Ancient DNA in the Australian context: Investigating evolutionary and ecological responses to environmental change.

Vicki Ann Thomson

Australian Centre for Ancient DNA
School of Earth and Environmental Sciences
University of Adelaide

A thesis submitted for the degree of Doctor of Philosophy at the University of Adelaide

November 2012

Table of Contents

Table of Contents iii
Abstract ix
Thesis declaration xi
Acknowledgements xii
Chapter 1: Introduction 1
Phylogeography 1
Definitions and background 1
Spatial arrangement of genetic diversity 1
Incorporating other fields into phylogeography 2
Ancient DNA 3
Definitions and background 3
The history of ancient DNA 3
Polymerase Chain Reaction (PCR) 4
Contamination 4
Degradation processes 5
DNA preservation 5
Natural History Collections 7
Ancient DNA in the Australasian region 8
The Australasian region 8
Australia's arid zone 9
Tropical sites in the Pacific 10
Caves as repositories of faunal remains 12
Applications of ancient DNA 12
Taxonomy, Natural History and Evolutionary biologists 13
The Emu 13
The Emu: morphology 13
The conservation of vulnerable species 14
The Ghost Bat: extinction hypotheses 14
Anthropology and Archaeology 15
Colonisation of the Pacific 15
The Commensal Model 15
Haplogroups D and E: the Pacific 16
Polynesian chickens in South America 16
Thesis outline 18
References 19
Chapter 2 35
Methodology 35
Ancient DNA laboratory methods 35

Extraction method ... 36
Population genetic statistics ... 37
The coalescent ... 39
Bayesian Evolutionary Analysis of Sampling Trees (BEAST) ... 40
Approximate Bayesian Computation (ABC) ... 44
Bayesian Serial SimCoal (BayeSSC) ... 45
Demographic and Range expansion tests in Arlequin ... 47
Transforming expansion time estimates to a demographic scale ... 50
Ecological Niche Modelling ... 51
Theory ... 51
Reducing complexity using Ecological Niche Factor Analysis (ENFA) ... 52
Maximum Entropy (MaxEnt) ... 54
Statistical tests of significance ... 55
Conclusion ... 57
References ... 58
Chapter 3: The Emu ... 65
Postglacial expansions from a south-western refugium: reconstructing the recent evolutionary history of Emus (Aves: *Dromaius*) using ancient DNA ... 67
Abstract: ... 67
Introduction ... 68
Materials and Methods ... 76
Samples ... 76
aDNA Extraction, Amplification, and Sequencing. ... 76
Morphometric analyses. ... 76
Population Genetic Analyses. ... 77
Phylogenetic and phylogeographic analyses. ... 77
Population demographic structure. ... 78
Results ... 78
Phylogenetic analyses of island taxa vs. the mainland Emu ... 79
Morphometrics of dwarfed island vs. mainland Emu populations ... 83
Phylogenetic and phylogeographic analyses of the mainland Emu ... 84
Reconstructing Emu demographic history ... 87
Discussion ... 96
The evolutionary origin of dwarf island Emus ... 96
Demography - west to east spatial expansion at 350 kya ... 98
Demography – western expansion at 100 kya ... 100
Conclusion ... 101
Acknowledgements ... 102
Supplementary Materials and Methods ... 103

DNA extraction and sequencing 103
Morphometric analyses 104
Population Genetic Analyses 105
Phylogenetic and phylogeographic analyses 106
Population demographic structure 107
Supplementary Information 109
Supplementary Tables and Figures 109
References 131
Chapter 4: the Ghost Bat 141
A Ghost of a chance? Evolutionary history of the Ghost Bat (*Macroderma gigas*) and its chances of surviving future climate change 145
Abstract 146
Introduction 148
Methods 152
Dataset collection (M. gigas presence localities) 152
DNA extraction and sequencing 153
Phylogenetic and phylogeographic analyses 154
Population genetic and demographic analyses 154
Climatic Envelope reconstruction 155
Presence and pseudo-absence dataset generation 157
Results 157
Population genetic and demographic analyses 157
Intra-specific Diversity 157
Environmental variables involved in extinctions 161
Climatic modelling 165
Last Glacial Maximum (LGM) 165
Last Inter-Glacial (LIG) 166
Future 166
Discussion 168
Genetic substructuring 168
Northern colonies 168
Central West 168
Central East 169
Temporal signatures 170
Climate modelling signatures 174
Assumptions inherent in the ecological niche modelling used 175
Future distributions 175
Conclusion 176
Supplementary Information 178

Chapter 5: The Polynesian chicken 193

Don't count your chickens: Resolving migration patterns of humans and commensals across the Pacific. 197

Abstract 198

Introduction 199

Using chickens to investigate Polynesian contact with South America 202

Results 203

Damaged nature of ancient Pacific chicken samples 204

Phylogeography 206

Discussion 210

Phylogeography of D_{AP} and D_{WP} 210

South American contact 212

Materials and Methods 213

Supplementary Information – Materials and Methods 215

Replication. 217

Sequence analysis 218

Supplementary Information 220

Ancient Pacific samples 220

Independent replication 220

Cloning 220

Shrimp DNase experiment to test ACAD3890 and ACAD9060 samples 221

Phylogenetic analyses including Storey *et al.* (2007) data 222

References 271

Chapter 6 279

Discussion and Conclusion 279

Overall summary of this thesis 279

Chapter 2: Methodology of this thesis 280

Chapter 3: the Emu (Dromaius spp.) 281

Issues encountered 281

Sample size 281

Sample preservation 282

Independence of sampling 283

Island formation as calibration dates 284

Expansion tests in Arlequin 286

Linking dates to climate events 291

Other demographic events 292

Significance of the work 294

Future directions 295

Chapter 4: the Ghost Bat (Macroderma gigas) ... 297
Issues Encountered ... 297
Sample Type ... 297
Calibrations ... 297
BEAST vs. *BEAST: the Ghost Bat ... 298
Assumptions ... 302
Sample size ... 304
Expansion tests from Arlequin ... 304
Significance ... 305
Future work ... 306
In the short-term ... 306
In the long-term ... 306
Chapter 5: the Polynesian Chicken (Gallus gallus) ... 307
Issues Encountered ... 307
Sample size ... 308
Replication ... 308
Independence of sampling ... 308
Significance ... 309
Future work ... 310
In the short-term ... 310
In the long-term ... 311
Future work in general ... 312
Concluding remarks ... 313
References ... 315

Ancient DNA in the Australian context: Investigating evolutionary and ecological responses to environmental change.

Abstract

This thesis investigates how ancient DNA (aDNA) can provide insight into past (and potentially lost) genetic diversity, and how this information can be used to reconstruct evolutionary history. Specifically, it examines how the combination of aDNA data with other analytical techniques can reveal information unavailable from modern genetic studies of mitochondrial DNA, and help to refine models and test hypotheses about the history of individual species through time. Understanding 'patterns' of past genetic diversity and the 'processes' involved in forming these patterns at different temporal and spatial scales is of critical importance in the management of biodiversity, and for predicting future impacts of on-going environmental changes.

This thesis has particular significance for the ancient DNA community in Australia as it demonstrates how DNA may be preserved even in sub-optimal environmental conditions, such as tropical, sand-blown, and arid environments. The genetic information successfully extracted from these sub-optimal samples has allowed a range of phylogeographic questions to be addressed. The overarching question I ask in these studies is: what insights can aDNA provide into evolutionary history and, where these insights contrast with existing models, which additional analytical techniques help clarify or refine these models?

Chapter 2 (Methods) details the optimisation of extraction methods for the isolation of ancient DNA from sub-optimal samples (where often only minimal sample volumes are available for study). I then review the analytical techniques used to investigate the evolutionary history of the three species described in the following chapters.

The first species examined (chapter 3), is the Australian Emu species complex (*Dromaius novaehollandiae, D. baudinianus, D. ater,* and *D. novaehollandiae diemenensis*). I use both morphological methods and coalescent analysis to investigate the evolutionary history of emu, and the basis for separate taxonomic designations of the Tasmanian Emu, and the dwarf Kangaroo Island and King Island Emus. The morphological differences of the latter two species are first used to investigate potential causes of their dwarfism. Based on this assessment, I use genetic data to examine the demographic/dispersal history and phylogeographic structuring of the mainland Emu across its range.

In chapter 4, I investigate the range restricted and vulnerable Australian Ghost Bat (*Macroderma gigas*). Ancient DNA from subfossil cave deposits is used to describe the genetic relationship between extinct southern and extant northern populations and highlight how ancient DNA is essential in examining ancient patterns of gene flow. I then use species distribution modelling to test whether changing climates are responsible for the isolation and extirpations of these southern populations. I approximated the ecological niche currently occupied by Ghost Bats and then hindcast to the last glacial maximum (LGM) and last interglacial (LIG) to examine potential dispersal opportunities in the past, and forecast using models of projected future climates to investigate likely responses to anthropogenic climate change.

In chapter 5, I explore the use of chicken (*Gallus gallus*) genetic diversity as a record of the migration of early Polynesians across the Pacific. The genetic analysis of commensals allows identification of population bottlenecks and examination of potential source populations, and also a separate means to establish the route that their human-counterparts travelled. I incorporate archaeological models of Polynesian history into patterns of chicken diversity across the Pacific, and test hypotheses of potential dispersal routes of this human-commensal dyad using statistical significance testing.

Thesis declaration

I, Vicki Thomson certify that this work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by any other person, except where due reference has been made in the text.

I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library catalogue, and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

____

Vicki Thomson

16th November 2012

Acknowledgements

Firstly, I would like to thank my supervisors, Professor Alan Cooper, Professor Steve Donnellan, and Dr. Ken Aplin. Their assistance and encouragement during my PhD was greatly appreciated. Also I would like to thank the School of Earth and Environmental Sciences at the University of Adelaide for supporting my candidature.

I would like to greatly thank all the members of the Australian Centre for Ancient DNA (ACAD) lab for their help, support, provision of alcohol, and wide range of knowledge, which was freely shared. In particular, I would like to thank my two thesis mates, Clio Der Sarkissian for help with BayeSSC scripts and late night chats, and Julien Soubrier for general discussions about BEAST, manuscript and thesis editing, and late night chats. Going through the final stages of a PhD is helped enormously when you have others to share it with. Also I'd like to thank those who've provided great support in the lab and/or over lunches and after-work drinks, including Maria Lekis (our unofficial social director), Bastien Llamas, Wolfgang Haak, Marta Kasper (Nanna Kasper), Jessica Metcalfe, and Janette Edson. In addition, I'd like to thank the ACAD thesis writing group for reading and providing feedback on various parts of my thesis.

My final thank you is to family and friends, especially those that have put up with missed birthdays, and emails with no response, during the final stages of the PhD.

Chapter 1: Introduction

Phylogeography

Definitions and background

Phylogeography is a field that combines diverse microevolutionary and macroevolutionary disciplines to investigate the 'patterns' and 'processes' creating geographic distributions of genetic lineages (Avise 2000). Avise (2000) recognises seven well-established fields to describe where phylogeography lies as a scientific discipline: Historical geography, Palaeontology, Phylogenetic biology, Molecular genetics, Population genetics, Demography, and Ethology. Traditionally, phylogeography provides a deeper temporal context to the contemporary ecological pressures that affect the spatial distributions of genetic diversity by also investigating the historical 'processes' involved. However, phylogeography utilises purely modern data, where the genetic signal of past events is often overwritten by more recent events. By taking past (and potentially lost) genetic diversity into account through ancient DNA (aDNA) analysis, the broader temporal context of phylogeography can be extended to deeper temporal scales (past the most recent demographic events), and increasing the accuracy and detail of reconstructions of species' evolutionary history. In this thesis, the advantages of using aDNA in comparison to purely modern mitochondrial DNA is examined in the context of investigating the evolutionary history of species. Specifically, aDNA is combined with other analytical techniques to test different hypotheses about three diverse taxa – the Emu (chapter 3), lacks any prior studies of phylogeographic patterns; the Ghost Bat (chapter 4), has had phylogeographic patterns examined using purely modern data; and the Polynesian Chicken (chapter 5), where previous aDNA studies have produced only preliminary phylogeographic hypotheses lacking in robustness. Ancient DNA illuminates unknown aspects of the evolutionary history of each species, with surprising results.

Spatial arrangement of genetic diversity

Many historical processes may explain the phylogeographic patterns of spatially disjunct taxa. Two of the most common processes are dispersal and vicariance. A dispersalist interpretation of phylogeographic patterns describes contemporary occupation of a range by either active or passive dispersal from ancestral refugia. In contrast, a vicariant interpretation describes the spatial arrangement of species diversity as due to separation of formerly continuous ancestral distributions by environmental factors (such as the formation of deserts, sea level rise isolating offshore islands *etc.*). Each of the species studied in this thesis provide contrasting levels of vicariance and dispersal over their recent evolutionary history. In chapter 3, the high dispersal ability of the Emu has likely erased previous phylogeographic structure of the mainland population to some degree, as changing environmental conditions allowed repeated population and/or range expansions into and out of refugial areas. Furthermore, as the majority of Emu

inhabit the southern regions of Australia, certain populations became isolated by rising sea levels after the Last Glacial Maxima (on Tasmania, King Island, and Kangaroo Island), and have evolved as separate taxa as a direct function of this barrier to gene flow.

In contrast, although Ghost Bats (chapter 4) have experienced many of the same environmental changes as the Emu during the Quaternary, Ghost Bats have experienced different vicariant and dispersal processes due to species-specific life history traits. For example, geographically isolated Ghost Bat populations across both the northern (extant) and southern (extinct) regions of Australia may stem from either vicariance over long timescales or relatively recent dispersal. As we know that modern Ghost Bat populations exhibit low dispersal abilities (which has resulted in the high degree of population isolation in northern regions), the expectation is that the Ghost Bat would have relatively high levels of genetic diversity across the species as a whole, due to isolated populations preserving ancestral lineages (Avise 2000). Therefore, using the genetic diversity of these populations and theories of the coalescent, I have reconstructed the timescale over which these Ghost Bat populations likely became isolated, and have drawn conclusions about the mechanisms (vicariance or dispersal) responsible for the current phylogeographic patterns.

In chapter 5, I use a commensal species (the Polynesian chicken) to investigate human movement in the Pacific, which is not possible within the Australian context due to the very early arrival of modern humans. Chapter 5 highlights the major effect human mediated dispersal can have on the spatial arrangement of genetic diversity in a species – especially where non-volant species are transported across sea barriers. Where such passive dispersal has occurred, the human-commensal pair (or dyad) can be treated as one cohesive unit when examining vicariance/dispersal processes, facilitating analysis in areas with limited human archaeological remains. For example, populations of both human and chickens on the Asian mainland and in island south-east Asia (ISEA) have developed genetic diversity in-situ as natural and cultural barriers to gene flow formed over millennia. Changing climate conditions during the Holocene fostered the appropriate pressures for the human-commensal dyad to disperse across the Pacific, leaving a genetic trail marking the route, and timing of this process.

Consequently, ancient DNA allows investigation of the interaction of vicariant and dispersal processes that have formed the current patterns of genetic diversity seen in the three Australasian species examined in this thesis.

Incorporating other fields into phylogeography

If the 19th and 20th century's scientists focused on documenting and defining the world's species (via taxonomic and systematic descriptions), then the 21st century's scientists need to focus on building upon this knowledge to gain an understanding about evolutionary processes, and ultimately aim to manage the diversity of these species for future generations. A paucity of

accurate data exists about the evolutionary history of the three taxa studied in this thesis, so addressing questions about the historical processes involved in forming spatial arrangements of genetic diversity will aid in extending our knowledge base for management decisions involving these species. However, phylogeographic research which uses purely contemporary mitochondrial data to examine the ecological and historical processes responsible for biogeographic patterns can only access a small portion of the evolutionary history. By expanding the repertoire to include ancient DNA, coalescent-based analytical techniques, and ecological niche modelling (ENM), researchers can more fully interpret the evolutionary footprints left by historical processes in both ancient and modern patterns of genetic diversity. The addition of new techniques to the traditional fields of taxonomy, phylogeography, population genetics, and archaeology permit a range of new questions to be addressed which were not previously accessible – and we can now place phylogeography in an even more important juncture in the 21st century.

Ancient DNA

Many developments have occurred in the field of molecular biology over the last three decades that have greatly improved our ability to address questions about species' evolutionary history and ecological responses to changing environments. This is especially true of the significant advances in extracting and amplifying DNA from suboptimal biological resources, such as organisms preserved in the palaeontological record. Studies using this 'ancient' DNA, such as those from fossilised plant or animal remains, have expanded researchers' ability to examine the history of organisms and taxa long since dead (Vilà *et al.* 2001; Hofreiter *et al.* 2002; Krause *et al.* 2008). We can now use ancient DNA (aDNA) to place extinct species within a taxonomic framework (Hofreiter 2003; Bunce *et al.* 2005; Bunce *et al.* 2009); investigate the geographic distribution of a species' past genetic diversity (Aubry *et al.* 2009; Barnett *et al.* 2009); and put a temporal scale on speciation and other divergence events (Hofreiter *et al.* 2002). The addition of aDNA to studies of evolutionary history and human impacts on biogeographic patterns aids our knowledge of the complex biotic and abiotic interactions affecting species during past glacial cycles. Moreover, aDNA studies of past genetic diversity can provide information critical to improving our ability to preserve taxa and their diversity during future environmental changes.

Definitions and background

The history of ancient DNA

The field of ancient DNA (aDNA) can be loosely defined as focusing on the isolation of genetic material from plant or animal remains that were not deliberately preserved for DNA extraction. The origins of this discipline have its roots in an early study of an extinct South African species, the Quagga (Higuchi *et al.* 1984). Three years later the development of the

polymerase chain reaction (PCR) technique (Mullis and Faloona 1987) allowed the field to expand exponentially, increasing both the variety of species studied, *e.g.,* Bison, Woolly Mammoth, Iberian desman, European spotted hyena, Holarctic shrew (Shapiro *et al.* 2004; Nogues-Bravo *et al.* 2008; Morueta-Holme *et al.* 2010; Varela *et al.* 2010; Hope *et al.* 2011; Lorenzen *et al.* 2011), as well as the age of the samples used, *i.e.,* from 50 years old museum samples up to > 300 kya ice cores from Greenland (Willerslev *et al.* 2003; Willerslev *et al.* 2007; Rowe *et al.* 2011).

Polymerase Chain Reaction (PCR)

The Polymerase Chain Reaction (PCR) uses a naturally occurring thermostable DNA polymerase, *Thermus aquaticus* (Taq) to amplify targeted segments of DNA. PCR does this by cycling through a series of temperature-dependent reactions: first, a denaturation step of 94° C which separates the double-stranded helical structure of DNA; second, an annealing step (generally between 50-65° C) to bind an oligonucleotide primer to each single stranded DNA molecule; and lastly, an extension step (between 64-72° C) that allows the Taq polymerase enzyme to incorporate single complementary deoxyribonucleotide bases (dNTP's) onto the end of the primer molecule ad infinitum. Repetition through these temperature cycles allows the amplification process to exponentially increase the copies of the target DNA until hundreds of thousands of copies exist. Although this technique was fundamental to the development of many areas of molecular biology, it was particularly beneficial to the field of aDNA because previous methods such as bacterial cloning were poorly suited to the low copy number and highly damaged and fragmented nature of the DNA in these ancient samples.

Contamination

Although the invention of this pivotal technique has permitted the retrieval of genetic material from ancient remains, PCR also has disadvantages. Any modern DNA that inadvertently enters either the extraction or amplification reaction will be preferentially amplified over the ancient fragmentary DNA. Until this issue could be quantified and dealt with appropriately many publications claimed the successful recovery of Miocene- and even Cretaceous-age DNA: from plant chloroplasts (Golenberg *et al.* 1990), insects in amber (Cano and Poinar 1993; Cano *et al.* 1993), and dinosaur bones (Woodward *et al.* 1994). Serious doubts were expressed about the authenticity of these ancient sequences (Paabo and Wilson 1991; Graur and Pupko 2001), with subsequent work on these (and other) million-year-old samples finding that they could not be replicated (Austin *et al.* 1997), and therefore possibly stemmed from bacterial contamination (Sidow *et al.* 1991; Desalle *et al.* 1993). In response to these early difficulties, the aDNA field has developed strict criteria to reduce contamination issues and promote authentication of results (Handt *et al.* 1994; Cooper and Poinar 2000). These aDNA criteria include: 1) a physically isolated work area – separate from other modern genetic labs, with one-way movement of

labware and personnel; 2) inclusion of control amplifications – in both extraction and PCR reactions; 3) appropriate molecular behaviour – aDNA is usually shorter than modern contaminating DNA (Malmstrom *et al.* 2007); 4) reproducibility – at least two successful PCR and sequencing reactions per fragment is required to confirm authentic sequence; 5) cloning – for examination of damage patterns; 6) independent replication – so that systemic contamination of lab reagents can be identified; 7) biochemical preservation – if no collagen is present for radiocarbon dating, then it is unlikely that DNA will be preserved; 8) quantification – quantitative PCR is often used to assess the level of DNA present; and 9) associated remains – if other animal remains from the same layer also yield aDNA then the results are more likely to stem from endogenous DNA (Cooper and Poinar 2000). When such precautions are taken, it is possible to retrieve believable and accurate evidence of genetic diversity thousands of years old that can help address issues about modern species' ability to survive current and future adversities. However, our ability to successfully extract DNA in these instances is largely dependent on the specific taphonomic processes that individual samples have been subjected to.

Degradation processes

During the life of an organism, enzymes repair DNA from the continual damage suffered due to chemical degradation (*e.g.,* oxidative and hydrolytic processes), radiation, and free radicals *etc*. These innate repair mechanisms cease after death, allowing various degradation processes to proceed unchecked, resulting in 1) DNA strand breakages; 2) the formation of interstrand cross-links (ICLs); and 3) the removal or alteration of nucleotide bases (Paabo 1989). All three of these damage types can prevent the DNA polymerase enzyme used in PCR from continuing amplification past the damage site (resulting in shortened or fragmented DNA). However, only the alteration of bases (*i.e.,* from cytosine to uracil) tends to create 'miscoding' lesions that actually allows the polymerase to continue replication but with the incorporation of erroneous nucleotide bases (Paabo 1989).

DNA preservation

Knowledge of bone diagenesis (the post-mortem chemical and physical processes bones go through in the depositional environment) is central to understanding how DNA is preserved in its biomolecular environment as cells and their accompanying proteins and minerals decompose (Nielsen-Marsh 2002). However, little is understood about the details of these processes. Bone diagenesis appears to follow three main pathways: slow chemical degradation of the proteins, resulting in bones with good histology; fast chemical degradation of the mineral component, resulting in bones with poor histology; and/or fast microbial degradation of the proteins, resulting in complete loss of bone from the depositional environment (see Table 1; Nielsen-Marsh 2002). The extent to which each proceeds in isolation or concurrently with the others is also not well known.

Table 1. Summary of current understanding about bone diagenesis pathways (from Nielsen-Marsh 2002)

Pathway 1	Pathway 2	Pathway 3
Slow loss of protein through chemical degradation	Rapid dissolution of mineral through chemical degradation	Rapid loss of protein through microbial degradation
Assumed slow mineral replacement	Rapid mineral replacement	No mineral replacement
Resultant fossil has good histology	Resultant fossil has poor histology	Bone lost from the paleontological record

Water is one of the main sources of degradation in bone remains, as it dissolves the mineral crystals (*i.e.,* bone apatite) and promotes the growth of microbes (Bollongino *et al.* 2008). Increased temperatures also enhance chemical degradation and microbial attack, while low pH (acidic) soils also dissolve the mineral structure (*i.e.,* calcium phosphate) of bone apatite (Bollongino *et al.* 2008). Therefore, it appears that as long as bone mineral is stable and microbial attack is limited (*i.e.,* pathway 1 dominates) bones have a higher chance of containing well-preserved biomolecules (including DNA).

Bones (and teeth) preserved in permafrost sites provide excellent conditions where DNA is often preserved for 10,000-100,000 years. These cold/dry depositional environments generally result in high success rates of DNA recovery (Bison sp. success rate 352/442 in Shapiro *et al.* 2004; Ovibos moschatus success rate 207/446 in Campos *et al.* 2010). In fact quantitative PCR studies in Malmstrom *et al.* (2007) and Schwarz *et al*. (2009) have shown that permafrost mammoth bones dated between 11,900-48,000 ya BP have similar preservation levels as Neolithic/Middle Ages archaeological bone remains from non-permafrost conditions, based on lambda values of damage frequency from linear regression models developed by Deagle *et al*. (2006). However, in recent years the field of aDNA has diversified into a variety of alternate sample types, including coprolites (Poinar *et al.* 1998; Hofreiter *et al.* 2000), hair (Gilbert *et al.* 2004; Gilbert *et al.* 2007a), nails (Gilbert *et al.* 2007b), feathers (Rawlence *et al.* 2009; Hartnup *et al.* 2011), regurgitated owl pellets (Poulakakis *et al.* 2005), eggshell (Oskam *et al.* 2010; Oskam *et al.* 2011), sediments (Haile *et al.* 2009; Madeja *et al.* 2010), seeds (Gyulai *et al.* 2006; Nasab *et al.* 2010; Varela *et al.* 2010; Fan *et al.* 2011), plant exocarp (Erickson *et al.* 2005), and pathogens, such as *Yersinia pestis* (which causes the plague; Bos *et al.* 2011) and *Bordetella pertussis* (which causes whooping cough; Theves *et al.* 2011). Each of these new substrates differs in its own particular inherent abilities to repel water and microbes, two of the major damaging agents of DNA. Each substrate is therefore also likely to exhibit differential uptake of bacterial contamination, degree of DNA damage, template concentration, and level of PCR inhibitors.

Where these damaging agents are minor, these sub-optimal sample types can potentially be used for examining species' history over evolutionary timeframes.

During the last few years in particular, the likelihood that less-than-ideal samples can provide useful genetic material has increased through the optimisation of extraction techniques and the development of next generation sequencing (NGS) technology (Oskam *et al.* 2010; Oskam *et al.* 2011; Paplinska *et al.* 2011; Rowe *et al.* 2011). These sub-optimal samples include those from non-permafrost sites, such as the semi-arid, temperate, and/or tropical regions that have in the past rarely produced replicable DNA of any great age (Bollongino *et al.* 2008; Schwarz *et al.* 2009). A focus on finding the location of preserved DNA within ancient bone/eggshell samples, and therefore the best way to extract it (Oskam *et al.* 2010; Campos *et al.* 2011), has improved knowledge of why certain extraction methods give higher DNA yields. In particular, Oskam *et al.* (2010) has highlighted the exceptional DNA preservation ability of eggshell due to this substrate's stable intra-crystalline organic matrix (Oskam *et al.* 2010; Oskam *et al.* 2011). As eggshell, in particular, is often preserved in Australian archaeological and palaeozoological sites, this research has opened the door to molecular data from Pleistocene age fossils from Australian sites otherwise out of reach. At the other end of this temporal spectrum, the increasing use of museum samples (collected in the last few hundred years) in NGS studies, has highlighted the utility of this otherwise untapped genetic resource preserved in museums around the world (Rohland *et al.* 2004; Wandeler *et al.* 2007; Paplinska *et al.* 2011; Rowe *et al.* 2011).

Natural History Collections

The natural history collections held in museums around the world are repositories of a unique and irreplaceable record of past biodiversity (Wandeler *et al.* 2007; Rowe *et al.* 2011). The traditional specimen types (skins, skulls, and alcohol-preserved whole bodies) provide a record of phenotypic diversity and historical geographic distributions, with type specimens offering the key to defining and re-defining taxa long after their date of initial collection. Although the recent museum practice of preserving modern tissue samples increases our ability to easily extract genetic information, the traditional specimen types offer a longer time series, sometimes spanning centuries, but without optimal storage conditions. The degraded nature of the DNA in these traditional specimen types means a more expensive and laborious extraction process, and shorter amplicons (often of < 200 bp in length; Wandeler *et al.* 2007). In this way, many aspects of natural history collections resemble aDNA samples

The main difference between 'true' aDNA specimens and natural history collections is the time since death of the organisms. However, as many museums' sample storage areas are not optimal for DNA preservation, even the significantly younger age of these samples cannot negate the oxidation, heat and chemical degradation that occurs over time and result in fragmented DNA

and damaged nucleotide bases. These sub-optimal conditions generally include storage at ambient temperatures with temperature and humidity swings, rather than in fridges or freezers. An example of this is seen in the relatively high level of DNA damage in a permafrost mammoth bone kept at low ambient temperatures for 30 years, compared to those stored below 0° C for the 10 years since they were excavated (Schwarz *et al.* 2009). Australian museums do not use air-conditioning of storage space as standard practice, and consequently many specimens have been exposed to repeated seasonal temperature fluctuations and have advanced stages of DNA degradation.

Ancient DNA in the Australasian region

As the generally hot and humid/arid climate of the Australasian region is generally unsuitable for DNA preservation, there have been few aDNA studies conducted here. The temperate regions of New Zealand are an exception, with studies of moa dominating (Cooper *et al.* 1992; Cooper *et al.* 2001; Huynen *et al.* 2003; Bunce *et al.* 2005; Huynen *et al.* 2008; Allentoft *et al.* 2009; Bunce *et al.* 2009; Allentoft *et al.* 2010), but also individual studies of the Haast eagle (Bunce *et al.* 2005) and the Kiwi (Shepherd and Lambert 2008). However, the predominantly tropical, sub-tropical, and/or arid climates of the rest of the Australasian region are generally bad for DNA preservation.

The Australasian region

The poor preservation conditions in Australian environments partially explain why knowledge about the environmental effects of climatic cycles on species in the southern hemisphere lags behind that of the northern hemisphere (Beheregaray 2008; Byrne *et al.* 2008). The role that late Quaternary climate and environmental change have played in the diversity and distribution of present day plants and animals is a worldwide phenomenon. In the northern hemisphere, the impact of glacial/interglacial cycles on species has been investigated extensively, *e.g.,* North America, the Arctic, and Beringia (Taberlet *et al.* 1998; Hewitt 2000; Barnes *et al.* 2002; Hadly *et al.* 2004; Shapiro *et al.* 2004; Bhagwat and Willis 2008; Debruyne *et al.* 2008; Aubry *et al.* 2009; Campos *et al.* 2010; Shafer *et al.* 2010), with the environmental effects associated with fluctuations of glaciers well understood (Muller and MacDonald 1997; Clark *et al.* 2009).

In contrast to the Northern Hemisphere, the manifestation of past glacial cycles in Australia has been in fluctuations of aridity (Barrows *et al.* 2002). This cyclical desertification is thought to have led to repeated movement and reassembling of ecosystems during the Quaternary, which seems to have created distinctive phylogeographic patterns in the modern biota of Australia (Byrne 2008). One such pattern is seen in some widespread species that exhibit low genetic diversity and little geographic structuring, including those with suggestions of recent range expansions (Byrne 2008). These low diversity taxa include many bird species, *e.g.,* ducks,

parrots, cuckoos, and a range of passerines (Joseph *et al.* 2002; Joseph and Wilke 2006; Joseph *et al.* 2006; Toon *et al.* 2007; Kearns *et al.* 2009; Guay *et al.* 2010; Kearns *et al.* 2010), as well as some reptile and amphibian (Kuch *et al.* 2005; Burns *et al.* 2007), and plant species (Sandalwood; Byrne *et al.* 2000). This lack of genetic diversity has also been highlighted in many of Australia's unique and iconic marsupials (Koala, Eastern Grey Kangaroo, Yellow-bellied Glider, Brush-tailed Phascogale; Houlden *et al.* 1999; Spencer *et al.* 2001; Zenger *et al.* 2003; Brown *et al.* 2006). For these highly vagile and/or relatively widespread species, reconstructing demographic histories is a complex issue because few barriers have restricted their dispersal. Limited genetic diversity is one of the few signs to indicate the impact of past climatic and environmental change. The addition of ancient DNA (aDNA) to these phylogeographic studies (termed phylochronology; Hadly *et al.* 2004), provides a potential to focus directly on the genetic evidence of demographic changes as they occurred. Unfortunately, the retrieval of aDNA from sub-fossil bone is hampered by the extreme Australian climatic conditions that created the very demographic histories of interest.

Australia's arid zone

In general, our ability to reconstruct the evolutionary history of a species is based on an understanding of climate and geology of the region in which they live: we expect some level of agreement between physical and biological histories, such that physical barriers to gene flow coincide with populations isolated on either side of the barrier. In many instances this concordance has been detected in the contemporary genetic diversity of many species. As knowledge about the origins and changes in past continental biomes is gained through palaeoclimatic studies, the patterns in more Australian species are being investigated. So far, the majority of Australian studies have focused on species of the wet tropics rainforests (Joseph *et al.* 1995; Schneider *et al.* 1998; Bell *et al.* 2010) and temperate eucalypt forests of the south-east (Chapple *et al.* 2005; Dubey and Shine 2010), with species of the arid zone largely neglected (Byrne *et al.* 2008).

The Australian arid zone is defined as the region with a moisture index (rainfall/evaporation) of <0.4 (Byrne *et al.* 2008). The highly variable rainfall and soil infertility of the Australian arid zone means that it differs substantially from the well-studied North American and European correlates (Etten 2009). The lack of reliable rainfall is caused by interactions of the varying ocean climates surrounding the Australian continent: including the El-Nino Southern Oscillations, inter-decadal Pacific Oscillations, and the Indian Ocean Dipole (Quigley *et al.* 2010). The lack of soil nutrients, on the other hand, stems from the ancient nature of the Australian landscape: the highly weathered geology is exacerbated by the lack of renewal by volcanism, glacial activity, or wind-borne nutrients (Morton *et al.* 2011). The combined lack of rainfall and soil nutrients creates phylogeographic patterns in the flora that match rainfall

predictability and nutrient quality, with concomitant patterns of digestibility to herbivores (Morton *et al.* 2011). Shared floral and faunal patterns can be seen in the geographic distribution of plants that express the toxin fluoroacetate where 33 of the 34 Australian plant species that naturally produce this toxin are restricted to the southwest corner of Western Australia. In those animal species that are susceptible to this toxin, fluoroacetate blocks the tricarboxylic acid cycle, which allows citrate to accumulate in the tissue and blood, eventually causing death (Twigg *et al.* 1988). However, many native fauna have developed a tolerance to fluoroacetate and this region seems to have provided a refugium for plants and fluoroacetate-tolerant fauna in the past (King *et al.* 1978; Oliver *et al.* 1979; McIlroy 1984), which includes the Emu (Twigg *et al.* 1988). Where species are fluoroacetate-tolerant, but now inhabit regions outside the southwest corner, a deep history of co-evolution in the southwest can be inferred.

More generally, foodwebs in the Australian landscape ultimately mirror levels of water and nutrient access, with intermittent heavy rainfall allowing rapid population size increase in rodent and invertebrate species. After these ephemeral but intense rainfall events, endothermic vertebrate species that feed on the newly sprouting flora (or on the expanding small mammal and insect species) also flourish, either by large-scale population increases or by migration into these prosperous areas (Dickman 1999; Burbidge and Fuller 2007). Two Australian species, the Emu and the Ghost Bat are both likely to have responded to high rainfall events such as these. As a nomadic omnivore, the Emu is constantly on the move following the intermittently abundant food resources that respond to this rainfall; seeds, flowers, young sprouting annuals and perennials, as well as insects (Davies 2002). In fact, Emus seem to be conscious of the cues signalling heavy rain, as in western Australia they have been detected not only migrating according to the direction of predominant seasonal rainfall (north in summer, south in winter), but when a reversal in the normal seasonal rainfall occurs they reverse their migration patterns (Davies 2002). In contrast, and despite the Ghost Bat's capacity for flight, it has neither a large home range, nor the ability to occupying a wide range of landscapes/habitats that the Emu has. However, given the large distribution of the Ghost Bat it appears to have also experienced long-distance dispersal at some time in the past. The genetic signatures of these demographic and spatial expansions are testable by coalescent-based analysis techniques such as tests of expansion in Arlequin, Baysian Skyline Plots as well as Bayesian Serial SimCoal (see chapter 2: Methodology).

Tropical sites in the Pacific

Although age appears to have had the greatest impact on DNA preservation in permafrost or temperate environments (*i.e.,* more samples go through pathway 1 in Table 1), in the tropics the specific depositional environment seems to have played more of a role in DNA preservation (*i.e.,* pathways 2 and 3 appear to dominate). The combination of hot and wet conditions in tropical areas, found on many Pacific Islands, has restricted the age range of samples that have

yielded DNA. However, as many of the questions investigated with archaeological material in the Pacific Islands were within the last few thousand years (related to the cultures that colonised these islands during the Holocene), there is a chance DNA has been preserved sufficiently to allow amplification. The fact that many bones have been excavated from archaeological sites in the Pacific Islands (Steadman and Intoh 1994; Allen and Craig 2006; Barnes *et al.* 2006; Storey *et al.* 2008a) has indicated at least some depositional environments were conducive to biomolecular survival – suggesting that a subset of these sites may have had temperatures and pH levels suitable for DNA preservation (Nielsen-Marsh 2002).

Many of the archaeological sites on Pacific Islands were exposed sand dunes or middens, which have likely had water draining through them on a regular basis, which has affected DNA preservation in those remains buried deep within the dune/midden (*i.e.,* protected to a greater degree from extremes of temperature), as well as those that were closer to the surface (*i.e.,* exposed to higher ambient temperatures). Furthermore, the soil pH of these sand dunes/middens were likely alkaline from the coralline sand and calcareous soils dominating many Pacific Islands (Morrison 1990; Horrocks and Bedford 2005). Although alkaline soils have been shown to preserve bones better than acidic soils, alkaline soils have not been completely beneficial for bone minerals either, as under alkaline conditions and where carbon dioxide has been present, bicarbonate has be released that can further destroy bone apatite, exposing endogenous DNA to the elements (Bollongino *et al.* 2008). Taking all these dune/midden characteristics into account – water movement, temperature buffering ability, and pH of soil matrix – these open environments have likely provided highly variable DNA preservation conditions (Robins *et al.* 2001).

As a result of the variable preservation conditions, caution needs to be taken in authenticating DNA sequences from any ancient Pacific samples. This is particular relevant where commensal species are the focus of ancient DNA studies, as these taxa are also the likely source of contaminating DNA in both laboratories and field sites. Although few studies have investigated this issue, field contamination has likely occurred in the past, especially where modern populations of the same species have still occupied the archaeological site, as has been the case with cattle in Morocco (Bollongino *et al.* 2008). Contamination of laboratory reagents and consumables is a more accepted form of contamination, especially by commensal animals as DNA from these species are the most likely in factories/laboratories where items of plasticware and reagents are prepared and packaged (Gilbert *et al.* 2005; Leonard *et al.* 2007). Yet, many studies on commensal animals ignore relatively simple contamination reduction measures, such as adding a heat-labile endonuclease (which cleaves phosphodiester bonds in double stranded DNA) to PCR reactions to eliminate potential reagent contamination (Champlot *et al.* 2010). Shrimp DNase (USB 2006) is one example of such an endonuclease, and is added to PCR master mixes to degrade any contaminating double stranded DNA in the PCR reagents and plasticware,

prior to addition of the DNA extract to be amplified. Unfortunately this treatment will not remove any contaminating DNA that may have been present in the extraction reagents/plasticware, nor in the sample itself from field contamination. However, the addition of this simple procedure to ancient DNA studies goes some way to increasing confidence in the authenticity of results involving commensal species (Storey *et al.* 2007; Storey *et al.* 2010).

Caves as repositories of faunal remains

In both arid and tropical regions, the majority of older material yielding DNA comes from the buffered environments of caves. Caves have long been known as a rich repository of Quaternary faunal remains, including hominins such as the Neanderthal and associated Mousterian faunal remains from Mezmaiskaya cave, Russia (Baryshnikov *et al.* 1996; Ovchinnikov *et al.* 2000); the Denisovan finger bone from Denisova Cave, Siberia (Reich *et al.* 2010); and *Homo floresiensis* from Liang Bua Cave, Flores; island fauna, such as giant rodents and dwarfed stegodons also from Liang Bua, on Flores, (Morwood *et al.* 2004; van den Bergh *et al.* 2009); as well as remains of extinct megafaunal species from the Middle Pleistocene from caves in Naracoorte, South Australia and the Nullarbor Plain, Western Australia (Reed and Bourne 2000; Prideaux *et al.* 2007). Where these cave sites are in temperate climates, ancient DNA has been successfully retrieved from important samples in the age range of 30-50,000 ya, *e.g.,* Neanderthal and Denisovan remains (Ovchinnikov *et al.* 2000; Smith *et al.* 2001; Reich *et al.* 2010). As yet no authentic DNA has been generated from *Homo floresiensis* from a tropical site nor from anything older than 19,000 ya Emu eggshell samples from Tunnel Cave, Western Australia (Oskam *et al.* 2010). It appears that not even cave environments can buffer the high temperatures these regions experience over millennial time-scales. As with the tropical environments of the Pacific Islands, researchers may need to lower their expectations of the age at which DNA can be preserved from these Island Southeast Asian and Australian sites.

Applications of ancient DNA

Scientists are increasingly using aDNA to provide critical data for investigations of evolutionary patterns and processes in single species and biotas. Questions about past responses to Pleistocene climate cycles can be elucidated when evolutionary biologists are investigating the ability of extant species to adapt to future climate change (Hadly *et al.* 2004; Shapiro *et al.* 2004; Chan *et al.* 2005; Smith and Betancourt 2006; Jones 2008; Nogues-Bravo *et al.* 2008; Hofreiter and Stewart 2009; de Bruyn *et al.* 2011). Conservation biologists that are interested in the effects of habitat fragmentation on native taxa, can use aDNA to address how best to conserve remaining diversity (Miller and Waits 2003; Barnett *et al.* 2006; Martinez-Cruz and Godoy 2007; Paplinska *et al.* 2011). And finally, when archaeological contexts are under study, questions regarding domestication events can be investigated using aDNA (Allen *et al.* 2001; Jaenicke-Despres *et al.* 2003; Larson *et al.* 2005a; Larson *et al.* 2005b; Dobney and Larson 2006; Larson *et al.* 2007a;

Larson *et al.* 2007b; Storey *et al.* 2007; Dobney *et al.* 2008; Storey *et al.* 2008b; Larson *et al.* 2010). Each of these fields of science (and others) can benefit from the inclusion of aDNA from fossil samples in their research.

Taxonomy, Natural History and Evolutionary biologists

The Emu

Although native to Australia, the ubiquity and widespread nature of the Emu has meant reduced scientific interest in its population genetic and demographic history (although see Hammond *et al.* 2002; Heupink *et al.* 2011). This is important as many of our large and iconic species exhibit low levels of diversity, which poses a danger to their health and persistence, *e.g.,* the Tasmanian Devil (Miller *et al.* 2011). Devil Facial Tumour Disease (DFTD) in the Tasmanian Devil (*Sarcophilus harrisii*) is one example of how low genetic diversity within a species can reduce the ability of the immune system of an organism to fight disease, resulting in rapid declines and the possibility of extinction (Hawkins *et al.* 2006). Therefore, at a minimum, the genetic diversity level of all Australian species needs to be better characterised, including taxa that are common and/or widespread.

There have been few studies of the modern genetic diversity of Emus. Microsatellite diversity levels have been investigated for farmed Emus (from Australian and Thai farms) vs. wild Emus (Hammond *et al.* 2002). As the Emu has been extensively farmed for food, meat and oil, farmers have a vested interest in maintaining good genetic diversity levels for breeding purposes (Hammond *et al.* 2002), however this may also have affected the level of genetic similarity that farmed Emus have with wild populations (*e.g.,* farmed Emus may have been subject to bottlenecks, genetic drift, and outbreeding). Another study has compared extinct King Island Emu mitochondrial DNA to farmed mainland Emus (from Australian and New Zealand farms; Heupink *et al.* 2011). The study by Heupink *et al.* (2011) found the King Island Emu to lack genetic differentiation from the mainland Emu, however this study has been restricted to just four specimens of only one of the island species studied here (chapter 3), and with only farmed mainland Emus of unknown provenance for comparison. By compiling an extensive mainland Emu dataset for comparing against all island Emu species/subspecies, the study in chapter 3 provides a more complete picture of the evolutionary history of the *Dromaius* genus in Australia.

The Emu: morphology

The study by Heupink *et al.* (2011) also investigated the morphological differences between the King Island Emu and mainland Emus by examining skull shape, which they compared by eye. Furthermore, the morphological comparison lacked rigorous statistical analysis, such as principal component analysis. Although Heupink *et al.* (2011) proposed that skull shape has been used to distinguish these taxa (with no citation to support this point), they noted that the morphological differences in cranial contours between these two taxa appeared to resemble age-

specific differences within the mainland Emu, as the same dome shape seen in the King Island Emu was also present in immature mainland Emus. Heupink *et al.* (2011) also ruled out two of the other supposed distinguishing traits between the King Island Emu and the mainland Emu: the distal foramen of the tarso-metatarsus, which was also variable within the mainland Emu (again with no citation in support of this point); and plumage colouration, which Heupink *et al.* (2011) cast doubt on due to a lack of variation in 57bp of the melanocortin 1 receptor (MC1R) coding region (see chapter 3 for discussion on using only one melanin coding gene to definitively exclude a genetic basis for plumage colouration differences). However, the discussion on morphological differences between the dwarfed King Island Emu and the standard-sized mainland Emu by Heupink *et al.* (2011), did not focus on the main morphological difference – lower limb size. The study in chapter 3 has addressed this gap by analysing morphological analyses of limb bones from both island dwarfed Emus and the full-sized mainland Emu.

The conservation of vulnerable species

The Ghost Bat: extinction hypotheses

Many Australian taxa have already become extinct due to a combination of climatic change and human-induced ecosystem disturbance - a third of all worldwide mammal extinctions since 1600 AD have occurred in Australia (Short and Smith 1994). In order to preserve our remaining biodiversity now and into the future, knowledge about causes of population extinctions in vulnerable species is needed. One such species under threat is Australia's largest carnivorous bat, the Ghost Bat (*Macroderma gigas*). The Ghost Bat is a monotypic Microchiropteran bat restricted to Australia that has been both abundant and widespread in the past: it is dominant in 25Mya deposits at Riversleigh and is present in more recent subfossil cave remains that span the continent (Churchill 2008). However, the Ghost Bat's distribution has contracted to the northern coastal fringe since European settlement (Worthington Wilmer *et al.* 1994). All historic sightings of live bats were of single individuals or small colonies (Douglas 1956; 1957; Butler 1961; Douglas 1967), suggesting these populations declined rapidly or had already declined substantially as Europeans colonised the continent.

Hypotheses posed to explain the contracting distribution of the Ghost Bat include increasing aridification during the Pleistocene/Holocene; changes in fire regimes when Aboriginal people left their country causing a decline in small mammal species that are a food source during the dry season (Burbidge *et al.* 1988); or it may be a combination of climatic and anthropogenic factors. Until the timing of the initial gene flow between the extant and extinct populations and their subsequent range reductions, as well as the ecological niche of the Ghost Bat are known, these hypotheses cannot be tested. Therefore, in chapter 4, I examine the ancient patterns of gene flow in the Ghost Bat by including samples of extinct populations in a

phylogeographic study of genetic diversity and estimate the ecological niche from occurence data.

Anthropology and Archaeology

Colonisation of the Pacific

The movement of Polynesians across the Pacific was one of the last great human migrations; the nature, timing and route of which have been debated for centuries (Meggers 1975; Kirch 2000; Hurles *et al.* 2003). However, controversy over the existence, or extent of prehistoric contact between the Pacific and South America is ongoing. The presence of South American domesticates, the bottle gourd and sweet potato, in early sites across the Pacific (Hather and Kirch 1991; Clarke *et al.* 2006), appear to support this prehistoric contact. However, fundamental issues with these studies remain. Modern molecular work on the bottle gourd and sweet potato have been complicated by admixture with later/multiple introductions of the same species; post-European hybridisation has occurred between Asian and American bottle gourds in the Polynesian populations studied (Zhang *et al.* 2004); and 16th century transportation of the sweet potato from Mexico to the Philippines has influenced both the Filipino and potentially Oceanian gene pools analysed (Montenegro *et al.* 2008). Also, simulation studies have not excluded natural or 'rafting' dispersals in these species; bottle gourds float, and experimental studies have found them to still be viable after >7 months in seawater (Whitaker and Carter 1954). Sweet potato seeds are known to be spread by birds (O'Brien 1972), and being buoyant, their seed capsules may also have floated, or attached to other floating debris, from South America (or more likely Central America, as the Polynesian sweet potato shares more genetic affinity, and similarly high diversity levels, with Mexican rather than Peruvian/Ecuadorian sweet potatoes) (Montenegro *et al.* 2008). Given the multitude of equivocal evidence, what is needed is a definitive genetic link between the Pacific and South America. Zooarchaeological remains of commensal species provide the greatest opportunity to address this issue. This strategy is otherwise known as the 'commensal model'.

The Commensal Model

The basis of the commensal model is that the phylogeographic patterns of species that live in close association with humans should track the movement of cultures across the landscape. In the strict sense, commensalism can be defined as an association between two species, one of which benefits while the other derives no benefit or harm. In the Pacific, commensal species are generally seen as those animals that have been part of the transported landscape of the Polynesians, such as the pig (*Sus scrofa*), dog (*Canis familiarus*), rat (*Rattus exulans*) and chicken (*Gallus gallus*). The Polynesians are thought to have transported these species in order to provide a portable source of protein. Of the four Polynesian commensals, only the chicken has relatively high levels of mitochondrial genetic variability. This high level of genetic diversity stems from

the fact that chickens were domesticated multiple times in different geographic centres of Asia (Liu *et al.* 2006; Kanginakudru *et al.* 2008), promoting the isolation of different haplogroups in various geographical centres and the movement of specific lineages with human cultures around the world. Each of the nine chicken haplogroups has a slightly different biogeographic history, which has both advantages and disadvantages in reconstructing the movement of the human-commensal dyad (Liu *et al.* 2006).

Haplogroups D and E: the Pacific

Each of the nine highly divergent chicken haplogroups (designated A-I) have distributions influenced by the original natural diversity of the chicken, overlaid by separate domestication events within Asia, and the subsequent transportation by humans (Liu *et al.* 2006). Haplogroups D and E have both been found in chicken bones from archaeological deposits in the Pacific, however conjecture about the authenticity of the latter haplogroup from ancient chicken remains exists (Gongora *et al.* 2008). As the domesticated chicken lineages that were transported to Europe prehistorically, and around the world during the 'Age of Exploration' appear to belong to haplogroup E, the high proportion of haplogroup E in current chicken populations worldwide is not unexpected (Liu *et al.* 2006). The pervasiveness of this haplogroup in post- and pre-European chicken populations worldwide makes it phylogeographically uninformative. On the other hand, haplogroup D has a more 'localised' distribution in the Asia-Pacific region. Haplogroup D has an hypothesised homeland in India or Indonesia, and an acknowledged anthropogenic transportation route following the cultural activity of cock-fighting (including gamecocks from China and Japan) (Liu *et al.* 2006). Furthermore, no known haplogroup D chickens have been detected in European chicken stocks. Therefore, the 'localised' nature of haplogroup D gives it an advantage over the 'worldwide' haplogroup E for accurately reconstructing the movement of Polynesians into, and across, the Pacific.

Polynesian chickens in South America

Studies investigating ancient chickens as a proxy for Polynesian migration patterns have proposed evidence in support of trans-Pacific contact (Storey *et al.* 2007), which consists of a single pre-Columbian archaeological chicken bone from El-Arenal in Chile sharing an E haplotype (which is common worldwide) with samples from two Polynesian islands: American Samoa (n=1, Fatu-ma-Futi site) and Tonga (n=1, Mele Havea site); plus being one base pair (bp) different (but still common worldwide) from samples from another site in Tonga (n=1, Ha'ateiho site), and from sites on Hawaii (n=1, Kualoa site), Niue (n=1, Paluki site), and Rapa Nui (n=1, Anakena site). However, not only do the limited number of these samples pose problems but, like in other high profile studies in Pacific archaeology (Wilmshurst *et al.* 2008), the radiocarbon dating of the ancient South American chicken samples (n=3) have been brought into question (Gongora *et al.* 2008). This doubt over the chronological authenticity has important

ramifications, as without irrefutable pre-European dates, the ubiquity of this particular haplogroup means these chickens could have been introduced by: 1) Polynesians (ca. 1300 A.D.; Wilmshurst *et al.* 2011); 2) the first authentic Portuguese colonisers in 1500 A.D. (Carter 1971); or 3) by earlier explorers responsible for the first map showing South America, drawn in 1448 (Batalha-Reis 1897). It is also salient to note that any contamination from modern domestic animal DNA in lab materials, which is a common problem in aDNA work (Leonard *et al.* 2007) (yet often underappreciated outside it), could generate these worldwide chicken haplotypes.

This background therefore provides a context for the study detailed in chapter 5, where I investigate Polynesian migration/trade routes, by using new aDNA protocols and contamination reduction measures to examine a range of modern and archaeological samples from across the Pacific. In addition, a selection of Storey *et al*'s (2007) Pacific archaeological samples were re-examined, including the important Rapa Nui specimen linking east Polynesian and South American chickens, to explore the authenticity of the common worldwide haplotype appearing in ancient Polynesian chickens.

Thesis outline

The use of ancient DNA analyses, in concert with coalescent-based analyses and ecological niche modelling, has greatly increased our capacity to answer questions on species evolutionary and ecological history. The research presented here highlights the ability of ancient samples from non-permafrost regions (and therefore with non-optimal preservation conditions) to increase our knowledge about Australasian species' evolutionary and ecological responses to climate and sea-level change. In particular, I investigate Pleistocene-age patterns and potential barriers to gene flow in a widespread non-flying bird species (the Emu; chapter 3), and a range-restricted flying mammal species (the Ghost Bat, chapter 4) in the Australian context. In addition, I also answer questions regarding the spread of a human-commensal species pair in response to the lowered sea levels and climate amelioration in the Pacific during the late Holocene (the Polynesian Chicken; chapter 5).

References

Allen MS and Craig JA (2006). Dynamics of Polynesian subsistence: Insights from archaeofauna and stable isotope studies, Aitutaki, Southern Cook Islands. *Pacific Science* 63: 477-506.

Allen MS, Matisoo-Smith E and Horsburgh A (2001). Pacific 'Babes': Issues in the origins and dispersal of Pacific pigs and the potential of mitochondrial DNA analysis. *International Journal of Osteoarchaeology* 11(1-2): 4-13.

Allentoft M, Schuster SC, Holdaway R, Hale M, McLay E, Oskam C, Gilbert MT, Spencer P, Willerslev E and Bunce M (2009). Identification of microsatellites from an extinct moa species using high-throughput (454) sequence data. *BioTechniques* 46: 195-200.

Allentoft ME, Bunce M, Scofield RP, Hale ML and Holdaway RN (2010). Highly skewed sex ratios and biased fossil deposition of moa: ancient DNA provides new insight on New Zealand's extinct megafauna. *Quaternary Science Reviews* 29: 753-762.

Aubry KB, Statham MJ, Sacks BN, Perrine JD and Wisely SM (2009). Phylogeography of the North American red fox: vicariance in Pleistocene forest refugia. *Molecular Ecology* 18(12): 2668-2686.

Austin JJ, Ross AJ, Smith AB, Fortey RA and Thomas RH (1997). Problems of reproducibility - Does geologically ancient DNA survive in amber-preserved insects? *Proceedings of the Royal Society of London Series B-Biological Sciences* 264(1381): 467-474.

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

Barnes I, Matheus P, Shapiro B, Jensen D and Cooper A (2002). Dynamics of Pleistocene population extinctions in Beringian Brown Bears. *Science* 295(5563): 2267-2270.

Barnes SS, Matisoo-Smith E and Hunt TL (2006). Ancient DNA of the Pacific rat (*Rattus exulans*) from Rapa Nui (Easter Island). *Journal of Archaeological Science* 33(11): 1536-1540.

Barnett R, Shapiro B, Barnes I, Ho SYW, Burger J, Yamaguchi N, Higham TFG, Wheeler HT, Rosendahl W, Sher AV, Sotnikova M, Kuznetsova T, Baryshnikov GF, Martin LD, Harington CR, Burns JA and Cooper A (2009). Phylogeography of lions (*Panthera leo ssp.*) reveals three distinct taxa and a late Pleistocene reduction in genetic diversity. *Molecular Ecology* 18(8): 1668-1677.

Barnett R, Yamaguchi N, Barnes I and Cooper A (2006). Lost populations and preserving genetic diversity in the lion *Panthera leo*: Implications for its ex situ conservation. *Conservation Genetics* 7(4): 507-514.

Barrows TT, Stone JO, Fifield LK and Cresswell RG (2002). The timing of the last glacial maximum in Australia. *Quaternary Science Reviews* 21(1-3): 159-173.

Baryshnikov G, Hoffecker JF and Burgess RL (1996). Palaeontology and zooarchaeology of Mezmaiskaya Cave (northwestern Caucasus, Russia). *Journal of Archaeological Science* 23(3): 313-335.

Batalha-Reis J (1897). The supposed discovery of South America before 1448, and the critical methods of the historians of geographical discovery. *The Geographical Journal* 9(2): 185-210.

Beheregaray LB (2008). Twenty years of phylogeography: the state of the field and the challenges for the Southern Hemisphere. *Molecular Ecology* 17(17): 3754-3774.

Bell RC, Parra JL, Tonione M, Hoskin CJ, Mackenzie JB, Williams SE and Moritz C (2010). Patterns of persistence and isolation indicate resilience to climate change in montane rainforest lizards. *Molecular Ecology* 19(12): 2531-2544.

Bhagwat SA and Willis KJ (2008). Species persistence in northerly glacial refugia of Europe: a matter of chance or biogeographical traits? *Journal of Biogeography* 35(3): 464-482.

Bollongino R, Tresset A and Vigne J-D (2008). Environment and excavation: Pre-lab impacts on ancient DNA analyses. *Comptes Rendus Palevol* 7(2-3): 91-98.

Bos KI, Schuenemann VJ, Golding GB, Burbano HA, Waglechner N, Coombes BK, McPhee JB, DeWitte SN, Meyer M, Schmedes S, Wood J, Earn DJD, Herring DA, Bauer P, Poinar HN and Krause J (2011). A draft genome of *Yersinia pestis* from victims of the Black Death. *Nature* 478(7370): 506-510.

Brown M, Cooksley H, Carthew SM and Cooper SJB (2006). Conservation units and phylogeographic structure of an arboreal marsupial, the yellow-bellied glider (*Petaurus australis*). *Australian Journal of Zoology* 54(5): 305-317.

Bunce M, Szulkin M, Lerner HRL, Barnes I, Shapiro B, Cooper A and Holdaway RN (2005). Ancient DNA provides new insights into the evolutionary history of New Zealand's extinct giant eagle. *PloS Biology* 3(1): 44-46.

Bunce M, Worthy TH, Phillips MJ, Holdaway RN, Willerslev E, Haile J, Shapiro B, Scofield RP, Drummond AJ, Kamp PJJ and Cooper A (2009). The evolutionary history of the extinct ratite Moa and New Zealand Neogene paleogeography. *Proceedings of the National Academy of Sciences of the United States of America* 106(49): 20646-20651.

Burbidge AA and Fuller PJ (2007). Gibson Desert birds: responses to drought and plenty. *Emu* 107(2): 126-134.

Burbidge AA, Johnson KA, Fuller PJ and Southgate RI (1988). Aboriginal knowledge of the mammals of the central deserts of Australia. *Australian Wildlife Research* 15: 9-39.

Burns E, Eldridge M, Crayn D and Houlden B (2007). Low phylogeographic structure in a wide spread endangered Australian frog *Litoria aurea* (Anura: Hylidae). *Conservation Genetics* 8(1): 17-32.

Butler WH (1961). Occurance of the Ghost bat, *Macroderma gigas*, in the Great Victoria Desert, W.A. *Western Australian Naturalist* 8(1): 42-43.

Byrne M (2008). Evidence for multiple refugia at different time scales during Pleistocene climatic oscillations in southern Australia inferred from phylogeography. *Quaternary Science Reviews* 27(27-28): 2576-2585.

Byrne M, Macdonald B and Brand J (2000). Phylogeography and divergence in the chloroplast genome of Western Australian Sandalwood (*Santalum spicatum*). *Heredity* 91(4): 389-395.

Byrne M, Yeates DK, Joseph L, Kearney M, Bowler J, Williams MAJ, Cooper S, Donnellan SC, Keogh JS, Leys R, Melville J, Murphy DJ, Porch N and Wyrwoll KH (2008). Birth of a biome: insights into the assembly and maintenance of the Australian arid zone biota. *Molecular Ecology* 17(20): 4398-4417.

Campos PF, Craig OE, Turner-Walker G, Peacock E, Willerslev E and Gilbert MTP (2011). DNA in ancient bone - Where is it located and how should we extract it? *Annals of Anatomy* 194(1): 7-16.

Campos PF, Kristensen T, Orlando L, Sher A, Kholodova MV, Gotherstrom A, Hofreiter M, Drucker DG, Kosintsev P, Tikhonov A, Baryshnikov GF, Willerslev E and Gilbert MTP (2010). Ancient DNA sequences point to a large loss of mitochondrial genetic diversity in the saiga antelope (*Saiga tatarica*) since the Pleistocene. *Molecular Ecology* 19(22): 4863-4875.

Cano RJ and Poinar HN (1993). Rapid isolation of DNA from fossil and museum specimens suitable for PCR. *Biotechniques* 15(3): 432.

Cano RJ, Poinar HN, Pieniazek NJ, Acra A and Poinar GO (1993). Amplification and sequencing of DNA from a 120-135-Million-year-old Weevil. *Nature* 363(6429): 536-538.

Carter G (1971). Pre-Columbian chickens in America. Man across the sea: Problems of pre-Columbian contacts. Riley C, Kelley J, Pennington C and Rands R. Austin, University of Texas Press: 178-218.

Champlot S, Berthelot C, Pruvost M, Bennett EA, Grange T and Geigl EM (2010). An efficient multistrategy DNA decontamination procedure of PCR reagents for hypersensitive PCR applications. *PLOS ONE* 5(9): e13042.

Chan YL, Lacey EA, Pearson OP and Hadly EA (2005). Ancient DNA reveals Holocene loss of genetic diversity in a South American rodent. *Biology Letters* 1(4): 423-426.

Chapple DG, Keogh JS and Hutchinson MN (2005). Substantial genetic substructuring in southeastern and alpine Australia revealed by molecular phylogeography of the *Egernia whitii* (Lacertilia : Scincidae) species group. *Molecular Ecology* 14(5): 1279-1292.

Churchill S (2008). *Australian Bats*. Crows Nest, Allen and Unwin.

Clark PU, Dyke AS, Shakun JD, Carlson AE, Clark J, Wohlfarth B, Mitrovica JX, Hostetler SW and McCabe AM (2009). The Last Glacial Maximum. *Science* 325(5941): 710-714.

Clarke AC, Burtenshaw MK, McLenachan Pa, Erickson DL and Penny D (2006). Proceedings of the SMBE Tri-National Young Investigators' Workshop 2005. Reconstructing the origins and dispersal of the Polynesian bottle gourd (*Lagenaria siceraria*). *Molecular Biology and Evolution* 23(5): 893-900.

Cooper A, Lalueza-Fox C, Anderson S, Rambaut A, Austin J and Ward R (2001). Complete mitochondrial genome sequences of two extinct Moas clarify ratite evolution. *Nature* 409(6821): 704-707.

Cooper A, Mourerchauvire C, Chambers GK, Vonhaeseler A, Wilson AC and Paabo S (1992). Independent origins of New Zealand Moas and Kiwis. *Proceedings of the National Academy of Sciences of the United States of America* 89(18): 8741-8744.

Cooper A and Poinar HN (2000). Ancient DNA: Do it right or not at all. *Science* 289(5482): 1139.

Davies SJJF (2002). *Ratites and Tinamous: Tinamidae, Rheidae, Dromaiidae, Casuaiidae, Apterygidae, Struthionidae*. Oxford, Oxford University Press.

de Bruyn M, Hoelzel AR, Carvalho GR and Hofreiter M (2011). Faunal histories from Holocene ancient DNA. *Trends in ecology & evolution (Personal edition)* 26(8): 405-413.

Deagle BE, Eveson JP and Jarman SN (2006). Quantification of damage in DNA recovered from highly degraded samples - a case study on DNA in faeces. *Frontiers in Zoology* 3: 11.

Debruyne R, Chu G, King CE, Bos K, Kuch M, Schwarz C, Szpak P, Grocke DR, Matheus P, Zazula G, Guthrie D, Froese D, Buigues B, de Marliave C, Flemming C, Poinar D, Fisher D, Southon J, Tikhonov AN, MacPhee RDE and Poinar HN (2008). Out of America: Ancient DNA evidence for a New World origin of Late Quaternary woolly mammoths. *Current Biology* 18(17): 1320-1326.

Desalle R, Barcia M and Wray C (1993). PCR jumping in clones of 30-Million-year-old DNA fragments from amber preserved termites (*Mastotermes electrodominicus*). *Experientia* 49(10): 906-909.

Dickman CR (1999). Rodent-Ecosystem Relationships: A Review. Singleton GR, Hinds LA, Leirs H and Zhang Z. Canberra, Australian Centre for International Agricultural Research. 59: 113-133.

Dobney K, Cucchi T and Larson G (2008). The pigs of Island Southeast Asia and the Pacific: New evidence for taxonomic status and human-mediated dispersal. *Asian Perspectives* 47(1): 59-74.

Dobney K and Larson G (2006). Genetics and animal domestication: new windows on an elusive process. *Journal of Zoology* 269.

Douglas AM (1956). A recent Western Australia record of the Ghost Bat (*Macroderma gigas*). *The Western Australian Naturalist* 5(6): 140-141.

Douglas AM (1957). The Ghost Bat in the North-west. *The Western Australian Naturalist* 5(8): 232.

Douglas AM (1967). The natural history of the Ghost bat, *Macroderma gigas* (Microchiroptera, Megadermatidae), in Western Australia. *The Western Australian Naturalist* 10(6): 125-137.

Dubey S and Shine R (2010). Evolutionary diversification of the lizard genus *Bassiana* (Scincidae) across Southern Australia. *PLOS ONE* 5(9).

Erickson DL, Smith BD, Clarke AC, Sandweiss DH and Tuross N (2005). An Asian origin for a 10,000-year-old domesticated plant in the Americas. *Proceedings of the National Academy of Sciences of the United States of America* 102(51): 18315-18320.

Etten EJB (2009). Inter-annual rainfall variability of arid Australia: greater than elsewhere? *Australian Geographer* 40(1): 109-120.

Fan LJ, Gui YJ, Zheng YF, Wang Y, Cai DG and You XL (2011). Ancient DNA sequences of rice from the low Yangtze reveal significant genotypic divergence. *Chinese Science Bulletin* 56(28-29): 3108-3113.

Gilbert MT, Bandelt HJ, Hofreiter M and Barnes I (2005). Assessing ancient DNA studies. *Trends in ecology & evolution* 20(10): 541-544.

Gilbert MTP, Tomsho LP, Rendulic S, Packard M, Drautz DI, Sher A, Tikhonov A, Dalen L, Kuznetsova T, Kosintsev P, Campos PF, Higham T, Collins MJ, Wilson AS, Shidlovskiy F, Buigues B, Ericson PGP, Germonpre M, Gotherstrom A, Iacumin P, Nikolaev V, Nowak-Kemp M, Willerslev E, Knight JR, Irzyk GP, Perbost CS, Fredrikson KM, Harkins TT, Sheridan S, Miller W and Schuster SC (2007a). Whole-genome shotgun sequencing of mitochondria from ancient hair shafts. *Science* 317(5846): 1927-1930.

Gilbert MTP, Wilson AS, Bunce M, Hansen AJ, Willerslev E, Shapiro B, Higham TFG, Richards MP, O'Connell TC, Tobin DJ, Janaway RC and Cooper A (2004). Ancient mitochondrial DNA from hair. *Current Biology* 14(12): 463-464.

Gilbert MTR, Djurhuus D, Melchior L, Lynnerup N, Worobey M, Wilson AS, Andreasen B and Dissing J (2007b). MtDNA from hair and nail clarifies the genetic relationship of the 15th century Qilakitsoq inuit mummies. *American Journal of Physical Anthropology* 133(2): 847-853.

Golenberg EM, Giannasi DE, Clegg MT, Smiley CJ, Durbin M, Henderson D and Zurawski G (1990). Chloroplast DNA-sequence from a Miocene magnolia species. *Nature* 344(6267): 656-658.

Gongora J, Rawlence NJ, Mobegi VA, Jianlin H, Alcalde JA, Matus JT, Hanotte O, Moran C, Austin JJ, Ulm S, Anderson AJ, Larson G and Cooper A (2008). Indo-European and Asian origins for Chilean and Pacific chickens revealed by mtDNA. *Proceedings of the National Academy of Sciences of the United States of America* 105(30): 10308-10313.

Graur D and Pupko T (2001). The Permian bacterium that isn't. *Molecular Biology and Evolution* 18(6): 1143-1146.

Guay PJ, Chesser R, Mulder R, Afton A, Paton D and McCracken K (2010). East-west genetic differentiation in Musk Ducks (*Biziura lobata*) of Australia suggests late Pleistocene divergence at the Nullarbor Plain. *Conservation Genetics* 11(6): 2105-2120.

Gyulai G, Humphreys M, Lagler R, Szabo Z, Toth Z, Bittsanszky A, Gyulai F and Heszky L (2006). Seed remains of common millet from the 4th (Mongolia) and 15th (Hungary) centuries: AFLP, SSR and mtDNA sequence recoveries. *Seed Science Research* 16(3): 179-191.

Hadly EA, Ramakrishnan U, Chan YL, Tuinen MV, Keefe K, Spaeth PA and Conroy CJ (2004). Genetic response to climatic change: Insights from ancient DNA and phylochronology. *PloS Biology* 2(10): e290.

Haile J, Froese DG, MacPhee RDE, Roberts RG, Arnold LJ, Reyes AV, Rasmussen M, Nielsen R, Brook BW, Robinson S, Demuro M, Gilbert MTP, Munch K, Austin JJ, Cooper A, Barnes I, Moller P and Willerslev E (2009). Ancient DNA reveals late survival of mammoth and horse in interior Alaska. *Proceedings of the National Academy of Sciences of the United States of America* 106(52): 22352-22357.

Hammond EL, Lymbery AJ, Martin GB, Groth D and Wetherall JD (2002). Microsatellite analysis of genetic diversity in wild and farmed Emus (*Dromaius novaehollandiae*). *The Journal of Heredity* 93(5): 376-380.

Handt O, Hoss M, Krings M and Paabo S (1994). Ancient DNA - Methodological challenges. *Experientia* 50(6): 524-529.

Hartnup K, Huynen L, Kanawa RT, Shepherd LD, Millar CD and Lambert DM (2011). Ancient DNA recovers the origins of Maori feather cloaks. *Molecular Biology and Evolution* 28(10): 2741-2750.

Hather J and Kirch PV (1991). Prehistoric sweet potato (*Ipomoea batatas*) from Mangaia Island, Central Polynesia. *Antiquity* 65: 887-893.

Hawkins CE, Baars C, Hesterman H, Hocking GJ, Jones ME, Lazenby B, Mann D, Mooney N, Pemberton D, Pyecroft S, Restani M and Wiersma J (2006). Emerging disease and population decline of an island endemic, the Tasmanian devil *Sarcophilus harrisii*. *Biological Conservation* 131(2): 307-324.

Heupink TH, Huynen L and Lambert DM (2011). Ancient DNA suggests dwarf and 'Giant' Emu are conspecific. *PLOS ONE* 6(4).

Hewitt G (2000). The genetic legacy of the Quaternary ice ages. *Nature* 405(6789): 907-913.

Higuchi R, Bowman B, Freiberger M, Ryder OA and Wilson AC (1984). DNA sequences from the *Quagga*, an extinct member of the horse family. *Nature* 312: 282-284.

Hofreiter M (2003). Phylogeny, diet, and habitat of an extinct ground sloth from Cuchillo Curo, Neuquen Province, southwest Argentina. *Quaternary Research* 59(3): 364-378.

Hofreiter M, Capelli C, Krings M, Waits L, Conard N, Munzel S, Rabeder G, Nagel D, Paunovic M, Jambresic G, Meyer S, Weiss G and Paabo S (2002). Ancient DNA analyses reveal high mitochondrial DNA sequence diversity and parallel morphological evolution of late Pleistocene cave bears. *Molecular Biology and Evolution* 19(8): 1244-1250.

Hofreiter M, Poinar HN, Spaulding WG, Bauer K, Martin PS, Possnert G and Paabo S (2000). A molecular analysis of ground sloth diet through the last glaciation. *Molecular Ecology* 9: 1975-1984.

Hofreiter M and Stewart J (2009). Ecological change, range fluctuations and population dynamics during the Pleistocene. *Current biology* 19(14): 584-594.

Hope AG, Waltari E, Fedorov VB, Goropashnaya AV, Talbot SL and Cook JA (2011). Persistence and diversification of the Holarctic shrew, *Sorex tundrensis* (Family Soricidae), in response to climate change. *Molecular Ecology* 20(20): 4346-4370.

Horrocks M and Bedford S (2005). Microfossil analysis of Lapita deposits in Vanuatu reveals introduced *Araceae* (Aroids). *Archaeology in Oceania* 40(2): 67-74.

Houlden BA, Costello BH, Sharkey D, Fowler EV, Melzer A, Ellis W, Carrick F, Baverstock PR and Elphinstone MS (1999). Phylogeographic differentiation in the mitochondrial control region in the koala, *Phascolarctos cinereus* (Goldfuss 1817). *Molecular Ecology* 8(6): 999-1011.

Hurles ME, Matisoo-Smith E, Gray RD and Penny D (2003). Untangling Oceanic settlement: the edge of the knowable. *Trends in Ecology & Evolution* 18(10): 531-540.

Huynen L, Lissone I, Sawyer S and Lambert D (2008). Genetic identification of Moa remains recovered from Tiniroto, Gisborne. *Journal of the Royal Society of New Zealand* 38(4): 231-235.

Huynen L, Millar CD, Scofield RP and Lambert DM (2003). Nuclear DNA sequences detect species limits in ancient Moa. *Nature* 425(6954): 175-178.

Jaenicke-Despres V, Buckler ES, Smith BD, Gilbert MTP, Cooper A, Doebley J and Paabo S (2003). Early allelic selection in maize as revealed by ancient DNA. *Science* 302(5648): 1206-1208.

Jones AG (2008). A theoretical quantitative genetic study of negative ecological interactions and extinction times in changing environments. *BMC Evolutionary Biology* 8.

Joseph L, Moritz C and Hugall A (1995). Molecular support for vicariance as a source of diversity in rainforest. *Proceedings of the Royal Society of London Series B-Biological Sciences* 260(1358): 177-182.

Joseph L and Wilke T (2006). Molecular resolution of population history, systematics and historical biogeography of the Australian ringneck parrots *Barnardius*: are we there yet? *Emu* 106(1): 49-62.

Joseph L, Wilke T and Alpers D (2002). Reconciling genetic expectations from host specificity with historical population dynamics in an avian brood parasite, Horsfield's Bronze-Cuckoo *Chalcites basalis* of Australia. *Molecular Ecology* 11(4): 829-837.

Joseph L, Wilke T, Ten Have J and Terry Chesser R (2006). Implications of mitochondrial DNA polyphyly in two ecologically undifferentiated but morphologically distinct migratory birds, the masked and white-browed woodswallows *Artamus spp*. of inland Australia. *Journal of Avian Biology* 37(6): 625-636.

Kanginakudru S, Metta M, Jakati RD and Nagaraju J (2008). Genetic evidence from Indian red jungle fowl corroborates multiple domestication of modern day chicken. *BMC Evolutionary Biology* 8: 174-174.

Kearns AM, Joseph L and Cook LG (2010). The impact of Pleistocene changes of climate and landscape on Australian birds: a test using the Pied Butcherbird (*Cracticus nigrogularis*). *Emu* 110(4): 285-295.

Kearns AM, Joseph L, Edwards SV and Double MC (2009). Inferring the phylogeography and evolutionary history of the splendid fairy-wren *Malurus splendens* from mitochondrial DNA and spectrophotometry. *Journal of Avian Biology* 40(1): 7-17.

King D, Oliver A and Mead R (1978). The adaptation of some Western Australian mammals to food plants containing fluoroacetate. *Australian Journal of Zoology* 26(4): 699-712.

Kirch P (2000). *On the Road of the Winds: An Archaeological History of the Pacific Islands Before European Contact*. Berkeley, University of California Press.

Krause J, Unger T, Nocon A, Malaspinas AS, Kolokotronis SO, Stiller M, Soibelzon L, Spriggs H, Dear PH, Briggs AW, Bray SCE, O'Brien SJ, Rabeder G, Matheus P, Cooper A, Slatkin M, Paabo S and Hofreiter M (2008). Mitochondrial genomes reveal an explosive radiation of extinct and extant bears near the Miocene-Pliocene boundary. *BMC Evolutionary Biology* 8.

Kuch U, Keogh JS, Weigel J, Smith LA and Mebs D (2005). Phylogeography of Australia's king brown snake (*Pseudechis australis*) reveals Pliocene divergence and Pleistocene dispersal of a top predator. *Naturwissenschaften* 92(3): 121-127.

Larson G, Albarella U, Dobney K, Rowley-Conwy P, Schibler J, Tresset A, Vigne JD, Edwards CJ, Schlumbaum A, Dinu A, Balacsescu A, Dolman G, Tagliacozzo A, Manaseryan N, Miracle P, Van Wijngaarden-Bakker L, Masseti M, Bradley DG and Cooper A (2007a). Ancient DNA, pig domestication, and the spread of the Neolithic into Europe. *Proceedings of the National Academy of Sciences of the United States of America* 104(39): 15276-15281.

Larson G, Cucchi T, Fujita M, Matisoo-Smith E, Robins J, Anderson A, Rolett B, Spriggs M, Dolman G, Kim TH, Thuy NT, Randi E, Doherty M, Due RA, Bollt R, Djubiantono T, Griffin B, Intoh M, Keane E, Kirch P, Li KT, Morwood M, Pedrina LM, Piper PJ, Rabett RJ, Shooter P, Van den Bergh G, West E, Wickler S, Yuan J, Cooper A and Dobney K (2007b). Phylogeny and ancient DNA of *Sus* provides insights into neolithic expansion in Island Southeast Asia and Oceania. *Proceedings of the National Academy of Sciences of the United States of America* 104(12): 4834-4839.

Larson G, Dobney K, Albarella U, Fang MY, Matisoo-Smith E, Robins J, Lowden S, Finlayson H, Brand T, Willerslev E, Rowley-Conwy P, Andersson L and Cooper A (2005a). Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science* 307(5715): 1618-1621.

Larson G, Dobney K, Albarella U, Matisoo-Smith E, Robins J, Lowden S, Rowley-Conwy P, Andersson L and Cooper A (2005b). Domesticated pigs in eastern Indonesia - Response. *Science* 309(5733): 381-381.

Larson G, Liu R, Zhao X, Yuan J, Fuller D, Barton L, Dobney K, Fan Q, Gu Z, Liu XH, Luo Y, Lv P, Andersson L and Li N (2010). Patterns of East Asian pig domestication, migration, and turnover revealed by modern and ancient DNA. *Proceedings of the National Academy of Sciences of the United States of America* 107(17): 7686-7691.

Leonard J, Shanks O, Hofreiter M, Kreuz E, Hodges L, Ream W, Wayne R and Fleischer R (2007). Animal DNA in PCR reagents plagues ancient DNA research. *Journal of Archaeological Science* 34(9): 1361-1366.

Liu YP, Wu GS, Yao YG, Miao YW, Luikart G, Baig M, Beja-Pereira A, Ding ZL, Palanichamy MG and Zhang YP (2006). Multiple maternal origins of chickens: out of the Asian jungles. *Molecular Phylogenetics and Evolution* 38(1): 12-19.

