S. K., Pedersen, B. V., and Dabelsteen, T. (2011). Increased DNA amplification success of non-invasive genetic samples by successful removal of inhibitors from faecal samples collected in the field. *Conservation Genetics Resources* **3**, 41–43. doi:10.1007/s12686-010-9280-8

Iyengar, A., Babu, V. N., Hedges, S., Venkataraman, B., Maclean, N., and Morin, P. A. (2005). Phylogeography, genetic structure, and diversity in the dhole (*Cuon alpinus*). *Molecular Ecology* **14**, 2281–2297. doi:10.1111/j.1365-294X.2005.02582.x

Jones, R., Cable, J., and Bruford, M. W. (2008). An evaluation of non-invasive sampling for genetic analysis in northern European reptiles. *The Herpetological Journal* **18**, 32–39.

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

Kohn, M. H., and Wayne, R. K. (1997). Facts from feces revisited. *Trends in Ecology & Evolution* **12**, 223–227. doi:10.1016/S0169-5347(97)01050-1

Lanham, E. J., and Bull, C. M. (2004). Enhanced vigilance in groups in *Egernia stokesii*, a lizard with stable social aggregations. *Journal of Zoology* **263**, 95–99. doi:10.1017/S0952836904004923

Main, A. R., and Bull, C. M. (1996). Mother–offspring recognition in two Australian lizards, *Tiliqua rugosa* and *Egernia stokesii*. *Animal Behaviour* **52**, 193–200. doi:10.1006/anbe.1996.0164

Marrero, P., Fregel, R., Cabrera, V. M., and Nogales, M. (2009). Extraction of high-quality host DNA from feces and regurgitated seeds: a useful tool for vertebrate ecological studies. *Biological Research* **42**, 147–151. doi:10.4067/S0716-97602009000200002

Monroe, C., Grier, C., and Kemp, B. M. (2013). Evaluating the efficiency of various thermo-stable polymerases against co-extracted PCR inhibitors in ancient DNA samples. *Forensic Science International* **228**, 142–153. doi:10.1016/j.forsciint.2013.02.029

Monteiro, L., Bonnemaison, D., Vekris, A., Petry, K. G., Bonnet, J., Vidal, R., Cabrita, J., and Mégraud, F. (1997). Complex polysaccharides as PCR inhibitors in feces: *Helicobacter pylori* model. *Journal of Clinical Microbiology* **35**, 995–998.

Morin, P. A., Chambers, K. E., Boesch, C., and Vigilant, L. (2001). Quantitative polymerase chain reaction analysis of DNA from noninvasive samples for accurate microsatellite genotyping of wild chimpanzees (*Pan troglodytes verus*). *Molecular Ecology* **10**, 1835–1844. doi:10.1046/j.0962-1083.2001.01308.x

Murphy, M. A., Kendall, K. C., Robinson, A., and Waits, L. P. (2007). The impact of time and field conditions on brown bear (*Ursus arctos*) faecal DNA amplification. *Conservation Genetics* **8**, 1219–1224. doi:10.1007/s10592-006-9264-0

Nagy, Z. T. (2010). A hands-on overview of tissue preservation methods for molecular genetic analyses. *Organisms, Diversity & Evolution* **10**, 91–105. doi:10.1007/s13127-010-0012-4

Navidi, W., Arnheim, N., and Waterman, M. S. (1992). A multiple-tubes approach for accurate genotyping of very small DNA samples by using PCR: statistical considerations. *American Journal of Human Genetics* **50**, 347–359.

Nsubuga, A. M., Robbins, M. M., Roeder, A. D., Morin, A., Boesch, C., and Vigilant, L. (2004). Factors affecting the amount of genomic DNA extracted from ape faeces and the identification of an improved sample storage method. *Molecular Ecology* **13**, 2089–2094. doi:10.1111/j.1365-294X.2004.02207.x

Panasci, M., Ballard, W. B., Breck, S., Rodriguez, D., Densmore, L. D. III, Wester, D. B., and Baker, R. J. (2011). Evaluation of fecal DNA preservation techniques and effects of sample age and diet on genotyping success. *Journal of Wildlife Management* **75**, 1616–1624. doi:10.1002/jwmg.221

Pavey, C. R., Burwell, C. J., and Nano, C. E. M. (2010). Foraging ecology and habitat use of Slater's skink (*Egernia slateri*): an endangered Australian desert lizard. *Journal of Herpetology* **44**, 563–571. doi:10.1670/09-102.1

Piggott, M. P. (2004). Effect of sample age and season of collection on the reliability of microsatellite genotyping of faecal DNA. *Wildlife Research* **31**, 485–493. doi:10.1071/WR03096

Piggott, M. P., and Taylor, A. C. (2003). Extensive evaluation of faecal preservation and DNA extraction methods in Australian native and introduced species. *Australian Journal of Zoology* **51**, 341–355. doi:10.1071/ZO03012

Renan, S., Speyer, E., Shahar, N., Gueta, T., Templeton, A. R., and Bar-David, S. (2012). A factorial design experiment as a pilot study for noninvasive genetic sampling. *Molecular Ecology Resources* **12**, 1040–1047. doi:10.1111/j.1755-0998.2012.03170.x

Roeder, A. D., Archer, F. I., Poinar, H. N., and Morin, P. A. (2004). A novel method for collection and preservation of faeces for genetic studies. *Molecular Ecology* **4**, 761–764. doi:10.1111/j.1471-8286.2004.00737.x

Smith, A. L., Fenner, A. L., Bull, C. M., and Gardner, M. G. (2009). Genotypes and nematode infestations in an endangered lizard, *Tiliqua adelaidensis*. *Applied Herpetology* **6**, 300–305. doi:10.1163/157075408X386150

Solberg, K. H., Bellemain, E., Drageset, O.-M., Taberlet, P., and Swenson, J. E. (2006). An evaluation of field and non-invasive genetic methods to estimate brown bear (*Ursus arctos*) population size. *Biological Conservation* **128**, 158–168. doi:10.1016/j.biocon.2005.09.025

Stenglein, J. L., Waits, L. P., Ausband, D. E., Zager, P., and Mack, C. M. (2011). Estimating gray wolf pack size and family relationships using noninvasive genetic sampling at rendevous sites. *Journal of Mammalogy* **92**, 784–795. doi:10.1644/10-MAMM-A-200.1

Taberlet, P., and Luikart, G. (1999). Non-invasive genetic sampling and individual identification. *Biological Journal of the Linnean Society* **68**, 41–55. doi:10.1111/j.1095-8312.1999.tb01157.x

Taberlet, P., Griffin, S., Goossens, B., Questiau, S., Manceau, V., Escaravage, N., Waits, L. P., and Bouvet, J. (1996). Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research* **24**, 3189–3194. doi:10.1093/nar/24.16.3189

Taberlet, P., Waits, L. P., and Luikart, G. (1999). Noninvasive genetic sampling: look before you leap. *Trends in Ecology & Evolution* **14**, 323–327. doi:10.1016/S0169-5347(99)01637-7

Turner, F. B., and Medica, P. A. (1982). The distribution and abundance of the flat-tailed horned lizard (*Phrynosoma mcallii*). *Copeia* 815–823. doi:10.2307/1444091

Valiere, N. (2002). GIMLET: a computer program for analysing genetic individual identification data. *Molecular Ecology Notes* **2**, 377–379.

Valiere, N., Bonenfant, C., Toigo, C., Luikart, G., Gaillard, J.-M., and Klein, F. (2007). Importance of a pilot study for non-invasive genetic sampling: genotyping errors and population size estimation in red deer. *Conservation Genetics* **8**, 69–78. doi:10.1007/s10592-006-9149-2

Vynne, C., Baker, M. R., Breuer, Z. K., and Wasser, S. K. (2012). Factors influencing degradation of DNA and hormones in maned wolf scat. *Animal Conservation* **15**, 184–194. doi:10.1111/j.1469-1795.2011.00503.x

Waits, L. P., and Paetkau, D. (2005). Noninvasive genetic sampling tools for wildlife biologists: a review of applications and recommendations for accurate data collection. *Journal of Wildlife Management* **69**, 1419–1433. doi:10.2193/0022-541X(2005)69[1419:NGSTFW]2.0.CO;2

Watts, H. E., Scribner, K. T., Garcia, H. A., and Holekamp, K. E. (2011). Genetic diversity and structure in two spotted hyena populations reflects

social organization and male dispersal. *Journal of Zoology* **285**, 281–291. doi:10.1111/j.1469-7998.2011.00842.x

Wehausen, J. D., Ramey, R. R. II, and Epps, C. W. (2004). Experiments in DNA extraction and PCR amplification from bighorn sheet feces: the importance of DNA extraction method. *The Journal of Heredity* **95**, 503–509. doi:10.1093/jhered/esh068

Wilgers, D. J., and Horne, E. A. (2009). Discrimination of chemical stimuli in conspecific fecal pellets by a visually adept iguanid lizard, *Crotaphytus collaris*. *Journal of Ethology* **27**, 157–163. doi:10.1007/s10164-008-0100-5

Handling Editor: Steven Cooper

Appendix 3: ***Egernia stokesii*** **MHC alleles**

Egernia stokesii **MHC I alleles**

Source file: Estokesii_MHCI_renamed.fas

>Egst-UB*06_CAM05286

GGGCTGCCCCTGTTCACCGCTATGGCATACTTGGATGACCAGTTCATTGGTTTCT

ATGACAGCAGTCGGAGAAGGATCGTGACTTCTGTGGCCTGGATGCAGAAGACTG

AGAAGGAGGACCCCCAGCTCTGGGCCATGCACACGAAGATGGCACAGCGCGAG

GAGCTGTTTCTCAGGAGGGACCTAGTGACTGTGCAGAATGTTCACAACCAGAGT

>Egst-UB*07_CAM05343

GGGCTGCCCCTGTTCACCGCTATGGCATACTTGGATGACCAGTTCATTGGTTTCT

ATGACAGCAGTTGGGGAAGGATCCTGTCTCCTGTAACCTGGGGGCAGAAGACTG

AGAAGGAGGACCCCCAGCTCTGGGCCACGCACACGCAGTTGGCACAGAACAATG

AGCTGTTTCGCAGGAAGGACCTAGCGACTGTGCAGAACCTTCACAACCAGAGT

>Egst-UB*08_CAM05404

GGGCTGCCCCTGTTCACCGCTATGGCATACTTGGATGACCAGTTCATTGGTTTCT

ATGACAGCAGTTGGGGAAGGATCCTGTCTCCTGTAACCTGGGGGCAGAAGACTG

AGAAGGAGGACCCCCAGCTCTGGGCCATGCACACGAAGATGGCACAGCGCGAG

GAGCTGTTTCTCAGGAGGGACCTAGTGACTGTGCAGAATGTTCACAACCAGAGT

>Egst-UB*09_CAM05616

GGGCTGCCCCTGTTCACTGCTATGGCATACTTGGATGACCAGCTCATTGGTTTCT

ATGACAGCAGTCGGAGAAGGATCGTGACTTCTGTGGCCTGGATGCAGAAGACTG

AGAAGGAGGACCCCCAGCTCTGGGCCACGCACACGCAGTTGGCACAGAACAATG

AGCTGTTTCTCAGGAGGGACCTAGCGACTGTGCAGAATCTTCACAACCAGAGT

>Egst-UB*10_HAWKER11037

GAGCTGCCCTGGTTCCTCATTGTGGGCTACGTGGACGGCCAGCGCTTCTTCCACT

ACGACAGCGCCGGGAGGAGGGCAAGGCCTGCAGCGCCCTGGATGGATAAGGTG

TGCCAGGAGGACCCCCACTACTGGGACCAGGAGACCCAGAGCGCACGGGGCGA

GGAGGCGACACTCAGATCTTACCTGGAGATTGCAAGGCAGCGCTACAACCAGAG

C

>Egst-UB*11_HAWKER11046

GAGCTGCCCTGGTTCCTCATTGTGGGCTACGTGGACGGCCAGCGCTTCTTCCACT

ACGACAGCGCTGGGAGGAGGGCAATGCCTGCAGCGCCCTGGATGGATAAGGTGT

GCCAGGAGGACCCCCACTACTGGGACCAGGAGACCCAGAGCGCACGGGGCGAG

GAGGCAACACTCAGATCTTACCTGGAGATTGCAAGGCAGCGCTACAACCAGAGT

>Egst-UB*12_HAWKER11054

GAGCTGCCCTGGTTCCTCATTGTGGGCTACGTGGACGGCCAGCGCTTCTTCCACT

ATGACAGCGCTGGGAGGAGGGCAATGCCTGCAGTGCCCTGGATGGATAAGGTGT

GCCAGGAGGACCCCCACTACTGGGACCAGGAGACCCAGAGCGCACAGGGCGAG

GAGGCGACACTCAGATCTTACCTGGAGATTGCAAGGCAGCGTTACAACCAGAGT

>Egst-UB*13_HAWKER13034

GCGCAGCCCCGGTTCATCAAGGAGGGCTACTTGGATGGCCAGTGCATCATGTACT

ATGACAACAACAGGAACAGGGTGATGCATCGAGTACCCTGGATGGAGAAGGCA

GACAAGGAGACTTCCCCGTCCTGGGATGGACAGCCCCAGAGTGAAATGGGCACT

GAAGCGATCTTCAGAGCAAACCTGGAGATCCTGCAGAAGCGGTACAACCAGAGT

>Egst-UB*14_HAWKER13257

GCGCAGCCCCGGTTCATCATGGAGGGCTACTTGGATGGCCAGTGCATCATGTATT

ATGACAACCACAGGAACAGGGTGATGCATCGAGTACCCTGGATGGAGAAGGTGG

ACAAGAAGACTTCCCTGTCCTGGGATGTACAGACCCAGAGTGCATTTGGCAGTG

AAGCGATCTTCACAGTGAACCTGGAGATTCTGCAGAAGCGCCACAACCACAGC

>Egst-UB*15_HAWKER13269

GCGCAGCCCTGGTTCATCATGGAGGGCTACTTGGATGGGCAGTGCATCATGTACT

ATGACAACCACAGGAACAGGGTGATGCATCGAGTACCCTGGATGGAGAAGGCG

GACAAGAAGACTTCCCTGTCCTGGGATGTACAGACCCAGAGTGCATTTGGCAGT

GAAGCGATCTTCACAGTGAACCTGGAGATTCTGCAGAAGCGCCACAACCAGAGT

>Egst-UB*16_HAWKER15156

GGGCTGCCCCAGTTCATGGCGTTGGCATATACGGATGATCAGCTGATGGGTGTCT

ATGACAGCAATAGGAGGAGGATTGTACCTTATTTGCCCTGGGTGAAGAAGATTG

AGAAGGAGGACCCCCAATTCTGGGCCATGCCTACACAGTTTATGCAGAACCAGA

AACTGTATTTGGGGGTGGATCTGGACACTGTGCAGGAACACTACAACCAGAGC

>Egst-UB*17_HAWKER15312

GGGCTGCCCCTGTTCACCGCCACGGCATACTTGGATGACCAGCTCATTGGTTTCT

ATGACAGCAGTGGGGGAAGGATCGTGTCTCCTGTAGCCTGGGGGCAGAAGACTG

AGAAGGAAGACCCTGAGCTCTGGGCCATGCACACGCGGATGGCACAGAATGAG

GAGCTGAGTCTCAGGGTGCACCTAGCGACTGTGCAGAATCTTCACAACCAGAGT

>Egst-UB*18_HAWKER15452

GGGCTGCCCCTGTTCACCGCCAGGGCATACTTGGATGACCAGCTCATTGGTTTCT

ATGACAGCAGTGGGGGAAGGATCGTGTCTCCTGTGGCCTGGGGGCAGAAGACTG

AGAAGGAAGACCCTGAGCTCTGGGCCATGCACACGCGGATGGCACAGAATGAG

GAGCTGAGTCTCAGGGTGCACCTAGCGACTTTGCAGAATCTTCACAACCAGAGT

>Egst-UB*19_HAWKER15780

GGGCTGCCCCTGTTCACCGCTACGGCATACTTGGATGACCAGCTCATTGGTTTCT

ATGACAGCAGTAGGGGAAGGATTGTGTCTCCTGTGGCCTTGGGGCAGAAGACTG

AGAAGGAGGACCCCCATCTCTGGGTCAGGCACACGCAGATGGCACAGAACAATG

AGCTGTTTCTCAGGAGGGACCTAGCGACTGTGCAGAATCTTCACAACCAGAGT

>Egst-UB*20_HAWKER15808

GGGCTGCCCCTGTTCACCGCTACGGCATACTTGGATGACCAGCTCATTGGTTTCT

ATGACAGCAGTAGGGGAAGGATTGTGTCTCCTGTGGCCTTGGGGCAGAAGACTG

AGAAGGAGGACCCCCATCTCTGGGTCAGGCACACGCAGATGGCACAGAACAATG

AGCTGTTTCTCAGGAGGGACCTAGTGACTGTGCAGAATGTTCACAACCAGAGT

>Egst-UB*21_HAWKER16045

GGGCTGCCCCTGTTCACCGCTACGGCATACTTGGATGACCAGCTCATTGGTTTCT

ATGACAGCAGTTGGGGAAGGATCCTGTCTCCTGTGGCCTGGGGGCAGAAGACTG

AGAAGGAGGACCCCCAGCTCTGGGACATGCACATGCAGATGGCACAGCGTGAGG

AGCTGGTTCTCAGGAGGCACCTAGCGACTGTGCAGAATCTTCACAACCAGAGT

>Egst-UB*22_HAWKER16385

GGGCTGCCCCTGTTCACCGCTATGGCATACTTGGATGACCAGCTCATTGGATTCT

ATGACAGCAGTCGGAGAAGGATCGTGACTTCTGTGGCCTGGATGCAGAAGACTG

AGAAGGAGGACCCCCAGCTCTGGGACATGCACACGAAGATGGCACAGCGCGAG

GAGCTGTTTCTCAGGAGGGACCTAGTGACTGTGCAGAATGTTCACAACCAGAGT

>Egst-UB*23_HAWKER16582

GGGCTGCCCCTGTTCACCGCTATGGCATACTTGGATGACCAGCTCATTGGTTTCT

ATGACAGCAGTAGGGGAAGGATTGTGTCTCCTGTGGCCTTGGGGCAGAAGACTG

AGAAGGAGGACCCCCATCTCTGGGTCAGTCACACGCAGATGGCACAGAACAATG

AGCTGTTTCTCAGGAGGGACCTAGTGACTGTGCAGAATGTTCACAACCAGAGT

>Egst-UB*24_HAWKER16665

GGGCTGCCCCTGTTCACCGCTATGGCATACTTGGATGACCAGCTCATTGGTTTCT

ATGACAGCAGTCGGAGAAGGATCATGACTTCTGTGGCCTGGATGCAGAAGACTG

AGAAGGAGGACCCCCAGCTCTGGGACATGCACACGAAGATGGCACAGCGCGAG

GAGCTGTTTCTCAGGAGGGACCTAGTGACTGTGCAGAATGTTCACAACCAGAGT

>Egst-UB*25_HAWKER16846

GGGCTGCCCCTGTTCACCGCTATGGCATACTTGGATGACCAGCTCATTGGTTTCT

ATGACAGCAGTGGGGGAAGGATCGTGTCTCCTGTGGCCTGGGGGCAGAAGACTG

AGAAGGAAGACCCTGAGCTCTGGGCCATGCACACGCGGATGGCACAGAATGAG

GAGCTGAGTCTCAGGGTGCACCTAGCGACTTTGCAGAATCTTCACAACCAGAGT

>Egst-UB*26_HAWKER16952

GGGCTGCCCCTGTTCACCGCTATGGCATACTTGGATGACCAGCTCATTGGTTTCT

ATGACAGCAGTTGGAGAAGCATCGTGTCTCCTGTGGCCTGGGTACAGAATACTG

AGAAGGAGGACCCCGAGCTCTGGGCCATGCACACGGAGATGGCACAGAACGAG

GAGCTGAGTCTCAGGAGGCACCTAGCGACTGTGCAGAATCTTCACAACCAGAGT

>Egst-UB*27_HAWKER17175

GGGCTGCCCCTGTTCACCGCTATGGCATACTTGGATGACCAGCTCATTGGTTTCT

ATGACAGCAGTTGGGAAAGGATCGTGTCTCCTGTGGCCTGGGTGCAGAAGGCTG

AGAAGGAGGACCCCGAGCTCTGGGCCATGCACACGCGGATGGCACAGAACACTG

AGCTGTTTCGCAGGAAGGACCTAGCGACTGTGCAGAACCTTCACAACCAGAGT

>Egst-UB*28_HAWKER17352

GGGCTGCCCCTGTTCACCGCTATGGCATACTTGGATGACCAGCTCATTGGTTTCT

ATGACAGCAGTTGGGGAAGGATCCTGCCTCCTGTGGCCTGGGGGCAGAAGACTG

AGAAGGAGGACCCCCAGCTCTGGGCCATGCACACGGAGATGGCACAGAATGAG

GAGCTGAGTCTCAGGAGGCACGTAGCGACTGTGCAGAATGTTCACAACCAGAGT

>Egst-UB*29_HAWKER17498

GGGCTGCCCCTGTTCACCGCTATGGCATACTTGGATGACCAGCTCATTGGTTTCT

ATGACAGCAGTTGGGGAAGGATCGTGTCTCCTGTGGCCTGGGTGCAGAAGACTG

AGAAGGAGGACCCCGAGCTCTGGGCCATGCACACACAGAGAGCAAAGAACAAT

GAGCTGTTTCTCAGGGGGCACCTAGCAACTGTGCAGAATCTTCACAACCAGAGT

>Egst-UB*30_HAWKER17595

GGGCTGCCCCTGTTCACCGCTATGGCATACTTGGATGACCAGCTCATTGGTTTCT

ATGACAGCAGTTGGGGAAGGATCGTGTCTCCTGTGGCCTGGGTGCAGAAGGCTG

AGAAGGAGGACCCCGAGCTCTGGGCCATGCACACGCGGATGGCACAGAACACTG

AGCTGTTTCGCAGGAAGGACCTAGCGACTGTGCAGAACCTTCACAACCAGAGT

>Egst-UB*31_HAWKER17693

GGGCTGCCCCTGTTCACCGCTATGGCATACTTGGATGACCAGCTCATTGGTTTCT

GTGACAGCAGTGGGGGAAGGATCGTGTCTCCTGTGGCCTGGGTGCAGAAGACTG

AGAGGAAGGAATACCAGCTCTGGGCCATGCACACGCAGGTGGCACAGAAGGAG

GAGCTGAGTCTCAGGAGGGACCTAGCGACTGTGCAGAATCTTCACAACCAGAGT

>Egst-UB*32_HAWKER17895

GGGCTGCCCCTGTTCACCGCTATGGCATACTTGGATGACCAGTTCATTGGTTTCT

ATGACAGCAGTTGGGGAAGGATCCTGTCTCCTGTAACCTGGGGGCAGAAGACTG

AGAAGGAGGACCCCCAGCTCTGGGCCACGCACACGCAGTTGGCACAGAACAATG

AGCTGTTTCTCAGGAGGGACCTAGCGACTGTGCAGAATCTTCACAACCAGAGT

>Egst-UB*33_HAWKER17977

GGGCTGCCCCTGTTCACCGCTATGGCATACTTGGATGACCAGTTCATTGGTTTCT

ATGACAGCAGTTGGGGAAGGATCCTGTCTCCTGTAACCTGGGGGCAGAAGACTG

AGAAGGAGGACCCCCATCTCTGGGTCAGGCACACGCAGATGGCACAGAACAATG

AGCTGTTTCTCAGGAGGGACCTAGTGACTGTGCAGAATGTTCACAACCAGAGT

>Egst-UB*34_HAWKER18104

GGGCTGCCCCTGTTCACCGCTATGGCATACTTGGATGACCAGTTCATTGGTTTCT

ATGACAGCAGTTGGGGAAGGATCGTGTCTGCTGTGGCCTGGATGCAGAAGACTG

AGAAGGAGGACCCCCAGCTCTGGGACATGCACACGAAGATGGCACAGAACAAT

GAGCTGTTTCGCAGGAGGGACCTAGCGACTGTGCAGAATCTTCACAACCAGAGT

>Egst-UB*35_HAWKER18416

GGGCTGCCCCTGTTCACTGCTACGGCATACTTGGATGACCAGCTCATTGGTTTCT

ATGACAGCAGTAGGGGAAGGATTGTGTCTCCTGTGGCCTTGGGGCAGAAGACTG

AGAAGGAGGACCCCCATCTCTGGGTCAGTCACACGCAGATGGCACAGAACAATG

AGCTGTTTCTCAGGAGGGACCTAGTGACTGTGCAGAATGTTCACAACCAGAGT

>Egst-UB*36_HAWKER18531

GGGCTGCCCCTGTTCACTGCTATGGCATACTTGGATGACCAGCTCATTGGTTTCT

ATGACAGCAGCTGGGGAAGGATCGTGTCTCCTGTGGCCTGGGTGCAGAAGGCTG

AGAAGGAGGACCCCGAGCTCTGGGCCATGCACGCGCGGCTGGCACAGAACAATG

AGCTCTTTCGCAGGAAGGACCTAGCGACTGTGCAGAACCTTCACAACCAGAGT

>Egst-UB*37_HAWKER18533

GGGCTGCCCCTGTTCACTGCTATGGCATACTTGGATGACCAGCTCATTGGTTTCT

ATGACAGCAGCTGGGGAAGGATCGTGTCTCCTGTGGCCTGGGTGCAGAAGGCTG

AGAAGGAGGACCCCGAGCTCTGGGCCATGCACGTGCGGCTGGCACAGAACAATG

AGCTGTTTCGCAGGAAGGACCTAGCGACTGTGCAGAACCTTCACAACCAGAGT

>Egst-UB*38_HAWKER18698

GGGCTGCCCCTGTTCACTGCTATGGCATACTTGGATGACCAGCTCATTGGTTTCT

ATGACAGCAGTCGGAGAAGGATCGTGACTTCTGTGGCCTGGATGCAGAAGACTG

AGAAGGAGGACCCCCAGCTCTGGGCCATGCACACGAAGATGGCACAGCGCGAG

GAGCTGTTTCTCAGGAGGGACCTAGTGACTGTGCAGAATGTTCACAACCAGAGT

>Egst-UB*39_HAWKER18885

GGGCTGCCCCTGTTCACTGCTATGGCATACTTGGATGACCAGCTCATTGGTTTCT

ATGACAGCAGTGGGGGAAGGATCGTGTCTCCTGTGGCCTGGGTGCAGAAGGCTG

AGAAGGAGGACCCCGAGCTCTGGGCCATGCACACGCGGATGGCACAGAACACTG

AGCTGTTTCGCAGGAAGGACCTAGCGACTGTGCAGAACCTTCACAACCAGAGT

>Egst-UB*40_HAWKER19087

GGGCTGCCCCTGTTCACTGCTATGGCATACTTGGATGACCAGCTCATTGGTTTCT

ATGACAGCAGTTGGGGAAGGATCGTGTCTCCTGTGGCCTGGGTGCAGAAGGCTG

AGAAGGAGGACCCCGAGCTCTGGGCCATGCACACGCGGATGGCACAGAACAAT

GAGCTGTTTCGCAGGAAGGACCTAGCGACTGTGCAGAACCTTCACAACCAGAGT

>Egst-UB*41_HAWKER19446

GGGCTGCCTCTGTTCACCGCTATGGCATACTTGGATGACCAGCTCATTGGTTTCT

ATGACAGCAGTGGGGGAAGGATCGTGTCTCCTGTGGCCTGGGTACAGAATACTG

AGAAGGAGGACCCCGAGCTCTGGGCCATGCACACGCAGATGGCACAGCGTGAG