Lorenzen ED, Nogues-Bravo D, Orlando L, Weinstock J, Binladen J, Marske KA, Ugan A, Borregaard MK, Gilbert MTP, Nielsen R, Ho SYW, Goebel T, Graf KE, Byers D, Stenderup JT, Rasmussen M, Campos PF, Leonard JA, Koepfli KP, Froese D, Zazula G, Stafford TW, Aaris-Sorensen K, Batra P, Haywood AM, Singarayer JS, Valdes PJ, Boeskorov G, Burns JA, Davydov SP, Haile J, Jenkins DL, Kosintsev P, Kuznetsova T, Lai XL, Martin LD, McDonald HG, Mol D, Meldgaard M, Munch K, Stephan E, Sablin

M, Sommer RS, Sipko T, Scott E, Suchard MA, Tikhonov A, Willerslev R, Wayne RK, Cooper A, Hofreiter M, Sher A, Shapiro B, Rahbek C and Willerslev E (2011). Species-specific responses of Late Quaternary megafauna to climate and humans. *Nature* 479(7373): 359-U195.

Madeja J, Wacnik A, Wypasek E, Chandran A and Stankiewicz E (2010). Integrated palynological and molecular analyses of late Holocene deposits from Lake Milkowskie (NE Poland): Verification of local human impact on environment. *Quaternary International* 220(1-2): 147-152.

Malmstrom H, Svensson EM, Gilbert MTP, Willerslev E, Gotherstrom A and Holmlund G (2007). More on contamination: The use of asymmetric molecular behavior to identify authentic ancient human DNA. *Molecular Biology and Evolution* 24(4): 998-1004.

Martinez-Cruz B and Godoy JA (2007). Genetic evidence for a recent divergence and subsequent gene flow between Spanish and Eastern imperial eagles. *BMC Evolutionary Biology* 7.

McIlroy J (1984). The sensitivity of Australian animals to 1080 poison. Vii. Native and introduced birds. *Wildlife Research* 11(2): 373-385.

Meggers BJ (1975). The transpacific origin of Mesoamerican civilization: A preliminary review of the evidence and its theoretical implications. *American Anthropologist* 77(1): 1-27.

Miller CR and Waits LP (2003). The history of effective population size and genetic diversity in the Yellowstone grizzly (*Ursus arctos*): implications for conservation. *Proceedings of the National Academy of Sciences of the United States of America* 100(7): 4334-4339.

Miller W, Hayes VM, Ratan A, Petersen DC, Wittekindt NE, Miller J, Walenz B, Knight J, Qi J, Zhao FQ, Wang QY, Bedoya-Reina OC, Katiyar N, Tomsho LP, Kasson LM, Hardie RA, Woodbridge P, Tindall EA, Bertelsen MF, Dixon D, Pyecroft S, Helgen KM, Lesk AM, Pringle TH, Patterson N, Zhang Y, Kreiss A, Woods GM, Jones ME and Schuster SC (2011). Genetic diversity and population structure of the endangered marsupial *Sarcophilus harrisii* (Tasmanian devil). *Proceedings of the National Academy of Sciences of the United States of America* 108(30): 12348-12353.

Montenegro A, Avis C and Weaver A (2008). Modeling the prehistoric arrival of the sweet potato in Polynesia. *Journal of Archaeological Science* 35(2): 355-367.

Morrison RJ (1990). *Pacific atoll soils: Chemistry, mineralogy and classification*. Washington D.C., National Museum of Natural History.

Morton SR, Smith DMS, Dickman CR, Dunkerley DL, Friedel MH, McAllister RRJ, Reid JRW, Roshier DA, Smith MA, Walsh FJ, Wardle GM, Watson IW and Westoby M (2011). A fresh framework for the ecology of arid Australia. *Journal of Arid Environments* 75: 313-329.

Morueta-Holme N, Flojgaard C and Svenning J-C (2010). Climate change risks and conservation implications for a threatened small-range mammal species. *PLOS ONE* 5(4): e10360.

Morwood MJ, Soejono RP, Roberts RG, Sutikna T, Turney CSM, Westaway KE, Rink WJ, Zhao JX, van Den Bergh GD, Due RA, Hobbs DR, Moore MW, Bird MI and Fifield LK (2004). Archaeology and age of a new hominin from Flores in eastern Indonesia. *Nature* 431(7012): 1087-1091.

Muller RA and MacDonald GJ (1997). Glacial cycles and astronomical forcing. *Science* 277(5323): 215-218.

Mullis KB and Faloona FA (1987). Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain-reaction. *Methods in Enzymology* 155: 335-350.

Nasab HM, Mardi M, Talaee H, Nashli HF, Pirseyedi SM, Nobari AH and Mowla SJ (2010). Molecular analysis of ancient DNA extracted from 3250-3450 year-old plant seeds excavated from Tepe Sagz Abad in Iran. *Journal of Agricultural Science and Technology* 12(4): 459-470.

Nielsen-Marsh C (2002). Biomolecules in fossil remains: Multidisciplinary approach to endurance. *Geology*(June): 12-14.

Nogues-Bravo D, Rodriguez J, Hortal J, Batra P and Araujo MB (2008). Climate change, humans, and the extinction of the Woolly Mammoth. *PloS Biology* 6(4): e79.

O'Brien PJ (1972). The sweet potato: Its origin and dispersal. *American Anthropologist* 74(3): 342-365.

Oliver A, King D and Mead R (1979). Fluoroacetate tolerance, a genetic marker in some Australian mammals. *Australian Journal of Zoology* 27(3): 363-363.

Oskam CL, Haile J, McLay E, Rigby P, Allentoft ME, Olsen ME, Bengtsson C, Miller GH, Schwenninger JL, Jacomb C, Walter R, Baynes A, Dortch J, Parker-Pearson M, Gilbert MTP, Holdaway RN, Willerslev E and Bunce M (2010). Fossil avian eggshell preserves ancient DNA. *Proceedings of the Royal Society of London Series B-Biological Sciences* 277(1690): 1991-2000.

Oskam CL, Jacomb C, Allentoft ME, Walter R, Scofield RP, Haile J, Holdaway RN and Bunce M (2011). Molecular and morphological analyses of avian eggshell excavated from a late thirteenth century earth oven. *Journal of Archaeological Science* 38(10): 2589-2595.

Ovchinnikov IV, Gotherstrom A, Romanova GP, Kharitonov VM, Liden K and Goodwin W (2000). Molecular analysis of Neanderthal DNA from the northern Caucasus. *Nature* 404(6777): 490-493.

Paabo S (1989). Ancient DNA: Extraction, characterization, molecular cloning, and enzymatic amplification. *Proceedings of the National Academy of Sciences of the United States of America* 86: 1939-1943.

Paabo S and Wilson A (1991). Miocene DNA sequences - a dream come true? *Current Biology* 1(1): 45-46.

Paplinska JZ, Taggart DA, Corrigan T, Eldridge MDB and Austin JJ (2011). Using DNA from museum specimens to preserve the integrity of evolutionarily significant unit boundaries in threatened species. *Biological Conservation* 144(1): 290-297.

Poinar HN, Hofreiter M, Spaulding WG, Martin PS, Stankiewicz BA, Bland H, Evershed RP, Possnert G and Paabo S (1998). Molecular coproscopy: Dung and diet of the extinct ground sloth *Nothrotheriops shastensis*. *Science* 281(5375): 402-406.

Poulakakis N, Lymberakis P, Paragamian K and Mylonas M (2005). Isolation and amplification of shrew DNA from barn owl pellets. *Biological Journal of the Linnean Society* 85(3): 331-340.

Prideaux GJ, Long Ja, Ayliffe LK, Hellstrom JC, Pillans B, Boles WE, Hutchinson MN, Roberts RG, Cupper ML, Arnold LJ, Devine PD and Warburton NM (2007). An arid-adapted middle Pleistocene vertebrate fauna from south-central Australia. *Nature* 445(7126): 422-425.

Quigley MC, Horton T, Hellstrom JC, Cupper ML and Sandiford M (2010). Holocene climate change in arid Australia from speleothem and alluvial records. *The Holocene* 20(7): 1093-1104.

Rawlence NJ, Wood JR, Armstrong KN and Cooper A (2009). DNA content and distribution in ancient feathers and potential to reconstruct the plumage of extinct avian taxa. *Proceedings of the Royal Society of London Series B-Biological Sciences* 276(1672): 3395-3402.

Reed EH and Bourne SJ (2000). Pleistocene fossil vertebrate sites of the South East region of South Australia. *Transactions of the Royal Society of South Australia* 124: 61-90.

Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, Viola B, Briggs AW, Stenzel U, Johnson PLF, Maricic T, Good JM, Marques-Bonet T, Alkan C, Fu Q, Mallick S, Li H, Meyer M, Eichler EE, Stoneking M, Richards M, Talamo S, Shunkov MV, Derevianko AP, Hublin J-J, Kelso J, Slatkin M and Paabo S (2010). Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* 468(7327): 1053-1060.

Robins JH, Matisoo-Smith EA and Furey L (2001). Hit or miss? Factors affecting DNA preservation in Pacific archaeological material. Australasian Connections and New Directions: Proceedings of the 7th Australasian Archaeometry Conference, Auckland, University of Auckland.

Rohland N, Siedel H and Hofreiter M (2004). Nondestructive DNA extraction method for mitochondrial DNA analyses of museum specimens. *Biotechniques* 36(5): 814.

Rowe KC, Singhal S, MacManes MD, Ayroles JF, Morelli TL, Rubidge EM, Bi K and Moritz CC (2011). Museum genomics: low-cost and high-accuracy genetic data from historical specimens. *Molecular Ecology Resources* 11: 1082-1092.

Schneider CJ, Cunningham M and Moritz C (1998). Comparative phylogeography and the history of endemic vertebrates in the Wet Tropics rainforests of Australia. *Molecular Ecology* 7(4): 487-498.

Schwarz C, Debruyne R, Kuch M, McNally E, Schwarcz H, Aubrey AD, Bada J and Poinar H (2009). New insights from old bones: DNA preservation and degradation in permafrost preserved mammoth remains. *Nucleic acids research* 37(10): 3215-3229.

Shafer ABA, Cullingham CI, Cote SD and Coltman DW (2010). Of glaciers and refugia: a decade of study sheds new light on the phylogeography of northwestern North America. *Molecular Ecology* 19(21): 4589-4621.

Shapiro B, Drummond AJ, Rambaut A, Wilson MC, Matheus PE, Sher AV, Pybus OG, Gilbert MTP, Barnes I, Binladen J, Willerslev E, Hansen AJ, Baryshnikov GF, Burns JA, Davydov S, Driver JC, Froese DG, Harington CR, Keddie G, Kosintsev P, Kunz ML, Martin LD, Stephenson RO, Storer J, Tedford R, Zimov S and Cooper A (2004). Rise and fall of the Beringian Steppe Bison. *Science* 306(5701): 1561-1565.

Shepherd LD and Lambert DM (2008). Ancient DNA and conservation: lessons from the endangered kiwi of New Zealand. *Molecular Ecology* 17(9): 2174-2184.

Short J and Smith A (1994). Mammal decline and recovery in Australia. *Journal of Mammalogy* 75(2): 288-297.

Sidow A, Wilson AC and Paabo S (1991). Bacterial-DNA in Clarkia fossils. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* 333(1268): 429-433.

Smith CI, Chamberlain AT, Riley MS, Cooper A, Stringer CB and Collins MJ (2001). Not just old but old and cold? *Nature* 410(6830): 771-772.

Smith FA and Betancourt JL (2006). Predicting woodrat (*Neotoma*) responses to anthropogenic warming from studies of the palaeomidden record. *Journal of Biogeography* 33(12): 2061-2076.

Spencer PBS, Rhind SG and Eldridge MDB (2001). Phylogeographic structure within *Phascogale* (Marsupialia: Dasyuridae) based on partial cytochrome b sequence. *Australian Journal of Zoology* 49(4): 369-377.

Steadman DW and Intoh M (1994). Biogeography and prehistoric exploitation of birds from Fais Island, Yap State, Federated States of Micronesia. *Pacific Science* 48(2): 116-135.

Storey AA, Ladefoged T and Matisoo-Smith EA (2008a). Counting your chickens: Density and distribution of chicken remains in archaeological sites of Oceania. *International Journal of Osteoarchaeology* 18(3): 240-261.

Storey AA, Quiroz D, Ramirez JM, Beavan-Athfield N, Addison DJ, Walter R, Hunt T, Athens JS, Huynen L and Matisoo-Smith EA (2008b). Pre-Columbian chickens, dates, isotopes, and mtDNA. *Proceedings of the National Academy of Sciences of the United States of America* 105(48): E99.

Storey AA, Ramirez JM, Quiroz D, Burley DV, Addison DJ, Walter R, Anderson AJ, Hunt TL, Athens JS, Huynen L and Matisoo-smith EA (2007). Radiocarbon and DNA evidence for a pre-Columbian introduction of Polynesian chickens to Chile. *Proceedings of the National Academy of Sciences of the United States of America* 104(25): 10335-10339.

Storey AA, Spriggs M, Bedford S, Hawkins SC, Robins JH, Huynen L and Matisoo-Smith E (2010). Mitochondrial DNA from 3000-year old chickens at the Teouma site, Vanuatu. *Journal of Archaeological Science* 37(10): 2459-2468.

Taberlet P, Fumagalli L, Wust-Saucy A and Cosson J (1998). Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology* 7(4): 453-464.

Theves C, Senescau A, Vanin S, Keyser C, Ricaut FX, Alekseev AN, Dabernat H, Ludes B, Fabre R and Crubezy E (2011). Molecular identification of bacteria by total sequence screening: Determining the cause of death in ancient human subjects. *PLOS ONE* 6(7).

Toon A, Mather PB, Baker AM, Durrant KL and Hughes JM (2007). Pleistocene refugia in an arid landscape: analysis of a widely distributed Australian passerine. *Molecular Ecology* 16(12): 2525-2541.

Twigg L, King D, Davis H, Saunders D and Mead R (1988). Tolerance to, and metabolism of, fluoroacetate in the Emu. *Wildlife Research* 15(3): 239-247.

USB (2006). Shrimp DNase, Recombinant. Corporation U. Cleveland, USB Corporation: 1-2.

van den Bergh GD, Meijer HJM, Due Awe R, Morwood MJ, Szabo K, van den Hoek Ostende LW, Sutikna T, Saptomo EW, Piper PJ and Dobney KM (2009). The Liang Bua faunal remains: a 95k.yr. sequence from Flores, East Indonesia. *Journal of human evolution* 57(5): 527-537.

Varela S, Lobo JM, Rodriguez J and Batra P (2010). Were the Late Pleistocene climatic changes responsible for the disappearance of the European spotted hyena populations? Hindcasting a species geographic distribution across time. *Quaternary Science Reviews* 29: 2027-2035.

Vilà C, Leonard JA, Götherström A, Marklund S, Sandberg K, Lidén K, Wayne RK and Ellegren H (2001). Widespread origins of domestic horse lineages. *Science* 291(5503): 474-477.

Wandeler P, Hoeck PEA and Keller LF (2007). Back to the future: museum specimens in population genetics. *Trends in Ecology & Evolution* 22(12): 634-642.

Whitaker T and Carter G (1954). Oceanic drift of gourds-experimental observations. *American Journal of Botany* 41: 697-700.

Willerslev E, Cappellini E, Boomsma W, Nielsen R, Hebsgaard MB, Brand TB, Hofreiter M, Bunce M, Poinar HN, Dahl-Jensen D, Johnsen S, Steffensen JP, Bennike O, Schwenninger JL, Nathan R, Armitage S, de Hoog CJ, Alfimov V, Christl M, Beer J, Muscheler R, Barker J, Sharp M, Penkman KEH, Haile J, Taberlet P, Gilbert MTP, Casoli A, Campani E and Collins MJ (2007). Ancient biomolecules from deep ice cores reveal a forested Southern Greenland. *Science* 317(5834): 111-114.

Willerslev E, Hansen AJ, Binladen J, Brand TB, Gilbert MTP, Shapiro B, Bunce M, Wiuf C, Gilichinsky DA and Cooper A (2003). Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science* 300(5620): 791-795.

Wilmshurst JM, Anderson AJ, Higham T and Worthy TH (2008). Dating the late prehistoric dispersal of Polynesians to New Zealand using the commensal Pacific rat. *Proceedings of the National Academy of Sciences of the United States of America* 105(22): 7676-7680.

Wilmshurst JM, Hunt TL, Lipo CP and Anderson AJ (2011). High-precision radiocarbon dating shows recent and rapid initial human colonization of East Polynesia. *Proceedings of the National Academy of Sciences of the United States of America* 108(5): 1815-1820.

Woodward SR, Weyand NJ and Bunnell M (1994). DNA sequence from Cretaceous period bone fragments. *Science* 266(5188): 1229-1232.

Worthington Wilmer J, Moritz C, Hall L and Toop J (1994). Extreme population structuring in the threatened Ghost Bat, *Macroderma gigas* - Evidence from mitochondrial DNA. *Proceedings of the Royal Society of London Series B-Biological Sciences* 257(1349): 193-198.

Zenger KR, Eldridge MDB and Cooper DW (2003). Intraspecific variation, sex-biased dispersal and phylogeography of the eastern grey kangaroo (*Macropus giganteus*). *Heredity* 91(2): 153-162.

Zhang D, Rossel G, Kriegner A and Hijmans R (2004). AFLP assessment of diversity in sweetpotato from Latin America and the Pacific region: Its implications on the dispersal of the crop. *Genetic Resources and Crop Evolution* 51(2): 115-120.

Chapter 2

Methodology

This thesis aims to test the ability of ancient DNA (aDNA), in combination with other techniques, to reconstruct the recent evolutionary history of a range of species: the Australian Ghost Bat (chapter 4) and Emu (chapter 3) and the Polynesian chicken (chapter 5). These species were chosen because they exemplify evolution across a range of timescales and geographic situations, and are represented by well-preserved ancient bones in caves, archaeological sites and museums.

This chapter, on Methodology, outlines the aDNA methods used to extract genetic data from the Emu, Ghost Bat, and Chicken samples, before describing the coalescent-based techniques used to reconstruct their evolutionary history, such as Bayesian Evolutionary Analysis of Sampling Trees (BEAST), Bayesian Skyline Plots (BSP), Bayesian Serial Simcoal (BayeSSC), and the use of Arlequin to perform tests of demographic and range expansions. This chapter then discusses the Ecological Niche Modelling (ENM) methods, such as Ecological Niche Factor Analysis (ENFA) and Maximum Entropy (MaxEnt), used to predict the Ghost Bat's ecological niche and geographic distributions in current, past, and future environments. Lastly, the statistical tests used to examine the haplotype frequencies recording the dispersal of the Polynesian Chicken' will be described.

Ancient DNA laboratory methods

In the early days of the aDNA field, DNA was first extracted from fossil remains using a phenol/chloroform protocol with centrifugal dialysis (Paabo *et al.* 1988), but this was soon replaced by silica-based methods that use guanidinium thiocyanate (GuSCN), a chaotropic salt that reduces co-extraction of inhibitors (Hoss and Paabo 1993; Rohland and Hofreiter 2007a; 2007b). All aDNA methods aim to maximise the extraction of endogenous DNA, while minimising the co-extraction of inhibitors from the surrounding environment, such as humic and fulvic acids (Wilson 1997), and Maillard products (Paabo 1989). Silica-based kits for specific sample types that utilise the inhibitor reduction potential of GuSCN are now commercially available. The application of these commercial kits to aDNA samples has been proven to significantly reduce the per-sample cost of extracting DNA without affecting the efficiency (Yang *et al.* 1998).

The optimisation of these silica-GuSCN extraction methods for different types of sample is still underway (eggshell; Oskam *et al.* 2010; museum samples; Paplinska *et al.* 2011). The extraction method I developed for the work outlined in this thesis is another instance of optimisation of a silica-GuSCN extraction method, this time specifically developed for specimens where minimal sample volumes are available, which is common in studies of small mammals or

birds. Instead of extracting DNA from the collagen-containing fraction only (after removal of the hydroxyapatite-containing fraction), my extraction method digests and purifies DNA from both fractions simultaneously using a modification of the commercial DNeasy Blood and Tissue Kit (Qiagen, Valencia, CA). Recent research quantifying the differential preservation of DNA in each of these fractions has found that an approach using both fractions gives higher DNA yield than those which discard the hydroxyapatite-containing EDTA supernatant (Yang *et al.* 1998; Rohland and Hofreiter 2007b; Campos *et al.* 2011)

Extraction method

Each bone was ground to fine powder in a Mikrodismembrator (5000 rpm, for 10 seconds). For specimens with relatively large sample volumes, approximately 70 mg of bone powder was decalcified concurrently with protein digestion by incubation at 55 °C overnight in 1mL of extraction buffer (consisting of 0.4725 M EDTA (pH=8.0), 0.2 % sodium dodecyl sulphate (SDS), and 0.7 $mg.ml^{-1}$ Proteinase K). After digestion, samples were centrifuged at 10,000 rpm for 5 mins and the supernatant was transferred to an Amicon ultra-4 (Millipore), which was centrifuged at 4000 xg until only 100 µL supernatant remained. The supernatant was washed with 1 mL molecular grade water and centrifuged again (at 4000 xg until only 100 µl remained). An equal volume of ATL buffer (Qiagen DNeasy kit) was then added, mixed, and the supernatant removed to a 2 mL screw-cap tube. The supernatant was incubated for 10–60 mins at room temperature on a rotary mixer after the addition of an equal volume of AL buffer (Qiagen DNeasy kit) and 0.02 $\mu g.\mu l^{-1}$ of carrier RNA. After the incubation, an equal volume of ethanol (100 %) was added, and then the total volume was transferred to a Qiagen DNeasy spin column where it was incubated at room temperature for 10–60 mins. The extraction then followed the Qiagen DNeasy kit instructions, with the following exceptions at the elution stage: 100–150 µL of warmed AE buffer was added and then incubated at room temperature for 10–30 mins, before being centrifuged at 8,000 rpm for 1 min, this step was repeated to finish with 200–300 µL of total volume.

In other studies (Thomson *et al.* unpublished) this protocol was modified such that individual teeth are digested whole, *i.e.,* without the need for powdering. This reduces the need for sample handling (eliminating another potential source of contamination) and promotes the retention of more starting material (avoiding the powdering step means no potential loss of material either into the air due to static electricity, or on equipment used to powder the material). This method is ideal for small mammals, such as rodents, where teeth are small enough that the reagent volumes required for complete digestion will fit into a standard 2ml screw cap tube. Experience has shown that many larger tubes can leak if placed on a rotary mixer at 55 °C overnight, causing potential cross-contamination between samples.

All other laboratory methods are detailed within each chapter, including PCR and post-PCR Sanger sequencing. The major analytical techniques used in this thesis are outlined below, with the inherent problems, limitations, and advantages of each discussed.

Population genetic statistics

Even before PCR and Sanger sequencing technologies provided easy access to the genome, the pioneers of population genetics, R. A. Fisher, Sewell Wright, and G. Watterson were studying the microevolutionary processes occurring in populations. These three researchers were at the forefront of the traditional population genetic theories/statistics that were being developed in the 1950's and 1960's, such as F-statistics (Wright 1951; 1965), number of segregating sites (Watterson 1975), and tests of neutrality (Watterson 1978). In the post-PCR world, new statistics to describe microevolutionary processes was still being uncovered, such as inbreeding coefficients (Slatkin and Hudson 1991), mismatch distributions (Rogers and Harpending 1992), AMOVA (Excoffier *et al.* 1992), linkage disequilibrium (Slatkin 1994), and exact tests of population differentiation (Raymond and Rousset 1995).

Many of these population genetic statistics have been used for decades to describe the evolutionary history of organisms (long before the field of aDNA was developed). However, population genetic statistics are all based on synchronous genetic signals (*i.e,* data from one time period, *e.g.,* contemporary). When heterochronous data (*i.e.,* data from across multiple time points) are included in calculations of these population genetic statistics, the additional mutations that occur on branches between the ancient and contemporary time points may affect the inferred demographic history (Depaulis *et al.* 2009). When the heterochronous samples represent a considerable part of the tree depth (even just 10%) and are included with the contemporary samples to estimate these population statistics, the effects of the ancient mutations on these population statistics may substantially alter the demographic conclusions drawn from a study (Depaulis *et al.* 2009). However, when the heterochrony is only across a few generations (in fact when ancient samples are < 0.1 x 2Ne generations), the additional mutations may have a negligible effect on these population statistics. Also, when the ratio of ancient to contemporary samples is low (*e.g.,* < 0.1), the effect on population statistics is minimal (Depaulis *et al.* 2009). This is an important issue and one that has been overlooked in ancient population studies (Valdiosera *et al.* 2007; Valdiosera *et al.* 2008; O'Keefe *et al.* 2009). The latter study by Valdiosera *et al.* (2008) potentially overestimated Pleistocene and Holocene estimates of θ (population mutation parameter) from highly heterochronous Iberian brown bear datasets (Pleistocene dataset spanned 70 kya which represents >2Ne, and Holocene dataset spanned 7 kya), leading to likely erroneous conclusions about a severe bottleneck at the Pleistocene-Holocene boundary and a lesser bottleneck from Holocene-recent.

However in this thesis, neither the Emu (chapter 3), nor the Ghost Bat (chapter 4) datasets are as heterochronous as many aDNA studies. Although the ancient Emu samples range from 8,800-16,500 ya (which represents 25% of the tree depth), the same haplotypes are present in both the ancient and contemporary samples (Figure 1), suggesting that only a few mutations have arisen over the last few tens of thousands of years. Whether or not the Emu population can be classes as a 'measurably evolving population' as defined by Drummond *et al.* (2003) is evaluated in the Discussion in chapter 6. However, it is interesting to note that the Emu dataset did pass the date randomisation test of Ho *et al.* (2011b) suggesting there is enough signal in the heterochronous samples to estimate molecular rates and node dates.

With respect to the Ghost Bat, the relatively recent age of the ancient populations (inferred from the stratigraphic position as surface finds on the cave floors) suggests the ancient Ghost Bat samples are within the ‘few generations’ category rather than tens of thousands of years. Considering the tMRCA of the Central West is 342 kya, even if the ancient NAR populations spanned ten thousand years old, the NAR population would still only represent 0.058Ne, suggesting there would be little effect on any population genetic statistics, such as S (segregating sites) and π (average number of differences between two random sequences), which are the basis for the expansion tests in Arlequin. If a slight effect of heterochrony was experienced in the population genetic statistics of the NAR population, the effect would be an overestimation of expansion signal (Depaulis *et al.* 2009). However, as noted above, studies like that of Valdiosera *et al.* (2008) on Iberian brown bears that had highly heterochronous datasets (~1-7Ne), are much more susceptible to drawing erroneous conclusions about population events.

However, even given the additional power within heterochronous dataset, population genetic statistics have the disadvantage of being purely descriptive in nature; they can reveal patterns (Wright’s F_{ST} may reveal significant population differentiation) but they cannot test between competing hypotheses about evolutionary processes. In contrast, coalescent theory (below) provides the ability to explicitly test competing demographic scenarios, such as rates and direction of migration, presence of bottlenecks, and genetic admixture.

The coalescent

Coalescent theory provides the basis for the mathematical modelling of a genealogical history for a dataset. The coalescence process, most simply, allows the reconstruction of all genetic lineages backwards in time to a common ancestor – and this process is also equivalent to the process of genetic drift forwards in time (Figure 2). Within a population the coalescent process can be used to calculate the mathematics of the expected distribution of these ancestral coalescent times (Kingman 2000), which means our knowledge about the relationship between patterns of common ancestry and population demography (population size, growth rate, patterns of gene flow *etc.*) can be used to simulate datasets undergoing different demographic histories. As the coalescent process reconstructs population histories backwards from the individuals actually present (haplotypes not detected in a population do not need to be reconstructed), the reconstruction is much more efficient than traditional population genetics techniques which track every potential individual forward in time. Even so, reconstructing the coalescent process is computationally intensive, and the user must take into account the limitations and assumptions that underlie this technique. These assumptions include: no recombination within loci (ideal for mtDNA), selective neutrality (thought to make mtDNA ideal, however this is increasingly being debated; Gerber *et al.* 2001), the topology of the coalescent tree is independent from the mutational process (allowing faster computation), and the sequence data follows an infinite-site model (a mutational model where sites are assumed to mutate only once in its history; Tajima 1996). This last point results in many analyses requiring the removal of sites that display homoplasy (*i.e.,* two allele states), which occurs in ancient samples due to either real mutations occurring in some but not all mitochondrial genomes, or damaged sites masquerading as mutations.

Coalescent theory underlies many analytical techniques used in population genetics and demographic reconstructions, *e.g.,* Bayesian Evolutionary Analysis of Sampling Trees (BEAST; Drummond and Rambaut 2007), Bayesian Serial SimCoal (BayeSSC; Anderson *et al.*), and tests for demographic and range expansions (in Arlequin; Excoffier and Lischer 2010).

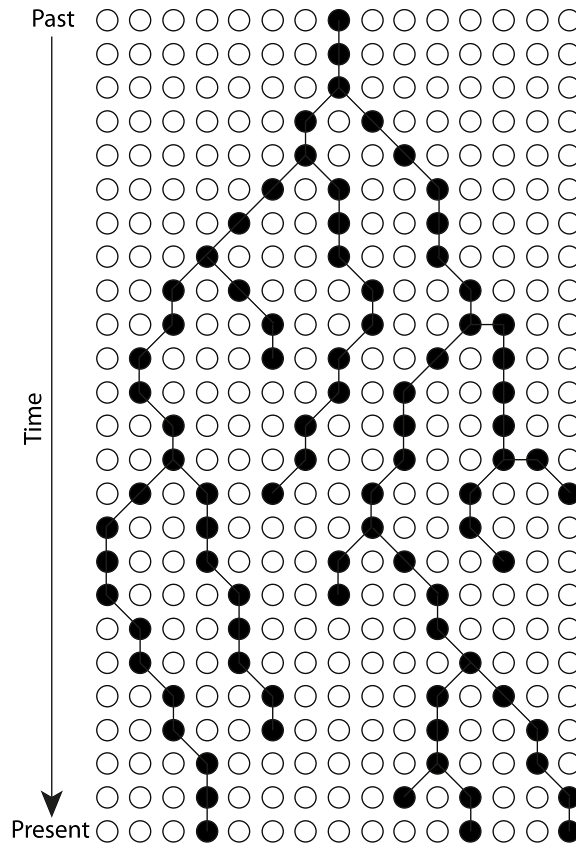


Figure 2. Illustration of the coalescent process where haplotypes are traced backward to the time of the most recent common ancestor (tMRCA). Coalescent theory uses only modern-day samples (n=3 in this example), therefore reconstruction of this process is much more efficient than traditional population genetics techniques which track every individual forwards in time.

<u>**Bayesian Evolutionary Analysis of Sampling Trees (BEAST)**</u>

The software program BEAST (Bayesian Evolutionary Analysis of Sampling Trees) computes a Bayesian analysis of molecular sequences (Drummond and Rambaut 2007). BEAST was designed to use a front-end software, BEAUti, to input data, set up the parameters and models to be tested, and output an .xml file for running in BEAST. BEAST itself uses the Metropolis-Hastings Markov Chain Monte Carlo algorithm to sample from probability distributions, based on a Markov chain stepping semi-randomly through 'tree space' evaluating and rejecting proposed moves based on the probability distribution of the sampled trees. Each tree is weighted according to its posterior probability. After a set number of steps, the chain stops and is evaluated for convergence to the stationary distribution. A good chain will achieve convergence quickly from any random starting tree, but knowledge about how long the chain will take to reach convergence is not known *a priori*. If the tree space for a particular analysis has many peaks separated by deep valleys, the chain may get stuck in local maxima without reaching the global maxima that will give a good representation of the posterior distribution. Therefore, multiple chains are usually run as a check that convergence has occurred to the global, and not a

local, maxima. As BEAST can estimate multiple demographic parameters simultaneously, 'tree space' does not only represent the posterior distribution for a phylogenetic tree but also nucleotide substitution models, demographic models, tree shape priors, relaxed clock models, node calibration models *etc*. Complex evolutionary models can now be estimated by calculating a range of these simple models at the same time (Drummond and Rambaut 2007). By calculating these parameters simultaneously, hypothesis testing of evolutionary models can be done without being biased by selection of a single tree topology *a priori* (Drummond and Rambaut 2007).

In addition to phylogenetic and genealogical tree building, BEAST also produces supplementary plots – Bayesian Skyline Plots (BSP) – describing changes in population size (NeT) over time (Drummond *et al.* 2005). The demographic history described in a BSP is represented as a median line with 95% High Posterior Distribution (HPD; equivalent to margins of error), which measures changes in population size over time. Reconstructing demographic changes using BSPs (Figure 3) requires specification of a coalescent-Bayesian Skyline tree prior and the number of changes in effective population size in BEAUti. However, even though setting up a BSP run is simple, the BSP is highly parameterised and computationally intensive; in fact, depending on the chain length of the run, the visualisation software (Tracer v1.5) may require additional memory just to reconstruct the BSP from the BEAST log and tree files.

Many large iconic species have been analysed using BSPs, showing: expansions mid-Holocene (Horses; Lippold *et al.* 2011), expansions after the LGM (Native Americans; Kumar *et al.* 2011), declines after the LGM (Bison; Drummond *et al.* 2005), and historical crashes due to human hunting (Fur seal; Hoffman *et al.* 2011). However, the BSP is only useful when the dataset is strongly informative about population history, *i.e.,* requires a highly polymorphic dataset with a history of relatively simple and recent, yet dramatic, demographic change (Drummond and Rambaut 2007). This is not the case for the Emu (chapter 3). The Emu lacks genetic diversity, which makes reconstructing a BSP of demographic change difficult. Figure 3 shows a BSP of the Emu dataset where a straight line can be drawn through the 95% HPD (in purple), indicating a constant population size could not be ruled out. Comparing the BEAST reconstruction of a constant population size and the BSP of demographic change using Bayes Factors also indicates that a constant population size is strongly supported (BF = 56.65). In addition, the Emu exhibits evidence of purely range expansions, the genetic signal of which will not be detected in a BSP as skyline plots only estimate numerical not geographic changes in populations (see chapter 3). The range expansions were detected using Arlequin expansion tests, which also found evidence against a demographic expansion (see below). The BayeSSC analysis also refutes a demographically expanding single population, with the Emu population more likely a single population that became two sub-structured populations with low levels of migration (see chapter 3).

A BSP was reconstructed for the Emu, as the BEAST assumption of panmixia was not violated. Panmixia is defined as random mating within a population. The large dispersal distances that the Emu can traverse make it plausible that the Emu is acting as one interbreeding population at the continent-scale. However, as the Ghost Bat has a very restricted distribution, one panmictic population cannot be assumed. No BSP was attempted in chapter 4.

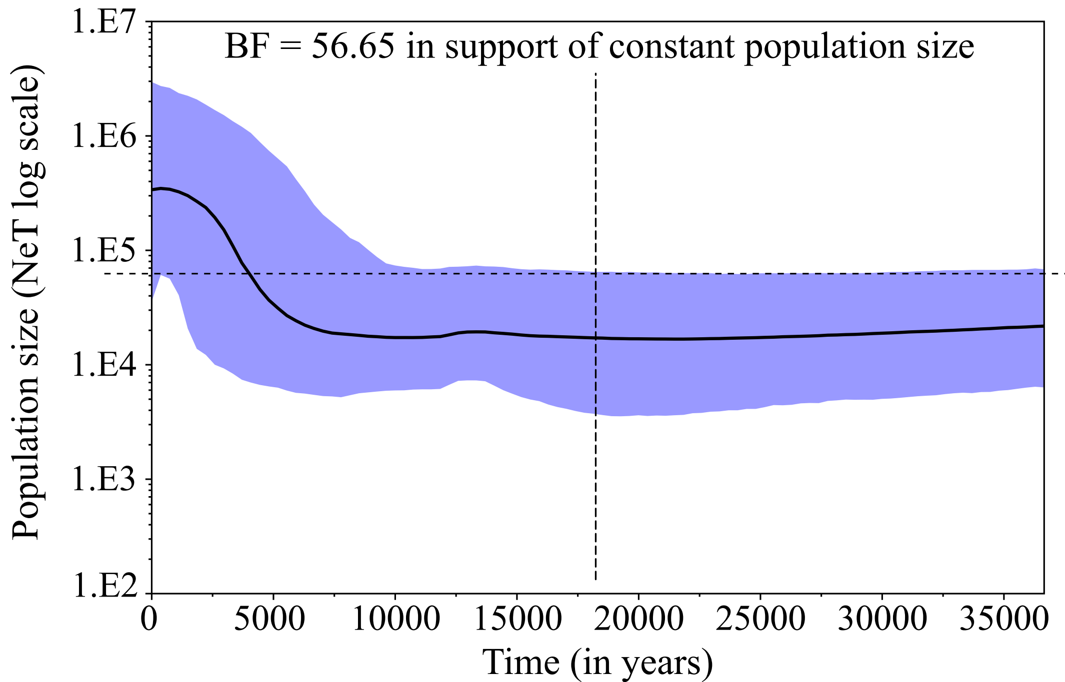


Figure 3. Bayesian Skyline Plot (BSP) for an example dataset of the mainland Emu, showing how the demographic history can be estimated from sequence data (a constant population size could not be ruled out however, as a straight line can still be drawn through the centre of the upper and lower 95% HPD interval). The x-axis is in units of time before present, and the y-axis is equal to Neτ (the product of effective population size and the generation time in years). The Bayes Factor (BF) supporting either constant population size or BSP is indicated in upper right hand corner (BF <1 is negative, 1<BF<3 is barely worth mentioning, 3<BF<20 is positive, 20<BF<150 is strong, and BF>150 is very strong evidence; Sibon Li and Drummond 2011), with the horizontal dashed line indicating the mean population size estimated from BEAST runs of constant population size. The vertical line indicates date of earliest sample for each dataset, with demography to the right of the line speculative.

Another useful tool within BEAST is the ability to recreate a multispecies coalescent using *BEAST (pronounced 'StarBEAST') (Heled and Drummond 2010). *BEAST was used to assess the mutation rate and phylogeny of the Ghost Bat dataset (see discussion in chapter 6) for comparison to the BEAST mutation rate and phylogeny presented in chapter 4. Where BEAST can reconstruct either phylogenetic relationships (using single individuals across multiple species) or population histories (demographic events within a single interbreeding population/species), *BEAST uses multiple individuals per species to co-estimate many gene trees within a shared species tree in combination with the demography of both extant and ancestral species. In *BEAST analyses, increasing the number of individuals per species increases the accuracy of the parameter and decreases the confidence intervals (95% HPD), thereby achieving more accurate estimates of divergence times (Drummond *et al.* 2005). Another way to decrease the error and credible interval sizes is by increasing the loci number, however this does not rule out using single locus datasets.

One advantage of the BEAST and *BEAST algorithms is that they allow the user to include as much temporal information as is available: purely modern sequence data, known fossil dates for particular nodes on the tree, and heterochronous data such as aDNA samples with radiocarbon dates. This allows the software program to be very flexible; if only modern data is available it is still possible to calculate phylogenies in units of mutation, while fossil calibrations can be used to date deep nodes on the tree. However, as in all analytical techniques, the use of calibration points requires an understanding of the assumptions inherent in the algorithms: *i.e.,* what are the consequences of using each type of calibration, what are the errors associated with extrapolating over such large temporal scales (tens of millions of years), and what affect do assumptions about the constancy of evolutionary rate have while species are being subject to differing selection pressures over time. Additionally, when very deep node dates are used to calibrate the tree, even with a relaxed molecular clock, the resulting evolutionary rate will appear much slower than rates calibrated at intermediate parts of the tree (calibrated with aDNA). At the other extreme is genealogical data, which has the fastest rate, as those polymorphisms not yet fixed within a species also contribute to the rate. This effect is known as the 'rates curve', which was first described by Ho *et al.* (2005), and has been related to a number of biases including inherent differences in the evolutionary rates of polymorphisms versus substitutions (Ho *et al.* 2011a). There is currently no way to avoid or correct for this bias, undermining the validity of many studies that have used deep fossil calibrations (*e.g.,* 65 Mya calibration of ostrich – emu/cassowary split) on trees to infer species or population level events (speciation events within Kiwis; Burbidge *et al.* 2003). However, a major advantage of the BEAST software suite is the ability to include radiocarbon dates associated with aDNA samples, allowing calibration points to

be applied to the tips of the tree, ensuring more accurate calculation of the mutation rate over more recent timescales of species or population or level events.

One of the best examples of the potential applications of BEAST and heterochronous samples is the Bison (*Bison bison sp.*) study of Shapiro *et al*. (2004) where 38% of the 442 samples had radiocarbon dates, and a BSP analysis reconstructed a dramatic decline in the Pleistocene Bison population. This species is one of the few worldwide where significant amounts of both time and money have been focused on answering questions about species' responses to Pleistocene climate change. Where radiocarbon dated samples are not available – either due to a lack of collagen in samples or lack of funds for radiocarbon dates – fossil calibrations remain the one of the few ways to add a temporal dimension to demographic and divergence events. Provided researchers are aware of the inherent bias in rate estimates, of issues extrapolating over evolutionary time-scales, and the uncertainty surrounding divergence dates that the rates curve creates, then the field of evolutionary biology will be open to solutions to solve this issue.

Approximate Bayesian Computation (ABC)

Although the coalescent theory allows hypotheses based on competing demographic scenarios to be tested, the software programs that use the coalescent usually examine a single type of evolutionary model, such as migration (*e.g.,* MIGRATE), population divergence (*e.g.,* IMa), or population size change (*e.g.,* BEAST's BSP). In reality, populations undergo all of these evolutionary processes over time (not just one in isolation), yet the coalescent alone cannot separate these signals, nor establish when or in what order each process is likely to have occurred within a population. Approximate Bayesian Computation (ABC) is a new technique that allows complex scenarios to be tested using coalescent simulations (Figure 4).

The ABC process starts by the user: 1) defining the evolutionary models to be tested; 2) choosing summary statistics that will best describe the model within the particular population/dataset; 3) simulating datasets that are equivalent to those observed in the population under study (based on the same number of samples, and same age of samples if the dataset is heterochronous); 4) calculating the same summary statistics for each simulated dataset; 5) selecting a subset of the simulations with the closest summary statistics to that of the empirical dataset; 6) choosing amongst those models with the highest probability; 7) estimating the parameter values; and 8) examining the quality of the model selection and parameter estimates (Grunwald 2011). Therefore, comparing the summary statistics from the simulated populations to the sampled population, allows conclusions to be drawn about the most likely demographic history of the species in question.

Bayesian Serial SimCoal (BayeSSC)

SimCoal (Excoffier *et al.* 2000) is a program that simulates molecular genetic diversity (step 3 above) for populations under different demographic scenarios (such as bottlenecks, complex instances of admixture, drift, migration *etc*.), which are then evaluated using ABC. The simulations of gene genealogies allow empirical distributions to be generated for any population genetic statistic, which are then tested against the population genetic statistics of the empirical dataset under study to evaluate the most likely of any number of competing hypotheses. BayesSSC (Anderson *et al.* 2004) is a modification of SimCoal that allows the extra information from serial sampling (*i.e.,* aDNA samples or virus samples) to be included in reconstructing simulated populations. In essence this means BayeSSC can simulate what occurred prior to an event, such as a bottleneck. Hypotheses to be tested can involve parameters such as the number of interbreeding populations within a dataset, the size of each population, the growth rates for each population (in % per generation), migration rates between the different populations (in % per generation), time since an 'event', *e.g.,* an expansion or divergence (in generations ago), mutation rate (in mutations per sequence length per generation). Although the number of possible population events that can be tested appear almost limitless (see Figure 4 for an example), the BayeSSC simulation process is very computationally intensive, with the number of simulations required for statistical robustness increasing with the number of parameters to simulate:

$$S = 10^{(3+p)}$$

where S is the number of simulations, and p is the number of parameters.

Due to computational limitations this number of simulations is rarely performed, for example, only 5000 simulations were performed in a study on ground squirrels where up to 6 parameters were estimated (O'Keefe *et al.* 2009); and 1000 simulations of up to 12 parameters for a study investigating ancestral relationships between Etruscan and modern Tuscans (Belle *et al.* 2006). However, other studies have performed larger numbers of simulations: Haak *et al.* (2010) analysed 100,000 simulations for up to 9 parameters, and Campos *et al.* (2010) used 200,000 simulations for 5 parameters. In chapter 3, 100,000 simulations were performed to estimate up to 4 parameters, which is closer to statistical certainty than many previous studies have used. Currently, the main limitation of BayeSSC is the computing power available to perform sufficient numbers of simulations for statistical robustness.

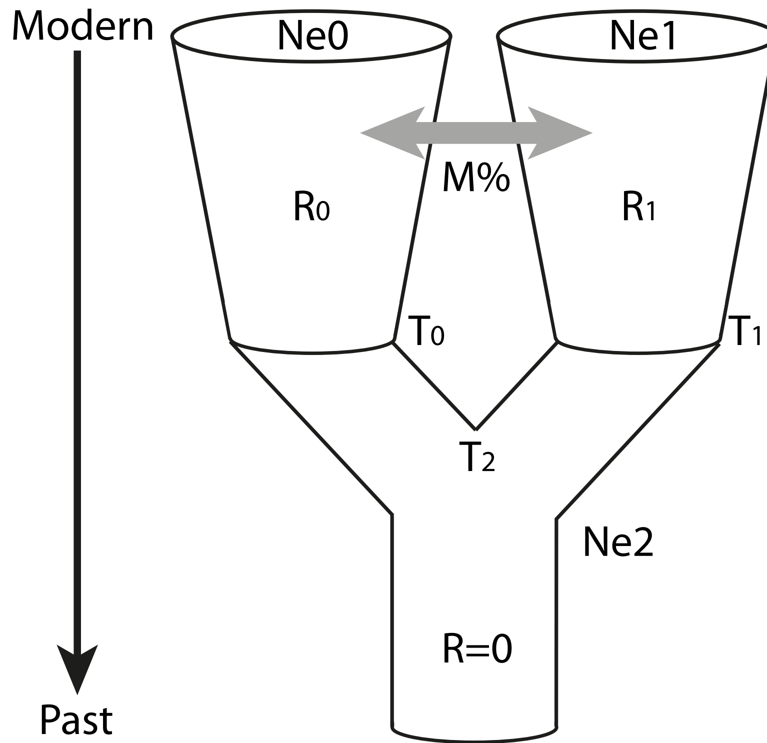


Figure 4. An example of a scenario to be tested in Bayesian Serial SimCoal (BayeSSC). Two present-day populations (of population size Ne0 and Ne1 with M% per generation bi-directional migration) that have expanded (at rate R0 and R1) since some time in the past (T0 and T1). This is prior to some 'event' (T2) when they were one constant-sized ancestral population (of size Ne2); some or all of these parameters can be simulated and used to generate population genetic statistics for this simulated population. These statistics can be compared against population genetic statistics for an empirical dataset to assess whether this model is more or less likely than another model to explain the genetic patterns observed in the empirical dataset.

Another issue is that two of the population genetics statistics (haplotype diversity and F_{ST}) estimated by BayeSSC for the simulations are not calculated in the same way as Arlequin, even though BayeSSC is based on SimCoal, also written by the authors of Arlequin (Excoffier *et al.* 2000; Excoffier and Lischer 2010). Although the BayeSSC website (http://www.stanford.edu/group/hadlylab/ssc/index.html) mentions these two differences, not enough emphasis is placed on the need to recalculate these statistics using the same methodology as BayeSSC. The method that BayeSSC uses to calculate haplotype diversity is biased by a factor of ${}^{(n-1)}/_{n}$, n = number of samples. The recalculation is relatively simple to convert from the haplotype diversity statistics that are output from Arlequin. However, the fixation index F_{ST} (a measure of genetic divergence between populations) is calculated using a different formula. BayeSSC uses the F_{ST} formula of Hudson *et al.* (1992), whereby:

$$F_{ST} = 1 - {}^{H_w}/_{H_b}$$

where H_w is the mean number of differences between different sequences sampled from the same subpopulation, and H_b is the mean number of differences between sequences sampled from the two different subpopulations sampled.

However, Arlequin uses the standard F_{ST} formula of Wright (1951) whereby:

$$F_{ST} = {}^{1}/_{(4Nm+1)}$$

where N is the population effective size, and m is the rate of migration into the population.

Using ABC, the posterior distributions of the parameters can be evaluated against the prior distribution from the empirical values for the summary statistics using the locfit, akima, and lattice packages (R development core team 2011), plus the reject function from http://www.stanford.edu/group/hadlylab/ssc/eval.r. The euclidean distance between the simulated and empirical values is calculated and those within *delta* units (*delta* chosen to usually include between 0.1-10% of simulations provided this is greater than 50 simulations) are retained. The reject function will return a cumulative density function for each parameter being estimated, so that the range of values for which the posterior density surpasses the prior distribution can be selected for a 2nd round of BayeSSC for more simulations. An Akaike Information Criterion (AIC) value for each scenario being tested will also be output – AIC is an heuristic calculation used to assess whether the improvement in fit of a more complex model justifies the increase in parameters it requires. The AIC values of multiple scenarios can be compared and the hypothesis with the lowest AIC value is the more likely evolutionary history (of those being tested) for the observed population.

Therefore, while BayeSSC provides a powerful ability to test multiple complex hypotheses about population/species history using samples dated through aDNA, it is necessary to recognise and deal with several bugs in the current version.

Demographic and Range expansion tests in Arlequin

Initially, the coalescent was developed for use in simple models in single populations, however it has now been expanded to include subdivided populations (Wakeley 1999; 2001) and those with some level of migration (Slatkin 1995). In addition, population demography (Li 1977) is no longer the only aspect of a species evolutionary history that can be reconstructed. The recent work of Ray *et al.* (2003) linking range expansions with specific ancestry patterns now allows parameters of the range expansion to be estimated through simulations.

In Arlequin there are tests for two different types of expansions: a sudden step-wise demographic expansion, and a sudden spatial or range expansion. Populations undergoing each of

these expansion types have particular signals in summary statistics and mismatch distribution; a mismatch distribution is a frequency graph of the pairwise nucleotide differences between any two random individuals of a population. Any population in equilibrium has an 'L' shaped mismatch distribution with a high frequency of individuals with the same haplotype (*i.e.,* zero differences), while populations that have experienced an expansion will have one or more peaks (see Figure 5).

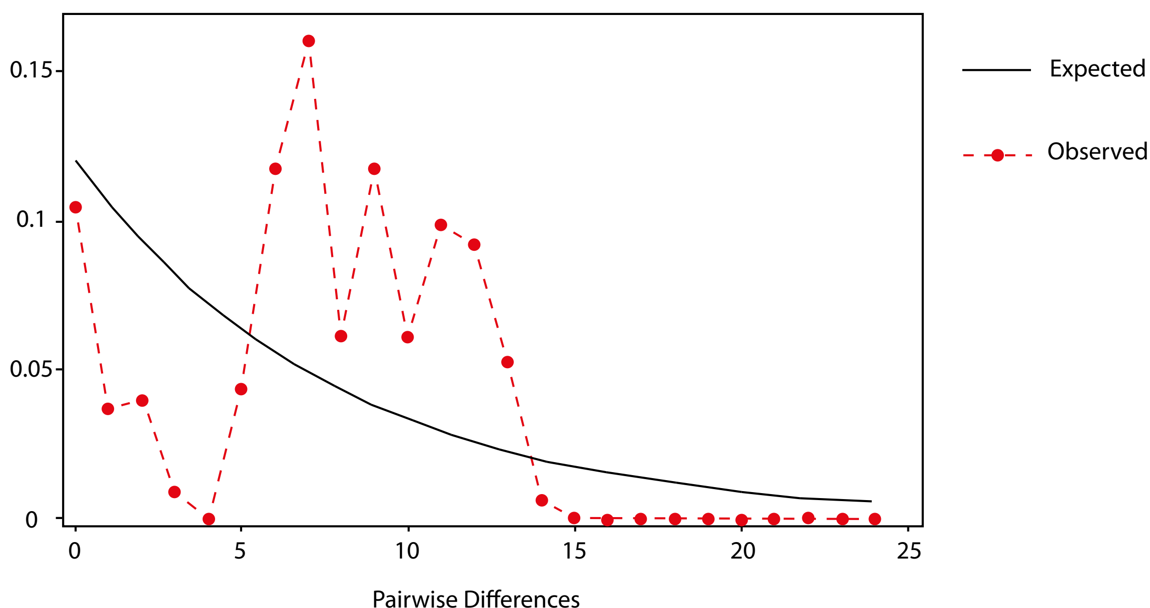


Figure 5. Example mismatch distribution showing pairwise differences observed in a dataset (red dashed line) and that expected in a population at equilibrium (*e.g.,* constant population size; black solid line).

Populations experiencing a sudden demographic expansion have: (1) gene genealogies that exhibit a star-shape with short internal branches, which reflects an excess of rare mutations (Harpending and Rogers 2000); and (2) mismatch distributions that are unimodal (*e.g.,* a single peak or wave; see Figure 5). As the x-axis of these mismatch distributions are measured in number of pairwise nucleotide differences, the location of the peak or wave is dependent on time since the expansion. Furthermore, the vertical intercept of the curve (tailing or left-hand edge) is inversely proportional to the degree of population increase, and the steepness of the leading edge of the peak (right-hand edge) is inversely proportional to the initial population size (Rogers and Harpending 1992). All three of these characteristics of the mismatch distributions allow the reconstruction of certain demographic parameters: population size before, and after the expansion, as well as the time since expansion. Although the calculation of the initial (θ_0) and final population sizes (θ_1) cannot be reconstructed with much accuracy, there is reasonable precision about the time since expansion (τ) (Schneider and Excoffier 1999).

The formulae describing a demographic expansion was initially developed by Li (1977); where we assume that if a stationary population at equilibrium increased in population size from *N0* to *N1* at τ (*tau*) generations ago, then the probability of observing *S* differences between two randomly selected (non-recombining) haplotypes is defined by:

$$F_S(\tau, \theta_0, \theta_1) = F_S(\theta_1) + exp\left(-\tau\frac{\theta_1 + 1}{\theta_1}\right)\sum_{j=0}^{S}\frac{\tau^j}{j!}\left[F_{S-j}(\theta_0) - F_{S-j}(\theta_1)\right]$$

where $F_S(\theta_1) = \frac{\theta^s}{(\theta+1)^{S+1}}$ is the probability of observing two random haplotypes with S differences in a stationary population, $\theta_0 = 2\mu N_0$, $\theta_1 = 2\mu N_1$, $\tau = 2\mu T$, and μ is the mutation rate for the whole haplotype.

Ray *et al.* (2003) also present a method that allows a test for range expansions. Although range expansions may also involve increases in population size, they do not necessarily result in the same genetic signature as purely demographic expansions (see below). The formulae describing a range expansion was developed by Excoffier (2004); where genes are sampled from a single deme belonging to a population subdivided into an infinite number of demes, each with population size '*N*', exchanging a fraction '*M'* of migrants with other demes, then some τ (*tau*) generations in the past, the system was reduced to a single deme of size N_0. Using this simple model, the probability that two genes sampled in a small deme of size *N* differ at *S* sites as below:

$$F_0(S; M, \theta_0; \theta_1, \tau) = \frac{\theta_1^S}{A^{S+1}} + \sum_{j=0}^{S}\left(\frac{(Me^{-\tau} + C)\theta_0^j\tau^{S-j}}{(M+1)(\theta_0+1)^{j+1}(S-j)!} - \frac{\tau^j\theta_1^{S-j}C}{j!\,A^{S-j+1}}\right)$$

where $\theta_0 = 2\mu N_0$, $\theta_1 = 2\mu N_1$, $\tau = 2\mu T$, and $A = \theta_1 + M + 1$, and $C = e^{-\tau A/\theta_1}$.

The signal of a demographic expansion differs most from that of a range expansion when the number of migrants (*Nm*) is small, however when *Nm* is large many characteristics are similar: unimodal mismatch distributions and star-shaped genealogies. In contrast, when few migrants are involved in a range expansion, the mismatch distributions are multimodal, with both very short and long branch lengths, and low levels of segregating sites (*S*), resulting from the majority of coalescent events occurring more recently. In addition to these mismatch distributions and genealogical shape differences between large and small *Nm* range expansions, large *Nm* range expansions exhibit high levels of *S*, very significantly negative Tajima's D and Fu's F_S, which are all related to the majority of coalescent events occurring early in the population history, around the onset of the range expansion (Ray *et al.* 2003). As the tests for both the demographic and range expansions in Arlequin resemble a 'black box', little more can be gleaned about how

the tests differentiate between demographic expansions and range expansions with large Nm, without writing a script to calculate the estimates of *tau* explicitly and simulating populations with large Nm under different expansion types and severities.

Even in the recent review paper by Excoffier *et al.* (2009), no clear explanation is proposed to differentiate the similar patterns of these two quite different population histories. For example, where range expansions occur with low levels of migration (*M*) the mismatch distribution resembles instances where a demographic expansion increases the population size dramatically (Excoffier 2004). Similarly, where range expansions exhibit high levels of migration the mismatch distributions resembles cases where demographic expansions result in very small or negligible increases in population size (Excoffier 2004). This has important ramifications for the Emu in chapter 3, as this dataset apparently exhibits evidence against demographic expansions, but support for quite ancient range expansions. The ecology of the Emu may provide a clue as to the more likely demographic history: the demographic expansion test is based on the assumption that the population has remained stationary, which in the Emu's case is unlikely given their highly mobile nature and often long distance dispersal, sometimes on the scale of hundreds-to-thousands of kilometres (Davies 2002).

Transforming expansion time estimates to a demographic scale

Very few studies have utilised the new complement of these expansion tests (Banguera-Hinestroza *et al.* 2010). Where these expansion tests have been used, even fewer have transformed the *tau* values of time of expansion into years. A recent study has noted that many issues exist in the conclusions drawn from studies that do carry out this transformation: errors in the use of mutation rates vs. divergence rates, incorrect units of mutation, and/or inappropriate generation times for the species of interest (Schenekar and Weiss 2011). Furthermore, even when the mutation rate and generation time used are as accurate as possible, researchers still need to realise that these parameters are actually averages over evolutionary time periods during which life history traits and the species' environment can change, inevitably resulting in inaccuracies. In each of the chapters of this thesis where these tests are used, the Emu (chapter 3) and the Ghost Bat (chapter 4), generation times are only approximate. The formula used in this thesis to calculate generation time is:

$$\frac{(T_1 + T_2)}{2}$$

where T_1 is age at first breeding and T_2 is age at last breeding. Accurate estimates of age at first and last breeding from wild populations are ideal, however this information is not always available. Any estimate of these ages from captive animals will be less accurate, although age at first breeding will be less inaccurate as it is likely related more to physiological constants. Furthermore, age at death in the wild, where known, is probably a more accurate estimate of age at last breeding than age at last breeding for captive birds.

As few long-term studies of Ghost Bats have been undertaken on wild populations – partially due to the species' sensitivity to disturbance – the age at first breeding (1 year) and last breeding (14 years) was based on captive animals (Jones 2008). Similarly, little information is known about Emu generation time. The age at first breeding is known from captive birds (2 years), which should not change substantially compared to wild birds. Fortunately the age at death in wild Emus is known from banding studies (of 350 Emus fitted with bands with a 14% recovery rate, no birds older than 6 years were recovered).

Ecological Niche Modelling

Theory

Another technique rarely combined with aDNA is ecological niche modelling (ENM) or species distribution modelling (SDM) – both terms are used interchangeably. These types of predictive modelling of species' ecosystem requirements and geographic distributions are becoming more common as environmental data has become more detailed and widely available. These ENM techniques relate the occurrence data for a species to their abiotic or biotic environments, *i.e.,* these methods model the ecological niche of a species (*e.g.,* where can the species colonise?). However, only recently has it been pointed out that these SDM methods are based on the assumption that the probability of occurrence is related to the quality of habitat (Basille *et al.* 2008). What is possibly of equal importance is an understanding of the factors determining the species distribution (*e.g.,* what is the species searching for?). If we cannot define a species' ecological niche (*i.e.,* the processes that drive species' distributions) and thereby *a priori* reduce the complexity of the many explanatory variables (*i.e.,* bioclimatic layers) available for input into ENM modelling, then any ecological models developed from the ENM will be of limited use and non-robust. Furthermore, conservation efforts based on these models will be for naught. The body of work describing the need to reduce the quantity of potential predictor variables is gaining momentum (Rushton *et al.* 2004; Basille *et al.* 2008; Varela *et al.* 2011; Warren and Seifert 2011). Rushton *et al*. (2004) highlight the fact that the number of potential ENM models increase by a factor of two factorial of the number of predictor variables used, and therefore selecting amongst these potential models for the 'best' model may be influenced greatly by the pathway taken on the search. Warren et al. (2011) concurs with a need to avoid overparameterisation as they find that not only does model complexity affect inferences with respect to habitat suitability, and the relative importance of ecogeographical variables in determining species distributions, but also the ability to transfer these predictions into other temporal and spatial arenas.

Reducing complexity using Ecological Niche Factor Analysis (ENFA)

In chapter 4 on the Ghost Bat, ENM is used to examine predicted distributions under various spatio-temporal scenarios. Although the Ghost Bat is known to require certain temperature and humidity levels in the cave roosts they inhabit (Baudinette *et al.* 2000), it was important not to bias the model by assuming that these are the same biogeographic variables influencing their distribution at the continental-scale. Therefore, it was desirable to avoid culling variables arbitrarily in an attempt to reduce the complexity of using all 22 ecogeographical variables (from climatic and topographical layers) available online for Australia. Basille *et al.* (2008) suggest using Ecological Niche Factor Analysis (ENFA, within the 'adehabitat' package in R; Hirzel *et al.* 2002) to establish the ecological niche of the species in question: ENFA explores ecological space to maximise 1) the difference between the species' optimal conditions and the conditions available in the study area (the marginality, see Figure 6), and 2) the ratio of variance of available conditions over variance of conditions used by the species (the specialisation, see Figure 6). In essence, the marginality of a species' ecological niche is defined as how much the niche deviates from the average environment/habitat of the study area, and the specialisation of a species' ecological niche is defined as the narrowness of the niche (Basille *et al.* 2008). Furthermore, Varela *et al.* (2011) explicitly state that the use of ENFA to select the most biologically meaningful variables for modelling the distribution of species is especially beneficial for paleo-species distribution modelling (PSDM) because the larger the number of predictor variables, the narrower the predicted niche will be.

By first using ENFA to select the ecogeographical variables that most closely determine the ecological niche of the Ghost Bat – 22 ecogeographical variables were reduced down to the nine variables with highest marginality. A Kruskal-Wallace analysis of variance was then used to preferentially remove those annual or monthly variables correlated with more than one other variable, with the remaining six bioclimatic variables highly predictive of the Ghost Bat ecological niche. The six bioclimatic variables included: annual mean temperature (bio1), isothermality (temperature variability, bio3), mean temperature of coldest quarter (bio11), precipitation seasonality (bio15), precipitation of wettest quarter (bio16), and precipitation of warmest quarter (bio18). By evaluating the six highly predictive variables, the Ghost Bat's ecological niche appears to be warm and humid (Figure 7). The six variables were subsequently used in the MaxEnt modelling performed for the Ghost Bat in chapter 4.

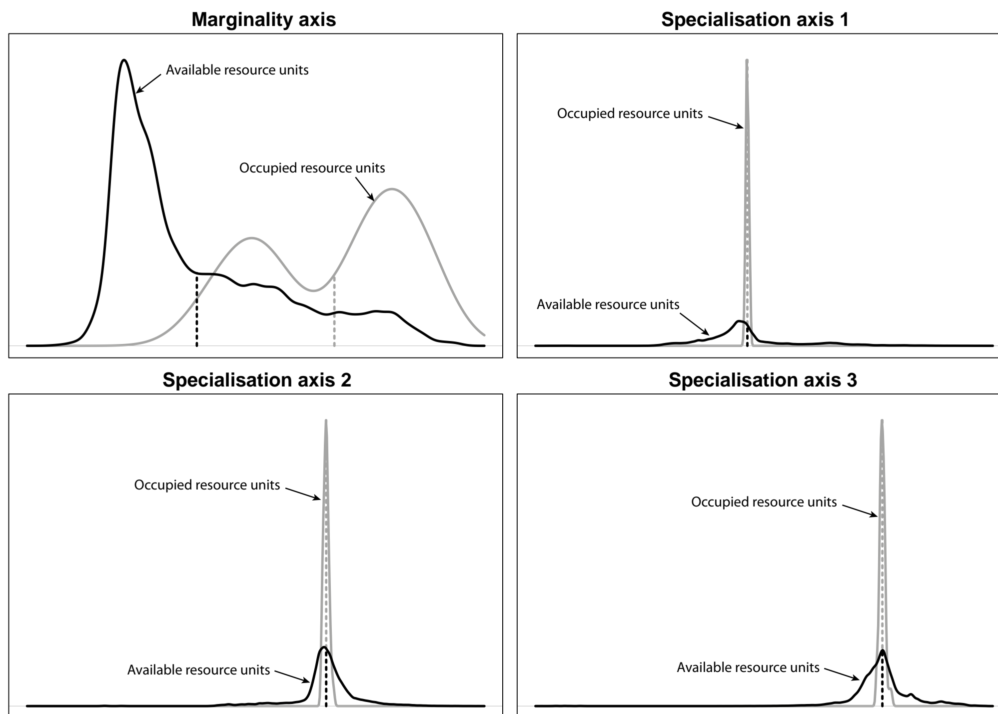


Figure 6. ENFA density plots of marginality axis and first three specialisation axes for the Ghost Bat, showing available resource units (in black) and the occupied resource units (in grey). The Ghost Bat's ecological niche is positively skewed and deviates to a reasonable extent from the available habitat in the study area, while all three specialisation axes showing how narrow the Ghost Bat ecological niche is.

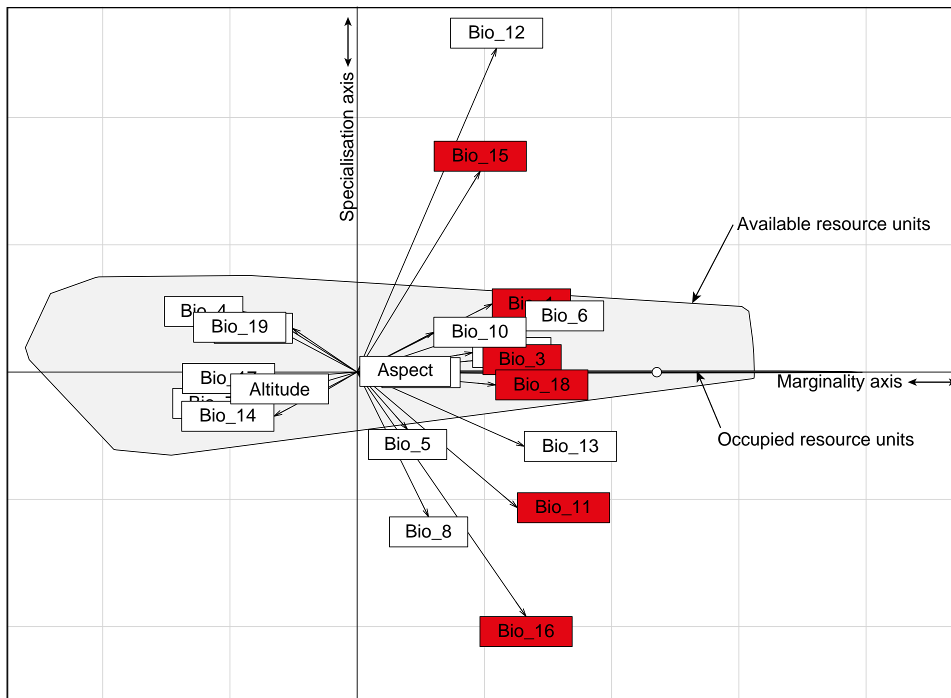


Figure 7. Biplot of the ENFA, showing the marginality axis (x-axis) and first specialisation axis (y-axis). The light grey polygon represents the distribution of available habitats, and the dark grey polygon (thin line on top of x-axis) represents the distribution of occupied habitats. This thinness in the y-axis of the occupied habitats again highlights the very narrow ecological niche of the Ghost Bat compared to that of the available habitats. The ecogeographic variables in red represent the six selected for further modelling using MaxEnt.

Maximum Entropy (MaxEnt)

Where species data has been systematically collected without sampling biases *e.g.,* via biological surveys), then regression methods are best for SDM, as absence data (true absence of a species from a region) is more likely to be known with some confidence. However, much of what we know about species occurrences is sourced from museum records that are subject to many biases (*e.g.,* collections from sites along the route of early explorers, or within easy access from roads). These museum data are also often sparse and limited in coverage, with little information about absence data. Specific SDM methods are therefore required that use the presence-only data of museums. One of the more rigorous SDM methods using presence-only data is maximum entropy (MaxEnt; Elith *et al.* 2006). In studies comparing SDM methods, the predictive performance of MaxEnt was shown to be highly robust compared to other methods (Elith *et al.* 2006). Maximum entropy methods have been used in spectral analysis for at least 20 years (Jaynes 1982), but have only recently been adapted for species distribution modelling (Phillips *et al.* 2006).

MaxEnt is an information-theoretic approach that uses the principles of maximum entropy to estimate a set of functions that relate environmental variables to suitable habitat, such that a model can be developed of a species' ecological niche and potential distributions under different

temporal and spatial scenarios. More detail on MaxEnt and its use in modelling the ecological niche of the Ghost Bat is found in chapter 4.

Statistical tests of significance

The domesticated chicken lineages that were transported to Europe prehistorically, and around the world during the ‘Age of Exploration’ appear to belong to Haplogroup E (Liu *et al.* 2006), with Haplogroup A also now present in Europe. Another domesticated lineage subsequently transported into the Pacific comprises Haplogroup D, which has an hypothesised homeland in India or Indonesia, and with a transportation route following the cultural activity of cock-fighting (including gamecocks from China and Japan) (Liu *et al.* 2006). Based on the different frequencies of these two haplogroups in different chicken populations, binomial probability distributions, binomial tests of statistical significance, and the Fisher’s exact test were all used in chapter 5, to propose that modern South American chickens had more similar haplotype patterns to European chickens than to modern and ancient chickens in the Pacific. Binomial distributions are based on a random process that has two possible outcomes (success or failure), with this discrete probability distribution representing the number of successes from n independent trials (with a constant probability p of success). The act of excavating chicken bones from the archaeological record is classed as n independent trials, with the particular haplotype/haplogroup of that chicken bone being success (*e.g.,* haplogroup E) or failure (*e.g.,* not haplogroup E), and the probability p of success is what is being tested. For example, what is the probability of finding no haplogroup E chickens (success = 0) out of 22 ancient chicken samples from the Pacific Islands ($n = 22$) if haplogroup E is actually present at a frequency of 55% ($p = 0.55$)? This is displayed as a density plot in the supplementary information of chapter 5. Another example is shown in Figure 8 (below), where the chicken haplogroup frequencies found in Chile, South America were found to be similar to those in Europe, rather than in either modern or ancient Pacific chickens.

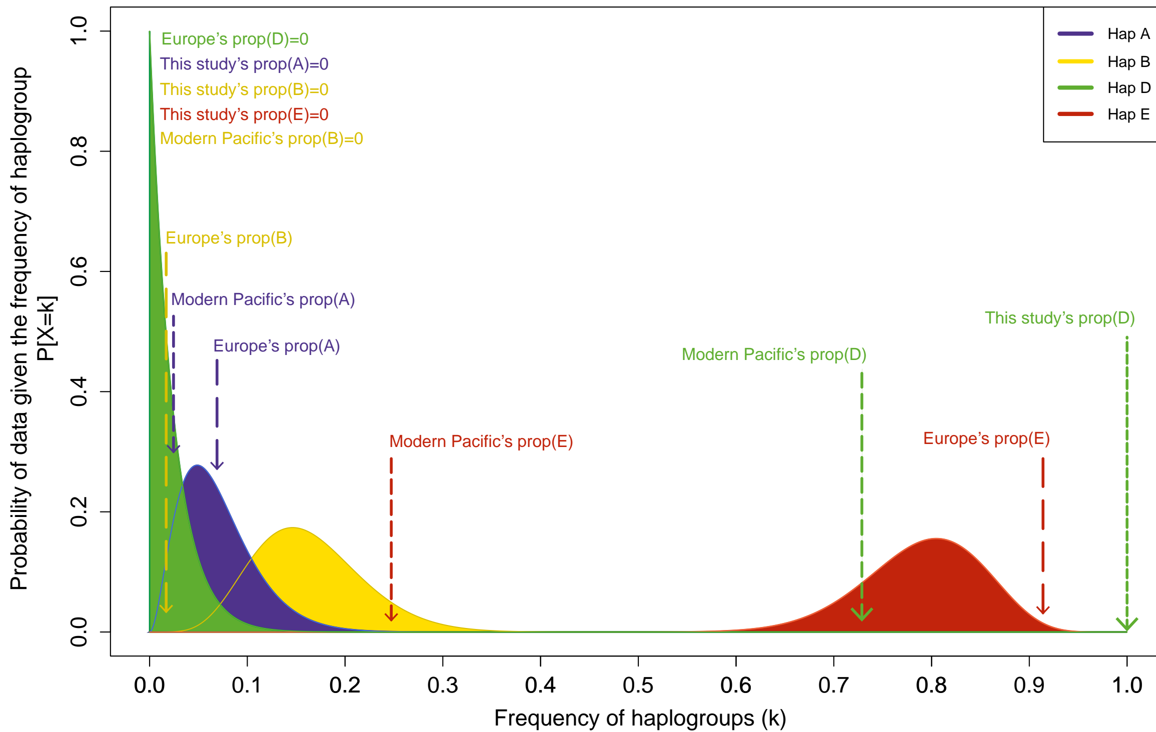


Figure 8. Binomial probability distribution showing the probability of each chicken haplogroup from modern European, modern Pacific, and ancient Pacific chicken populations matching those from modern chickens in Chile, South America. The binomial distributions reflect the proportions of each haplogroup in the Chilean chicken samples from (Gongora *et al.* 2008) with the arrows representing the other chicken populations. Haplogroup A is denoted in purple, haplogroup B is denoted in yellow, haplogroup D is denoted in green, and haplogroup E is denoted in red.

This binomial probability distribution is the basis of the binomial test of statistical significance. Given a number of successes x, a number of trials n, and a hypothesised probability of success p, what is the evidence against the null hypothesis (*e.g.,* a null hypothesis is that South American chickens have the same haplogroup E frequency as Pacific chickens). A similar test of significance is the Fisher's exact test, however this significance test can examine all haplogroup frequencies at once. For example, if haplogroups A, B, D, and E were all detected in chickens from South American and the Pacific, Fisher's exact test could assess whether overall the haplogroup frequencies were significantly different between these two locations. Both of these statistical tests were performed in R (R development core team 2011) using the binom.test and fisher.test command in the 'stats' package.

To my knowledge, no other aDNA studies have performed these types of tests of significance to explore similarities in populations for testing competing hypotheses about migration and source populations.

Conclusion

The methods described here have either been at the cutting edge of the field and incorporated aDNA in their analysis (BEAST, BSP, BayeSSC), or have complemented the insights that aDNA bring to investigating species' evolutionary history (ENM, statistical tests of significance). These methods are used extensively in the following three chapters: the Emu (chapter 3), the Ghost Bat (chapter 4), and the Polynesian Chicken (chapter 5), and will be further discussed in the Discussion (chapter 6).

References

Anderson CNK, Ramakrishnan U, Chan YL and Hadly EA (2004). Serial SimCoal: A population genetics model for data from multiple populations and points in time. *Bioinformatics* 21(8): 1733-1734.

Banguera-Hinestroza E, Bjorge A, Reid RJ, Jepson P and Hoelzel AR (2010). The influence of glacial epochs and habitat dependence on the diversity and phylogeography of a coastal dolphin species: *Lagenorhynchus albirostris*. *Conservation Genetics* 11(5): 1823-1836.

Basille M, Calenge C, Marboutin E, Anderson R and Gaillard J-M (2008). Assessing habitat selection using multivariate statistics: Some refinements of the ecological-niche factor analysis. *Ecological Modelling* 211: 233-240.

Baudinette RV, Churchill SK, Christian KA, Nelson JE and Hudson PJ (2000). Energy, water balance and the roost microenvironment in three Australian cave-dwelling bats (Microchiroptera). *Journal of Comparative Physiology B: Biochemical, Systemic, and Environmental Physiology* 170(5): 439-446.

Belle EMS, Ramakrishnan U, Mountain JL and Barbujani G (2006). Serial coalescent simulations suggest a weak genealogical relationship between Etruscans and modern Tuscans. *Proceedings of the National Academy of Sciences of the United States of America* 103(21): 8012-8017.

Burbidge ML, Colbourne RM, Robertson HA and Baker AJ (2003). Molecular and other biological evidence supports the recognition of at least three species of Brown Kiwi. *Conservation Genetics* 4: 167-177.

Campos PF, Craig OE, Turner-Walker G, Peacock E, Willerslev E and Gilbert MTP (2011). DNA in ancient bone - Where is it located and how should we extract it? *Annals of Anatomy* 194(1): 7-16.

Campos PF, Kristensen T, Orlando L, Sher A, Kholodova MV, Gotherstrom A, Hofreiter M, Drucker DG, Kosintsev P, Tikhonov A, Baryshnikov GF, Willerslev E and Gilbert MTP (2010). Ancient DNA sequences point to a large loss of mitochondrial genetic diversity in the saiga antelope (*Saiga tatarica*) since the Pleistocene. *Molecular Ecology* 19(22): 4863-4875.

Davies SJJF (2002). *Ratites and Tinamous: Tinamidae, Rheidae, Dromaiidae, Casuaiidae, Apterygidae, Struthionidae*. Oxford, Oxford University Press.

Depaulis F, Orlando L and Hanni C (2009). Using classical population genetics tools with heterochroneous data: time matters! *PLOS ONE* 4(5): e5541-e5541.

Drummond A and Rambaut A (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* 7(1): 214.

Drummond AJ, Pybus OG, Rambaut A, Forsberg R and Rodrigo AG (2003). Measurably evolving populations. *Trends in Ecology and Evolution* 18(9): 481-488.

Drummond AJ, Rambaut A, Shapiro B and Pybus OG (2005). Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution* 22(5): 1185-1192.

Elith J, Graham CH, Anderson RP, Dudík M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, Li J, Lohmann LG, Loiselle BA, Manion G, Moritz C, Nakamura M, Nakazawa Y, McC. M. Overton J, Townsend Peterson A, Phillips SJ, Richardson K, Scachetti-Pereira R, Schapire RE, Soberón J, Williams S, Wisz MS and Zimmermann NE (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29(2): 129-151.

Excoffier L (2004). Patterns of DNA sequence diversity and genetic structure after a range expansion: lessons from the infinite-island model. *Molecular Ecology* 13(4): 853-864.

Excoffier L, Foll M and Petit RJ (2009). Genetic consequences of range expansions. *Annual Review of Ecology Evolution and Systematics* 40: 481-501.

Excoffier L and Lischer H (2010). Arlequin suite ver3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* 10: 564-567.

Excoffier L, Novembre J and Schneider S (2000). SIMCOAL: A general coalescent program for the simulation of molecular data in interconnected populations with arbitrary demography. *Journal of Heredity* 91(6): 506-509.

Excoffier L, Smouse PE and Quattro JM (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes - application to human mitochondrial DNA restriction data. *Genetics* 131(2): 479-491.

Gerber AS, Loggins R, Kumar S and Dowling TE (2001). Does nonneutral evolution shape observed patterns of DNA variation in animal mitochondrial genomes? *Annual Review of Genetics* 35: 539-566.

Gongora J, Rawlence NJ, Mobegi VA, Jianlin H, Alcalde JA, Matus JT, Hanotte O, Moran C, Austin JJ, Ulm S, Anderson AJ, Larson G and Cooper A (2008). Indo-European and Asian origins for Chilean and Pacific chickens revealed by mtDNA. *Proceedings of the National Academy of Sciences of the United States of America* 105(30): 10308-10313.