GAGCTGGTTCTCAGGAGGCACCTAGCGACTGTGCAGAATCTTCACAACCAGAGT

>Egst-UB*42_HAWKER19536

GGGCTGCCTCTGTTCACCGCTATGGCATACTTGGATGACCAGCTCATTGGTTTCT

ATGACAGCAGTTGGGGAAGGATCATGTCTCCTGTGGCCTGGGTACAGAATACTG

AGAAGGAGGACCCCGAGCTCTGGGCCATGCACACACAGATGGCACAGAACAAT

GAGCTTTTTCTCAGGAGGCACCTAGCGACTGTGCAGAATCTTCACAACCAGAGT

>Egst-UB*43_HAWKER19594

GGGCTGCCTCTGTTCACCGCTATGGCATACTTGGATGACCAGCTCATTGGTTTCT

ATGACAGCAGTTGGGGAAGGATCATGTCTCCTGTGGCCTGGGTACAGAATACTG

AGAAGGAGGACCCCGAGCTCTGGGCCATGCACACGCAGATGGCACAGCGTGAG

GAGCTGGTTCTCAGGAGGCACCTAGCGACTGTGCAGAATCTTCACAACCAGAGT

>Egst-UB*44_HAWKER19668

GGGCTGCCTCTGTTCACCGCTATGGCATACTTGGATGACCAGCTCATTGGTTTCT

ATGACAGCAGTTGGGGAAGGATCATGTCTCCTGTGGCCTGGGTGCAGAAGGCTG

AGAAGGAGGACCCCGAGCTCTGGGCCATGCACACGCGGATGGCACAGAACAAT

GAGCTGAGTCTCAGGAGGCACCTAGCGACTGTGCAGAATCTTCACAACCAGAGT

***Egernia stokesii* MHC II alleles**

Source file: Estokesii_MHCII_renamed.fas

>Egst-DRB*01_211

CGCCGCGGGAGCTTCGAGGCCGTCACAGAGCTGGGCGAGCTTGAAGCCCGCGTC

TGGAACAGCCAGACAGAGTTTCTGGAAGCGATGCGGACCGACGTGAAC

>Egst-DRB*02_795

CGCCGCGGGAGCTTCGAGGCCATCACGGAGCTGGGCGAGCCTGAAGCCCGCGTC

TGGAACAGCCAGACGGAGTTTCTGGAAGCGATGCGGACCGACGTGGAC

>Egst-DRB*03_108

CGCCGAGGGACCTACGAGGCCGTCACCGCGCTGGGCGAGCCCTCAGCGCGCTAC

TGGAACAGCCTGACGGACTACATGGAGCGGAAGCGGACTGAAGTGGAC

>Egst-DRB*04_280

CGCCGCGGGAGCTTCGAGGCCGTCACGGAGCTGGGCAAGCCTGAAGCCCGCGTC

TGGAACAGCCAGACAGAGTTTCTGGAAGCGATGCGGACCGACGTGGAC

>Egst-DRB*05_444

CGCCGTGGGAGCTTCGAGGCCATCACAGAGCTGGGCGAGCCTGAAGCCCGCATC

TGGAACAGCCAGACAGAGTTTCTGGAAGCGATGCGGACCGACGTGGAC

Appendix 4: ***Egernia stokesii*** **MHC allele evolutionary relationships**

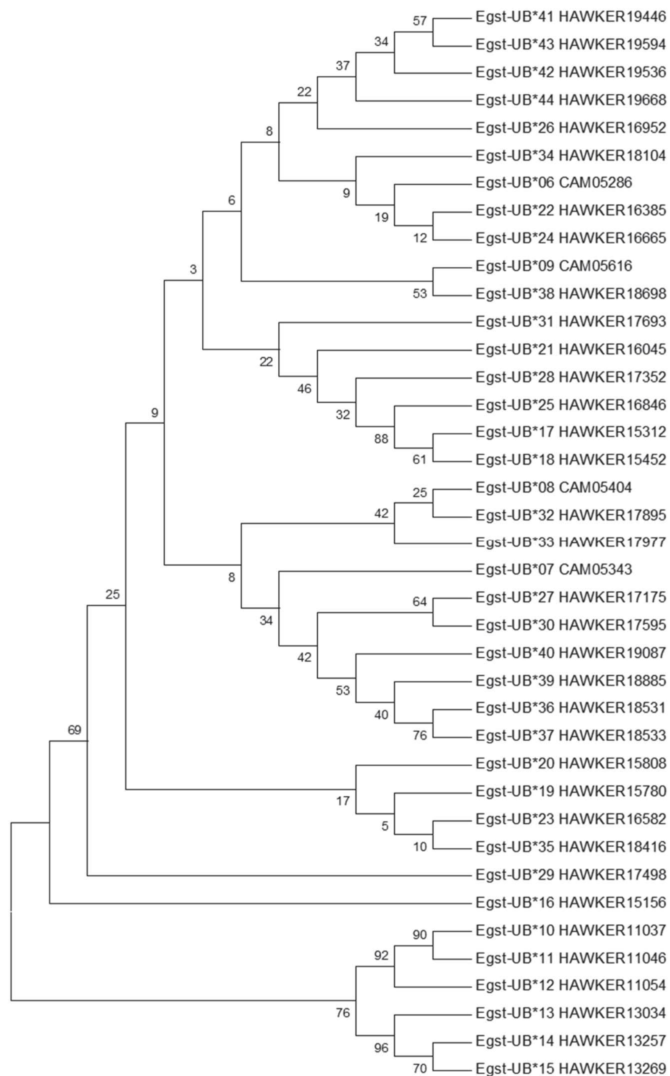


Figure 1: *Egernia stokesii* neighbour joining tree derived from 39 MHC I alleles. The evolutionary history was inferred using the Neighbor-Joining method. The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analysed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The evolutionary distances were computed using the Nei-Gojobori method and are in the

units of the number of synonymous substitutions per synonymous site. The analysis involved 39 nucleotide sequences. All ambiguous positions were removed for each sequence pair. There were a total of 72 positions in the final dataset. Evolutionary analyses were conducted in MEGA6. (Source file: Estokesii_MHCI_renamed.meg)

MHC II

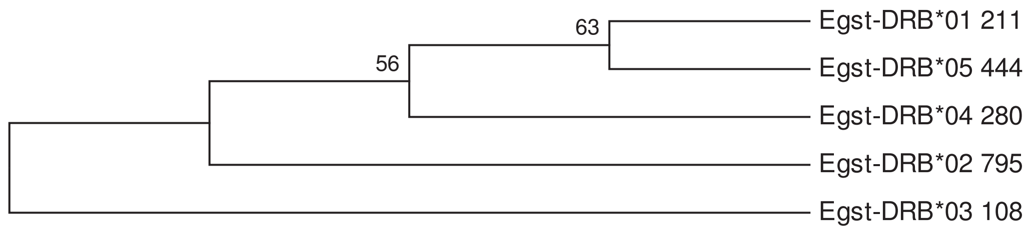


Figure 2: *Egernia stokesii* neighbour joining tree derived from five MHC II alleles. The evolutionary history was inferred using the Neighbor-Joining method. The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analysed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The evolutionary distances were computed using the Nei-Gojobori method and are in the units of the number of synonymous substitutions per synonymous site. The analysis involved 5 nucleotide sequences. All ambiguous positions were removed for each sequence pair. There were a total of 34 positions in the final dataset. Evolutionary analyses were conducted in MEGA6. (Source file: Estokesii_MHCII_renamed.meg)

Appendix 5: ***Egernia stokesii*** **MHC genotypes**

Egernia stokesii MHC I genotypes 1994-98
Source file: MHCI genotypes & attributes V2 worksheet in \Data\CAM_1994-98\Estokesii_MGSamples_Dataset.xls
MHC I genotypes from Samples & Genotypes >T1 worksheet in CAM9498_MHCI_jMHC.xlsx
Group and sex from Group members details worksheet Dr Michael Gardner thesis chapter 6
Grey shaded rows were subadults therefore were excluded from analysis
Bold text in grey highlighted cells indicate reads >2 = called as an allele

			Allele name	Egst-UB*11	Egst-UB*44	Egst-UB*09	Egst-UB*08	Egst-UB*06	Egst-UB*37	Egst-UB*36	Egst-UB*07	Egst-UB*12	Egst-UB*16	Egst-UB*28	Egst-UB*17	Egst-UB*31	Egst-UB*22	Egst-UB*24	Egst-UB*35	Egst-UB*34	Egst-UB*30	Egst-UB*39	Egst-UB*42	Egst-UB*18	Egst-UB*40	Egst-UB*43	Egst-UB*38	Egst-UB*27	Egst-UB*32	Egst-UB*21	Egst-UB*20	Egst-UB*13	Egst-UB*26	Egst-UB*29
			Sequence ID	3353	6127	5636	5404	5286	5577	5576	5343	3356	4480	5114	4469	5262	4887	4943	5530	5452	5203	5777	6065	4505	5872	6090	5664	5094	5364	4769	4655	3881	5043	5159
ID	Group	Age	Sex																															
10	D	Adults	F	2	0	0	2	2	0	0	0	0	0	0	0	0	8	0	0	8	0	0	0	0	33	0	40	2	19	1	17	37	0	14
32	L	Adults	F	0	1	1	3	2	1	3	0	0	0	6	1	0	1	0	0	2	0	0	0	25	1	0	28	2	23	5	12	28	0	1
73	D	Adults	F	1	0	4	0	0	0	22	1	60	0	0	42	0	0	125	0	0	279	0	197	0	0	0	467	0	249	0	163	279	116	0
79	C	Adults	F	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	12	2	13	10	12	22	0	26	0	14	12	3	0
85	J2	Adults	F	0	0	1	0	0	0	0	0	4	0	0	0	0	2	3	0	0	11	0	1	0	16	5	15	4	16	0	10	13	5	2
88	B1	Adults	F	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	1	0	0	0	1	20	0	17	0	8	0	4	21	0	6
89	G2	Adults	F	2	0	8	4	0	0	1	0	0	0	0	0	0	0	0	0	0	2	0	0	0	14	0	83	0	71	0	19	35	0	0
92	C	Adults	F	0	0	0	0	0	3	0	0	0	0	0	0	0	1	0	0	1	0	0	0	4	3	0	2	0	1	0	2	44	2	1
107	G2	Adults	F	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	2	0	1	0	0	0	3	4	5	0	4	2	0	0
112	C	Adults	F	8	0	0	0	0	0	0	0	0	0	0	0	0	203	0	0	193	0	0	5	0	472	0	1	1	1	0	181	376	4	220
117	C	Adults	F	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	2	0	0	0	0	0	0	0	5	0	0	3	2	0	1
142	A	Adults	F	0	0	0	0	1	0	1	2	0	0	0	0	0	1	0	0	0	0	0	0	0	4	0	5	4	2	1	3	1	1	2
166	A	Adults	F	0	4	0	0	0	5	4	0	0	0	0	11	0	9	6	0	5	0	0	12	0	9	0	1	0	1	0	18	27	10	3
238	E5	Adults	F	0	0	0	0	0	0	0	0	0	0	0	0	0	0	241	0	0	1	0	1	2	0	0	6	0	2	1206	154	3	156	0
239	N4	Adults	F	4	0	0	0	0	0	3	0	0	0	0	21	0	0	26	0	0	0	0	98	0	30	2	1	0	0	0	26	87	60	0
504	G2	Adults	F	0	3	2	4	1	2	0	0	0	0	33	0	0	1	0	0	6	0	7	0	32	16	4	61	0	68	0	36	101	9	8
520	L/N7/8	Adults	F	0	0	5	0	0	0	1	0	0	0	0	0	0	14	0	0	10	0	0	0	0	30	0	58	0	61	0	30	114	0	17
532	D	Adults	F	0	0	0	0	0	0	0	0	0	2	0	0	0	18	0	0	12	6	89	0	40	1	79	1	86	0	0	31	39	43	34
552	N1	Adults	F	2	0	6	7	3	0	27	1	130	0	0	148	0	3	193	0	2	618	0	415	0	0	0	803	12	703	0	291	466	397	1
554	B8	Adults	F	16	0	4	3	2	0	0	4	161	22	0	0	1	149	0	0	131	561	0	2	0	0	0	641	514	545	0	224	440	1	162
690	E1	Adults	F	0	0	6	5	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	29	0	49	61	0	52	101	18	40	22	0
704	H	Adults	F	1	0	38	29	11	0	0	11	0	0	0	0	0	3	0	0	1	3	2	0	0	220	1	457	0	392	0	42	219	0	1
26	J2	Adults	M	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	4	0	8	0	11	0	4	2	0	0
31	M	Adults	M	0	0	0	1	0	0	0	0	11	0	0	0	0	0	6	0	0	19	0	1	1	0	0	20	0	18	38	21	3	8	1
68	D	Adults	M	0	0	0	2	1	0	0	1	0	0	0	0	0	3	2	0	5	0	0	1	0	24	0	30	0	23	0	12	29	0	13
69	C	Adults	M	0	0	0	0	0	0	0	0	0	0	0	0	0	34	0	0	23	1	0	0	1	80	0	0	0	0	0	41	70	0	36
70	E5	Adults	M	0	0	5	1	0	0	0	3	0	0	0	0	0	0	0	0	0	0	268	0	135	209	291	379	0	240	0	202	176	160	0
76	B8	Adults	M	5	0	2	1	0	0	0	1	0	0	0	0	0	11	0	0	7	0	0	0	0	44	0	22	0	28	1	24	27	0	5
86	D	Adults	M	0	0	0	0	0	0	11	0	0	0	0	0	0	0	3	0	0	0	4	0	114	18	0	2	0	16	5	4	144	78	0
103	B8	Adults	M	0	0	0	0	0	0	0	0	0	0	0	0	0	25	0	0	0	0	0	0	1	13	98	4	2	4	0	36	21	12	37
106	H	Adults	M	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	7	0	0	0	63	1	212	16	0	9
141	D/J floater	Adults	M	0	0	1	0	0	0	45	0	0	0	0	22	0	1	70	0	0	0	0	176	0	2	0	328	0	351	4	78	430	145	5
165	L/M	Adults	M	0	0	6	9	4	0	8	1	0	0	3	0	1	2	0	0	0	0	0	0	0	98	0	525	0	356	0	77	222	1	1
201	E1	Adults	M	0	0	0	0	0	4	0	0	0	0	15	0	21	17	0	44	14	0	0	0	26	6	0	0	2	0	0	31	87	0	15
202	D	Adults	M	0	0	0	0	0	0	0	0	0	0	4	0	0	3	0	0	1	0	0	0	5	4	1	14	2	1	0	8	11	0	4
221	C	Adults	M	1	0	2	0	0	0	2	0	0	0	0	0	1	2	10	0	1	0	0	0	2	17	3	22	0	24	1	12	48	22	4
226	A	Adults	M	0	0	1	5	7	5	0	0	0	3	0	0	0	0	61	0	0	0	0	0	0	114	0	232	2	142	0	58	103	61	0
241	D	Adults	M	0	0	0	1	0	0	0	3	0	3	0	0	0	12	0	0	11	1	0	0	0	13	0	43	22	31	0	12	80	0	15
244	G2	Adults	M	0	1	0	0	0	3	0	0	0	0	26	15	0	0	21	0	0	0	0	42	37	49	0	0	0	0	1	18	106	30	0
268	N4	Adults	M	0	0	0	0	0	0	1	0	0	0	0	0	0	4	0	0	2	0	3	0	1	7	3	3	2	0	0	3	15	0	5
472	H/J floater	Adults	M	0	0	0	0	0	0	3	0	0	0	0	0	0	36	0	0	16	0	1	0	0	47	2	0	3	2	0	30	168	0	26
507	B1	Adults	M	10	0	8	3	0	0	0	0	0	0	0	0	0	1	18	0	0	0	0	0	0	115	0	109	0	86	1	40	201	44	0
521	N1	Adults	M	14	0	11	5	2	0	0	1	0	0	0	0	63	21	0	92	22	1	0	0	0	64	0	138	0	130	0	40	184	0	19
537	B8	Adults	M	0	0	0	0	0	0	0	0	0	0	0	0	0	0	81	0	0	0	0	0	126	1	284	0	0	0	990	130	0	217	0
591	J3	Adults	M	11	0	0	0	0	0	0	0	0	0	0	0	0	423	0	0	397	0	3	0	0	982	2	0	1	5	0	300	989	2	406
701	A	Adults	M	1	0	4	4	4	0	0	6	0	0	0	0	0	27	0	0	23	0	0	0	0	185	0	119	0	143	0	78	105	0	18
901	M	Adults	M	0	0	12	13	2	0	0	2	1	0	0	0	0	0	0	0	0	106	1	0	55	0	110	163	0	135	0	22	96	53	0
682	N7/8	95/96 coh	?	0	0	0	0	0	0	0	0	0	0	0	11	0	9	5	0	9	0	0	21	0	0	0	0	2	0	0	5	37	18	14
582	B1	93/94 coh	?	0	0	2	0	2	0	0	0	0	0	0	1	0	0	7	0	0	0	0	73	0	69	1	105	0	41	0	119	62	19	0
82	C/E1	92/93 coh	F	0	0	0	0	0	12	0	0	0	0	54	0	1	6	2	1	3	0	0	0	196	4	149	6	0	2	229	141	62	100	1
256	C	92/93 coh	F	0	0	1	0	0	0	7	0	0	0	0	0	0	2	0	0	1	0	0	0	0	12	0	48	1	18	0	4	16	0	10
412	C	92/93 coh	F	0	0	0	0	0	0	0	0	0	0	0	0	7	3	2	6	2	0	0	0	2	1	6	1	0	0	10	7	9	8	0
250	C	92/93 coh	M	0	0	0	1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	1	0	2	0	6	0	32	6	0	4	1	0
527	B8	91/92 coh	F	4	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	22	0	78	0	69	0	7	28	0	0
584	B1/J floate	89/90 coh	F	0	0	0	0	0	0	0	0	0	0	0	0	0	101	1	0	118	0	370	0	194	245	328	4	0	2	0	215	271	231	110