Haak W, Balanovsky O, Sanchez JJ, Koshel S, Zaporozhchenko V, Adler CJ, Der Sarkissian CSI, Brandt G, Schwarz C, Nicklisch N, Dresely V, Fritsch B, Balanovska E, Villems R, Meller H, Alt KW, Cooper A and Members of the Genographic C (2010). Ancient DNA from European early neolithic farmers reveals their near eastern affinities. *PLoS Biology* 8(11): e1000536.

Harpending H and Rogers A (2000). Genetic perspectives on human origins and differentiation. *Annual Review of Genomics and Human Genetics* 1: 361-385.

Heled J and Drummond AJ (2010). Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution* 27(3): 570-580.

Hirzel AH, Hausser J, Chessel D and Perrin N (2002). Ecological-niche factor analysis: How to compute habitat-suitability maps without absence data? *Ecology* 83(7): 2027-2036.

Ho SY, Phillips MJ, Cooper A and Drummond AJ (2005). Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Molecular Biology and Evolution* 22(7): 1561-1568.

Ho SYW, Lanfear R, Bromham L, Phillips MJ, Soubrier J, Rodrigo AG and Cooper A (2011a). Time-dependent rates of molecular evolution. *Molecular Ecology*.

Ho SYW, Lanfear R, Phillips MJ, Barnes I, Thomas JA, Kolokotronis SO and Shapiro B (2011b). Bayesian estimation of substitution rates from ancient DNA sequences with low information content. *Systematic Biology* 60(3): 366-374.

Hoffman JI, Grant SM, Forcada J and Phillips CD (2011). Bayesian inference of a historical bottleneck in a heavily exploited marine mammal. *Molecular Ecology* 20(19): 3989-4008.

Hoss M and Paabo S (1993). DNA extraction from Pleistocene bones by a silica-based purification method. *Nucleic Acids Research* 21(16): 3913-3914.

Hudson RR, Slatkin M and Maddison WP (1992). Estimation of levels of gene flow from DNA sequence data. *Genetics* 132(2): 583-589.

Jaynes ET (1982). On the rationale of Maximum-Entropy methods. *Proceedings of the IEEE* 70(9): 939-952.

Jones K. (2008). "Husbandry Manual for Ghost bat *Macroderma gigas* Mammalia: Megadermatidae." Western Sydney Institute of TAFE Retrieved 09 December 2011, 2011, from http://www.australianzookeeping.org/Husbandry_Manuals.htm - Order_Chiroptera_(Bats)_.

Kingman JFC (2000). Origins of the coalescent: 1974-1982. *Genetics* 156(4): 1461-1463.

Kumar S, Bellis C, Zlojutro M, Melton PE, Blangero J and Curran JE (2011). Large scale mitochondrial sequencing in Mexican Americans suggests a reappraisal of Native American origins. *BMC Evolutionary Biology* 11.

Li WH (1977). Distribution of nucleotide differences between 2 randomly chosen cistrons in a finite population. *Genetics* 85(2): 331-337.

Lippold S, Matzke NJ, Reissmann M and Hofreiter M (2011). Whole mitochondrial genome sequencing of domestic horses reveals incorporation of extensive wild horse diversity during domestication. *BMC Evolutionary Biology* 11.

Liu YP, Wu GS, Yao YG, Miao YW, Luikart G, Baig M, Beja-Pereira A, Ding ZL, Palanichamy MG and Zhang YP (2006). Multiple maternal origins of chickens: out of the Asian jungles. *Molecular Phylogenetics and Evolution* 38(1): 12-19.

O'Keefe K, Ramakrishnan U, Van Tuinen M and Hadly Ea (2009). Source-sink dynamics structure a common montane mammal. *Molecular Ecology* 18(23): 4775-4789.

Oskam CL, Haile J, McLay E, Rigby P, Allentoft ME, Olsen ME, Bengtsson C, Miller GH, Schwenninger JL, Jacomb C, Walter R, Baynes A, Dortch J, Parker-Pearson M, Gilbert MTP, Holdaway RN, Willerslev E and Bunce M (2010). Fossil avian eggshell preserves ancient DNA. *Proceedings of the Royal Society of London Series B-Biological Sciences* 277(1690): 1991-2000.

Paabo S (1989). Ancient DNA: Extraction, characterization, molecular cloning, and enzymatic amplification. *Proceedings of the National Academy of Sciences of the United States of America* 86: 1939-1943.

Paabo S, Gifford JA and Wilson AC (1988). Mitochondrial sequences from a 7,000-year-old brain. *Nucleic Acids Research* 16: 9775-9787.

Paplinska JZ, Taggart DA, Corrigan T, Eldridge MDB and Austin JJ (2011). Using DNA from museum specimens to preserve the integrity of evolutionarily significant unit boundaries in threatened species. *Biological Conservation* 144(1): 290-297.

Phillips SJ, Anderson RP and Schapire RE (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190: 231-259.

R development core team (2011). R: A language and environment for statistical computing. R foundation for statistical computing

Ray N, Currat M and Excoffier L (2003). Intra-deme molecular diversity in spatially expanding populations. *Molecular Biology and Evolution* 20(1): 76-86.

Raymond M and Rousset F (1995). An exact test for population differentiation. *Evolution* 49(6): 1280-1283.

Rogers AR and Harpending H (1992). Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution* 9(3): 552-569.

Rohland N and Hofreiter M (2007a). Ancient DNA extraction from bones and teeth. *Nature Protocols* 2(7): 1756-1762.

Rohland N and Hofreiter M (2007b). Comparison and optimization of ancient DNA extraction. *Biotechniques* 42(3): 343-352.

Rushton SP, Ormerod SJ and Kerby G (2004). New paradigms for modelling species distributions? *Journal of Applied Ecology* 41(2): 193-200.

Schenekar T and Weiss S (2011). High rate of calculation errors in mismatch distribution analysis results in numerous false inferences of biological importance. *Heredity* 107(6): 511-512.

Schneider S and Excoffier L (1999). Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates very among sites: Application to human mitochondrial DNA. *Genetics* 152(3): 1079-1089.

Shapiro B, Drummond AJ, Rambaut A, Wilson MC, Matheus PE, Sher AV, Pybus OG, Gilbert MTP, Barnes I, Binladen J, Willerslev E, Hansen AJ, Baryshnikov GF, Burns JA, Davydov S, Driver JC, Froese DG, Harington CR, Keddie G, Kosintsev P, Kunz ML, Martin LD, Stephenson RO, Storer J, Tedford R, Zimov S and Cooper A (2004). Rise and fall of the Beringian Steppe Bison. *Science* 306(5701): 1561-1565.

Sibon Li WL and Drummond AJ (2011). Model averaging and Bayes factor calculation of relaxed molecular clocks in Bayesian phylogenetics. *Molecular Biology and Evolution*.

Slatkin M (1994). Linkage disequilibrium in growing and stable populations. *Genetics* 137(1): 331-336.

Slatkin M (1995). A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139(1): 457-462.

Slatkin M and Hudson RR (1991). Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* 129(2): 555-562.

Tajima F (1996). Infinite-allele model and infinite-site model in population genetics. *Journal of Genetics* 75(1): 27-31.

Valdiosera CE, Garcia N, Anderung C, Dalen L, Cregut-Bonnoure E, Kahlke RD, Stiller M, Brandstrom M, Thomas MG, Arsuaga JL, Gotherstrom A and Barnes I (2007). Staying out in the cold: glacial refugia and mitochondrial DNA phylogeography in ancient European brown bears. *Molecular Ecology* 16(24): 5140-5148.

Valdiosera CE, Garcia-Garitagoitia JL, Garcia N, Doadrio I, Thomas MG, Hanni C, Arsuaga JL, Barnes I, Hofreiter M, Orlando L and Gotherstorm A (2008). Surprising migration and population size dynamics in ancient Iberian Brown Bears (*Ursus arctos*). *Proceedings of the National Academy of Sciences of the United States of America* 105(13): 5123-5128.

Varela S, Lobo JM and Hortal J (2011). Using species distribution models in paleobiogeography: A matter of data, predictors and concepts. *Palaeogeography, Palaeoclimatology, Palaeoecology* 310: 451-463.

Wakeley J (1999). Nonequilibrium migration in human history. *Genetics* 153(4): 1863-1871.

Wakeley J (2001). The coalescent in an island model of population subdivision with variation among demes. *Theoretical Population Biology* 59(2): 133-144.

Warren DL and Seifert SN (2011). Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. *Ecological Applications* 21(2): 335-342.

Watterson G (1975). On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology* 7(2): 256-276.

Watterson G (1978). The homozygosity test of neutrality. *Genetics* 88(2): 405-417.

Wilson IG (1997). Inhibition and facilitation of nucleic acid amplification. *Applied and Environmental Microbiology* 63(10): 3741-3751.

Wright S (1951). The genetical structure of populations. *Annals of Eugenics* 15(4): 323-354.

Wright S (1965). The interpretation of population structure by F statistics with special regard to systems of mating. *Evolution* 19(3): 395-420.

Yang DY, Eng B, Waye JS, Dudar JC and Saunders SR (1998). Technical note: Improved DNA extraction from ancient bones using silica-based spin columns. *American Journal of Physical Anthropology* 105(4): 539-543.

Chapter 3: The Emu

Postglacial expansions from a south-western refugium: reconstructing the recent evolutionary history of Emus (Aves: *Dromaius*) using ancient DNA

STATEMENT OF AUTHORSHIP

Postglacial expansions from a south-western refugium: reconstructing the recent evolutionary history of Emus (Aves: *Dromaius*) using ancient DNA

Text in manuscript

THOMSON, V.A. (Candidate)

Performed lab work on all samples, performed all analysis, interpreted the data, created figures, and wrote manuscript.

I hereby certify that the statement of contribution is accurate.

Signed. ....*Date* 10/5/2012

COOPER, A.

Supervised development of work, helped in data interpretation and manuscript writing.

I hereby certify that the statement of contribution is accurate and I give permission for the inclusion of the manuscript in the thesis.

Signed. ....*Date* 10/5/2012

Postglacial expansions from a south-western refugium: reconstructing the recent evolutionary history of Emus (Aves: *Dromaius*) using ancient DNA

Authors:

VA Thomson, A Cooper.

Abstract:

The Emu (*Dromaius novaehollandiae novaehollandiae*) is an iconic Australian taxon, yet little is known about its recent evolutionary history. We use mitochondrial DNA from over a hundred specimens to investigate the evolutionary history of *Dromaius*, including almost 30 ancient samples of two dwarf putative species (from King Island and Kangaroo Island), an island subspecies from Tasmania (of indeterminate size), and Pleistocene eggshell from mainland individuals. We find the three island emu populations show no genetic divergence from mainland emus, but that the extent of dwarfing is correlated with island size and period of isolation, and interpret these populations as isolated remnants of mainland emu diversity. Within mainland populations we find evidence that *D. n. novaehollandiae* became restricted to the southwest of the continent prior to the Middle Pleistocene. A major postglacial expansion out of this southwest refugium occurred just prior to the start of Marine Isotope Stage (MIS) 9 (~ 350 kya), probably in response to interglacial conditions that enabled *D. n. novaehollandiae* to cross the Nullarbor Plain (previously a barrier to gene flow), and colonise the eastern half of the continent. A secondary (minor) postglacial expansion occurred at MIS 5 (~ 100 kya) within the western population, coincident with the last interglacial.

Introduction

Climatic and environmental changes during the Late Quaternary have been key determinants of the diversity and distribution of present day plants and animals. While patterns in the Northern Hemisphere are well known, most notably the Arctic, temperate Europe, and Beringia (Muller and MacDonald 1997; Taberlet *et al.* 1998; Hewitt 2000; Barnes *et al.* 2002; Hadly *et al.* 2004; Shapiro *et al.* 2004; Bhagwat and Willis 2008; Debruyne *et al.* 2008; Aubry *et al.* 2009; Clark *et al.* 2009; Campos *et al.* 2010; Shafer *et al.* 2010) the effects of climatic cycles on ecosystems in the southern Hemisphere are comparatively poorly understood. This is especially true in Australia, where actual glaciation was restricted in extent and climatic cycles were expressed mainly as changing levels of aridity (Barrows *et al.* 2002). The cyclical desertification of Australia led to repeated movement and reassembling of ecosystems during the Quaternary, which seems to have created at least two distinct categories of phylogeographic patterning in the modern biota of Australia (Byrne 2008): (1) species with deep phylogeographic structuring; and (2) species that exhibit low genetic diversity and little phylogeographic structure. The reconstruction of evolutionary histories for species with deep structuring is relatively straightforward, *e.g.,* barriers to gene flow include the Canning basin in geckos (Pepper *et al.* 2011a; Pepper *et al.* 2011b); the Nullarbor Plain in the Brush tailed Phascogale (Spencer *et al.* 2001); the Carpentarian gap in finches (Cracraft 1986); the Burdekin gap in Bowerbirds (Nicholls and Austin 2005) and frogs (James and Moritz 2000); the Eyrean barrier in the Australian Ringneck and the Splendid Fairy-wren (Ford 1987a). These barriers to gene flow act to varying degrees on different species and are usually due to disjunctions in habitat that prevent dispersal at the regional scale. However until recently, the genetic profiles of species lacking phylogeogeographic structure have provided little insight into their evolutionary histories and responses to the changing climates and environments of the Quaternary.

Many of the taxa which exhibit a lack of phylogeographic structure are highly vagile and/or relatively widespread: including birds, *e.g.,* ducks, parrots, cuckoos, and a range of passerines (Joseph *et al.* 2002; Joseph and Wilke 2006; Joseph *et al.* 2006; Toon *et al.* 2007; Kearns *et al.* 2009; Guay *et al.* 2010; Kearns *et al.* 2010); reptiles, amphibians, and plants, *e.g.,* snakes, frogs, and sandalwood (Byrne *et al.* 2000; Kuch *et al.* 2005; Burns *et al.* 2007); and many of Australia's unique and iconic marsupials, *e.g.,* Koalas, Eastern Grey Kangaroos, and the Yellow-bellied Glider (Houlden *et al.* 1999; Zenger *et al.* 2003; Brown *et al.* 2006). The widespread and mobile nature of these taxa complicates the reconstruction of demographic histories from modern data alone, as few barriers have restricted dispersal. The addition of ancient DNA (aDNA) to these phylogeographic studies (termed phylochronology; Hadly *et al.* 2004), can provide direct insight into genetic distribution patterns prior to recent contact events or dispersals.

Unfortunately, the extreme climatic conditions of the Australia environment are not well suited to aDNA survival. Hot and arid conditions interrupted by episodic and sometimes extreme rainfall events are not likely to be conducive to DNA preservation, as generally high temperatures and water degrade DNA (Bollongino *et al.* 2008). However, eggshell has recently been shown to preserve DNA under adverse conditions (Oskam *et al.* 2011). In contrast to bone, where DNA is bound to hydroxyapatite in a calcium matrix, eggshell is a calcite biomineral of exceptional preservation potential, where DNA can be stable for millennia (Oskam *et al.* 2010). Eggshell is preserved in many Australian palaeontological cave sites and a large proportion is derived from the Emu (*Dromaius novaehollandiae novaehollandiae*) (Miller *et al.* 1999; Miller *et al.* 2005). The emu is a large flightless ratite bird which remains widespread in Australia but has two recently extinct, dwarf relatives on several islands to the south of the continent. Modern emu populations are highly vagile and likely to show limited phylogeographic structure. The availability of palaeontological eggshell samples therefore provides a rare opportunity to assess the phylogeographic history of Australian taxa using a combination of modern and ancient samples. Furthermore, the isolation of emu populations on continental islands by post-glacial sea level rise provides an opportunity to calibrate the genetic divergences between island and mainland populations.

The Emu belongs to one of two extant lineages of large flightless ratite birds in Australia (the other being the Cassowary, *Casuarius spp.*, inhabitants of tropical rainforests). Four recent taxa are usually distinguished: the extant mainland Emu, *D. n. novaehollandiae*; and three recently extinct island forms: the King Island Emu (*D. ater*), the Kangaroo Island Emu (*D. baudinianus*) which were both distinguished by their short stature, osteological differences in the limb bones and plumage colouration (Vieillot 1817; Morgan and Sutton 1928; Parker 1984); and the Tasmanian Emu (*Dromaius novaehollandiae diemenensis*), which was close or identical in size to mainland populations, although there are conflicting accounts of its size (Dove 1924; Dove 1926a; Dove 1926b). Gene flow between these island taxa and the mainland Emu (*D. n. novaehollandiae*) has been halted for many thousands of years, as Tasmania was isolated from the mainland 14,000 years ago (ya), King Island was isolated from Tasmania 12,000 ya, and Kangaroo Island was isolated from the South Australian mainland 8,800 ya (Hope *et al.* 1977; Lambeck and Chappell 2001). Emu fossils from Late Pleistocene sites on the mainland include a few diminutive specimens (most notably, those from Lake Menindee reported by Patterson and Rich, 1987) and these remain taxonomically unallocated.

Based on the palaeontological data there are at least two hypotheses that might explain the presence of dwarf forms of Emu restricted to offshore islands and a larger form of the Emu on the Australian mainland. Hypothesis one: the island taxa are remnants of formerly widespread dwarfed species of the Pleistocene, that have since become extinct on the mainland (reflecting

phyletic nanism; Gould and MacFadden 2004). If this hypothesis is correct, we would expect to see evidence of a deep split between the dwarfed and large species. Hypothesis two: the island forms may represent populations of *D. n. novaehollandiae* that became dwarfed after isolation from mainland gene flow through phenotypic plasticity or genetic adaptation (reflecting autapomorphic nanism; Gould and MacFadden 2004). If hypothesis two is correct, we would expect evidence that the dwarfed and large forms are simply morphological variants of the same species. For completeness, note that a third hypothesis does exist: the mainland form may stem from an ancestral dwarf form that has since expanded in body size to fill the larger niches on the mainland, whilst the ancestral dwarf body size has been preserved on the offshore islands. The deep fossil history of the large bodied *D. n. novaehollandiae* on the mainland since the Late Pliocene suggests this is unlikely (Patterson and Rich 1987). Hypothesis two is termed island dwarfism and is potentially the most likely explanation.

The phenomenon of island dwarfism in large taxa is common worldwide, with density-dependent factors (*i.e.,* resource limited systems) and/or relaxed predator pressures creating conditions that favour smaller individuals (Foster 1964; Lomolino 1985; Vartanyan *et al.* 1993; Fritz *et al.* 2005; McNab 2010). Where selection pressures are relaxed, the major variables affecting size evolution in island populations of non-volant taxa include island size and length of isolation (Anderson and Handley 2002). Island size is expected to influence optimal body size due to differences in intra- and inter-specific competitive interactions for resources, while length of isolation (island age) should be inversely proportional to body size, until such time as equilibrium is reached (Anderson and Handley 2002). However, dwarfism has also been hypothesised to occur in response to climatically induced resource limitations – 'temporal dwarfism' rather than 'island dwarfism' – such as proposed for the modern koala (Price 2008) and agile wallaby (Marshall and Corruccini 1978).

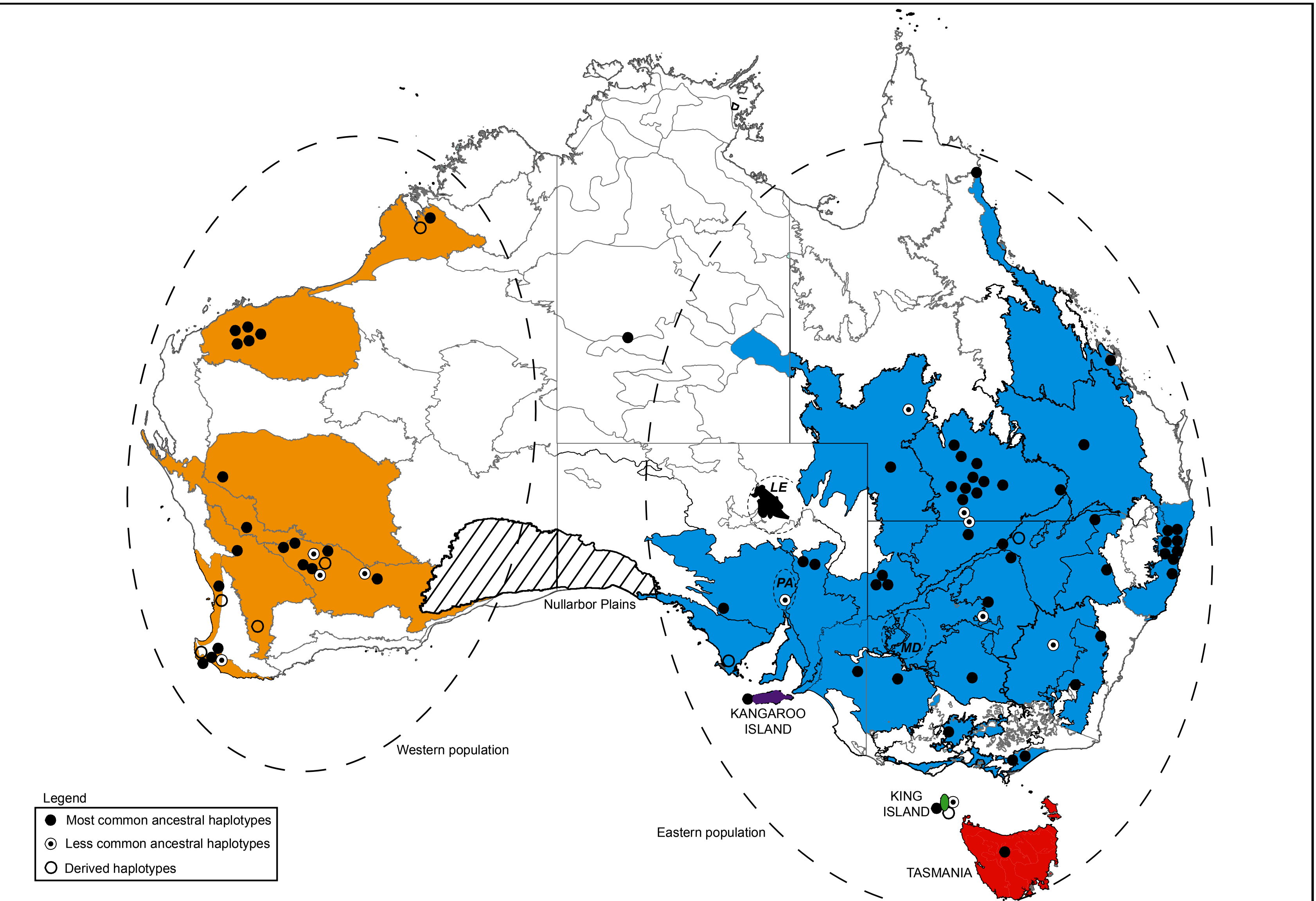
The complex nature of these factors and their relative interactions in causing dwarfism leads to uncertainty surrounding the evolutionary history of *D. ater* and *D. baudinianus.* Although all island *Dromaius* were driven extinct by human hunting in the 19th century, limb bones of *D. ater* and *D. baudinianus* have been preserved in the palaeontological record on these southern temperate islands, making it possible to test these alternative hypotheses using ancient DNA (aDNA) methods. Furthermore, *D. ater* was isolated from *D. n. novaehollandiae* for approximately 5,200 years longer than *D. baudinianus*, so that morphometric comparisons of the limb bones of living and extinct emus may provide information about the process of evolutionary divergence in size and shape, and also help to resolve issues around taxonomic status. Such morphometric comparisons include tests for allometric versus non-allometric scaling. Allometric scaling within organisms deals with the physiological shape changes required to maintain body function in different sized individuals of the same or phylogenetically similar species, however

organisms can also show shape-uncorrelated size differences (non-allometric scaling) if the different body part performs different physiological functions, or performs the physiological functions in a different way, which is likely in deeply divergent species. If island taxa are truly remnants of a formerly widespread dwarfed species that has become extinct on the mainland (Hypothesis 1), then we would expect relatively high genetic divergence and potentially non-allometric scaling of limb bones. However if the island taxa are populations of *D. n. novaehollandiae* that have became dwarfed through insularization (Hypothesis 2), then we would expect the island forms to contain a subset of the genetic diversity found on the mainland and to reflect allometric scaling.

In order to test these two hypotheses, a survey of the current genetic diversity of *D. n. novaehollandiae* is required, as no phylogeographic studies have been performed on the mainland Emu to investigate its recent evolutionary history. Another key question of interest is how the Emu survived the environmental extremes of the late Quaternary, in particular the period around 50,000 years ago when many Australian megafauna, including another large flightless bird (*Genyornis newtoni*), became extinct (Miller *et al.* 1999; Brook *et al.* 2007; Faith and O'Connell 2011). Although, as a nomadic omnivore, the Emu is constantly on the move following the intermittently abundant food resources that respond to ephemeral rainfall events, such as seeds, flowers, young sprouting annuals and perennials, as well as insects (Davies 2002), uncertainty exists about how Emus deal with highly arid periods. Important clues may be found by examining how other, highly mobile Australian species responded to past climatic cycles. For example, even widespread and/or highly mobile species appear to show signs of historic subdivision by the Nullarbor Plain, an extensive barren expanse of chenopod shrub steppe on the southern coast of Australia which forms a significant barrier to gene flow (Figure 1; Crisp and Cook 2007; Toon *et al.* 2007; Neaves *et al.* 2009; Salinas *et al.* 2009). However, examination of the Pleistocene record of caves in the Nullarbor Plain has found that an extensive assemblage of woodland fauna inhabited the Nullarbor region from approximately 200-400 kya (Prideaux *et al.* 2007), so this arid steppe has not always prevented habitation in the region, and in fact many species may have simply skirted the northern edge of the Nullarbor Plain. Within the arid zone, the ephemeral nature of the inland arid and semi-arid lakes has played a large role in sustaining gene flow and/or creating divergent populations (*e.g.* Lake Eyre, Murray-Darling lakes, Lake Amadeus *etc.*). The inland drainage basins are often flat, with a patchwork of meandering rivers that open up to many shallow yet ephemeral waterholes, and are surrounded by large sparsely vegetated swampy regions. The Eyrean Barrier, an arid barrier starting at the top of the Spencer Gulf and running through Lake Eyre basin and Simpson Desert, has also caused vicariance to develop between many bird species or subspecies pairs (Ford 1974; Ford 1987b; 1987a). In addition, areas that putatively acted as major refugia from the arid centre in Australian species have been identified in

the continent's more temperate south-eastern (Victoria/Tasmania) and south-western corners (Western Australia). The southwest, in particular, is characterised by much higher (10-fold) plant biodiversity levels than south-eastern Australia (Dodson 2001). All three of these regions (the Nullarbor/Eyrean barrier, and the south-east and south-western corners) are potentially important for understanding the recent history of the Emu. Previous workers have speculated that *D. n. novaehollandiae* from western Australia might belong to a separate subspecies based on differences in plumage colouration (Le Souef 1904) and others have suggested a long evolutionary history for the Emu in south-western Australia, as it is one of only a few Australian species that can tolerate fluoroacetate-bearing plants of the genera *Gastrolobium* and *Oxylobium*, which are confined to southwestern Australia (King *et al.* 1978; McIlroy 1984; Twigg *et al.* 1988).

We undertook a combined phylogenetic and phylogeographic study of Emus (*Dromaius* spp.) across their range on the Australian continent and offshore islands. In particular, we aimed to determine the evolutionary history and status of the putative dwarf species *D. ater* and *D. baudinianus*, and of the Tasmanian 'subspecies' *D. n. diemenensis*. We examined both the genetic divergence of *D. ater, D. baudinianus, D. n. diemenensis,* from *D. n. novaehollandiae*, and the relationship between shape and size variation in limb bones of *D. ater, D. baudinianus,* and *D. n. novaehollandiae*. Our study also provides substantial geographic coverage of *D. n. novaehollandiae* from eastern and western Australia for an in-depth phylogeographic study, and includes aDNA samples dating up to approximately 16,500 ya, allowing us to add a temporal component to the study of the evolutionary history of mainland Emus. We hypothesise that like other widespread or highly mobile species, *D. n. novaehollandiae* populations were affected by Australia's major recent barrier to gene flow, the Nullarbor Plain, and took advantage of one of Australia's major Pleistocene refugia, the south-west corner.



LE
PA
MD
Nullarbor Plains
KANGAROO ISLAND
KING ISLAND
TASMANIA
Western population
Eastern population
Legend
Most common ancestral haplotypes
Less common ancestral haplotypes
Derived haplotypes

Figure 1. Map showing Emu sampling locations as circles: black circles represent Emu samples with most common ancestral mitochondrial haplotypes (H2, H3, H7), white circle with dots represent Emu samples with less common ancestral haplotypes (H1, H5, H6, H11, H12, H18), and unfilled circles represent Emu samples with one or more variable sites from the ancestral haplotypes (H4, H5, H9, H10, H13, H14, H16, H17). Inland arid and semi-arid lakes that supported Emu populations during the Pleistocene are outlined by small dashed circles: Lake Eyre (LE), Port Augusta (PA), and Murray-Darling (MD); large dashed ellipses indicate the western and eastern populations; hashed bioregion indicates the Nullarbor Plain. Black lines outline biogeographic regions according to Interim Biogeographic Regionalisation for Australia (IBRA), which describe large geographically distinct areas with similar vegetation, geology, and climate. Colours represent populations: western (orange), eastern (blue), Tasmania (red), Kangaroo Island (purple), King Island (green).

Materials and Methods

Samples.

A total of 24 ancient Emu bone and eggshell samples were obtained from museums and collaborators (Table S5). These ancient samples included specimens of *Dromaius baudinianus* (Kelly Hill Caves, Kangaroo Island), *D. ater* (King Island), and *D. n. novaehollandiae* (Tunnel Cave, WA). We also obtained 19 non-invasive opportunistically collected samples from across Australia, including 1 scat, and 18 modern feather samples from emus killed by vehicles; and 67 museum specimens, including 10 historically collected bone samples, 21 toepads, 5 feathers, 2 skin or mummified tissue specimens, and 29 modern tissue samples (Table S5). These 110 newly sampled individuals were the basis of the dataset for this study, however as seven of the museum samples lacked detailed location information they were excluded from the analyses. In addition the two *D. ater* samples from Heupink *et al.* (2011) were included for completeness. Note due to a lack of sampling between the Nullarbor Plain and the Eyrean Barrier the effects of these two common barriers could not be distinguished.

aDNA Extraction, Amplification, and Sequencing.

Samples were extracted, amplified, and sequenced in specialist aDNA laboratories at the Australian Centre for Ancient DNA (ACAD) in Adelaide, South Australia, according to a range of strict protocols, including numerous controls (Cooper and Poinar 2000). A modified Qiagen silica extraction protocol was used to extract DNA (Supplementary Materials and Methods). Eighteen randomly selected modern individuals spanning the geographic spread of the continent were sequenced for both hypervariable regions (HVSI and HVSII) of the mtDNA *CR* to identify variable sites. When these initial 18 individuals indicated low levels of genetic variation (6 SNP's in 1000bp), 100-200bp regions surrounding the parsimony informative SNP's were sequenced in the balance of the samples while attempting to cover both hypervariable regions (350bp of non-contiguous sequence was concatenated together). Primers were combined into two multiplexes: Multiplex A, comprising primers A1360/A1407 (174bp), A1414/A1361 (105bp), A1691/A1692 (149bp), A1695/A1363 (143bp); and multiplex B, comprising primers A1410/A1411 (175bp), A1539/A1440 (135bp), A1693/A1694 (147bp) (Table S6).

Morphometric analyses.

A morphometric analysis of femur, tibia and tarso-metatarsi bones was undertaken for *D. ater, D. baudinianus*, and *D. n. novaehollandiae*, although the tibia dataset was excluded from further analyses due to measurement error. The size and shape of the two leg bones as overall length (L), proximal width (PW), proximal depth (PD), distal width (DW), distal depth (DD), shaft width (SW), and shaft depth (SD) were measured. These raw measurements as well as shape indices and log shape variables calculated from them have all been shown to vary significantly between ratite species according to latitude (Worthy 1994). To explore these data, plots of the

measurements against each other were prepared using R and 'plotmatrix' in ggplot2 package. Univariate analysis of variance (ANOVA) of the pairwise comparisons of measurements between taxa was assessed using Bonferroni's correction with $\alpha = 0.05$ to establish whether *D. ater* and *D. baudinianus* were significantly different in size. Multivariate analyses such as principal components analyses (PCA) and regression analyses of PC1 versus island size and time since isolation were done on the 'log shape' variables to indicate the overall relationships between the standard-sized and dwarfed Emu. A Pearson correlation analysis of the size-adjusted data (as per Gilbert and Rossie 2007) was done to test for a significant influence of allometry on the Emu taxa. Additionally, two new tests of allometry were performed on the log shape variables: PCA and Allometry ratio spectra (as per Baur and Leuenberger 2011), which were run using the R script from Baur and Leuenberger (2011).

Population Genetic Analyses.

Population diversity indices (No. of haplotypes, No. of segregating sites, haplotype diversity, nucleotide diversity, Tajima's D statistic, and F_{ST} statistics) were estimated using Arlequin v3.5.1.2 (Excoffier and Lischer 2010). The panmixia assumption was tested with an Allelic Aggregation Index Analysis (AAIA) using Alleles in Space (AIS) v1.0 (Miller 2005). AIS assesses the correlation between genetic distance, which it calculates using simple mismatch distances calculated from raw genetic data and geographic distance, using a Delaunay triangulation-based connectivity network that it generates from the latitude and longitude co-ordinates for the dataset. The dataset was also analysed using Genetic Landscape Shape in AIS in order to visually identify genetic discontinuities or landscape regions where relatively low or high genetic distances occur. In order to formally characterise the patterns observed in the Genetic Landscape Shape analysis, we performed post hoc linear regression analyses using longitude as the independent variable and various genetic diversity variables in Ostats v4.0.2 (http://www.maths.uq.edu.au/~mrb/ostats/index.php).

Phylogenetic and phylogeographic analyses.

An unrooted haplotype network was also constructed using the Median Joining network in the software Network v4.6 (with gaps excluded; Bandelt *et al.* 1999), as initial data visualisation in SplitsTree v4 using the NeighborNet algorithm (Huson and Bryant 2006) indicated the Emu dataset was not treelike. The evolutionary rate was estimated using the Bayesian algorithm in BEAST v1.6.1 (Drummond and Rambaut 2007) using the HKY model, which was identified as the best substitution model in ModelGenerator v0.85 (Keane *et al.* 2006). In order to test whether our dataset had sufficient phylogenetic and temporal signal from tip calibration ages to estimate the evolutionary rate with any accuracy, 10 randomisation experiments were conducted using BEAST runs with identical priors as the optimal BEAST runs for our real dataset. Furthermore, a binomial test of significance was performed to establish how likely it is to detect three of the five

ancient eggshell samples as the more common haplotypes in modern populations (H3 and H7) if the frequency of all current day haplotypes were equal in the ancient populations of 12,000-16,500 ya (*e.g.,* how likely is it that H3 and H7 have been common for a long time?).

Population demographic structure.

In order to investigate the demographic history of *Dromaius*, mismatch distributions of the modern and ancient eastern populations were generated using Arlequin v3.5 (Excoffier and Lischer 2010) to test for evidence of demographic expansion (Rogers and Harpending 1992) or spatial expansion (Excoffier 2004). However, the ancient western samples were excluded from the expansion tests as the sample size was too low (n=5). The demographic history of *Dromaius* was also inferred using the Bayesian coalescent approach as implemented in BEAST using the Bayesian Skyline Plot (BSP). Initially, panmixia was assumed for the entire dataset with the timing of coalescent events estimated by carbon dates of associated layers for *D. n. novaehollandiae* eggshell samples and the estimated island separation age for each island taxon (Kangaroo Island separated from the South Australian mainland 8,800 ya; Hope *et al.* 1977; King Island separated from Tasmania 12,000 ya and Tasmania separated from the Victorian mainland 14,000 ya; Lambeck and Chappell 2001). As BEAST is limited to reconstructing demographic histories of single populations under relatively simplistic demographic processes, we also modelled increasingly more complex scenarios using Bayesian Serial Simcoal (BayeSSC; Anderson et al. 2004) against which the observed data was compared. Four different hypotheses were tested (see Figure 4 and Table 1) where effective population size (N_E), growth rate (r), date of events (T_E), and migration rate (m) were estimated. By performing two iterative rounds of BayeSSC, the dataset can be used to refine the more likely effective population sizes, growth rates, and date of the events, while different migration rates were tested in increments of 5% per-generation in both directions, as well as in only one direction, eastwards or westwards (see Supplementary Methods for more details).

Results

We were able to successfully extract, amplify and sequence 350 bp of mitochondrial *control region* DNA (mtDNA *CR*) for 103 of 185 Emu samples that had detailed location information. These samples encompass 80 *D. n. novaehollandiae* (including 5 late Pleistocene eggshell samples), 8 ancient *D. ater*, 11 ancient *D. baudinianus*, and 4 ancient *D. n. diemenensis* specimens (Figure 2), which we combined with 2 additional King Island sequences from Heupink *et al.* (2011). Unfortunately, no suitable samples were available from Lake Menindee to directly test the identity of possible mainland ‘dwarf’ emus in the Late Pleistocene. Of the initial 1,000 bp analysed for a geographically widespread subset of the data (18 individuals), only 6 nucleotide

sites were found to be variable (see Materials and Methods), and when the full dataset (105 samples) was sequenced for the regions surrounding the parsimony informative SNPs and concatenated to the same 350bp, 17 unique haplotypes were delineated (Figure 2). See Chapter 6: the Discussion for more discussion on potential reasons behind the success rates of each sample type.

Phylogenetic analyses of island taxa vs. the mainland Emu

The Emu dataset as a whole shows relatively low nucleotide diversity but higher haplotype diversity (Table 2). Although the haplotype network (Figure 2) includes all four putative *Dromaius* taxa and encompasses *D. n. novaehollandiae* from across mesic and semi-arid Australia, there is no evidence that the island emus form distinct mitochondrial DNA lineages. Instead, haplotypes from each of the putative island taxa are fully intermixed with haplotypes present in mainland Emus. The presence of unique haplotypes (H4 and H5 in Figure 2) in each of the Kangaroo and King Island emu populations may be a reflection of 'neotypy' where novel mutations are just starting to appear in isolated populations that otherwise only share common ancestral haplotypes with the larger source population (Omland *et al.* 2006). However, sample sizes would need to be much larger to exclude the more likely possibility that these represent shared but rare haplotypes that have not yet been detected in the mainland Emu population.

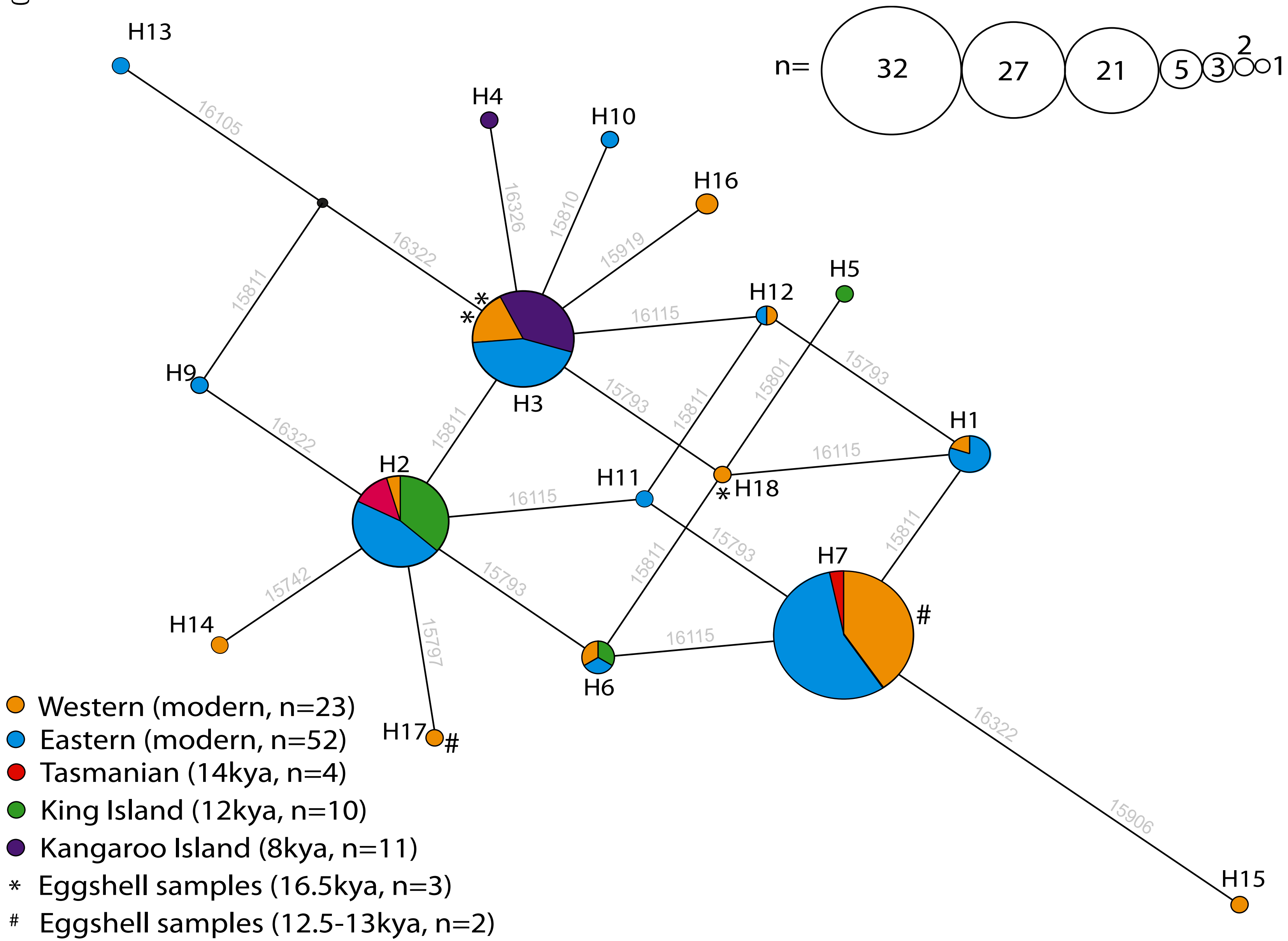

n=
32
27
21
5
3
2
1
H1
H2
H3
H4
H5
H6
H7
H9
H10
H11
H12
H13
H14
H15
H16
H17
H18
16105
16322
16326
15810
15919
16115
15793
15811
15801
15742
15797
15906
Western (modern, n=23)
Eastern (modern, n=52)
Tasmanian (14kya, n=4)
King Island (12kya, n=10)
Kangaroo Island (8kya, n=11)
* Eggshell samples (16.5kya, n=3)
# Eggshell samples (12.5-13kya, n=2)

Figure 2. **Median-joining Network based on 105 individuals with 350bp concatenated mtDNA *CR* sequence, showing frequency and location of haplotypes across Australia (those individuals without location details were excluded). The more common haplotypes form a reticulated network, or central cube (CC) shape, with those haplotypes with singletons extending outwards from the reticulation. The numbers on branches refer to the relative nucleotide position of mutations separating haplotypes compared to a reference sequence AF338711, and the node size reflects number of haplotypes (n). Note, all haplotypes from the eggshell samples that comprise the ancient western population are still commonly found in modern western and eastern populations.**

Table 1. Results of Akaike Information Criteria (AIC) and finite sample corrected AIC (AIC_C) analysis for the Emu population models simulated in Bayesian Serial Simcoal (BayeSSC) and evaluated using Approximate Bayesian Computation (ABC), and summarized in Figure 4. The model with the lowest AIC value is the more likely model given the data.

Model	No. Par_i	Ln (L_i)	AIC	Δ_i (AIC)	w_i (AIC)	AIC_C	Δ_i (AIC_C)	w_i (AIC_C)
H_0	1	-183.977	184.977	3.539	0.0625	370.001	6.642	0.0168
H_1 (10%, 50%)	2	-181.751	183.751	2.312	0.1154	367.984	4.625	0.0459
H_1 (90%, 90%)	2	-182.138	184.138	2.700	0.0951	368.758	5.399	0.0312
H_2	4	-187.285	191.285	9.846	0.0027	383.051	19.692	0.000
H_3 (20%, 0%)	4	-177.487	181.487	0.048	0.3578	363.456	0.097	0.4421
H_3 (35%, 0%) *	4	-177.439	181.439	0.000	0.3666	363.359	0.000	0.4640

* - Model with the maximum Ln (likelihood) and minimum AIC value.

No. Par_i = number of independent estimated parameters for model i; Ln (L_i) = natural logarithm of the maximum likelihood for model i; AIC = Akaike Information Criteria for model i; Δ_i (AIC) = [AIC_i – min(AIC)]; w_i (AIC) = rounded Akaike weights; AIC_C = finite sample corrected Akaike Information Criteria for model i; Δ_i (AIC_C) = [AIC_{Ci} – min(AIC_C)]; w_i (AIC_C) = rounded finite sample corrected Akaike weights. The total sample number used in BayeSSC simulations and AIC_C calculations was n=88.

Table 2. Population genetic diversity indices, neutrality test statistics, and population differentiation statistics as calculated in Arlequin.

Population	n	#H	PD	Diversity indices				Neutrality statistics	
				S	π (%)	Hd	Θ (S)	Tajima's D	Fu's Fs
Pilot *Dromaius spp.* dataset (1000bp)	18	8	1.91	6	0.19	0.88	1.74	0.31	-2.66
Entire *Dromaius spp.* dataset (350bp)	105	17	1.68	12	0.48	0.79	2.30	-0.70	-8.11*
Two population hypothesis (western vs. eastern) using 350bp dataset									
W	28	11	1.90	8	0.54	0.77	2.06	-0.23	-4.79*
E	77	12	1.56	8	0.44	0.78	1.63	-0.11	-4.06
Two population hypothesis (western vs. eastern) separated temporally									
W (modern)	23	9	1.82	7	0.52	0.72	1.90	-0.13	-3.27
W (ancient)	5	4	2.00	4	0.57	0.90	1.92	0.27	-1.01
E (modern)	52	10	1.64	6	0.47	0.78	1.33	0.60	-2.74
E (ancient) KAN/KIN/TAS	25	6	0.98	5	0.28	0.67	1.32	-0.75	-1.97

n - number of samples; #H – number of haplotypes; PD – mean number of pairwise differences between haplotypes; S – number of segregating sites; π – nucleotide diversity; Hd – haplotype diversity; θ(S) – Theta population parameter estimated from the infinite-site equilibrium relationship. * - P-value <0.05 (for Fu's F_S statistic, significance at the 5% level is reflected by a P-value <0.02); ** - P-value <0.005.

Morphometrics of dwarfed island vs. mainland Emu populations

Limb bones of the Kangaroo Island Emu (*D. baudinianus*) were significantly larger than those of the King Island Emu (*D. ater*) for all variables except tarso-metatarsi distal and shaft width (Table S1 for details of the measurements taken). Furthermore, the Emus formed a body-size ranking, with the Kangaroo Island Emu limb bones intermediate in size between the King Island Emu and mainland Emu (*D. n. novaehollandiae*) limb bones (Figure S1 and Table S1). As no skeletal material of the Tasmanian Emu are represented in museum collections, no limb bones of the Tasmanian 'subspecies' were available to compare against the King Island and Kangaroo Island Emu bones for relative size differences.

The PCA and allometry ratio spectra, *sensu* Baur and Leuenberger (2011), which test for evidence of allometry found the most evidence for allometric scaling in the proximal width and shaft depth of the femora bones, and distal width and shaft depth in the tarso-metatarsi across the mainland, King Island and Kangaroo Island Emu populations (see supplementary info, Figure S3 & S4). Furthermore, the Pearson correlation comparison between the standardised measurements and the geometric mean of the individual (as a proxy for overall size) found significant correlations between the same limb bone proportions as the PCA and allometry ratio spectra above (Table S2). Allometric scaling within organisms deals with the physiological shape changes required to maintain body function in different sized individuals of the same or phylogenetically similar species, however organisms can also show shape-uncorrelated size differences if the different body part performs different physiological functions, or performs the physiological functions in a different way. The Principal Component Analyses (PCAs) of the femur and tarso-metatarsi bones also highlight a lack of significant shape differences, other than those related to size, between *D. ater, D. baudinianus* and *D. n. novaehollandiae* (Figure S5).

The body-size ranking of the two island populations and mainland emus (from large to small: mainland Emu, Kangaroo Island Emu, King Island Emu) is positively associated with the size of the island (natural logarithm of island size in square kilometres) and negatively associated with the length of isolation from mainland Emu populations (date of separation by rising sea levels; Figure 3). These associations suggest a direct link between island habitat and dwarfing, but the sample size prevents a rigorous test of this relationship.

The lack of molecular genetic and morphological differentiation (other than size) of *D. ater* and *D. baudinianus* from mainland Emu indicates that these putative taxa are simply recent geographic isolates, and represent subsets of the diversity of adjacent mainland populations immediately prior to the Holocene rise in sea levels. Furthermore, the date of isolation can be used to calibrate phylogeographic analyses of mainland populations.

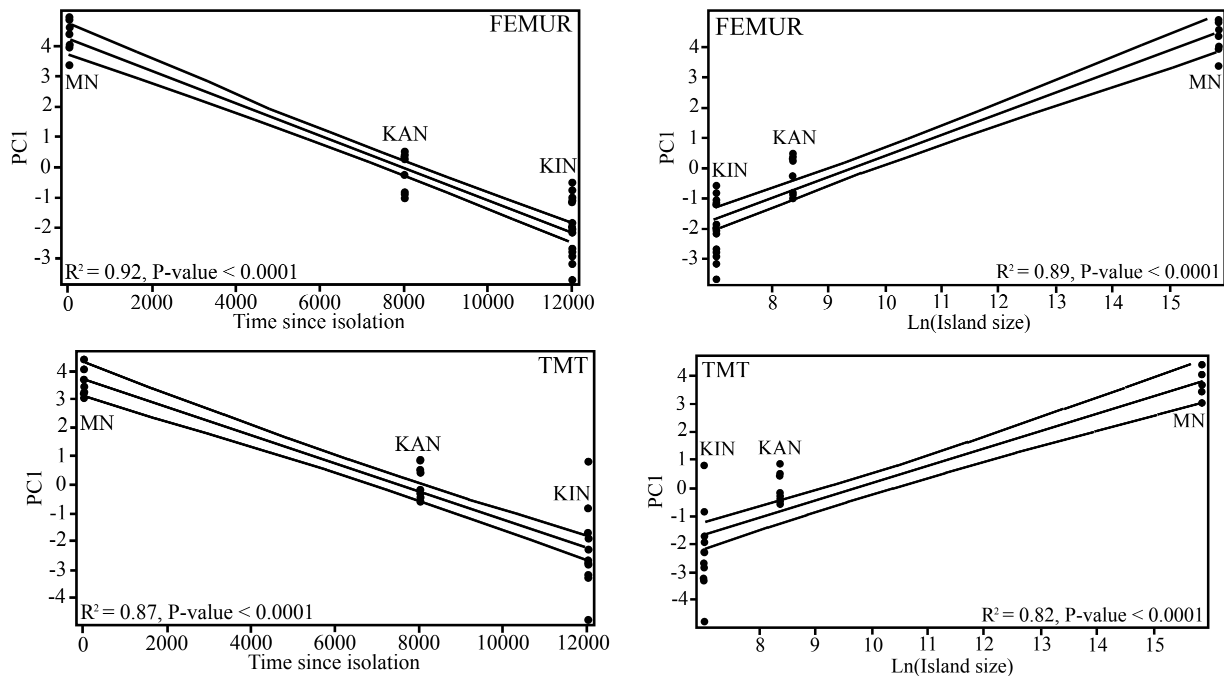


Figure 3. Linear regression of femur and tarso-metatarsi (TMT) PC1 versus 'Time since isolation' and 'Ln (island size)' of mainland (MN), Kangaroo Island (KAN), and King Island (KIN) emus.

Phylogenetic and phylogeographic analyses of the mainland Emu

The median-joining network also highlights the lack of geographic structure in the genetic diversity of mainland Emu across Australia (Figure 2). The reticulation of haplotypes in the centre of the network – the central cube (CC) – contains the three most common haplotypes (H2, H3, H7), which are distributed across the entire geographic range of the emu (Figure 1 and S2: black dots). CC haplotypes also represent the majority of the ancient samples (three of the five ancient eggshell samples are H3 or H7), suggesting haplotypes common in modern populations have been common since at least the late Pleistocene (the probability of obtaining two of the three most common modern haplotypes if they were not more common than just random in the late Pleistocene is 0.003; Table S3).

As *D. n. novaehollandiae* are known to disperse long distances (up to 300km in 8 months; Davies *et al.* 1971) isolation by distance (IBD) was not expected, so panmixia across the continent was initially used as a null hypothesis. When this assumption was tested in an Allelic Aggregation Index Analysis (AAIA tests for evidence against the null hypothesis that each allele is distributed randomly across a geographic landscape relative to the aggregation of an empirical dataset) there was evidence supporting this assumption of panmixia across the continent (Miller 2005). However, when different hypotheses about the Nullarbor Plain as a barrier to gene flow were tested in the BayeSSC analysis (one constant sized population vs. one population that split into two vs. two isolated populations), the observed data best fit a model of one population that split into two isolated populations but had a continuously low level of gene flow (25-35%) from west to east (Figure 4 & Table 1). This result suggests that the Nullarbor Plain/Eyrean Barrier

potentially acted as a semi-permeable barrier to gene flow at some point in the history of the Emu. Although not obvious from the haplotype network (Figure 2), the eastern and western Emu populations show significantly different haplotype frequencies when both modern and ancient populations are considered (p-value = 0.001787 from the Fisher's exact test in Table S3). However, when only modern populations are considered, the haplotype frequencies are not significantly different (at the 5% level; p-value = 0.07821 from the Fisher's exact test in Table S3). This suggests the recent gene flow may have been erasing the signal of this population divergence, since the dates of the ancient samples (*i.e.,* 8,800-16,500 ya).

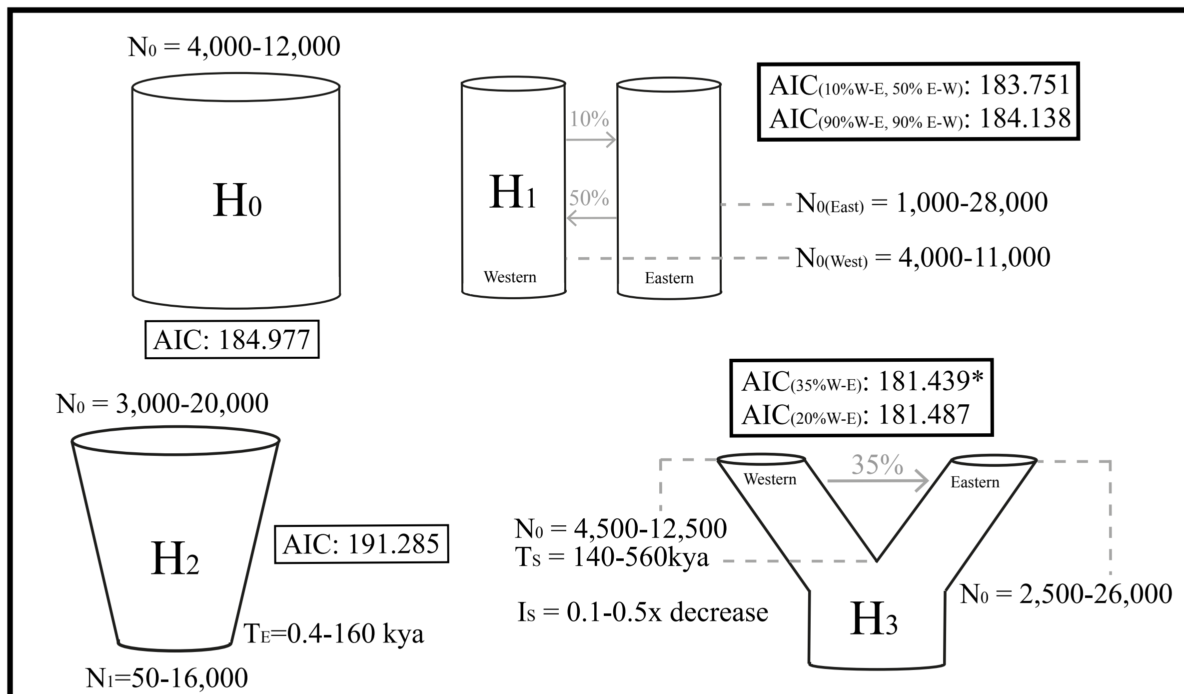


Figure 4. Population models used in Bayesian Serial Simcoal (BayeSSC) simulations followed by Approximate Bayesian Computation (ABC) to test the possible role of the Nullarbor Plain as a barrier to gene flow. Model H_0 is the null hypothesis of a single panmictic population of constant size. An alternate hypothesis is model H_1, where the Nullarbor Plain acted as a semi-permeable barrier over very long time-scales, separating a western population from an eastern population with different levels of migration/generation between them. Model H_2 is a single population that has undergone an expansion event T_E years ago from an effective population size of N_1 to N_0; Model H_3 starts out as a single panmictic population that is separated by the increasing aridity of the Nullarbor Plain, yet with some level of gene flow between them. Although H1 and H3 are similar models, H3 was included to establish whether there had ever been a time when the mainland Emu was acting as one interbreeding population. The level of gene flow was tested from 0-100% in both directions, plus from 0-100% in either direction only (in increments of 5% migration per generation). These four models are rather simplistic, however more complex models were not tested due to concern that the lack of genetic diversity in the dataset would hinder reconstructing the population parameters accurately, and because of a lack of computing power needed to perform the adequate number of simulations required for statistical robustness (see Chapter 2: Methodology for a discussion on the issue of number of simulations).

Reconstructing Emu demographic history

Given the deep evolutionary timescale of *Dromaius* in the Australian landscape (Patterson and Rich 1987; Boles 1992; Field and Boles 1998; Miller *et al.* 1999; Miller *et al.* 2005), the relatively low level of genetic diversity in both modern and late Pleistocene Emu populations is strongly suggestive of significant recent demographic perturbations in this lineage. We were therefore concerned that any genetic signal of the evolutionary history in *Dromaius* may have been obscured by their prolific breeding ability, short generation time, and high vagility (Dawson *et al.* 1984). However, the combined Emu dataset passes the date randomisation test (Figure S8), which suggests there is enough temporal signal in the ancient sequences for accurate reconstruction of the evolutionary rate used in inferring many evolutionary events.

Neither of the western or eastern populations show evidence of demographic expansions, yet both show strong evidence of spatial expansions. The population genetic statistics in Table 2 show evidence of population expansion in the western population and in the entire *Dromaius* dataset (indicated by the combination of moderate haplotype diversity and low nucleotide diversity, and in the significantly negative Tajima's D and/or Fu's F_S). Moderate haplotype diversity and low nucleotide diversity signify a history of population expansion (Lowe *et al.* 2004). Signs of this expansion being spatial (and not demographic) in nature are seen in both the BayeSSC simulations and the Bayesian Skyline Plots. The BayeSSC simulations, which tested between hypotheses of one panmictic population, vs. two vicariant populations with a shared vs. separate histories, found evidence of one constant sized population that became separated into two constant sized populations (140-560 kya) with 20-35% migration from the western into the eastern population as being more likely given the data (Figure 4). The Bayesian Skyline Plots (BSP, which is used to reconstruct the demographic history of a population) found that no demographic expansion was supported; moreover, a constant (yet stationary) population size could not be rejected for the entire *Dromaius* dataset, or for either the western or eastern populations, based on the Bayes Factor (Figure S6). Furthermore, both the western and eastern populations show no evidence for demographic expansion in the mismatch distribution-based expansion tests; yet both show support for spatial expansions, at around 100 thousand years ago (kya) (44-162 kya, 95% quartile) and around 348 kya (210-440 kya, 95% quartile) respectively (Table 3 and Figure 6). Given the overwhelming evidence for population spatial expansions, the bimodal mismatch distribution of the western population contrasted with the more unimodal mismatch distribution of the eastern population, is most likely due to differences in the deme size of each population or in the time since the expansion (Figure 5; Ray *et al.* 2003). The Genetic Landscape Shape plot created in Alleles in Space (AIS; Miller 2005), which graphically represents geographic areas of high genetic inter-individual distance, highlights the longitudinal

cline in genetic diversity; western samples have higher genetic diversity (*i.e.,* higher inter-individual distance) relative to eastern samples (Figure 7). Post-hoc linear regression analyses also revealed distinct longitudinal trends (Figure S7), where the mean numbers of pairwise differences were negatively correlated with longitude ($R^2 = 0.9198$, $P = 0.0006$), *i.e.,* the western population has higher genetic diversity.

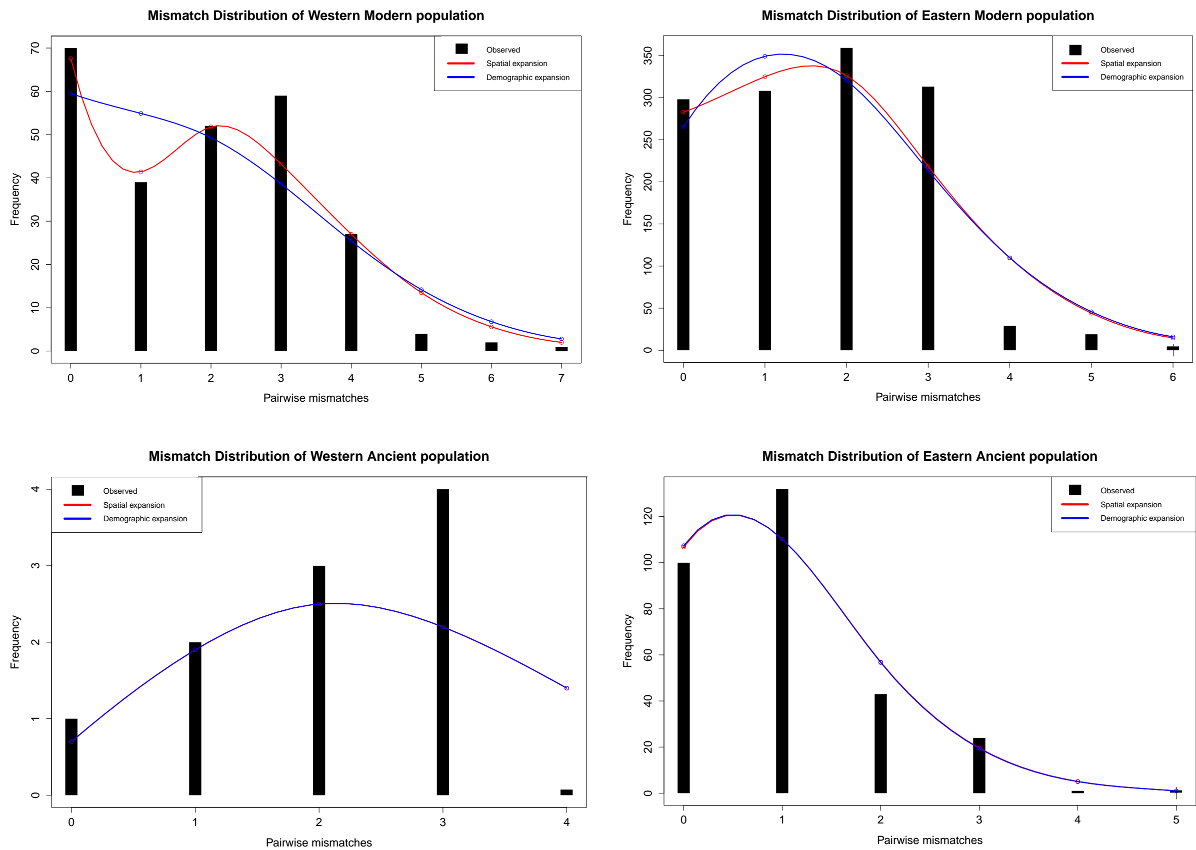


Figure 5. Pairwise mismatch distributions observed for each population (western vs. eastern), and for each time period (ancient vs. modern) compared to the pattern expected under demographic and spatial expansions. The western ancient population mismatch distribution is shown, however it was excluded from most further analyses as the sample size was too low (n=5).

Table 3. Population expansion tests as calculated in Arlequin (excluding West Ancient due to small sample size).

Population	n	Parameters estimated under the demographic expansion model						Test of Goodness of fit (demographic expansion model)		Parameters estimated under the spatial expansion model						Test of Goodness of fit (spatial expansion model)	
		τ	T (ya)	2.5% qt T (ya)	97.5% qt T (ya)	Θ (0)	Θ (1)	SSD	Hg	τ	T (ya)	2.5% qt T (ya)	97.5% qt T (ya)	Θ (0)	M	SSD	Hg
Two population hypothesis (Western vs. Eastern)																	
W (modern)	23	29.8	-	-	-	-	-	0.10	0.30*	8.08	108,895	44,226	161,700	0.32	2.02	0.05	0.18
E (modern)	52	27.8	-	-	-	-	-	0.10**	0.19**	25.9	349,057	210,533	433,540	0.49	2.7	0.07	0.19
E (ancient)^	25	22.5	-	-	-	-	-	0.21*	0.35	22.0	348,143	226,478	439,192	0.34	1.66	0.11	0.34

n - number of samples; τ – age of expansion (in units of mutation); Θ (0) - Theta population parameter estimated at time = 0; Θ (1) - Theta population parameter estimated at time = 1; SSD – sum of squared deviations between observed and expected mismatches; Hg – Harpending's Raggedness index. The Arlequin input file was based on 564bp of mtDNA *CR* sequence (uses some missing data so that the maximum number of samples could be included) and the mean molecular rate of 1.06×10^{-7} substitutions/site/year (estimated from the BEAST runs). For calculation of T (time since expansion in years) the value of μ is 2.39×10^{-4} substitutions/seq/generation, and for those populations with ancient samples the average age of the ancient dataset was added to the value output from Arlequin. ^ Eastern (ancient) population has an average age of 11,564 ya.
*** - P-value <0.05; ** - P-value <0.005. The P-values for the Test of Goodness of fit indicate those datasets that do not fit the assumption being tested (*i.e.*, is evidence against the assumption of a demographic/spatial expansion**

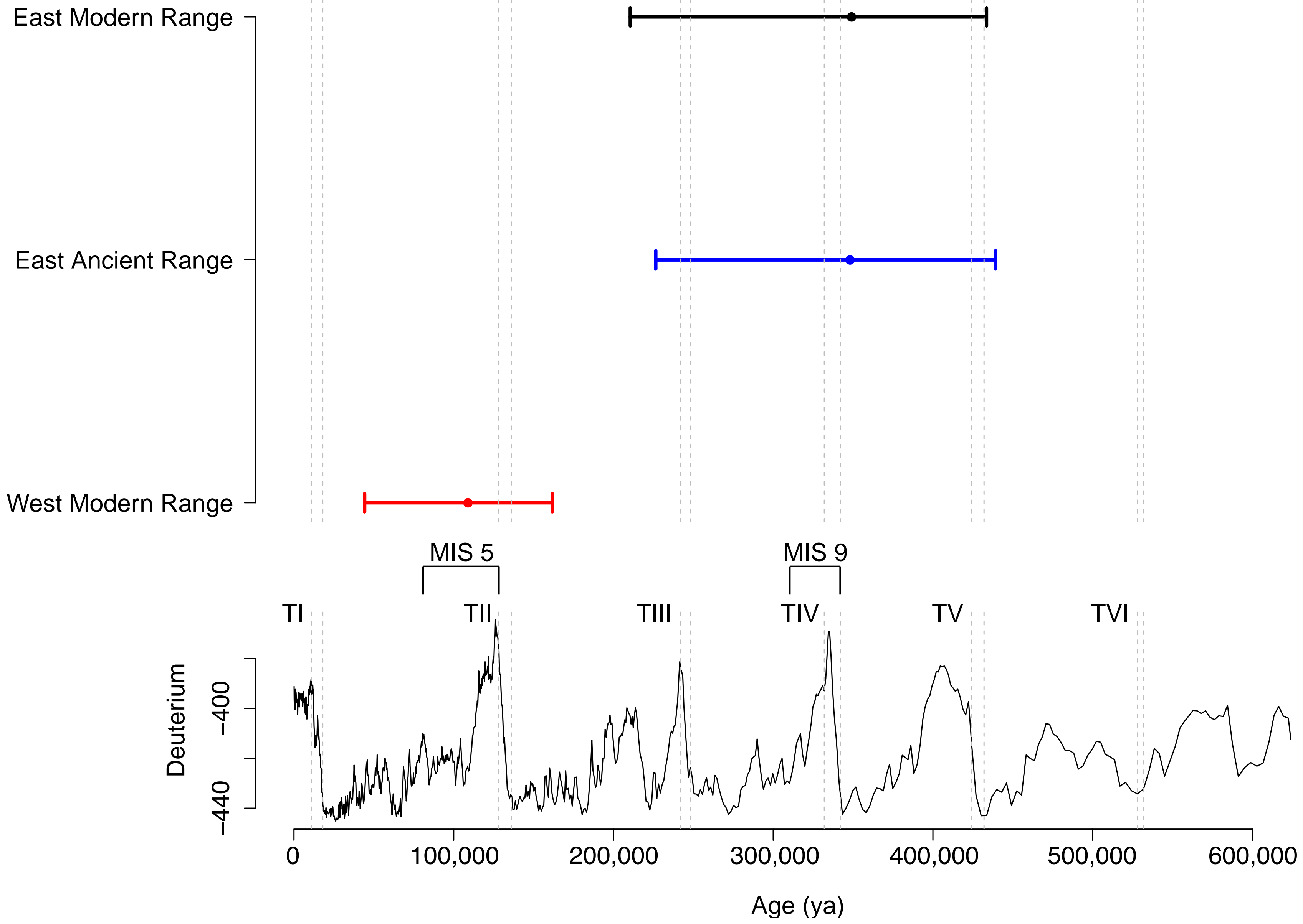
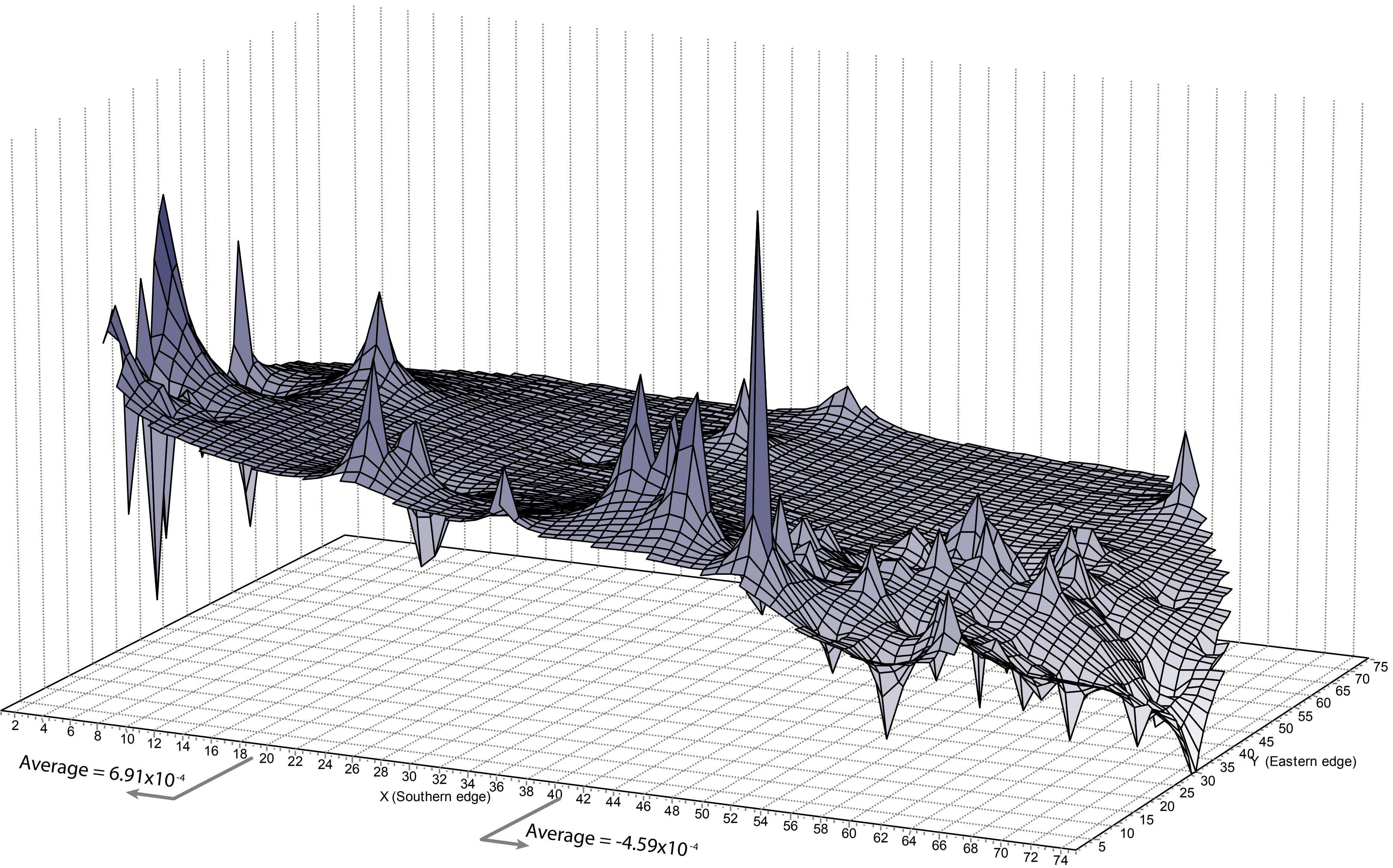

East Modern Range
East Ancient Range
West Modern Range
MIS 5
MIS 9
TI
TII
TIII
TIV
TV
TVI
Deuterium
−400
−440
0
100,000
200,000
300,000
400,000
500,000
600,000
Age (ya)

Figure 6. Range expansion signal as calculated for each population (excluding West Ancient due to small sample size) using mismatch distributions in Arlequin, compared to Antarctic deuterium levels as a proxy for changing temperature in the southern hemisphere over the Late Pleistocene (EPICA *et al.* 2004). Note the West Ancient Range population only has n=5 samples, so was excluded from the Arlequin expansion tests.



Y (Eastern edge)
5 10 15 20 25 30 35 40 45 50 55 60 65 70 75
X (Southern edge)
2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44 46 48 50 52 54 56 58 60 62 64 66 68 70 72 74
Average = $6.91x10^{-4}$
Average = $-4.59x10^{-4}$

Figure 7. Three dimensional representation of spatial patterns of modern *D. n. novaehollandiae* genetic diversity across the sampled landscape of the Australian continent (note, if the 3D shape landscape plot is seen from the top, the outline of the Australian continent is more obvious). The X and Y axes show 75 x 75 grid based on longitude (Southern edge) and latitude (Eastern edge), and the Z axis shows residual genetic distances between sampling sites calculated in program Alleles in Space. Peaks (dark shading) indicate large genetic diversity, and troughs (light shading) indicate high genetic similarity between sampling locations. The cline in residual genetic distance from west to east is indicated by the average interpolated residual genetic distance of the western vs eastern sampling localities (western sampling localities are defined as <124° longitude, with an average of 6.91×10^{-4}; and eastern sampling localities are defined as >135° longitude, with an average of -4.59×10^{-4}).

Discussion

The nucleotide diversity of the Emu is relatively low compared to other Australian bird species with a widespread range and high dispersal ability, such as the Grey Teal (*Anas gracilis;* π: 1.4% for control region) and Australian Magpie (*Gymnorhina tibicen;* π: 0.87-4.64% for control region), but is similar to other Australian birds suspected of experiencing population expansion following climate-mediated bottlenecks, such as the Horsfield's Bronze-Cuckoo (*Calcites basalis;* π: 0.44% for control region), the Pied Butcherbird (*Cracticus nigrogularis;* π: 0.2% for ND2), the Black-faced Woodswallow (*Artamus cinereus;* π: 0.4%), and the Spiny-cheeked Honeyeater (*Acanthagenys rufogularis;* π: 0.22%) which all have nucleotide diversities an order of magnitude lower (Joseph *et al.* 2002; Joseph and Wilke 2007; Toon *et al.* 2007; Joseph *et al.* 2009; Kearns *et al.* 2010). Given the deep evolutionary timescale of *Dromaius* in the Australian landscape (Patterson and Rich 1987; Boles 1992; Field and Boles 1998; Miller *et al.* 1999; Miller *et al.* 2005), the relatively low level of genetic diversity in both modern and late Pleistocene Emu populations is strongly suggestive of significant recent demographic perturbations in this lineage. We were therefore concerned that any genetic signal of the evolutionary history in *Dromaius* may have been obscured by their prolific breeding ability, short generation time, and high vagility (Dawson *et al.* 1984). However, the combined Emu dataset passes the date randomisation test (Figure S8); additionally, recent advances in serial coalescent analyses allowed us to test the likelihoods of alternative population histories given the observed data, and standard population genetic techniques confirmed the history of spatial expansions in *Dromaius.*

The evolutionary origin of dwarf island Emus

The three main island populations of Emus (*D. baudinianus, D. ater* and *D. n. diemenensis*), now all extinct, have long been regarded as separate taxa, the first two mainly on account of their dwarf stature. However, the high proportion of shared haplotypes with *D. n. novaehollandiae* indicates that they are all recent derivatives of a single, more widely distributed ancestral population, and all probably owe their special characteristics to the effects of insularization. The apparent links between limb size (*i.e.,* as a proxy for overall body size) and both island size and length of isolation, suggest that *D. ater* and *D. baudinianus* were undergoing directional selection but had possibly not yet reached optimal limb size. The lack of genetic differentiation suggests these morphological differences are due to phenotypic plasticity. This may also be the case for the reported dark colouration of the island Emu taxa (Le Souef 1904), as Heupink *et al*. (2011) found no evidence in the *MC1R* gene to support a genetic basis for plumage colouration differences between *D. ater* and *D. n. novaehollandiae*; however, other genes that regulate the melanocortin system may be involved (Roulin *et al.* 2011). In addition, the few

morphological differences reported for the island taxa include the distal foramen of the tarsometatarsi being open in *D. baudinianus* compared to *D. ater* (Parker 1984), and the dome shaped skull in *D. baudinianus* and *D. ater* compared to *D. n. novaehollandiae* (Heupink et al. 2011), which on closer inspection are all found within the variability of *D. n. novaehollandiae* (Patterson and Rich 1987; Heupink et al. 2011). Parker's (1984) lack of comparison with *D. n. novaehollandiae* when describing *D. baudinianus* throws doubt on his judgement of the dwarfed emus from Kangaroo Island as a separate species, and in fact the allometric scaling and lack of genetic differentiation shown here may necessitate a re-description of *D. baudinianus* to the subspecific level within *D. novaehollandiae*.

The molecular data presented here argues against the notion that the various island dwarf emus were relicts of a formerly more widespread dwarf emu that also occurred on the mainland, however this cannot be completely ruled out on the basis of the samples analysed to date. There are several dwarf species of Emu that have been described from the fossil record, however even the youngest of these are of 'Late Pleistocene' age and the preservation conditions of the sand lunette where they were discovered at Lake Menindee in NSW makes the retrieval of aDNA unlikely (Patterson and Rich 1987). In fact, the few pre-Holocene Emu samples to yield aDNA in Australia have all been found in cave sites (this study; Oskam *et al.* 2010), so until well preserved remains of relatively recent dwarfed mainland Emus are found this hypothesis cannot be tested. Furthermore, it is possible that a former dwarf mainland relative of *D. n. novaehollandiae* might have been recently derived and only weakly differentiated genetically, or that mitochondrial introgression had occurred between two more fully differentiated species.

Therefore, the combined genetic and morphometric evidence suggests that the island taxa are populations of *D. n. novaehollandiae* that became dwarfed after isolation from mainland gene flow. As island dwarfism can result from density-dependent factors (*i.e.,* resource limited systems) and/or relaxed predator pressures creating conditions that favour smaller individuals, a discussion of these factors in the Australian context is warranted. It is likely that predator pressures similar to the mainland were experienced on the islands as Aborigines were known to hunt Emus on the mainland with net traps, and Aboriginal populations occupied Tasmania continuously until European settlement (Flood 2004). Furthermore, Aborigines inhabited Kangaroo Island until approximately 5 kya (Hope *et al.* 1977), and were at least present on King Island at 14 kya (Flood 2004). In addition Thylacines (*Thylacinus cynocephalus*) are known to have inhabited both Tasmania (Jones and Stoddart 1998) and Kangaroo Island (Prideaux pers. comm.), but not King Island (Flood 2004). The hunting style of the thylacine is as an ambush predator, but it does not usually hunts animals as large as full sized Emus (Jones and Stoddart 1998; Wroe *et al.* 2007; Figueirido & Janis 2011), however the dwarfed nature of the island Emus may have made easier prey for thylacines. Although no data is available on the mechanics of the

gait of dwarf Emus, the high manoeuvrability of the mainland Emu may have been retained in the island Emus if hunting pressures by Aborigines/thylacines would have selected for it. Therefore dwarfed Emus may still have posed a challenge to the thylacine due to their large body size, and ability to kick and escape predators. The relaxed selection pressures that caused the island dwarfism on Kangaroo and King islands were therefore possibly not a lack of predators, but more likely due to a relaxation of selection for extreme mobility (*i.e.,* long legs) needed on mainland Australia to locate scarce resources such as water and high quality food items (Dawson *et al.* 1984). On small islands such mobility is less likely to be a selective advantage while the reduced resource requirements of smaller animals may be better suited to the environment (Lomolino 1985).

Demography - west to east spatial expansion at 350 kya

We have identified signals of a spatial expansion of emus from west to east across the Australian continent in multiple analyses. The BayeSSC analyses support the scenario of a single population that became divided sometime in the Middle Pleistocene (560-140,000 ya) but with low levels of uni-directional gene flow between them (20-35% per generation). This gene flow was from west to east across a semi-permeable barrier (Figure 4). As we have few mainland Emu samples between the Nullarbor Plain and Eyrean Barrier, at this time it is not possible to state definitively which of these two barriers to gene flow caused the west-east split in Emu evolutionary history prior to 350 kya. However, Emu eggshell remains excavated from the marginal dunes around Lake Eyre and from near Port Augusta (west of the Eyrean barrier, yet east of the Nullarbor Plain) extend continuously back in time until at least 140 kya (Miller *et al.* 2005), suggesting the Eyrean Barrier is not likely to have acted to prevent gene flow during the last glacial cycle. This history of Emu presence (Miller *et al.* 2005) also highlights the fact that although the Lake Eyre playa lake may have supported Emu occupation of the semi-arid region during the last glacial cycle, no evidence exists as to whether the playa basins provided sufficient refugial conditions during previous glacial cycles.

The lack of clear phylogeographic structure in the haplotype network suggests that most of the evidence of past isolation between the western and eastern populations has likely been erased by more recent gene flow across the Nullarbor Plain/Eyrean Barrier. Neither of the western or eastern populations show evidence of demographic expansions, yet both show strong evidence of spatial expansions. The expansion signal in the eastern population dates to approximately 350,000 ya – roughly coincident with Termination IV, just prior to the start of Marine Isotope Stage (MIS) 9 (Becquey and Gersonde 2003); and in the Western population to approximately 100,000 ya – just after Termination II, during MIS 5 (Becquey and Gersonde 2003; EPICA *et al.* 2004; Table 3 and Figure 6). However, the upper and lower bounds of the

expansion events cover a substantial period of time. In the case of the eastern population expansion, the margins of error include Terminations III and V, and the intervening glacial periods. Given the substantial temporal overlap with other glacial and interglacial periods for the eastern expansion, we are concerned the uncertainty involved in these calculations obscures any authentic signal of the timing of a population range expansion. The BayeSSC simulations also detected evidence of this divergence from one population into vicariant western and eastern populations, however the error margins involved in the BayeSSC analysis were even wider (140-560 kya). Even taking into account the uncertainties involved in the expansion tests, the point estimate of the both eastern and western population expansion times do appear to roughly coincide with periods of climate amelioration in Termination IV and Termination II, respectively (Pahnke *et al.* 2003). This response of the Emu during periods of increasing moisture availability is not surprising in a taxon with such well-attuned sensitivities to detecting rainfall over long distances, which has provided the Emu with an ability to respond rapidly to improving environmental conditions. Research has shown that Emus seem to be conscious of the cues signalling heavy rain, as in western Australia they have been detected not only migrating according to the direction of predominant seasonal rainfall (north in summer, south in winter), but when a reversal in the normal seasonal rainfall occurs they reverse their migration patterns (Davies 2002). The more surprising result is that the Nullarbor Plain/Eyrean Barrier would act as a semi-permeable barrier to such a large species with a high dispersing ability. Although Emu populations may have skirted the northern border of the Nullarbor Plain to cross the continent, the current distribution map (Figure S2) for the Emu shows the sparse nature of recorded sightings in this very arid region. The lack of Emu presence in this region during the current interglacial period may suggest that both current (humid) and past (arid) climates may have prevented even such a generalist as the Emu from successfully inhabiting the arid region north of the Nullarbor Plain.

The population genetic analyses suggest that the western population of *D. n. novaehollandiae* is ancestral and identify the southwest corner as a likely recurrent refugium for the species. This conclusion is supported by the higher nucleotide diversity level in the western population (Table 2) and by the west-east longitudinal cline in genetic diversity seen in the Landscape Shape analysis (Figures 7 & S7). Furthermore, an east to west spatial expansion is not likely due to the strong genetic differentiation (with highly significant F_{ST} values) between both modern populations and the ancient eastern samples (8,800-14,000 ya) – contrasting with the lack of differentiation between the modern populations and the ancient western samples (12,500-16,500 ya; see Table S3). The BayeSSC and BSP simulation (Figure 4 and S6) also refute the alternative scenarios of *in situ* demographic changes in both eastern and western populations. The southwest is a long term refugium for many groups of plants of animals; its flora is characterised

by much higher (10-fold) biodiversity and endemism than south-eastern Australia (Dodson 2001), and there are several endemic mammals and birds, with varying levels of divergence from eastern Australian relatives. One way to further test the hypothesis that the south-west corner was a critical refugium for the Emu would be to compare the tolerance level for fluoroacetate between *D. n. novaehollandiae* in the east with those in the west; similar levels between east and west would support the idea of a western refugium, but if disparate levels were detected in the east compared to the west, then a separate eastern refugium would be more likely.

Postglacial expansions of Emus out of a southwest refugium may have occurred repeatedly during previous glacial cycles, yet been overwritten by the most recent expansion signals detected from the Middle Pleistocene. The one major expansion detected around 350 kya saw the western population of *D. n. novaehollandiae* cross the Nullarbor Plain and expand over the eastern half of the continent. However, recently discovered Middle Pleistocene cave remains highlights an extensive assemblage of woodland fauna that inhabited the Nullarbor region from approximately 200-400 kya (Prideaux *et al.* 2007) around the same time as the west-east postglacial expansion in *D. n. novaehollandiae*. One likely explanation for this change in habitation level is the post-Termination IV interglacial appears to have been more warm/humid than previous interglacials, potentially allowing a broader expansion of flora and fauna biodiversity into the Nullarbor region.

Demography – western expansion at 100 kya

A second minor post-glacial expansion probably occurred around 100,000 ya (post-Termination II) but this extended across the western half of the continent. Interestingly, Termination II (in addition to Termination IV) was also substantially more warm/humid than other interglacials, suggesting *D. n. novaehollandiae* requires relatively warm/humid conditions to expand out from refugia – the climatic/ecological niche of *D. n. novaehollandiae* may be narrower than assumed. Although the two expansions around the time of Terminations II and IV were the predominant signal detected in the *D. n. novaehollandiae* dataset, both the western and eastern populations are likely to have experienced small-scale within-population contraction/expansion events in response to other glacial/interglacial cycles. The eastern population in particular would have experienced different ecological conditions (after expanding into the east post-Termination IV) compared to previous glacial cycles in the west. However, even in the east there would have been refugia to provide sanctuary during glacial maxima: the eastern inland arid and semi-arid lakes, and/or the southeast corner. Evidence that these lakes have supported life during extreme conditions is evidenced by continuous deposits of *D. n. novaehollandiae* eggshell samples from 130,000 ya to the present at Lake Eyre, the Murray-Darling basin and Port Augusta. Emu persisted in the Lake Eyre and Murray-Darling regions

even through sudden and substantial ecosystem change that precipitated the extinction of *Genyornis newtoni*, another ratite species known to have inhabited this region with *D. n. novaehollandiae* (Miller *et al.* 1997; Miller *et al.* 1999; Miller *et al.* 2005). The omnivorous and opportunistic feeding behaviour of *D. n. novaehollandiae* may have allowed it to persist around these playa lakes as long as there was some water available (the simple gastrointestinal tract of the Emu necessitates frequent water consumption - daily in the case of chicks; Dawson *et al.* 1984). The Australian southeast (Victoria/Tasmania) probably also provided suitable refugium conditions, as it lies at similar latitudes to the southwest refugium and contains much topographic and habitat heterogeneity. Finally, the Eyre Peninsula, Yorke Peninsula, Mt. Lofty Ranges, and Kangaroo Island are thought to have formed one continuous refuge (South Australia) during at least the last glacial maxima (Ford 1987a). Whether or not this South Australia refugium and the southeast corner would have been periodically contiguous is not known, however hydrologic zonation maps projected to the LGM (using the climatic index of Prescott based on precipitation and evaporation ratios) show they both have similar climatic index values, suggesting gene flow between these refugia was possible (Bowler 1982).

It seems extraordinary that such a high dispersing, omnivorous taxa that currently inhabits arid sections of the Australian landscape would be limited geographically by the increasing aridity of past glacial cycles. However, ecological studies have shown the Emu to have highly attuned sensitivities to cues signalling rainfall, even at quite a large distance (Davies 2002). It is possibly these ties to rainfall events, and the associated flourishing food sources (such as seeds, flowers, sprouting plants, and insects) that ultimately determine the distributions of Emu across the landscape.

Conclusion

Dromaius has a complex demographic history that can only be reconstructed through a comprehensive survey of both modern and ancient genetic diversity (via fossil bone and eggshell samples). Our study identifies the Nullarbor Plain as likely to have acted as a semi-permeable genetic barrier for *D. n. novaehollandiae*. It also identifies a likely refugium in the humid southwest of Australia, where a degree of genetic diversity was preserved through multiple glacial cycles; and provides evidence for at least two postglacial expansions: one eastwards across the Nullarbor Plain around Termination IV, and the other within the western population around Termination II. We find no genetic evidence that the extinct island emu taxa of King Island, Kangaroo Island, or Tasmania (*D. ater, D. baudinianus*, or *D. n. diemenensis*) warrant separate taxonomic status. Each of these populations appears to be recent isolates of *D. n. novaehollandiae,* which have undergone rapid morphological change in response to insularization. An association between limb length and island size/length of isolation in these

dwarfed taxa suggests the relaxation of selection pressures accompanying a reduction in home range size/resources was responsible for the island dwarfism via phenotypic plasticity.

The low genetic diversity found within the Emu appears to stem from multiple genetic bottlenecks caused by contraction to the refugial zone in the southwestern corner of the continent during arid glacial cycles, prior to range expansions out of this refugia during interglacials (MIS 5 and MIS 9). One consequence of low genetic diversity is a potential susceptibility to disease, such as seen in the Tasmanian Devil (*Sarcophilus harrisii*) and facial tumour disease (Hawkins *et al.* 2006; Miller *et al.* 2011). The lack of genetic variation may (especially if the mitochondrial control region's lack of diversity is mirrored in immune response genes as seen in the Tasmanian Devil; Siddle *et al.* 2007) expose the Emu, a common and widespread species, to the risk of decline and potentially eventual extinction if faced with a similarly strong environmental challenge.

We note that although future work in this area could concentrate on increasing the sample size and geographic spread of ancient specimens (in particular eggshell samples), given the lack of genetic diversity in the mtDNA *CR* we do not expect that increasing the depth of sequencing to the entire mitogenome will substantially increase the resolution of *Dromaius* evolutionary history. On the other hand, the nuclear genome may hold promise in reconstructing *Dromaius* evolutionary history, and this is likely to be accessible through recent advances in next generation sequencing and library preparation of ancient samples.

Acknowledgements

We thank the following people and institutions for samples: Walter Boles and Yong Yi Zhen at the Australian Museum (AM), Robert Palmer and Leo Joseph from the Australian National Wildlife Collection (ANWC) at CSIRO, Aaron Camens and the Field Naturalists Society, Michael Bunce and Joe Dortch from Murdoch University, Wayne Longmore and David Pickering at Museum Victoria, Mark Adams from the Natural History Museum at Tring, Gavin Dally from the Northern Territory Museum and Art Gallery, Rohan Wallace and Richard Johnson from the Queensland Department of Environment and Resource Management (DERM), Heather Janetzki from the Queensland Museum (QM), Lisa Gershwin and Craig Reid from the Queen Victoria Museum and Art Gallery (QVMAG), Cath Kemper and Philippa Horton at the South Australian Museum (SAM), Claire Stevenson from the Western Australian Museum (WAM), and Andrew Woolnough.

Supplementary Materials and Methods

DNA extraction and sequencing

Ancient material. A total of 28 ancient Emu bone, eggshell and feather samples were obtained from museums and collaborators (Table S4). These ancient samples included specimens of *Dromaius baudinianus* (Kelly Hill Caves, Kangaroo Island), *D. ater* (King Island), *D. n. diemenensis* (Tasmania), and *D. n. novaehollandiae* (Tunnel Cave, WA). DNA was extracted from 0.1-0.7 g of bone powder using a modified Qiagen silica-column based kit (QIAGEN DNeasy Blood and Tissue kit), with DNA extracted from the eggshell samples following a modified version of Oskam *et al.* (2010). All ancient samples were extracted in a specialised ancient DNA laboratory at the Australian Centre for Ancient DNA (ACAD) that is physically remote from modern genetic laboratories. The ACAD ancient laboratory is kept at positive air pressure and uses standard ancient DNA precautions (clean body suits, face shields, multiple pairs of gloves), universal UV irradiation, and regular cleaning with oxidising agents to prevent contamination. All reagents and materials are supplied DNA-free and used only once. Separate hoods are used for DNA extractions and to set up PCR experiments, and the amplification and post-PCR procedures are performed in a post-PCR laboratory located in a separate, physically remote building to minimise the potential for contamination by PCR amplicons. The uni-directional movement of researchers and materials is carefully controlled to further reduce contamination risk.

Modern material. We obtained 19 non-invasive opportunistically collected samples from across Australia (including 1 scat, and 18 modern feather samples from emus killed by vehicles) and 67 museum specimens (including 10 historically collected bone samples, 21 toepads, 5 feathers, 2 skin or mummified tissue specimens, and 29 modern tissue samples). DNA was extracted from bone samples using the same method as for ancient bone samples in the specialised ancient DNA laboratory. DNA was extracted from the scat sample using an ISOLATE Fecal DNA Kit (Bioline), feathers were extracted using a modified version of the method from Rawlence *et al.* (2009), and toepads and skin/mummified tissue samples were extracted with a modified DNeasy Blood and Tissue Kit (Qiagen) in a pre-PCR clean-room laboratory. The ACAD pre-PCR clean-room laboratory uses ancient DNA precautions (laboratory coats, face masks, multiple pairs of gloves) and regular cleaning with oxidising agents to prevent contamination. The modern tissue samples were extracted in a modern laboratory using a salting-out method (Miller *et al.* 1988).

PCR and Sequencing. To estimate the genetic relationships amongst *Dromaius*, we successfully amplified and sequenced variable portions of the mitochondrial *control region* (mtDNA *CR*) from 110 individuals, comprising 87 *D. n. novaehollandiae*, 11 *D. baudinianus*, 8 *D. ater*, and 4 *D. n. diemenensis,* which we combined with seven already published by Heupink *et al.* (2011). Eighteen randomly selected modern individuals spanning the geographic spread of the continent were sequenced for both hypervariable regions (HVSI and HVSII) of the mtDNA *CR* to identify variable sites. When these initial 18 individuals indicated very low levels of genetic variation (6 SNP's in 1000bp), 100-200bp regions surrounding the parsimony informative SNP's were sequenced in the balance of the samples while attempting to cover both hypervariable regions. Primers were combined into two multiplexes: Multiplex A, comprising primers A1360/A1407 (174bp), A1414/A1361 (105bp), A1691/A1692 (149bp), A1695/A1363 (143bp); and multiplex B, comprising primers A1410/A1411 (175bp), A1539/A1440 (135bp), A1693/A1694 (147bp) (Table S5). PCRs were set up using 25 uL volumes containing a final concentration of 1U Platinum Taq DNA Polymerase High Fidelity (Invitrogen), 1x PCR Buffer (Platinum, Invitrogen), 3mM $MgSO_4$, 200 uM each dNTP, 2 mg ml^{-1} rabbit serum albumin (Sigma), 1 uM forward and reverse primers and 2-3 uL of template DNA. PCR reactions were done on a Corbett Research Palm Cycler with the following cycling conditions: 94°C for 2 min, 55 cycles of 94°C for 30 s, 55°C for 30 s, 68°C for 30 s, and a final extension of 10 min at 68°C. Amplifications of extraction and PCR controls were performed in all experiments to monitor contamination. PCR products were separated by electrophoresis on a 3.5% agarose gel. Singleplex reactions were amplified from the multiplex reactions if warranted. Successful PCR products were purified using EXOSAP (Fermentas), after which the forward and reverse complement of each fragment was sequenced from the same PCR reaction with the PCR primers using Big Dye Terminator v3.1 cycle-sequencing chemistry and cleaned up using a Multiscreen®$_{384}$ SEQ plate (Millipore). Sequencing was conducted on an ABI 3130XC capillary sequencer. Sequences were edited and the forward and reverse strands aligned using Geneious v5.4.6 (Drummond *et al.* 2011).

Morphometric analyses

A review of published morphometric measurements of *D. ater* and *D. baudinianus* revealed only basic measurements of leg bone length and shaft diameter (Scott 1924; Morgan and Sutton 1928). Therefore, a more substantial morphometric analysis of femur, tibia and tarso-metatarsi bones was undertaken for *D. ater, D. baudinianus*, and *D. n. novaehollandiae* (Table S6), although the tibia dataset was excluded from further analyses due to measurement error. As the ancient specimens were not excavated as articulated skeletal elements, measurements of only right hand femora/tibia/tarso-metatarsi were taken to ensure no individual was measured twice. While every effort was made to take measurements from specimens sampled for DNA purposes,

this often was not possible due to damage to the bone architecture or visual inspection indicating poor DNA preservation. Furthermore, sample sizes were too low to successfully separate any effects of sexual dimorphism within each population.

We measured the size and shape of the two leg bones as overall length (L), proximal width (PW), proximal depth (PD), distal width (DW), distal depth (DD), shaft width (SW), and shaft depth (SD). These measurements were chosen to best reflect those aspects of limb size and proportions that vary with climate according to Allen's rule, specifically limb bones from the three major limb segments. These raw measurements as well as shape indices (Table S6) and log shape variables calculated from them have all been shown to vary significantly between ratite species according to latitude (Worthy 1994). To explore these data, plots of the measurements against each other were prepared using R and 'plotmatrix' in ggplot2 package (Figure S1). Univariate analysis of variance (ANOVA) of the pairwise comparisons of measurements between taxa was assessed using Bonferroni's correction with $\alpha = 0.05$ to establish whether *D. ater* and *D. baudinianus* were significantly different in size. Multivariate analyses, such as principal components analyses (PCA) and regression analyses of PC1 versus island size and time since isolation were done on the 'log shape' variables. The log shape variables are simply the log-transformed ratios of each variable with the geometric mean of all measurements for that individual (or the geometric population mean – geometric mean of that measurement for all individuals). These ratios are 'scale free' in that an isometric size component has been removed from the equation. These tests were performed to test hypotheses regarding isometric vs. allometric size differences, optimal body size and evolutionary equilibrium. However, this type of isometric size correction does not account for all shape differences associated with size; therefore, size adjusted data may still be influenced by allometry. The PCA was carried out using the prcomp function in R with a covariance matrix (which is more appropriate than correlation matrices for small sample sizes; Perktas and Gosler 2010). A Pearson correlation analysis of the size-adjusted data (as per Gilbert and Rossie 2007) was done to test for a significant influence of allometry on the Emu taxa (Table S2). Additionally, PCA and Allometry ratio spectra (as per Baur and Leuenberger 2011) were also undertaken on the log shape variables, and were run using the R script from Baur and Leuenberger (2011).

Population Genetic Analyses

Population diversity indices (No. of haplotypes, No. of segregating sites, haplotype diversity, nucleotide diversity, Tajima's D statistic, and Fu's FS) and differentiation statistics (F_{ST}) were estimated using Arlequin v3.5.1.2 (Excoffier and Lischer 2010), and are presented in Table 2 and S4. As *D. n. novaehollandiae* are known to disperse long distances (up to 300km in 8 months; Davies *et al.* 1971) isolation by distance (IBD) was not expected, so panmixia across the

continent was initially used as a null hypothesis. The panmixia assumption was tested with an Allelic Aggregation Index Analysis (AAIA) using Alleles in Space (AIS) v1.0 (Miller 2005). AIS calculates genetic distance using simple mismatch distances calculated from raw genetic data, which it assigns to the midpoints of a Delaunay triangulation-based connectivity network that it generates.

The dataset was also analysed using Genetic Landscape Shape in AIS in order to visually identify genetic discontinuities or landscape regions where relatively low or high genetic distances occur. Given the large spatial scale across which sampling was conducted (*i.e.,* the entire continent), the residual genetic distances derived from the linear regression of pairwise genetic distances was checked against mismatch distances from the raw genetic data. Residual genetic distances are usually used when IBD is present in a dataset. It also ensures that large interpolation peaks are not over-inferred due to one or more samples being geographically isolated (*i.e.,* Tasmania, King Island, Kangaroo Island). In order to formally characterise the patterns observed in the Genetic Landscape Shape analysis, we performed post hoc linear regression analyses using longitude as the independent variable and various genetic diversity variables in Ostats v4.0.2 (http://www.maths.uq.edu.au/~mrb/ostats/index.php). As *D. n. novaehollandiae* do not belong to well-defined populations, samples were grouped by location according to Interim Biogeographic Regionalisation of Australia (IBRA) boundaries to simplify the calculation of pairwise genetic differences in this post-hoc analysis of correlations between genetic and geographic distances.

Phylogenetic and phylogeographic analyses

An unrooted haplotype network was also constructed using the Median Joining network in the software Network v4.6 (with gaps excluded; Bandelt *et al.* 1999), as initial data visualisation in SplitsTree v4 using the NeighborNet algorithm (Huson and Bryant 2006) indicated the Emu dataset was not treelike.

The evolutionary rate was estimated using the Bayesian algorithm in BEAST v1.6.1 (Drummond and Rambaut 2007). The genealogy was reconstructed under the HKY model. Although HKY+I was identified as the best fitting nucleotide substitution model using Akaike Information Criteria (AIC) and Bayesian Information Criterion implemented in ModelGenerator v0.85 (Keane *et al.* 2006), no proportion of invariant sites was included as this analysis was intra-specific (Ho *et al.* 2008). Posterior population parameters were estimated with two Markov chain Monte Carlo (MCMC) runs of 200,000,000 iterations each, with sampling every 20,000 iterations and discarding the first 2,000,000 as burnin. The effective sample size and convergence of the MCMC runs were checked in TRACER v1.5 (Drummond and Rambaut 2007). In order to test whether our dataset had sufficient phylogenetic and temporal signal from tip calibration ages to

estimate the evolutionary rate with any accuracy, 10 randomisation experiments were conducted using BEAST runs with identical priors as the optimal BEAST runs for our real dataset (Ho *et al.* 2011). This entailed creating 10 replicate datasets by randomising the tip ages across the tree using the randomisation function in Microsoft Excel v14.1.2. To rule out that the posterior distributions were being strongly influenced by the prior distributions, 'prior only' BEAST runs were done using the same conditions as the real dataset.

Population demographic structure

In order to investigate the demographic history of *Dromaius*, mismatch distributions of the modern and ancient eastern populations were generated using Arlequin v3.5 (Excoffier and Lischer 2010) to test for evidence of demographic expansion (Rogers and Harpending 1992) or spatial expansion (Excoffier 2004). Demographic expansion has been defined as a sudden increase in population size or sustained exponential growth (Rogers and Harpending 1992), while spatial expansion is defined as a range expansion in a subdivided population (Excoffier 2004). The ancient western samples were excluded, as the sample size was too low (n=5). The expansion dates from the tau statistic (in units of mutation) reported from Arlequin were transformed into units of years, using the evolutionary rate calculated from BEAST and the generation time. Generation time was set at 4 years, which was calculated using the formula $(T_1 + T_2)/2$; T_1 is age of first breeding (2 years; Davies *et al.* 1971); T_2 is age of last breeding (6 years based on age at death in the wild; Davies 2002). As an accurate generation time is particularly important in these calculations, it must be noted that the calculated expansion dates are based on these approximated breeding ages. The estimates of expansion times were plotted against the EPICA deuterium levels as measured from the Antarctic Vostok ice core (EPICA *et al.* 2004), as a proxy for southern hemisphere glacial/interglacial temperature cycles.

The demographic history of *Dromaius* was inferred also using the Bayesian coalescent approach as implemented in BEAST using the Bayesian Skyline Plot (BSP). Initially, panmixia was assumed for the entire dataset with the timing of coalescent events estimated by carbon dates of associated layers for *D. n. novaehollandiae* eggshell samples and the estimated island separation age for each island taxon (Kangaroo Island separated from the South Australian mainland 8,800 ya; Hope *et al.* 1977; King Island separated from Tasmania 12,000 ya and Tasmania separated from the Victorian mainland 14,000 ya; Lambeck and Chappell 2001). Due to the Holocene/Late Pleistocene age of these islands, the mechanism of formation (lowered sea level) is well understood, therefore *D. ater, D. baudinianus* and *D. n. diemenensis* were judged to represent larger source populations on the adjacent mainland just prior to separation of their island. To test *Dromaius* population expansion hypothesis against the null hypothesis of constant population size, we reconstructed the Emu genealogy assuming a constant population size (see BEAST description above) and compared it to BSP of the same dataset using Bayes Factors.

As BEAST is limited to reconstructing demographic histories of single populations under relatively simplistic demographic processes, we also modelled increasingly more complex scenarios using Bayesian Serial Simcoal (BayeSSC; Anderson *et al.* 2004) against which the observed data was compared. A null hypothesis of constant population size was initially modelled to test whether the pairwise differences, haplotype diversity, Tajima's D and F_{ST} statistic could be explained by stochastic effects from our sampling scheme. To implement this null hypothesis, all *Dromaius* were treated as a single population with constant size, as there is no evidence that the modern *D. n. novaehollandiae, D. ater, D. baudinianus* and *D. n. diemenensis* are distinct lineages at the mtDNA level. The generation time was calculated as for the expansion tests above. BayeSSC was then used to model increasingly more complex demographic scenarios, including different population growth scenarios as (1) a single population; (2) one population that became separated into two populations (west and east of the Nullarbor Plain) with some level of migration; or (3) two separate populations (west and east of the Nullarbor Plain) with some level of migration (see Figure 4). An initial BayeSSC run (100,000 simulations) was used to estimate posterior distributions of the parameters according to an Approximate Bayesian Computation (ABC) calculation. We used the observed values of the parameters shown in Table 2 (except Haplotype diversity, which was biased by a factor of (n-1)/n, and F_{ST} which was calculated as per BayeSSC), packages in R (locfit, akima, and lattice; R development core team 2011) and the reject function (available at http://www.stanford.edu/group/hadlylab/ssc/eval.r). The posterior distributions of these initial runs were used as the prior distributions for a second run for each demographic scenario. The demographic scenarios were tested against the null hypothesis using Akaike Information Criterion (AIC: Figure 4 & Table 1).

Supplementary Information

Morphometrics of island dwarf taxa vs. the mainland Emu

The linear relationship between *D. n. novaehollandiae* (largest), *D. baudinianus* (intermediate), and *D. ater* (smallest) is indicative of allometric scaling (Groeneveld *et al.* 2011), which was tested statistically using a Pearson correlation comparison of the shape indices (the ratio of one raw measurement over another raw measurement - which represents limb proportions) against the geometric mean for each individual (as a proxy for overall size; Table S6), and by calculating the PCA and Allometry ratio spectrums (Baur and Leuenberger 2011). The shape indices consistently identified as showing allometry were: for femur, shaft depth vs. proximal width (SD:PW); and for Tarso-metatarsi (TMT), length/shaft depth vs. distal width (L/SD:DW) (Table S2, and Figures S3 & S4). This indicates that these limb proportions contributed substantially to the variation of PC1. The PCA and Allometry ratio spectra were undertaken with log-shape variables and are therefore ‘size free’, and identifying shape differences independent of (isometric) size. The Pearson correlation comparison was done with raw shape indices, which again removes isometric size. However, this type of isometric size correction does not account for all shape differences associated with size; therefore, size adjusted data may still be influenced by allometry. Yet both types of analyses produced the same limb proportions as displaying shape differences.

Supplementary Tables and Figures

Table S1. Simple summary statistics of femur (F) and tarso-metatarsi (TMT) measurements of *D. n. novaehollandiae*, *D. baudinianus*, and *D. ater*, including estimation of the mean difference between *D. baudinianus* and *D. ater* using univariate ANOVA.

Femur	n	L	PW	PD	DW	DD	SW	SD
		Mean (mm) ± Std Dev	Mean (mm) ± Std Dev	Mean (mm) ± Std Dev	Mean (mm) ± Std Dev	Mean (mm) ± Std Dev	Mean (mm) ± Std Dev	Mean (mm) ± Std Dev
D. novaehollandiae (Mainland Emu)	8	232.5 (±8.5)	55.9 (±2.2)	60.0 (±2.1)	72.1 (±2.0)	71.5 (±2.5)	30.8 (±2.1)	28.9 (±2.4)
D. baudinianus (Kangaroo Island Emu)	13	177.9 (±5.9)	46.4 (±2.1)	45.4 (±2.8)	55.2 (±1.5)	52.8 (±1.3)	23.6 (±1.1)	21.8 (±1.3)
D. ater (King Island Emu)	16	164.8 (±11.6)	40.4 (±2.1)	38.9 (±3.7)	49.5 (±2.5)	47.5 (±2.8)	21.7 (±1.6)	18.0 (±1.6)
		Estimate (95% CI)	Estimate (95% CI)	Estimate (95% CI)	Estimate (95% CI)	Estimate (95% CI)	Estimate (95% CI)	Estimate (95% CI)
D. baudinianus – D. ater (Kangaroo Island Emu - King Island Emu)		13.1 (21.8, 4.4)*	6 (8.0, 4.0)*	6.5 (9.5, 3.6)*	5.7 (7.7, 3.7)*	5.4 (7.5, 3.2)*	1.9 (3.4, 0.4)*	3.7 (5.4, 2.1)*
Tarso-metatarsi	**n**	**L**	**PW**	**PD**	**DW**	**DD**	**SW**	**SD**
		Mean (mm) ± Std Dev	Mean (mm) ± Std Dev	Mean (mm) ± Std Dev	Mean (mm) ± Std Dev	Mean (mm) ± Std Dev	Mean (mm) ± Std Dev	Mean (mm) ± Std Dev
D. novaehollandiae (Mainland Emu)	8	435.9 (±20.1)	52.3 (±2.7)	41.7 (±1.7)	52.2 (±1.6)	27.4 (±1.4)	20.3 (±1.7)	22.9 (±1.8)
D. baudinianus (Kangaroo Island Emu)	13	329.1 (±6.9)	40.2 (±1.1)	32.5 (±1.6)	44.4 (±5.8)	21.9 (±0.8)	15.5 (±0.3)	17.1 (±0.9)
D. ater (King Island Emu)	12	295.0 (±27.5)	36.7 (±3.8)	27.0 (±3.5)	40.3 (±3.6)	17.7 (±2.8)	14.4 (±1.8)	14.2 (±1.7)
		Estimate (95% CI)	Estimate (95% CI)	Estimate (95% CI)	Estimate (95% CI)	Estimate (95% CI)	Estimate (95% CI)	Estimate (95% CI)
D. baudinianus – D. ater (Kangaroo Island Emu - King Island Emu)		45 (28.3, 61.6)*	3.54 (0.8, 6.3)*	5.52 (3.0, 8.1)*	4.11 (-0.3, 8.5)	4.17 (2.3, 6.1)*	1.11 (-0.3, 2.5)	2.87 (1.4, 4.4)*

L – Length; PW – Proximal Width; PD – Proximal Depth; DW – Distal Width; DD – Distal Depth; SW – Shaft Width; SD – Shaft Depth.

* - significant difference between *D. ater* and *D. baudinianus* measurements at the 95% level using a Bonferroni correction.

Table S2. Pearson correlation coefficient (r) of all indices (ratio of one raw measurement over another raw measurement) compared to the geometric mean for each individual, which represents different body proportions.

Femur	r	P-value	TMT	r	P	Tibia	r	P
PW/L	0.0550	0.1624	PW/L	0.4253	<0.0001	PW/L	0.3312	0.0002
PD/L	0.2663	0.0011	PD/L	0.0682	0.1421	PD/L	0.3704	<0.0001
DW/L	0.0196	0.4087	DW/L	0.7315*	<0.0001	DW/L	0.3079	0.0003
DD/L	0.2454	0.0018	DD/L	0.1026	0.0692	DD/L	0.0041	0.7008
SW/L	0.0072	0.6177	SW/L	0.2941	0.0011	SW/L	0.4861	<0.0001
SD/L	0.4171	<0.0001	SD/L	0.0119	0.5454	SD/L	0.0254	0.3394
L/PW	0.0657	0.1256	L/PW	0.4487	<0.0001	L/PW	0.3344	0.0001
PD/PW	0.3894	<0.0001	PD/PW	0.1885	0.0116	PD/PW	0.3658	<0.0001
DW/PW	0.1344	0.0256	DW/PW	0.4020	<0.0001	DW/PW	0.2739	0.0007
DD/PW	0.3478	0.0001	DD/PW	0.0921	0.0860	DD/PW	0.2726	0.0008
SW/PW	0.0854	0.0792	SW/PW	0.0093	0.5935	SW/PW	0.1682	0.0105
SD/PW	0.6011*	<0.0001	SD/PW	0.2387	0.0039	SD/PW	0.4231	<0.0001
L/PD	0.2586	0.0013	L/PD	0.0329	0.3126	L/PD	0.3693	<0.0001
PW/PD	0.3625	<0.0001	PW/PD	0.1983	0.0094	PW/PD	0.3583	<0.0001
DW/PD	0.2497	0.0016	DW/PD	0.3854	0.0001	DW/PD	0.3894	<0.0001
DD/PD	0.0631	0.1336	DD/PD	0.0044	0.7143	DD/PD	0.3862	<0.0001
SW/PD	0.1533	0.0165	SW/PD	0.0993	0.0741	SW/PD	0.4277	<0.0001
SD/PD	0.0350	0.2675	SD/PD	0.0124	0.5373	SD/PD	0.2926	0.0005
L/DW	0.0194	0.4108	L/DW	0.7720*	<0.0001	L/DW	0.3177	0.0002
PW/DW	0.1248	0.0320	PW/DW	0.4112	<0.0001	PW/DW	0.2840	0.0006
PD/DW	0.2453	0.0018	PD/DW	0.4447	<0.0001	PD/DW	0.4021	<0.0001
DD/DW	0.1894	0.0071	DD/DW	0.4402	<0.0001	DD/DW	0.1672	0.0108
SW/DW	0.0019	0.7969	SW/DW	0.1226	0.0457	SW/DW	0.1485	0.0169
SD/DW	0.2963	0.0005	SD/DW	0.5203*	<0.0001	SD/DW	0.3116	0.0003
L/DD	0.2156	0.0038	L/DD	0.0990	0.0746	L/DD	0.0026	0.7624
PW/DD	0.3170	0.0003	PW/DD	0.1062	0.0642	PW/DD	0.2936	0.0004
PD/DD	0.0543	0.1652	PD/DD	0.0000	0.9766	PD/DD	0.4260	<0.0001
DW/DD	0.1958	0.0061	DW/DD	0.4436	<0.0001	DW/DD	0.1738	0.0092
SW/DD	0.0851	0.0799	SW/DD	0.0906	0.0887	SW/DD	0.2790	0.0007
SD/DD	0.1686	0.0116	SD/DD	0.0254	0.3759	SD/DD	0.0369	0.2482
L/SW	0.0052	0.6710	L/SW	0.3603	0.0002	L/SW	0.5092*	<0.0001
PW/SW	0.0815	0.0868	PW/SW	0.0042	0.7205	PW/SW	0.1897	0.0063
PD/SW	0.1536	0.0164	PD/SW	0.0774	0.1170	PD/SW	0.4752	<0.0001
DW/SW	0.0022	0.7849	DW/SW	0.1309	0.0385	DW/SW	0.1538	0.0149
DD/SW	0.0959	0.0622	DD/SW	0.0959	0.0795	DD/SW	0.3130	0.0003
SD/SW	0.3776	<0.0001	SD/SW	0.1797	0.0140	SD/SW	0.4731	<0.0001
L/SD	0.4136	<0.0001	L/SD	0.0122	0.5411	L/SD	0.0307	0.2924
PW/SD	0.5840*	<0.0001	PW/SD	0.2427	0.0036	PW/SD	0.3691	<0.0001

PD/SD	0.0435	0.2156	PD/SD	0.0250	0.3792	PD/SD	0.2615	0.001
DW/SD	0.3009	0.0004	DW/SD	0.5015*	<0.0001	DW/SD	0.3127	0.0003
DD/SD	0.1733	0.0104	DD/SD	0.0341	0.3034	DD/SD	0.0370	0.2473
SW/SD	0.3571	<0.0001	SW/SD	0.1624	0.0201	SW/SD	0.4621	<0.0001

* - Correlation coefficients (r) >0.5, which indicates these measurements are influenced by allometry.

Table S3. Binomial test of statistical significance for continuity in the frequency of two of the three most common haplotypes (H3 and H7) between the ancient Western population and the modern Emu populations.

Haplotypes	***Proportion of haplotype in population***	***p-value***
Two common (H3, H7)	0.590	0.345
All haplotypes equal (1/14)	0.071	0.003**

* - P-value <0.05; ** - P-value <0.005

This Binomial test of significance assesses the likelihood of detecting three of the five Western ancient (eggshell) samples as two of the three most common haplotypes, if H3 and H7 were not already common at this time (12,500-16,500 ya). Assuming the selection of the eggshell samples from the cave deposit was independent (*i.e.,* they were not from the same egg/clutch and were not laid by the same individual emu), then the probability that the frequencies of H3 and H7 at 12,500-16,500 ya are the same as in modern populations is 0.345 (not significant). However, the probability that all 14 haplotypes present in the modern populations had equal frequencies at 12,500-16,500 ya is 0.003 (significant at the 5% level). Therefore, it is likely that at least H3 and H7 haplotypes were at high frequencies in the Western Emu population at 12,500-16,500 ya.

Table S4. Fixation index F_{ST} for two-population hypothesis (modern and ancient) as calculated in Arlequin.

	$WESTERN_M$	$EASTERN_M$	$WESTERN_A$	$EASTERN_A$
$WESTERN_M$	0			
$EASTERN_M$	0.01650	0		
$WESTERN_A$	0.09061	-0.00172	0	
$EASTERN_A$	0.31926**	0.16727**	0.02301	0

* - P-value <0.05; ** - P-value <0.005

M – modern samples; A – ancient samples

Table S5. Details of samples used in this study, with location data and source information.

Code	*ACAD #*	*Lat*	*Long*	*State*	*Area*	*Source*	*Accession #*	*H*	*Dates**
ATIDB	8689	-35.4	148.9	ACT	Tidbinbilla Nature Reserve	ANWC	7	2	Modern
KAIS01	8145	-36	136.93	SA	Kangaroo Island, Kelly Hill caves	SAM	B6814	3	8,800
KAIS02	8146	-36	136.9	SA	Kangaroo Island, Kelly Hill caves	SAM	B6816c	3	8,800
KAIS03	8147	-36	136.9	SA	Kangaroo Island, Kelly Hill caves	SAM	B6816f	3	8,800
KAIS04	8148	-36	136.9	SA	Kangaroo Island, Kelly Hill caves	SAM	B6839d	3	8,800
KAIS05	8149	-36	136.9	SA	Kangaroo Island, Kelly Hill caves	SAM	B6890c	4	8,800
KAIS06	10034	-36	136.9	SA	Kangaroo Island, Kelly Hill caves	SAM	B6816a	3	8,800
KAIS07	10035	-36	136.9	SA	Kangaroo Island, Kelly Hill caves	SAM	B6839c	3	8,800
KAIS08	10036	-36	136.9	SA	Kangaroo Island, Kelly Hill caves	SAM	B6890a	3	8,800
KAIS09	10037	-36	136.9	SA	Kangaroo Island, Kelly Hill caves	SAM	B11762a	3	8,800
KAIS10	10038	-36	136.9	SA	Kangaroo Island, Kelly Hill caves	SAM	B11762b	3	8,800
KAIS11	9794	-36	136.9	SA	Kangaroo Island	MV	P150273	3	8,800
KNGI07	9772	-39.9	143.9	TAS	King Island,	AM	F11021 (C)	2	12,000
KNGI09	9774	-39.9	143.9	TAS	King Island	MV	BE04808	2	12,000
KNGI10	9783	-39.9	143.9	TAS	King Island	MV	B22642	5	12,000
KNGI13	9791	-39.9	143.9	TAS	King Island	MV	P28816 PP	2	12,000
KNGI15	9793	-39.9	143.9	TAS	King Island, blowout in sand dunes	MV	P158629	2	12,000
KNGI16	9795	-39.9	143.9	TAS	King Island	MV	Uncatalogued (photo of 6 - bottom)	6	12,000
KNGI17	9796	-39.9	143.9	TAS	King Island, Porky River, material mixed	MV	Uncatalogued (photo of 6 - 2nd from bottom)	2	12,000
KNGI19	10960	-39.9	143.9	TAS	King Island	QM	O 10985	2	12,000
KNGI21		-39.9	143.9	TAS	King Island	Heupink	HQ910430	2	12,000
KNGI22		-39.9	143.9	TAS	King Island	Heupink	HQ910431	2	12,000
MAIN1	*9776*	-	-	-	-	*MV*	*B31685*	*2*	*Modern*
MAIN2	*9780*	-	-	*VIC*	-	*MV*	*R6947*	*2*	*Modern*
MAIN3	*9788*	-	-	-	-	*MV*	*B19058*	*7*	*Modern*
MAIN4	*9779*	-	-	*VIC*	-	*MV*	*R1815*	*8*	*Modern*
MAIN5	*9785*	-	-	-	-	*MV*	*B23642*	*7*	*Modern*
MAIN6	*9786*	-	-	-	-	*MV*	*B23643*	*2*	*Modern*
MAIN7						*Heupink*	*HQ910419*	*2*	*Modern*

MAIN8						*Heupink*	*HQ910420*	*2*	*Modern*
MAIN9						*Heupink*	*HQ910421*	*1*	*Modern*
MAIN10						*Heupink*	*HQ910422*	*7*	*Modern*
MAIN11						*Heupink*	*HQ910423*	*7*	*Modern*
NBLUE	8786	-33.6	150.3	NSW	Blue Mountains	AM	O.1141	2	63
NBOU1	8789	-30.3	146.1	NSW	Boorindal, Bourke	AM	O.28093	7	Modern
NBOU2	8790	-30.3	146.1	NSW	Boorindal, Bourke	AM	O.28094	7	25
NBOU3	10543	-30.3	146.1	NSW	Boorindal, Bourke	DERM	RW4	2	Modern
NCOON	9179	-30.9	149.4	NSW	Near Coonabarabrin	RK	EMU_NSW_29	2	Modern
NDENQ	8793	-35.5	145.0	NSW	Deniliquin	AM	O.53936	7	Modern
NFOR2	10562	-29.6	145.5	NSW	Fords Bridge, near Bourke	DERM	RW52	9	Modern
NFOW1	9777	-31.1	141.7	NSW	Fowler's Gap near Broken Hill	MV	B12823	3	Modern
NFOW2	9787	-31.1	141.7	NSW	Fowler's Gap near Broken Hill	MV	B12822	2	Modern
NFOW3	9781	-31.1	141.7	NSW	Fowler's Gap near Broken Hill	MV	B12821	3	Modern
NHUN1	10541	-29.5	144.4	NSW	South of Hungerford	DERM	RW2	2	Modern
NHUN2	10542	-29.1	144.6	NSW	Hungerford	DERM	RW3	1	Modern
NLKCO	8748	-33.6	147.4	NSW	Lake Cowal	AM	O.62409	6	Modern
NMIN1	8795	-29.8	153.2	NSW	Minnie Water Road	AM	O.70485	3	Modern
NMIN3	8797	-29.8	153.2	NSW	Minnie Water Road	AM	O.70487	3	Modern
NMIN4	8798	-29.8	153.2	NSW	Minnie Water Road	AM	O.70488	3	Modern
NMIN5	8801	-29.8	153.2	NSW	8km W of Minnie Water	AM	O.71099	3	Modern
NMORE	8788	-29.1	149.6	NSW	Garah, near Moree	AM	O.26392	1	30
NTUC	8802	-29.7	153.1	NSW	Tucabia, Summerville Road	AM	O.71207	3	Modern
NTYUE	*8808*	*-22.1*	*131.6*	*NT*	*Yuendumu*	*NTMAG*	*NTM T.2504*	*7*	*Modern*
NWALL	8800	-29.6	153.3	NSW	Gulmarrad, Wallaby Lane	AM	O.70526	3	Modern
NWOOL	8799	-29.9	153.3	NSW	Wooli, Rogers Bridge	AM	O.70489	3	Modern
NYATH	8691	-32.7	145.4	NSW	Yathong Nature Reserve, c. 130km S of Cobar	ANWC	31650	1	Modern
NYUR	8794	-29.9	153.2	NSW	Yuraygir National Park: North Illaroo	AM	O.65121	3	Modern
QCAPB	8792	-15.4	145.2	QLD	Cape Bedford	AM	O.38678	7	Modern
QCOOP	8690	-27.9	141.5	QLD	Cooper Creek Drainage, 50km SE of Nappa Merrie Station	ANWC	20822	2	Modern
QCUN4	10551	-28.1	145.9	QLD	Cunnamulla	DERM	RW12	1	Modern
QCUN6	10569	-28	146.1	QLD	Cunnamulla, 35km from Cunnamulla, between St George & Cunnamulla	DERM	RW59	2	Modern
QEUL11	10568	-28.1	144.4	QLD	Eulo, 60km from Thargomundah,	DERM	RW58	7	Modern

QEUL12	10570	-28	146.6	QLD	Cunnamulla, 90km from Cunnamulla, between St George & Cunnamulla	DERM	RW60	7	Modern
QEUL02	10554	-28.0	145.1	QLD	Eulo, north of Eulo	DERM	RW15	7	Modern
QEUL04	10556	-27	145.1	QLD	Eulo, north of Eulo	DERM	RW17	7	Modern
QEUL08	10560	-28	145.1	QLD	Eulo, north of Eulo	DERM	RW21	7	Modern
QEUL09	10566	-28.7	144.8	QLD	Eulo, south of Eulo on Eulo-Hundergord Rd	DERM	RW56	1	Modern
QNAR1	10548	-27.9	145.7	QLD	Cunnamulla, north of Cunnamulla	DERM	RW9	3	Modern
QQUL1	10546	-26.8	145.1	QLD	South of Chaepie	DERM	RW7	7	Modern
QROM1	10563	-26.6	148.8	QLD	Roma, 5km from Roma	DERM	RW53	3	Modern
QSHOA	8688	-22.8	150.5	QLD	North of Rockhampton, Shoalwater Bay Army Training Reserve	ANWC	44264	2	Modern
QSTG1	10564	-28.0	148.3	QLD	St George	DERM	RW54	2	Modern
QWIN1	10565	-25.4	142.5	QLD	Windorah, 18km W of Windorah	DERM	RW55	1	Modern
QWYA1	10550	-27.2	145.8	QLD	Wyandra, Mount Alfred near Wyandra	DERM	RW11	7	Modern
SBIL2	9673	-35.0	140.5	SA	Near Billiatt	AC	1	7	Modern
SFLIN1	8290	-31.5	138.6	SA	Flinders Ranges	SAM	ABTC83954	7	Modern
SFLIN2	8687	-31.5	138.6	SA	Flinders Ranges	SAM	ABTC83950	7	Modern
SGAWL	8283	-32.5	135.5	SA	Gawler Ranges	SAM	ABTC02818	7	Modern
SPAUG	8284	-32.1	137.9	SA	57k N Port Augusta	SAM	ABTC12924	1	Modern
STULK	8291	-34.8	135.7	SA	11.9k WNW Tulka	SAM	ABTC93845	1	Modern
TAS02	8803	-41.8	146.4	TAS	Tasmania	QVMA	#2	2	14,000
TAS04	8805	-41.8	146.4	TAS	Tasmania (Syntype)	BMNH	BMNH 1838.1.15.203	2	14,000
TAS06	8807	-41.8	146.4	TAS	Tasmania	BMNH	BMNH 1894.8.21.1	7	14,000
TAS07	9439	-41.8	146.4	TAS	Tasmania	QVMA	#3	2	14,000
VIC1	*9778*	-	-	*VIC*	-	*MV*	*R4240*	*7*	*47*
VPAT1	9775	35.3	142.2	VIC	Patchewollock North	MV	B12543	2	Modern
VWOO1	9782	-38.5	146.9	VIC	Woodside	MV	R6156	7	35
VWOO2	9789	-38.5	146.9	VIC	Woodside	MV	R6155	7	Modern
VWOO3	9799	-38.5	146.9	VIC	Woodside	MV	R6099	7	Modern
WBON1	8285	-30.4	118.3	WA	N Bonnie Rock	SAM	ABTC65041	7	Modern
WBON2	8682	-30.4	118.3	WA	N Bonnie Rock	SAM	ABTC65042	7	Modern
WDER	9800	-17.3	123.6	WA	Near Derby	MV	R12274	3	Modern
WGIDG	9171	-31.8	116.2	WA	Gidgegannup, 14 miles NE of Midland	WAM	8932	7	Modern
WGRAC	9173	-33.7	117.6	WA	Gracefield	WAM	29603	1	46
WLAT2	9175	-29.8	116.4	WA	Latham	WAM	35661	7	9
WMAR	8288	-31.5	119.5	WA	E Marvelock	SAM	ABTC65048	6	Modern
WMED	9176	-17.4	124	WA	Meda	WAM	35682	1	Modern
WMIL2	9165	-21.6	117.1	WA	Millstream	WAM	8324	7	Modern
WMIL3	9166	-21.6	117.1	WA	Millstream	WAM	8325	7	Modern
WMIL4	9167	-21.6	117.1	WA	Millstream	WAM	8326	7	Modern
WMIL5	9168	-21.6	117.1	WA	Millstream	WAM	8327	7	Modern

WMIL6	9169	-21.6	117.1	WA	Millstream	WAM	8328	7	Modern
WMOOR	9170	-30.5	116.2	WA	Moore River	WAM	8329	7	Modern
WMOSP	9172	-32.0	115.8	WA	Mosman Park, Perth	WAM	18614	1	Modern
WSTH1	8286	-31.2	119.3	WA	N Southern Cross	SAM	ABTC65043	1	Modern
WSTH2	8287	-31.2	119.3	WA	E Southern Cross	SAM	ABTC65046	7	Modern
WSTH3	8683	-31.2	119.3	WA	N Southern Cross	SAM	ABTC65044	2	Modern
WSTH4	8684	-31.2	119.3	WA	N Southern Cross	SAM	ABTC65045	7	Modern
WSTH5	8685	-31.2	119.4	WA	E Southern Cross	SAM	ABTC65047	1	Modern
WTUN1	9668	-34.1	115.0	WA	Tunnel Cave	JD	AD213	7	13000
WTUN2	9669	-34.1	115.0	WA	Tunnel Cave	JD	AD214	1	12500
WTUN6	9972	-34.1	115.0	WA	Tunnel Cave	JD	G10.79.7D	1	16500
WTUN8	9973	-34.1	115.0	WA	Tunnel Cave	JD	G10.59.7	3	16500
WTUN9	9974	-34.1	115.0	WA	Tunnel Cave	JD	G10.63.7A1	3	16500
WWSF1	8289	-31.3	118.7	WA	N Westonia Farmland	SAM	ABTC65049	1	Modern
WWSF2	8686	-31.3	118.7	WA	N Westonia Farmland	SAM	ABTC65050	3	Modern
WYUIN4	9932	-27.8	116.5	WA	Yuin station, near state barrier fence	RK	WA_EMU_4	3	Modern

* Dates are in B.P. (years before 1950 A.D.), either as radiocarbon dates, approximate date of island separation or museum collection date.

Samples in italics are those samples excluded from the analyses due to a lack of locality information, or in the case of the NTYUE specimen a lack of sample numbers.

Table S6. Primer details used.

Multiplex A		
A1360	Forward	GGGCTATGTATTATTGTACA
A1407	Reverse	CAATATATGGTTTGGGCACGTA
A1414	Forward	GGAACCCGTATTTCCCCTAA
A1361	Reverse	CTGTTAAAGGGGAACGTGGA
A1691	Forward	GCGTATTTGGGGAATCAT
A1692	Reverse	TGGTGTTTAGGAAATTAGAGG
A1695	Forward	TCACAAATTTATTATAGCAACTCC
A1363	Reverse	GTCCTTAGGTAGGGTGGGTA

Multiplex B		
A1410	Forward	TATATACGTGCCCAAACCAT
A1411	Reverse	GAGGATGGTGAATCAAGAAA
A1539	Forward	TCAACGTACCCCCATAAGTA
A1440	Reverse	AGAGCCCGAATGATTCCCCA
A1693	Forward	GCTTGATGGGCATAATTCTA
A1694	Reverse	ACTCCAGTACTGATGACAAAAA

Table S7. Raw measurements and shape indices of *Dromaius* limb bones used in Principal Component Analysis (PCA).

	TIBIA				*Raw measurements*								*Shape indices*			
Indiv.	*Specimen*	*Location*	*L*	*PW*	*PD*	*DW*	*DD*	*SW*	*SD*	*L/GM*	*PW/GM*	*PD/GM*	*DW/GM*	*DD/GM*	*SW/GM*	*SD/GM*
MN01	B12543	Mainland	440	52.6	84.7	47.1	42.1	27.5	24	7.425	0.888	1.429	0.795	0.71	0.464	0.405
MN02	B12822	Mainland	420	52.9	84.1	44.7	41.6	24.4	21.5	7.442	0.937	1.49	0.792	0.737	0.432	0.381
MN03	B12823	Mainland	423	52.5	84	47.9	44.16	25.4	24.8	7.171	0.89	1.424	0.812	0.749	0.431	0.42
MN04	B19058	Mainland	455	55.3	88.9	49.6	43.8	28	25.7	7.346	0.893	1.435	0.801	0.707	0.452	0.415
MN05	B31685	Mainland	444	51.3	86.2	46.6	42.7	26.5	23.2	7.564	0.874	1.468	0.794	0.727	0.451	0.395
MN06	R1815	Mainland	445	54	89.3	47.4	43.2	26.3	24	7.426	0.901	1.49	0.791	0.721	0.439	0.401
MN07	R4240	Mainland	400	51.6	86.1	44.9	41.8	26.6	27	6.817	0.879	1.467	0.765	0.712	0.453	0.46
MN08	R6947	Mainland	460	55.5	86.6	50.6	47.4	27.6	27.6	7.275	0.878	1.37	0.8	0.75	0.437	0.437
KA30	6814a	Kangaroo Is.	332	39.6	61.5	37.3	34.3	22	18	7.282	0.869	1.349	0.818	0.752	0.483	0.395
KA14	11762a	Kangaroo Is.	317	40.2	60.4	34.5	32.4	21.4	17.2	7.213	0.915	1.374	0.785	0.737	0.487	0.391
KA15	11762b	Kangaroo Is.	337	42.3	66.2	39.6	37.2	23.2	18	7.032	0.883	1.381	0.826	0.776	0.484	0.376
KA16	11762c	Kangaroo Is.	340	43	66.7	39.8	37	23.2	18.7	7.023	0.888	1.378	0.822	0.764	0.479	0.386
KA17	6814b	Kangaroo Is.	330	39.4	62.2	36	34.3	22.3	17.5	7.289	0.87	1.374	0.795	0.758	0.493	0.387
KA18	6816a	Kangaroo Is.	330	41.7	67.4	37.5	36.6	22	17	7.084	0.895	1.447	0.805	0.786	0.472	0.365
KA19	6816b	Kangaroo Is.	335	40.5	66.8	37.5	35.6	20.3	17	7.327	0.886	1.461	0.82	0.779	0.444	0.372
KA20	6816c	Kangaroo Is.	335	41	62	36	33.4	21.5	18	7.382	0.904	1.366	0.793	0.736	0.474	0.397
KA21	6816d	Kangaroo Is.	338	41.6	63.3	37	33	21.3	18	7.395	0.91	1.385	0.81	0.722	0.466	0.394
KA22	6816f	Kangaroo Is.	328	41.6	65.4	36.6	35	22.4	17.4	7.108	0.902	1.417	0.793	0.758	0.485	0.377
KA23	6839a	Kangaroo Is.	319	37	62.3	36	32.5	20	16.5	7.373	0.855	1.44	0.832	0.751	0.462	0.381
KA24	6839b	Kangaroo Is.	320	35.5	61.1	36	32.5	20	17.8	7.377	0.818	1.408	0.83	0.749	0.461	0.41
KA25	6839c	Kangaroo Is.	320	41.1	62.7	33.8	32.4	21.4	17.2	7.231	0.929	1.417	0.764	0.732	0.484	0.389
KA26	6839d	Kangaroo Is.	328	42.3	63.6	36.5	32.8	21.1	17.1	7.268	0.937	1.409	0.809	0.727	0.468	0.379
KA27	6890a	Kangaroo Is.	330	42.5	67.4	37.2	35.3	21.5	18	7.075	0.911	1.445	0.798	0.757	0.461	0.386
KA28	6890b	Kangaroo Is.	330	42	68	36.7	34.3	21.6	18	7.116	0.906	1.466	0.791	0.74	0.466	0.388
KA29	6890c	Kangaroo Is.	326	41.4	64.7	35.7	34	23	18	7.08	0.899	1.405	0.775	0.738	0.5	0.391
KI17	F11023L	King Is.	275	54.9	32.7	33.1	28.2	19.7	15.1	7.052	1.408	0.839	0.849	0.723	0.505	0.387
KI18	F11023M	King Is.	265	48.8	29.3	31.2	24.3	19.7	13.6	7.38	1.359	0.816	0.869	0.677	0.549	0.379
KI19	F11024L	King Is.	270	51.3	31.7	35	30.2	19.4	16.3	6.857	1.303	0.805	0.889	0.767	0.493	0.414
KI20	F11024M	King Is.	280	50	33	33.1	27	19.7	15.5	7.267	1.298	0.856	0.859	0.701	0.511	0.402
KI21	F15560	King Is.	350	66.5	43	41.2	36.2	22.5	19.1	7.2	1.368	0.885	0.848	0.745	0.463	0.393
KI22	F15561	King Is.	310	59.4	38.2	36.5	27.5	23.8	16.8	7.17	1.374	0.884	0.844	0.636	0.551	0.389
KI23	F15562	King Is.	260	51.5	31	33.3	27.9	18.5	17.9	6.736	1.334	0.803	0.863	0.723	0.479	0.464
KI24	F15563	King Is.	275	52.4	32.2	32.9	26.4	19.6	15.2	7.187	1.369	0.842	0.86	0.69	0.512	0.397
KI25	F15564	King Is.	325	57	35.7	37.4	32.8	23	17.6	7.357	1.29	0.808	0.847	0.742	0.521	0.398
KI26	F15565	King Is.	310	58.3	38	34.1	31.7	22.7	18.3	7.08	1.332	0.868	0.779	0.724	0.518	0.418
KI27	F15566	King Is.	315	63.7	39.8	38.7	32.2	21	16.1	7.106	1.437	0.898	0.873	0.726	0.474	0.363
KI28	F15567	King Is.	315	61.1	38.8	36.8	33.3	23.1	16.7	7.057	1.369	0.869	0.824	0.746	0.518	0.374
KI29	F16468	King Is.	285	51	33.9	32.1	28.6	18.6	14.3	7.446	1.332	0.886	0.838	0.747	0.486	0.374

	TIBIA				Raw measurements								Shape indices			
Indiv.	***Specimen***	***Location***	***L***	***PW***	***PD***	***DW***	***DD***	***SW***	***SD***	***L/GM***	***PW/GM***	***PD/GM***	***DW/GM***	***DD/GM***	***SW/GM***	***SD/GM***
MN01	B12543	Mainland	440	52.6	84.7	47.1	42.1	27.5	24	7.425	0.888	1.429	0.795	0.71	0.464	0.405
MN02	B12822	Mainland	420	52.9	84.1	44.7	41.6	24.4	21.5	7.442	0.937	1.49	0.792	0.737	0.432	0.381
MN03	B12823	Mainland	423	52.5	84	47.9	44.16	25.4	24.8	7.171	0.89	1.424	0.812	0.749	0.431	0.42
MN04	B19058	Mainland	455	55.3	88.9	49.6	43.8	28	25.7	7.346	0.893	1.435	0.801	0.707	0.452	0.415
MN05	B31685	Mainland	444	51.3	86.2	46.6	42.7	26.5	23.2	7.564	0.874	1.468	0.794	0.727	0.451	0.395
MN06	R1815	Mainland	445	54	89.3	47.4	43.2	26.3	24	7.426	0.901	1.49	0.791	0.721	0.439	0.401
MN07	R4240	Mainland	400	51.6	86.1	44.9	41.8	26.6	27	6.817	0.879	1.467	0.765	0.712	0.453	0.46
MN08	R6947	Mainland	460	55.5	86.6	50.6	47.4	27.6	27.6	7.275	0.878	1.37	0.8	0.75	0.437	0.437
KA30	6814a	Kangaroo Is.	332	39.6	61.5	37.3	34.3	22	18	7.282	0.869	1.349	0.818	0.752	0.483	0.395
KA14	11762a	Kangaroo Is.	317	40.2	60.4	34.5	32.4	21.4	17.2	7.213	0.915	1.374	0.785	0.737	0.487	0.391
KA15	11762b	Kangaroo Is.	337	42.3	66.2	39.6	37.2	23.2	18	7.032	0.883	1.381	0.826	0.776	0.484	0.376
KA16	11762c	Kangaroo Is.	340	43	66.7	39.8	37	23.2	18.7	7.023	0.888	1.378	0.822	0.764	0.479	0.386
KA17	6814b	Kangaroo Is.	330	39.4	62.2	36	34.3	22.3	17.5	7.289	0.87	1.374	0.795	0.758	0.493	0.387
KA18	6816a	Kangaroo Is.	330	41.7	67.4	37.5	36.6	22	17	7.084	0.895	1.447	0.805	0.786	0.472	0.365
KA19	6816b	Kangaroo Is.	335	40.5	66.8	37.5	35.6	20.3	17	7.327	0.886	1.461	0.82	0.779	0.444	0.372
KA20	6816c	Kangaroo Is.	335	41	62	36	33.4	21.5	18	7.382	0.904	1.366	0.793	0.736	0.474	0.397
KA21	6816d	Kangaroo Is.	338	41.6	63.3	37	33	21.3	18	7.395	0.91	1.385	0.81	0.722	0.466	0.394
KA22	6816f	Kangaroo Is.	328	41.6	65.4	36.6	35	22.4	17.4	7.108	0.902	1.417	0.793	0.758	0.485	0.377
KA23	6839a	Kangaroo Is.	319	37	62.3	36	32.5	20	16.5	7.373	0.855	1.44	0.832	0.751	0.462	0.381
KA24	6839b	Kangaroo Is.	320	35.5	61.1	36	32.5	20	17.8	7.377	0.818	1.408	0.83	0.749	0.461	0.41
KA25	6839c	Kangaroo Is.	320	41.1	62.7	33.8	32.4	21.4	17.2	7.231	0.929	1.417	0.764	0.732	0.484	0.389
KA26	6839d	Kangaroo Is.	328	42.3	63.6	36.5	32.8	21.1	17.1	7.268	0.937	1.409	0.809	0.727	0.468	0.379
KA27	6890a	Kangaroo Is.	330	42.5	67.4	37.2	35.3	21.5	18	7.075	0.911	1.445	0.798	0.757	0.461	0.386
KA28	6890b	Kangaroo Is.	330	42	68	36.7	34.3	21.6	18	7.116	0.906	1.466	0.791	0.74	0.466	0.388
KA29	6890c	Kangaroo Is.	326	41.4	64.7	35.7	34	23	18	7.08	0.899	1.405	0.775	0.738	0.5	0.391
KI17	F11023L	King Is.	275	54.9	32.7	33.1	28.2	19.7	15.1	7.052	1.408	0.839	0.849	0.723	0.505	0.387
KI18	F11023M	King Is.	265	48.8	29.3	31.2	24.3	19.7	13.6	7.38	1.359	0.816	0.869	0.677	0.549	0.379
KI19	F11024L	King Is.	270	51.3	31.7	35	30.2	19.4	16.3	6.857	1.303	0.805	0.889	0.767	0.493	0.414
KI20	F11024M	King Is.	280	50	33	33.1	27	19.7	15.5	7.267	1.298	0.856	0.859	0.701	0.511	0.402
KI21	F15560	King Is.	350	66.5	43	41.2	36.2	22.5	19.1	7.2	1.368	0.885	0.848	0.745	0.463	0.393
KI22	F15561	King Is.	310	59.4	38.2	36.5	27.5	23.8	16.8	7.17	1.374	0.884	0.844	0.636	0.551	0.389
KI23	F15562	King Is.	260	51.5	31	33.3	27.9	18.5	17.9	6.736	1.334	0.803	0.863	0.723	0.479	0.464
KI24	F15563	King Is.	275	52.4	32.2	32.9	26.4	19.6	15.2	7.187	1.369	0.842	0.86	0.69	0.512	0.397
KI25	F15564	King Is.	325	57	35.7	37.4	32.8	23	17.6	7.357	1.29	0.808	0.847	0.742	0.521	0.398
KI26	F15565	King Is.	310	58.3	38	34.1	31.7	22.7	18.3	7.08	1.332	0.868	0.779	0.724	0.518	0.418
KI27	F15566	King Is.	315	63.7	39.8	38.7	32.2	21	16.1	7.106	1.437	0.898	0.873	0.726	0.474	0.363
KI28	F15567	King Is.	315	61.1	38.8	36.8	33.3	23.1	16.7	7.057	1.369	0.869	0.824	0.746	0.518	0.374
KI29	F16468	King Is.	285	51	33.9	32.1	28.6	18.6	14.3	7.446	1.332	0.886	0.838	0.747	0.486	0.374

TARSO-METATARSI			Raw measurements							Shape indices						
Indiv.	Specimen	Location	L	PW	PD	DW	DD	SW	SD	L/GM	PW/GM	PD/GM	DW/GM	DD/GM	SW/GM	SD/GM
MN01	B12543	Mainland	390	49.9	39.5	52.1	29.2	19.7	25.4	8.129	1.04	0.823	1.086	0.609	0.411	0.529
MN02	B12822	Mainland	380	49.8	41.8	50.5	26.3	20.3	24.5	8.05	1.055	0.885	1.07	0.557	0.43	0.519
MN03	B12823	Mainland	380	51.6	41.5	51.5	27	18.4	22.2	8.192	1.112	0.895	1.11	0.582	0.397	0.479
MN04	B19058	Mainland	383	55.5	44.4	53.3	26.4	23	22	7.827	1.134	0.907	1.089	0.539	0.47	0.45
MN05	B31685	Mainland	390	49.7	40.3	50.4	27.3	18.7	21.2	8.506	1.084	0.879	1.099	0.595	0.408	0.462
MN06	R1815	Mainland	395	54.1	41.3	54.4	28.9	19.4	21.2	7.079	0.97	0.74	0.975	0.518	0.348	0.38
MN07	R4240	Mainland	370	51.3	40.8	51.3	25.3	22.5	21.8	7.903	1.096	0.871	1.096	0.54	0.481	0.466
MN08	R6947	Mainland	415	56.3	43.8	54.1	28.8	20.4	25	8.253	1.12	0.871	1.076	0.573	0.406	0.497
KA31	6814	Kangaroo Is.	274	38.7	31.6	42.4	21.9	15.4	17.3	7.55	1.066	0.871	1.168	0.603	0.424	0.477
KA32	6814a	Kangaroo Is.	272	38.6	32	42.9	21.6	15.4	16.2	7.565	1.074	0.89	1.193	0.601	0.428	0.451
KA33	11764a	Kangaroo Is.	273	41.1	33.4	43.4	21.7	15.1	16.3	7.473	1.125	0.914	1.188	0.594	0.413	0.446
KA34	11764b	Kangaroo Is.	287	42	35.4	43.6	22.9	16	18	7.478	1.094	0.922	1.136	0.597	0.417	0.469
KA35	11764d	Kangaroo Is.	287	42	35.4	43.9	23	16	18.7	7.425	1.087	0.916	1.136	0.595	0.414	0.484
KA36	11764e	Kangaroo Is.	285	39	30	43	23	16	16.5	7.799	1.067	0.821	1.177	0.629	0.438	0.451
KA37	6817a	Kangaroo Is.	285	40.6	32.2	63.1	23	15.6	16.4	7.299	1.04	0.825	1.616	0.589	0.4	0.42
KA38	6817b	Kangaroo Is.	282	39.5	32.2	43.3	21.3	15.2	18.6	7.638	1.07	0.872	1.173	0.577	0.412	0.504
KA39	6817c	Kangaroo Is.	277	40.5	32.3	43.4	21.2	15.8	16	7.615	1.113	0.888	1.193	0.583	0.434	0.44
KA40	6817d	Kangaroo Is.	282	39.5	32	43.7	20.9	15.3	17.2	7.734	1.083	0.878	1.199	0.573	0.42	0.472
KA41	6817e	Kangaroo Is.	278	40.6	30.4	43.3	20.9	15.3	17.5	7.657	1.118	0.837	1.193	0.576	0.421	0.482
KA42	6817f	Kangaroo Is.	272	41	33.7	43.1	21.3	15.2	16.3	7.462	1.125	0.925	1.182	0.584	0.417	0.447
KA43	6840c	Kangaroo Is.	276	40	31.4	38	21.8	15.7	17	7.691	1.115	0.875	1.059	0.607	0.437	0.474
KI30	F11025-1	King Is.	225	37.8	25.7	41.2	17.3	18.3	13.2	6.94	1.166	0.793	1.271	0.534	0.564	0.407
KI31	F11025-4	King Is.	215	36.8	26.1	36.8	14.9	13.4	14.1	7.189	1.231	0.873	1.231	0.498	0.448	0.471
KI32	F11025-5	King Is.	235	36.5	28.4	39.8	17.5	14.4	13.3	7.403	1.15	0.895	1.254	0.551	0.454	0.419
KI33	F11025-6	King Is.	253	33.5	23.4	38.7	17.5	13.5	16	8.1	1.072	0.749	1.239	0.56	0.432	0.512
KI34	F15594	King Is.	220	38	26.2	41.3	18.1	13.3	15.2	6.912	1.194	0.823	1.298	0.569	0.418	0.478
KI35	F15595	King Is.	275	38.1	28.4	42.6	20.8	16.1	15.6	7.825	1.084	0.808	1.212	0.592	0.458	0.444
KI36	F15598	King Is.	220	36	28.4	35.5	13.5	13.2	14.4	7.402	1.211	0.956	1.194	0.454	0.444	0.484
KI37	F15600	King Is.	205	29.7	18.1	36.6	15.1	15	10.8	7.656	1.109	0.676	1.367	0.564	0.56	0.403
KI38	F15601	King Is.	245	36.1	28.9	43.7	19.1	13.3	15.2	7.408	1.092	0.875	1.321	0.579	0.402	0.46
KI39	F15602	King Is.	280	46.3	32	48.4	23.7	17.1	17	7.172	1.186	0.82	1.24	0.608	0.437	0.435
KI40	F16464	King Is.	223	37.1	29.8	41.2	16	12.5	12.7	7.26	1.208	0.97	1.341	0.521	0.407	0.413
KI41	F16471	King Is.	215	34.4	27.9	37.6	19	13.1	13	7.027	1.126	0.912	1.228	0.621	0.428	0.425

Indiv. – Unique short identifier name for each individual. GM – Geometric Mean = $(L \times PW \times PD \times DW \times DD \times SW \times SD)^{1/7}$

L – Overall length. PW – Proximal width. PD – Proximal depth. DW – Distal width. DD – Distal depth. SW – Shaft width. SD – Shaft depth.

Table S7. Summary statistics of *Dromaius* limb bones, Femur, Tibia, and Tarso-metatarsi.

	Femur	*n*	*Mean*	*Median*	*Minimum*	*Maximum*	*CoV*
Mainland	Length	8	232.50	230.0	220.0	245.0	0.0364
	Proximal Width	8	55.93	55.5	53.3	59.8	0.0401
	Proximal Depth	8	60.04	59.2	58.2	63.9	0.0345
	Distal Width	8	72.13	71.8	70.2	76.4	0.0279
	Distal Depth	8	71.46	70.7	68.5	75.8	0.0349
	Shaft Width	8	30.76	30.8	27.5	34.5	0.0695
	Shaft Depth	8	28.89	29.1	25.7	32.4	0.0814
Kangaroo Island	Length	13	177.85	180.0	165.0	185.0	0.0334
	Proximal Width	13	46.38	46.4	42.6	49.6	0.0449
	Proximal Depth	13	45.44	46.0	39.4	48.2	0.0623
	Distal Width	13	55.17	55.0	52.5	58.0	0.0274
	Distal Depth	13	52.85	52.5	51.2	55.0	0.0237
	Shaft Width	13	23.62	23.5	21.6	25.7	0.0478
	Shaft Depth	13	21.78	22.0	19.8	24.0	0.0604
King Island	Length	16	164.75	160.0	150.0	190.0	0.0704
	Proximal Width	16	40.40	40.2	35.7	43.5	0.0509
	Proximal Depth	16	38.92	38.6	33.5	44.9	0.0948
	Distal Width	16	49.48	49.2	45.8	54.5	0.0513
	Distal Depth	16	47.47	47.3	42.5	52.1	0.0587
	Shaft Width	16	21.70	21.4	19.4	24.2	0.0716
	Shaft Depth	16	18.03	18.0	15.3	22.0	0.0904
	Tibia	***n***	***Mean***	***Median***	***Minimum***	***Maximum***	***CoV***
Mainland	Length	8	435.88	442.0	400.0	460.0	0.0460
	Proximal Width	8	53.21	52.8	51.3	55.5	0.0297
	Proximal Depth	8	86.24	86.2	84.0	89.3	0.0234
	Distal Width	8	47.35	47.3	44.7	50.6	0.0434
	Distal Depth	8	43.35	43.0	41.6	47.4	0.0434
	Shaft Width	8	26.54	26.6	24.4	28.0	0.0453
	Shaft Depth	8	24.73	24.4	21.5	27.6	0.0811
Kangaroo Island	Length	18	329.44	330.0	317.0	340.0	0.0208
	Proximal Width	18	40.73	41.3	35.5	43.0	0.0471
	Proximal Depth	18	64.32	64.2	60.4	68.0	0.0384
	Distal Width	18	36.75	36.7	33.8	39.8	0.0404
	Distal Depth	18	34.37	34.3	32.4	37.2	0.0476
	Shaft Width	18	21.59	21.5	20.0	23.2	0.0463
	Shaft Depth	18	17.59	17.7	16.5	18.7	0.0311
King Island	Length	13	295.00	285.0	260.0	350.0	0.0931
	Proximal Width	13	55.84	54.9	48.8	66.5	0.1011
	Proximal Depth	13	35.18	33.9	29.3	43.0	0.1156
	Distal Width	13	35.03	34.1	31.2	41.2	0.0831
	Distal Depth	13	29.72	28.6	24.3	36.2	0.1125
	Shaft Width	13	20.87	19.7	18.5	23.8	0.0905
	Shaft Depth	13	16.35	16.3	13.6	19.1	0.0983

	Tarso-metatarsi	*n*	*Mean*	*Median*	*Minimum*	*Maximum*	*CoV*
Mainland	Length	8	387.88	386.5	370.0	415.0	0.0346
	Proximal Width	8	52.28	51.5	49.7	56.3	0.0510
	Proximal Depth	8	41.68	41.4	39.5	44.4	0.0401
	Distal Width	8	52.20	51.8	50.4	54.4	0.0299
	Distal Depth	8	27.40	27.2	25.3	29.2	0.0521
	Shaft Width	8	20.30	20.0	18.4	23.0	0.0822
	Shaft Depth	8	22.91	22.1	21.2	25.4	0.0765
Kangaroo Island	Length	13	279.23	278.0	272.0	287.0	0.0204
	Proximal Width	13	40.24	40.5	38.6	42.0	0.0283
	Proximal Depth	13	32.46	32.2	30.0	35.4	0.0507
	Distal Width	13	44.39	43.3	38.0	63.1	0.1311
	Distal Depth	13	21.88	21.7	20.9	23.0	0.0373
	Shaft Width	13	15.54	15.4	15.1	16.0	0.0213
	Shaft Depth	13	17.08	17.0	16.0	18.7	0.0534
King Island	Length	12	234.25	224.0	205.0	280.0	0.1033
	Proximal Width	12	36.70	36.7	29.7	46.3	0.1047
	Proximal Depth	12	26.95	28.2	18.1	32.0	0.1317
	Distal Width	12	40.28	40.5	35.5	48.4	0.0899
	Distal Depth	12	17.72	17.5	13.5	23.7	0.1573
	Shaft Width	12	14.43	13.5	12.5	18.3	0.1262
	Shaft Depth	12	14.21	14.3	10.8	17.0	0.1203

n – number of samples;

CoV – Coefficient of Variation (a measure of the variability of the data).