MHC II genotypes originally taken from Samples & Genotypes >T1 worksheet in CAM9498_MHCII_jMHC.xlsx
Group and sex from *Group members details* worksheet copied/sorted from Dr Michael Gardner thesis chapter 6
Individual 141 amended from '?' to Adult based on detail in MG thesis
Grey shaded rows were subadults therefore excluded from analysis
Bold text in grey highlighted cells indicate reads >2 = called as an allele

			Allele name	Egst-DRB*02	Egst-DRB*05	Egst-DRB*04	Egst-DRB*03	Egst-DRB*01
			Sequence ID	795_538	444_741	280	108	211
ID	Group	Age	Sex					
10	D	Adults	F	**42**	**3332**	**6**	0	0
32	L	Adults	F	**693**	**2112**	0	1	**17**
73	D	Adults	F	**1104**	**6**	1	0	**1317**
79	C	Adults	F	**3**	**3367**	0	0	1
85	J2	Adults	F	**6**	**2756**	0	0	**297**
88	B1	Adults	F	**16**	**3057**	0	0	0
89	G2	Adults	F	2	**3793**	0	0	2
92	C	Adults	F	**3**	**1234**	0	0	0
112	C	Adults	F	**51**	**3139**	0	0	0
166	A	Adults	F	**2905**	**7**	0	0	0
239	N4	Adults	F	**503**	**1464**	0	0	0
504	G2	Adults	F	**2203**	**2352**	0	0	**6**
520	L/N7/8	Adults	F	1	**2791**	0	0	**505**
530	D/M	Adults	F	**2273**	**1866**	0	0	0
532	D	Adults	F	**516**	**58**	**8**	0	**5**
552	N1	Adults	F	**1459**	**4**	0	0	**1921**
554	B8	Adults	F	**2437**	**41**	0	0	**1190**
560	H	Adults	F	**3**	**4405**	**3**	0	0
690	E1	Adults	F	0	**4372**	**417**	0	0
704	H	Adults	F	**4**	**3935**	0	0	**73**
778	DAM	Adults	F	**1254**	1	0	0	0
26	J2	Adults	M	1	**3510**	0	**353**	**7**
31	M	Adults	M	**25**	**17**	0	**272**	**4068**
68	D	Adults	M	0	**3551**	0	0	1
69	C	Adults	M	0	**2938**	0	0	1
70	E5	Adults	M	0	**4197**	0	0	0
76	B8	Adults	M	0	**3519**	0	0	0
141	J FLOATERS	Adults	M	**288**	**3316**	0	0	0
165	L/M	Adults	M	**3**	**3650**	0	0	0
201	E1	Adults	M	**3469**	**26**	0	0	0
202	D	Adults	M	**35**	**973**	2	0	0
221	C	Adults	M	**28**	**799**	1	0	0
226	A/G2	Adults	M	**4**	**3870**	0	0	0
241	D	Adults	M	**156**	**2150**	0	0	0
244	G2	Adults	M	**15**	**4431**	0	0	0
254	D	Adults	M	**4**	**3972**	0	0	0
268	N4	Adults	M	**1790**	**401**	**166**	0	**30**
298	G2	Adults	M	**902**	**868**	0	0	1
400	D	Adults	M	**1212**	**2790**	0	0	0
407	M/N7/8	Adults	M	1	**3501**	0	2	0
472	H/J floater?	Adults	M	**21**	**3424**	**3**	0	**275**
507	B1	Adults	M	0	**4665**	0	0	0
521	N1	Adults	M	**2912**	**2424**	0	0	1
591	J3	Adults	M	**12**	**5607**	0	0	**3**
701	A	Adults	M	1	**5431**	0	0	0
901	M	Adults	M	**3**	**4**	0	0	**4269**
682	N7/8	95/96 cohort	?	932	3	0	0	749
582	B1	93/94 cohort	?	0	1359	0	0	0
82	C/E1	92/93 cohort	F	837	160	43	0	0
256	C	92/93 cohort	F	0	2726	0	0	0
412	C/E1	92/93 cohort	F	1099	31	80	0	0
250	C	92/93 cohort	M	2	905	0	0	0
527	B8	91/92 cohort	F	0	1925	0	0	0
311	D/E1	89/90 cohort	F	0	4035	0	0	0
584	B1/J floater?	89/90 cohort	F	7	4426	0	0	0

Egernia stokesii MHC I genotypes 2012-13
Source file: FINAL genotypes & samples worksheet in Hawker_MHCI_Genotyping.xlsx at \Estokesii_MHCdownstream_17112015\jMHC\Hawker2012-13
Grey shaded rows were subadults therefore were excluded from analysis
Final list of samples and MHC I genotypes where samples have been filtered based on T1 coverage threshold of max T1 = 5328 (alleles called where reads > 2)
Includes addition of 3 alleles to 15312 ALLELE for TO 388167, 412968, 389630, AND 406817 - FROM SEQ 15333 WHICH TRANSLATES TO THE SAME AMINO ACID

	Allele name	Egst-UB*10	Egst-UB*11	Egst-UB*12	Egst-UB*13	Egst-UB*14	Egst-UB*15	Egst-UB*16	Egst-UB*17	Egst-UB*18	Egst-UB*19	Egst-UB*20	Egst-UB*21	Egst-UB*22	Egst-UB*23	Egst-UB*24	Egst-UB*25	Egst-UB*26	Egst-UB*27	Egst-UB*28	Egst-UB*29	Egst-UB*30	Egst-UB*31	Egst-UB*32	Egst-UB*33	Egst-UB*34	Egst-UB*35	Egst-UB*36	Egst-UB*37	Egst-UB*38	Egst-UB*39	Egst-UB*40	Egst-UB*41	Egst-UB*42	Egst-UB*43	Egst-UB*44
	Sequence ID	11037	11046	11054	13034	13257	13269	15156	15312	15452	15780	15808	16045	16385	16582	16665	16846	16952	17175	17352	17498	17595	17693	17895	17977	18104	18416	18531	18533	18698	18885	19087	19446	19536	19594	19668
ID	Site																																			
1046168	CAM	0	0	2	50	0	36	0	0	0	3	79	0	0	0	0	0	0	2	0	9	251	0	244	0	0	0	0	0	267	0	42	0	0	0	0
1055885	CAM	0	11	0	148	0	0	0	0	2	1	58	0	35	0	5	0	1	0	1	27	0	0	99	1	47	0	3	1	149	0	54	0	1	1	0
1062514	CAM	0	0	9	21	0	1	0	0	0	0	7	0	4	0	0	0	0	17	0	4	23	1	19	0	4	0	0	0	29	0	0	0	0	0	0
1068267	CAM	0	0	0	31	0	0	0	1	2	0	2	0	0	0	0	0	1	2	0	0	0	0	0	0	0	0	0	0	1	1	2	0	0	0	0
1128694	CAM	0	0	0	164	0	0	0	26	1	0	45	0	21	5	6	0	81	8	0	24	0	0	21	0	171	0	11	0	23	0	41	0	29	4	0
1180852	CAM	0	16	0	8	0	0	0	0	81	0	57	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	1	0	3	0	0	0	14	0
2002	CAM	0	0	2	5	0	0	0	0	0	0	3	0	1	0	0	0	0	6	2	4	6	0	5	0	2	0	0	0	17	0	1	0	0	28	0
2006	CAM	0	0	0	118	0	0	0	0	0	0	7	1	0	0	0	0	0	1	0	0	0	0	110	0	0	0	0	0	210	0	120	0	0	1	0
2008	CAM	0	0	0	99	0	0	0	7	45	0	28	0	0	1	45	3	45	0	0	0	0	0	0	0	0	0	0	1	0	0	38	0	22	0	0
201010	CAM	0	0	0	64	0	0	0	0	0	0	42	0	0	0	0	0	0	3	0	2	0	0	160	0	4	0	0	0	159	0	108	0	5	0	0
2018	CAM	0	0	2	1	0	0	0	0	0	0	34	14	2	0	2	0	1	0	0	32	0	0	3	0	0	1	0	0	4	0	2	0	0	0	0
2019	CAM	0	0	0	159	0	0	0	0	0	7	191	0	0	0	0	0	1	3	0	0	0	0	521	1	0	0	0	0	237	0	320	0	0	0	0
2021	CAM	0	0	11	0	0	0	0	0	0	2	22	1	2	0	0	0	0	0	0	0	4	0	34	4	0	0	0	0	27	0	3	0	0	0	0
2022	CAM	0	0	0	5	0	0	0	0	4	2	0	1	2	0	0	0	0	2	0	1	0	0	2	0	3	1	0	0	4	1	1	0	2	1	0
2023	CAM	0	0	0	514	0	0	0	0	248	4	184	0	1	0	0	10	0	0	1	1	0	0	357	2	0	0	0	2	634	0	1	0	0	4	3
2024	CAM	0	0	0	248	0	0	0	0	159	1	39	0	0	0	0	1	0	0	2	0	0	0	162	0	0	0	0	4	151	0	0	0	0	0	2
2025	CAM	0	0	0	181	0	0	1	0	139	0	34	23	94	0	6	1	5	17	0	79	0	12	2	0	32	8	0	3	7	0	103	0	0	1	4
2026	CAM	0	0	0	199	0	0	0	1	265	0	139	1	102	4	0	5	1	5	0	32	0	280	1	0	53	223	0	33	0	0	0	0	6	0	3
2067	CAM	0	0	0	39	0	0	0	0	47	0	11	68	0	0	0	1	32	0	0	0	0	0	21	0	0	0	0	0	39	0	11	0	0	30	0
2068	CAM	0	0	0	191	0	0	0	0	105	1	125	0	0	0	0	3	97	0	0	0	0	0	148	0	0	0	10	0	174	184	0	0	0	151	0
2069	CAM	0	0	0	28	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	30	0	0	0	0	0	679	0	0	0	0	1	0
2078	CAM	0	0	0	221	0	0	0	0	0	0	115	1	89	0	12	0	8	16	0	6	0	1	212	1	10	1	0	0	102	0	118	0	2	1	0
2079	CAM	0	0	0	213	0	0	0	0	0	1	189	0	22	0	127	0	41	2	0	1	0	0	1	0	15	0	0	0	2	0	103	0	65	0	0
2080	CAM	0	0	0	231	0	0	0	0	0	6	58	0	0	0	0	0	0	0	0	0	0	0	537	0	1	0	0	0	287	0	580	0	0	1	0
2081	CAM	0	0	0	150	0	0	0	0	0	0	35	0	20	0	0	0	0	0	0	15	0	0	147	0	13	0	0	0	144	0	95	0	0	0	0
2082	CAM	0	0	1	10	0	0	0	0	0	0	1	0	3	0	1	0	0	3	0	1	2	0	2	0	0	0	0	0	9	0	144	0	1	0	0
2092	CAM	24	0	0	27	0	0	1	0	35	0	5	0	0	0	2	0	92	0	0	0	0	0	1	0	0	0	0	0	0	11	0	0	21	5	0
2093	CAM	0	0	0	196	0	0	0	0	0	0	99	0	11	0	0	0	0	1	0	44	0	0	15	0	26	0	0	0	198	0	37	0	0	0	0
2094	CAM	0	0	0	561	0	0	0	0	71	0	166	0	156	0	1	0	0	3	1	118	0	0	1	0	8	0	0	0	7	0	59	0	0	0	0
2301	CAM	0	2	0	144	0	0	0	0	59	0	100	0	1	0	0	1	0	0	0	0	0	0	32	0	0	0	0	30	88	0	472	0	0	0	163
2311	CAM	9	0	0	184	0	0	0	2	0	0	104	0	2	0	0	0	2	1	0	1	3	0	361	0	1	0	1	0	460	0	180	0	2	0	0
2312	CAM	0	0	0	87	0	0	0	0	0	0	149	0	0	0	58	0	0	0	0	0	0	0	6	0	0	0	0	0	166	0	19	0	81	0	0
2313	CAM	0	0	51	183	0	19	0	0	0	0	133	0	1	0	0	0	0	0	0	0	330	0	389	3	0	0	0	0	798	0	374	0	0	0	0
2314	CAM	0	1	0	447	0	0	0	0	197	3	284	0	80	0	0	1	0	0	0	42	0	0	3	0	52	0	0	8	10	0	198	0	0	0	27
23322	CAM	0	0	0	250	0	0	1	0	126	1	284	0	70	0	0	6	1	1	0	104	0	0	0	0	77	0	0	0	4	0	178	0	0	0	2
2333	CAM	0	0	0	292	0	0	1	0	0	1	225	0	80	2	1	0	0	6	0	66	0	241	3	0	32	141	0	0	1	0	91	0	0	0	0
2334	CAM	0	0	0	32	0	0	2	21	0	0	31	0	3	0	3	0	31	7	0	9	0	0	7	0	3	0	1	0	8	0	8	0	35	0	0
2335	CAM	0	0	0	114	0	0	5	0	1	0	46	0	14	1	2	0	3	90	0	16	0	0	51	0	24	0	0	0	139	0	100	0	0	0	0
2336	CAM	1	0	82	101	0	7	0	0	0	0	96	0	0	0	0	0	0	0	0	0	144	0	334	0	0	0	0	0	503	0	208	0	0	0	0
338724	CAM	0	0	0	160	0	0	0	1	0	4	221	0	3	2	6	0	1	4	1	5	0	0	364	0	4	0	2	0	373	0	91	1	1	1	0
384967	CAM	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	2	5	0	5	0	1	0	0	0	4	0	1	0	0	0	0
388167	CAM	0	0	0	148	0	0	0	3	0	3	122	494	0	0	45	0	66	0	0	0	0	0	255	0	0	0	0	1	378	0	0	0	0	0	0
388642	CAM	0	4	0	147	0	0	0	0	0	0	43	14	2	0	4	0	0	0	0	5	0	0	20	0	0	0	0	1	22	0	9	0	0	0	0
389835	CAM	0	0	0	38	0	0	0	0	0	0	122	0	0	0	0	0	0	0	0	424	0	0	31	0	4	0	0	0	532	0	78	0	0	0	0
389883	CAM	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	6	0	1	0	0	0	0
390804	CAM	0	0	0	80	0	0	0	0	42	0	48	40	0	0	0	0	27	0	0	0	0	0	49	0	0	0	0	0	132	0	77	0	0	114	0
405240	CAM	12	0	0	52	0	0	0	0	0	0	70	0	21	0	0	0	0	0	0	26	0	0	0	0	44	0	1	0	0	0	18	0	0	0	0
405242	CAM	0	0	0	260	0	0	0	0	0	3	60	1	1	0	2	0	1	0	0	4	0	0	384	1	2	0	8	0	428	0	140	0	0	0	0
406701	CAM	0	0	0	31	0	0	0	32	0	0	23	0	0	0	0	0	14	0	0	0	0	0	67	0	0	0	0	0	171	0	104	0	182	0	0
406853	CAM	0	0	0	139	0	0	9	0	0	0	10	0	23	0	0	0	0	269	0	61	1	0	0	0	41	0	0	0	0	0	80	0	0	0	0
412968	CAM	0	0	0	70	0	0	0	3	0	0	117	87	50	0	11	0	35	80	0	98	0	0	0	0	12	0	0	0	0	0	0	0	0	0	0
413399	CAM	0	0	0	156	0	0	0	0	56	1	208	0	0	0	0	4	96	0	1	0	0	0	85	0	0	0	0	0	297	162	103	0	0	81	0
413675	CAM	0	0	0	65	0	0	0	0	0	0	94	73	27	0	137	0	2	0	0	20	0	0	1	0	55	0	0	0	1	0	97	0	0	0	0
537730	CAM	0	0	0	64	0	0	0	0	0	0	26	0	5	0	0	0	0	0	0	12	0	0	61	2	10	0	1	0	75	0	18	0	0	0	0
680339	CAM	2	0	0	117	0	0	8	0	0	1	45	0	127	0	0	0	0	27	0	42	0	0	1	0	104	0	2	0	0	0	0	0	0	0	0
695125	CAM	0	0	0	114	0	0	0	0	0	3	203	0	44	0	0	0	0	0	0	110	0	0	78	2	39	0	0	3	192	0	52	0	0	0	0
726636	CAM	0	0	0	525	0	0	0	0	473	0	208	0	58	0	0	8	0	0	0	29	0	0	0	0	0	0	0	29	0	0	57	0	0	0	18
727148	CAM	0	0	0	52	0	0	0	0	0	0	35	0	0	0	0	0	1	1	0	114	0	0	0	0	12	5	0	0	0	0	28	0	0	0	0
7433814	CAM	0	0	0	32	0	0	0	0	0	0	1	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0
7433935	CAM	0	0	0	75	0	0	0	21	0	0	33	0	1	0	51	0	73	0	0	2	0	0	1	0	0	0	4	0	1	0	0	0	66	1	0
7434842	CAM	0	0	0	196	0	0	0	13	1	0	39	0	10	4	83	0	32	0	0	131	0	239	2	0	11	288	0	0	1	0	0	0	13	0	0
7435198	CAM	0	0	0	167	0	0	0	0	0	0	35	0	59	0	0	0	0	0	0	36	0	99	0	0	114	41	0	0	0	0	90	0	0	0	0
7435210	CAM	0	0	0	53	0	0	0	2	0	1	28	0	0	0	6	0	15	0	0	0	0	0	153	0	0	0	0	0	188	0	103	0	17	0	0
744840	CAM	0	0	12	100	0	3	0	0	2	4	190	4	1	1	1	0	0	0	0	0	92	0	284	5	2	0	0	0	459	0	97	0	0	2	1
1069851	CAS	0	0	0	159	0	0	22	0	0	0	34	0	65	0	1	0	0	154	0	63	0	0	7	0	48	0	0	0	7	0	4	0	0	0	0
1139525	CAS	0	0	0	372	0	0	7	0	62	0	122	0	148	0	0	1	1	89	1	31	1	0	3	0	50	0	0	35	1	0	1	0	0	0	0
1188598	CAS	0	0	0	81	0	0	0	0	0	0	235	0	17	0	0	0	0	7	0	22	0	0	38	0	6	0	0	0	41	0	48	0	0	0	0
1211451	CAS	1	0	0	87	0	0	0	22	0	0	104	8	22	0	35	0	91	11	0	39	1	0	10	0	15	0	9	0	20	0	8	0	47	1	0