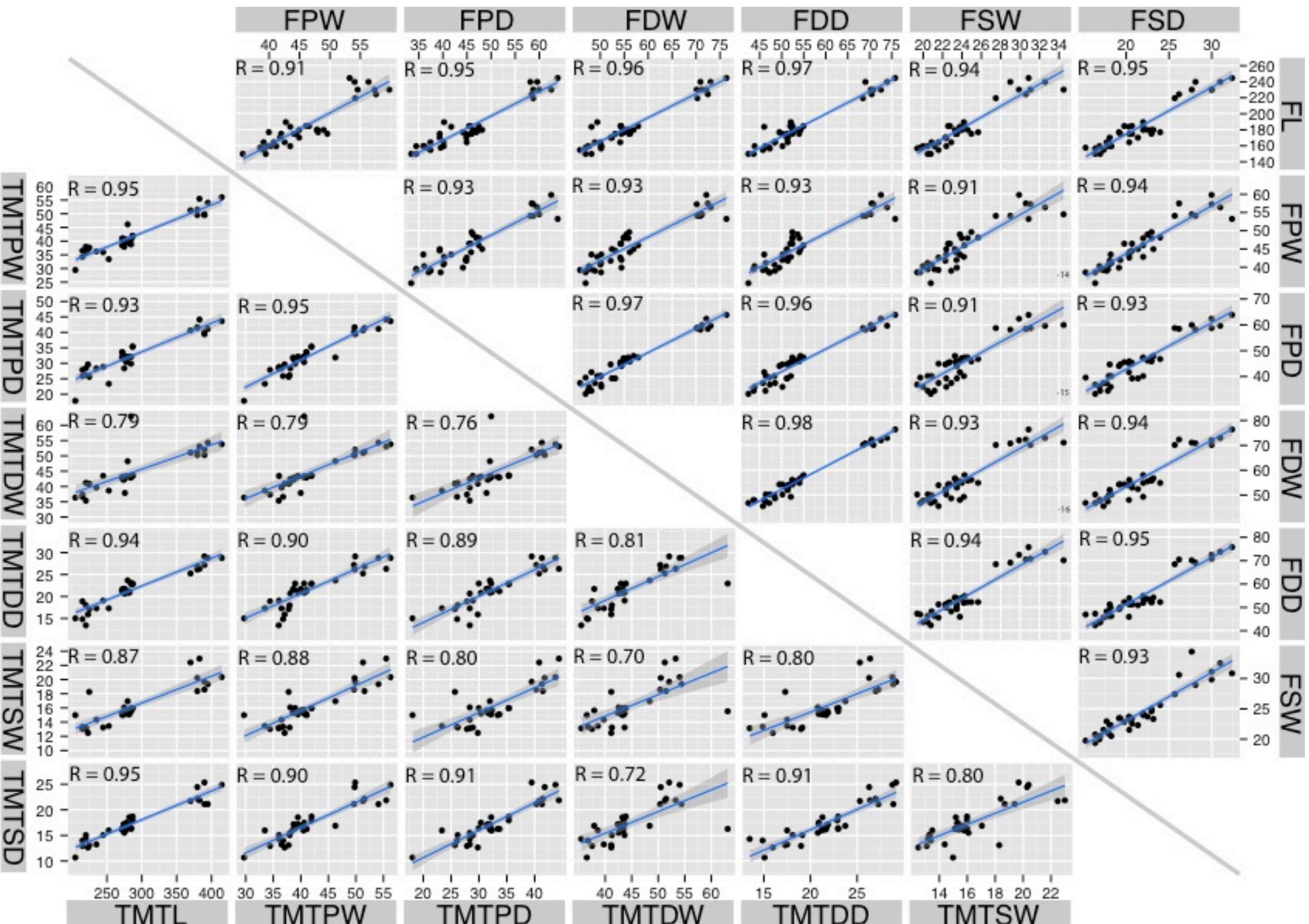


Figure S1. Matrix of scatterplots showing pairwise linear regression analyses of *D. ater, D. baudinianus* and *D. n. novaehollandiae* limb bone raw measurements, with femur measurements above the diagonal and tarso-metatarsi measurements below the diagonal (the measurement used in each pairwise comparison is identified in the margins). The Pearson correlation coefficient between the pairwise measurements is noted in upper left of each panel, and all the P-values < 0.0024 (the Bonferroni corrected P-value equivalent of 0.05). The linear arrangement of the relationship between *D. ater, D. baudinianus* and *D. novaehollandiae* in each panel indicates allometry (Groeneveld *et al.* 2011).

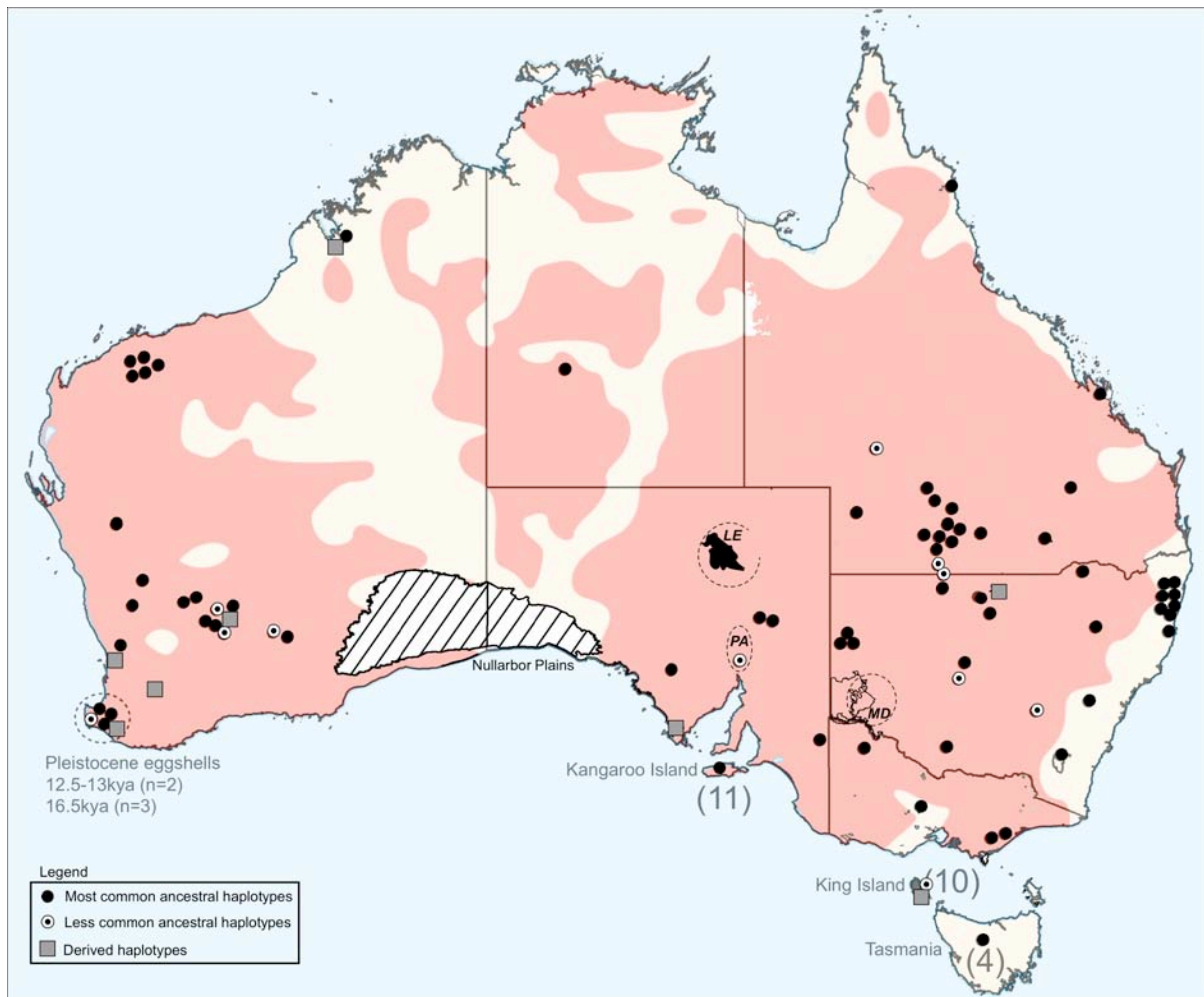


Figure S2. Map showing modern Emu distributions (modified from Blakers *et al.* 1985) with the dark colouration indicating areas where Emus have been recorded and the samples in this study marked by circles and squares. The Lake Eyre (LE), Port Augusta (PA), and Murray-Darling (MD) basins where Miller *et al.* (1999; 2005) sampled ancient Emu eggshell from up to 140 kya are noted.

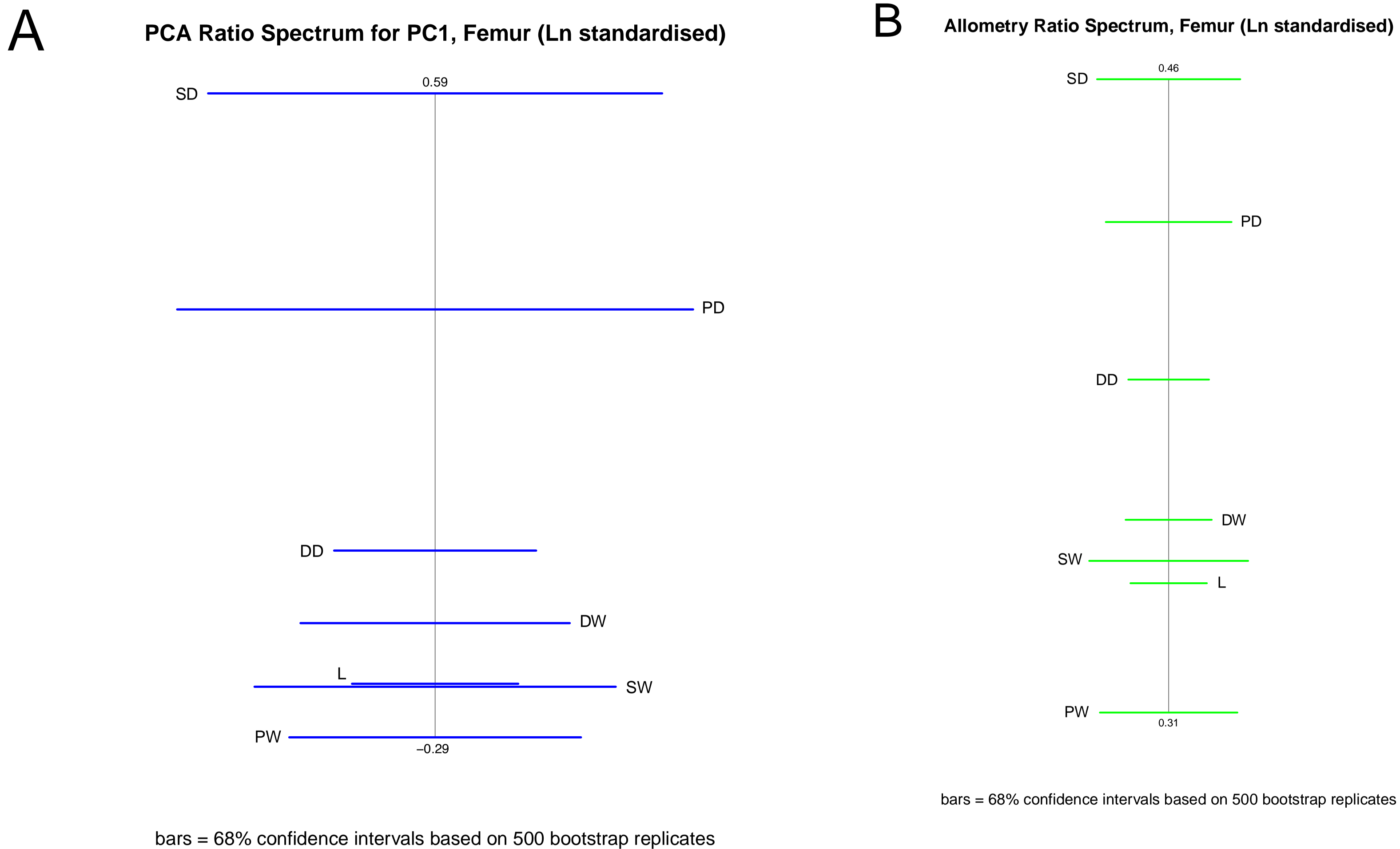
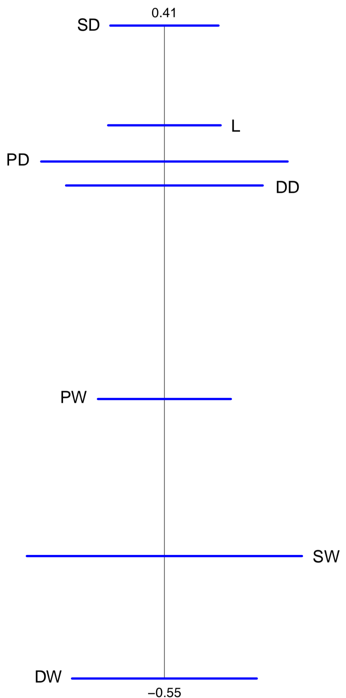


Figure S3. PCA ratio spectrum and Allometry ratio spectrum for PC1 of femur dataset. **The PCA ratio spectrum indicates the contribution that different body proportions play in the variation of PC1; distant variables on the spectrum indicate a major contribution to the variation of PC1 while variables close together indicate little or no contribution to PC1. The Allometry ratio spectrum informs on the body proportions that show strongest allometric growth; distant variables on the spectrum indicate the body proportions whose covariance with overall body size is maximal, while those variables close together indicate little or no allometric growth.**

A



B

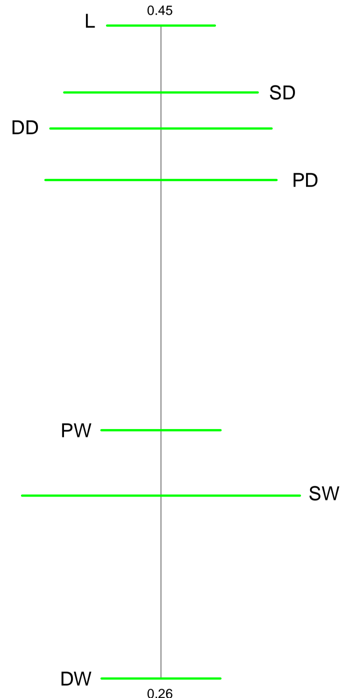
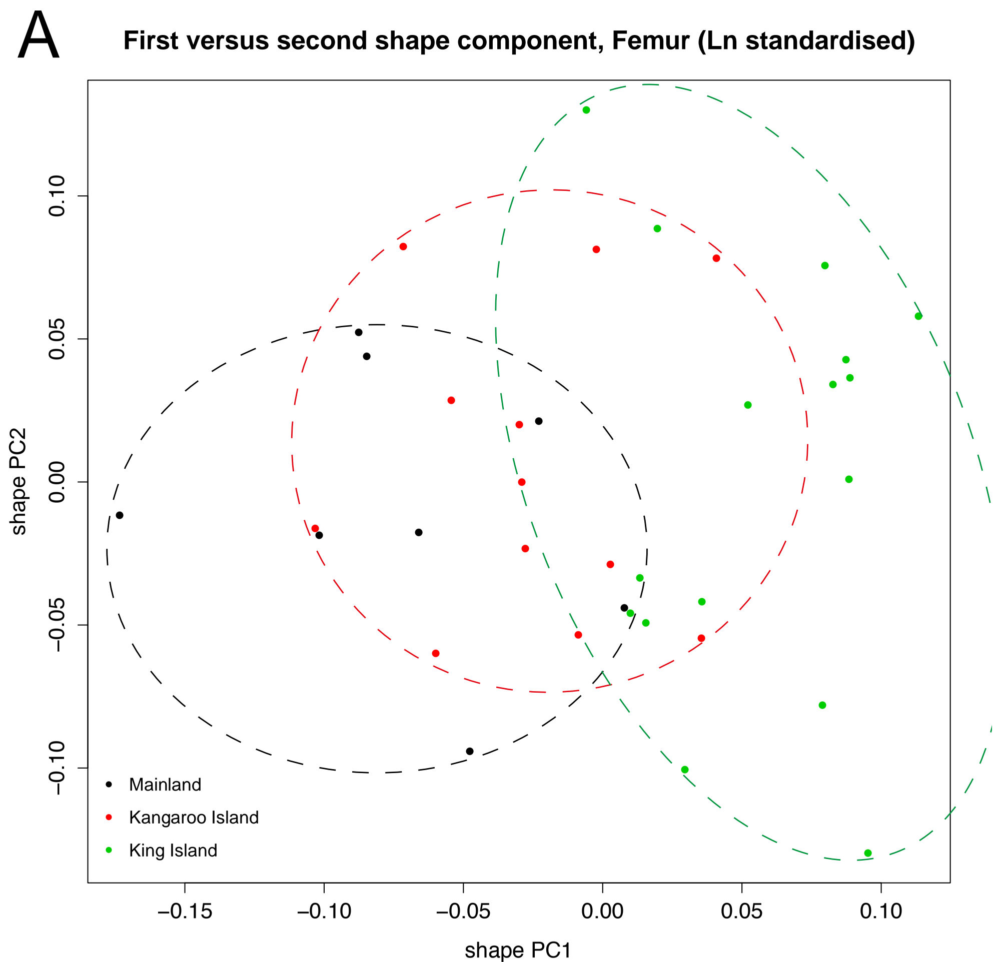


Figure S4. PCA ratio spectrum and Allometry ratio spectrum for PC1 of TMT dataset. The PCA ratio spectrum indicates the contribution that different body proportions play in the variation of PC1; distant variables on the spectrum indicate a major contribution to the variation of PC1 while variables close together indicate little or no contribution to PC1. The Allometry ratio spectrum informs on the body proportions that show strongest allometric growth; distant variables on the spectrum indicate the body proportions whose covariance with overall body size is maximal. while those variables close together indicate little or no allometric growth.



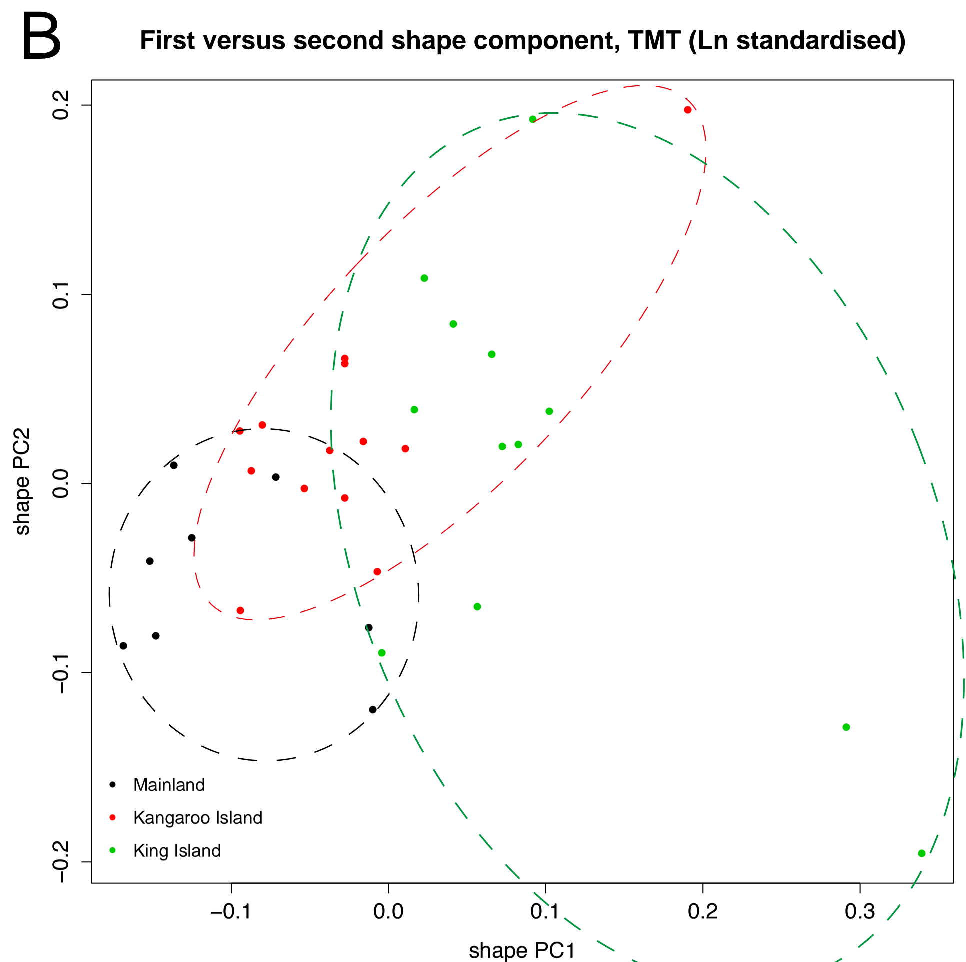


Figure S5. Principal component analysis (PCA) of 'log shape' indices for A) femur and B) tarso-metatarsi measurements from *D. ater* (green), *D. baudinianus* (red), and *D. n. novaehollandiae* (black). To estimate the change in shape between the dwarfed emus and *D. n. novaehollandiae*, raw measurements of these bones were divided by geometric population mean to remove isometric size from the PCA

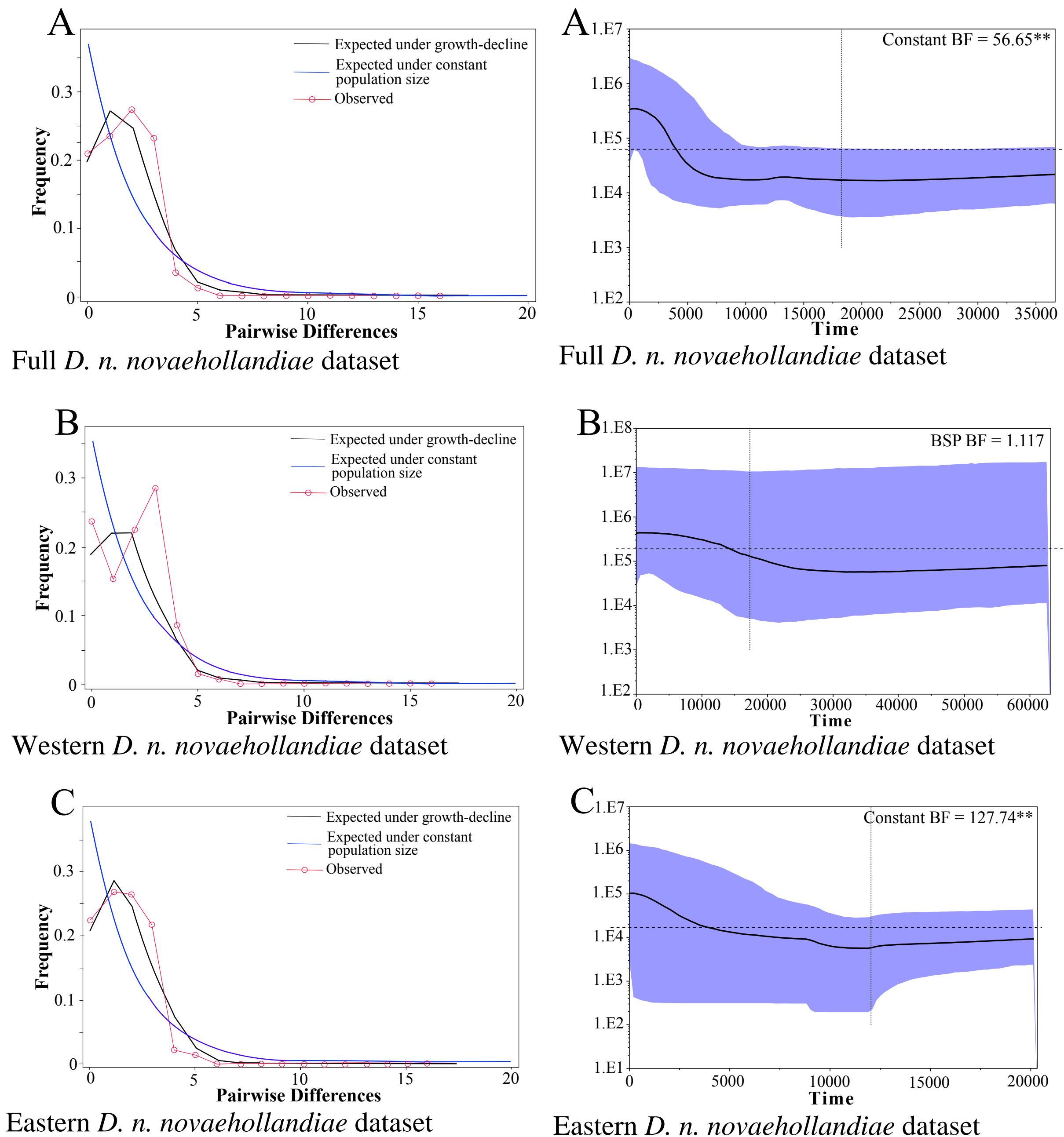


Figure S6. Mismatch distributions and Bayesian Skyline Plot (BSP) estimate of Emu demographic history for A) Entire dataset, B) western population, and C) eastern population. The unimodal mismatch distributions (left hand panels) for the entire dataset and the eastern population indicate evidence for demographic expansions, however the bimodality of the western population suggests it is at equilibrium. This is in direct contrast to the BSP's (right hand panels) – where the entire dataset and eastern populations had very strong or decisive evidence that a constant population size was more likely, versus the western population where there was weak evidence that the demographic history shown in the BSP was more likely; however even in this case a constant population size could not be rejected. The Bayes Factor (BF) supporting either constant population size or BSP is indicated in upper right hand corner (BF <1 is negative, 1<BF<3 is barely worth mentioning, 3<BF<20 is positive, 20<BF<150 is strong, and BF>150 is very strong evidence; Sibon Li and Drummond 2011), with the horizontal dashed line indicating the mean population size estimated from BEAST runs of constant population size. The vertical line indicates date of earliest sample for each dataset, with demography to the right of the line speculative.

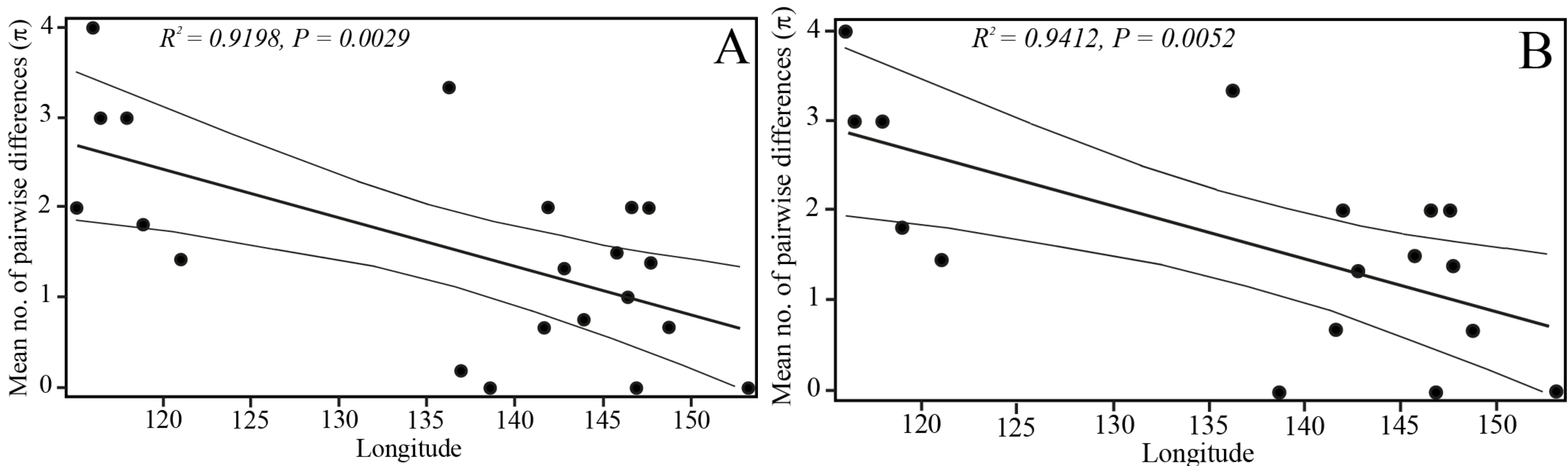


Figure S7. Results of linear regression analyses of the *D. n. novaehollandiae* individuals grouped by the Interim Biogeographic Regionalisation of Australia (IBRA) bioregion 'population' in which they were located, comparing mean number of pairwise differences within each 'population' (π) vs. longitudinal gradient, using A) both modern and ancient data, and B) modern data only. Both show a significant negative association between mean number of pairwise differences and longitude, confirming that populations of *D. n. novaehollandiae* in western Australia have higher genetic diversity than those in eastern Australia. In contrast to these results, haplotype diversity and nucleotide diversity statistics did not show significant associations with longitude. Straight lines show linear regression association, with 95% confidence intervals shown as curved lines.alisation of Australia (IBRA) bioregion 'population' in which they were located, comparing mean number of pairwise differences within each 'population' (π) vs. longitudinal gradient, using A) both modern and ancient data, and B) modern data only. Both show a significant negative association between mean number of pairwise differences and longitude, confirming that populations of *D. novaehollandiae* in western Australia have higher genetic diversity than those in eastern Australia. In contrast to these results, haplotype diversity and nucleotide diversity statistics did not show significant associations with longitude. Straight lines show linear regression association, with 95% confidence intervals shown as curved lines.

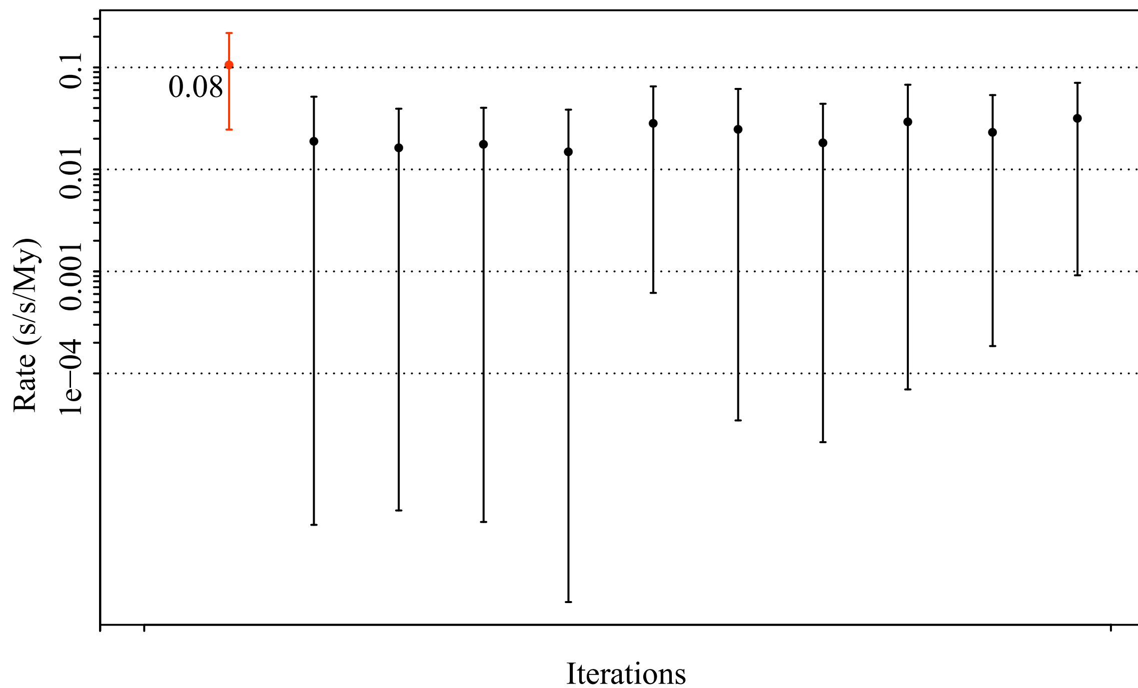


Figure S8. Results of date randomisation test, showing actual evolutionary rate with 95% HPD (shown in red) calculated in BEAST using a constant population size, HKY+G substitution model and uncorrelated LogNormal clock model. Ten replicates were undertaken by randomising the dates associated with each sample (shown in black). The 95% HPD of the randomisations do not overlap the point estimate of the actual evolutionary rate, therefore there is enough phylogenetic signal associated with the dated samples for the evolutionary rate to be considered accurate enough for use in further analyses.

References

Anderson CNK, Ramakrishnan U, Chan YL and Hadly EA (2004). Serial SimCoal: A population genetics model for data from multiple populations and points in time. *Bioinformatics* 21(8): 1733-1734.

Anderson RP and Handley CO (2002). Dwarfism in insular sloths: Biogeography, selection, and evolutionary rate. *Evolution* 56(5): 1045-1058.

Aubry KB, Statham MJ, Sacks BN, Perrine JD and Wisely SM (2009). Phylogeography of the North American red fox: vicariance in Pleistocene forest refugia. *Molecular Ecology* 18(12): 2668-2686.

Bandelt HJ, Forster P and Rohl A (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* 16(1): 37-48.

Barnes I, Matheus P, Shapiro B, Jensen D and Cooper A (2002). Dynamics of Pleistocene population extinctions in Beringian Brown Bears. *Science* 295(5563): 2267-2270.

Barrows TT, Stone JO, Fifield LK and Cresswell RG (2002). The timing of the Last Glacial Maximum in Australia. *Quaternary Science Reviews* 21(1-3): 159-173.

Baur H and Leuenberger C (2011). Analysis of ratios in multivariate morphometry. *Systematic Biology* 60(2011): 1-13.

Becquey S and Gersonde R (2003). A 0.55-Ma paleotemperature record from the Subantarctic zone: Implications for Antarctic Circumpolar Current development. *Paleoceanography* 18(1).

Bhagwat SA and Willis KJ (2008). Species persistence in northerly glacial refugia of Europe: a matter of chance or biogeographical traits? *Journal of Biogeography* 35(3): 464-482.

Blakers M, Davies SJJF and Reilly PN (1985). *The Atlas of Australian Birds*. Melbourne, Melbourne University Press.

Boles W (1992). Revision of the *Dromaius gidju* Patterson and Rich 1987 from Riversleigh, northwestern Queensland, Australia, with a reassessment of its generic position. *Los Angeles County Museum Science Series* 36: 195-208.

Bollongino R, Tresset A and Vigne J-D (2008). Environment and excavation: Pre-lab impacts on ancient DNA analyses. *Comptes Rendus Palevol* 7(2-3): 91-98.

Bowler JM (1982). Aridity in the late Tertiary and Quaternary of Australia. Evolution of the Flora and Fauna of Arid Australia. Barker WR and Greenslade PJM. Frewville, Australian Systematic Botany Society and ANZAAS: 35-46.

Brook BW, Bowman D, Burney DA, Flannery TF, Gagan MK, Gillespie R, Johnson CN, Kershaw P, Magee JW, Martin PS, Miller GH, Peiser B and Roberts RG (2007). Would the Australian megafauna have become extinct if humans had never colonised the continent? Comments on "A review of the evidence for a human role in the extinction of

Australian megafauna and an alternative explanation" by S Wroe and J Field. *Quaternary Science Reviews* 26(3-4): 560-564.

Brown M, Cooksley H, Carthew SM and Cooper SJB (2006). Conservation units and phylogeographic structure of an arboreal marsupial, the yellow-bellied glider (*Petaurus australis*). *Australian Journal of Zoology* 54(5): 305-317.

Burns E, Eldridge M, Crayn D and Houlden B (2007). Low phylogeographic structure in a wide spread endangered Australian frog *Litoria aurea* (Anura: Hylidae). *Conservation Genetics* 8(1): 17-32.

Byrne M (2008). Evidence for multiple refugia at different time scales during Pleistocene climatic oscillations in southern Australia inferred from phylogeography. *Quaternary Science Reviews* 27(27-28): 2576-2585.

Byrne M, Macdonald B and Brand J (2000). Phylogeography and divergence in the chloroplast genome of Western Australian Sandalwood (*Santalum spicatum*). *Heredity* 91(4): 389-395.

Campos PF, Kristensen T, Orlando L, Sher A, Kholodova MV, Gotherstrom A, Hofreiter M, Drucker DG, Kosintsev P, Tikhonov A, Baryshnikov GF, Willerslev E and Gilbert MTP (2010). Ancient DNA sequences point to a large loss of mitochondrial genetic diversity in the saiga antelope (*Saiga tatarica*) since the Pleistocene. *Molecular Ecology* 19(22): 4863-4875.

Clark PU, Dyke AS, Shakun JD, Carlson AE, Clark J, Wohlfarth B, Mitrovica JX, Hostetler SW and McCabe AM (2009). The Last Glacial Maximum. *Science* 325(5941): 710-714.

Cooper A and Poinar HN (2000). Ancient DNA: Do It Right or Not at All. *Science* 289(5482): 1139.

Cracraft J (1986). Origin and evolution of continental biotas: speciation and historical congruence within the Australian avifauna. *Evolution* 40(5): 977-996.

Crisp MD and Cook LG (2007). A congruent molecular signature of vicariance across multiple plant lineages. *Molecular Phylogenetics and Evolution* 43(3): 1106-1117.

Davies S, Beck M and Kruiskamp J (1971). The results of banding 154 Emus in Western Australia. *CSIRO Wildlife Research* 16: 77-79.

Davies SJJF (2002). *Ratites and Tinamous: Tinamidae, Rheidae, Dromaiidae, Casuaiidae, Apterygidae, Struthionidae*. Oxford, Oxford University Press.

Dawson TJ, Read D, Russell EM and Herd RM (1984). Seasonal variation in daily activity patterns, water relations and diet of Emus. *Emu* 84: 93-102.

Debruyne R, Chu G, King CE, Bos K, Kuch M, Schwarz C, Szpak P, Grocke DR, Matheus P, Zazula G, Guthrie D, Froese D, Buigues B, de Marliave C, Flemming C, Poinar D, Fisher D, Southon J, Tikhonov AN, MacPhee RDE and Poinar HN (2008). Out of America:

Ancient DNA evidence for a new world origin of Late Quaternary woolly mammoths. *Current Biology* 18(17): 1320-1326.

Dodson J (2001). Holocene vegetation change in the Mediterranean-type climate regions of Australia. *The Holocene* 11(6): 673-680.

Dove H (1926a). How Tasmania lost the Emu. *The Emu* XXV: 213.

Dove S (1924). Notes on the Tasmanian Emu. *The Emu* XXIII: 221-222.

Dove S (1926b). The Tasmanian Emu. *The Emu* XXV: 290-291.

Drummond A and Rambaut A (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* 7(1): 214.

Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A, Duran C, Field M, Heled J, Kearse M, Markowitz S, Moir R, Stones-Havas S, Sturrock S, Thierer T and Wilson AC (2011). Geneious v5.4. *Available from* http://www.geneious.com/.

EPICA, Barbante C, Barnes PRF, Barnola J-M, Bigler M, Castellano E, Cattani O, Chappellaz J, Dahl-Jensen D, Delmonte B, Dreyfus G, Durand G, Falourd S, Fischer H, Fluckiger J, Hansson ME, Huybrechts P, Jugie G, Johnsen SJ, Jouzel J, Kaufmann P, Kipfstuhl J, Lambert F, Lipenkov VY, Littot GC, Longinelli A, Lorrain R, Maggi V, Masson-Delmotte V, Miller H, Mulvaney R, Oerlemans J, Oerter H, Orombelli G, Parrenin F, Peel DA, Petit J-R, Raynaud D, Ritz C, Ruth U, Schwander J, Siegenthaler U, Souchez R, Stauffer B, Steffensen JP, Stenni B, Stocker TF, Tabacco IE, Udisti R, van de Wal RSW, van den Broeke M, Weiss J, Wilhelms F, Winther J-G, Wolff EW and Zucchelli M (2004). Eight glacial cycles from an Antarctic ice core. *Nature* 429(6992): 623-628.

Excoffier L (2004). Patterns of DNA sequence diversity and genetic structure after a range expansion: lessons from the infinite-island model. *Molecular Ecology* 13(4): 853-864.

Excoffier L and Lischer H (2010). Arlequin suite ver3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* 10: 564-567.

Faith JT and O'Connell JF (2011). Revisiting the late Pleistocene mammal extinction record at Tight Entrance Cave, southwestern Australia. *Quaternary Research* 76(3): 397-400.

Field JH and Boles WE (1998). *Genyornis newtoni* and *Dromaius novaehollandiae* at 30,000 B.P. in central northern New South Wales. *Alcheringa: An Australasian Journal of Palaeontology* 22(2): 177-188.

Flood J (2004). *Archaeology of the Dreamtime*, J. B. Publishing.

Ford J (1987a). Hybrid zones in Australian birds. *Emu* 87: 158-178.

Ford J (1987b). Minor isolates and minor geographical barriers in avian speciation in continental Australia. *Emu* 87: 90-102.

Ford JR (1974). Speciation in Australian birds adapted to arid habitats. *Emu* 74(3): 161-168.

Foster JB (1964). Evolution of mammals on islands. *Nature* 202(4929): 234-235.

Fritz U, Siroky P, Kami H and Wink M (2005). Environmentally caused dwarfism or a valid species--is *Testudo weissingeri* Bour, 1996 a distinct evolutionary lineage? New evidence from mitochondrial and nuclear genomic markers. *Molecular Phylogenetics and Evolution* 37(2): 389-401.

Gilbert CC and Rossie JB (2007). Congruence of molecules and morphology using a narrow allometric approach. *Proceedings of the National Academy of Sciences of the United States of America* 104(29): 11910-11914.

Gould GC and MacFadden BJ (2004). Chapter 17: Gigantism, dwarfism, and Cope's Rule: Nothing in evolution makes sense without a phylogeny. *Bulletin of the American Museum of Natural History*: 219-237.

Groeneveld LF, Rasoloarison RM and Kappeler PM (2011). Morphometrics confirm taxonomic deflation in dwarf lemurs (Primates: Cheirogaleidae), as suggested by genetics. *Zoological Journal of the Linnean Society* 161(1): 229-244.

Guay PJ, Chesser R, Mulder R, Afton A, Paton D and McCracken K (2010). East-west genetic differentiation in Musk Ducks (*Biziura lobata*) of Australia suggests late Pleistocene divergence at the Nullarbor Plain. *Conservation Genetics* 11(6): 2105-2120.

Hadly EA, Ramakrishnan U, Chan YL, Tuinen MV, Keefe K, Spaeth PA and Conroy CJ (2004). Genetic response to climatic change: Insights from ancient DNA and phylochronology. *PloS Biology* 2(10): e290.

Hawkins CE, Baars C, Hesterman H, Hocking GJ, Jones ME, Lazenby B, Mann D, Mooney N, Pemberton D, Pyecroft S, Restani M and Wiersma J (2006). Emerging disease and population decline of an island endemic, the Tasmanian devil *Sarcophilus harrisii*. *Biological Conservation* 131(2): 307-324.

Heupink TH, Huynen L and Lambert DM (2011). Ancient DNA suggests dwarf and 'Giant' Emu are conspecific. *PloS One* 6(4).

Hewitt G (2000). The genetic legacy of the Quaternary ice ages. *Nature* 405(6789): 907-913.

Ho SYW, Lanfear R, Phillips MJ, Barnes I, Thomas JA, Kolokotronis SO and Shapiro B (2011). Bayesian estimation of substitution rates from ancient DNA sequences with low information content. *Systematic Biology* 60(3): 366-374.

Ho SYW, Larson G, Edwards CJ, Heupink TH, Lakin KE, Holland PWH and Shapiro B (2008). Correlating Bayesian date estimates with climatic events and domestication using a bovine case study. *Biology Letters* 4(4): 370-374.

Hope JH, Lampert RJ, Edmondson E, Smith MJ and Van Tets GF (1977). Late Pleistocene faunal remains from Seton Rock Shelter, Kangaroo Island, South Australia. *Journal of Biogeography* 4(4): 363-385.

Houlden BA, Costello BH, Sharkey D, Fowler EV, Melzer A, Ellis W, Carrick F, Baverstock PR and Elphinstone MS (1999). Phylogeographic differentiation in the mitochondrial control region in the koala, *Phascolarctos cinereus* (Goldfuss 1817). *Molecular Ecology* 8(6): 999-1011.

Huson DH and Bryant D (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* 23(2): 254-267.

James CH and Moritz C (2000). Intraspecific phylogeography in the sedge frog *Litoria fallax* (Hylidae) indicates pre-Pleistocene vicariance of an open forest species from eastern Australia. *Molecular Ecology* 9(3): 349-358.

Jones ME and Stoddart DM (1998). Reconstruction of the predatory behaviour of the extinct marsupial thylacine (*Thylacinus cynocephalus*). *Journal of Zoology* 246: 239-246.

Joseph L, Adcock GJ, Linde C, Omland KE, Heinsohn R, Chesser RT and Roshier DA (2009). A tangled tale of two teal: population history of the grey *Anas gracilis* and chestnut teal *A. castanea* of Australia. *Journal of Avian Biology* 40: 430-439.

Joseph L and Wilke T (2006). Molecular resolution of population history, systematics and historical biogeography of the Australian ringneck parrots *Barnardius*: are we there yet? *Emu* 106(1): 49-62.

Joseph L and Wilke T (2007). Lack of phylogeographic structure in three widespread Australian birds reinforces emerging challenges in Australian historical biogeography. *Journal of Biogeography* 34(4): 612-624.

Joseph L, Wilke T and Alpers D (2002). Reconciling genetic expectations from host specificity with historical population dynamics in an avian brood parasite, Horsfield's Bronze-Cuckoo *Chalcites basalis* of Australia. *Molecular Ecology* 11(4): 829-837.

Joseph L, Wilke T, Ten Have J and Terry Chesser R (2006). Implications of mitochondrial DNA polyphyly in two ecologically undifferentiated but morphologically distinct migratory birds, the masked and white-browed woodswallows *Artamus spp*. of inland Australia. *Journal of Avian Biology* 37(6): 625-636.

Keane TM, Creevey CJ, Pentony MM, Naughton TJ and McLnerney JO (2006). Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evolutionary Biology* 6(29): 1-17.

Kearns AM, Joseph L and Cook LG (2010). The impact of Pleistocene changes of climate and landscape on Australian birds: a test using the Pied Butcherbird (*Cracticus nigrogularis*). *Emu* 110(4): 285-295.

Kearns AM, Joseph L, Edwards SV and Double MC (2009). Inferring the phylogeography and evolutionary history of the splendid fairy-wren *Malurus splendens* from mitochondrial DNA and spectrophotometry. *Journal of Avian Biology* 40(1): 7-17.

King D, Oliver A and Mead R (1978). The adaptation of some Western Australian mammals to food plants containing fluoroacetate. *Australian Journal of Zoology* 26(4): 699-712.

Kuch U, Keogh JS, Weigel J, Smith LA and Mebs D (2005). Phylogeography of Australia's king brown snake (*Pseudechis australis*) reveals Pliocene divergence and Pleistocene dispersal of a top predator. *Naturwissenschaften* 92(3): 121-127.

Lambeck K and Chappell J (2001). Sea level change through the last glacial cycle. *Science* 292(5517): 679-686.

Le Souef D (1904). Emu (*Dromaius novaehollandaie*) Feathers. *The Emu* 25: 213.

Lomolino MV (1985). Body size of mammals on islands - The island rule reexamined. *American Naturalist* 125(2): 310-316.

Lowe A, Harris S and Ashton P (2004). *Ecological Genetics: Design, Analysis, and Application*. Oxford, Blackwell Science Ltd.

Marshall LG and Corruccini RS (1978). Variability, evolutionary rates, and allometry in dwarfing lineages. *Paleobiology* 4: 101-119.

McIlroy J (1984). The sensitivity of Australian animals to 1080 poison. Vii. Native and introduced birds. *Wildlife Research* 11(2): 373-385.

McNab B (2010). Geographic and temporal correlations of mammalian size reconsidered: a resource rule. *Oecologia* 164(1): 13-23.

Meyer M, Kircher M, Gansauge M-T, Li H, Racimo F, Mallick S, Schraiber JG, Jay F, Prufer K, de Filippo C, Sudmant PH, Alkan C, Fu Q, Do R, Rohland N, Tandon A, Siebauer M, Green RE, Bryc K, Briggs AW, Stenzel U, Dabney J, Shendure J, Kitzman J, Hammer MF, Shunkov MV, Derevianko AP, Patterson N, Andres AM, Eichler EE, Slatkin M, Reich D, Kelso J and Paabo S (2012). A high-coverage genome sequence from an Archaic Denisovan individual. *Science* 338(6104): 222-226.

Miller GH, Fogel ML, Magee JW, Gagan MK, Clarke SJ and Johnson BJ (2005). Ecosystem collapse in Pleistocene Australia and a human role in megafaunal extinction. *Science* 309(5732): 287-290.

Miller GH, Magee JW, Johnson BJ, Fogel ML, Spooner NA, McCulloch MT and Ayliffe LK (1999). Pleistocene extinction of *Genyornis newtoni*: Human impact on Australian megafauna. *Science* 283(1999): 205-205.

Miller GH, Magee JW and Jull AJT (1997). Low-latitude glacial cooling in the Southern Hemisphere from amino-acid racemization in emu eggshells. *Nature* 385: 241-244.

Miller MP (2005). Alleles In Space (AIS): Computer software for the joint analysis of interindividual spatial and genetic information. *Journal of Heredity* 96(6): 722-724.

Miller SA, Dykes DD and Polesky HF (1988). A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Research* 16(3): 1215.

Miller W, Hayes VM, Ratan A, Petersen DC, Wittekindt NE, Miller J, Walenz B, Knight J, Qi J, Zhao FQ, Wang QY, Bedoya-Reina OC, Katiyar N, Tomsho LP, Kasson LM, Hardie RA, Woodbridge P, Tindall EA, Bertelsen MF, Dixon D, Pyecroft S, Helgen KM, Lesk AM, Pringle TH, Patterson N, Zhang Y, Kreiss A, Woods GM, Jones ME and Schuster SC (2011). Genetic diversity and population structure of the endangered marsupial Sarcophilus harrisii (Tasmanian devil). *Proceedings of the National Academy of Sciences of the United States of America* 108(30): 12348-12353.

Morgan AM and Sutton J (1928). A critical description of some recently discovered bones of the extinct Kangaroo Island Emu (*Dromaius diemenianus*). *The Emu* 28(1): 1-19.

Muller RA and MacDonald GJ (1997). Glacial cycles and astronomical forcing. *Science* 277(5323): 215-218.

Neaves LE, Zenger KR, Prince RI, Eldridge MD and Cooper DW (2009). Landscape discontinuities influence gene flow and genetic structure in a large, vagile Australian mammal, *Macropus fuliginosus*. *Molecular Ecology* 18(16): 3363-3378.

Nicholls JA and Austin JJ (2005). Phylogeography of an east Australian wet-forest bird, the satin bowerbird (*Ptilonorhynchus violaceus*), derived from mtDNA, and its relationship to morphology. *Molecular Ecology* 14(5): 1485-1496.

Omland KE, Baker JM and Peters JL (2006). Genetic signatures of intermediate divergence: population history of Old and New World Holarctic ravens (*Corvus corax*). *Molecular Ecology* 15(3): 795-808.

Oskam CL, Haile J, McLay E, Rigby P, Allentoft ME, Olsen ME, Bengtsson C, Miller GH, Schwenninger JL, Jacomb C, Walter R, Baynes A, Dortch J, Parker-Pearson M, Gilbert MTP, Holdaway RN, Willerslev E and Bunce M (2010). Fossil avian eggshell preserves ancient DNA. *Proceedings of the Royal Society B-Biological Sciences* 277(1690): 1991-2000.

Oskam CL, Jacomb C, Allentoft ME, Walter R, Scofield RP, Haile J, Holdaway RN and Bunce M (2011). Molecular and morphological analyses of avian eggshell excavated from a late thirteenth century earth oven. *Journal of Archaeological Science* 38(10): 2589-2595.

Pahnke K, Zahn R, Elderfield H and Schulz M (2003). 340,000-year centennial-scale marine record of Southern Hemisphere climatic oscillation. *Science* 301(5635): 948-952.

Parker S (1984). The extinct Kangaroo Island Emu, a hitherto unrecognised species. *Bulletin of the British Ornithological Club* 104: 19-22.

Patterson C and Rich P (1987). The fossil history of the Emus, *Dromaius* (Aves: Dromaiinae). *Records of the South Australian Museum* 21(2): 85-117.

Pepper M, Fujita MK, Moritz C and Keogh JS (2011a). Palaeoclimate change drove diversification among isolated mountain refugia in the Australian arid zone. *Molecular Ecology* 20(7): 1529-1545.

Pepper M, Ho SYW, Fujita MK and Keogh JS (2011b). The genetic legacy of aridification: Climate cycling fostered lizard diversification in Australian montane refugia and left low-lying deserts genetically depauperate. *Molecular Phylogenetics and Evolution* 61(3): 750-759.

Perktas U and Gosler AG (2010). Measurement error revisited: its importance for the analysis of size and shape of birds. *Acta Ornithologica* 45(2): 161-172.

Price GJ (2008). Is the modern koala (*Phascolarctos cinereus*) a derived dwarf of a Pleistocene giant? Implications for testing megafauna extinction hypotheses. *Quaternary Science Reviews* 27(27-28): 2516-2521.

Prideaux GJ, Long Ja, Ayliffe LK, Hellstrom JC, Pillans B, Boles WE, Hutchinson MN, Roberts RG, Cupper ML, Arnold LJ, Devine PD and Warburton NM (2007). An arid-adapted middle Pleistocene vertebrate fauna from south-central Australia. *Nature* 445(7126): 422-425.

R development core team (2011). R: A language and environment for statistical computing. R foundation for statistical computing

Rawlence NJ, Wood JR, Armstrong KN and Cooper A (2009). DNA content and distribution in ancient feathers and potential to reconstruct the plumage of extinct avian taxa. *Proceedings of the Royal Society B-Biological Sciences* 276(1672): 3395-3402.

Ray N, Currat M and Excoffier L (2003). Intra-deme molecular diversity in spatially expanding populations. *Molecular Biology and Evolution* 20(1): 76-86.

Rogers AR and Harpending H (1992). Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution* 9(3): 552-569.

Roulin A, Emaresi G, Bize P, Gasparini J, Piault R and Ducrest AL (2011). Pale and dark reddish melanic tawny owls differentially regulate the level of blood circulating POMC prohormone in relation to environmental conditions. *Oecologia* 166(4): 913-921.

Salinas M, Bunce M, Cancilla D, Alpers DL and Spencer PBS (2009). Divergent lineages in the heath mouse (*Pseudomys shortridgei*) are indicative of major contraction to geographically isolated refugia on the eastern and western sides of Australia during the early Pleistocene. *Australian Journal of Zoology* 57(1): 41-47.

Scott H (1924). A note on the King Island Emu. *Papers and Proceedings of the Royal Society of Tasmania*: 103-107.

Shafer ABA, Cullingham CI, Cote SD and Coltman DW (2010). Of glaciers and refugia: a decade of study sheds new light on the phylogeography of northwestern North America. *Molecular Ecology* 19(21): 4589-4621.

Shapiro B, Drummond AJ, Rambaut A, Wilson MC, Matheus PE, Sher AV, Pybus OG, Gilbert MTP, Barnes I, Binladen J, Willerslev E, Hansen AJ, Baryshnikov GF, Burns JA, Davydov S, Driver JC, Froese DG, Harington CR, Keddie G, Kosintsev P, Kunz ML, Martin LD, Stephenson RO, Storer J, Tedford R, Zimov S and Cooper A (2004). Rise and fall of the Beringian Steppe Bison. *Science* 306(5701): 1561-1565.

Sibon Li WL and Drummond AJ (2011). Model averaging and Bayes factor calculation of relaxed molecular clocks in Bayesian phylogenetics. *Molecular Biology and Evolution*.

Siddle HV, Kreiss A, Eldridge MDB, Noonan E, Clarke CJ, Pyecroft S, Woods GM and 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(41): 16221-16226.

Spencer PBS, Rhind SG and Eldridge MDB (2001). Phylogeographic structure within *Phascogale* (Marsupialia: Dasyuridae) based on partial cytochrome b sequence. *Australian Journal of Zoology* 49(4): 369-377.

Taberlet P, Fumagalli L, Wust-Saucy A and Cosson J (1998). Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology* 7(4): 453-464.

Toon A, Mather PB, Baker AM, Durrant KL and Hughes JM (2007). Pleistocene refugia in an arid landscape: analysis of a widely distributed Australian passerine. *Molecular Ecology* 16(12): 2525-2541.

Twigg L, King D, Davis H, Saunders D and Mead R (1988). Tolerance to, and metabolism of, fluoroacetate in the Emu. *Wildlife Research* 15(3): 239-247.

Vartanyan SL, Garutt VE and Sher AV (1993). Holocene dwarf mammoths from Wrangel Island in the Siberian Arctic. *Nature* 362: 337-340.

Vieillot L (1817). Description of *Dromaius ater*. *Nouveau Dictionaire d'Histoire Naturelle* 11: 212.

Worthy TH (1994). Reappraisal of *Dinornis* (Aves: Dinornithiformes) species - a morphometric analysis. *New Zealand Journal of Zoology* 21: 113-134.

Zenger KR, Eldridge MDB and Cooper DW (2003). Intraspecific variation, sex-biased dispersal and phylogeography of the Eastern Grey Kangaroo (*Macropus giganteus*). *Heredity* 91(2): 153-162.

Chapter 4: the Ghost Bat

A ghost of a chance? Evolutionary history of the Ghost Bat (*Macroderma gigas*) and its chances of surviving future climate change

STATEMENT OF AUTHORSHIP

A Ghost of a chance: evolutionary history of the Ghost Bat (*Macroderma gigas*) and its chances of surviving future climate change

Text in manuscript

THOMSON, V.A. (Candidate)

Performed ancient DNA extractions, PCR amplifications, and sequencing, plus downstream processing and analysis on all samples, performed the phylogeographic analysis, interpreted the data, created figures, and wrote manuscript.

I hereby certify that the statement of contribution is accurate.

Signed. ....*Date* 10/5/2012

MEDLIN, GRAHAM

Gave advice on fieldwork and evaluated the manuscript.

I hereby certify that the statement of contribution is accurate and I give permission for the inclusion of the manuscript in the thesis.

Signed.. ....*Date* 24-04-2012

ARMSTRONG, KYLE

Performed modern DNA extractions, PCR amplification and sequencing, and evaluation of the manuscript.

I hereby certify that the statement of contribution is accurate and I give permission for the inclusion of the manuscript in the thesis.

Signed. ....*Date* 2/5/2012

COOPER, A.

Provided funding for laboratory work and evaluation of the manuscript.

I hereby certify that the statement of contribution is accurate and I give permission for the inclusion of the manuscript in the thesis.

Signed. …. ………………….*Date* 10/5/2012

A Ghost of a chance? Evolutionary history of the Ghost Bat (*Macroderma gigas*) and its chances of surviving future climate change

Vicki A. Thomson[a], Kyle N. Armstrong[a], Graham Medlin[b], Alan Cooper[a].

[a]Australian Centre for Ancient DNA, University of Adelaide, Adelaide SA 5005;[b]South Australian Museum.

Journal: Diversity and Distributions (5000 words)

Abstract

Aim: An understanding of species' past distributions and response to climate change can help conservationists to prioritise management strategies for maximum biodiversity preservation in the face of future anthropogenic environmental change. This is particularly important for Australia, where mammal extinctions since the time of European colonisation account for half of the global total over this period. Many other Australian mammals have suffered major range contractions and this can confound our view of a species' ecological requirements and of its demographic history as reconstructed from genetic signals. On both counts, a more balanced picture can emerge through investigation of the distribution and genetic signature of subfossil remains representing the early historic period and deeper into prehistory. An excellent case study in the Australian context is the Ghost Bat (*Macroderma gigas*), which has undergone a substantial range contraction since European colonisation, is represented in many recent cave deposits, and is currently rated Vulnerable by IUCN. Our aim was to reconstruct past distributions and the connectedness between extant and extinct colonies of Australian Ghost Bats (*Macroderma gigas*) using modern samples and ancient DNA (aDNA) from Holocene subfossil deposits. By combining this modern and palaeogenetic data with ecological niche modelling (ENM) of present and past climatic conditions, we inferred the response of *M. gigas* to past climate change and investigate the mechanisms behind the extinction of the southern colonies. Furthermore, by modelling future climate projections we attempt to assess the ability of *M. gigas* to survive future climate change.

Location The Australian continent.

Methods We sequenced a 255 bp fragment of the mitochondrial *CR* from cave subfossil material of extinct colonies using aDNA techniques, and fresh tissues from extant colonies, and combined with existing sequences to establish the timings of genetic divergence and expansion events of colonies. Furthermore, we implemented ENM to predict suitable current environmental conditions and to model past and future distributional patterns.

Results We found evidence for widespread expansion after population bottlenecks during the Middle Pleistocene (0.126-0.781 Mya). In addition, two colonies were found to exhibit signs of expansion after the last interglacial (LIG; 120-140ka). Phylogenetic reconstructions suggest that southern colonies, as represented by Holocene subfossil deposits, underwent periodic small-scale expansion out of multiple refugia through the repeated glacial cycles of the Pleistocene, resulting in deep divergence between regional clades.

Main conclusions Our study is the first to investigate the temporal timeframe over which gene flow occurred amongst *M. gigas* colonies, highlighting that *M. gigas* has had contrasting regional responses to global climatic events. We found evidence that colonies of *M. gigas* have expanded southwards during periods of higher humidity (interglacials), and contracted

northwards in response to increasing aridity (e.g. in the lead up to the last glacial maximum). Based on the combined analyses, we concur with previous views that *M. gigas* is a relictual species, both phylogenetically and geographically. However, its propensity to utilise marginal habitats and establish new colonies during interglacial periods (and possibly at other times during ephemeral rainfall events) confers a degree of resilience not present in many non-volant Australian mammals. Furthermore, provided conservation efforts are made to ensure its short-term survival, there is the potential that future climate change may allow *M. gigas* to re-invade currently unoccupied parts of its former range in southern Australia. However, uncertainty surrounds whether these re-invading *M. gigas* populations can sustain viable colonies, given the changes in ecosystem services that these currently unoccupied regions will have experienced since historical extirpations of *M. gigas* from these southern regions.

Introduction

Climatic and environmental change during the Late Quaternary played a key role in shaping the diversity and distribution of present day plants and animals worldwide (Irwin 2002; Jones 2008; Maclean and Wilson 2011; Campos *et al.* 2011; Braconnot *et al.* 2007; Fujioka *et al.* 2009). In Australia, these climate fluctuations were characterised by repeated periods of increased aridity (Barrows *et al.* 2002), which caused a temporal flux in ecological communities (Chapple *et al.* 2004; Kearney and Blacket 2008). Individual species within the modern biota carry signals of these events in their pattern of distribution and in their genetic signature. In particular, ecosystems of the Australian arid zone harbour many species with distinctive phylogeographic patterns influenced by the formation of deserts (*e.g.* the Simpson and Western Deserts), and potential refugia offered by mountain ranges (*e.g.* the Pilbara, Kimberley and Arunta blocks) (Byrne *et al.* 2008). In addition, the ephemeral nature of the inland arid and semi-arid lakes has played a large role in sustaining gene flow and/or creating divergent populations (*e.g.* Lake Eyre, Murray-Darling lakes, Lake Amadeus *etc.*). The inland drainage basins are often flat, with a patchwork of meandering rivers that open up to many shallow yet ephemeral waterholes, and are surrounded by large sparsely vegetated swampy regions. Each playa lake system will respond differently to rainfall events, depending on depth of lake basin, height of water table, strength of connection to nearby river systems, size and run-off features of the catchment area, and each system will sustain varying patterns of flora and fauna temporally and spatially. In combination, the mixture of deserts, mountain ranges and lakes across the Australian arid and semi-arid zones have structured both past and present biodiversity. Therefore, the deep phylogeographic structuring in many arid zone taxa facilitates reconstruction of biotic responses to past climate change (Blacket *et al.* 2001; Chapple *et al.* 2004; Cooper *et al.* 2000; Garrick *et al.* 2004; Oliver *et al.* 2007; Perez *et al.* 2005; Smith *et al.* 2007), and may help in predicting the ability of these same species to survive future climate change.

The Australian biota has suffered an exceptionally high rate of extinction since European settlement due to a combination of climatic change, human-induced ecosystem disturbance, and invasive species introductions (Burbidge *et al.* 2008; Bilney *et al.* 2010; Faith and O'Connell 2011). Australia has the dubious honour of contributing to half of all recent mammal extinctions globally (Short and Smith 1994). In order to preserve the remaining biodiversity, a deeper understanding is needed about the causes of population declines in vulnerable species and factors that might influence survival under future climate change. For this study we chose to investigate Australia's largest carnivorous bat, *Macroderma gigas* (the Ghost Bat). *Macroderma gigas* is the sole surviving species of a formerly more diverse radiation of megadermatid bats in Australia (Worthington Wilmer *et al.* 1994; Worthington Wilmer *et al.* 1999; Hand 1996). It is commonly associated with bluffs, faceted slopes and caves/overhangs formed by weathering of specific

geological formations to create suitable cave microclimates, as well as old subterranean mine workings (Armstrong 2003; Armstrong and Anstee 2000). Accordingly, extant colonies of *M. gigas* occur mostly as small disjunct isolates in terrains that provide natural and human-made shelters. At present they are restricted to the northern regions of the continent, at latitudes above 24 °S (Worthington Wilmer *et al.* 1994). However, historically and in recent prehistoric times *M. gigas* occurred more widely across the continent and subfossil remains occur in many caves across southern Australia (Churchill and Helman 1990; Figure S1; Molnar *et al.* 1984; reviewed in Armstrong and Anstee 2000). In Western Australia some Ghost Bat colonies persisted over long time periods, from at least the Holocene/Pleistocene boundary (Hastings Cave, WA: 11,000ya; Devil's Lair, WA: 12,000ya; Bridge 1975) through into the historic period (1850's-1950's; Douglas 1967; 1956; 1957; Molnar *et al.* 1984). However, all historic sightings at these now-extinct colonies were of single individuals or small colonies (Butler 1961; Douglas 1956; 1957; 1967), suggesting that they declined rapidly or were already in decline at the time of European settlement. The causes of these declines are as yet unknown, although changing climatic conditions and declines in available prey species have been highlighted as possible mechanisms (Molnar *et al.* 1984; Burbidge *et al.* 1988).

The distribution and demographic history of *M. gigas* are likely influenced by its biotic interactions and life history traits (Pearson and Dawson 2003). *Macroderma gigas*' major biotic interactions are with its prey species, as it has few natural predators (although competitive interactions do occur with medium sized owls; Hudson and Wilson 1986). In extant colonies, the prey of *M. gigas* are predominantly insects, especially during the wet season, with birds, other bats, small mammals, and frogs supplementing their diet during the dry season (Tidemann *et al.* 1985). Several life-history traits may be important in determining the size and resilience of *M. gigas* colonies. For example, the species has a low reproductive rate with females producing one young per year, and exhibiting a high level of natal roost philopatry (Worthington Wilmer *et al.* 1994; Worthington Wilmer *et al.* 1999). It does not enter torpor and thus requires a continuous food supply; and being homeothermic, expends energy to maintain a constant body temperature (Kulzer *et al.* 1970; Leitner and Nelson 1967). The species has a preferred roosting microclimate of around 28° C and humidity level between 70–90% (Baudinette *et al.* 2000; Churchill 1991). Availability of suitable cave formations clearly limits the location of colonies and *M. gigas* is thought to be very sensitive to disturbance of its roosts (Churchill and Helman 1990; Churchill 1991; Worthington Wilmer *et al.* 1999; Worthington Wilmer *et al.* 1994; Armstrong and Anstee 2000). All of these factors may be important for understanding the decline and extinction of southern colonies of *M. gigas*.

Several authors have attempted to explain the extinction of *M. gigas* colonies in southern and central Australia (Molnar *et al.* 1984; Bridge 1975; Butler 1961; Finlayson 1961; Burbidge *et*

al. 1988). We incorporated these prior views into three alternative, testable hypotheses that might explain the current and past distribution of *M. gigas*:

- Hypothesis #1: The core distribution was always in the northern regions of the continent (i.e. where the extant colonies are located) and extinct southern colonies were the product of episodic range expansions during more productive warm/wet periods *i.e.* the ephemeral filling of many inland arid and semi-arid lakes (Periodic Southern Expansion Hypothesis).
- Hypothesis #2: The original core distribution was southern Australia but the species slowly contracted northwards in response to increased aridity, with southern populations persisting into historic times in locally productive refugia with suitable cave microclimates (Northern Relocation Hypothesis).
- Hypothesis #3: The species was effectively pan-continental, limited primarily by the distribution of caves with suitable microclimates, and southern colonies declined due to human-induced factors (Pleistocene Continental Distribution Hypothesis).

Each of these hypotheses has a different set of predicted genetic, temporal, and climatic modelling signatures (outlined in Table 1). We considered the hypotheses in light of phylogeographic reconstructions based on mitochondrial *CR* data that included samples from previously published data (Worthington Wilmer *et al.* 1999; Worthington Wilmer *et al.* 1994), extant northern colonies and sub-fossil remains from extinct southern colonies, the latter using methods appropriate for ancient DNA (aDNA). We further explored the implications of these results through ecological niche modelling, under present and past climatic conditions, and under various future climate projections to assess the vulnerability of *M. gigas* to climate change.

Table 1. The hypotheses being tested in this study, with the genetic (mtDNA), temporal and ecological niche modelling signatures expected under each scenario, and assuming the signal stems from more recent glacial cycles.

Hypotheses	Genetic signature	Temporal signature	Climatic modelling signature
Periodic Southern Expansion Hypothesis: Core distribution in north; episodic southward expansion during more productive wet/warm periods: predicts expansions during Last Inter-Glacial (LIG) and Holocene	Some northern haplotypes present in southern colonies; gene flow between colonies connected by inland arid and semi-arid lakes (*e.g.* Lake Eyre and Lake Amadeus); evidence of expansions in southern colonies	Nodes date from before last warm/wet expansion event: predicts nodes date from just prior to the LIG or Holocene	Extant modelling predicts suitable climate in south during last expansion event: predicts suitable climate to occur during LIG and Holocene. Assumes a centre of origin for the species in tropical northern Australia, *e.g.* see Hand (1996)
Northern Relocation Hypothesis: Core distribution in south; contraction northwards during dry/arid periods including Last Glacial Maximum (LGM); later survival in locally productive refugia with suitable cave microclimates.	Some southern haplotypes present in northern colonies; gene flow between colonies connected by caves with suitable microclimate; evidence of expansion in northern colonies	Nodes date from before last dry/arid period contraction event: predicts nodes date from just prior to LGM	Extinct modelling predicts suitable climate in south prior to last contraction event in LGM: predicts suitable climate to occur prior to LGM, therefore likely to be during LIG
Pleistocene Continental Distribution Hypothesis: Originally widespread, limited primarily by availability of caves, with regional extinctions due to human-induced ecological change	Extinct colonies only distantly related to extant colonies; gene flow random	Node dates very old; stem from more humid period prior to 1Mya.	Modelling shows suitable habitat localised around caves with *M. gigas* presence, but less suitable habitat in intervening areas or karst areas that lack sub-fossil *M. gigas*

Methods

Dataset collection (M. gigas presence localities)

Subfossil *M. gigas* remains were obtained from caves in the Northern Agricultural Region (NAR), Western Australia (Figure 1 and Table S1). Limb bones and teeth were collected from surface deposits from caves E-24 and J-03, situated near Eneabba and Jurien, respectively (along the coast north of Perth). Additional samples of rostral cartilage and wing tissue were sampled from mummified specimens in the South Australian Museum (SAM), which were collected as surface cave finds from Alice Springs in central Australia (CA) and the Flinders Ranges (FR). Although no radiocarbon dates were obtained from any subfossil or mummified individuals, the physical condition of these specimens strongly suggests that they derive from some of the last remaining individuals occupying these caves (*e.g.* within the last 200 years). Modern tissue/blood samples were collected from live *M. gigas* from the Pilbara region and combined with sequences from Worthington Wilmer *et al.* (1999; 1994).

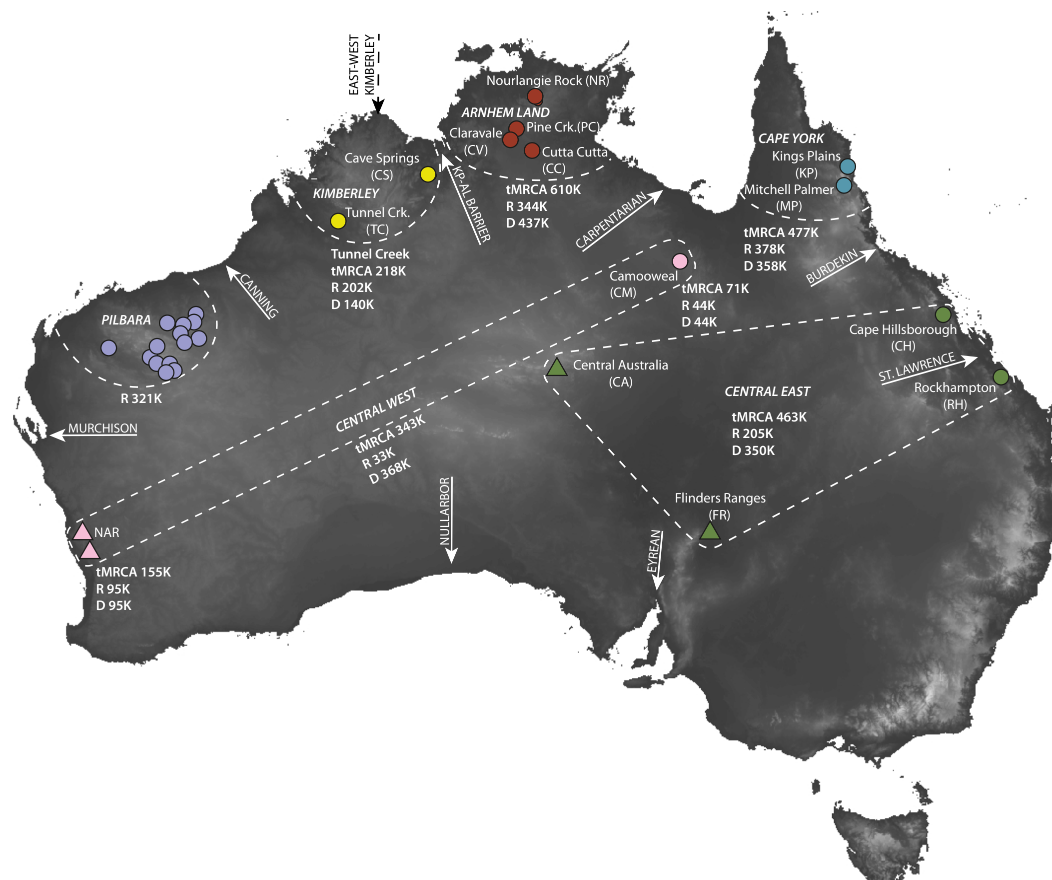


Figure 1. Map of *M. gigas* colonies sampled in this study (see Figure S1 for extinct populations not sampled in this study). Circles correspond to extant colonies; triangles correspond to extinct colonies; colours refer to clades denoted in BEAST phylogeny in Figure 2. The time to most recent common ancestor (tMRCA), range expansions (R), and demographic expansions (D) of regions are shown. Major geographic barriers to gene flow in Australian birds are indicated by arrows (Ford 1987). Biogeographic barriers are noted by arrows, including Kimberley Plateau-Arnhem Land (KP-AL) barrier.

DNA extraction and sequencing

Samples: Seventeen bone/teeth from the NAR and eight cartilage samples from FR and CA were extracted, amplified and sequenced for the mitochondrial control region (*CR*). DNA was extracted from single teeth or 0.2 g of bone powder using a modified version of the DNeasy Tissue kit (Qiagen) (see Chapter 5 methods). All bone and teeth samples were extracted in a specialised ancient DNA laboratory that is physically remote from laboratories handling modern samples and PCR products. The Australian Centre for Ancient DNA (ACAD) aDNA laboratory is kept at positive air pressure, and uses standard ancient DNA precautions (clean body suits, face shields, multiple pairs of gloves), universal UV irradiation, and regular cleaning with oxidising agents to prevent contamination. All reagents and materials are supplied DNA-free and used only once. Separate hoods are used for DNA extractions and to set up PCR experiments, and the amplification and post-PCR procedures are performed in a post-PCR laboratory located in a separate, physically remote building, to minimise the potential for contamination by PCR amplicons. The movement of researchers and materials is carefully controlled to further reduce this risk. The rostral cartilage and mummified tissue were extracted and PCR reactions constructed in a pre-PCR clean-room laboratory. The modern tissue samples were extracted in a separate laboratory using a MasterPure DNA purification kit (Epicentre Biotechnologies).

PCR and Sequencing. The same 255 bp of mitochondrial *CR* was amplified using primers designed inside the 300 bp fragment generated in the studies of Worthington-Wilmer *et al.* (1999; 1994): A605F – 5'-CTACTTAAACTACCCCTTGC-3', A606R – 5'-GTTATGTCTTATGTACTGTCATG-3', A1542 – 5'-CCCCATGCATATAAGCAAG-3', and A1543 – 5'-AGAGATAACCCATTGGAGGTGA-3'. These primers were designed specifically to amplify *M. gigas*, as the original primers (Worthington Wilmer *et al.* 1994) were in highly conserved tRNA regions and would potentially co-amplify other contaminating mammal DNA. PCRs were set up in 25 µL volumes containing a final concentration of 1 U Platinum Taq DNA Polymerase High Fidelity (Invitrogen), 1x PCR Buffer (Platinum, Invitrogen), 3 mM $MgSO_4$, 200 µM each dNTP, 2 mg ml^{-1} rabbit serum albumin (Sigma), 1 µM forward and reverse primers and 2–3 µL of template DNA. Thermocycling was conducted on a Corbett Research Palm Cycler under the following conditions: 94° C for 2 min, followed by 55 cycles of 94° C for 30 s, 55° C for 30 s, 68° C for 30 s, and a final extension of 10 min at 68°C. Extraction and PCR negative controls were included in all amplifications to detect possible contamination. PCR products were separated by electrophoresis on a 3.5% agarose gel. Successful PCR amplifications were purified using EXOSAP (Fermentas), and the forward and reverse complement of each fragment was sequenced from the same PCR reaction with the PCR primers using Big Dye Terminator v3.1 cycle-sequencing chemistry, and cleaned up on a Multiscreen®$_{384}$ SEQ plate (Millipore). The sequencing run was conducted on an ABI 3130XC capillary sequencer. The sequences were

edited and the forward and reverse strands aligned using Geneious (Drummond *et al.* 2011). The modern tissues from the Pilbara and Kimberley were amplified using the primers in Worthington Wilmer *et al.* (1999; 1994), and sequenced in a similar manner to the ancient samples above. All the edited sequences were aligned to 59 sequences generated by Worthington-Wilmer *et al.* (1999) to form a dataset for 235 individual *M. gigas*.

Phylogenetic and phylogeographic analyses

Phylogenetic relationships were estimated using the BEAST (Bayesian evolutionary analysis by sampling trees) algorithm (Drummond and Rambaut 2007). BEAST was used to estimate intra-specific phylogeographic relationships among colonies (n=235 *M. gigas* individuals, plus 2 individuals of the closest relative, *Megaderma lyra*, as an outgroup), using a secondary fossil calibration point of 16.5 Mya (± 2.75Mya) for the split between Macroderma and Megaderma genera (2005) to date the major clades. A proportion of invariant sites was considered unnecessary for the BEAST analysis of intra-specific diversity as over the shorter time scale invariant sites will likely be due to a lack of opportunity for mutations to occur, rather than evolutionary constraints forcing sites to be conserved (Ho *et al.* 2008). Therefore, when ModelGenerator v0.85 (Keane *et al.* 2006) was used to estimate which substitution model best fit the data, the HKY+G model was the highest ranking model without a proportion of invariant sites (based on Akaike Information Criteria 1 and 2, and the Bayesian Information Criterion). Posterior population parameters were estimated with multiple Markov chain Monte Carlo (MCMC) runs (2 runs of 100 million iterations), with sampling every 10,000 iterations and discarding the first 1 million as burnin. The effective sample size and convergence of the MCMC runs were checked in TRACER v1.5 (Drummond and Rambaut 2007), and TreeAnnotator v1.6.0 (Drummond and Rambaut 2007) was used to produce a consensus tree from the 10,000 trees from the BEAST runs which had the highest Bayes Factor. Based on the high degree of philopatry and phylogeographic structuring from Worthington Wilmer *et al.* (1999; 1994) we expected the phylogeographic structuring of *M. gigas* to be present at multiple levels, for example at the individual cave level, at the level of colonies and at the regional level. Table S1 lists the individual caves, colonies and regions for all samples used in this analysis.

Population genetic and demographic analyses

Population statistics, F_{ST} differentiation statistics, and the parameters of demographic and range expansions were estimated using Arlequin v3.5.1.2 (Excoffier and Lischer 2010). Population genetic statistics in the large Pilbara sample will be reported in a separate collaborative study. Since we believed the cave sub-fossil remains to be of historic or late prehistoric age and therefore each colony to form a cohesive temporal group, it is appropriate to apply these tests to both the extinct and extant colonies. The time since an expansion was converted from *tau* (mutational units) to years using the evolutionary rate calculated from the

BEAST analysis and a generation time of 7.5 years. The generation time of *M. gigas* was estimated using the formula $(T_1 + T_2)/2$; T_1 is age of first breeding (1 year in females; Jones 2008); T_2 is age of last breeding (14 years in females; Jones 2008). It should be noted that these values are based on observations of captive animals, as there is no relevant information available from wild populations, and we also do not know whether these life history parameters varied geographically or through time. As expansion date calculations are particularly sensitive to generation time estimates, this may represent a significant source of error in these calculations (in addition to the errors associated with using a secondary calibration point to estimate the evolutionary rate in BEAST).