2046	CAS	0	0	0	94	0	0	2	0	154	0	92	0	50	0	4	0	32	92	0	3	1	0	0	0	26	1	0	1	0	156	2	0	6	113	0
2048	CAS	0	0	0	249	0	0	25	0	1	0	90	3	105	0	0	0	0	466	0	130	3	0	0	0	140	0	0	0	0	0	0	0	0	0	0
2051	CAS	0	0	3	39	1	0	0	0	0	0	55	1	15	1	5	0	0	2	1	13	0	0	173	1	12	0	0	2	237	0	100	0	1	0	0
2052	CAS	0	0	0	78	0	0	9	0	0	0	204	0	46	0	0	0	0	136	0	199	1	0	0	0	29	0	0	19	0	0	0	0	0	0	0
2053	CAS	0	0	1	0	33	0	0	0	157	0	97	0	0	85	0	2	240	0	102	2	0	0	6	0	1	0	0	0	12	188	301	0	0	888	0
2054	CAS	0	0	0	5	11	0	0	0	4	0	58	8	63	36	7	1	8	1	81	35	1	0	5	0	84	0	2	0	6	5	110	0	0	222	0
2055	CAS	0	0	0	306	0	0	0	85	1	1	77	0	137	0	45	0	18	1	0	7	0	2	8	0	91	0	31	0	8	0	102	0	121	0	0
2056	CAS	0	0	0	18	0	0	0	0	0	0	0	0	2	0	0	0	63	99	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	83	0
2057	CAS	0	0	1	69	2	0	58	0	0	0	27	0	113	0	0	0	7	292	2	59	3	0	8	1	110	0	0	3	17	0	22	0	4	0	0
2058	CAS	0	0	0	162	0	0	12	0	0	3	61	3	48	0	3	0	0	187	0	19	3	0	175	1	24	0	0	1	116	0	84	1	0	0	0
2059	CAS	0	0	0	335	0	0	0	0	0	1	106	0	1	0	0	0	0	0	0	0	0	0	559	1	0	0	0	0	574	0	156	0	0	0	0
2060	CAS	0	0	0	73	0	0	0	18	1	0	54	2	9	1	19	0	15	22	0	39	0	0	1	0	28	0	0	0	0	0	1	0	53	6	0
2061	CAS	0	0	0	217	0	0	7	63	2	0	51	0	77	0	62	0	140	91	0	10	0	0	0	0	96	0	0	0	2	7	51	0	319	4	0
2062	CAS	0	0	0	99	0	0	0	55	1	0	12	0	0	0	46	0	98	0	0	0	0	0	89	0	0	0	0	0	293	0	124	0	96	0	0
2064	CAS	0	0	0	77	0	0	11	0	0	0	58	0	103	0	0	0	0	168	0	99	0	0	0	0	55	0	0	0	0	0	0	0	0	0	0
2065	CAS	0	0	0	283	0	0	11	0	0	0	177	0	222	0	0	0	0	408	0	218	3	0	0	0	185	0	11	0	0	1	0	0	0	0	0
2071	CAS	0	0	0	190	0	0	6	0	0	1	97	0	24	0	0	0	0	19	0	63	1	0	175	0	5	0	0	0	77	0	39	0	0	0	0
2072	CAS	0	0	0	54	0	0	1	0	0	0	12	0	36	0	24	0	30	120	0	7	0	0	1	0	40	0	0	5	2	0	0	0	1	0	0
2074	CAS	0	2	0	846	0	0	0	112	294	0	170	0	0	0	103	5	216	0	0	0	0	0	1	0	0	0	17	41	1	0	0	0	288	0	21
2075	CAS	0	0	3	169	4	0	7	0	0	0	58	0	85	45	0	0	0	263	43	32	0	0	0	0	24	0	0	0	2	5	231	0	0	230	0
2076	CAS	0	0	0	176	0	0	0	0	0	1	14	0	0	0	13	0	169	0	0	0	0	0	304	0	0	0	0	7	223	0	108	0	0	0	0
2077	CAS	0	0	0	129	0	0	0	10	0	0	195	0	31	0	6	0	230	258	0	49	0	0	2	0	35	0	0	0	0	0	177	0	59	0	0
2086	CAS	0	0	0	1	0	0	0	0	0	0	4	141	0	0	35	0	10	0	0	101	0	0	3	0	1	0	0	7	6	0	0	0	0	0	0
2087	CAS	0	0	0	160	0	0	4	0	0	6	217	1	3	0	0	0	0	143	0	4	0	0	315	1	13	0	0	1	151	0	32	0	3	0	0
2088	CAS	0	0	0	152	0	0	0	0	216	0	109	0	21	0	0	3	96	0	0	11	0	0	0	0	108	0	0	113	0	0	70	0	0	110	0
2089	CAS	0	0	0	15	0	0	1	0	0	0	217	0	7	0	1	0	98	0	0	2	0	0	2	0	4	0	0	0	89	0	3	0	2	0	0
2090	CAS	0	0	0	86	0	0	0	83	0	0	36	335	0	0	135	0	100	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	59	0	0
2300	CAS	0	0	0	236	0	0	42	0	0	0	172	0	170	0	0	0	0	352	0	119	0	0	0	0	181	0	5	0	0	0	1	0	0	0	0
2303	CAS	0	0	0	162	0	0	32	0	0	0	103	1	81	0	0	0	0	189	0	125	3	0	0	0	80	0	0	0	0	0	1	0	0	0	0
2304	CAS	0	0	0	122	0	0	13	0	0	0	40	0	27	0	0	0	1	207	0	77	0	0	1	0	51	0	0	0	1	0	0	0	0	0	0
2305	CAS	0	0	0	37	0	0	0	0	0	2	108	0	6	0	2	0	4	155	0	3	1	2	0	0	128	0	0	0	1	0	5	0	2	0	0
2306	CAS	0	0	0	118	0	0	0	1	0	0	12	0	1	0	136	0	102	5	0	1	0	0	0	0	3	0	2	0	0	0	92	1	36	0	0
2307	CAS	0	0	0	58	0	0	141	0	0	0	26	1	2	44	0	0	0	57	2	9	0	0	0	0	6	0	0	0	0	0	511	0	0	18	0
2308	CAS	0	0	0	247	0	0	0	48	0	0	81	1	0	0	17	0	174	0	0	0	0	0	197	1	0	1	0	1	329	0	285	0	173	0	0
2321	CAS	0	0	0	394	0	0	20	0	0	0	134	0	139	0	0	0	0	197	0	25	0	0	226	0	15	1	0	0	493	1	91	0	0	3	0
2743	CAS	0	0	0	62	0	0	0	0	0	0	19	0	7	0	0	0	6	52	0	20	1	0	0	0	7	0	0	0	0	0	0	0	0	0	0
3000	CAS	2	0	0	107	0	0	0	0	92	0	79	0	0	0	0	1	133	1	0	0	0	0	294	0	0	0	0	0	233	175	127	0	0	80	0
338077	CAS	0	0	0	11	0	0	0	0	0	0	11	0	0	0	0	0	0	49	0	15	0	0	0	0	6	0	0	3	1	0	0	0	0	0	0
338112	CAS	0	0	0	60	0	0	0	1	0	0	18	0	0	0	57	0	0	1	0	0	0	0	183	1	0	0	0	0	118	0	188	11	0	0	0
338733	CAS	0	0	0	18	0	0	0	0	0	0	70	0	7	0	2	0	3	0	0	3	0	0	0	0	57	0	0	0	1	0	0	0	24	0	0
338857	CAS	0	0	2	69	2	0	0	0	0	2	68	0	0	49	0	0	0	0	77	0	0	0	42	0	0	0	0	0	318	0	137	0	0	65	0
387100	CAS	0	0	0	37	0	0	0	0	0	1	12	0	0	0	14	0	73	0	0	0	0	0	85	0	0	0	0	2	29	0	131	0	0	0	0
388783	CAS	0	0	0	80	0	0	14	0	0	0	29	0	7	0	5	0	8	10	0	7	0	0	0	0	32	0	0	17	0	0	0	0	0	0	0
388918	CAS	0	0	0	159	0	0	0	0	0	11	258	0	0	0	0	0	0	1	0	1	0	0	304	4	1	0	0	0	524	0	179	0	0	0	0
389332	CAS	0	0	0	4	0	0	0	0	102	0	94	0	37	0	0	1	86	1	0	10	0	0	0	0	61	0	0	0	1	142	0	0	0	62	0
389361	CAS	0	0	0	67	0	1	0	8	36	1	166	0	1	0	48	0	75	4	0	1	2	0	4	0	0	0	0	0	5	153	36	0	87	81	0
389630	CAS	0	0	1	48	0	0	2	3	1	0	24	0	6	0	16	0	43	0	3	9	0	0	4	0	17	0	0	10	3	1	44	0	0	0	0
405994	CAS	0	0	0	22	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	1	0	0	2	0	0	0	0	0	20	0	4	0	0	0	0
406443	CAS	0	0	0	429	0	0	0	71	0	0	214	1	0	0	217	0	352	0	0	1	0	0	5	0	1	0	19	1	4	0	185	0	161	0	0
406817	CAS	0	69	0	112	0	0	0	3	101	0	24	0	0	0	10	0	5	0	0	0	0	0	0	0	0	0	0	46	0	0	0	0	0	0	0
408967	CAS	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	9	0	0	0	0	4	0	0	0	0	0	1	0	0	0	0
409374	CAS	0	0	24	0	0	0	0	0	0	0	55	0	2	286	0	0	0	0	146	40	0	0	0	0	71	0	0	0	0	0	95	0	0	363	0
409419	CAS	0	0	76	39	6	0	15	0	1	0	17	0	46	16	4	0	9	38	13	11	0	0	17	0	10	0	0	2	27	1	27	0	2	56	0
413195	CAS	0	0	0	0	0	0	0	39	0	0	1	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
413453	CAS	0	0	0	194	0	0	0	0	0	0	36	0	20	0	115	0	138	0	0	24	0	0	0	0	53	0	0	23	0	0	0	0	59	0	0
528385	CAS	0	0	0	309	0	0	0	200	1	0	79	0	0	0	75	0	114	0	0	0	0	0	151	0	0	0	0	0	129	0	370	0	236	0	0
7431836	CAS	0	0	0	76	0	0	5	7	0	1	393	3	61	0	4	0	4	34	0	38	0	0	15	0	223	0	3	0	23	0	11	0	8	0	0
7433111	CAS	0	0	0	97	0	0	0	21	0	0	42	0	30	1	26	0	92	0	1	49	0	0	0	0	65	0	2	7	3	0	2	0	39	2	0
7434345	CAS	0	0	0	8	0	0	0	0	0	0	3	0	1	0	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7434859	CAS	0	0	1	108	0	0	0	5	0	0	308	0	82	1	131	0	88	1	1	51	0	0	0	0	20	0	1	27	1	0	1	0	146	0	0
7435152	CAS	0	0	0	37	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	3	0	0	12	0	0	0	0	0	0	0	1	0	0	0	0
7435313	CAS	0	0	40	69	21	0	0	0	0	0	34	0	0	82	0	0	2	0	157	0	0	0	14	1	0	0	0	0	219	0	127	0	0	33	0
743824	CAS	0	0	0	0	0	0	0	0	0	0	0	0	16	0	0	0	0	15	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
761171	CAS	0	0	0	94	0	0	0	0	0	0	35	0	0	0	0	0	0	0	0	1	0	0	96	0	11	0	0	0	164	0	47	0	0	0	0
763205	CAS	0	0	0	395	0	0	54	0	0	3	189	0	140	0	0	0	1	610	0	137	4	0	0	0	139	0	0	0	0	0	2	0	0	0	0
436591	CAS	0	0	0	120	0	0	6	0	1	0	95	0	143	0	1	0	0	175	0	164	0	0	4	0	122	0	0	1	2	0	1	0	0	0	0
1050417	CRR	0	0	0	24	0	0	0	0	0	4	33	0	0	0	0	0	0	0	0	0	0	0	123	4	0	0	0	0	182	0	48	0	0	0	0
1053251	CRR	0	0	8	6	0	1	0	0	0	0	12	0	8	0	0	0	2	2	0	7	16	0	16	1	6	0	0	0	12	0	0	0	0	2	0
1166665	CRR	0	0	0	170	0	2	8	1	7	0	148	429	29	4	24	0	42	314	5	177	13	5	78	0	40	4	0	0	90	0	36	0	5	10	0
1225866	CRR	0	0	0	4	0	0	0	0	0	0	4	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	9	2	2	0	0	0	0
1226553	CRR	0	0	0	320	1	0	37	0	0	2	236	2	58	0	0	0	0	41	0	49	0	0	437	1	82	0	0	0	367	0	54	0	0	0	0
2027	CRR	0	0	0	0	0	0	0	0	0	0	145	520	0	0	191	0	172	3	0	0	0	0	0	0	0	0	0	7	1	0	1	0	0	1	0
2028	CRR	0	0	0	92	0	0	0	0	0	0	55	0	84	0	42	0	41	143	0	9	1	0	0	0	58	0	0	6	0	0	0	0	0	0	0
2029	CRR	0	0	0	92	0	0	0	0	0	0	9	0	69	0	0	0	1	180	0	91	1	0	1	0	13	0	0	0	3	0	0	0	0	0	0
2030	CRR	0	0	0	178	0	0	0	0	1	0	77	0	8	0	1	0	0	0	0	25	0	0	199	0	5	0	0	0	248	0	50	0	0	0	0
2031	CRR	0	0	0	570	0	0	0	74	1	0	68	2	253	1	105	0	152	329	0	52	0	3	0	0	19	2	0	0	0	0	0	0	115	1	0
2032	CRR	0	0	0	76	3	0	0	0	0	0	15	0	74	5	0	0	0	0	6	63	0	1	0	0	1	0	0	0	0	0	58	0	0	84	0
2033	CRR	0	0	0	6	0	0	2	0	0	0	32	170	1	0	109	0	192	1	0	1	0	0	0	0	0	0	0	20	2	1	0	28	0	0	0
2034	CRR	0	0	63	4	54	0	0	1	0	0	54	0	1	42	187	0	100	0	346	0	0	0	1	0	0	0	0	0	4	0	2	0	0	65	0
2035	CRR	0	0	0	184	0	0	0	2	0	0	39	160	101	0	135	0	11	125	0	5	0	0	1	0	52	0	0	0	3	0	2	0	1	0	0
2036	CRR	0	0	0	321	0	0	8	87	0	0	182	2	112	0	142	0	138	208	0	88	0	0	0	0	47	0	12	0	0	0	0	0	130	0	0
2038	CRR	0	0	5	224	0	14	0	0	0	5	123	0	0	0	0	0	1	1	0	0	348	0	916	4	1	0	0	0	1032	0	251	0	0	0	0

2039	CRR	0	0	0	137	0	0	32	0	3	0	28	0	127	0	0	0	2	213	0	103	1	0	1	0	44	0	0	0	0	0	2	0	0	2	0
2040	CRR	0	0	0	160	0	0	18	0	0	1	101	311	78	0	71	0	46	333	0	94	2	0	1	0	31	0	0	0	0	0	0	0	0	0	0
2041	CRR	0	0	0	18	0	0	1	0	2	0	54	246	40	0	61	0	69	7	2	84	0	0	0	0	25	1	0	14	0	1	2	0	0	7	0
2042	CRR	0	0	0	418	0	0	15	0	5	1	103	6	413	1	0	0	0	487	0	396	8	1	2	0	255	1	0	0	1	0	0	0	0	0	0
2043	CRR	0	0	0	271	0	0	5	0	0	0	106	0	158	0	1	0	0	431	0	151	2	0	0	0	46	0	0	0	0	0	1	0	0	0	0
2044	CRR	0	0	0	51	0	1	0	0	0	0	36	0	0	0	0	0	0	0	0	0	66	0	174	0	0	0	0	0	227	0	31	0	0	0	0
2070	CRR	0	0	0	130	0	0	0	0	1	0	112	0	15	0	0	0	0	167	1	101	0	0	0	0	219	0	0	0	0	0	0	0	0	0	0
2085	CRR	0	0	0	73	0	0	1	0	0	2	53	103	1	0	62	0	97	145	0	9	1	0	0	0	3	0	0	0	0	0	1	0	0	0	0
2096	CRR	0	0	0	203	0	0	0	0	0	3	104	0	52	1	1	0	0	160	0	70	1	0	190	2	20	0	0	0	340	0	102	1	0	0	0
2098	CRR	7	1	0	81	0	0	15	15	0	0	99	0	45	0	26	0	38	192	0	9	2	0	0	0	6	0	0	0	0	0	1	0	79	0	0
2099	CRR	0	0	0	61	0	0	0	1	0	0	51	0	36	2	0	0	12	3	0	28	0	0	65	0	19	0	0	32	87	0	57	0	0	10	0
2302	CRR	0	0	0	230	0	0	0	0	0	1	206	331	79	0	77	0	66	0	0	46	0	0	0	0	70	0	7	0	0	0	0	0	0	0	0
2315	CRR	0	0	0	59	0	0	0	0	0	0	211	0	2	0	56	0	12	3	0	3	0	0	96	0	1	0	0	0	88	0	86	0	0	0	0
2317	CRR	0	0	0	0	0	0	0	0	1	1	58	275	5	29	150	0	28	1	45	1	0	0	0	0	1	0	0	0	1	0	31	0	0	288	0
2318	CRR	0	0	2	243	8	0	18	2	0	0	57	0	60	127	0	0	0	425	202	99	1	0	6	0	41	0	0	1	2	0	158	0	0	224	0
2337	CRR	0	0	0	2	0	0	0	0	0	0	47	233	89	0	44	0	38	1	0	23	0	0	0	0	60	0	0	10	0	0	1	0	1	1	0
2338	CRR	0	0	0	44	0	0	3	46	0	0	51	4	29	0	4	0	25	105	0	3	0	0	0	0	45	0	0	0	3	0	0	0	72	0	0
2339	CRR	0	0	0	62	0	0	0	100	0	0	26	0	5	0	94	0	113	0	0	176	0	0	0	0	17	0	0	0	0	0	0	0	66	0	0
2340	CRR	0	0	0	69	0	0	2	0	0	0	11	1	47	0	0	0	0	432	0	173	6	0	0	0	89	0	0	0	0	0	0	0	0	0	0
2400	CRR	0	0	0	219	0	0	1	9	0	0	36	0	26	0	18	0	250	308	0	139	3	0	0	0	37	0	5	0	0	0	0	0	19	0	0
2401	CRR	0	0	0	4	0	0	0	0	1	0	6	0	0	0	5	0	115	0	0	0	0	0	3	0	0	0	0	0	2	1	0	0	0	0	0
386733	CRR	0	0	0	101	0	0	101	0	0	0	42	0	48	0	1	0	0	172	0	104	3	0	0	0	19	0	0	0	2	0	0	0	0	0	0
389012	CRR	0	0	1	273	0	0	0	132	0	1	44	0	1	8	64	0	74	0	2	2	1	1	69	0	3	0	7	0	112	0	46	0	21	2	0
389316	CRR	0	0	0	41	0	0	0	0	1	1	16	0	0	0	0	0	0	0	0	0	0	0	45	0	0	0	0	0	54	0	30	0	0	0	0
389777	CRR	0	0	0	45	0	0	0	0	0	1	101	0	11	0	0	0	0	0	0	60	0	0	110	1	22	0	0	0	128	0	185	0	0	0	0
389833	CRR	0	0	1	2	0	0	0	0	0	0	0	0	0	0	60	0	0	1	0	0	2	0	3	0	0	0	0	0	2	0	0	0	0	0	0
389902	CRR	0	0	0	1	0	0	0	0	0	0	32	35	1	0	5	0	0	6	0	2	0	0	2	0	3	0	0	0	1	0	0	0	0	1	0
390562	CRR	0	0	0	137	0	0	7	0	1	0	9	0	56	0	0	0	1	130	0	10	0	0	0	0	50	0	0	0	0	0	1	0	0	0	0
406175	CRR	0	0	0	3	0	0	0	0	0	0	1	0	0	0	0	0	0	2	0	0	0	0	1	0	10	0	0	0	0	0	3	0	0	0	0
406298	CRR	0	0	0	0	0	0	0	0	0	0	2	30	2	0	1	0	17	0	0	3	0	0	0	0	4	0	0	1	0	0	0	0	0	0	0
412998	CRR	0	0	0	4	0	0	0	0	0	0	0	8	0	0	1	0	20	556	0	9	1	0	1	0	3	0	0	0	0	0	0	0	0	0	0
413423	CRR	0	0	36	134	0	8	2	0	0	0	26	0	86	0	1	0	0	138	0	92	185	0	240	1	53	0	0	0	396	0	0	0	0	0	0
413492	CRR	0	0	0	84	0	0	4	3	1	0	35	0	10	1	11	0	31	49	0	11	3	0	0	0	25	0	3	0	0	0	0	0	42	0	0
413720	CRR	0	0	0	8	0	0	0	0	0	0	0	1	3	0	0	0	0	2	0	9	2	0	0	0	10	0	0	0	0	0	0	0	0	0	0
413732	CRR	0	0	0	3	0	0	0	0	0	0	1	0	1	0	0	0	0	2	0	0	0	0	2	0	0	0	0	0	26	0	0	0	0	0	0
413760	CRR	0	0	0	79	0	0	0	0	0	3	30	0	0	0	0	0	19	0	0	0	0	0	50	1	0	0	0	0	7	0	24	0	4	1	0
413849	CRR	0	0	1	1	0	0	0	0	0	0	0	0	1	1	0	0	0	4	0	0	0	0	0	0	1	0	0	0	0	0	6	0	0	24	0
535887	CRR	0	0	16	0	2	0	0	0	0	0	1	0	11	131	0	0	0	0	30	35	0	0	0	0	31	0	0	0	0	0	107	0	0	60	0
536503	CRR	0	0	0	76	0	0	0	0	0	5	217	251	0	0	43	0	38	0	0	0	0	0	263	1	0	0	0	0	421	0	134	0	0	0	0
717048	CRR	0	0	0	54	0	0	0	0	0	2	107	141	0	0	26	0	40	0	0	0	0	0	133	1	0	0	0	0	141	0	90	0	0	0	0
722442	CRR	0	0	14	12	0	0	0	0	0	0	2	0	1	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	99	0
726439	CRR	0	0	0	5	0	0	3	0	0	0	11	0	7	0	0	0	0	21	1	8	18	0	15	0	7	0	0	0	73	0	0	0	0	0	0
7431549	CRR	0	0	0	133	0	0	3	0	0	0	91	0	4	0	42	0	40	91	0	6	0	0	0	0	58	0	0	0	0	0	0	0	0	0	0
7432946	CRR	0	0	0	0	0	0	0	0	0	0	0	114	0	0	0	0	43	0	0	55	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0
7433129	CRR	0	0	21	164	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	23	0	88	0	0	0	0	0	344	0	62	0	0	0	0
7433582	CRR	0	0	0	359	0	0	0	0	0	3	137	0	17	0	0	0	0	0	0	11	0	0	114	0	10	0	1	0	33	0	45	0	0	0	0
7434569	CRR	0	0	0	398	0	0	0	0	0	11	151	0	0	0	0	0	0	0	0	0	2	0	688	2	1	0	0	0	688	0	170	0	1	0	0
7434657	CRR	0	0	0	15	0	0	0	1	4	0	3	0	0	0	3	0	4	0	0	0	0	0	0	0	0	0	2	3	0	0	0	0	5	0	0