Climatic Envelope reconstruction

We constructed models of the distribution of *M. gigas* based on climatic envelopes to investigate Pleistocene/Holocene patterns of suitable habitat in the context of our three hypotheses. The species' current distribution, and its distribution during both past and future conditions were modelled under the assumptions that: 1) the climatic niche of *M. gigas* has remained constant over time; 2) climate is an important factor determining the broad distribution pattern of *M. gigas*; and 3) ambient climate can be used as an approximation for cave microclimates, based on an expectation that the number of caves providing suitable microclimate will vary in accordance with ambient conditions. The latter assumption may be the most important with respect to the robustness of species distribution modelling for *M. gigas*. However, it must be acknowledged that any effect of climate-derived responses of vegetation habitats that ultimately determines prey availability and the climatic preferences of introduced competitors can not be separated from climate itself as a factor.

As microclimate data from extant and extinct colonies in cave sites are not readily available, nor easily gathered, 19 commonly used atmospheric bioclimatic variables, in addition to altitude, were used as a proxy for cave microclimates. Conditions in caves vary according to passage diameter and degree of ventilation, as well as external climate (Smithson 1991), and cave temperatures have generally been found to fluctuate around the annual mean ambient temperature (Churchill 1991; Smithson 1991), though temperature and RH are generally higher and more stable with increasing depth into the cave (Churchill 1991; Armstrong 2000; 2001). In this study, we made a reasonable assumption that regional climate had a role in firstly creating suitable subterranean habitats and secondly in supplying adequate groundwater to elevate RH inside caves. Even though M. gigas is essentially relictual in arid areas such as the Pilbara (suggesting certain climatic parameters are unimportant), there are still seasonal weather events that provide sufficient rainfall to make habitable by this species. Interpretations of the importance of various climatic parameters used in the models are important in this context.

We collated environmental data at the finest scale available for Australia via the Worldclim website (www.worldclim.org). This included 19 climatic layers at current conditions: interpolated from observed data, representing the years 1950–2000; past conditions: last glacial maximum (LGM) at ~21,000 ya (Braconnot *et al.* 2007) and last inter-glacial (LIG) at ~ 120,000–140,000 ya (Otto-Bliesner *et al.* 2006); and future climate model scenarios for 2030 and 2080 as the CSIRO Mk3.0 SRES A1B emission scenario (Canestrelli *et al.* 2010). The A1B emission scenario is based on high-energy requirements from balanced energy sources, but with emissions less than the A1Fl (Fossil Intensive) scenario. The LIG layers were available at multiple resolutions, so the highest resolution of 30 arc-seconds (equivalent to ~ 1 km) was used; however, the LGM and future layers were only available at 2.5 arc-minute resolution (~5 km). The altitude layer for each resolution was used to generate slope and aspect layers using the Terrain package in R. A total of 22 of these ecogeographical variables (EGV) was used to select those with the greatest influence (highest marginality) on the niche of *M. gigas* using ENFA in the 'adehabitat' package in R (Ecological Niche Factor Analysis; Hirzel *et al.* 2002). The ENFA uses only presence data, so is therefore not biased by the method of selecting absence localities (Hirzel *et al.* 2002). As not all of these EGV layers were normally distributed, each layer was transformed by taking the square root, the square or cube of the data, as ENFA requires normal and symmetrical variables (data not shown). The nine layers with the highest absolute value of marginality were tested for pairwise correlations on values extracted from the occurrence records, with any two variables considered highly correlated when the correlation coefficient was > 0.9. We excluded those variables that were highly correlated with more than one other variable, such that those variables on a monthly or annual scale were excluded over quarterly scales. This reduced the variables to six climatic variables (Bio1, Bio3, Bio11, Bio15, Bio16, Bio18) that were used further in the ecological niche modelling (ENM) in MaxEnt (Phillips *et al.* 2006). The MaxEnt program uses maximum entropy to model species geographic distributions, and its predictive performance has been shown to be highly robust compared to other methods (Elith *et al.* 2006). To assess the ability of MaxEnt to accurately predict the current distribution, and to project the past and future distributions of *M. gigas*, we used 10–fold cross-validation and assessed the predictive performance on the held-out folds using the area under the receiver operating characteristic curve (AUC; Hanley and McNeil 1982). Cross-validation selects 10 different subsets of the presence data (without replacement), with the model run 10 times using each of the subsets as test data with the rest used for training the model; the AUC measures the ability of the predictions to discriminate presence from absence. In order to create simpler, smoother functions of each variable, only hinge features were used with standard regularisation parameters (beta = 1), and a logistic output.

Presence and pseudo-absence dataset generation

Both the ENFA and Maxent analyses were undertaken using two datasets. One dataset comprised extant colonies, which included the genetically sampled colonies plus the locations of additional specimens from Ozcam records (n=48 samples, termed 'extant' hereafter). The second dataset included the extinct southern colonies in addition to the extant colonies (n=53 samples, termed 'extinct' hereafter) under the premise that their extinction was caused by purely human-mediated factors. So therefore their inclusion could be compared against the extant dataset to assess their pre-European ecological niche. However, not all subfossil localities reported in previous literature were included in the 'extinct' dataset as some are represented by remains of uncertain age (i.e. potentially not 'modern') or they are un-vouchered (Churchill and Helman 1990; Molnar *et al.* 1984; Dobson 1880). Two pseudo-absence, or background, datasets ('background-extant' and 'background-extinct' represent locations where *M. gigas* are not found) were selected randomly across the entire Australian continent (approximately 1000 sites) excluding sites within 1° latitude and longitude of presence localities to allow for inaccuracies in presence records. The values of climatic variables for the background dataset (*e.g.* annual temperature where *M. gigas* is not found) is used to contrast against the values of presence climatic variables (*e.g.* annual temperature where *M. gigas* is found) in order to build the model describing the distributional patterns of *M. gigas*. The background-extant dataset included the extinct colonies as absence points as these are sites currently known to lack the presence of *M. gigas*.

Results

The phylogeographic analysis of *M. gigas* mtDNA *CR* sequences reveals a number of highly divergent and largely reciprocally monophyletic clades restricted to different geographic regions, including Cape York, the Kimberley, Arnhem Land, Pilbara, and two new regions: the 'central east' and 'central west' (Figure 2).

Population genetic and demographic analyses

Intra-specific Diversity

The Cape York, Arnhem Land, Northern Agricultural Region (NAR), Rockhampton/Cape Hillsborough (RC), Pilbara, and Camooweal (CM) regions are genetically the most diverse, with high levels of haplotype diversity but low levels of nucleotide diversity (Table 2). These characteristics suggest bottlenecks followed by population growth (Lowe *et al.* 2004). This population growth evidently occurred within regions and did not result in substantial gene flow between regions (based on significant levels of genetic differentiation; Table 3). Moreover, the population growth occurred a significant time ago, as the time to most recent common ancestor (tMRCA) of clades, plus signs for demographic and range expansions of colonies and regions, all pre-date the last glacial maximum (LGM), most by a significant margin (Tables 4 and 5).

Table 2. Diversity summary statistics for each region and/or colonies within regions.

Region	*Colonies within region*	*n*	*#H*	*Tajima's D*	*Fu's FS*	*Haplotype diversity*	*Nucleotide diversity*
Cape York	(KP, MP)	11	4	1.131	3.445	0.746	0.0202
Arnhem land	(CC, CV, NR, PC)	27	10	0.955	0.774	0.889	0.0283
	PC+CV	21	10	1.463	1.683	0.833	0.0240
	NR	6	3	-0.447	0.117	0.733	0.0047
Pilbara		138	14	0.002	2.196	0.781	0.0207
Kimberley	(TC, CS)	17	4	-1.949*	1.089	0.331	0.0067
	TC	16	3	-1.55*	-0.014	0.242	0.0024
Central West	(NAR, CM)	22	11	0.212	-2.417	0.918	0.0148
	NAR (E24, J03)	17	9	-0.191	-4.022*	0.890	0.0077
	CM	5	2	1.225	0.626	0.600	0.0020
Central East	(CA, RC, FR)	20	10	-0.807	0.212	0.863	0.0272
	CA	5	5	-0.764	-0.190	1.000	0.0440
	RC	12	3	1.967	5.059	0.621	0.0174
	FR	3	3	0.000	0.588	1.000	0.0183

* indicates p-value < 0.05. n – number of samples; #H – number of haplotypes.

CA – Central Australia; CC – Cutta Cutta; CH – Cape Hillsborough; CM – Camooweal; CS – Cave Springs; CV – Claravale Station; FR – Flinders Ranges; KP – Kings Plains; MP – Mitchell Palmer; NAR – Northern Agricultural Region; NR – Nourlangie Rock; PC – Pine Creek; RC – Rockhampton and Cape Hillsborough; TC – Tunnel Creek.

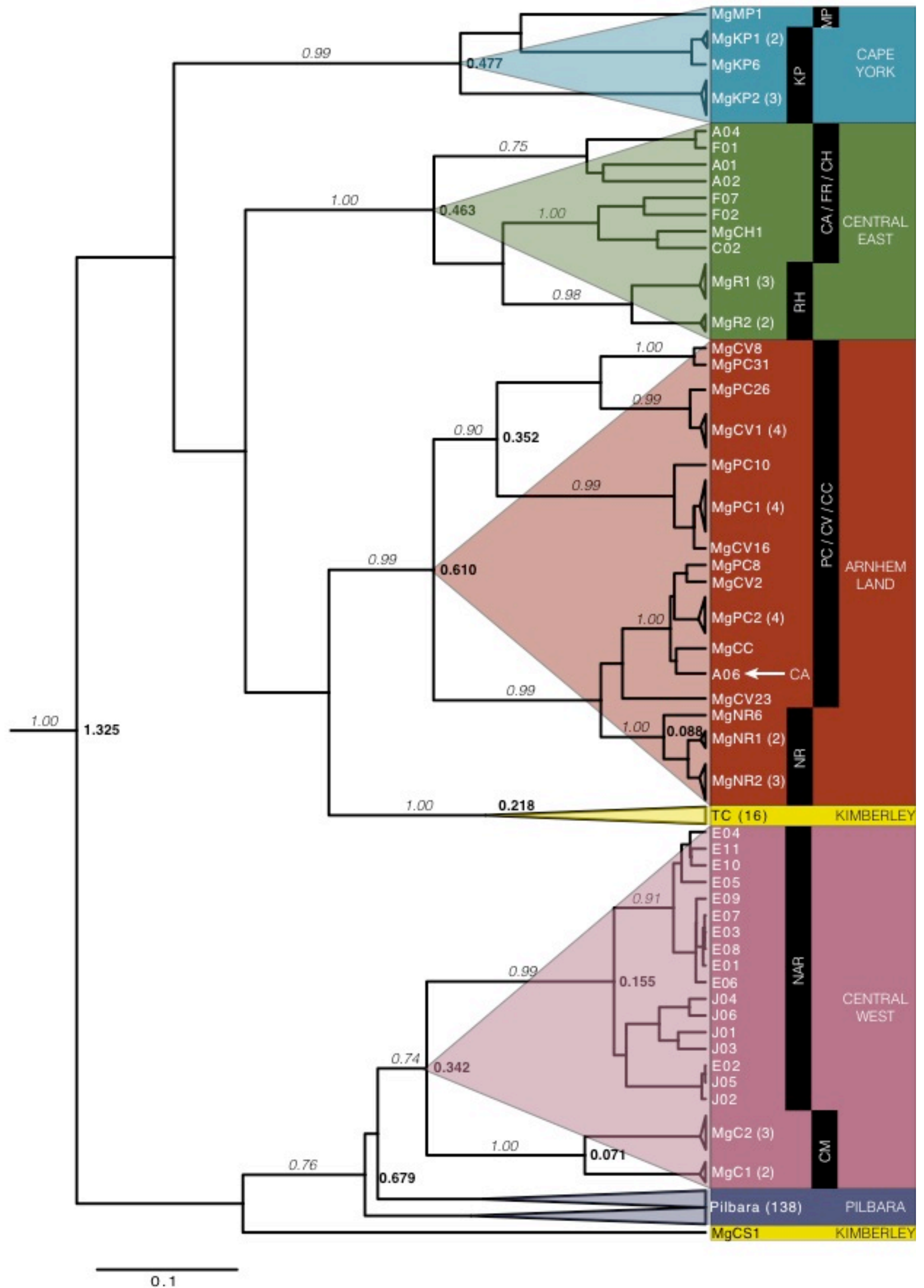


Figure 2. BEAST rooted phylogeny of intra-species diversity using HKY+G substitution model and showing all 235 sequences (outgroups have been removed for clarity). Numbers (in *italics*) on branches refer to posterior probabilities, and numbers on nodes (in bold) represent the time to most recent common ancestor (tMRCA) of node in Millions of years ago (Mya). CA – Central Australia; CC – Cutta Cutta; CH – Cape Hillsborough; CM – Camooweal; CS – Cave Springs; CV – Claravale Station; FR – Flinders Ranges; KP – Kings Plains; MP – Mitchell Palmer; NAR – Northern Agricultural Region; NR – Nourlangie Rock; PC – Pine Creek; RH – Rockhampton; CH – Cape Hillsborough; TC – Tunnel Creek.

Table 3. F_{ST} indices between colonies.

REGION:	*Arnhem Land*		*Central East*			*Kimberley*	*Cape York*	*Central West*		*Pilbara*
CAVES:	***PC/CV/CC***	***NR***	***CA***	***FR***	***RC (RH/C***	***TC/CS***	***KM (KP/MP)***	***NAR***	***CM***	
PC/CV/CC	-									
NR	0.51*	-								
CA	0.48*	0.69*	-							
FR	0.67*	0.90*	0.00	-						
RC (RH/CH)	0.67*	0.89*	0.17*	0.31*	-					
TC/CS	0.72*	0.91*	0.70*	0.84*	0.84*	-				
KM (KP/MP)	0.56*	0.79*	0.42*	0.63*	0.66*	0.81*	-			
NAR	0.72*	0.89*	0.69*	0.81*	0.82*	0.87*	0.78*	-		
CM	0.64*	0.94*	0.58*	0.85*	0.83*	0.90*	0.73*	0.77*	-	
PILBARA	0.61*	0.72*	0.54*	0.61*	0.62*	0.66*	0.56*	0.57	0.55*	-

* indicates p-value < 0.05, which represent colonies with significant genetic differentiation.

CA – Central Australia; CM – Camooweal; CV – Claravale Station; FR – Flinders Ranges; KM – Kings Plains and Mitchell Palmer; NAR – Northern Agricultural Region; NR – Nourlangie Rock; PC – Pine Creek; RC – Rockhampton and Cape Hillsborough; TC – Tunnel Creek.

The F_{ST} indices between individual caves within colonies/regions were all non-significant (PC:CV = 0.0265; PC:CC = -0.126; CV:CC = 0.131; RH:CH = 0.85; KP:MP = 0.295)

Table 4. Mean time estimates and associated 95% HPD for time to most recent common ancestor (tMRCA) of clades (defining regions) and/or colonies within clades (regions) from the BEAST run (HKY+G).

Region	*Caves/colonies within region*	*Mean tMRCA*	*2.5% qt*	*97.5% qt*	*PP*
Cape York	(KP, MP)	476,600	125,100	954,200	0.99
Arnhem Land	(CC, CV, NR, PC)	610,200	180,300	1,133,000	0.99
	TC	218,100	33,400	473,600	1.00
Central West	(NAR, CM)	342,300	93,900	661,600	0.75
	CM	70,600	4,700	179,500	1.00
	NAR (E24, J03)	155,000	37,600	309,400	0.98
Pilbara + Central West		679,600	212,100	1,231,500	0.76
Central East	(CA, RC, FR)	462,700	144,000	887,800	1.00

Evolutionary rate – 4.217 x 10^{-8} substitutions per site per year = 4.217% per Million years.

CA – Central Australia; CC – Cutta Cutta; CM – Camooweal; CV – Claravale Station; FR – Flinders Ranges; KP – Kings Plains; MP – Mitchell Palmer; NAR – Northern Agricultural Region; NR – Nourlangie Rock; PC – Pine Creek; RC – Rockhampton and Cape Hillsborough; TC – Tunnel Creek.

Despite the overwhelmingly regional signature in the phylogeographic pattern of *M. gigas* mtDNA, some colonies contained individuals from multiple clades (e.g. similar haplotypes shared between individuals from Arnhem Land and Central Australia, see arrowed sample in Figure 2), indicating either incomplete lineage sorting, or ongoing dispersal. In general, shared haplotypes were observed most often in the largest colonies, such as those at Pine Creek (PC), and Claravale (CV). Some of these colonies occur in natural caves but others are in abandoned mine adits (e.g. PC in Arnhem Land).

Special phylogenetic relationships between extinct southern colonies and geographically distant extant colonies indicate ancient patterns of divergence at the regional level within *M. gigas* (Figures 1 and 2). A shared genetic history is indicated prior to 343 thousand years ago (kya) (95% HPD: 94–662 kya; Table 4) between extinct colonies in the NAR (near Perth) and the extant colonies at CM (in western Queensland; located 2500 km away). An older connection, with an estimated age of at least 679 kya (95% HPD: 212 kya–1.23 Million years ago (Mya); Table 4), is indicated between the combined NAR/CM clade (designated 'Central West; Figure 1) and colonies in the Pilbara (only 1100 km away from the NAR). Additionally, the extinct colonies of CA and FR shared a genetic history with the extant RC colonies (designated 'Central East'; Figure 1) at least 463 kya (95% HPD: 144–888 kya; and at a distance of 1360–1740 km) rather than with the geographically more proximate CM colonies in western Queensland (which are 635–1150 km away). While the colonies in the Central West region are all reciprocally monophyletic, those of the Central East clade are not: *M. gigas* from the Flinders Ranges share similar haplotypes with individuals from both the Central Australia and Cape Hillsborough colonies.

Environmental variables involved in extinctions

Of the 22 ecogeographical variables used in the ENFA, the six with the highest power for predicting *M. gigas* presence (but with correlation coefficients < 0.9) included: annual mean temperature (bio1), isothermality (bio3), mean temperature of coldest quarter (bio11), precipitation seasonality (bio15), precipitation of wettest quarter (bio16), and precipitation of warmest quarter (bio18). Based on these variables it appears that all extant *M. gigas* inhabit regions that fall within a distinct climatic niche (Table 5), mainly limited by cold and dry stress. For example, mean temperatures of the coldest quarter range between 15–26° C for presence localities compared to 9–21° C for absence localities; precipitation of the wettest quarter is between 100–1100 mm for presence compared to < 500 mm for absence localities; and precipitation of the warmest quarter is between 100–600 mm for presence vs. < 300 mm for absence localities. Environmental conditions in arid regions of the country where *M. gigas* colonies have undergone recent extinction consistently resemble the pseudo-absence localities generated for this study; we infer that these extinct colonies existed under 'marginally suitable'

ambient climates for this species. Moreover, some extant colonies also resemble the pseudo-absence localities for one or a few variables (Figure S2) and also might be inferred to exist under marginal conditions for sustaining viable *M. gigas* colonies. Accordingly, we define those cyan-green colourations on the modelling maps to indicate ‘marginal’ habitats, and the yellow-red colourations to indicate ‘suitable’ habitat (Figure 3).

Table 5. Tests for demographic and range expansion for each clade (region) and/or colonies within clades (region).

Region	Colonies within regjon	n	Test for demographic expansion								Test for range expansion							
			Tau	T (yrs)	2.5% qt T	97.5% qt T	Theta 0	Theta 1	SSD	Hg	Tau	T	2.5% qt T	97.5% qt T	Theta 0	M	SSD	Hg
Cape York (KP, MP)		7	8.12	377,557	0	487,675	0	29.72	0.079	0.19	7.66	358,028	112,881	626,983	0	4.98	0.05	0.19
Arnhem land (PC, CV, NR)		27	9.39	436,608	192,527	608,549	0	23.36	0.021	0.037	7.4	344,079	180,247	582,483	1.58	6.3	0.014	0.04
Kimberley	(TC, CS)	17	3	139,491	19,071	139,491	0	0.39	0.037	0.374	10.7	494,826	17,948	3,966,788	0.21	0.3	0.018	0.373
Central West (NAR, CM)		22	7.92	368,257	0	480,773	0	6.13	0.026	0.04	0.69	32,548	7,603	3,237,714	3.91	99999	0.02	0.04
	NAR	17	2.05	95,319	30,241	143,033	0	99999	0.002	0.057	2.05	95,323	20,104	141,349	0	99999	0	0.06
	CM	5	0.94	43,707	0	131,046	0	99999	0.054	0.4	0.94	43,927	0	4,219,616	0	99999	0.05	0.4
Pilbara		138	-	-	-	-	-	-	0.075*	0.136*	7.08	320,830	170,077	501,316	0.27	1.12	0.04	0.14
Central East (CA, FR, RC)		14	7.52	349,658	104,346	507,745	0	18.2	0.047	0.136^	4.4	204,587	106,478	448,317	3.21	40.89	0.05	0.136^
	(CA, FR)	8	*5.29*	*245,784*	*53,704*	*2,681,118*	*5.18*	*99999*	*0.222*	*0.646*	*5.29*	*245,784*	*60,307*	*422,473*	*5.18*	*99999*	*0.222*	*0.646*
	CA	5	3.81	177,279	0	606,643	9.58	99999	0.071	0.18	3.81	177,232	120,065	561,346	9.58	99999	0.071	0.18
	FR	3	6.39	297,055	236,754	326,754	0	38.28	0.201	0.667	6.23	289,706	196,221	321,819	0.01	99999	0.203	0.667

* indicates p-value < 0.05, ^ indicates p-value < 0.10 – representing evidence against null hypothesis of expansion.

SSD – Sum of squared deviations (one measure of fit of model to the data); Hg – Harpending's raggedness index (another measure of fit of model to the data).

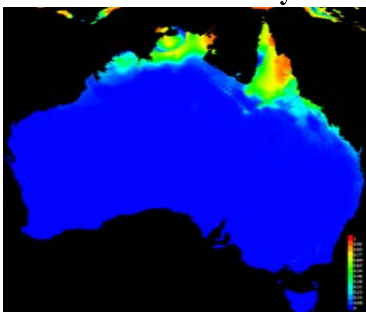
Underlined – Regions; *Underlined italics* – sub-region.

1

Figure 3. Ecological Niche Modelling results, showing for both extant and extinct datasets (in rows): predicted current, past and future distributions. Predicted distributions are logistic outputs, from low values (blue, 0-0.2) to high values (red, 0.8-1.0). The cyan-green colourations indicate 'marginal' habitat, while the yellow-red colourations indicate 'suitable' habitat

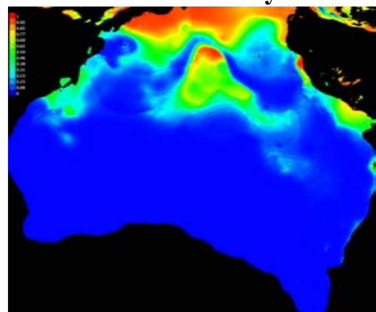
Extant dataset

LIG at ~120-140 kya



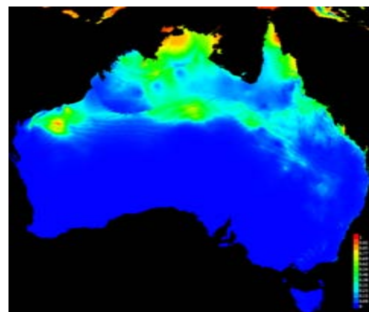
Resolution: 30 arc-second

LGM at ~21 kya



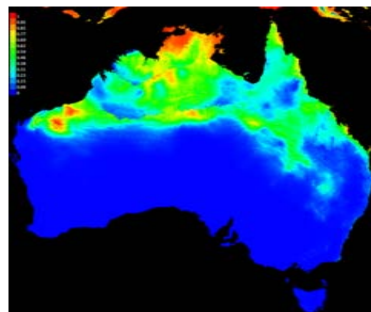
Resolution: 2.5 arc-minute

Current



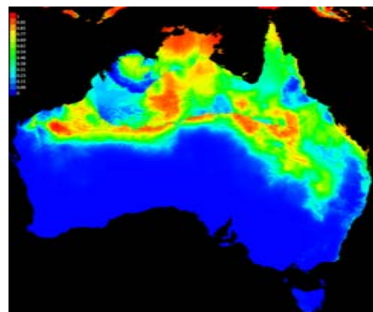
Resolution: 30 arc-second

2030



Resolution: 2.5 arc-minute CSIRO Mk3.0 A1B 2030

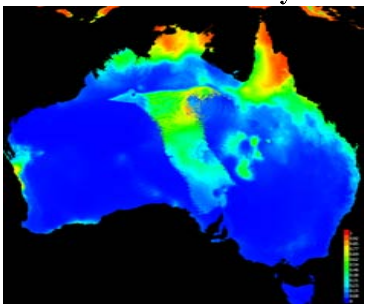
2080



Resolution: 2.5 arc-minute CSIRO Mk3.0 A1B 2080

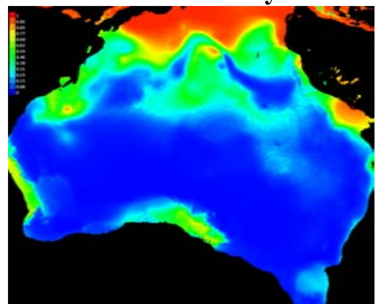
Extinct dataset

LIG at ~120-140 kya



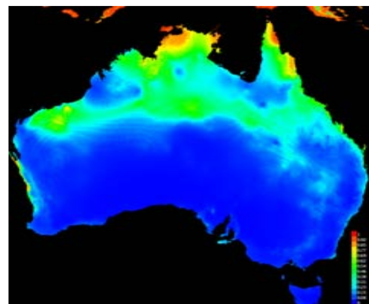
Resolution: 30 arc-second

LGM at ~21 kya



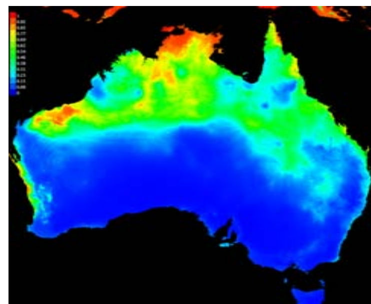
Resolution: 2.5 arc-minute

Current



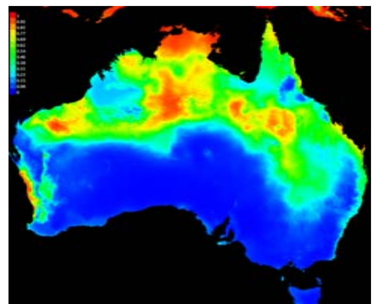
Resolution: 30 arc-second

2030



Resolution: 2.5 arc-minute CSIRO Mk3.0 A1B 2030

2080



Resolution: 2.5 arc-minute CSIRO Mk3.0 A1B 2080

The fit of the MaxEnt model to the distribution of the species was high (AUC values = 0.943 ± 0.022 for the extant dataset; AUC = 0.880 ± 0.083 for the extinct dataset), therefore we feel confident there is an association between the bioclimatic variables used and the current distribution of *M. gigas* (a random prediction corresponds to an AUC of 0.5). However, this might reflect the influence of the climate in promoting the availability of suitable microclimates inside caves, or another indirect cause, such as climate-derived responses of vegetation habitats that ultimately determine prey availability, the climatic preferences of introduced competitors; or even some interaction of all of these.

With Kruskal-Wallis analysis of variance we found a significant difference in values for each of the 9 important bioclimatic variables between locations where *M. gigas* are present vs. absent (Table 5 and Figure S2). However, of the 22 original bioclimatic variables, even the variable with the lowest absolute marginality score (maximum temperature of the warmest month) still had significantly different values between presence and absence locations (Figure S2), suggesting that many aspects of the ambient climatic conditions play a role in determining *M. gigas* presence in the landscape. When the extinct presence locations were used to model the current distribution of *M. gigas*, the limiting factors predicting *M. gigas* extinctions were mainly precipitation of the warmest quarter (i.e. <130 mm for the Flinders Ranges, NAR, and Central Australia colonies vs. an average of 324 mm for the extant colonies), and mean temperature of the coldest quarter (i.e. 11° C for the Central Australia colony vs. an average of 20° C for the extant colonies; Figure S2).

Climatic modelling

When reconstructions of past climate conditions were used to hindcast *M. gigas* distributions into the LGM (21 kya) and LIG (120 kya), we found the hindcast distribution to not differ substantially from the current distribution as based on the extant dataset (Figure 3). However, when the current distribution is based on the extinct dataset, the LIG and LGM distributions differ significantly from the modern distribution (Figure 3).

Last Glacial Maximum (LGM)

The ENM reconstruction using the extinct dataset predicted a slight increase in the area of extent of 'suitable' habitat in southern Australia from current to hindcast distributions. The main increase in 'suitable' regions during the LGM compared to current day is a strip of 'suitable' habitat along the coast margin of the Nullarbor Plain, the continental shelf currently submerged under Torres Strait and off the southern Queensland coast, and a small area along the WA coast; the Bassian Plain is identified as marginally suitable habitat (Figure 3). Although no recent colonies of *M. gigas* are known from most of these regions, the NAR colonies fall within the predicted 'suitable' habitat along the WA coast; this finding suggests that the NAR colonies may have remained viable even during the aridity of the LGM.

Last Inter-Glacial (LIG)

During the LIG, *M. gigas* distributions were predicted to span more of central Australia, with regions of high suitability dotted across the Lake Eyre Basin and potentially linking Cape York and the Camooweal/coastal Queensland sites with Central Australia and the Flinders Ranges (Figure 3). However, even during the favourable climatic conditions of the LIG, no increase in 'suitable' habitat linked the NAR region with any other *M. gigas* colonies (Figure 3). This long period of isolation for the NAR colonies is consistent with the time to most recent common ancestor (tMRCA) of the NAR colonies (155,000 ya; 95% HPD 37,600–309,400 ya). The last active gene flow between the NAR and other populations must have occurred under conditions of more widespread warmth and humidity even than predicted for the LIG, though perhaps only ephemeral in nature.

Future

When the predicted bioclimatic variables for the years of 2030 and 2080 are used with the extant dataset to model the extent of the *M. gigas'* ecological niche into the future, there is a slight increase in 'suitable' habitat exhibited for inland and coastal Queensland. However, it is only when the extinct dataset is used that an additional increase in 'suitable' habitat is seen along the WA and SA coastlines.

Table 5. Kruskal-Wallis analysis of variance between extant presence and absence locations for each of the nine most important bioclimatic variables.

Bioclimatic variables with highest marginality values	Presence		Background		Difference between Presence and Background (95% CI)	Kruskal-Wallis	
	Mean	Std deviation	Mean	Std deviation		$X^2(1)$	p-value
Bio1 Annual mean temp (°C)	25.992	1.815	21.226	3.811	4.766 (4.190, 5.342)	76.1	2.2e-06 ***
Bio3 Isothermality	55.021	5.029	49.224	3.597	5.796 (4.320, 7.273)	53.7	2.4e-13 ***
Bio6 Min temp of coldest month (•C)	*13.085*	*3.395*	*6.684*	*3.374*	*6.402 (5.395, 7.408)*	*90.1*	*2.2e-06* ***
Bio11 Mean temp of coldest quarter (°C)	21.025	3.16	14.305	3.931	6.720 (5.772, 7.668)	87.8	2.2e-16 ***
Bio12 Annual precipitation (mm)	*952.042*	*476.011*	*429.669*	*302.057*	*522.372 (382.949, 661.796)*	*58.5*	*2.0e-14* ***
Bio13 Precipitation of wettest month (mm)	*221.313*	*109.5*	*79.669*	*67.979*	*141.643 (109.583, 173.704)*	*76.5*	*2.2e-16* ***
Bio15 Precipitation seasonality (Coefficient of Variation)	96.792	16.029	56.386	31.939	40.405 (35.362, 45.449)	62.1	3.3e-15 ***
Bio16 Precipitation of wettest quarter (mm)	599.625	320.869	208.138	182.702	391.488 (297.663, 485.312)	73.4	2.2e-16 ***
Bio18 Precipitation of warmest quarter (mm)	344.417	147.245	155.422	125.571	188.995 (145.568, 232.422)	72.8	2.2e-16 ***
Bioclimatic variable with lowest marginality for comparison							
Bio5 Maximum temperature of warmest month (°C)	36.81	2.845	35.251	3.771	1.559 (0.702, 2.417)	8.4	0.003691**

* indicates p-value < 0.05; ** indicates p-value < 0.01; *** indicates p-value < 0.001

Italicised variables indicate those variables identified in ENFA as important to predicting *M. gigas* presence but removed due to high correlation with other variables.

Discussion

Despite its capacity for dispersal via flight and a wide distribution on the Australian continent, *M. gigas* exhibits surprising levels of phylogeographic structure. This suggests long-term isolation and deep divergence between allopatric populations and supports Finlayson's (1961) characterisation of *M. gigas* as a relictual taxon.

Genetic substructuring

Previous population genetic studies have suggested strong female philopatry (Worthington Wilmer *et al.* 1999; Worthington Wilmer *et al.* 1994). This interpretation is reinforced by the highly significant F_{ST} values between the majority of all colonies sampled in our study. The pattern of phylogenetic relationships among colonies of *M. gigas* at the continental scale indicates very ancient patterns of shared genetic history, possibly dating back to more humid periods of the Pleistocene, as landscapes were eroding and current habitats were developing into the forms we see today.

<u>Northern colonies</u>

The Cape York and Arnhem Land colonies are located within regions that appear to be to climatically stable regions for *M. gigas* (Figure 3). This interpretation is reinforced by the observation that these are some of the genetically most diverse *M. gigas* colonies, with long histories of genetic isolation but also with signs of population expansions since the Middle Pleistocene (Figure 4). Although our results support the notion that *M. gigas* females are highly philopatric (Worthington Wilmer *et al.* 1999), the observation of haplotypes belonging to multiple mtDNA *CR* clades in the larger colonies at Pine Creek/Claravale and in the Pilbara does imply occasional female dispersal between colonies. We suspect that this may not reflect natural behaviour for the species. Extant mine roosts of *M. gigas* in Arnhem Land and the Pilbara are likely subject to a high level of human disturbance (Armstrong and Anstee 2000) and this may be disrupting the usually strong site-fidelity of these colonies, encouraging dispersal between caves at a local level. However, this pattern of infrequent gene flow between colonies has only been detected in the better-sampled colonies with large sample sizes (e.g. the Pilbara, Arnhem Land, NAR), and it may be more widespread than suggested by present datasets.

<u>Central West</u>

Each of the Central West colonies appears to represent an ancient lineage, with NAR and CM having diverged > 300 kya, and the NAR and CM colonies together from those in the Pilbara even earlier at > 680 kya. The genetic link between the NAR and CM colonies spans 2500 km of habitat with currently relatively low relief and therefore a low cave forming tendency. Where subfossil remains of *M. gigas* have been sampled from suitable intervening caves (Alice Springs/Central Australia), quite different haplotypes are currently found (haplotypes belong to the Central East clade). However, these individuals belonging to the Central East clade found in

recent times may be remnants of a later expansion westwards from coastal Queensland, replacing any Central West clade individuals still remaining from an earlier expansion from CM to NAR. The ENM also suggests ancestors of the Pilbara/NAR/CM colonies may have actually inhabited central Australia during past interglacial cycles, as a lack of 'suitable' habitat in the Pilbara region but a large region of 'climatically suitable' habitat in central Australia is suggested for *M. gigas* during the LIG. Subsequent dispersal of *M. gigas* into the Pilbara region appears to have occurred around Terminal IV (range expansion at 320 kya); colonisation of the NAR probably after the LIG (range expansion at 95 kya); and CM perhaps occupied only in the lead up to the LGM (range expansion at 44 kya). Although the 95% HPD are quite wide for all these population expansions – thereby creating uncertainty in correlating expansion events with particular climatic cycles – the suggested expansion and colonization events are consistent with the changing distribution of suitable *M. gigas* habitat under glacial and interglacial conditions (Figure 3). In order to fully explore the possibility of central Australia as an interglacial refugium, the extinct *M. gigas* colonies in the Rawlinson and Kathleen Ranges, and near Warburton in Western Australia, would need to be sampled (Churchill and Helman 1990).

The persistence of the NAR colonies through the highly arid LGM into the 19th century is also consistent with the modelling results. Furthermore, it serves to highlight the importance of the southwest corner of Australia, which was a refugium for many species during periods of peak aridity (Byrne *et al.* 2008; Dodson 2001). However, in the case of *M. gigas* the paleontological record suggests that the core refuge lay not in the forested region of the southwest but somewhat to the north. For while the northern NAR colonies sampled in this study contain both skeletal remains and large piles of guano (indicating large colonies), some caves south of NAR, such as those at Wanneroo-Yanchep, contain only sporadic skeletal remains as surface deposits (Bridge 1975). Our findings support Bridge's (1975) suggestion that colonies south of NAR represented relatively recent southward expansions of *M. gigas*. Ancient DNA studies of material from the more southerly caves would provide a direct test of this proposition.

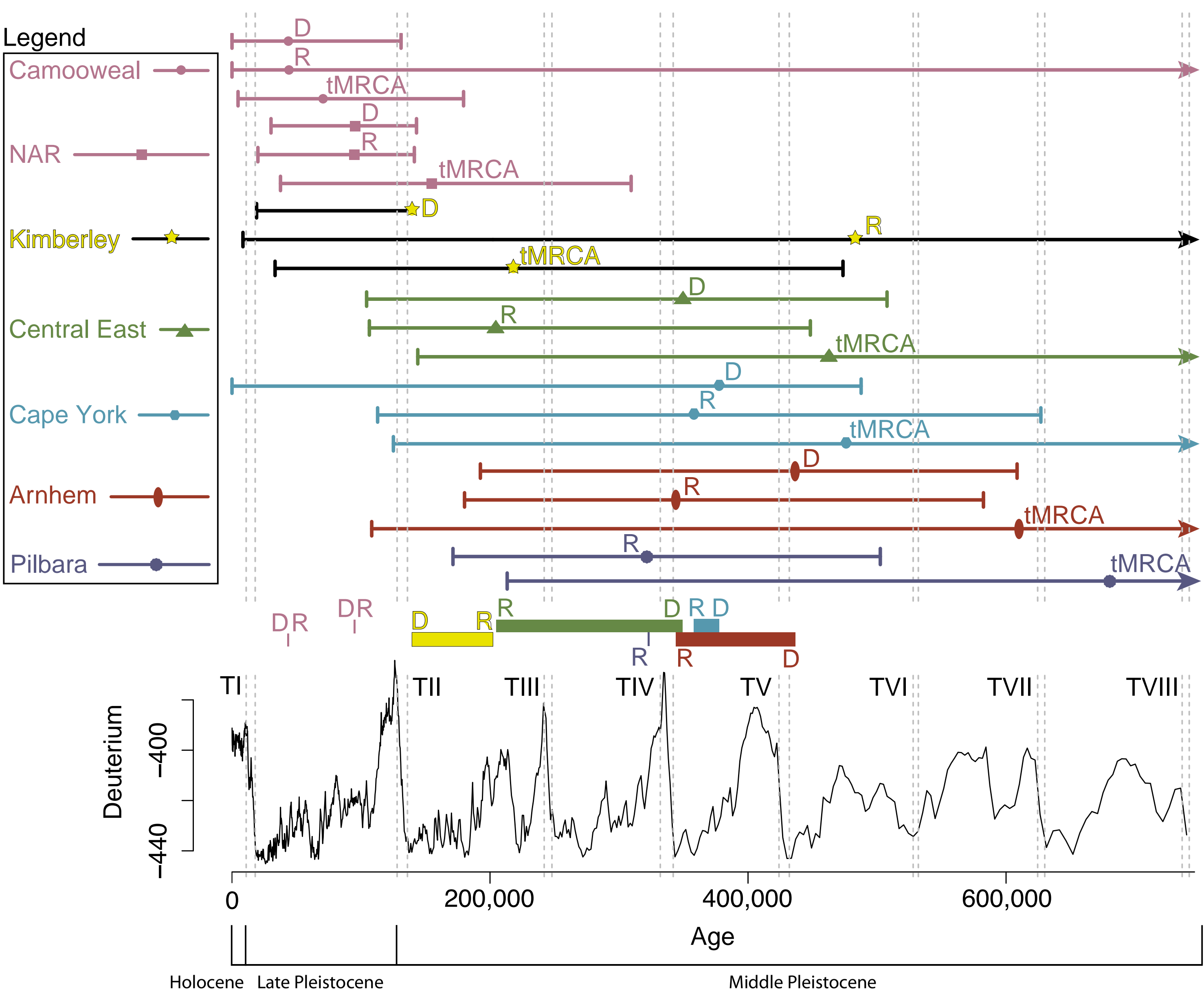
Central East

The low F_{ST} values between the CA and FR samples suggest some level of female dispersal, although possibly only on a regional scale (Table 3). However, the lack of reciprocal monophyly for CA, FR and CH colonies of the Central East clade clearly indicates secondary dispersal since the establishment of the CA and FR colonies by at least 100 kya. This is not surprising, as the ephemeral, yet cyclical, nature of flooding in the Lake Eyre basin would provide suitably warm/humid conditions and an expanded regional prey base to support repeated long-distance dispersal events. To test this theory, colonies around the type locality at Mount Margaret in western Queensland (Dobson 1880), and at any recently extinct colonies in NSW, would have to be sampled for genetic analysis (Molnar *et al.* 1984).

The evidence for expansions in both extant northern and extinct southern colonies do not consistently support any one of the three hypotheses. The lack of genetic links between the extinct colonies and their geographically proximate extant neighbour appears to support hypothesis three. However, taking into account the biogeographic barriers highlighted in Figure 1 and the potential for central Australian playa lakes to act as corridors for gene flow – for example Lake Eyre basin in eastern Australia linking RC to CA/FR colonies and Lake Amadeus/Lake Mackay in western Australia linking CM to NAR colonies – we regard hypothesis one (Periodic Southern Expansion Hypothesis) to be more likely than the others. Furthermore, where single individuals show haplotypes typical of another region/colony (i.e. the shared haplotypes between: Central Australia and Arnhem Land, the Flinders Ranges and Central Australia, the Pilbara and Kimberley; Cape Hillsborough and Central Australian individuals), the majority of the exotic haplotypes are of northern rather than southern origin, as would be predicted under hypothesis one (Periodic Southern Expansion Hypothesis).

Temporal signatures

Macroderma gigas has a deep evolutionary history in Australia – tMRCA is 1.32 Mya from the BEAST analysis – and the species clearly has been impacted by repeated glacial cycles. This long history is supported by the fossil record (Churchill 2008) and is consistent with the fact that many of the landscape features they inhabit have been tectonically stable since at least 80 Mya (the Pilbara Block; Kimberley Block; and Alice Springs sites in the Arunta Block; Kranendonk *et al.* 2002). Nevertheless, it is clear that regular gene flow between the major regions and even between relatively close colonies in most cases ceased in the order of hundreds of thousands of years ago. This history of long term isolations may reflect the formation and reinforcement of the major barriers/impediments to gene flow in arid and northern Australia, including: the Simpson Desert (initially formed 1 Mya; Fujioka *et al.* 2009); the stony deserts of central Australia (initially formed 3-2 Mya; Fujioka *et al.* 2005); the Murchison River gorge (Edwards 2007); the Carpentarian semi-arid barrier (Lee *et al.* 2008); the Canning basin (Pepper *et al.* 2011); the Kimberley Plateau-Arnhem Land Barrier (Eldridge *et al.* 2012); the Burdekin dry habitat gap (Chapple *et al.* 2011); and to a lesser extent the St. Lawrence-Gladstone dry corridor (Webb and Tracey 1981); and the East-West Kimberley Gap formed by a basalt range (Potter *et al.* 2012). Although the tMRCA for the NAR colonies is just prior to the LIG, consistent with hypothesis one (Periodic Southern Expansion Hypothesis), the 95% HPD are too wide for direct correlation between divergence of clades and climatic cycles (Table 4 and Figure 4). Accordingly, the temporal signal of *M. gigas* does not exclusively support any one of the three hypotheses being investigated.



Legend
Camooweal
NAR
Kimberley
Central East
Cape York
Arnhem
Pilbara
D
R
tMRCA
DR
Deuterium
−400
−440
TI
TII
TIII
TIV
TV
TVI
TVII
TVIII
0
200,000
400,000
600,000
Age
Holocene
Late Pleistocene
Middle Pleistocene

Figure 4. Correlation of time to most recent common ancestor (tMRCA) of *M. gigas* clades from BEAST analysis, plus demographic (D) and range (R) expansion dates from Arlequin analysis, with EPICA climatic cycles using changing deuterium levels in Antarctica as a proxy for changing temperature. TI-TVIII represent the termination of glacial periods in the southern hemisphere (EPICA *et al.* 2004).

Climate modelling signatures

The modelling of current distributions with the extant dataset does not identify the locations of extinct colonies as having 'suitable' habitat. Furthermore, the hindcast distributions using the extant datasets also failed to identify these extinct colony locations as having 'suitable' habitat during either arid (LGM) or humid (LIG) environmental conditions. Four possible reasons can be posed to explain these discrepancies. First, the proxy we used (of ambient climate for cave microclimate) may not be particularly relevant to understanding the distribution of *M. gigas*. However, the strong correlation between cold/dry stress and the *M. gigas* absence datasets concurs with the known physiological constraints of this species. Second, the use of the extant dataset may not provide a complete picture of the species' ecological niche. Yet, the fit of the MaxEnt models to the current distribution of the species was high (AUC values much greater than 0.5), suggesting that we are capturing an accurate representation of the ecological niche. Third, we may not be modelling the correct temporal period when these expansions occurred. However, population expansions predicted from the genetic data do appear to coincide with some of the time periods modelled using the ENM (i.e. LIG). Last, *M. gigas* may not need permanent or even persistent 'suitable' habitat to expand its range. Instead, it may be able to take advantage of ephemeral climate changes producing transient conditions that are 'suitable' for dispersal. Of the four, the last explanation seems the most plausible but is perhaps the most difficult to test.

The NAR colonies appear to have been established just prior to the LIG and experienced an expansion just after the LIG; yet the modelling fails to identify this area as having 'suitable' habitat during the LIG. All observations may be correct if intermittent and ephemeral late summer rainfall in the NAR provided opportunities for *M. gigas* to disperse out of large breeding colonies and find temporary refuge in rock-overhangs or shallow caves lacking an appropriate microclimate. These types of heavy rainfall events may have provided both the high humidity level and erupting insect/rodent populations required for *M. gigas* to survive and colonise the arid centre of the continent. Furthermore, modelling for the LIG indicates that regions in the Lake Eyre basin may have had habitat 'suitable' for supporting small *M. gigas* colonies, thereby forming the ancestors of the Central East clade. The evolutionary histories of NAR, CA and FR colonies thus provide support for hypothesis one (Periodic Southern Expansion Hypothesis). On the other hand, the more recent (post-LIG) establishment and expansion of the Camooweal caves colony points to possible northward expansion of *M. gigas*, perhaps in response to increasing aridity in central Australia. If this is the case, one could postulate that the CA, FR, and NAR colonies may have survived the increasing aridity of the last glacial cycle by using caves with suitable microclimates. The evolutionary history of the CM colony therefore offers some support for hypothesis two (Northern Relocation Hypothesis). Support for multiple hypotheses seems contradictory. However, it may actually be correct if the recent evolutionary history of *M. gigas*

has been one of contrasting regional responses to global climatic events. Indeed, given the evidence of deep genetic divergences within *M. gigas*, there is no *a priori* reason to expect all populations of the species to possess a common demographic history.

Assumptions inherent in the ecological niche modelling used

The type of climate envelope modelling undertaken here estimates the realised niche of *M. gigas*, as it utilises presence records which already reflect the impacts of biotic interactions including resource competition, predator-prey dynamics, etc. Moreover, as we have not investigated caves with appropriate microclimates within the 'suitable' habitat regions, the 'true' realised niche of *M. gigas* will be a subset of the projected 'suitable' habitat as defined by the model. The underlying geology and degree of weathering across the Australian landscape may have profoundly influenced the actual distribution of *M. gigas* in each of the time periods projected here. Although altitude, slope and aspect were all included in the ENFA (which selected most important variables determining *M. gigas* presence) the coarse-scale resolution of these layers (2.5' and 30") may have obscured any value they may have had in predicting both the current distribution, and into other temporal periods. Furthermore, it should be noted that the probability of presence predicted by the model does not strictly reflect 'true' probability of *M. gigas* colonies 'on the ground'. Many more factors need to be taken into account, including but not restricted to: availability of suitable cave sites; the level of human disturbance; colony sizes; and the operation of other detrimental factors.

It is interesting to note that while the extinct dataset gave a lower fit to current *M. gigas* distribution, it predicted larger areas of 'suitable' habitat when both hindcasting and forecasting. Overall, these predictions were in better agreement with the sub-fossil record. However, they also suggested that the recently extinct colonies were situated in 'marginal' areas for *M. gigas* and highlighted a possible danger with including extinct populations when doing forecast and hindcast modelling – that predicted distributions will include substantial areas of 'marginal' habitat in addition to genuinely 'suitable' habitat. This may not be the case for less vagile species that have suffered historical declines.

Future distributions

Whether or not *M. gigas* colonies are able to take advantage of the projected increase in 'marginal' habitat in the future will depend on several factors: whether the extant colonies can survive human encroachment into its current habitat over the next hundred years (have current populations been reduced in size too much to rebound?); whether there are appropriately wet/warm periods for expansion of prey populations to allow *M. gigas* to colonise these southern regions (will ENSO cycles retain their warm/wet characteristics?); and ultimately, on whether these regions contain caves with microclimates suitable to support maternal roosting behaviour of *M. gigas* (will caves in these regions develop acceptable microclimates as ambient climates

change?). However, climate is not the only consideration, either in terms of providing conditions for prey eruptions in central Australia, or for *M. gigas* cave microclimate requirements. If *M. gigas* can re-invade currently unoccupied parts of its former range in southern Australia, will there be sufficient ecological communities to sustain them, given the historical decline in small mammals in southern Australia (potentially due to human-mediated landscape modifications and the introductions of alien species; Bilney *et al.* 2010; Burbidge *et al.* 1988; Letnic and Dickman 2006; Menkhorst and Knight 2004; Short and Smith 1994)? Although anthropogenic factors may have contributed to the downward spiral of southern *M. gigas* extinctions historically, how will the continued effects of these factors in the intervening 200 years have combined with anthropogenic climate change to impact these potentially new colonisations of *M. gigas*?

Conclusion

The species distribution modelling for *M. gigas* is an idealised representation of the climatic envelope, with limitations inherent in projecting climates into past and future periods adding additional uncertainty to our forecasting/hindcasting ability. A major limitation of our study arises from the use of atmospheric climate as a proxy for suitable cave microclimate. In particular, the lack of landscape-scale information on rock type and its associated erosional structure (i.e. whether a formation was likely to contain caves), and of measured cave microclimate data in the modelling dataset, potentially restricts our conclusions to generalisations at the continental-scale. The interpretations made were limited to detecting connectedness between broad regions without a detailed analysis of whether the geological terrains would have provided caves over their associated long history of erosion into the features present today. This being said, by combining perspectives of genetic and continental occupancy patterns, we suggest that the past distribution of *M. gigas* has not simply been static with a contraction northwards following changing conditions in the Holocene and more recently after the arrival of Europeans (as suggested by Churchill and Helman 1990; Molnar *et al.* 1984). Rather, we present a more dynamic view where the warm, wet conditions of interglacial episodes allowed the species to expand southwards to occupy regions that became uninhabitable under glacial conditions (i.e. the majority of evidence supports the Periodic Southern Expansion Hypothesis). We suppose that rainfall events in interglacial periods that promoted biomass increases (including prey species) and provided the groundwater for maintaining humidity in caves allowed *M. gigas* to expand across the large swathes of otherwise inhospitable habitat covering the arid centre of Australia. However, this dependence on prey species, such as small mammals that are vulnerable to human disturbance, has also likely been a factor in its decline in the south where species and populations have disappeared from many areas. Genetic evidence from the Camooweal colony suggests that *M. gigas* has also responded to periods of extreme aridity by contracting out of the arid centre (providing minor support for the Northern Relocation Hypothesis).

In the context of ensuring the continued survival of *M. gigas* under future anthropogenic climate change, this new understanding of the past response of *M. gigas* might be applied to current conservation and management efforts, but with care. The ecological niche modelling predicts future climate change may allow *M. gigas* to re-invade currently unoccupied parts of its former range in southern Australia. However, even if predictions of a potentially positive outcome involving an expansion of distribution are accurate, we will also need to consider 1) whether in each geographic population (*sensu* Worthington Wilmer *et al.* 1994, 1999) the abundance of individuals and the number of secure and adequately protected colonies is sufficient for both *in situ* persistence and expansion of abundance and distribution boundaries in the long term; and 2) whether we have modified the landscape and its concomitant ecological structures and function in these currently un-occupied parts of its former range too much for natural processes to sustain these newly colonising populations of *M. gigas*.

Acknowledgements: We thank the Rio Tinto Iron Ore (WA) for funding and permission to use the Pilbara sequence data that were used as context for the extinct populations; and also to Molhar Pty Ltd for providing this data. We would also like to thank the Western Australian Speleological Group for providing access to the NAR caves and to those members who gave generous assistance in collecting the cave subfossil remains: Robert Susac, Greg Thomas, Ian Collette, Melanie Roberts, and Barbara Zakrzewska. Thanks also to Aaron Camens for paleontological expertise, and Neil Anderson (and friends) for providing caving expertise on the Flinders Ranges field trip. This work was supported by the Biological Society of South Australia through the Lirabenda Endowment Fund, the Nature Conservation Society of South Australia through the Conservation Biology Grant, and Australian Research Council grants to Alan Cooper.

The following permits were obtained to enter National Parks, and collect and transport the faunal (subfossil) material: permit no. C25769 from the Government of South Australia, Department for Environment and Heritage; permit nos. CE002666, SF007068 and export permit No. ES002086 from the Government of Western Australia, Department of Environment and Conservation; and import permit no. I9900 from the Government of South Australia, Department for Environment and Heritage.

The names and location details of the NAR caves have been withheld by the request of the Western Australian Speleological Group and the Pilbara cave site information has also been withheld at the request of KA as this data was initially generated for a population level study of the Ghost Bat funded by Rio Tinto Iron Ore (WA), and was only included to add context to the ancient samples.

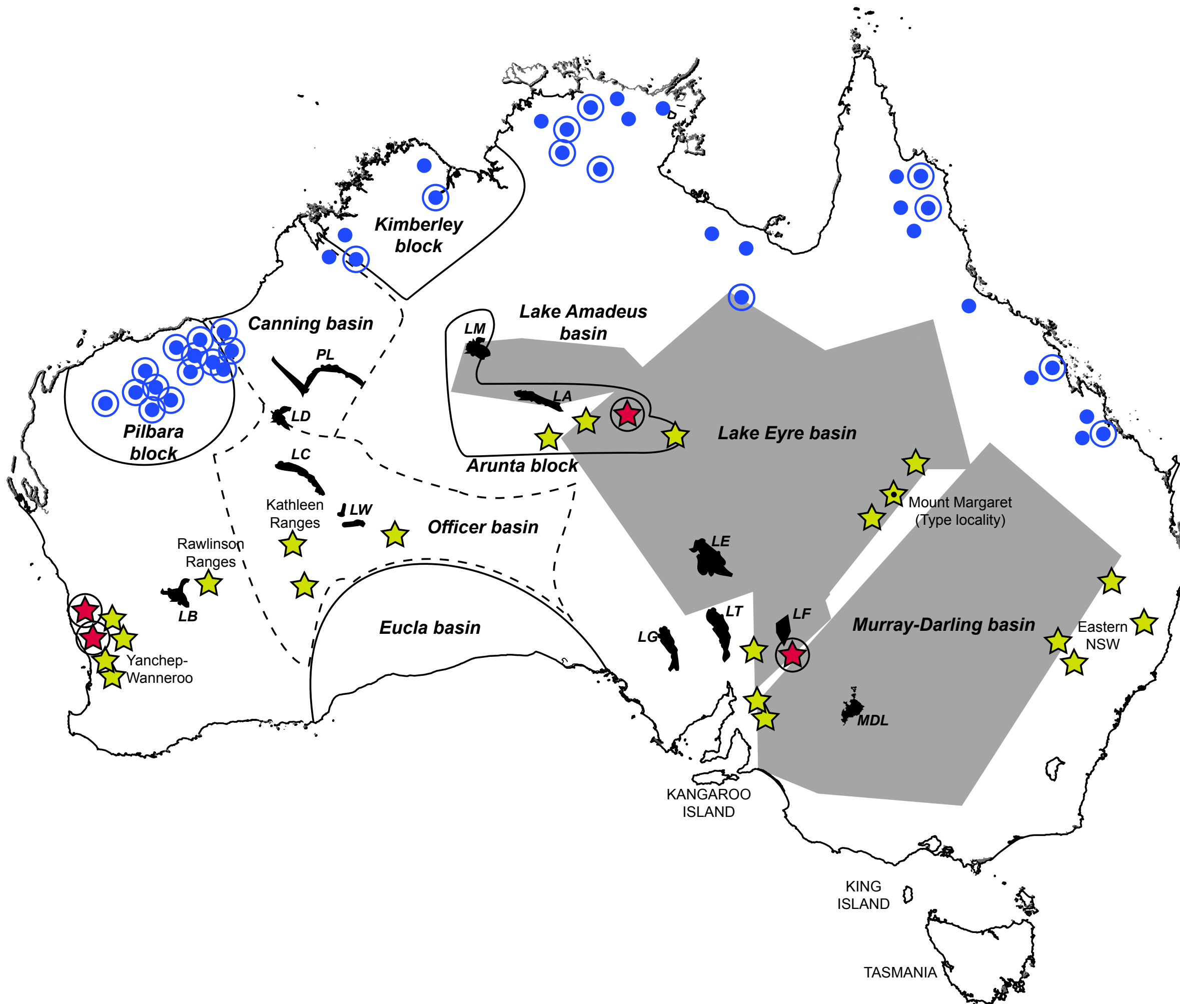


Figure S1. Map showing past and present distribution of *Macroderma gigas* colonies. Blue circles represent extant populations; those with an outer ring are those sampled by Worthington Wilmer *et al.* (1999; 1994) and included in this study. Stars represent extinct populations; red circled stars are the extinct populations sampled in this study; yellow stars are extinct populations not sampled in this study (Eastern NSW, Kathleen Ranges, Rawlinson Ranges, Yanchep-Wanneroo), with a central dot indicating the type locality (Mount Margaret). The arid and semi-arid inland lakes and associated basins are noted: LA, Lake Amadeus; LB, Lake Barlee; LC, Lake Carnegie; LD, Lake Disappointment; LE, Lake Eyre; LF, Lake Frome; LG, Lake Gairdner; LM, Lake Mackay; LT, Lake Torrens; LW, Lake Wells; MDL, Murray-Darling lakes, including Menindee Lakes, Anabranch Lake, Willandra Lakes, Lake Victoria; PL, Percival Lakes.

Table S1. Sample details, including specimen number, museum accession number, latitude and longitude, cave/colonies/region designation.

REGION	*Colonies*	*Cave*	*Specimen*	*Sample details*	*Accession number*	*Location*
Arnhem land	CC	CC	MgCC	WW *et al.* (1999)	AJ130998	Cutta Cutta, NT
Arnhem land	CV	CV	MgCV16	WW *et al.* (1999)	AJ131008	Claravale, NT
Arnhem land	CV	CV	MgCV1a	WW *et al.* (1999)	AJ131005	Claravale, NT
Arnhem land	CV	CV	MgCV1b	WW *et al.* (1999)	AJ131005	Claravale, NT
Arnhem land	CV	CV	MgCV1c	WW *et al.* (1999)	AJ131005	Claravale, NT
Arnhem land	CV	CV	MgCV1d	WW *et al.* (1999)	AJ131005	Claravale, NT
Arnhem land	CV	CV	MgCV2	WW *et al.* (1999)	AJ131007	Claravale, NT
Arnhem land	CV	CV	MgCV23	WW *et al.* (1999)	AJ131009	Claravale, NT
Arnhem land	CV	CV	MgCV8	WW *et al.* (1999)	AJ131006	Claravale, NT
Arnhem land	NR	NR	MgNR1a	WW *et al.* (1999)	AJ131010	Nourlangie Rock, NT
Arnhem land	NR	NR	MgNR1b	WW *et al.* (1999)	AJ131010	Nourlangie Rock, NT
Arnhem land	NR	NR	MgNR2a	WW *et al.* (1999)	AJ131011	Nourlangie Rock, NT
Arnhem land	NR	NR	MgNR2b	WW *et al.* (1999)	AJ131011	Nourlangie Rock, NT
Arnhem land	NR	NR	MgNR2c	WW *et al.* (1999)	AJ131011	Nourlangie Rock, NT
Arnhem land	NR	NR	MgNR6	WW *et al.* (1999)	AJ131012	Nourlangie Rock, NT
Arnhem land	PC	PC	MgPC10	WW *et al.* (1999)	AJ131002	Pine Creek, NT
Arnhem land	PC	PC	MgPC1a	WW *et al.* (1999)	AJ130999	Pine Creek, NT
Arnhem land	PC	PC	MgPC1b	WW *et al.* (1999)	AJ130999	Pine Creek, NT
Arnhem land	PC	PC	MgPC1c	WW *et al.* (1999)	AJ130999	Pine Creek, NT
Arnhem land	PC	PC	MgPC1d	WW *et al.* (1999)	AJ130999	Pine Creek, NT
Arnhem land	PC	PC	MgPC26	WW *et al.* (1999)	AJ131003	Pine Creek, NT
Arnhem land	PC	PC	MgPC2a	WW *et al.* (1999)	AJ131000	Pine Creek, NT
Arnhem land	PC	PC	MgPC2b	WW *et al.* (1999)	AJ131000	Pine Creek, NT
Arnhem land	PC	PC	MgPC2c	WW *et al.* (1999)	AJ131000	Pine Creek, NT
Arnhem land	PC	PC	MgPC2d	WW *et al.* (1999)	AJ131000	Pine Creek, NT
Arnhem land	PC	PC	MgPC31	WW *et al.* (1999)	AJ131004	Pine Creek, NT
Arnhem land	PC	PC	MgPC8	WW *et al.* (1999)	AJ131001	Pine Creek, NT
Cape York	KM	KP	MgKP1a	WW *et al.* (1999)	AJ130995	Kings Plains, Qld
Cape York	KM	KP	MgKP1b	WW *et al.* (1999)	AJ130995	Kings Plains, Qld
Cape York	KM	KP	MgKP2a	WW *et al.* (1999)	AJ130996	Kings Plains, Qld
Cape York	KM	KP	MgKP2b	WW *et al.* (1999)	AJ130996	Kings Plains, Qld
Cape York	KM	KP	MgKP2c	WW *et al.* (1999)	AJ130996	Kings Plains, Qld
Cape York	KM	KP	MgKP6	WW *et al.* (1999)	AJ130997	Kings Plains, Qld
Cape York	KM	MP	MgMP1	WW *et al.* (1999)	AJ130994	Mitchell Palmer, Qld
Central East	CA	AS	A01	SAM	M525/001	Alice Springs, NT
Central East	CA	AS	A02	SAM	M525/002	Alice Springs, NT
Central East	CA	AS	A04	SAM	M526/002	Alice Springs, NT
Central East	CA	AS	A06	SAM	M526/004	Alice Springs, NT
Central East	CA	CA	C02	SAM	M1992	Central Australia, NT
Central East	FR	FR	F01	SAM	M785	Flinders Ranges, SA
Central East	FR	FR	F02	SAM	M2691	Flinders Ranges, SA
Central East	FR	FR	F07	SAM	M3791	Flinders Ranges, SA
Central East	RC	CH	MgCH1	WW *et al.* (1999)	AJ130993	Cape

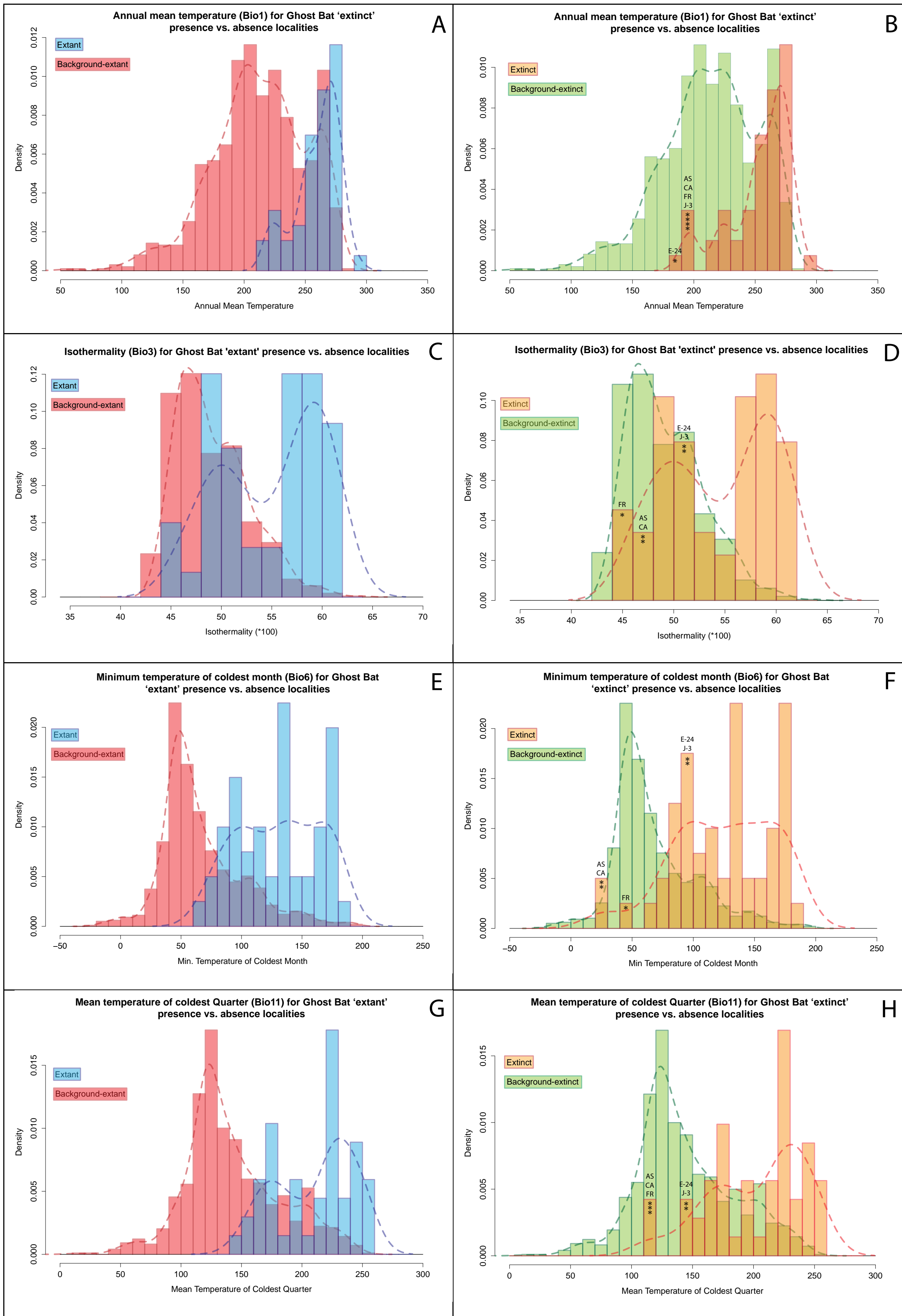
REGION	Colonies	Cave	Specimen	Sample details	Accession number	Location
						Hillsborough, Qld
Central East	RC	RH	MgR1a	WW *et al.* (1999)	AJ130991	Rockhampton, Qld
Central East	RC	RH	MgR1b	WW *et al.* (1999)	AJ130991	Rockhampton, Qld
Central East	RC	RH	MgR1c	WW *et al.* (1999)	AJ130991	Rockhampton, Qld
Central East	RC	RH	MgR2a	WW *et al.* (1999)	AJ130992	Rockhampton, Qld
Central East	RC	RH	MgR2b	WW *et al.* (1999)	AJ130992	Rockhampton, Qld
Central West	CM	CM	MgC1a	WW *et al.* (1999)	AJ130986	Camooweal, Qld
Central West	CM	CM	MgC1b	WW *et al.* (1999)	AJ130986	Camooweal, Qld
Central West	CM	CM	MgC2a	WW *et al.* (1999)	AJ130987	Camooweal, Qld
Central West	CM	CM	MgC2b	WW *et al.* (1999)	AJ130987	Camooweal, Qld
Central West	CM	CM	MgC2c	WW *et al.* (1999)	AJ130987	Camooweal, Qld
Central West	NAR	E-24	E01	Ancient	ACAD9467	E-24, WA
Central West	NAR	E-24	E02	Ancient	ACAD9469	E-24, WA
Central West	NAR	E-24	E03	Ancient	ACAD9471	E-24, WA
Central West	NAR	E-24	E04	Ancient	ACAD9473	E-24, WA
Central West	NAR	E-24	E05	Ancient	ACAD9475	E-24, WA
Central West	NAR	E-24	E06	Ancient	ACAD9476	E-24, WA
Central West	NAR	E-24	E07	Ancient	ACAD9549	E-24, WA
Central West	NAR	E-24	E08	Ancient	ACAD9550	E-24, WA
Central West	NAR	E-24	E09	Ancient	ACAD9551	E-24, WA
Central West	NAR	E-24	E10	Ancient	ACAD9552	E-24, WA
Central West	NAR	E-24	E11	Ancient	ACAD9554	E-24, WA
Central West	NAR	J-3	J01	Ancient	ACAD9290	J-3, WA
Central West	NAR	J-3	J02	Ancient	ACAD9292	J-3, WA
Central West	NAR	J-3	J03	Ancient	ACAD9294	J-3, WA
Central West	NAR	J-3	J04	Ancient	ACAD9555	J-3, WA
Central West	NAR	J-3	J05	Ancient	ACAD9556	J-3, WA
Central West	NAR	J-3	J06	Ancient	ACAD9557	J-3, WA
Kimberley	CS	CS	MgCS1	WW *et al.* (1999)	AJ130990	Cave Springs, WA
Kimberley	TC	TC	01TC	Modern		Tunnel Creek, WA
Kimberley	TC	TC	02TC	Modern		Tunnel Creek, WA
Kimberley	TC	TC	03TC	Modern		Tunnel Creek, WA
Kimberley	TC	TC	04TC	Modern		Tunnel Creek, WA
Kimberley	TC	TC	05TC	Modern		Tunnel Creek, WA
Kimberley	TC	TC	07TC	Modern		Tunnel Creek, WA
Kimberley	TC	TC	08TC	Modern		Tunnel Creek, WA
Kimberley	TC	TC	09TC	Modern		Tunnel Creek, WA
Kimberley	TC	TC	10TC	Modern		Tunnel Creek, WA
Kimberley	TC	TC	11TC	Modern		Tunnel Creek, WA
Kimberley	TC	TC	12TC	Modern		Tunnel Creek, WA
Kimberley	TC	TC	13TC	Modern		Tunnel Creek, WA
Kimberley	TC	TC	14TC	Modern		Tunnel Creek, WA
Kimberley	TC	TC	15TC	Modern		Tunnel Creek, WA
Kimberley	TC	TC	16TC	Modern		Tunnel Creek, WA
Kimberley	TC	TC	MgTC1	WW *et al.* (1999)	AJ130989	Tunnel Creek, WA
Pilbara	Pilbara	BB	BB11	Modern		Pilbara, WA
Pilbara	Pilbara	BB	BB12	Modern		Pilbara, WA
Pilbara	Pilbara	BB	BB13	Modern		Pilbara, WA
Pilbara	Pilbara	BB	BB14	Modern		Pilbara, WA
Pilbara	Pilbara	BB	BB15	Modern		Pilbara, WA
Pilbara	Pilbara	BB	BB16	Modern		Pilbara, WA
Pilbara	Pilbara	BB	BB17	Modern		Pilbara, WA

REGION	Colonies	Cave	Specimen	Sample details	Accession number	Location
Pilbara	Pilbara	BB	BB60	Modern		Pilbara, WA
Pilbara	Pilbara	BB	BB61	Modern		Pilbara, WA
Pilbara	Pilbara	BB	BB62	Modern		Pilbara, WA
Pilbara	Pilbara	BB	BB63	Modern		Pilbara, WA
Pilbara	Pilbara	BB	BB64	Modern		Pilbara, WA
Pilbara	Pilbara	BB	BB65	Modern		Pilbara, WA
Pilbara	Pilbara	BB	BB66	Modern		Pilbara, WA
Pilbara	Pilbara	BB	BB67	Modern		Pilbara, WA
Pilbara	Pilbara	BB	BB70	Modern		Pilbara, WA
Pilbara	Pilbara	BB	BB71	Modern		Pilbara, WA
Pilbara	Pilbara	BB	BB72	Modern		Pilbara, WA
Pilbara	Pilbara	BB	BB73	Modern		Pilbara, WA
Pilbara	Pilbara	BB	BB74	Modern		Pilbara, WA
Pilbara	Pilbara	BB	BB78	Modern		Pilbara, WA
Pilbara	Pilbara	BB	BB79	Modern		Pilbara, WA
Pilbara	Pilbara	CL	G01CL	Modern		Pilbara, WA
Pilbara	Pilbara	CL	G02CL	Modern		Pilbara, WA
Pilbara	Pilbara	CL	G03CL	Modern		Pilbara, WA
Pilbara	Pilbara	CL	G04CL	Modern		Pilbara, WA
Pilbara	Pilbara	CL	G05CL	Modern		Pilbara, WA
Pilbara	Pilbara	CL	G06CL	Modern		Pilbara, WA
Pilbara	Pilbara	CL	G07CL	Modern		Pilbara, WA
Pilbara	Pilbara	CL	G08CL	Modern		Pilbara, WA
Pilbara	Pilbara	CL	G09CL	Modern		Pilbara, WA
Pilbara	Pilbara	CL	G10CL	Modern		Pilbara, WA
Pilbara	Pilbara	CL	G11CL	Modern		Pilbara, WA
Pilbara	Pilbara	CL	G12CL	Modern		Pilbara, WA
Pilbara	Pilbara	CL	G13CL	Modern		Pilbara, WA
Pilbara	Pilbara	CL	G14CL	Modern		Pilbara, WA
Pilbara	Pilbara	CL	G15CL	Modern		Pilbara, WA
Pilbara	Pilbara	CL	G16CL	Modern		Pilbara, WA
Pilbara	Pilbara	CL	G17CL	Modern		Pilbara, WA
Pilbara	Pilbara	CL	G18CL	Modern		Pilbara, WA
Pilbara	Pilbara	CL	G19CL	Modern		Pilbara, WA
Pilbara	Pilbara	CL	G20CL	Modern		Pilbara, WA
Pilbara	Pilbara	CL	G21CL	Modern		Pilbara, WA
Pilbara	Pilbara	CL	G22CL	Modern		Pilbara, WA
Pilbara	Pilbara	CO	M48754	WAM	M48754	Pilbara, WA
Pilbara	Pilbara	CO	M48755	WAM	M48755	Pilbara, WA
Pilbara	Pilbara	CO	M49657	WAM	M49657	Pilbara, WA
Pilbara	Pilbara	CO	PKACOO	Modern		Pilbara, WA
Pilbara	Pilbara	EC	EC01EC	Modern		Pilbara, WA
Pilbara	Pilbara	KA	M47791	WAM	M47791	Pilbara, WA
Pilbara	Pilbara	KO	K75KO	Modern		Pilbara, WA
Pilbara	Pilbara	KQ	G1KQ	Modern		Pilbara, WA
Pilbara	Pilbara	KQ	G23KQ	Modern		Pilbara, WA
Pilbara	Pilbara	KQ	G24KQ	Modern		Pilbara, WA
Pilbara	Pilbara	KQ	G25KQ	Modern		Pilbara, WA
Pilbara	Pilbara	KQ	G26KQ	Modern		Pilbara, WA
Pilbara	Pilbara	KQ	G27KQ	Modern		Pilbara, WA
Pilbara	Pilbara	KQ	G28KQ	Modern		Pilbara, WA
Pilbara	Pilbara	KQ	G29KQ	Modern		Pilbara, WA