Egernia stokesii MHC II genotypes 2012-13
Source file: Genotypes & Samples filtered worksheet in Hawker_MHCII.xlsx at \Estokesii_MHCdownstream_17112015\jMHC\Hawker2012-13
Bold text in grey highlighted cells indicate reads >2 = called as an allele

			Allele name	Egst-DRB*03	Egst-DRB*04	Egst-DRB*01	Egst-DRB*05	Egst-DRB*02
			Sequence ID	274	522	415	785	1355
Sample ID	Site	Age_class	Sex					
338724	CAM	A	M	1	0	**3**	**1709**	**123**
388167	CAM	A	U	0	0	0	**2333**	2
406701	CAM	A	M	0	0	0	**3323**	**7**
406853	CAM	A	U	0	1	0	**1965**	**467**
412968	CAM	A	U	**18**	0	0	0	**3165**
413399	CAM	A	M	0	0	0	**2256**	1
413675	CAM	A	U	**12**	0	0	**2665**	1
537730	CAM	A	U	0	0	0	**4036**	1
680339	CAM	A	M	0	0	**364**	**5**	**1700**
695125	CAM	A	U	0	0	0	**2794**	**106**
1046168	CAM	A	U	0	1	**51**	**2088**	0
1055885	CAM	A	U	0	0	**19**	**4897**	**13**
1128694	CAM	A	M	0	0	**3**	**1888**	**125**
7433935	CAM	A	U	0	0	**3**	2	**2148**
2006	CAM	J	U	0	0	0	**2607**	1
2026	CAM	J	U	0	1	0	**30**	**3560**
2066	CAM	J	U	0	0	0	**1835**	0
2094	CAM	J	U	0	0	1	**1900**	**8**
2310	CAM	J	U	**3**	0	1	**2248**	**7**
2311	CAM	J	U	0	0	2	**3494**	**30**
2333	CAM	J	U	0	0	0	**2453**	**150**
2334	CAM	J	U	0	1	0	**181**	**2905**
2335	CAM	J	U	0	1	1	**944**	**833**
2008	CAM	S	U	0	0	0	**3070**	**87**
2018	CAM	S	U	**13**	1	**14**	**1689**	**6**
2020	CAM	S	U	0	0	0	**1842**	**226**
2021	CAM	S	U	0	0	**1118**	**731**	**50**
2023	CAM	S	U	0	0	2	**4463**	0
2025	CAM	S	U	0	0	0	**2109**	**580**
2068	CAM	S	U	0	0	0	**2778**	**4**
2314	CAM	S	U	0	0	1	**1839**	**355**
390804	CAM	S	M	1	**641**	0	**1673**	**3**
405240	CAM	S	M	0	0	**122**	**2889**	1
405242	CAM	S	U	0	0	1	**3048**	**5**
201010	CAM	S	U	0	0	0	**2184**	**27**
338077	CAS	A	U	0	0	0	2	**1456**
338733	CAS	A	M	0	0	0	0	**2301**
387100	CAS	A	M	0	0	0	**1198**	**517**
389332	CAS	A	M	0	0	0	2	**3166**
406443	CAS	A	U	0	0	0	**2422**	**323**
528385	CAS	A	U	0	0	0	**2019**	0
1069851	CAS	A	U	0	0	0	**43**	**1954**
1211451	CAS	A	U	0	0	1	**114**	**1874**
7431836	CAS	A	U	0	0	2	**291**	**2222**
436591	CAS	A	M	0	0	0	**9**	**3311**
2051	CAS	J	U	0	0	0	**2185**	**10**

2062	CAS	J	U	0	0	0	**2627**	2
2077	CAS	J	U	0	0	0	**1902**	2
2308	CAS	J	U	0	0	0	**4109**	**9**
2055	CAS	S	U	0	0	0	**2260**	**21**
2059	CAS	S	U	0	0	1	**2673**	**5**
2061	CAS	S	U	0	0	0	**2738**	**42**
2071	CAS	S	U	0	0	0	**1048**	**928**
2072	CAS	S	U	0	0	0	**9**	**2775**
2300	CAS	S	U	0	0	0	**4**	**3831**
2303	CAS	S	U	0	0	**3**	**11**	**2776**
2304	CAS	S	U	0	0	0	**6**	**2188**
2321	CAS	S	U	0	0	0	**3149**	**7**
3000	CAS	S	U	0	0	0	**3306**	**3**
413453	CAS	S	U	0	0	0	1	**2332**
7433111	CAS	S	U	0	2	0	**26**	**3202**
389316	CRR	A	U	0	0	0	**3357**	1
389902	CRR	A	U	**7**	1	**40**	**34**	**1964**
406175	CRR	A	U	0	0	1	**1755**	**7**
413760	CRR	A	U	0	0	1	**962**	**1467**
413849	CRR	A	U	0	**730**	**3**	**32**	**1785**
717048	CRR	A	M	**25**	0	0	**2959**	**4**
726439	CRR	A	U	0	0	**832**	0	**471**
1050417	CRR	A	M	0	1	**14**	**3632**	1
1053251	CRR	A	U	0	0	**4623**	**19**	**21**
1225866	CRR	A	M	0	**1107**	1	**907**	**6**
1226553	CRR	A	U	0	1	1	**2326**	**4**
2027	CRR	S	U	**52**	**5**	**3**	**16**	**2581**
2035	CRR	S	U	**5**	0	0	**131**	**1813**
2036	CRR	S	U	1	0	**15**	1	**3289**
2038	CRR	S	U	0	1	**80**	**1721**	**13**
2039	CRR	S	U	0	0	0	**13**	**3145**
2044	CRR	S	U	0	0	**141**	**2092**	1
2070	CRR	S	U	0	0	0	**8**	**2269**
2302	CRR	S	U	**130**	0	**2554**	2	**7**
389012	CRR	S	U	0	0	0	**1352**	**1070**
389777	CRR	S	M	0	0	0	**966**	**748**
389833	CRR	S	U	0	0	**4**	**29**	**2245**
406298	CRR	S	U	**266**	0	0	**4**	**3439**
413423	CRR	S	M	0	0	**650**	2	**802**
413492	CRR	S	M	0	0	0	0	**2680**
413732	CRR	S	U	0	0	2	**471**	**1175**
7433582	CRR	S	M	0	0	**470**	**1241**	**3**
7434569	CRR	S	U	0	0	1	**3290**	**11**

Appendix 6: *Egernia stokesii* longevity

During 2012-13 field surveys four lizards were captured that were first captured during 1993-98 surveys and again in 2003-04, which permitted a review of *E. stokesii* longevity estimates. Two lizards were subadults (sex unknown) when first caught in 1994/95 and two were adult males when first captured (1993/94, 1994/95). Adults mature at a minimum of six years of age (Duffield and Bull, 2002). Therefore minimum longevity is 25 years, which is consistent with previous estimates (Swan, 1990).

References

DUFFIELD, G. A. & BULL, C. M. 2002. Stable aggregations in an Australian lizard, *Egernia stokesii*. *Naturwissenschaften,* 89**,** 424-427.

SWAN, G. 1990. *A field guide to the snakes and lizards of New South Wales*, Three Sisters Productions.

Appendix 7: MHC allele clustering analysis

Background

While undertaking analyses for testing for a genetic basis of *E. stokesii* mate choice (Chapter 7), three clustering methods (k-means, multidimensional scaling, and Ward's hierarchical clustering) were used to explore if MHC alleles could be grouped into functional units or 'supertypes' (Doytchinova and Flower, 2005, Schwensow et al., 2007).

Method

Due to low variability in *E. stokesii* MHC II, clustering analyses was run for MHC I only. For each MHC I allele (n=31), each amino acid was described using five z-descriptors: z1 (hydorphobicity), z2 (steric bulk), z3 (polarity), and z4 and z5 (electronic effects) (Sandberg et al. 1998). First a matrix (Doytchinova and Flower, 2005) consisting of rows of amino acids and columns for five times the z-descriptors was constructed. We then used three clustering methods: 1) k-means, 2) multidimensional scaling (MDS), and 3) hierarchical clustering to determine if alleles could be grouped into clusters or supertypes. All clustering analyses were performed in R. For k-means we ran the *find.clusters* and *dapc* functions in the *adegenet* package. For MDS we used the *cmdscale* function and for hierarchical clustering we adopted Ward's method ("ward.D2") using the *hclust* function.

Results

The three approaches identified between nine and 13 clusters using 31 MHC I alleles. Most clusters contained between one and two alleles. None of the clustering methods resulted in distinct clusters of alleles. Therefore, the use of the supertype approach was not warranted.

References

DOYTCHINOVA, I. A. & FLOWER, D. R. 2005. In silico identification of supertypes for class II MHCs. *Journal of Immunology,* 174**,** 7085-7095.

SCHWENSOW, N., FIETZ, J., DAUSMANN, K. H. & SOMMER, S. 2007. Neutral versus adaptive genetic variation in parasite resistance: importance of major histocompatibility complex supertypes in a free-ranging primate. *Heredity,* 99**,** 265-277.

Appendix 8: *Egernia stokesii* microsatellite genotypes

2003-04 microsatellite genotypes

Data in Genepop format, in population order of Camel Hill, Castle Rock, Castle Rock Ridge.

Source file for 2003-04 data: Genepop_Hawker0304_Adults_3pops.xls at \Laboratory_Work\Microsatellites\Analyses\Genepop\2003-04

Title line:"Egernia stokesii Hawker 0304 Adults 3 pops"

Ecu2

Ecu5

Est3

Est8

Tr3.2

Tr4.11

Tr5.21

Ecu1

Ecu4

Est1

Est4

Tr5.20

pop

3 , 150168 116118 262325 106106 200210 113113 081081 111170 064068 270278 179186 153153

8 , 136136 115115 305334 106122 216216 113113 081081 000000 068068 246246 179183 000000

9 , 136150 113113 325348 106122 206206 113113 081081 113113 072072 238238 163163 153157

13 , 136150 113115 303321 106114 194194 113113 081083 111111 064068 250252 151157 000000

17 , 136136 113115 329348 114126 194194 113113 081081 113170 064068 242254 163163 181189

537730 , 136150 115117 317317 114122 194194 113113 081081 113113 068072 246255 139155 000000

547656 , 136136 113118 262309 110114 172172 113113 081081 111113 064068 250250 155155 153153

555867 , 146150 110114 266305 106114 188196 113113 079081 113113 064072 247250 155155 153153

557783 , 136168 113115 334348 114126 194206 113113 079081 113170 064068 242254 163163 181185

695125 , 136150 113113 325348 106122 000000 000000 000000 113177 072072 238238 163163 000000

722679 , 136148 115116 321321 106122 000000 000000 000000 113113 064068 242255 159159 000000

727148	,	150150	115116	305305	110122	208212	113113	081081	113113	064068	255270	163163	179179
728602	,	136158	115116	305313	114114	198198	000000	000000	113113	064068	250262	155155	000000
740884	,	136158	116118	317325	106130	194198	113113	081081	113113	068072	255255	155155	000000
764862	,	148150	115116	348348	106126	206208	113113	081081	113113	064072	242254	159159	000000
1046168	,	136150	115115	321325	106110	000000	000000	000000	000000	064068	000000	159159	000000
1051638	,	136158	116118	317325	106106	194198	113113	081081	111111	072072	234246	167167	191191
1055885	,	136150	114115	317325	114122	202216	113113	081081	111111	068072	254254	163175	185189
1062514	,	136136	115116	317325	110114	194194	113113	075075	177183	072072	254266	155163	187189
1064549	,	148150	115115	309309	106120	208208	113113	079081	111162	064064	242250	183183	157157
1068267	,	150154	114121	266348	110114	196196	113113	081081	113162	068072	247250	183183	000000
1128694	,	134136	115115	305305	106122	000000	000000	000000	000000	064064	250250	163163	000000
1221712	,	150154	114121	305305	114130	196196	113113	075081	113162	068072	247252	155163	000000
pop													
7	,	150150	115115	317329	110110	196196	000000	000000	182182	068072	230255	155155	000000
41	,	150150	115117	317325	110114	000000	000000	000000	111111	068068	250255	183187	000000

42	,	150150	115116	348348	106110	208208	113113	081081	170177	068072	238242	155155	157157
528385	,	150158	114115	309309	106110	204212	113113	081081	186186	064072	254270	155171	191191
692367	,	136136	113114	309325	110110	190194	113113	081081	111111	072072	234255	171171	000000
692554	,	150150	115117	293317	106110	208208	113113	081081	113113	064064	254270	179179	000000
695336	,	148150	117117	309325	122126	190190	113113	081081	183186	072072	255262	155159	000000
722730	,	150150	113114	309309	106114	190190	113113	081081	111111	064068	234238	155155	000000
723039	,	136150	115116	305305	110126	194194	113113	081081	175177	072072	238238	171171	157157
724852	,	150150	115115	329329	110110	196196	113113	081081	186186	064072	266266	167171	000000
726205	,	150150	113117	307307	110110	000000	113113	090090	113113	064064	255270	163163	000000
737727	,	148150	115117	321321	106126	190190	113113	081081	183186	072072	255262	155159	000000
743824	,	148150	115117	309309	110110	190190	113113	000000	111111	072072	255262	171171	187189
763205	,	150150	115117	301317	110126	196200	113113	081081	111113	068072	250270	171171	153153
1049471	,	150150	115117	301317	106110	208208	113113	000000	175177	068072	238254	151151	000000
1060612	,	136148	113117	309313	110122	190194	113113	081081	183186	072072	262262	155155	000000
1061198	,	150150	115115	301317	106114	208212	113113	081081	113113	064064	255270	155155	153153

1069851 , 136148 114117 309325 110110 190190 113113 081081 111113 072072 262266 155171 000000

1188598 , 150150 115117 293293 110110 212212 113113 081081 175183 064072 250254 167171 000000

1211451 , 150150 115116 313313 106126 202212 113113 081081 111111 072072 234234 155171 000000

1236178 , 148150 115115 325338 106110 208208 113113 081081 111111 064072 250255 155155 181189

pop

1 , 136136 114116 293301 110110 190190 113113 081081 111111 064072 250266 163163 189189

5 , 150150 116117 305313 110110 202202 113113 081081 111111 072072 250266 155155 187189

22 , 136150 114115 325348 110110 190190 113113 081081 111111 068072 234255 171171 181189

74299 , 136150 113117 317317 106122 194194 000000 000000 111111 072072 000000 171171 000000

535887 , 136150 110112 286293 106106 206208 113113 079081 173177 064072 234238 163171 157157

536503 , 150150 115115 258293 106110 200208 113113 081081 111111 064072 238250 171171 157157

687591 , 148150 115115 325329 106106 208208 113113 081081 111111 064072 250262 159171 189189

692453 , 150150 115115 293344 110114 000000 113113 081081 111113 068068 234234 155163 153191

696183 , 150170 112118 313313 110122 172172 113113 081081 111113 064068 242246 155167 000000

710781 , 130136 108113 309309 106122 000000 113113 081083 111163 064068 234238 163163 000000

717048	,	146150	109109	309309	114122	190190	113113	077077	111113	064068	234234	155155	153153
721450	,	142150	110112	305305	110114	000000	113113	077081	113113	062072	232270	167167	000000
722442	,	146150	113114	305313	110110	202202	113113	079081	186186	064072	266266	155155	000000
726439	,	144150	109111	262262	110122	000000	110113	079079	111113	064072	262270	163163	153153
746686	,	150170	116118	313321	110110	172172	000000	000000	111113	068072	250250	155155	000000
1043823	,	150150	115117	313317	110110	212212	113113	081081	111113	072072	250270	155155	000000
1050417	,	150150	108112	262313	106110	172172	113113	081081	111113	064068	234236	167167	000000
1053251	,	144150	107110	262262	110110	202202	113113	081081	111111	064068	234234	175179	191191
1056061	,	150150	108108	301329	114114	168168	113113	081081	137137	064068	252252	167167	000000
1061450	,	136158	113118	305317	106110	198206	113113	081081	175177	064068	234242	139171	000000
1066889	,	148150	114115	309325	110114	190190	113113	081081	111113	068072	238238	159163	000000
1131024	,	136158	116118	301313	106110	198208	113113	081081	111113	064072	255270	159159	000000
1166665	,	150150	113117	301329	122126	000000	113113	081081	111111	072072	250250	171171	181189
1223638	,	150170	116117	325325	110110	172172	113113	081081	186186	064068	242242	171183	000000
1225866	,	136150	116116	305305	102126	192192	113113	079083	113113	064068	233252	139153	000000

1235200	,	136136	115116	301301	110114	172172	113113	081081	111111	068072	266278	171171	000000

2012-13 microsatellite genotypes

Data in Genepop format, in population order of Camel Hill, Castle Rock, Castle Rock Ridge.

Source file for 2012-13 data: Genepop_Hawker1213_NonFullSibAdults.xlsx at \Laboratory_Work\Microsatellites\Analyses\Genepop\2012-13

Title line:"Egernia stokesii Hawker 1213 Non full sib adults"

Ecu2

Est1

Est3

Est4

Tr3.2

Tr5.20

TrL14

TrL28

TrL29

Trl35

TrL12

pop

1046168 , 136150 236240 319323 159181 207223 194194 155173 108112 175175 159171 122122

1055885 , 136150 315323 159159 201215 161173 172182 183189 112112 159181 123149 122122

1128694 , 136136 240260 305331 163173 215223 194228 177177 112116 157175 126173 122122

1180852 , 136158 244256 327327 159177 197215 182214 157173 108116 163174 124155 120148

1226414 , 150150 244244 304323 163163 211211 180188 161173 112112 159159 149193 118118

338724 , 136136 240244 319348 159163 193223 194228 155155 112116 159175 126159 120122

384967 , 148150 244256 315333 171171 205207 182214 159177 108112 155159 124155 120120

388167 , 136168 236244 315331 155185 193205 198228 161161 108112 159175 123125 120122

389835 , 136158 000000 323323 139185 197205 156178 000000 112112 155159 157157 120122

389883 , 136150 256260 304315 139181 201215 214214 157183 108112 155155 123125 122122

406701	,	150154	241244	264323	163181	211221	156188	161173	112112	157159	119149	118120
406853	,	150158	236252	315327	155173	203203	190194	173183	108108	159175	149155	120148
412968	,	136150	000000	291304	173185	205221	182214	155157	112112	175183	125173	120122
413399	,	150154	240256	315333	171181	211229	188214	157173	108112	155157	119159	122152
413675	,	136136	232244	323323	155173	215215	188188	157183	112116	159159	126149	120122
537730	,	136150	315331	149160	193193	137153	156182	199229	108112	175181	123173	118122
680339	,	136150	256256	300311	181185	211225	214214	173173	112116	155155	157209	118120
695125	,	136150	228260	323348	163185	205213	152156	157157	112112	155175	157171	120148
722765	,	136150	241244	323348	177181	193207	156188	157157	112112	157159	126149	118122
726636	,	136150	244244	291331	163173	207215	000000	161161	108112	157181	123171	118122
727148	,	150150	304331	159175	207211	161161	160172	181229	108112	157159	171193	118122
7433935	,	136150	163171	311348	171181	197201	182214	157157	108112	155159	123157	118120
7435198	,	136150	159171	323331	155163	211215	180214	161173	108112	155159	149171	122122

744840 , 136150 150155 319323 159181 207223 194194 155173 108112 175175 159171 122122

pop

1069851 , 136148 171171 307323 155171 189189 214234 173173 108112 157183 157157 120156

1139525 , 136148 252252 307311 155171 189193 234234 173173 112112 157157 157209 120120

1188598 , 150150 240244 291291 167171 211225 182188 177183 112116 157183 159183 122148

1211451 , 150150 136171 310339 155171 201211 214214 157173 108112 157181 125161 120122

1240719 , 136150 228248 327339 147171 193195 156182 157159 112112 155166 161173 120148

338049 , 136150 256256 323323 155171 000000 214214 173173 108108 155179 149171 120154

338077 , 136150 224241 307323 171181 189205 214234 173173 112112 157181 149157 120120

338112 , 150150 224240 291348 167185 211225 188214 173183 108112 157181 157159 148148

338378 , 150150 244252 291345 159185 211217 182234 157173 112116 157183 125177 120122

338857 , 150150 228261 256291 185185 189225 152156 169173 112112 159175 125161 120148

338894 , 136136 224256 323348 171185 189193 214214 157171 108116 181183 157171 138148

387100	,	148150	240252	319323	159171	207211	190190	173175	112112	157175	171183	120120
388783	,	136150	171175	300327	167171	193233	152214	173183	112112	157159	126157	120148
388918	,	150150	244256	256291	167185	199211	182214	157173	108112	157183	155159	118148
389361	,	150154	236240	305333	163181	187225	180188	157183	112112	155183	119175	148152
405994	,	150150	257269	256305	167171	195199	170214	157173	108112	147157	155209	118156
406443	,	148168	257261	311348	163171	205207	214228	183183	108116	147157	123155	120148
408967	,	150150	224236	305310	155163	187201	182214	173183	108112	181183	125175	122152
409374	,	150150	244260	291310	155177	211225	182214	157173	108116	157183	125125	120156
409497	,	136150	240244	300303	155185	000000	182188	183183	112112	183183	126209	120122
413195	,	150150	244260	300315	155171	211211	182190	157183	108112	155183	125209	122156
436591	,	150150	256260	307307	159181	201205	214214	173173	112112	155155	157157	120120
528385	,	150158	244261	307315	155171	203211	190228	157173	108112	155175	125125	120122
726843	,	150150	228269	256291	181185	225225	156170	159183	112112	147155	125171	156156

7431836 , 150150 160175 311348 171185 199211 152188 155169 112116 159183 155157 120120

7434345 , 148148 167167 307348 171177 189189 190234 173173 108112 157162 157209 120156

7434859 , 136150 167167 311348 155171 193211 234234 173173 108112 157157 157183 120156

7435152 , 148150 155167 291307 167171 189225 188234 173177 112116 157157 157159 120122

7435313 , 150150 163175 291305 147159 199207 152182 159173 112112 159166 125209 148156

743824 , 148150 245252 307348 171171 189211 188234 169173 108112 157183 157157 120156

761171 , 150150 257257 310315 155171 211211 214214 157173 108112 155157 125159 122148

763205 , 150150 155175 300315 171185 199225 152188 157157 108116 159183 155209 120146

pop

1050417 , 150170 260312 137183 171223 165181 158172 215231 112116 163181 157157 118120

1053251 , 150150 260312 137167 201223 173177 154158 155159 108116 147181 157157 118120

1166665 , 150150 155155 300327 171171 217225 188188 159159 112112 183183 171173 120120

1225866 , 136150 224256 335335 139171 000000 214214 157157 108112 155175 125171 120120

1226553	,	150150	159183	291305	171185	195225	168182	157157	112112	147155	125155	148156
386733	,	150150	240261	327348	167171	225229	152188	157169	112112	175183	126171	120146
388153	,	148150	224248	256327	155181	225233	182214	157183	112112	181183	126171	120120
389316	,	150150	228269	256291	167185	195211	156170	157183	112112	147155	155159	118148
390562	,	148150	240256	305327	155173	207207	188214	157169	108112	157183	125126	120156
406175	,	150150	240240	323323	171173	207209	188188	157157	108108	159159	149155	120148
412998	,	150150	228240	307327	163171	217225	156188	159183	108112	155183	171173	118120
413209	,	148150	228244	307307	155171	189217	156228	159169	112116	155157	171177	120120
413720	,	148150	140155	291327	181185	233233	156188	173183	112112	159183	125124	118120
413760	,	136150	240252	291311	155171	193211	188234	157177	108116	157183	181183	120148
413849	,	150150	240256	300335	155171	213225	188214	157157	108116	155183	157173	120120
535887	,	136150	224228	291348	163171	205207	156214	157183	108108	155175	125157	120122
536503	,	150150	140155	256291	171181	199207	156188	159173	112112	155183	125171	118120

717048	,	150150	136175	307331	155185	189213	152214	157173	112112	159181	161209	120120
722442	,	150150	256256	311335	155181	201225	214214	157173	112116	155155	126157	120120
726439	,	150150	260327	167175	207223	161173	154158	152189	112116	147159	126157	120120
7432629	,	136150	136167	260291	163181	189221	188214	157159	112116	147175	126157	118120

Appendix 9: Other publications I contributed to during my PhD

Gardner, M.G., **Pearson, S.K.**, Johnston, G.R. and Schwarz, M.P., 2015. Group living in squamate reptiles: a review of evidence for stable aggregations. *Biological Reviews*. DOI: 10.1111/brv.12201