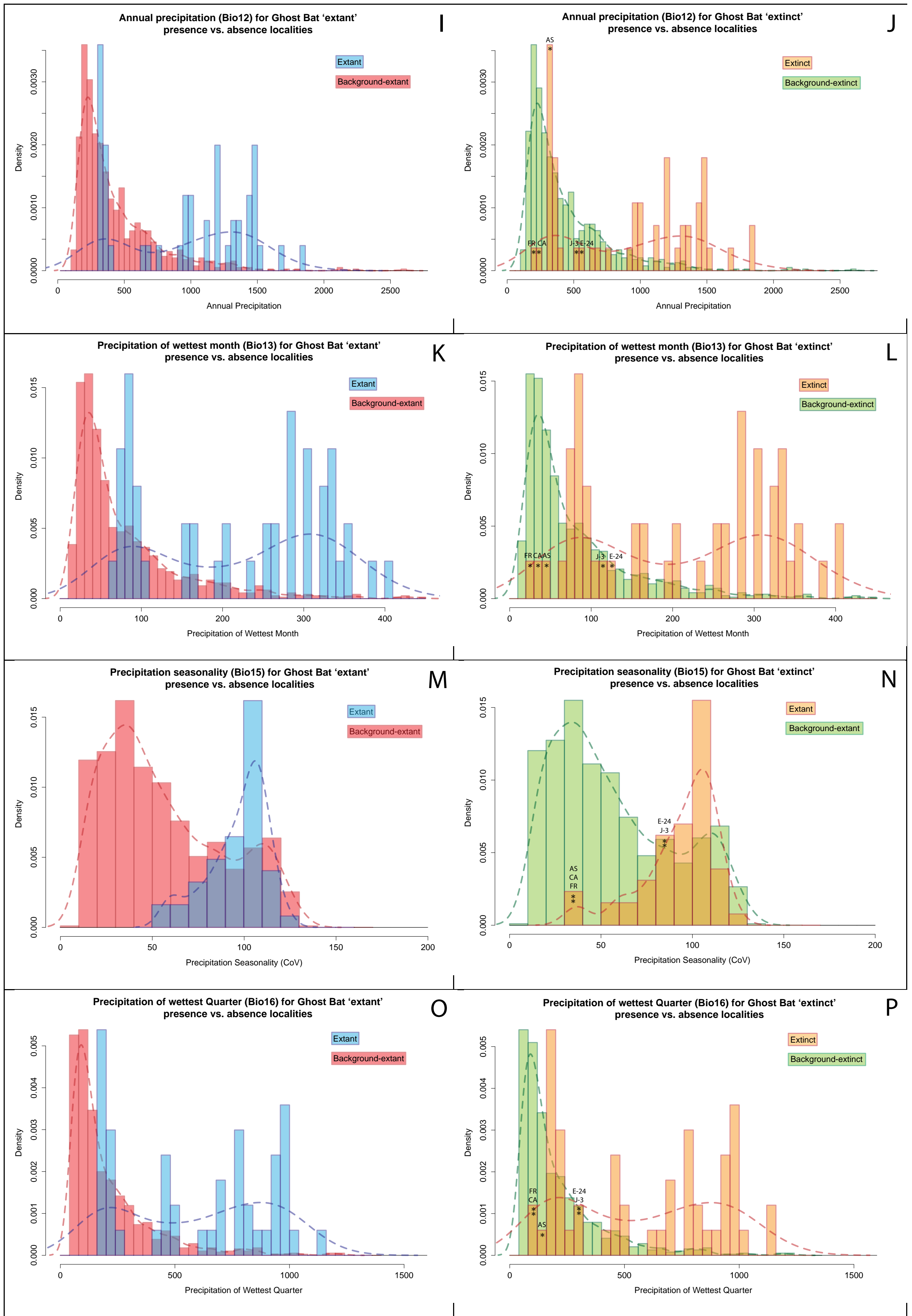
REGION	Colonies	Cave	Specimen	Sample details	Accession number	Location
Pilbara	Pilbara	KQ	G2KQ	Modern		Pilbara, WA
Pilbara	Pilbara	KQ	G30KQ	Modern		Pilbara, WA
Pilbara	Pilbara	KQ	G32KQ	Modern		Pilbara, WA
Pilbara	Pilbara	KQ	KQ02	Modern		Pilbara, WA
Pilbara	Pilbara	KQ	KQ03	Modern		Pilbara, WA
Pilbara	Pilbara	KQ	KQ04	Modern		Pilbara, WA
Pilbara	Pilbara	KQ	KQ05	Modern		Pilbara, WA
Pilbara	Pilbara	KQ	KQ06	Modern		Pilbara, WA
Pilbara	Pilbara	KQ	KQ07	Modern		Pilbara, WA
Pilbara	Pilbara	KQ	KQ08	Modern		Pilbara, WA
Pilbara	Pilbara	KQ	KQ09	Modern		Pilbara, WA
Pilbara	Pilbara	KQ	KQ10	Modern		Pilbara, WA
Pilbara	Pilbara	KQ	KQ11	Modern		Pilbara, WA
Pilbara	Pilbara	KQ	KQ12	Modern		Pilbara, WA
Pilbara	Pilbara	KQ	KQ13	Modern		Pilbara, WA
Pilbara	Pilbara	KQ	KQ14	Modern		Pilbara, WA
Pilbara	Pilbara	KQ	KQ15	Modern		Pilbara, WA
Pilbara	Pilbara	KQ	KQ20	Modern		Pilbara, WA
Pilbara	Pilbara	KQ	KQ21	Modern		Pilbara, WA
Pilbara	Pilbara	KQ	KQ22	Modern		Pilbara, WA
Pilbara	Pilbara	KQ	KQ24	Modern		Pilbara, WA
Pilbara	Pilbara	KQ	KQ25	Modern		Pilbara, WA
Pilbara	Pilbara	LR	LR01	Modern		Pilbara, WA
Pilbara	Pilbara	LR	LR02	Modern		Pilbara, WA
Pilbara	Pilbara	LR	LR03	Modern		Pilbara, WA
Pilbara	Pilbara	LR	LR04	Modern		Pilbara, WA
Pilbara	Pilbara	LR	LR05	Modern		Pilbara, WA
Pilbara	Pilbara	LR	LR06	Modern		Pilbara, WA
Pilbara	Pilbara	LR	LR07	Modern		Pilbara, WA
Pilbara	Pilbara	LR	LR08	Modern		Pilbara, WA
Pilbara	Pilbara	LR	LR09	Modern		Pilbara, WA
Pilbara	Pilbara	LR	LR10	Modern		Pilbara, WA
Pilbara	Pilbara	LR	LR11	Modern		Pilbara, WA
Pilbara	Pilbara	LR	LR12	Modern		Pilbara, WA
Pilbara	Pilbara	LR	LR13	Modern		Pilbara, WA
Pilbara	Pilbara	LR	LR14	Modern		Pilbara, WA
Pilbara	Pilbara	LR	LR15	Modern		Pilbara, WA
Pilbara	Pilbara	NU	NU01	Modern		Pilbara, WA
Pilbara	Pilbara	NU	NU02	Modern		Pilbara, WA
Pilbara	Pilbara	-	MgP1	WW *et al.* (1999)	AJ130998	Pilbara, WA
Pilbara	Pilbara	RR	AA14RR	Modern		Pilbara, WA
Pilbara	Pilbara	RR	BAKEN2	Modern		Pilbara, WA
Pilbara	Pilbara	RR	RHDA11	Modern		Pilbara, WA
Pilbara	Pilbara	RR	RHDA12	Modern		Pilbara, WA
Pilbara	Pilbara	RR	RHDA13	Modern		Pilbara, WA
Pilbara	Pilbara	RR	RHDA14	Modern		Pilbara, WA
Pilbara	Pilbara	RR	RHDA15	Modern		Pilbara, WA
Pilbara	Pilbara	RR	RHDA16	Modern		Pilbara, WA
Pilbara	Pilbara	RR	RHDA17	Modern		Pilbara, WA
Pilbara	Pilbara	RR	RHDA18	Modern		Pilbara, WA
Pilbara	Pilbara	RR	RHDA31	Modern		Pilbara, WA
Pilbara	Pilbara	RR	RHDA32	Modern		Pilbara, WA

REGION	*Colonies*	*Cave*	*Specimen*	*Sample details*	*Accession number*	*Location*
Pilbara	Pilbara	RR	RHDA33	Modern		Pilbara, WA
Pilbara	Pilbara	RR	RHDA34	Modern		Pilbara, WA
Pilbara	Pilbara	SG	KA_SG	Modern		Pilbara, WA
Pilbara	Pilbara	SG	KE_SG	Modern		Pilbara, WA
Pilbara	Pilbara	SG	KI_SG	Modern		Pilbara, WA
Pilbara	Pilbara	SG	KU_SG	Modern		Pilbara, WA
Pilbara	Pilbara	SG	SG0_SG	Modern		Pilbara, WA
Pilbara	Pilbara	SG	SG4_SG	Modern		Pilbara, WA
Pilbara	Pilbara	SG	SG5	Modern		Pilbara, WA
Pilbara	Pilbara	SG	TAG1SG	Modern		Pilbara, WA
Pilbara	Pilbara	SG	TAG2SG	Modern		Pilbara, WA
Pilbara	Pilbara	SG	TAG3SG	Modern		Pilbara, WA
Pilbara	Pilbara	TA	M3415	WAM	M3415	Pilbara, WA
Pilbara	Pilbara	TX	TEX08	Modern		Pilbara, WA
Pilbara	Pilbara	TX	TEX09	Modern		Pilbara, WA
Pilbara	Pilbara	TX	TEX1	Modern		Pilbara, WA
Pilbara	Pilbara	TX	TEX10	Modern		Pilbara, WA
Pilbara	Pilbara	TX	TEX2	Modern		Pilbara, WA
Pilbara	Pilbara	TX	TEX3	Modern		Pilbara, WA
Pilbara	Pilbara	TX	TEX4	Modern		Pilbara, WA
Pilbara	Pilbara	TX	TEX6	Modern		Pilbara, WA
Pilbara	Pilbara	TX	TEX7	Modern		Pilbara, WA
Pilbara	Pilbara	TX	TEX8	Modern		Pilbara, WA
Pilbara	Pilbara	WA	WA	Modern		Pilbara, WA
Pilbara	Pilbara	WA	WANA14	Modern		Pilbara, WA
Pilbara	Pilbara	WA	WANB11	Modern		Pilbara, WA
Pilbara	Pilbara	WA	WANG11	Modern		Pilbara, WA

Sample details abbreviations: SAM – South Australian Museum; WAM – Western Australian Museum; CA – Central Australia; CA – Central Australia; CC – Cutta Cutta; CH – Cape Hillsborough; CM – Camooweal; CS – Cave Springs; CV – Claravale Station; FR – Flinders Ranges and Central Australia; FR – Flinders Ranges; KP – Kings Plains; KM – Kings Plains and Mitchell Palmer; MP – Mitchell Palmer; NAR – Northern Agricultural Region; NR – Nourlangie Rock; PC – Pine Creek; RH – Rockhampton; RC – Rockhampton and Cape Hillsborough; TC – Tunnel Creek; E-24 – Cave near Eneabba, WA; J-3 – Cave near Jurien, WA. WW *et al.* (1999) - (Worthington Wilmer *et al.* 1999)



Annual mean temperature (Bio1) for Ghost Bat 'extinct' presence vs. absence localities
A
Extant
Background-extant
Density
Annual Mean Temperature
Annual mean temperature (Bio1) for Ghost Bat 'extinct' presence vs. absence localities
B
Extinct
Background-extinct
Annual Mean Temperature
Isothermality (Bio3) for Ghost Bat 'extant' presence vs. absence localities
C
Isothermality (*100)
Isothermality (Bio3) for Ghost Bat 'extinct' presence vs. absence localities
D
Minimum temperature of coldest month (Bio6) for Ghost Bat 'extant' presence vs. absence localities
E
Min. Temperature of Coldest Month
Minimum temperature of coldest month (Bio6) for Ghost Bat 'extinct' presence vs. absence localities
F
Min Temperature of Coldest Month
Mean temperature of coldest Quarter (Bio11) for Ghost Bat 'extant' presence vs. absence localities
G
Mean Temperature of Coldest Quarter
Mean temperature of coldest Quarter (Bio11) for Ghost Bat 'extinct' presence vs. absence localities
H
Mean Temperature of Coldest Quarter


Annual precipitation (Bio12) for Ghost Bat 'extant' presence vs. absence localities
I
Extant
Background-extant
Density
Annual Precipitation
Annual precipitation (Bio12) for Ghost Bat 'extinct' presence vs. absence localities
J
Extinct
Background-extinct
Annual Precipitation
Precipitation of wettest month (Bio13) for Ghost Bat 'extant' presence vs. absence localities
K
Precipitation of Wettest Month
Precipitation of wettest month (Bio13) for Ghost Bat 'extinct' presence vs. absence localities
L
Precipitation seasonality (Bio15) for Ghost Bat 'extant' presence vs. absence localities
M
Precipitation Seasonality (CoV)
Precipitation seasonality (Bio15) for Ghost Bat 'extinct' presence vs. absence localities
N
Precipitation of wettest Quarter (Bio16) for Ghost Bat 'extant' presence vs. absence localities
O
Precipitation of Wettest Quarter
Precipitation of wettest Quarter (Bio16) for Ghost Bat 'extinct' presence vs. absence localities
P

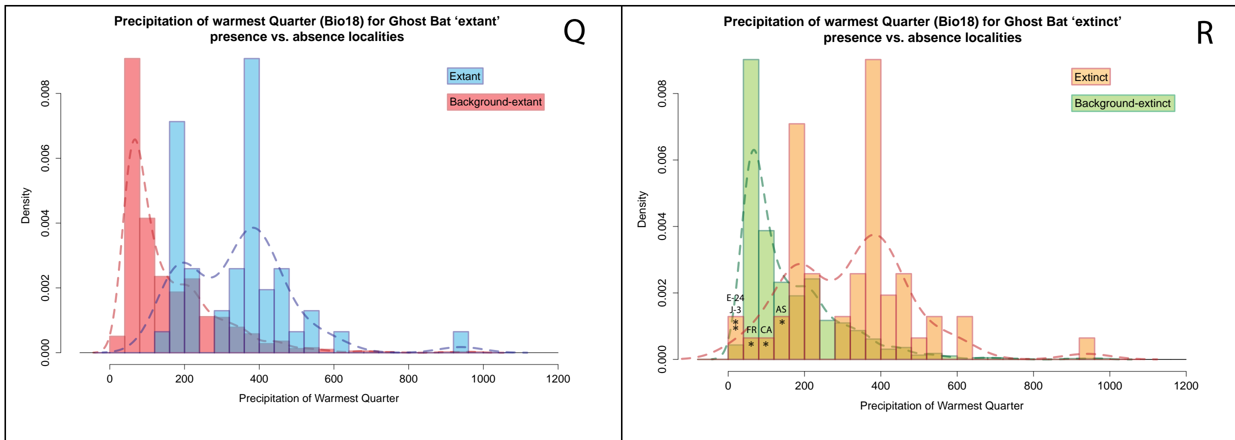


Figure S2. Nine bioclimatic variables identified as having a high impact on predicting Ghost Bat presence, showing variance between presence and pseudo-absence climate conditions for both extant and extinct datasets. A) annual mean temperature (Bio1) using extant dataset; B) annual mean temperature (Bio1) using extinct dataset; C) isothermality (Bio3) using extant dataset; D) isothermality (Bio3) using extinct dataset; E) minimum temperature of coldest month (Bio6) using extant dataset; F) minimum temperature of coldest month (Bio6) using extinct dataset; G) mean temperature of coldest quarter (Bio11) using extant dataset; H) mean temperature of coldest quarter (Bio11) using extinct dataset; I) annual precipitation (Bio12) using extant dataset; J) annual precipitation (Bio12) using extinct dataset; K) precipitation of wettest month (Bio13) using extant dataset; L) precipitation of wettest month (Bio13) using extinct dataset; M) precipitation seasonality (Bio15) using extant dataset; N) precipitation seasonality (Bio15) using extinct dataset; O) precipitation of wettest quarter (Bio16) using extant dataset; P) precipitation of wettest quarter (Bio16) using extinct dataset; Q) precipitation of warmest quarter (Bio18) using extant dataset; R) precipitation of warmest quarter (Bio18) using extinct dataset.

References

Armstrong KN (2000). Roost microclimates of the bat *Rhinonicteris aurantius* in a limestone cave in Geikie Gorge, Western Australia. *Australian Mammalogy* 22: 69-70.

Armstrong KN (2001). The roost habitat and distribution of the orange leaf-nosed bat, *Rhinonicteris aurantius*, in the Pilbara region of Western Australia. *Wildlife Research* 28: 95-104.

Armstrong KN (2003). *The bats that time forgot: the Orange Leaf-nosed Bat Rhinonicteris aurantius (Gray, 1845) (Microchiroptera: Hipposideridae) in the Pilbara region of Western Australia*. PhD, The University of Western Australia.

Armstrong KN and Anstee SD (2000). The Ghost bat in the Pilbara: 100 years on. *Australian Mammalogy* 22: 93-101.

Barrows TT, Stone JO, Fifield LK and Cresswell RG (2002). The timing of the last glacial maximum in Australia. *Quaternary Science Reviews* 21(1-3): 159-173.

Baudinette RV, Churchill SK, Christian KA, Nelson JE and Hudson PJ (2000). Energy, water balance and the roost microenvironment in three Australian cave-dwelling bats (Microchiroptera). *Journal of Comparative Physiology B: Biochemical, Systemic, and Environmental Physiology* 170(5): 439-446.

Bilney RJ, Cooke R and White JG (2010). Underestimated and severe: Small mammal decline from the forests of south-eastern Australia since European settlement, as revealed by a top-order predator. *Biological Conservation* 143(1): 52-59.

Blacket MJ, Adams M, Cooper SJB, Krajewski C and Westerman M (2001). Systematics and evolution of the Dasyurid marsupial genus *Sminthopsis*: I. The macroura species group. *Journal of Mammalian Evolution* 8(2): 149-170.

Braconnot P, Otto-Bliesner B, Harrison S, Joussaume S, Peterchmitt JY, Abe-Ouchi A, Crucifix M, Driesschaert E, Fichefet T, Hewitt CD, Kageyama M, Kitoh A, Laîné A, Loutre MF, Marti O, Merkel U, Ramstein G, Valdes P, Weber SL, Yu Y and Zhao Y (2007). Results of PMIP2 coupled simulations of the Mid-Holocene and last glacial maximum - Part 1: experiments and large-scale features. *Clim. Past* 3(2): 261-277.

Bridge PJ (1975). Paleo-distribution of *Macroderma gigas* in the southwest of Western Australia. *Helictite* 13: 35.

Burbidge AA, Johnson KA, Fuller PJ and Southgate RI (1988). Aboriginal knowledge of the mammals of the central deserts of Australia. *Australian Wildlife Research* 15: 9-39.

Burbidge AA, McKenzie NL, Brennan KEC, Woinarski JCZ, Dickman CR, Baynes A, Gordon G, Menkhorst PW and Robinson AC (2008). Conservation status and biogeography of Australia's terrestrial mammals. *Australian Journal of Zoology* 56(6): 411-422.

Butler WH (1961). Occurance of the Ghost bat, *Macroderma gigas*, in the Great Victoria Desert, W.A. *Western Australian Naturalist* 8(1): 42-43.

Byrne M, Yeates DK, Joseph L, Kearney M, Bowler J, Williams MAJ, Cooper S, Donnellan SC, Keogh JS, Leys R, Melville J, Murphy DJ, Porch N and Wyrwoll KH (2008). Birth of a biome: insights into the assembly and maintenance of the Australian arid zone biota. *Molecular Ecology* 17(20): 4398-4417.

Campos PF, Craig OE, Turner-Walker G, Peacock E, Willerslev E and Gilbert MTP (2011). DNA in ancient bone - Where is it located and how should we extract it? *Annals of Anatomy* 194(1): 7-16.

Canestrelli D, Aloise G, Cecchetti S and Nascetti G (2010). Birth of a hotspot of intraspecific genetic diversity: notes from the underground. *Molecular Ecology* 19: 5432-5451.

Chapple DG, Hoskin CJ, Chapple SNJ and Thompson MB (2011). Phylogeographic divergence in the widespread delicate skink (*Lampropholis delicata*) corresponds to dry habitat barriers in eastern Australia. *BMC Evolutionary Biology* 11.

Chapple DG, Keogh JS and Hutchinson MN (2004). Molecular phylogeography and systematics of the arid-zone members of the *Egernia whitii* (Lacertilia: Scincidae) species group. *Molecular Phylogenetics and Evolution* 33(3): 549-561.

Churchill S (2008). *Australian Bats*. Crows Nest, Allen and Unwin.

Churchill SK (1991). Distribution, abundance and roost selection of the Orange Horseshoe-bat, *Rhinonycteris aurantius*, a tropical cave-dweller. *Wildlife Research* 18: 343-353.

Churchill SK and Helman PM (1990). Distribution of the Ghost bat, *Macroderma gigas*, (Chiroptera: Megadermatidae) in central and South Australia. *Australian Mammalogy* 13: 149-156.

Cooper SJB, Adams M and Labrinidis A (2000). Phylogeography of the Australian dunnart *Sminthopsis crassicaudata* (Marsupialia: Dasyuridae). *Australian Journal of Zoology* 48(5): 461-473.

Dobson GE (1880). On some new or rare species of Chiroptera in the collection of the Gottingen Museum. *Proceedings of the Zoological Society of London* 48(3): 461-465.

Dodson J (2001). Holocene vegetation change in the Mediterranean-type climate regions of Australia. *The Holocene* 11(6): 673-680.

Douglas AM (1956). A recent Western Australia record of the Ghost Bat (*Macroderma gigas*). *The Western Australian Naturalist* 5(6): 140-141.

Douglas AM (1957). The Ghost Bat in the North-west. *The Western Australian Naturalist* 5(8): 232.

Douglas AM (1967). The natural history of the Ghost bat, *Macroderma gigas* (Microchiroptera, Megadermatidae), in Western Australia. *The Western Australian Naturalist* 10(6): 125-137.

Drummond A and Rambaut A (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* 7(1): 214.

Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A, Duran C, Field M, Heled J, Kearse M, Markowitz S, Moir R, Stones-Havas S, Sturrock S, Thierer T and Wilson AC (2011). Geneious v5.4. *Available from* http://www.geneious.com/.

Edwards DL (2007). Biogeography and speciation of a direct developing frog from the coastal arid zone of Western Australia. *Molecular Phylogenetics and Evolution* 45(2): 494-505.

Eldridge MDB, Potter S and Cooper SJB (2012). Biogeographic barriers in north-western Australia: an overview and standardisation of nomenclature. *Australian Journal of Zoology* 59(4): 270-272.

Elith J, Graham CH, Anderson RP, Dudík M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, Li J, Lohmann LG, Loiselle BA, Manion G, Moritz C, Nakamura M, Nakazawa Y, McC. M. Overton J, Townsend Peterson A, Phillips SJ, Richardson K, Scachetti-Pereira R, Schapire RE, Soberón J, Williams S, Wisz MS and Zimmermann NE (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29(2): 129-151.

EPICA, Barbante C, Barnes PRF, Barnola J-M, Bigler M, Castellano E, Cattani O, Chappellaz J, Dahl-Jensen D, Delmonte B, Dreyfus G, Durand G, Falourd S, Fischer H, Fluckiger J, Hansson ME, Huybrechts P, Jugie G, Johnsen SJ, Jouzel J, Kaufmann P, Kipfstuhl J, Lambert F, Lipenkov VY, Littot GC, Longinelli A, Lorrain R, Maggi V, Masson-Delmotte V, Miller H, Mulvaney R, Oerlemans J, Oerter H, Orombelli G, Parrenin F, Peel DA, Petit J-R, Raynaud D, Ritz C, Ruth U, Schwander J, Siegenthaler U, Souchez R, Stauffer B, Steffensen JP, Stenni B, Stocker TF, Tabacco IE, Udisti R, van de Wal RSW, van den Broeke M, Weiss J, Wilhelms F, Winther J-G, Wolff EW and Zucchelli M (2004). Eight glacial cycles from an Antarctic ice core. *Nature* 429(6992): 623-628.

Excoffier L and Lischer H (2010). Arlequin suite ver3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* 10: 564-567.

Faith JT and O'Connell JF (2011). Revisiting the late Pleistocene mammal extinction record at Tight Entrance Cave, southwestern Australia. *Quaternary Research* 76(3): 397-400.

Finlayson H (1961). On central Australian mammals Part IV - The distribution and status of central Australian species. *Records of the South Australian Museum* 14: 141-191.

Ford J (1987). Minor isolates and minor geographical barriers in avian speciation in continental Australia. *Emu* 87: 90-102.

Fujioka T, Chappell J, Fifield LK and Rhodes EJ (2009). Australian desert dune fields initiated with Pliocene–Pleistocene global climatic shift. *Geology* 37(1): 51-54.

Fujioka T, Chappell J, Honda M, Yatsevich I, Fifield K and Fabel D (2005). Global cooling initiated stony deserts in central Australia 2–4 Ma, dated by cosmogenic 21Ne-10Be. *Geology* 33(12): 993-996.

Garrick R, Sands C, Rowell D, Tait N, Greenslade P and Sunnucks P (2004). Phylogeography recapitulates topography: very fine-scale local endemism of a saproxylic 'giant' springtail at Tallaganda in the Great Dividing Range of south-east Australia. *Molecular Ecology* 13(11): 3329-3344.

Hand S and York A (1990). Morphological variation in the dentition and skull of the Australian Ghost Bat, *Macroderma gigas* (Microchiroptera, Megadermatidae). *Australian Journal of Zoology* 38(3): 263-286.

Hand SJ (1996). New Miocene and Pliocene megadermatids (Mammalia, Microchiroptera) from Australia, with comments on broader aspects of megadermatid evolution. *Geobios* 29(3): 365-377.

Hirzel AH, Hausser J, Chessel D and Perrin N (2002). Ecological-niche factor analysis: How to compute habitat-suitability maps without absence data? *Ecology* 83(7): 2027-2036.

Ho SY, Larson G, Edwards CJ, Heupink TH, Lakin KE, Holland PW and Shapiro B (2008). Correlating Bayesian date estimates with climatic events and domestication using a bovine case study. *Biology Letters* 4(4): 370-374.

Hudson WS and Wilson DE (1986). *Macroderma gigas*. *Mammalian Species*(260): 1-4.

Irwin DE (2002). Phylogeographic breaks without geographic barriers to gene flow. *Evolution* 56(2): 2383-2394.

Jones K. (2008). "Husbandry Manual for Ghost bat *Macroderma gigas* Mammalia: Megadermatidae." Western Sydney Institute of TAFE Retrieved 09 December 2011, 2011, from http://www.australianzookeeping.org/Husbandry_Manuals.htm - Order_Chiroptera_(Bats)_.

Keane TM, Creevey CJ, Pentony MM, Naughton TJ and McInerney JO (2006). Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evolutionary Biology* 6(29): 1-17.

Kearney M and Blacket MJ (2008). The evolution of sexual and parthenogenetic *Warramaba*: a window onto Plio-Pleistocene diversification processes in an arid biome. *Molecular Ecology* 17(24): 5257-5275.

Kranendonk MJV, Hickman AH, Smithies RH, Nelson DR and Pike G (2002). Geology and tectonic evolution of the Archean north Pilbara terrain, Pilbara Craton, Western Australia. *Economic Geology* 97(4): 695-732.

Kulzer E, Nelson JE, McKean JL and Mohres FP (1970). Temperature regulation in Australian bats (Microchiroptera). *Zeitschrift Fur Vergleichende Physiologie* 69(4): 426-&.

Lee JY, Edwards SV and Webster M (2008). Divergence across Australia's Carpentarian barrier: Statistical phylogeography of the Red-Backed Fairy Wren (*Malurus melanocephalus*). *Evolution* 62(12): 3117-3134.

Leitner P and Nelson JE (1967). Body temperature, oxygen consumption and heart rate in the Australian false vampire bat, *Macroderma gigas*. *Comparative Biochemistry and Physiology* 21: 75-81.

Letnic M and Dickman CR (2006). Boom means bust: Interactions between the El Nino/Southern Oscillation (ENSO), rainfall and the processes threatening mammal species in arid Australia. *Biodiversity and Conservation* 15: 3847-3880.

Lowe A, Harris S and Ashton P (2004). *Ecological Genetics: Design, Analysis, and Application*. Oxford, Blackwell Science Ltd.

Maclean IMD and Wilson RJ (2011). Recent ecological responses to climate change support predictions of high extinction risk. *Proceedings of the National Academy of Sciences of the United States of America* 108(30): 12337-12342.

McKenzie NL and Hall L (2008) *Macroderma gigas*. IUCN 2011. IUCN Red List of Threatened Species.

Menkhorst P and Knight F (2004). *A field guide to the mammals of Australia*. South Melbourne ET - 2, Oxford University Press.

Molnar RE, Hall LS and Mahoney JH (1984). New fossil localities for *Macroderma* Miller, 1906 (Chiroptera: Megadermatidae) in New South Wales and its past and present distribution in Australia. *Australian Mammalogy* 7: 63-73.

Oliver P, Hugall A, Adams M, Cooper SJB and Hutchinson M (2007). Genetic elucidation of cryptic and ancient diversity in a group of Australian diplodactyline geckos: The *Diplodactylus vittatus* complex. *Molecular Phylogenetics and Evolution* 44(1): 77-88.

Otto-Bliesner BL, Marshall SJ, Overpeck JT, Miller GH, Hu A and members CLIP (2006). Simulating arctic climate warmth and icefield retreat in the last interglaciation. *Science* 311(5768): 1751-1753.

Pearson RG and Dawson TP (2003). Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology and Biogeography* 12(5): 361-371.

Pepper M, Fujita MK, Moritz C and Keogh JS (2011). Palaeoclimate change drove diversification among isolated mountain refugia in the Australian arid zone. *Molecular Ecology* 20(7): 1529-1545.

Perez KE, Ponder WF, Colgan DJ, Clark SA and Lydeard C (2005). Molecular phylogeny and biogeography of spring-associated hydrobiid snails of the Great Artesian Basin, Australia. *Molecular Phylogenetics and Evolution* 34(3): 545-556.

Phillips SJ, Anderson RP and Schapire RE (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190(3-4): 231-259.

Potter S, Eldridge MDB, Taggart DA and Cooper SJB (2012). Multiple biogeographical barriers identified across the monsoon tropics of northern Australia: phylogeographic analysis of the brachyotis group of rock-wallabies. *Molecular Ecology* 21(9): 2254-2269.

Short J and Smith A (1994). Mammal decline and recovery in Australia. *Journal of Mammalogy* 75(2): 288-297.

Smith W, Scott IAW and Keogh JS (2007). Molecular phylogeography of Rosenberg's goanna (Reptilia: Varanidae: *Varanus rosenbergi*) and its conservation status in New South Wales. *Systematics and Biodiversity* 5(04): 361-369.

Smithson PA (1991). Inter-relationships between cave and outside air temperature. *Theoretical and Applied Climatology* 44: 65-73.

Teeling EC, Springer MS, Madsen O, Bates P, O'Brien SJ and Murphy WJ (2005). A molecular phylogeny for bats illuminates biogeography and the fossil record. *Science* 307(5709): 580-584.

Tidemann CR, Priddel DM, Nelson JE and Pettigrew JD (1985). Foraging behavior of the Australian Ghost Bat, *Macroderma gigas* (Microchiroptera, Megadermatidae). *Australian Journal of Zoology* 33(5): 705-713.

Webb LJ and Tracey JG (1981). Australian rainforests: Patterns and change. Ecological Biogeography of Australia. Keast A. The Hague, Dr. W. Junk. 1: 605-694.

Worthington Wilmer J, Hall L, Barratt E and Moritz C (1999). Genetic structure and male-mediated gene flow in the Ghost Bat (*Macroderma gigas*). *Evolution* 53(5): 1582-1591.

Worthington Wilmer J, Moritz C, Hall L and Toop J (1994). Extreme population structuring in the threatened Ghost Bat, *Macroderma gigas* - Evidence from mitochondrial DNA. *Proceedings of the Royal Society of London Series B-Biological Sciences* 257(1349): 193-198.

Chapter 5: The Polynesian chicken

Don’t count your chickens: Resolving migration patterns of humans and commensals across the Pacific

STATEMENT OF AUTHORSHIP

Don't count your chickens: Resolving migration patterns of humans and commensals across the Pacific.

Text in manuscript

THOMSON, V.A. (Candidate)

Performed ancient DNA extractions on majority of samples, performed majority of PCR amplifications and downstream processing and analysis, performed the phylogeographic analysis, interpreted the data, created figures, and wrote manuscript.

I hereby certify that the statement of contribution is accurate.

Signed. ....*Date* ..10/5/2012.........

AUSTIN, J.

Initiated research and acted as corresponding author.

I hereby certify that the statement of contribution is accurate and I give permission for the inclusion of the manuscript in the thesis.

Signed. ....*Date* ..10/5/2012.........

HUNT, T.

Provided samples and evaluation of the manuscript.

I hereby certify that the statement of contribution is accurate and I give permission for the inclusion of the manuscript in the thesis.

Signed. ....*Date* ...... 4 Oct 2011...

BURNEY, D.

Provided samples and evaluation of the manuscript.

I hereby certify that the statement of contribution is accurate and I give permission for the inclusion of the manuscript in the thesis.

Signed. ……………………………………………………………Date…10August 2011…

WOOD, J.

Performed extractions and PCR's on some of the samples and evaluated the manuscript.

I hereby certify that the statement of contribution is accurate and I give permission for the inclusion of the manuscript in the thesis.

Signed. ……………………………………………………………Date …3 August 2011…

GONGORA, J.

Contributed to conceptualization of the work, designed research and evaluated the manuscript.

I hereby certify that the statement of contribution is accurate and I give permission for the inclusion of the manuscript in the thesis.

Signed. ……………………………………………………………Date …03-08-2011…….

FLINK, L. G.

Performed independent replication of a critical sample.

I hereby certify that the statement of contribution is accurate and I give permission for the inclusion of the manuscript in the thesis.

Signed. ....*Date* ...17/07/11...........

LARSON, G.

Performed independent replication of a critical sample and evaluated the manuscript.

I hereby certify that the statement of contribution is accurate and I give permission for the inclusion of the manuscript in the thesis.

Signed. ……………………………………………………*Date*…*Aug. 3, 2011*……

DENHAM, T.

Provided comments on, and evaluation of, the manuscript.

I hereby certify that the statement of contribution is accurate and I give permission for the inclusion of the manuscript in the thesis.

Signed. … .*Date* ..20/09/2011..

COOPER, A.

Supervised development of work, helped in data interpretation and manuscript writing, and acted as corresponding author.

I hereby certify that the statement of contribution is accurate and I give permission for the inclusion of the manuscript in the thesis.

Signed. ……………………*Date* ..10/5/2012....

Don’t count your chickens: Resolving migration patterns of humans and commensals across the Pacific.

Authors:

Vicki A. Thomson[a], Jeremy J. Austin[a,b], Terry Hunt[c], David Burney[d], Tim Denham[e], Jamie R. Wood[f], Jaime Gongora[g], Linus Girdland Flink[h,i], Greger Larson[h], Alan Cooper[a].

[a]Australian Centre for Ancient DNA, University of Adelaide, Adelaide SA 5005; [b]Sciences Department, Museum Victoria, Carlton Gardens, Melbourne, VIC 3001; [c]Department of Anthropology, University of Hawai’i-Manoa, 2424 Maile Way, Saunders Hall 345, Honolulu HI 96822, USA; [d]National Tropical Botanical Garden, 3530 Papalina Road, Kalaheo, HI 96741 USA; [e]School of Geography and Environmental Science, Monash University, Clayton VIC 3800; [f]Landcare Research, PO Box 40, Lincoln 7640, New Zealand; [g]Faculty of Veterinary Science, University of Sydney, Sydney, NSW 2006; [h]Department of Archaeology, Durham University, South Road, Durham DH1 3LE, UK; [i]School of Biological and Biomedical Sciences, Durham University, South Road, Durham, DH1 3LE, UK.

Keywords: ancient DNA, Lapita, Pacific colonization, phylogeography, Polynesian chicken.

Classification (Major): Biological Sciences

Classification (Minor): Anthropology

The authors declare no conflict of interest.

To whom correspondence may be addressed:

Alan Cooper, Australian Centre for Ancient DNA, University of Adelaide, Adelaide SA 5005, 61 8 8313 5950, alan.cooper@adelaide.edu.au.

Jeremy Austin, Australian Centre for Ancient DNA, University of Adelaide, Adelaide SA 5005, 61 8 8313 4557, jeremy.austin@adelaide.edu.au.

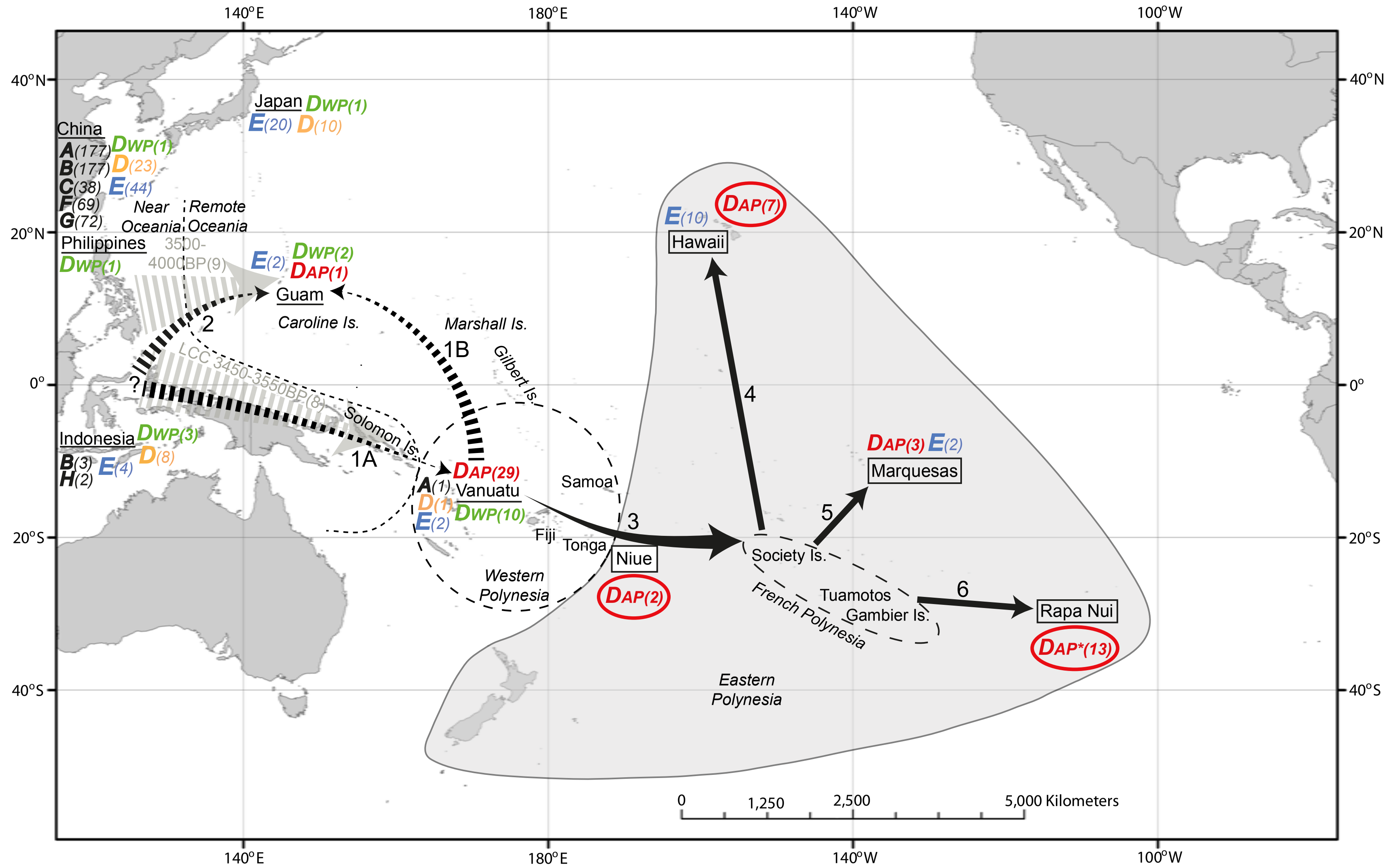
Abstract

The peopling of Remote Oceania was characterised by distinctive cultural traits and the dispersal of commensal animals, which proceeded east from Asia across the vast expanse of the Pacific Ocean. Despite more than a century of research, there is little conclusive evidence about the route and timing of this migration, or whether it included direct contact with South America. We use mitochondrial DNA from fifteen modern and twenty-two ancient chicken specimens, sampled across Polynesia, to investigate this issue. In contrast to previous reports, we find two clusters of Island South East Asian lineages widely distributed across the Pacific. One sub-clade is found in all archaeological chicken samples and is suggested to represent the original Polynesian domestic strain. It persists on several Polynesian islands at high frequencies. Neither Pacific group has been found in South American chickens, arguing against direct contact of Polynesian commensals with South America.

Introduction

The colonization of the remote Pacific was one of the last great prehistoric migrations, and the timing and route have been debated for centuries. A variety of hypotheses based on archaeological, linguistic and biological lines of evidence have been developed to explain this process, which culminated less than 800 years ago (ya) in the geographically dispersed islands of Hawai`i, New Zealand, and Rapa Nui (Easter Island) (Wilmshurst *et al.* 2011). There is considerable debate about whether these lines of evidence are congruent; clearly indicate the migration of Austronesian language speakers south and east from Taiwan; and/or highlight the degree to which their genes and Neolithic cultures replaced those of pre-existing populations in Island Southeast Asia (ISEA) (Shutler and Marck 1975; Diamond 2001; Diamond and Bellwood 2003; Specht 2007; Bulbeck 2008; Lemey *et al.* 2009; Donohue and Denham 2010). The Malayo-Polynesian sub-group (MP) of the Austronesian language family is widely accepted as originating in Taiwan, and replaced most pre-existing languages within ISEA (Blust 2000; Ross 2005). However, the genetics of modern human populations in the area show only a limited Taiwanese inheritance (Capelli *et al.* 2001; Hill *et al.* 2007; Lansing *et al.* 2007; Abdulla *et al.* 2009; Soares *et al.* 2011), and the fine scale spatio-temporal aspects of Neolithic material cultural traits are poorly understood (Donohue and Denham 2010), including commensal animals (Dobney *et al.* 2008).

In Remote Oceania (Figure 1), the eastward migration of Austronesian language speakers is commonly associated with the advent of the Lapita Cultural Complex (LCC) around 2800-3450 ya, which is characterised by distinctive dentate-stamped pottery (Green 1991; Kirch 1997; Spriggs 1997; Summerhayes 2000). Although the cultural and genetic patterns of Austronesian language speakers in ISEA and Near Oceania are complex, the first colonizers of Remote Oceania (or 'the Pacific') – initially the island groups comprising Vanuatu, New Caledonia, Fiji, Tonga and Samoa – were Lapita-bearing (Bedford *et al.* 2006; Clark and Anderson 2009). These colonists carried a material culture that combined elements from ISEA and Melanesia. For example, together with pottery, they brought commensal animals from ISEA such as the pig (*Sus scrofa*) (Allen *et al.* 2001; Larson *et al.* 2005; Larson *et al.* 2007), dog (*Canis familiaris*) (Savolainen *et al.* 2004; Oskarsson *et al.* 2011), Pacific rat (*Rattus exulans*) (Matisoo-Smith *et al.* 1998; Matisoo-Smith 2002; Matisoo-Smith and Robins 2004) and chicken (*Gallus gallus*) (Storey *et al.* 2007; Gongora *et al.* 2008). Similarly, they carried plants originally domesticated in the New Guinea region, including bananas (*Musa spp.*) (Perrier *et al.* 2011), breadfruit (*Artocarpus altilis*), sugarcane (*Saccharum* spp.) and some yams (*Dioscorea* spp.) (Lebot 1999). The descendants of these colonizers are presumed to have eventually settled the remainder of the Pacific.



140°E
180°E
140°W
100°W
40°N
20°N
0°
20°S
40°S
Japan DWP(1)
E(20) D(10)
China
A(177) DWP(1)
B(177) D(23)
C(38) E(44)
F(69)
G(72)
Near
Oceania
Remote
Oceania
Philippines
DWP(1)
3500-
4000BP(9)
E(2) DWP(2)
DAP(1)
Guam
Caroline Is.
Marshall Is.
2
1B
Gilbert Is.
LCC 3450-3550BP(8)
?
Solomon Is.
1A
Indonesia DWP(3)
B(3) E(4) D(8)
H(2)
DAP(29)
A(1) Vanuatu
D(1) DWP(10)
E(2)
Samoa
Fiji
Tonga
Niue
DAP(2)
Western
Polynesia
3
E(10) DAP(7)
Hawaii
4
DAP(3) E(2)
Marquesas
5
Society Is.
Tuamotos
Gambier Is.
French Polynesia
6
Rapa Nui
DAP*(13)
Eastern
Polynesia
0
1,250
2,500
5,000 Kilometers

Figure 1. **Map showing sites sampled in this study (boxed) and other locations used for reference sequences (Genbank sequence locations are underlined; Liu *et al.* 2006; Oka *et al.* 2007; Dancause *et al.* 2011). Geographic regions mentioned in text are shown (dashed line indicates demarcation between Near and Remote Oceania; Western Polynesia is defined by dashed circle and Eastern Polynesia is indicated by grey shaded triangle. Haplotypes are noted for each location: circled haplotypes denote ancient chicken samples with the number of samples in brackets, and the asterisk signifies the three Storey *et al.* (2007) samples that were re-examined. Colors refer to haplotype/haplogroup: D_{AP} haplogroup in red; D_{WP} haplogroup in green; other D haplotypes in yellow; E haplogroup in blue; all other haplotypes are in black (for reference only). All haplogroups are defined as per Liu *et al.* (2006). Arrows are representative of proposed migration routes: grey arrows represent movements inferred from archaeological data (Blust 2000; Specht 2007); black arrows represent transportation of commensal chickens: arrows that are dashed indicate hypothesised movements, while solid arrows are based on similarity of haplotypes. Numbers on arrows as discussed in text: 1A, introduction of chickens into Western Polynesia; 1B, possible late back-migration from Western Polynesia to Micronesia with a subset of diversity subsequently lost in Vanuatu; 2, alternate hypothesis to explain non-overlapping chicken diversity in Micronesia; 3, introduction of chickens from Western Polynesia into Eastern Polynesia; 4-6, spread of chickens within Eastern Polynesia**

The development of biomolecular markers and techniques have greatly improved our capacity to track the separate genetic histories of the biological elements associated with the expansion of Neolithic traits in ISEA, the development and spread of the LCC in Near Oceania, and the colonization of the Pacific. Molecular techniques can test elements of different models for these processes, whether 'Entangled Bank' (Terrell 1988), 'Slow Boat' (Oppenheimer and Richards 2001), 'Express Train' (Oppenheimer 2004) or 'Triple I' (Intrusion/ Integration/ Innovation) (Green 1991). Furthermore, the temporal and geographic analysis of ancient DNA (aDNA) from prehistoric remains of commensal animals and plants provides an even more powerful approach (Cooper and Poinar 2000). However, the evidence from ancient commensals in the Pacific has so far been equivocal. Mitochondrial DNA (mtDNA) sequences from the Pacific rat have provided limited resolution of migration patterns, probably due to limited amounts of sequence information, restricted sampling, and the likelihood that phylogeographic signals have been obscured by accidental back migrations (as stowaways on prehistoric vessels with inter-island interactions) (Matisoo-Smith *et al.* 1998; Matisoo-Smith 2002; Matisoo-Smith and Robins 2004). Though the movement of dogs into Australia has been explored, few Pacific dogs have yet to be characterized (Savolainen *et al.* 2004; Oskarsson *et al.* 2011). Lastly, although modern and ancient pig samples (Larson *et al.* 2005; Larson *et al.* 2007) support two migration trajectories from the East Asian mainland into the Pacific, the lack of resolution within populations involved in these two pathways has limited the understanding of interactions within Remote Oceania. In contrast, the domestic chicken has high levels of mitochondrial genetic variability in the Asia-Pacific region, and appears to be an ideal proxy to study human migration patterns in Oceania. Substantial phylogeographic structure exists within the mtDNA of chickens worldwide (Fumihito *et al.* 1996; Oka *et al.* 2007; Kanginakudru *et al.* 2008; Silva *et al.* 2009), with nine highly divergent haplogroups (designated A-I) identified (Liu *et al.* 2006). The current distribution of these clades is thought to reflect the original natural diversity, overlaid with separate domestication events within Asia, and subsequent transportation and trading further afield by both historic (Johannessen 1982) and prehistoric (West and Zhou 1988) human cultures.

Using chickens to investigate Polynesian contact with South America

Prehistoric contact between Asia and the Americas has long been indirectly suggested by the presence of South American domesticates (the bottle gourd and sweet potato) in early Pacific sites (Whitaker and Carter 1954; O'Brien 1972; Hather and Kirch 1991; Clarke *et al.* 2006). Recently, Storey *et al.* (2007) claimed that ancient archaeological chicken bones provided direct support for prehistoric contact, after finding a mitochondrial haplogroup E sequence in a single, immediately pre-Columbian, chicken bone (CHLARA001) from El Arenal-1 in Chile. The same genetic signature was found in several Polynesian archaeological samples, including Rapa Nui in

the far east of the Pacific Ocean (Storey *et al.* 2007; Storey *et al.* 2010). However, this conclusion was challenged as chicken haplogroup E is distributed widely throughout Eurasia (and worldwide), and represents the most common haplotypes in Europe (71%): the Middle East (91%) and India (55%) (Liu *et al.* 2006) so is not phylogenetically informative (Gongora *et al.* 2008). For example, Chilean archaeological specimens around the time of European contact could simply represent the first chickens introduced by Portuguese colonizers around 1500 A.D. (Carter 1971), rather than by Polynesians (A.D. ~1200-1300; Wilmshurst *et al.* 2011). It is also important to note that E haplotypes are likely to be produced from contamination with modern domestic chicken DNA, and contaminated laboratory consumables are a well-known problem in aDNA research (Leonard *et al.* 2007). Without a clear genetic connection, the only support provided by the chicken sequences for a prehistoric link between South America and the Pacific is the radiocarbon dates of three immediately pre-Columbian South American chicken samples (calibrated at two sigma to A.D. 1304-1424, A.D. 1427-1459, and A.D. 1426-1457; Storey *et al.* 2008). These have been queried on the basis that dietary marine carbon may have artificially added inbuilt age to the samples (Gongora *et al.* 2008). As a result, the implications of the chicken data generated by Storey *et al.* (2007) remain unresolved. To clarify the issue of whether E haplotypes are actually present in the Pacific prehistorically or result from contamination, we used new aDNA protocols, shrimp DNase anti-contamination measures and new samples to examine Pacific chicken diversity and further investigate Polynesian migration routes. In addition, we re-examined a selection of the previously published Pacific archaeological chicken samples, including the key Rapa Nui specimen (PAQANA011) used to link east Polynesia and South America (Storey *et al.* 2007).

Results

Thirty-seven archaeological chicken bones from Polynesia were analysed, 22 of which (59%) yielded positive PCR amplification and DNA sequencing results (Niue, n=2/8; Hawaii, n=7/11; Rapa Nui, n=13/18; Table 1). A 330 base pair (bp) segment of the mtDNA control region (CR) was amplified from each specimen in short overlapping fragments (either 250bp and 305bp; or 151bp, 187bp, and 192bp, Table S1, Figure S1, and Supplementary Material and Methods). All 22 of the ancient Polynesian chicken specimens with replicable results produced mitochondrial sequences belonging to clade D, a South East Asian lineage whose modern distribution closely follows the distribution of cockfighting in India, Indonesia, China, and Japan (i.e. is 'localised' to the Asia-Pacific region, see Figures 1 and S2; Liu *et al.* 2006). Additionally, two samples could not be reliably reproduced (one from Niue and one from Rapa Nui): they each generated a single PCR product that produced different non-D haplotypes (Tables 1 and S2). However, when double-strand specific Shrimp DNase was used to remove contaminating DNA from reagents (Champlot *et al.* 2010) these sequences were no longer detected (Supplementary Information).

We also analysed 17 naturally shed modern feather samples from the Marquesas and Hawaii, of which 15 (Marquesas, n=5; Hawaii, n=10; Table 1) gave positive amplification and DNA sequencing results for the same 330 bp of the mtDNA CR.

Damaged nature of ancient Pacific chicken samples

The sequences from the ancient Pacific chicken samples showed elevated levels of template fragmentation and miscoding lesions, which is not unexpected given the tropical and sub-tropical conditions of Pacific archaeological deposits (Robins *et al.* 2001). Elevated levels of template damage encourage the amplification of sporadic exogenous contamination, and indeed the laboratory consumables appeared to contain detectable amounts of both haplogroup E and A (the two most frequent haplogroups worldwide: common in South (Kanginakudru *et al.* 2008; Silva *et al.* 2009) and East Asia (Liu *et al.* 2006; Oka *et al.* 2007; Berthouly-Salazar *et al.* 2010; Dancause *et al.* 2011) but with haplotypes also in the Middle East, Europe, South America and Africa (Liu *et al.* 2006; Gongora *et al.* 2008; Muchadeyi *et al.* 2008; Adebambo *et al.* 2010). This is a potential explanation for the contrast between the Pacific-specific haplogroup D sequences found in this study and the 'worldwide' haplogroup E sequences obtained by Storey *et al.* from all Pacific archaeological specimens outside Rapa Nui (Storey *et al.* 2007). Interestingly, within Rapa Nui, Storey *et al.* also obtained haplogroup D sequences from five out of six archaeological bones, although this was not clear as phylogenetic analyses were not performed (Gongora *et al.* 2008).

Table 1. Pacific samples and corresponding haplotypes discussed in the text.

Locations		*# Samples*	*Haplotypes*					*References*
Island	*Site name*	*No. Samples attempted/Samples*	*D_1 sub-clade*	*D_2 sub-clade*	*D01**	*E*	*A*	
Ancient								
Rapa	Anakena	14 (1)/21	14	-	-	(1)	-	This study
Niue	Anatoloa	2/2	2	-	-	-	-	This study
	Paluki	0 (1)/6	-	-	-	-	(1)	This study
Hawaii	Makauwa	7/11	7	-	-	-	-	This study
Modern								
Marques	-	5/6	3	-	-	2	-	This study
Hawaii	Kokee	10/11	-	-	-	10	-	This study
Vanuatu	Efate	14	6	5	-	2	1	Dancause
	Tanna	3	2	-	1	-	-	
	Ambae	7	4	3	-	-	-	
	Aneityum	19	17	2	-	-	-	
Guam	-	5	1	2	-	2	-	Dancause

* - Haplotype from Liu *et al.* (2006)

(1) - Samples not replicable and concluded to be contamination by Shrimp DNAse experiment.

Efate has a population size of 42128, and an island size of 900km^2

Tanna has a population size of 25840, and an island size of 550km^2

Ambae has a population size of 9418, and an island size of 400km^2

Aneityum has a population size of 821, and an island size of 160km^2

Critically, the sixth Rapa Nui archaeological bone (PAQANA011) studied by Storey *et al.* (2007) yielded a haplogroup E sequence identical (apart from one ambiguous site) to their Chilean archaeological sample (CHLARA001), and this close genetic and geographic relationship was used to argue for pre-Columbian East Polynesian/South American contact. The elevated levels of template damage and risk of contaminating haplogroup E sequences in laboratory consumables suggest that this key result should be re-examined, especially given the otherwise ubiquitous nature of haplogroup D in Polynesian archaeological material. We obtained further samples from the original PAQANA011 bone, and performed independent DNA extraction and sequencing analysis in Adelaide and Durham. The analyses revealed that PAQANA011 exhibited high levels of template damage, with cloned sequences indicating that numerous PCR amplifications were initiated from single template molecules (Table S3). Direct sequencing and cloning analyses (Table S2 & S4) clearly placed PAQANA011 within the 'localised' D haplogroup, in contrast to the 'worldwide' haplotype E sequence reported by Storey *et al.* (2007). This finding indicates that all 16 Rapa Nui archaeological specimens examined to date belong to the D haplogroup. Within the cloning reactions, we note that one of ten amplicons sequenced at

Durham was contaminated by haplotype A35 exogenous chicken DNA (Table S4). Consequently, the PAQANA011 specimen appears to be particularly susceptible to contamination in PCR reactions (Supplementary Information; Table S3).

As the Storey *et al.* (2007) ancient Pacific dataset spans a similar geographic range to our dataset (from Niue/Tonga in the east to Rapa Nui in the west and as far north as Hawaii), the probability of detecting each haplogroup in the proportions observed in each study can be directly compared. If the E haplogroup is truly present within ancient Pacific chickens at the levels detected by Storey *et al.* (55%; 2007), then the probability that all 22 of our ancient samples would belong to the D haplogroup is practically zero (P-value = 2.9×10^{-8}). As Figure S3 shows, the probability distributions for each dataset are completely different – with our results showing 95% probability that if haplogroup E is present at all, it is in less than 13% of ancient Polynesian chickens. This contrasts with the Storey *et al.* paper (2007), as they show a 95% probability that the frequency of haplogroup E in their samples is greater than 29%. If, for arguments sake, haplogroup E chickens are truly present in low frequencies (e.g. 10%), then there is a small chance that they were not found in our 22 ancient samples simply due to stochastic effects (i.e. poor preservation conditions concealing haplogroup E from detection just by chance; P-value = 0.0985). However, if E is truly present at only 10% then it is highly unlikely that Storey *et al.* would observe 6/11 of their ancient Pacific chickens to be authentic haplogroup E sequences (P-value = 2.7×10^{-4}).

Phylogeography

The combined ancient and modern sequences were analysed along with data from modern Pacific chickens reported in Dancause *et al.* (2011), and global chicken sequences from Genbank. It was necessary to exclude sequences from Storey *et al.* (2007; 2010) from the Network analysis due to the short fragment lengths (167-201bp), and the lack of contamination removal procedures such as Shrimp DNase treatment. Figure 2 shows that the D haplotypes generated from our 22 ancient Polynesian samples clustered together to form a novel sub-clade (designated here as 'D_{AP}' or 'Ancient Pacific' sub-clade; Figures 2 and S2), which has not been detected outside the Pacific. This D_{AP} sub-clade remains common in modern chicken samples from the Marquesas (n=3/5) and Vanuatu (Dancause *et al.* 2011), suggesting a relatively high level of genetic continuity on these islands since prehistoric times. Dancause *et al.* (2011) also reported a modern D_{AP} sequence from Guam (n=1/5); although interestingly, this particular D_{AP} haplotype is not shared with Vanuatu. A cluster of haplotypes, designated here as D_{WP} (or 'Western Pacific'; see Figure 2), was found in all but one of the remaining modern D haplotypes from the Pacific (Dancause *et al.* 2011) and, similar to D_{AP}, neither the Guam nor Vanuatu D_{WP} haplotypes were shared between the island chains. In fact D haplotype frequencies between modern chickens from Guam and Vanuatu are significantly different (p-value = 0.00527; Table S5). The sole exception

to the D_{WP} or D_{AP} haplogroups was a D haplotype from Vanuatu (Dancause *et al.* 2011) previously defined as D01 (Liu *et al.* 2006), which is common in China, Myanmar, and India. The remainder of the modern Marquesas samples (n=2), and all the modern Hawaiian specimens (n=10) were found to belong to the worldwide E haplogroup, as did the balance of the Guam (n=2) and Vanuatu (n=2) sequences from Dancause *et al.* (2011) (Figure 1).

By comparing the observed data of modern South American chicken haplogroups (Gongora *et al.* 2008) to the haplogroup frequencies found in ancient Pacific (this study), modern Pacific (Dancause *et al.* 2011), and modern European chickens (Liu *et al.* 2006), we find that South American chicken haplogroups have more similar frequencies to those in European chickens (p-value = 0.059) than to ancient Pacific (p-value = 2.2×10^{-16}) or modern Pacific chickens (p-value = 2.2×10^{-16}; see Table S5 and Figure S5). In fact, in the PCA (Figure S6) principal component 1 (PC1) accounts for almost 70% of the variation between these chicken populations, with PC1 clearly separating modern European/South American chickens from all Pacific chickens. Haplogroups A, D and E have no significant difference in frequencies between modern South American and European sampled populations, while all our ancient Pacific and most of the modern Pacific samples had significantly different frequencies (p-values < 0.01 for all but haplogroup A where $p = 0.11$ for a match to modern chicken frequencies in Vanuatu). Although haplogroup B was present in significantly different frequencies between modern South American and European populations (15% vs. 2% respectively), the fact that haplogroup B chickens are present in both Europe and South America, but not in the Pacific to date highlights the close relationship between modern chicken haplogroups in South America and Europe.

Node size proportional to number of samples

132 35 14 8 2 1

D_WP
Western Pacific

D_AP
Ancient Pacific

Ancient Pacific 'D' samples
Modern Pacific 'D' samples
Modern non-Pacific 'D' samples
Haplotype E01 as outgroup

E01
D10 267
D24
D15 272
279 D25
D26 280
75
380
165
159 166
147
132 155 270 D13
157
268 D11
158
387
D06 52
153 D04
76
D23 73
278 D22
261 D01
365
364
386
384
385
381
171
D02
49 170
48
47
51
167
169
45
46
274 D17
273 D16
156 275 D18 D21
276 D19
277 D20
D08 265
D09 266
264 D07
D03 262

1bp
Line length proportional to sequence differences

Figure 2. Unrooted haplotype network generated using 330bp of mitochondrial control region for Polynesian chickens analysed in this study and other haplotype D sequences in Genbank (Liu *et al.* 2006; Oka *et al.* 2007; Muchadeyi *et al.* 2008; Silva *et al.* 2009; Berthouly-Salazar *et al.* 2010; Dancause *et al.* 2011; Mwacharo *et al.* 2011). Labels in larger font next to node circles are haplotype designations according to Liu *et al.* (Liu *et al.* 2006), while labels in smaller font next to (or within) node circles are haplotype numbers in this study (Dataset S1). Colors reflect sampling location, with sub-clade D_{AP} and grouping D_{WP} circled with dashed lines.

Discussion

The environmental conditions in the Pacific (and possibly the Chilean) archaeological deposits are clearly not optimal for long term DNA preservation (Robins *et al.* 2001), and this is reflected in the low amplification success rate. Just over a third of the 81 ancient chicken bones so far analysed from the Pacific (Storey *et al.* 2007; Storey *et al.* 2010, this paper) have given replicable amplification and sequencing results (32/81, 39%). The majority of these were D haplotypes (25/32 or 78%), with E haplotypes comprising the rest (7/32 or 22%) (Storey *et al.* 2007; Storey *et al.* 2010). None of the Clade E sequences (Storey *et al.* 2007; Storey *et al.* 2010) were produced from analyses using anti-contamination measures such as Shrimp DNase, and indeed were often obtained from a single isolated bone from the oldest layers of archaeological deposits. When shrimp DNase is used (this study) it is highly likely (at the 99% level) that haplogroup E ancient chickens, if present at all, were at a low frequency in the Pacific (< 19%; Figure S3), which is not consistent with the Storey *et al.* (2007) dataset (p-value = 0.0074). When no shrimp DNase is used to remove potential lab contamination (as in Storey *et al.* 2007), it is highly likely (at the 99% level) that haplogroup E sequences are present in at least 29% of ancient samples (Storey *et al.* 2007). Given that at least the Rapa Nui haplogroup E sequence is now thought to result from contamination, we suggest that the other E haplotypes from archaeological specimens (Storey *et al.* 2007) should be considered questionable until replicated using treatments to remove potential contamination.

Phylogeography of D_{AP} and D_{WP}

Sequences from mitochondrial haplogroup D dominate the genetic diversity of our ancient and modern Pacific chicken sequences plus the Genbank sequences from Vanuatu and Guam (Dancause *et al.* 2011) (Figures 1 and 2), suggesting a strong association with ISEA and/or mainland South East Asia (MSEA) (Liu *et al.* 2006). The new sub-clade D_{AP}, found in all 22 archaeological specimens examined, has not been found outside the Pacific and appears to represent a commensal chicken breed of the Ancient Pacific. The D_{AP} sub-clade survives in modern chickens found in Eastern Polynesia, Western Polynesia, and Micronesia (Dancause *et al.* 2011), despite likely subsequent introgression of European breeds during the colonial and modern periods. The sister group D_{WP} has been reported from modern chickens in Japan (n=14), Indonesia (n=3), the Philippines (n=1), Thailand (n=1), and China (n=1) (Liu *et al.* 2006; Oka *et al.* 2007). Furthermore, it is widespread in modern chickens of the western Pacific, but was not detected further east than Vanuatu. Therefore, while D_{WP} indicates an ISEA/MSEA origin it is currently restricted to Western Polynesia/Micronesia, and it appears only the D_{AP} sub-clade was carried further east.

The remaining haplogroup D sequence detected in the Pacific (D01; Dancause *et al.* 2011) was found only in modern chickens from western Polynesia, and falls within a clade associated with a high level of human-mediated transportation. Although D01 has a predominantly South and East Asian homeland (including India; Liu *et al.* 2006), closely related haplotypes extend as far west as east Africa (e.g. modern chickens in Madagascar, Kenya, Sudan, Zimbabwe and Malawi; Dataset S1; Liu *et al.* 2006; Muchadeyi *et al.* 2008; Mwacharo *et al.* 2011). While tempting to suggest these sequences might reflect the dispersal of Austronesian speakers westward to Madagascar (Bellwood 1991; Razafindrazaka *et al.* 2010), it will be necessary to examine ancient chickens from this African distribution to determine the antiquity of such a connection.

The vast majority of modern chickens sampled in Vanuatu by Dancause *et al.* (2011) fell within haplogroup D (93%, Table 1), and were found on all the islands sampled including the larger, more highly populated island of Efate (and to a lesser extent Tanna), as well as the smaller, more undeveloped islands of Ambae and Aneityum. The remaining modern Vanuatu samples fell within the globally widespread E and A haplogroups (n=2 and n=1, respectively), and importantly, were only found on the most populated island, Efate (Dancause *et al.* 2011). The high proportion of D_{AP}, and to a lesser extent D_{WP}, haplotypes found in Vanuatu (and in particular on the more remote islands), suggests that these D haplotypes represent the prehistoric chicken diversity. Furthermore, the more remote islands have significantly different haplogroup frequencies than the more populated island of Efate (p-value = 0.0295; Figure S4 and Table S5). In contrast, the few isolated A and E haplotypes restricted to Efate suggest more recent (possibly historic) introductions, as both haplogroups (but especially E) have been distributed globally via European colonial contact (Liu *et al.* 2006; Gongora *et al.* 2008; Muchadeyi *et al.* 2008; Adebambo *et al.* 2010). Similarly, we consider the archaeological E haplotypes obtained by Storey *et al.* (2007) to represent contamination because at least some of these specimens were actually D_{AP}, and none of our ancient samples were found to contain replicable E haplotypes. As a result, we suggest that only the D_{AP} and D_{WP} sequences reflect the pre-European spread of people into the Pacific, and this is strongly supported by the apparent restriction of D_{AP} haplotypes to the Pacific.

Previous studies of prehistoric Pacific chickens (Storey *et al.* 2007; Storey *et al.* 2010) have proposed a two-wave migration model in Remote Oceania/Western Polynesia, with each wave characterised by a different mtDNA haplogroup (clades D and E), with E used to demonstrate contact with South America. However, our sequences suggest only clade D sequences were moved across the Pacific, potentially in one or two waves. A one-wave scenario would be consistent with the movement of a single mixed population from MSEA/ISEA, containing both D_{AP} and D_{WP} Pacific chickens (and possibly D01), eastwards as far as

Vanuatu/western Polynesia (Figure 1: arrow 1A). Subsequent dispersals (potentially as separate phases) moved a subset of this genetic diversity (D_{AP}) into eastern Polynesia, where it became widely distributed (Figure 1: arrows 3-6); and another subset (both D_{AP} and D_{WP}) as a back migration into Micronesia (Figure 1: arrow 1B). Several alternative two-wave scenarios can be proposed. Perhaps the most likely being where both D_{AP} and D_{WP} sequences result from a single mixed population of ISEA origin transported separately to: Micronesia, but no further (Figure 1: arrow 2); and to western Polynesia (Figure 1: arrow 1A), with D_{AP} only being distributed widely across East Polynesia (Figure 1: arrows 3-6). Although a sampling bias may explain the non-overlapping D_{AP} and D_{WP} haplotypes between Guam and Vanuatu (and significant p-value of 0.00097; Table S5), the current data is more consistent with this two-wave hypothesis. Importantly, a similar two-wave genetic pattern was suggested for the human dispersal of pigs across the Pacific (Larson *et al.* 2007), again with the separate dispersal through Micronesia not appearing to reach Western Polynesia.

Recently, Addison and Matisoo-Smith (2010) further developed the two-wave model of Polynesian origins put forward by Storey *et al.* (2007; 2010), by compiling molecular data from all the Polynesian commensals. Addison and Matisoo-Smith (2010) suggest one wave moved through Micronesia to Western Polynesia (1,500-2,000 ya) where it mixed with an earlier Lapita migration wave (2,800-3,500 ya). Our chicken data would be consistent with such a pattern if Micronesian D_{WP} or D_{AP} sequences were present in Western or Eastern Polynesia. However, this is not the case. In addition, the lack of convincing archaeological evidence for early domestic animals across many areas of Micronesia (Wickler 2004; Anderson 2009) provides little support for this early Micronesian wave, and if anything suggests that the Micronesian chicken diversity potentially represents a late dispersal (<1200 ya; Wickler 2004). In this regard, genetic sequences from radiocarbon-dated chicken bones in the area have the potential to further resolve the role of Micronesia as a stepping stone, trading partner, or terminus of back-migrations in the prehistoric Pacific.

South American contact

The presence of D_{AP} sequences in ancient or modern South American samples would provide strong evidence that East Polynesians introduced chickens to South America in prehistoric times. However, no D_{AP} sequences have been detected in the single ancient sample (Storey *et al.* 2007) or 41 modern specimens examined, which include multiple native breeds (Gongora *et al.* 2008). Consequently, there is no genetic evidence in support of a prehistoric introduction of Polynesian chickens, and if anything the absence of Polynesian sequences in South America to date argues against it. In fact, the close resemblance in haplogroup frequencies between modern South American and European chickens sampled to date, suggests these South American chickens stem from post-Columbian introductions. The only way to reconcile these

results (lack of D haplogroups and the similar frequencies in ‘worldwide’ haplogroups between South America and Europe) with a prehistoric link between South America and the Pacific is based on wholesale replacement of the early Polynesian chickens with European chickens after colonisation. This is surprising as Polynesian-introduced chickens might be expected to have become rapidly and widely established through pre-Columbian trade links.

Materials and Methods

Samples. Eight chicken bones from Niue were provided by Richard Walter (Otago University Anthropology department) from archaeological sites in Paluki and Anatoloa; 11 ancient Hawaiian chicken bones from an excavation at Makauwahi Cave on Kauai, Hawaii (National Tropical Botanical Gardens collection); and 18 Rapa Nui chicken bones excavated from a site at Anakena (University of Hawaii Rapa Nui Field School). Sequences from five of these Rapa Nui samples were previously published by Storey *et al.* (2007) (Table S2). Modern chicken feathers of unknown age were also collected non-invasively: Eleven feathers from Kokee, Kauai, were obtained by DB, and six feathers from the Marquesas, French Polynesia, collected by TH.

aDNA Extraction, Amplification, and Sequencing. Samples were extracted, amplified, and sequenced in specialist aDNA laboratories at the Australian Centre for Ancient DNA (ACAD) in Adelaide, South Australia, according to a range of strict protocols, including numerous controls (Cooper and Poinar 2000). A modified Qiagen silica extraction protocol was used to extract DNA (Supplementary Materials and Methods). PCR was used to amplify overlapping sections of the mtDNA CR (trimmed to 330bp common to both ancient and modern datasets; Figure S1), and sequencing was performed using an ABI 3130 or 3730 (Supplementary Materials and Methods). Independent external replication was performed in a dedicated aDNA lab in the Archaeology Department at Durham University following strict laboratory procedures as per commonly used guidelines (Cooper and Poinar 2000). Cloning of both the initial and independently replicated PCR fragments from bone sample PAQANA011 was performed using Stratagene and/or Topo cloning kits using manufacturers instructions after an A-tailing procedure at the ACAD laboratories (Supplementary Material and Methods).

Sequence analysis. In addition to the 37 sequences generated in this study, we downloaded 866 mtDNA CR chicken sequences from Genbank (Liu *et al.* 2006; Oka *et al.* 2007; Storey *et al.* 2007; Gongora *et al.* 2008; Kanginakudru *et al.* 2008; Muchadeyi *et al.* 2008; Silva *et al.* 2009; Adebambo *et al.* 2010; Berthouly-Salazar *et al.* 2010; Dancause *et al.* 2011; Mwacharo *et al.* 2011) to establish the geographic distributional range for each chicken haplogroup. These 903 aligned chicken CR sequences (trimmed to 330bp) were collapsed to 404 unique haplotypes (Dataset S1). ModelGenerator (Keane *et al.* 2006) was used to establish the best model to fit the data (GTR+I+G). Identification of the Pacific chicken sequences to

haplogroup level was undertaken by generation of a Maximum Likelihood tree (PhyML; Figure S2; Guindon *et al.* 2010). We also explored the data in SplitsTree4 (Huson and Bryant 2006), using the NeighborNet algorithm, and found that the data appeared not to be tree-like, so a Median Joining Network (using Network v4.6; Bandelt *et al.* 1999) was also generated for just the D haplogroup (Supplementary Material and Methods; Figure 2).

Statistical analysis. The tests of statistical significance were performed using the binom.test and fisher.test commands and the probability distribution graphs were created using the dbinom command, in the R 'stats' package (R development core team 2011).

Acknowledgements: We thank Jessica Metcalf for assistance with sample extraction, John Terrell for manuscript discussions, and the generous assistance of the individuals and institutions listed in the *Materials and methods* section that provided access to samples. This work was supported by the Australian Research Council grants to AC and ARC Discovery grant DP110105187.

Supplementary Information – Materials and Methods

aDNA Extraction, Amplification, and Sequencing. All samples were extracted, amplified, and sequenced in specialist aDNA laboratories at the Australian Centre for Ancient DNA (ACAD) in Adelaide, South Australia, according to a range of strict protocols and including controls (Cooper and Poinar 2000). Ancient bone samples (n=37) were extracted and PCR experiments set up in the physically remote ACAD ancient laboratory, whereas the feathers (n=17) were extracted and PCR experiments set up in either the remote ACAD ancient lab or the physically remote ACAD pre-PCR clean-room laboratory. The ACAD ancient laboratory is kept at positive air pressure, and uses standard ancient DNA precautions (clean body suits, face shields, multiple pairs of gloves), universal UV irradiation, and regular cleaning with oxidising agents to prevent contamination. The ACAD pre-PCR cleanroom uses many of the same ancient DNA precautions (laboratory coats, face masks, multiple pairs of gloves) and undergoes regular cleaning with oxidising agents to prevent contamination. All reagents and materials are supplied DNA-free and used only once. Separate hoods are used for DNA extractions and to set up PCR experiments, and the amplification and post-PCR procedures are performed in a post-PCR laboratory located in a separate, physically remote building, to minimise the potential for contamination by PCR amplicons. Uni-directional movement of researchers and materials are carefully controlled to further reduce contamination risk.

Bone extractions. Each chicken bone was ground to fine powder in a Mikrodismembrator (5000 rpm, for 10 seconds). Approximately 70 mg of bone powder was decalcified concurrently with protein digestion by incubation at 55 °C overnight in 1mL of extraction buffer (consisting of 0.4725 M EDTA (pH=8.0), 0.2 % sodium dodecyl sulphate (SDS), and 0.7 $mg.ml^{-1}$ Proteinase K). After digestion, samples were centrifuged at 10,000 rpm for 5 mins and the supernatant was transferred to an Amicon ultra-4 (Millipore), which was centrifuged at 4000 xg until only 100 µL supernatant remained. The supernatant was washed with 1 mL molecular grade water and centrifuged again (at 4000 xg until only 100 µl remained). An equal volume of ATL buffer (Qiagen DNeasy kit) was then added, mixed, and the supernatant removed to a 2 mL screw-cap tube. The supernatant was incubated for 10–60 mins at room temperature on a rotary mixer after the addition of an equal volume of AL buffer (Qiagen DNeasy kit) and 0.02 $\mu g.\mu l^{-1}$ of carrier RNA. After the incubation, an equal volume of ethanol (100 %) was added, and then the total volume was transferred to a Qiagen DNeasy spin column where it was incubated at room temperature for 10–60 mins. The extraction then followed the Qiagen DNeasy kit instructions, with the following exceptions at the elution stage: 100–150 µL of warmed AE buffer was added and then incubated at room temperature for 10–30 mins, before being centrifuged at 8,000 rpm for 1 min, this step was repeated to finish with 200–300 µL of total volume.

Feather extractions. Approximately 5 mm of each feather tip was rehydrated overnight with 1 ml phosphate buffered saline (PBS) on a rotary mixer at room temperature. On day 2, the supernatant was removed, the feather tip was macerated using a clean scalpel blade, and the sample was digested in 440 µl of digestion buffer (comprising ATL buffer (Qiagen DNeasy kit) with 1.8 mg.ml^{-1} Proteinase K, and 90 mM Dithiothreitol) overnight at 55 °C on a rotary mixer. After digestion, 400 µL of AL buffer (Qiagen DNeasy kit) and 0.02 µg.µl^{-1} of carrier RNA was added and incubated at room temperature on a rotary mixer for 10–30 mins, after which 400 µL of 100 % ethanol was added. The supernatant (650 µl) was incubated on a Qiagen DNeasy spin column for 10–30 mins before being centrifuged at 8000rpm for 1 min. This incubation was then repeated until all of the supernatant had been centrifuged through the column. The feather extraction protocol then followed that of the bone extraction procedure above.

PCR amplification and sequencing. PCRs were set up using 25 µL volumes containing a final concentration of 1 U Platinum Taq DNA Polymerase High Fidelity (Invitrogen), 1 x PCR Buffer (Platinum, Invitrogen), 3 mM $MgSO_4$, 200 µM each dNTP, 2 mg.ml^{-1} rabbit serum albumin (Sigma), 1 µM forward and reverse primers and 2-3 µl of template DNA. Primers used to generate the D-loop sequences are listed in Table S1. See Figure S1 for diagram describing primer arrangement.

Table S1. Primer sequences

Primer name	Primer sequence (5' to 3')	Reference
GG144F	ACCCATTATATGTATACGGGCATTAA	(Storey *et al.* 2007)
GG387R	CGAGCATAACCAAATGGGTTAGA	(Storey *et al.* 2007)
GG316F	AACAAGTCACCTAACTATGAATGGTTAC	(Storey *et al.* 2007)
GG586R	AGTTATGCATGGGATGTGCCTGACCGA	(Storey *et al.* 2007)
A1780F	CAGCTCCAAACCACTACCAAG	This paper
A1781R	AGGTGACTTGTTGGGGGAAG	This paper
A1958F	TCTAACTCATTTGGTTATGCTCG	This paper
A1959R	AGTTATGTATGGGATGTGCCTGACCGA	This paper

PCR reactions were performed on a Corbett Research Palm Cycler using the following cycling conditions: 94 °C for 2 min, 55 cycles of 94 °C for 30 s, 55 °C for 30 s, 68 °C for 30 s, and a final extension of 10 min at 68 °C. Amplifications of extraction and PCR controls were performed in all experiments to monitor contamination. PCR products were separated by electrophoresis on a 3.5 % agarose gel. Successful PCR products (10 µl) were purified using 0.8 µl of EXOSAP (Fermentas) at final concentration of 0.38 U/µl Exonuclease I, and 0.05 U/µl Shrimp Alkaline Phosphotase, and thermal cycled at 37 °C for 30 mins, 80 °C for 15 mins, and

15 °C for 3 mins on a Corbett Research Palm Cycler. The forward and reverse complements of each fragment were sequenced from the same PCR reaction using the same primers as for the PCR, and Big Dye Terminator v3.1 cycle-sequencing chemistry, followed by vacuum clean up on a Multiscreen®$_{384}$ SEQ plate (Millipore). The sequencing run was conducted on an ABI 3130XC capillary sequencer.

Primers GG144F/GG387R (Storey *et al.* 2007) and GG316F/GG586R (Storey *et al.* 2007) were used initially but as the PCR products amplified from these primers (fragment 1 and 2) are 250bp and 305bp respectively; additional primers were designed to cover the same range of mtDNA control region. Primer GG144F was paired with A1781 (187bp as fragment 3) and A1780 was paired with GG387R (151bp as fragment 4) to cover the equivalent DNA sequence as fragment 1 but in two overlapping fragments. Primers A1958 and A1959 (192bp as fragment 5) were used to cover the balance of the mtDNA CR under study. The use of this alternative primer set meant that a sequence gap was introduced in some ancient sequences equivalent to the primer binding region, however as this region of 23bp had only 8/901 instances of a variable site across the entire reference dataset (but at the 5' end of the primer which should still allow the primer to bind and amplification to occur), this 23bp gap was excluded from the dataset. Further trimming to the sequence length shared across all chicken specimens resulted in a final sequence length of 330bp.

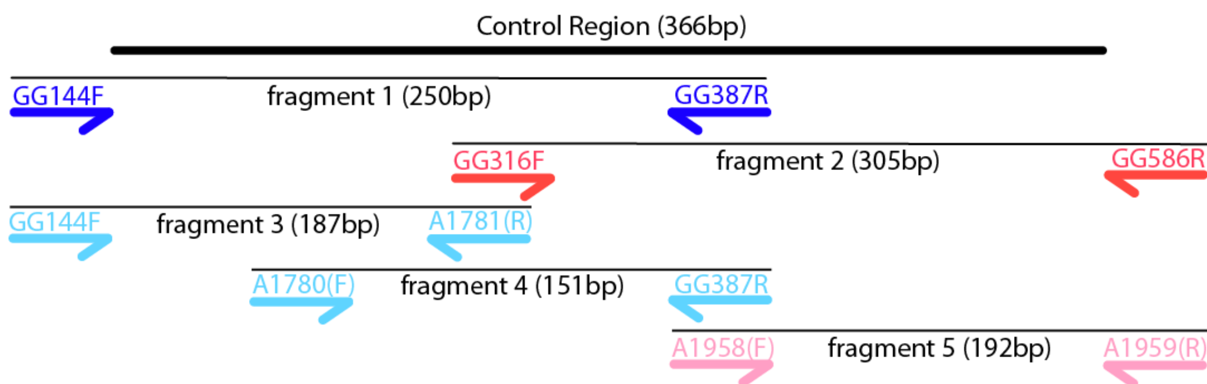


Figure S1. Details of primer arrangement showing the 366bp target region, prior to trimming sequences to the length common across both ancient and modern datasets (330bp).

Replication.

DNA extraction of the replicate ancient chicken bone fragment PAQANA011 was performed in a dedicated aDNA lab in the Archaeology department at Durham University following strict laboratory procedures as per commonly used guidelines (Cooper and Poinar 2000). All equipment and work surfaces were cleaned before and after each use with a dilute solution of bleach (10 %) followed by ethanol (99 %). The ancient chicken bone (~0.05 g) was pulverized in a Micro-dismembrator, digested in 0.425 M EDTA, 0.05 % SDS, 0.05 M Tris-HCI

and 0.333 $mg.ml^{-1}$ proteinase K and incubated overnight on a rotary mixer at 50 °C until fully dissolved. 2 ml of solution was then concentrated in a Millipore Amicon Ultra-4 30 KDa MWCO to a final volume of 100 µl. The concentrated extract was purified using the QIAquick PCR Purification Kit following manufacturers recommendations, except that the final elution step was performed twice to produce a final volume of 100 µl. A negative extraction control was performed alongside the ancient bone sample.

PCRs were setup in 25 µl reactions using 1.25 U Taq GOLD (Applied Biosystems), 1 x Gold buffer (Applied Biosystems), 2.5 mM $MgCl_2$, 0.5 $\mu g.\mu l^{-1}$ bovine serum albumine (BSA), 200 µM of each dNTP, 0.8 µM of each forward and reverse primers, and 2-5 µl of aDNA extract. We used PCR primers (5'-3') GG144F and GG387R (2); GG316F, and GG586R (2). One PCR negative control was included for every three aDNA template PCR tubes. We ran a total of 22 PCRs with aDNA template, eight PCR negative controls and two PCR negative extraction control. Neither the PCR negative controls nor the negative extraction control produced bands (PCR product) when analysed by gel-electrophoresis. PCR cycling conditions were 95°C for 5 min, 50 cycles of 94 °C for 45 sec, 54 °C for 45 sec and 72 °C for 45 sec, followed by 72 °C for 10 min. PCR products were stored at -20 °C. Sanger sequencing on the Applied Biosystems 3730 DNA Analyser was performed at the DNA sequencing service in the School of Biological and Biomedical Sciences at Durham University.

Cloning. The PCR products generated from bone sample PAQANA011 were cloned using Stratagene and/or Topo cloning kits using manufacturers instructions (after an A-tailing reaction). The A-tailing reaction consisted of a 20 µl volume reaction containing 0.125 U HotMaster Taq, 2.5 µM dATP, 10x HotMaster buffer, 17 µl cleaned PCR products. The Buffer, dATP's and Taq were activated at 94 °C for 2 mins prior to addition of the PCR products then a further incubation at 72 °C for 10 mins. The A-tailed PCR products were then cleaned up using an Isopropanol precipitation and resuspended in 10 µl of PCR grade water. Sanger sequencing of the cloned PCR products were performed according to the procedures outlined above.

Sequence analysis

The sequence chromatograms were formed into contigs and edited manually in Geneious (Drummond *et al.* 2011). Reference sequences downloaded from Genbank were aligned with our sequences using the ClustalW algorithm in Geneious (then refined by eye) to form a dataset of 903 sequences of 519bp in length. This dataset was then trimmed to the shortest common length of our ancient Pacific sequences and the 23bp gap created by the primer overlap was removed. This resulted in a full dataset of 330bp in length. Using Collapse v1.2 we reduced the full dataset down to haplotypes and for clarity, we distinguish 404 haplotypes, including A1-I3 designated by Liu *et al.* (2006) (Table S5). ModelGenerator (Keane *et al.* 2006) was used to establish the best model to fit the data, which was GTR+I+G based on the Akaike Information Criterion 2. In order

to establish which haplogroups the Pacific chickens belonged to we used PhyML (Guindon *et al.* 2010) to generate a Maximum Likelihood tree using the full dataset (903 sequences) and with a GTR site model and approximate Likelihood Ratio Test with Chi^2-based parametric branch supports on nodes. We explored the data in SplitsTree4 (Huson and Bryant 2006), using the NeighborNet algorithm, and found that the data appeared not to be tree-like, so a network was also generated of just the D haplogroup.

Network. All of the samples that fell within the D haplogroup from our phylogeographic analysis above (excluding the Storey *et al.* samples) were used to create a Median Joining network in the software Network v4.6 (Bandelt *et al.* 1999). DNASP v5.10.1 (Librado and Rozas 2009) was used to form the .rdf file for input into Network from the nexus file of the aligned sequence dataset, where gaps were considered. As all the ancient Pacific samples radiate out from the most common haplotype, this star cluster was designated sub-clade D_{AP}. The majority of the rest of the Pacific samples (all modern) fell within five haplotypes, each separated by individual SNP's in a linear progression, so these haplotypes were designated as group D_{WP}. However, as the D_{WP} grouping is within a reticulate area of the network, no monophyly was found within the phylogenetic tree. Therefore, the classification of it as a 'group' of haplotypes is based on purely geographic criteria (i.e. chickens with these five closely related haplotypes were found in Western Polynesia).

Supplementary Information

Ancient Pacific samples

Storey *et al.* (2007) found five of their six Rapa Nui samples fell within clade D (Table S2), with the sixth (PAQANA011) falling within clade E, which is found worldwide. Our re-extraction of this Rapa Nui bone sample (PAQANA011) contradicts this as we found it to be a D haplotype. Repeated amplifications and Sanger sequencing of this sample placed it within the D haplogroup (Table S4), however it also highlighted 15 C-to-T transitions between the 17 amplicons. This type of transition is commonly observed in aDNA because of post-mortem template damage, with the hydrolytic loss of amino-groups from cytosine converting the base to uracil, which DNA polymerases read as a thymine base (Paabo *et al.* 2004). As these internal PCR replications confirmed the discrepancy between our extraction (ACAD9068) and Storey *et al.*'s (2007) published sequence (EF535246) for bone sample PAQANA011, we had it independently replicated by another aDNA laboratory at Durham University (Table S4).

Independent replication

A subsample of the PAQANA011 bone was sent to Durham University, where it was extracted and three PCR amplifications were performed for each of fragments 1 and 2 of the mtDNA CR (as only fragment 1 is diagnostic to haplogroup level and below, it is this fragment that is compared to the Liu *et al.* (2006) dataset in the discussion below). Two different haplotypes were detected across the three amplicons of fragment 1 for PAQANA011. Two amplicons matched each other and fell within clade D. The third amplicon matched Haplotype A35, which is found in chickens from China and Japan (Liu *et al.* 2006), as well as two other Genbank samples: AM746039 (Muchadeyi *et al.* 2008) and AB263973, both of which are commercial breeds. Liu *et al.* (2006) found almost 95% of domestic chickens to belong to clades A, B, C, E, F, and G, which suggests the presence of clade A (this study) and clade E (Storey *et al.* 2007) amplicons from this sample may reflect lab consumable/reagent contamination by modern domestic chicken DNA. Although fragment 2 is not as phylogeographically informative as fragment 1, amplification of this fragment did reinforce the highly damaged/degraded nature of this particular sample, which may have allowed modern chicken DNA (at low levels in lab reagents/consumables) to occasionally outcompete the endogenous DNA. Across the three amplicons of fragment 2, seven randomly distributed C-to-T transitions were found to differentiate the sequences from our extract of this sample (ACAD9068; see Table S4). Although the damaged sites tended to be within the longer amplicons, this was not always the case.

Cloning

Cloning of the PCR amplicons from both labs (ACAD and Durham University) was undertaken at ACAD to confirm the Sanger sequencing results and to establish whether the C=>T

transitions were due to post-mortem damage. At the base pairs where C=>T transitions were initially detected, few differences were detected between the clones, which suggests that each amplicon was formed by amplification from a single damaged template (Table S3). However, at least four matching (non-damaged) amplicons were retrieved for both fragments 1 and 2, so that a consensus sequence could be generated (Table S3).

Shrimp DNase experiment to test ACAD3890 and ACAD9060 samples

From the 24 ancient samples successfully amplifying DNA, two samples (ACAD3890 from Niue, and ACAD9060 from Rapa Nui) yielded haplotypes other than from clade D. ACAD3890 matched Liu *et al.'s* haplotype 'A34', which is found in only one modern sample from Xinjiang, China (Liu *et al.* 2006). ACAD3890 had poor amplification/sequencing success (n=1/31), with only one amplicon (102bp) amplifying and sequencing successfully (using primers A1780 and GG387R). The second sample, ACAD9060, matched E01 found commonly worldwide (China, n=19; India, n=10; Sri Lanka, n=20; Japan, n=27; Iran, n=3; Turkmenistan, n=3; UK, n=2; Europe, n=34; Chile, n=25, Kenya, n=58) (Liu *et al.* 2006; Oka *et al.* 2007; Gongora *et al.* 2008; Mwacharo *et al.* 2011). This sample also did not amplify often (n=8/14), with the E01 haplotype only occurring once (all other PCR amplicons could not be successfully sequenced). Accordingly, contamination by modern chicken DNA was suspected as the source of both the A34 and E01 haplotypes, however due to the stringent aDNA procedures in place at the ACAD, there is limited opportunity for modern chicken DNA to enter the lab. The possibility of contamination in the laboratory consumables/reagents was tested by the addition of Shrimp DNase to three sets of subsequent PCR reactions for all 24 samples. Shrimp DNase is an endonuclease that cleaves phosphodiester bonds in double stranded DNA. It is often used to treat PCR master mixes prior to the addition of extracted DNA in order to break down contaminating modern DNA in PCR reagents. No DNA was successfully amplified after Shrimp DNase treatment of PCR's for the ACAD3890 sample, while Shrimp DNase treatment of PCR's for ACAD9060 sporadically gave haplotype D sequence across a variety of fragment sizes (210bp, 190bp, 129bp, and 90bp), plus one sequence (116bp) that could not be assigned to any haplogroup – it had 7 mismatches from its closest BLAST matches (94% identity). Both of these results suggest PCR reagent contamination by modern chicken DNA was the likely source of the original A34 and E01 haplotypes. When ACAD9060 did give a D haplotype, it matched the D haplotype from the other ancient Pacific samples for fragment 1, however fragment 2 could not be amplified so this sample was excluded from further analyses. The possibility of laboratory consumable/reagent contamination needs to be discussed more in aDNA studies, especially when the use of a simple PCR additive, such as Shrimp DNase, can rule out one source of possible contamination (i.e. PCR reagents; USB 2006). This additive (or similar) is essential for aDNA studies of commensal or domesticated species, where DNA from modern populations of the same

species may permeate factories where lab consumables and/or reagents are produced (Leonard *et al.* 2007).

Although the overall percentage of E haplotypes (7/81, 9%) detected in ancient Polynesian chicken samples is higher than the nominal 5% contamination rate of modern domestic species found in lab consumables (Leonard *et al.* 2007), low levels of preserved endogenous DNA may allow any contaminating modern chicken DNA to outcompete them in PCR reactions (unless an endonuclease such as Shrimp DNase is used to remove contaminating DNA in PCR lab reagents). Previously, some of the Storey *et al*. co-authors have reported low amplification success from Mele Havea (Tonga) and Paluki (Niue) (Robins *et al.* 2001) — and in our analyses of Paluki material we identified a non-D haplotype to be contamination from PCR reagents (Table S2).

Phylogenetic analyses including Storey *et al.* (2007) data

Although we did not include the Storey *et al.* (2007) samples in our Network analysis, we did include them in the Maximum Likelihood phylogenetic analysis (Figure S2; Table S5). For those Storey *et al.* (2007) samples we could independently replicate (i.e. there was sample left) and for those we could not (i.e. there was no sample left), phylogenetic analyses allowed us to confirm that their five D sequences from Rapa Nui form part of the Ancient Pacific D_{AP} sub-clade (see Table S2). This means that D_{AP} sequences have been independently determined from Pacific archaeological material multiple times.

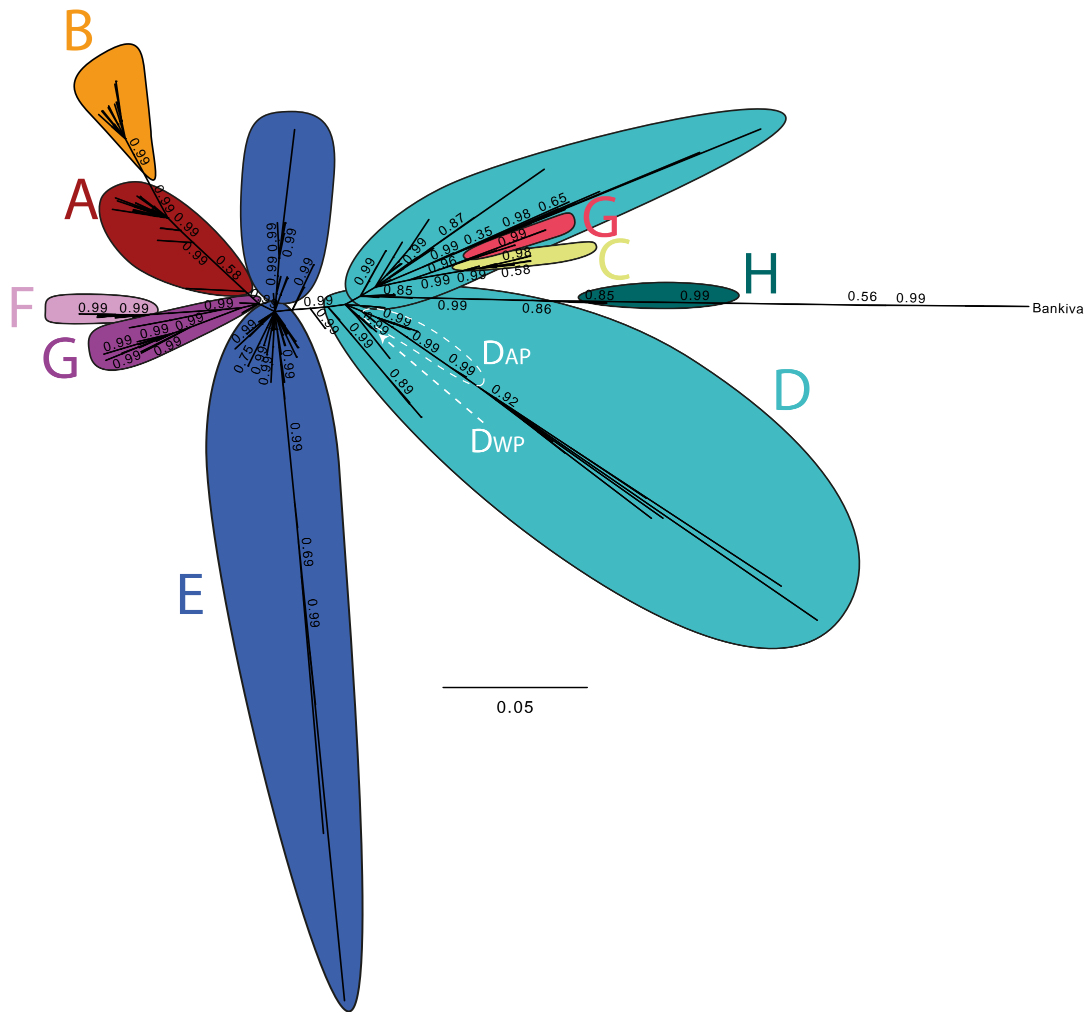


Figure S2. Maximum Likelihood tree constructed using PhyML with 330bp of mitochondrial control region for 903 *Gallus gallus* sequences worldwide (see Table S7 for list of samples), with *G. g. bankiva* as an outgroup. Colors and labels in this figure represent each of the nine worldwide chicken haplogroups initially identified in Liu *et al.* (2006), with the addition of our D_{AP} and D_{WP} groups. The support values on branches are estimated using a Chi^2-based approximate Likelihood Ratio Test (aLRT) – the D_{AP} and D_{WP} groups have branch support of 0.999 and 0.9999, respectively.

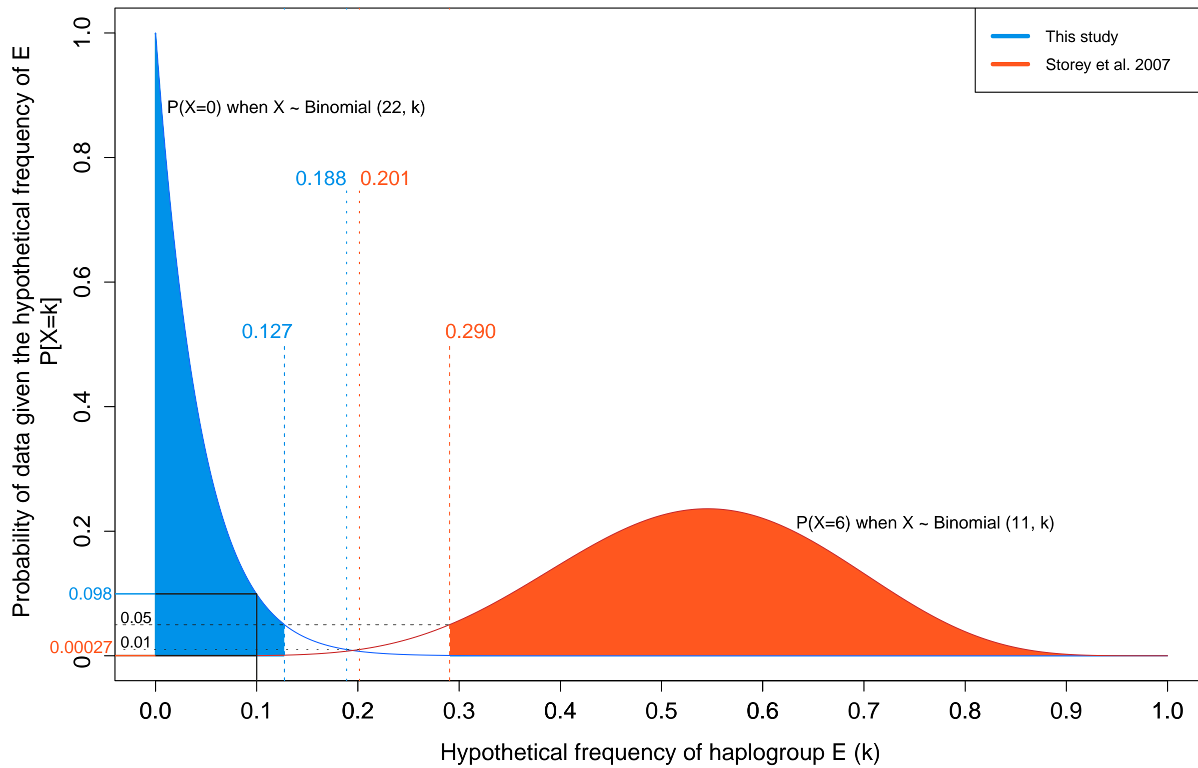


Figure S3. Binomial probability distribution showing the probability of detecting the observed number of haplogroup E sequences (for a range of hypothetical frequencies of E in the ancient Pacific chicken population). For example, if we assume that haplogroup E is actually present at a frequency of 0.1 (i.e. 10%) in the total ancient population, then the probability of detecting 0/22 haplogroup E sequences is 0.098 (this study), but the probability of Storey *et al.* (2007) detecting 6/11 haplogroup E sequences is 0.00027.

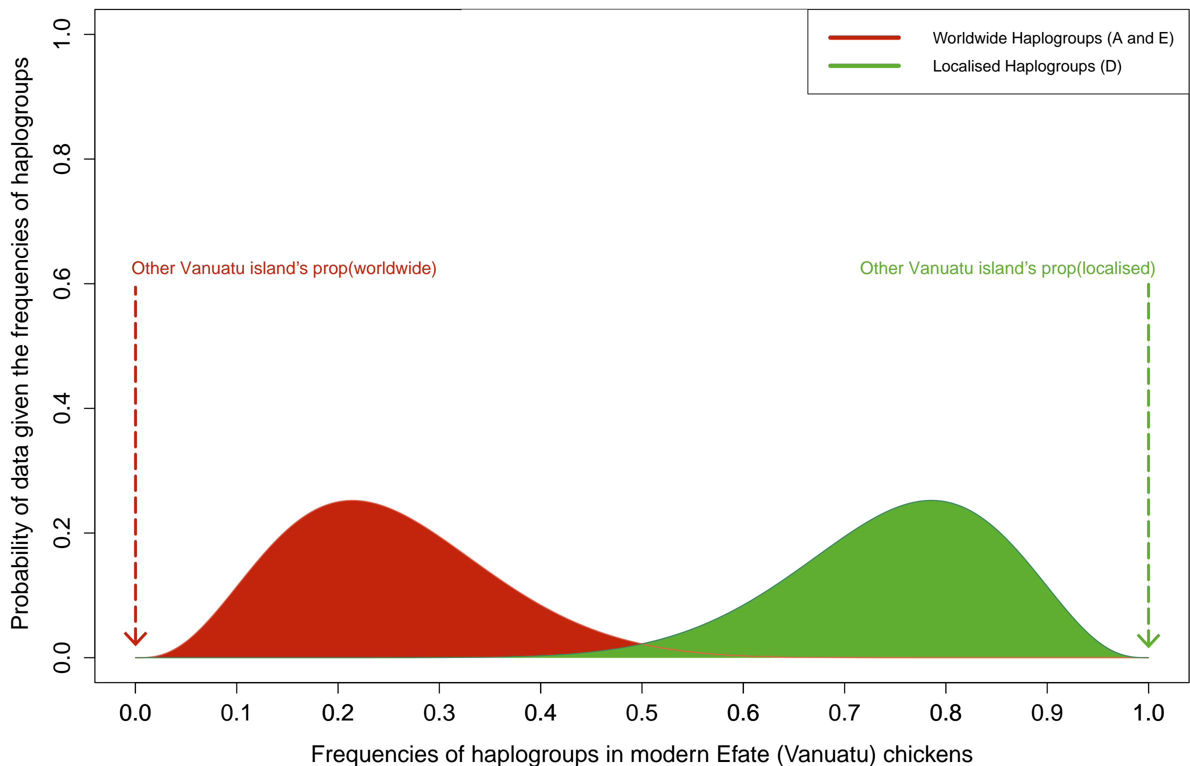


Figure S4. Probability distributions of 'worldwide' vs. 'localised' haplogroups from modern chickens from Efate and the proportion of these haplogroups on the other Vanuatu islands (Dancause *et al.* 2011). The modern Vanuatu chickens sampled from Efate include two worldwide haplogroups not present on the other less populated islands. The Fisher's exact test (in table S5) comparing Efate haplogroup frequencies to those on the other Vanuatu islands, found that the haplogroup frequencies in Efate are significantly different to those on the other islands (p-value = 0.0295). Furthermore, if we assume that the modern proportions of 'worldwide' haplotypes on Efate reflect the true chicken populations in Vanuatu generally, then the probability that the 'worldwide' haplogroups were missed just by chance on the less populated islands is very small (p-value = 0.00092). Therefore, we posit that the 'localised' (haplogroup D) chickens represent the prehistoric chickens of the Pacific, but that the 'worldwide' haplogroups are only present on Efate (the most populated island) due to later European introductions.

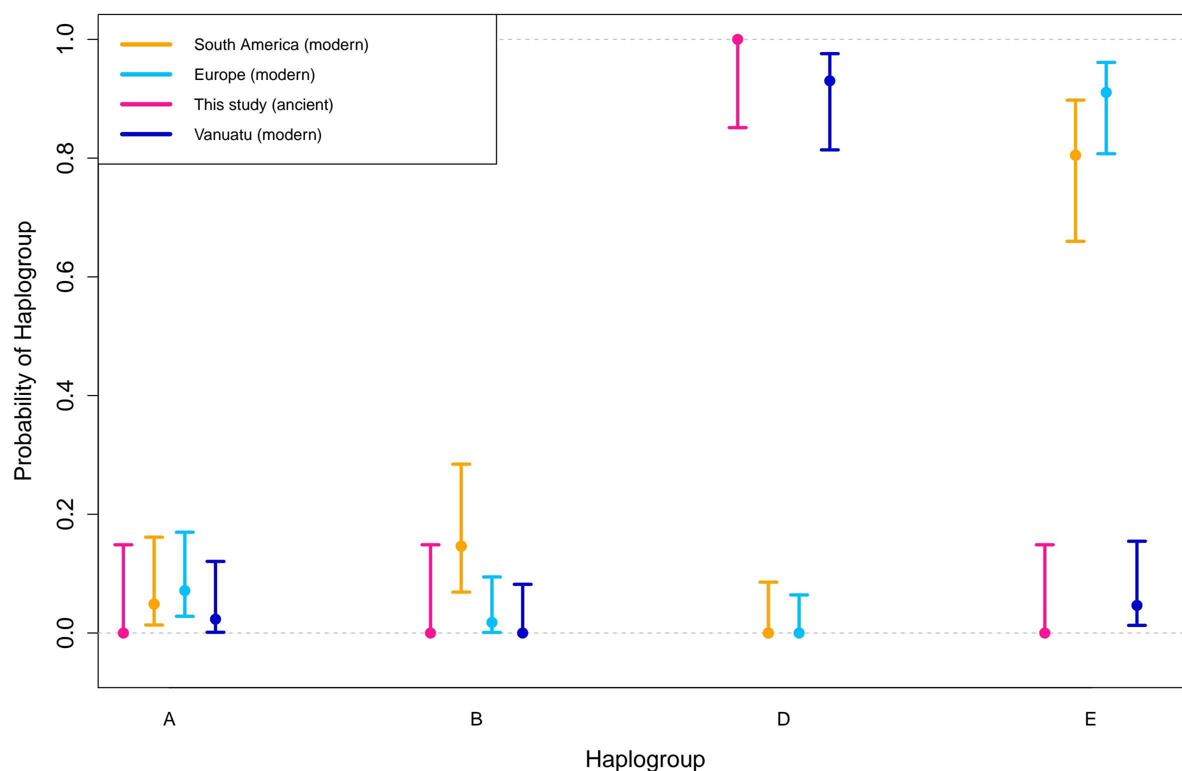


Figure S5. Binomial probability distributions with 95% confidence intervals for haplogroups A, B, D, and E in modern chicken populations in South America (Gongora *et al.* 2008), ancient Pacific (this study), modern Vanuatu (this study; Dancause *et al.* 2011), and modern European samples (Liu *et al.* 2006). The modern South American chickens sampled to date match the signature of European modern chickens more closely than that of either ancient or modern Pacific chickens (see Fisher's exact test data in Table S5). Modern South American chicken haplogroup frequencies are not statistically different at the 5% level from modern European haplogroup frequencies overall (p-value = 0.059), but are significantly different from ancient Pacific (this study, p-value = 2.2 x 10-16; and Storey *et al.* 2007, p-value = 0.00043), and modern Pacific chickens (p-value = 7.51 x 10^{-16}). For example, if we assume that the modern South American chickens sampled by Gongora *et al.* (2008) accurately represent the total chicken populations in South America today, then the proportion of haplogroups A, D, and E in South America more closely reflects the proportions of the same haplogroups in modern European chickens than it does the modern Pacific (this study, Dancause *et al.* 2011) or ancient Pacific chickens of our study (Table S6), which can be seen in this figure, where the 95% confidence intervals of the South American and European modern chickens overlap for haplogroups D and E. The frequency of haplogroup B in South America (15%) is considered significantly different from that in Europe (2%), but the fact haplogroup B is present in detectable levels in both South American and European chicken populations lends support for this link, as haplogroup B has not be found at all in either ancient or modern Pacific chickens.

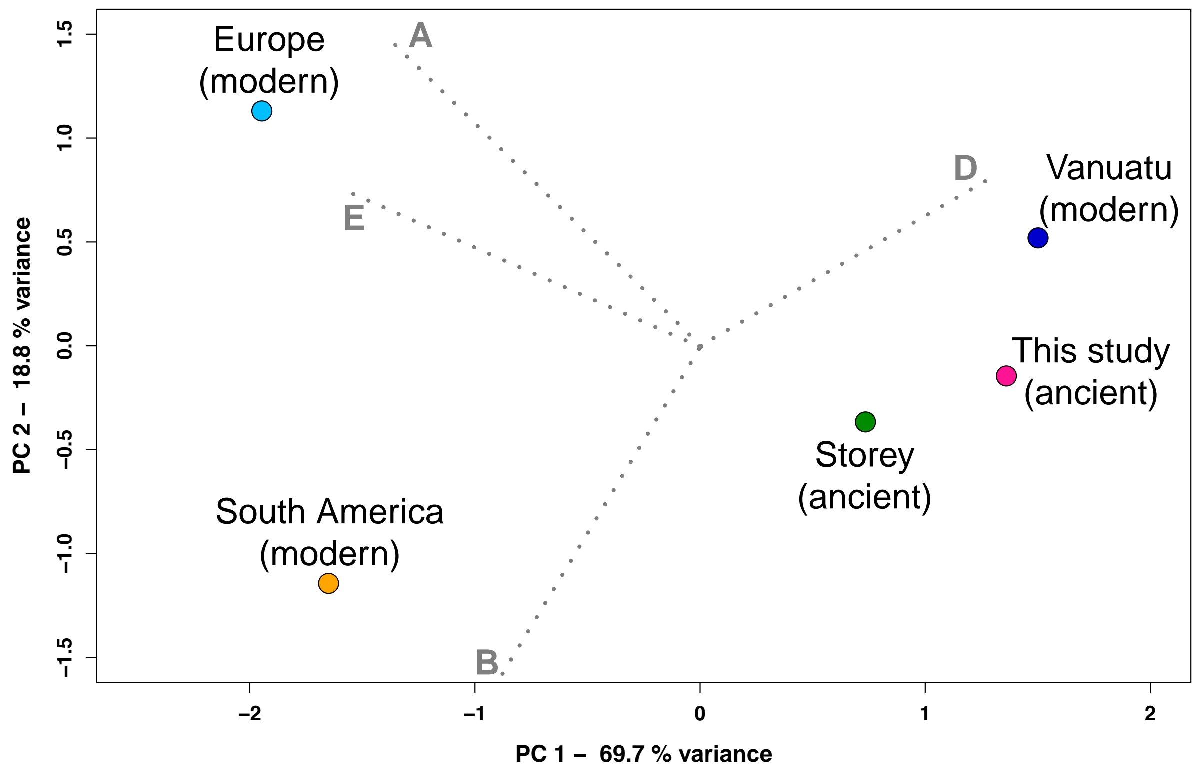


Figure S6. Results of principal component analysis (PCA) of haplogroup frequencies by population/location. Principal component 1 (PC1) accounts for almost 70% of the variation, with this principal component separating the modern chicken populations of Europe/South America from those in the Pacific (both modern and ancient). This figure highlights the fact that South American and European chickens (on the left hand side) are very different from Pacific chickens (on the right hand side). This dichotomy is strongly impacted by the high proportion of D haplotypes (which are localized to the Asia-Pacific region) in the Pacific, in comparison to South American and European chickens, which are influenced by A, B and E haplotypes (which are found worldwide).

Table S2 – Location and dating information for the chicken samples successfully analysed in this study, plus Storey *et al.*'s (2007) and Dancause *et al.*'s (2011) Pacific samples.

	Island	This study (or noted if otherwise)	D clade from Figure S2 or Haplotype from Liu et al. (2006) (n)	Samples analysed by Storey et al. (2007)	D clade from Figure S2 or Haplotype from Liu et al. (2006) (n)	Sample age	Site details	Genbank Accession No.
Ancient	Rapa Nui	ACAD9057/PAQANA009 (366bp)	D_{AP}	PAQANA009 (200bp)	D_{AP}	660-520 cal BP[&]	Unit 1, Layer IX, Level 10, Half: North, Anakena	EF535244
		ACAD9060/PAQANA007 (366bp)	E01				Unit 1, Layer VI, Level 6, Half: East, Anakena	
		ACAD9065/PAQANA003 (366bp)	D_{AP}				Unit 1, Layer VI, Level 6, Half: East, Anakena	
		ACAD9066/PAQANA013 (343bp)	D_{AP}				Unit 1, Layer VIII, Level 17, Half: South, Anakena	
		ACAD9067/PAQANA001 (366bp)	D_{AP}				Unit 1, Layer IX, Level 10, Half: North, Anakena	
		ACAD9068/PAQANA011 (366bp)	D_{AP}	PAQANA011 (185bp)	E^	680-550 cal BP[&]	Unit 1, Layer VIII, Level 18, Half: South, Anakena	EF535246
		ACAD9069/PAQANA005 (366bp)	D_{AP}				Unit 1, Layer VI, Level 3, Half: West, Anakena	
		ACAD9070/PAQANA002 (366bp)	D_{AP}				Unit 1, Layer VIII, Level 17, Half: South, Anakena	
		ACAD9071/PAQANA004 (201bp)	D_{AP}	PAQANA004 (201bp)	D_{AP}	520-690 and 570-580 cal BP[&]; 630-740 and 560-610 cal BP[&]	Unit 1, Layer VII, Level 7, Half: East, Anakena	EF535242
		ACAD9072/RNAN10 (366bp)	D_{AP}				Unit 1, Layer VIII, Level 18, Half: South, Anakena	
		ACAD9073/RNAN11 (357bp)	D_{AP}				Unit 1, Layer VIII, Level 16, Half: South, Anakena	
		ACAD9074/RNAN12 (366bp)	D_{AP}				Unit 1, Layer VIII, Level 16, Half: South, Anakena	
		ACAD9075/RNAN07 (366bp)	D_{AP}				Unit 5, Layer V, Level 5, Half: East, Anakena	
		ACAD9076/RNAN06 (366bp)	D_{AP}				Unit 5, Layer II, Level 2, Half: East, Anakena	
		ACAD9063	-	PAQANA006 (201bp)	D_{AP}	670-550 cal BP[&]	Unit 1, Layer V, Level 2, Half: West, Anakena	EF535243
		ACAD9064	-	PAQANA010 (185bp)	D_{AP}	670-550 & 700-540 cal BP[&]	Unit 1, Layer VII, Level 15, Half: South, Anakena	EF535245
		-	-	PAQHAN001 (193bp)	D_{AP}	Prehistoric[&]	Hanga Hahave, Rapa Nui	EF535247

Ancient	Hawaii	ACAD8136/HWAN08 (366bp)	D_{AP}				BAC-NW, HH75, 290-300, Makauwahi Cave, Kauai	
		ACAD8668/HWAN01 (300bp)	D_{AP}				BAC-NW, GG75, deep slump 2.5-4m below datum, Makauwahi Cave, Kauai	
		ACAD8670/HWAN02 (343bp)	D_{AP}				BAC-NW, GG74, 310mm below datum, Makauwahi Cave, Kauai	
		ACAD8671/HWAN03 (343bp)	D_{AP}				BAC-NW, GG75, deep slump 2.5-4m below datum, Makauwahi Cave, Kauai	
		ACAD8672/HWAN04 (343bp)	D_{AP}			534-478 cal BP	BAC-NW, GG76, 300mm below datum, Makauwahi Cave, Kauai	
		ACAD8674/HWAN05 (343bp)	D_{AP}				BAC-NW, GG76, 300mm below datum, Makauwahi Cave, Kauai	
		ACAD8675/HWAN07 (343bp)	D_{AP}				BAC-NW, GG76, 340mm below datum, Makauwahi Cave, Kauai	
				HWIKUA001 (185bp)	E	910-670 cal BP[&]	Kualoa, Hawaii	EF535238
	Niue	ACAD3890/NUAN03 (102bp)	A34				Paluki, Niue	
		ACAD3895/NUAN02 (366bp)	D_{AP}				Anatoloa, Niue	
		ACAD3896/NUAN01 (357bp)	D_{AP}			1285-1167 cal BP	Anatoloa, Niue	
				NIUPKI009 (185bp)	E	1590-1410 cal BP[&]	Paluki, Niue	EF535239
	Tonga			Tonga TD (167bp)	E	2000-1500 cal BP[&]	Ha'ateiho, Tonga	EF535236
				Tonga HB (167bp)	E	2000-1500 cal BP[&]	Mele Havea, Tonga	EF535237
	Samoa			AMSFTF001 (169bp)	E	1000-500 BP[&]	Fatu-ma-futi, American Samoa	EF535240

Modern	Hawaii	ACAD8993, 5 (353bp)	E01 (2)			Modern	Kokee, Kauai	
		ACAD8992, 4, 6, 7 (353bp)	E06 (4)			Modern	Kokee, Kauai	
		ACAD8989-91, 98 (353bp)	E* (4)			Modern	Kokee, Kauai	
	Marquesas	ACAD9030-31, 34 (353bp)	D_{AP} (3)			Modern	Marquesas	
		ACAD9029, 9033 (353bp)	E01 (2)			Modern	Marquesas	
	Guam	Dancause *et al.* (19)	D_{AP} (1)			Modern	Guam	FJ914363
		Dancause *et al.* (19)	D_{WP} (2)			Modern	Guam	FJ914362, 64
		Dancause *et al.* (19)	E01 (1)			Modern	Guam	FJ914366
		Dancause *et al.* (19)	E03 (1)			Modern	Guam	FJ914365
	Vanuatu	Dancause *et al.* (19)	D_{AP} (29)			Modern	Efate, Tanna, Ambae, Aneityum	FJ914319,20,22-3,25-35,38,40-3,46,48,50-1,54-6,58-9
		Dancause *et al.* (19)	D_{WP} (10)			Modern	Efate, Ambae, Aneityum	FJ914321,24,36,39,44-5,47,52-53,57
		Dancause *et al.* (19)	D01 (1)			Modern	Tanna	FJ914349
		Dancause *et al.* (19)	E* (1)			Modern	Efate	FJ914318
		Dancause *et al.* (19)	E* (1)			Modern	Efate	FJ914317
		Dancause *et al.* (19)	A34 (1)			Modern	Efate	FJ914337

* - falls within particular clade, but no exact match at haplotype level from Liu *et al.* (2006).

^ - haplotype could match many within particular clade from Liu *et al.* (2006).

[&] - Radiocarbon date is from Storey *et al.* (2007).

Table S3 - Cloning results of ACAD internal replication and external replication.

Base pair positions according to Liu *et al.*[6]	128	136	178	186	193	200	204	217	222	230	235	242	257*	267	271	281	292	303	323	324	350	358	372	386	394	399	407	432	449	463	476	478	479	480
CONSENSUS	T	C	T	C	C	A	C	C	T	C	A	G	T	C	T	A	T	G	C	C	C	C	G	C	C	A	C	C	C	C	C	C	C	G
ACAD9068A_shrimpA1_(562/1781)	T	C	T	C	C	A	C	C	T	C	A	G	T	C																				
ACAD9068A_shrimpA2_(1780/1781)									T	C	A	G																						
ACAD9068A_shrimpA3_(1780/1781)									T	C	A	G																						
ACAD9068A_(562/563)_SANGER_SEQ	T	T	T	C	T	A	C	C	T	T	A	G	T	C	T	A	T	G	C	C														
ACAD9068A_(562/563)_CLONE21	T	T	T	C	T	A	C	C	T	T	A	G	T	C	T	A	T	G	C	C														
ACAD9068A_(562/563)_CLONE30	T	T	T	C	T	A	C	C	T	T	A	G	T	C	T	A	T	G	C	C														
ACAD9068A_(562/563)_CLONE13	T	T	T	C	T	A	C	C	T	T	A	G	T	C	T	A	T	G	C	C														
ACAD9068A_(562/563)_CLONE15	T	T	T	C	T	A	C	C	T	T	A	G	T	C	T	A	T	G	C	C														
ACAD9068A_(562/563)_CLONE18	T	T	T	C	T	A	C	C	T	T	A	G	T	C	T	A	T	G	C	C														
ACAD9068A_(562/563)_CLONE24	T	T	T	C	T	A	C	C	T	T	A	G	T	C	T	A	T	G	C	C														
ACAD9068A_(562/563)_CLONE6	T	T	T	C	T	A	C	C	T	T	A	G	T	C	T	A	T	G	C	C														
ACAD9068A_(562/563)_CLONE14	T	T	T	C	T	A	C	C	T	T	A	G	T	C	T	A	T	G	C	C														
ACAD9068A_(562/563)_CLONE16	T	T	T	C	T	A	C	C	T	T	A	G	T	C	T	A	T	G	C	C														
ACAD9068A_(562/563)_CLONE18	T	T	T	C	T	A	C	C	T	T	A	G	T	C	T	A	T	G	C	C														
ACAD9068A_(562/563)_CLONE19	T	T	T	C	T	A	C	C	T	T	A	G	T	C	T	A	T	G	C	C														
ACAD9068A_(562/563)_CLONE22	T	T	T	C	T	A	C	C	T	T	A	G	T	C	T	A	T	G	C	C														
ACAD9068A_(562/563)_CLONE23	T	T	T	C	T	A	C	C	T	T	A	G	T	C	T	A	T	G	C	C														

ACAD9068A_(564/565)_SANGER_SEQ																			***C***	***C***	***C***	***C***	***G***	***T***	***C***	***A***	***C***	***C***	***C***	***T***	***T***	***C***	***C***	***G***
ACAD9068A_(564/565)_CLONE1																			C	C	C	C	G	***T***	C	A	C	C	C	***T***	***T***	C	C	G
ACAD9068A_(564/565)_CLONE10																			C	C	C	C	G	***T***	C	A	C	C	C	***T***	***T***	C	C	G
ACAD9068A_(564/565)_CLONE12																			C	C	C	C	G	***T***	C	A	C	C	C	***T***	***T***	C	C	G
ACAD9068A_(564/565)_CLONE13																			C	C	C	C	G	***T***	C	A	C	C	C	***T***	***T***	C	C	G
ACAD9068A_(564/565)_CLONE14																			C	C	C	C	G	***T***	C	A	C	C	C	***T***	***T***	C	C	G
ACAD9068A_(564/565)_CLONE16																			C	C	C	C	G	***T***	C	A	C	C	C	***T***	***T***	C	C	G
ACAD9068A_(564/565)_CLONE4																			C	C	C	C	G	***T***	C	A	C	C	C	***T***	***T***	C	C	G
ACAD9068A_(564/565)_CLONE7																			C	C	C	C	G	***T***	C	A	C	C	C	***T***	***T***	C	C	G
ACAD9068A_(564/565)_CLONE9																			C	C	C	C	G	***T***	C	A	C	C	C	***T***	***T***	C	C	G
ACAD9068A_(564/565)_CLONE3																			C	C	C	C	G	***T***	C	A	C	C	C	***T***	***T***	C	C	***A***
ACAD9068A_(564/565)_CLONE15																			C	C	C	C	***A***	***T***	C	A	C	C	C	***T***	***T***	C	C	G
ACAD9068A_(564/565)_CLONE8																			C	C	C	C	G	***T***	C	***G***	C	C	C	***T***	***T***	C	C	G
ACAD9068A_(562/1781)_SANGER_SEQ	***T***	***C***	***T***	***C***	***C***	***A***	***C***	***C***	***T***	***C***	***A***	***R***	***T***	***C***																				
ACAD9068A_(562/1781)_CLONE10 (E)	T	C	***C***	C	C	A	C	C	T	C	A	***A***	***C***	***T***																				
ACAD9068A_(562/1781)_CLONE7 (E)	T	C	***C***	C	C	A	C	C	T	C	A	***A***	***C***	***T***																				
ACAD9068A_(562/1781)_CLONE5 (E?)	T	C	T	C	C	A	C	C	T	C	A	***A***	***C***	***T***																				
ACAD9068A_(562/1781)_CLONE15	T	C	***C***	C	C	A	C	C	T	C	***G***	G	T	C																				
ACAD9068A_(562/1781)_CLONE1	T	C	T	C	C	A	C	C	T	C	A	G	T	C																				
ACAD9068A_(562/1781)_CLONE8	T	C	T	C	C	A	C	C	T	C	A	G	T	C																				
ACAD9068A_(562/1781)_CLONE9	T	C	T	C	C	A	C	C	T	C	A	G	T	C																				

ACAD9068A_(562/1781)_CLONE14	T	C	***C***	C	C	A	C	C	T	C	A	G	T	C																				
ACAD9068A_(562/1781)_CLONE16	T	C	***C***	C	C	A	C	C	T	C	A	G	T	C																				
ACAD9068A_(562/1781)_CLONE2	T	C	***C***	C	C	A	C	C	T	C	A	G	T	C																				
ACAD9068A_(1780/563)_SANGER_SEQ									***T***	***T***	***A***	***G***	***T***	***C***	***T***	***A***	***T***	***G***	***C***	***C***														
ACAD9068A_(1780/563)_CLONE1									T	C	A	G	T	C	T	A	T	G	C	C														
ACAD9068A_(1780/563)_CLONE10									T	C	A	G	T	C	T	A	T	G	C	C														
ACAD9068A_(1780/563)_CLONE11									T	C	A	G	T	C	T	A	T	G	C	C														
ACAD9068A_(1780/563)_CLONE12									T	C	A	G	T	C	T	A	T	G	C	C														
ACAD9068A_(1780/563)_CLONE13									T	C	A	G	T	C	T	A	T	G	C	C														
ACAD9068A_(1780/563)_CLONE14									T	C	A	G	T	C	T	A	T	G	C	C														
ACAD9068A_(1780/563)_CLONE15									T	C	A	G	T	C	T	A	T	G	C	C														
ACAD9068A_(1780/563)_CLONE16									T	C	A	G	T	C	T	A	T	G	C	C														
ACAD9068A_(1780/563)_CLONE2									T	C	A	G	T	C	T	A	T	G	C	C														
ACAD9068A_(1780/563)_CLONE3									T	C	A	G	T	C	T	A	T	G	C	C														
ACAD9068A_(1780/563)_CLONE4									T	C	A	G	T	C	T	A	T	G	C	C														
ACAD9068A_(1780/563)_CLONE6									T	C	A	G	T	C	T	A	T	G	C	C														
ACAD9068A_(1780/563)_CLONE8									T	C	A	G	T	C	T	A	T	G	C	C														
ACAD9068A_(1780/563)_CLONE9									T	C	A	G	T	C	T	***G***	T	G	C	C														
External_replication_FM009_G05/G06	***T***	***C***	***T***	***C***	***C***	***A***	***C***	***C***	***T***	***C***	***A***	***G***	***T***	***C***	***T***	***A***	***T***	***G***	***C***	***C***														
External_replication_SE009_D06/D12	***T***	***C***	***T***	***C***	***C***	***A***	***C***	***C***	***T***	***C***	***A***	***G***	***T***	***C***	***T***	***A***	***T***	***G***	***C***	***C***														
EXTREP7_(562/563)_CLONE44_E10	T	C	T	C	C	A	C	C	T	C	A	G	T	C	T	A	T	G	C	C														

EXTREP7_(562/563)_CLONE49_B11	T	C	T	C	C	A	C	C	T	C	A	G	T	C	T	A	T	G	C	C														
EXTREP7_(562/563)_CLONE42_C10	T	C	T	C	C	A	C	C	T	C	A	G	T	C	T	A	T	G	C	C														
EXTREP7_(562/563)_CLONE47_H10	T	C	T	C	C	A	C	C	T	C	A	G	T	C	T	A	T	G	C	C														
EXTREP7_(562/563)_CLONE48_A11	T	C	T	C	C	A	C	C	T	C	A	G	T	C	T	A	T	G	C	C														
EXTREP7_(562/563)_CLONE46_G10	T	C	T	C	C	A	C	C	T	C	A	G	T	C	T	A	T	G	C	C														
EXTREP7_(562/563)_CLONE45_F10	T	C	T	C	C	A	C	C	T	C	A	G	T	C	T	A	T	G	C	C														
EXTREP7_(562/563)_CLONE43_D10	T	C	T	C	C	A	C	C	T	C	A	G	T	C	T	A	***C***	G	C	C														
EXTREP3_(1780/1781)_CLONE35_E09									T	C	A	G	T	C																				
EXTREP3_(1780/1781)_CLONE33_C09									T	C	A	G	T	C																				
EXTREP3_(1780/1781)_CLONE34_D09									T	C	A	G	T	C																				
EXTREP3_(1780/1781)_CLONE30_H08									T	C	A	G	T	C																				
EXTREP3_(1780/1781)_CLONE32_B09									T	C	A	G	T	C																				
EXTREP3_(1780/1781)_CLONE37_F09									T	C	A	G	T	C																				
EXTREP3_(1780/1781)_CLONE31_A09									T	C	A	G	T	C																				
A35_AY465994	*C*	*C*	*T*	*T*	*C*	*G*	*T*	*T*	*C*	*C*	*A*	*A*	*C*	*T*	*C*	*A*	*T*	*A*	*C*	*C*	*C*	*C*	*G*	*C*	*C*	*A*	*C*	*C*	*C*	*C*	*C*	*C*	*C*	*G*
External_replication_SE009_E06/E12	***C***	***C***	***T***	***T***	***C***	***G***	***T***	***T***	***C***	***C***	***A***	***A***	***C***	***T***	***C***	***A***	***T***	***A***	***C***	***C***														
EXTREP5_(562/563)_CLONE38_G09	C	C	T	T	C	G	T	T	C	C	A	A	C	T	C	A	T	A	C	C														
EXTREP5_(562/563)_CLONE39_H09	C	C	T	T	C	G	T	T	C	C	A	A	C	T	C	A	T	A	C	C														
EXTREP5_(562/563)_CLONE40_A10	C	C	T	T	C	G	T	T	C	C	A	A	C	T	C	A	T	A	C	C														
EXTREP5_(562/563)_CLONE41_B10	C	C	T	T	C	G	T	T	C	C	A	A	C	T	C	A	T	A	C	C														
External_replication_SE009_B06/B12																			***C***	***C***	***T***	***T***	***G***	***C***	***C***	***A***	***C***	***T***	***T***	***C***	***C***	***T***	***T***	***G***

EXTREP9_(564/565)_CLONE56_D11																			C	C	***T***	***T***	G	C	C	A	C	***T***	***T***	C	C	***T***	***T***	G
EXTREP9_(564/565)_CLONE53_C11																			C	C	***T***	***T***	G	C	C	A	C	***T***	***T***	C	C	***T***	***T***	G
External_replication_E05/E06																			***C***	***C***	***C***	***C***	***G***	***C***	***T***	***A***	***C***	***C***	***C***	***C***	***C***	***C***	***C***	***G***
External_replication_F05/F06																			***C***	***C***	***C***	***C***	***G***	***C***	***C***	***A***	***C***	***C***	***C***	***C***	***C***	***C***	***C***	***G***

Underlined – Consensus sequence from all the data

Bold italised – Sanger sequenced data

Table S4 - Replication of PAQANA011 - internal (extract ACAD9068 using repeated PCR and Sanger sequencing) and external (at Durham University), and compared to Storey *et al.*'s (2007) EF535246 (PAQANA011) sequence

			130	138	180	188	195	202	206	219	224	232	244	259	269	273	305	352	360	388	396	434	451	465	478	480	481	Mismatch from consensus (bp)	Seq. length
CONSENSUS	Lab	Clade	T	C	T	C	C	A	C	C	T	C	G	T	C	T	G	C	C	C	C	C	C	C	C	C	C		
ACAD9068_562_1781_5OCT2010	ACAD	D	T	C	T	C	C	A	C	C	T	C	A	T	C	-	-	-	-	-	-	-	-	-	-	-	-	1	141bp
ACAD9068_562_1781_3Nov2010	ACAD	D	T	C	T	C	C	A	C	C	T	C	G	T	C	-	-	-	-	-	-	-	-	-	-	-	-	-	141bp
ACAD9068A_562_563_061001	ACAD	D	T	T	T	C	T	A	C	C	T	T	G	T	C	T	G	-	-	-	-	-	-	-	-	-	-	3	201bp
ACAD9068A1_562_1781(D)	ACAD	D	T	C	T	C	C	A	C	C	T	C	G	T	C	-	-	-	-	-	-	-	-	-	-	-	-	-	141bp
ACAD9068A3_A1031(D)	ACAD	D	-	-	-	-	-	-	-	-	T	C	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	32bp
ACAD9068A2_A1031(D)	ACAD	D	-	-	-	-	-	-	-	-	T	C	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	33bp
ACAD9068_1780_563_5OCT_2010	ACAD	D	-	-	-	-	-	-	-	-	T	C	G	T	C	T	G	-	-	-	-	-	-	-	-	-	-	-	97bp
ACAD9068A_564_565_V051002	ACAD	D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	C	C	C	C	C	C	C	C	C	C	1	217bp
ACAD9068_564_565_5OCT_2010	ACAD	D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	C	C	C	C	C	C	C	C	C	C	1	217bp
ACAD9068A_564_565_051001	ACAD	D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	C	C	T	C	C	C	T	T	C	C	4	195bp
ACAD9068_564_565_rep1_April11	ACAD	D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	C	C	C	C	C	C	C	-	134bp
ACAD9068_564_565_rep2_April11	ACAD	D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	C	C	C	C	C	C	C	C	-	142bp
EXTERNAL REPLICATION																													
Linus_G_562_563	Durham	D	T	C	T	C	C	A	C	C	T	C	G	T	C	T	G	-	-	-	-	-	-	-	-	-	-	-	201bp

SE009-D_Contig_edited	Durham	D	T	C	T	C	C	A	C	C	T	C	G	T	C	T	G	-	-	-	-	-	-	-	-	-	-	-	201bp
SE009-B_Contig_edited	Durham	D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	T	C	C	T	T	C	C	T	T	6	189bp
Linus_E_564_565	Durham	D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	C	C	T	C	C	C	C	C	C	1	189bp
Linus_F_564_565	Durham	D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	C	C	C	C	C	C	C	C	C	-	189bp
CONTAMINATION																													
SE009-E_Contig_edited	Durham	A35	C	C	T	T	C	G	T	T	C	C	A	C	T	C	A	-	-	-	-	-	-	-	-	-	-	11	201bp
EF535246 (PAQANA011)	Storey	E	-	-	C	C	C	A	C	C	T	C	A	C	T	T	A	-	-	-	-	-	-	-	-	-	-	5	192bp

Variable sites only are shown, and numbered according to Liu *et al.*'s (2006) data.

Shaded cells – mismatches to the consensus sequence for PAQANA011 analysed in this study

Dark shaded cells – type 2 transitions (C=>T or G=>A).

Table S5. Fisher exact test to detect significance of deviation from null hypothesis of equal haplogroup proportions between chicken populations (alternate hypothesis is not-equal).

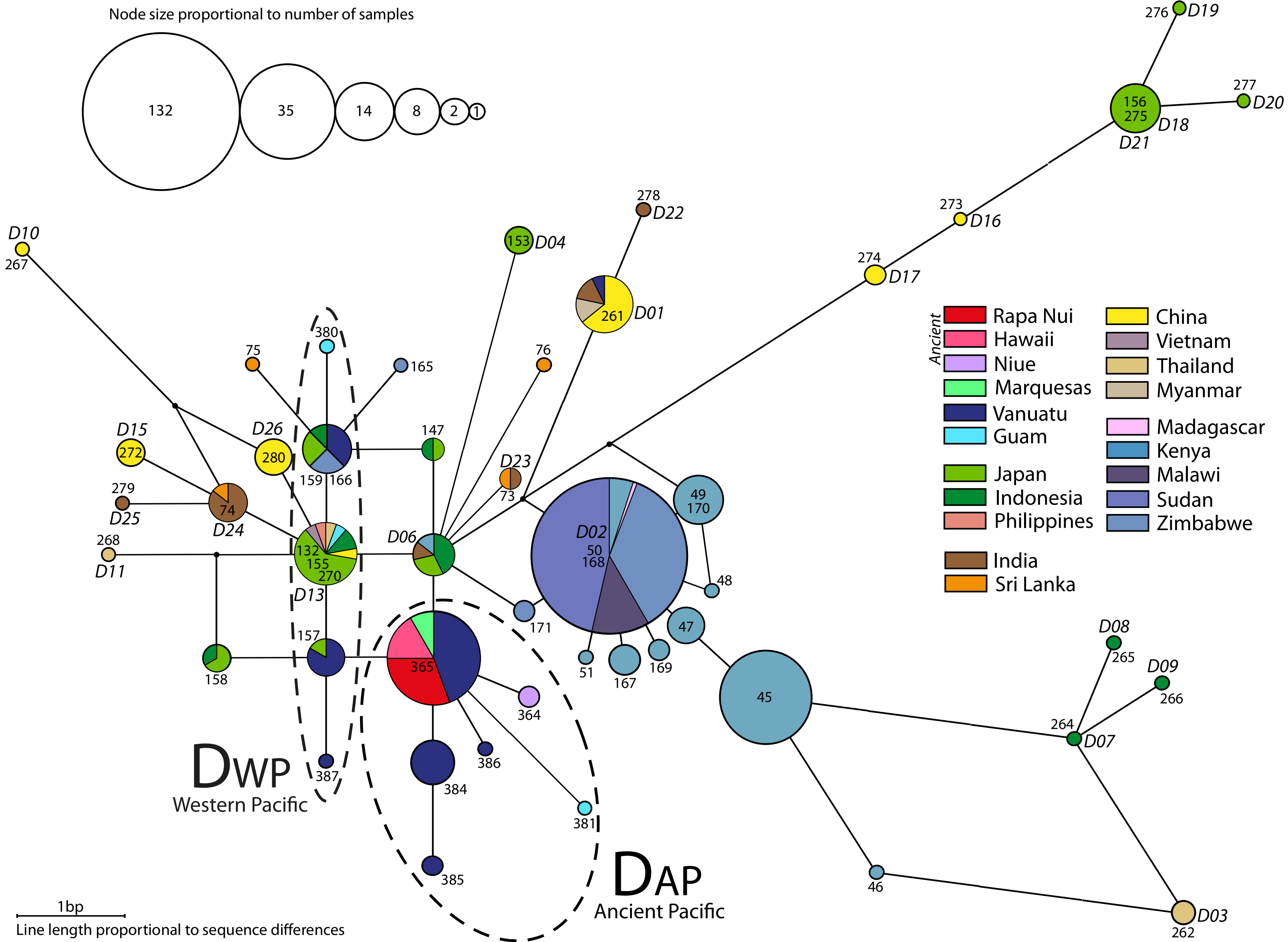
Chicken populations tested for equal proportions of haplogroups (A, B, D, E)		
Population 1	***Population 2***	***P-value***
Modern Chilean (Gongora et al. 2008)	Modern European (Liu et al. 2006)	0.05855
Modern Chilean	Modern Pacific (incl. this study)	7.5×10^{-16}*
Modern Chilean	Ancient Pacific (this study)	2.2×10^{-16}*
Modern Chilean	Ancient Pacific (Storey et al. 2007)	0.00043*
Efate (Dancause et al. 2011)	Other Vanuatu islands (Dancause et al. 2011)	0.0295*
Ancient Pacific (Storey et al. 2007)	Ancient Pacific (this study)	0.00042*
Chicken populations tested for equal proportions of sub-haplogroups (D_{AP}, D_{WP}, D01, E, A)		
Population 1	***Population 2***	***P-value***
Modern Pacific (this study, Dancause et al. 2011)	Ancient Pacific (this study)	0.00150*
Chicken populations tested for equal proportions of individual D haplotypes		
Population 1	***Population 2***	***P-value***
Modern Guam (Dancause et al. 2011)	Ancient Pacific (this study)	0.00049*
Modern Marquesas (this study)	Ancient Pacific (this study)	0.08058
Modern Vanuatu (Dancause et al. 2011)	Ancient Pacific (this study)	0.00677*
Modern Guam (Dancause et al. 2011)	Modern Vanuatu (Dancause et al. 2011)	0.00097*
Modern Marquesas (this study)	Modern Guam (Dancause et al. 2011)	0.1
Modern Vanuatu (Dancause et al. 2011)	Modern Marquesas (this study)	0.8753

* P-value < 0.05 indicates significant difference between haplogroup frequencies between population 1 and population 2.

Table S6. Binomial test of statistical significance for deviations in the frequency of each haplogroup in possible source populations compared to the expected distributions of each haplogroup in South American chickens observed in Gongora *et al.* (2008). The possible source populations tested include modern European chickens (Liu *et al.* 2006), ancient Pacific chickens (this study), and modern Pacific chickens (this study, Dancause *et al.* 2011).

Haplogroup	Frequency in modern Chilean chickens (n=41) (Gongora et al. 2008)	Frequency in modern European chickens (n=56) (Liu et al. 2006)	Frequency in ancient Pacific chickens (n=22) (this study)	Frequency in modern Pacific chickens (n=63) (this study, Dancause et al. 2011)
A	0.05	0.07	0.00*	0.02
B	0.15	0.02*	0.00*	0.00*
D	0	0	1.00*	0.73*
E	0.8	0.91	0.00*	0.25*

* P-value < 0.01 indicates significant deviation between a particular haplogroup frequency in possible source chicken population compared to that expected from the frequency of that haplotype observed in modern Chilean chickens.



Node size proportional to number of samples
132
35
14
8
2
1
Ancient
Rapa Nui
Hawaii
Niue
Marquesas
Vanuatu
Guam
Japan
Indonesia
Philippines
India
Sri Lanka
China
Vietnam
Thailand
Myanmar
Madagascar
Kenya
Malawi
Sudan
Zimbabwe
DWP
Western Pacific
DAP
Ancient Pacific
1bp
Line length proportional to sequence differences

Figure S7. Unrooted haplotype network generated using 330bp of mitochondrial control region for Polynesian chickens analysed in this study and other haplotype D sequences in Genbank (Liu *et al.* 2006; Oka *et al.* 2007; Muchadeyi *et al.* 2008; Silva *et al.* 2009; Berthouly-Salazar *et al.* 2010; Dancause *et al.* 2011; Mwacharo *et al.* 2011). Labels in larger font next to node circles are haplotype designations according to Liu *et al.* (Liu *et al.* 2006), while labels in smaller font next to (or within) node circles are haplotype numbers in this study (Dataset S1). Colors reflect sampling location, with sub-clade D_{AP} and grouping D_{WP} circled with dashed lines.

Table S7. Unique haplotypes based on 903 sequences of 330bp fragment of mitochondrial control region used in this study, prepared using Collapse v1.2.

Unique Haplotype No.	Haplogroup according to Liu *et al*'s (2006) designations	No. of samples per haplotype [n]	Genbank Accession No.	Code	Reference	Location	Ancient or Modern
1	**A**	**[1]**	**FJ851658**	**GIRIRAJA57**	**Adebambo *et al.* (2010)**	**Nigeria**	**Modern**
2	**E**	**[1]**	**FJ851679**	**SW32**	**Adebambo *et al.* (2010)**	**Nigeria**	**Modern**
3	**E**	**[1]**	**FJ851656**	**YAFFA113**	**Adebambo *et al.* (2010)**	**Nigeria**	**Modern**
4	**E**	**[1]**	**GU951752**	**NW59**	**Adebambo *et al.* (2010)**	**Nigeria**	**Modern**
5	**E**	**[2]**	**FJ851657**	**ANAK109**	**Adebambo *et al.* (2010)**	**Nigeria**	**Modern**
5	E		GU951755	SE8	Adebambo *et al.* (2010)	Nigeria	Modern
6	**E**	**[1]**	**FJ851678**	**SW30**	**Adebambo *et al.* (2010)**	**Nigeria**	**Modern**
7	**E**	**[3]**	**FJ851666**	**NE84**	**Adebambo *et al.* (2010)**	**Nigeria**	**Modern**
7	E		FJ851682	SW59	Adebambo *et al.* (2010)	Nigeria	Modern
7	E		FJ851684	SW62	Adebambo *et al.* (2010)	Nigeria	Modern
8	**E**	**[1]**	**FJ851675**	**SW15**	**Adebambo *et al.* (2010)**	**Nigeria**	**Modern**
9	**E**	**[1]**	**FJ851660**	**NE62**	**Adebambo *et al.* (2010)**	**Nigeria**	**Modern**
10	**E**	**[1]**	**FJ851662**	**NE73**	**Adebambo *et al.* (2010)**	**Nigeria**	**Modern**
11	**E**	**[1]**	**FJ851663**	**NE74**	**Adebambo *et al.* (2010)**	**Nigeria**	**Modern**
12	**E**	**[1]**	**FJ851664**	**NE79**	**Adebambo *et al.* (2010)**	**Nigeria**	**Modern**
13	**E**	**[1]**	**FJ851665**	**NE81**	**Adebambo *et al.* (2010)**	**Nigeria**	**Modern**
14		**[1]**	**FJ851667**	**NW7**	**Adebambo *et al.* (2010)**	**Nigeria**	**Modern**
15	**E**	**[1]**	**FJ851668**	**NW9**	**Adebambo *et al.* (2010)**	**Nigeria**	**Modern**
16	**E**	**[1]**	**FJ851669**	**NW18**	**Adebambo *et al.* (2010)**	**Nigeria**	**Modern**
17	**E**	**[1]**	**FJ851670**	**NW25**	**Adebambo *et al.* (2010)**	**Nigeria**	**Modern**
18	**E**	**[1]**	**FJ851671**	**NW35**	**Adebambo *et al.* (2010)**	**Nigeria**	**Modern**
19	**E**	**[1]**	**FJ851672**	**NW36**	**Adebambo *et al.* (2010)**	**Nigeria**	**Modern**
20	**E**	**[1]**	**FJ851673**	**SW10**	**Adebambo *et al.* (2010)**	**Nigeria**	**Modern**
21	**E**	**[1]**	**FJ851674**	**SW12**	**Adebambo *et al.* (2010)**	**Nigeria**	**Modern**
22	**E**	**[1]**	**FJ851677**	**SW27**	**Adebambo *et al.* (2010)**	**Nigeria**	**Modern**
23	**E**	**[1]**	**FJ851680**	**SW49**	**Adebambo *et al.* (2010)**	**Nigeria**	**Modern**
24	**E**	**[1]**	**FJ851681**	**SW55**	**Adebambo *et al.* (2010)**	**Nigeria**	**Modern**
25	**E**	**[1]**	**FJ851683**	**SW60**	**Adebambo *et al.* (2010)**	**Nigeria**	**Modern**
26	**E**	**[1]**	**FJ851685**	**SW77**	**Adebambo *et al.* (2010)**	**Nigeria**	**Modern**

27	**E**	**[1]**	**FJ851686**	**SW82**	**Adebambo *et al.* (2010)**	**Nigeria**	**Modern**
28	**E**	**[1]**	**GU951751**	**NE96**	**Adebambo *et al.* (2010)**	**Nigeria**	**Modern**
29	**E**	**[1]**	**GU951753**	**SE2**	**Adebambo *et al.* (2010)**	**Nigeria**	**Modern**
30	**E**	**[2]**	**GU951754**	**SE5**	**Adebambo *et al.* (2010)**	**Nigeria**	**Modern**
30	E		FJ851676	SW20	Adebambo *et al.* (2010)	Nigeria	Modern
31	**E**	**[1]**	**GU951756**	**SE10**	**Adebambo *et al.* (2010)**	**Nigeria**	**Modern**
32	**E**	**[1]**	**GU951757**	**SE18**	**Adebambo *et al.* (2010)**	**Nigeria**	**Modern**
33	**E**	**[1]**	**GU951758**	**SE32**	**Adebambo *et al.* (2010)**	**Nigeria**	**Modern**
34	**E**	**[1]**	**FJ851659**	**NE51**	**Adebambo *et al.* (2010)**	**Nigeria**	**Modern**
35	**E04**	**[15]**	**AF128325**		**Liu *et al.* (2006)**	**China (7), Japan (2), Europe (2)**	**Modern**
35	E		EF190850	CHL24	Gongora *et al.* (2008)	Chile	Modern
35	E		EU847816		Kanginakudru *et al.* (2008)	India	Modern
35	E		EU095147		Mwacharo *et al.* (2011)	Kenya	Modern
36	**E**	**[1]**	**EU095169**		**Mwacharo *et al.* (2011)**	**Kenya**	**Modern**
37	**E**	**[1]**	**EU095149**		**Mwacharo *et al.* (2011)**	**Kenya**	**Modern**
38	**E01**	**[199]**	**AB007734**		**Liu *et al.* (2006)**	**China (19), India (9), Japan (11), Iran (3), Turkmenistan (3), UK (2), Europe (34), Unknown (1)**	**Modern**
38	E		FJ914366	GUAM101	Dancause *et al.* (2011)	Guam	Modern
38	E		EF190838	CHL11	Gongora *et al.* (2008)	Chile	Modern
38	E		EF190839	CHL12	Gongora *et al.* (2008)	Chile	Modern
38	E		EF190840	CHL13	Gongora *et al.* (2008)	Chile	Modern
38	E		EF190841	CHL14	Gongora *et al.* (2008)	Chile	Modern
38	E		EF190842	CHL15	Gongora *et al.* (2008)	Chile	Modern
38	E		EF190844	CHL18	Gongora *et al.* (2008)	Chile	Modern
38	E		EF190845	CHL19	Gongora *et al.* (2008)	Chile	Modern
38	E		EF190831	CHL2	Gongora *et al.* (2008)	Chile	Modern
38	E		EF190847	CHL21	Gongora *et al.* (2008)	Chile	Modern
38	E		EF190851	CHL25	Gongora *et al.* (2008)	Chile	Modern

38	E		EF190853	CHL27	Gongora *et al.* (2008)	Chile	Modern
38	E		EF190854	CHL28	Gongora *et al.* (2008)	Chile	Modern
38	E		EF190856	CHL30	Gongora *et al.* (2008)	Chile	Modern
38	E		EF190859	CHL33	Gongora *et al.* (2008)	Chile	Modern
38	E		EF190860	CHL34	Gongora *et al.* (2008)	Chile	Modern
38	E		EF190861	CHL35	Gongora *et al.* (2008)	Chile	Modern
38	E		EF190864	CHL38	Gongora *et al.* (2008)	Chile	Modern
38	E		EF190865	CHL39	Gongora *et al.* (2008)	Chile	Modern
38	E		EF190866	CHL40	Gongora *et al.* (2008)	Chile	Modern
38	E		EF190867	CHL41	Gongora *et al.* (2008)	Chile	Modern
38	E		EF190868	CHL42	Gongora *et al.* (2008)	Chile	Modern
38	E		EF190869	CHL43	Gongora *et al.* (2008)	Chile	Modern
38	E		EF190834	CHL7	Gongora *et al.* (2008)	Chile	Modern
38	E		EF190835	CHL8	Gongora *et al.* (2008)	Chile	Modern
38	E		EF190836	CHL9	Gongora *et al.* (2008)	Chile	Modern
38	E16		AY704713		Liu *et al.* (2006)	India (1)	Modern
38	E		EU095075	MG17	Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095081	MR14	Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095083	MR22	Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095089	MR43	Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095102		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095105		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095107		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095112		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095123		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095124		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095126		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095127		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095128		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095129		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095130		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095131		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095132		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095133		Mwacharo *et al.* (2011)	Kenya	Modern

38	E		EU095134		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095135		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095136		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095137		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095139		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095140		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095142		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095143		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095144		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095145		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095146		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095148		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095151		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095155		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095157		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095160		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095164		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095165		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095166		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095167		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095168		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095170		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095171		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095172		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095173		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095174		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095175		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095178		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095179		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095180		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095181		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095183		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095184		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095185		Mwacharo *et al.* (2011)	Kenya	Modern

38	E		EU095187		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095188		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095189		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095190		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095191		Mwacharo *et al.* (2011)	Kenya	Modern
38	E		EU095192	NV44	Mwacharo *et al.* (2011)	Kenya	Modern
38	E		AB268508		Oka *et al.* (2007)	Japan (10)	Modern
38	E		AB268510		Oka *et al.* (2007)	Japan (1)	Modern
38	E		AB268513		Oka *et al.* (2007)	Japan (5)	Modern
38	E		EU199906		Silva *et al.* (2009)	Sri Lanka (14)	Modern
38	E		EU199907		Silva *et al.* (2009)	Sri Lanka (1)	Modern
38	E		EU199908		Silva *et al.* (2009)	Sri Lanka (1)	Modern
38	E		EU199909		Silva *et al.* (2009)	Sri Lanka (2)	Modern
38	E		EU199911		Silva *et al.* (2009)	Sri Lanka (1)	Modern
38	E		EU199912		Silva *et al.* (2009)	Sri Lanka (1)	Modern
38	E			HWMD05	This study	Hawaii	Modern
38	E			HWMD07	This study	Hawaii	Modern
38	E			MQMD01	This study	Marquesas	Modern
38	E			MQMD03	This study	Marquesas	Modern
39	**E08**	**[20]**	**AY704696**		**Liu *et al.* (2006)**	**Azerbaijan (1), Japan (2)**	**Modern**
39	E14			E14	Liu *et al.* (2006)	Japan (1)	Modern
39	E		EU095035	KF04	Mwacharo *et al.* (2011)	Kenya	Modern
39	E		EU095039	KF19	Mwacharo *et al.* (2011)	Kenya	Modern
39	E		EU095040	KF21	Mwacharo *et al.* (2011)	Kenya	Modern
39	E		EU095041	KF24	Mwacharo *et al.* (2011)	Kenya	Modern
39	E		EU095042	KF27	Mwacharo *et al.* (2011)	Kenya	Modern
39	E		EU095046	CH01	Mwacharo *et al.* (2011)	Kenya	Modern
39	E		EU095048	FF01	Mwacharo *et al.* (2011)	Kenya	Modern
39	E		EU095049	TT02	Mwacharo *et al.* (2011)	Kenya	Modern
39	E		EU095050	TT05	Mwacharo *et al.* (2011)	Kenya	Modern
39	E		EU095058	FF05	Mwacharo *et al.* (2011)	Kenya	Modern
39	E		EU095117		Mwacharo *et al.* (2011)	Kenya	Modern

39	E		EU095121		Mwacharo *et al.* (2011)	Kenya	Modern
39	E		EU095152		Mwacharo *et al.* (2011)	Kenya	Modern
39	E		AB268511		Oka *et al.* (2007)	Japan (5)	Modern
40	**E**	**[2]**	**EU095154**		**Mwacharo *et al.* (2011)**	**Kenya**	**Modern**
40	E		EU095125		Mwacharo *et al.* (2011)	Kenya	Modern
41	**E**	**[2]**	**AY644978**		**Liu *et al.* (2006)**	**Europe (1)**	**Modern**
41	E		EU095101		Mwacharo *et al.* (2011)	Kenya	Modern
42	**E**	**[1]**	**EU095122**		**Mwacharo *et al.* (2011)**	**Kenya**	**Modern**
43	**E**	**[15]**	**EU095109**		**Mwacharo *et al.* (2011)**	**Kenya**	**Modern**
43	E		EU095043	KF29	Mwacharo *et al.* (2011)	Kenya	Modern
43	E		EU095057	PN15	Mwacharo *et al.* (2011)	Kenya	Modern
43	E		EU095062	TL04	Mwacharo *et al.* (2011)	Kenya	Modern
43	E		EU095063	MG01	Mwacharo *et al.* (2011)	Kenya	Modern
43	E		EU095064	MG02	Mwacharo *et al.* (2011)	Kenya	Modern
43	E		EU095074	MG16	Mwacharo *et al.* (2011)	Kenya	Modern
43	E		EU095085	MR31	Mwacharo *et al.* (2011)	Kenya	Modern
43	E		EU095091	MR50	Mwacharo *et al.* (2011)	Kenya	Modern
43	E		EU095093	KT08	Mwacharo *et al.* (2011)	Kenya	Modern
43	E		EU095095	KT15	Mwacharo *et al.* (2011)	Kenya	Modern
43	E		EU095138		Mwacharo *et al.* (2011)	Kenya	Modern
43	E		EU095150		Mwacharo *et al.* (2011)	Kenya	Modern
43	E		EU095162		Mwacharo *et al.* (2011)	Kenya	Modern
43	E		EU095176		Mwacharo *et al.* (2011)	Kenya	Modern
44	**E**	**[1]**	**EU095059**	**CH07**	**Mwacharo *et al.* (2011)**	**Kenya**	**Modern**
45	**D**	**[35]**	**EU095092**	**KT02**	**Mwacharo *et al.* (2011)**	**Kenya**	**Modern**
45	D		EU095051	TT11	Mwacharo *et al.* (2011)	Kenya	Modern
45	D		EU095052	TT14	Mwacharo *et al.* (2011)	Kenya	Modern
45	D		EU095056	PN13	Mwacharo *et al.* (2011)	Kenya	Modern
45	D		EU095065	MG03	Mwacharo *et al.* (2011)	Kenya	Modern
45	D		EU095067	MG05	Mwacharo *et al.* (2011)	Kenya	Modern
45	D		EU095068	MG06	Mwacharo *et al.* (2011)	Kenya	Modern
45	D		EU095069	MG08	Mwacharo *et al.* (2011)	Kenya	Modern
45	D		EU095070	MG10	Mwacharo *et al.* (2011)	Kenya	Modern
45	D		EU095071	MG11	Mwacharo *et al.* (2011)	Kenya	Modern

45	D		EU095072	MG12	Mwacharo *et al.* (2011)	Kenya	Modern
45	D		EU095073	MG14	Mwacharo *et al.* (2011)	Kenya	Modern
45	D		EU095076	MG18	Mwacharo *et al.* (2011)	Kenya	Modern
45	D		EU095077	MG20	Mwacharo *et al.* (2011)	Kenya	Modern
45	D		EU095078	MR01	Mwacharo *et al.* (2011)	Kenya	Modern
45	D		EU095079	MR09	Mwacharo *et al.* (2011)	Kenya	Modern
45	D		EU095080	MR11	Mwacharo *et al.* (2011)	Kenya	Modern
45	D		EU095084	MR27	Mwacharo *et al.* (2011)	Kenya	Modern
45	D		EU095086	MR34	Mwacharo *et al.* (2011)	Kenya	Modern
45	D		EU095087	MR37	Mwacharo *et al.* (2011)	Kenya	Modern
45	D		EU095088	MR40	Mwacharo *et al.* (2011)	Kenya	Modern
45	D		EU095090	MR46	Mwacharo *et al.* (2011)	Kenya	Modern
45	D		EU095096	KT17	Mwacharo *et al.* (2011)	Kenya	Modern
45	D		EU095097	KT20	Mwacharo *et al.* (2011)	Kenya	Modern
45	D		EU095099	KT26	Mwacharo *et al.* (2011)	Kenya	Modern
45	D		EU095100		Mwacharo *et al.* (2011)	Kenya	Modern
45	D		EU095103		Mwacharo *et al.* (2011)	Kenya	Modern
45	D		EU095104		Mwacharo *et al.* (2011)	Kenya	Modern
45	D		EU095106		Mwacharo *et al.* (2011)	Kenya	Modern
45	D		EU095108		Mwacharo *et al.* (2011)	Kenya	Modern
45	D		EU095115		Mwacharo *et al.* (2011)	Kenya	Modern
45	D		EU095116		Mwacharo *et al.* (2011)	Kenya	Modern
45	D		EU095119		Mwacharo *et al.* (2011)	Kenya	Modern
45	D		EU095182		Mwacharo *et al.* (2011)	Kenya	Modern
45	D		EU095186		Mwacharo *et al.* (2011)	Kenya	Modern
46	**D**	**[1]**	**EU095094**	**KT12**	**Mwacharo *et al.* (2011)**	**Kenya**	**Modern**
47	**D**	**[5]**	**EU095111**		**Mwacharo *et al.* (2011)**	**Kenya**	**Modern**
47	D		EU095082	MR17	Mwacharo *et al.* (2011)	Kenya	Modern
47	D		EU095113		Mwacharo *et al.* (2011)	Kenya	Modern
47	D		EU095114		Mwacharo *et al.* (2011)	Kenya	Modern
47	D		EU095118		Mwacharo *et al.* (2011)	Kenya	Modern
48	**D**	**[1]**	**EU095037**	**KF13**	**Mwacharo *et al.* (2011)**	**Kenya**	**Modern**
49	**D**	**[1]**	**EU095054**	**TT21**	**Mwacharo *et al.* (2011)**	**Kenya**	**Modern**
50	**D02**	**[8]**	**AB007742**		**Liu *et al.* (2006)**	**Madagascar (1),**	**Modern**

						Unknown (2)	
50	D		EU095034	KF01	Mwacharo *et al.* (2011)	Kenya	Modern
50	D		EU095044	PN04	Mwacharo *et al.* (2011)	Kenya	Modern
50	D		EU095047	DC05	Mwacharo *et al.* (2011)	Kenya	Modern
50	D		EU095055	PN10	Mwacharo *et al.* (2011)	Kenya	Modern
50	D		EU095098	KT23	Mwacharo *et al.* (2011)	Kenya	Modern
50	D		EU095110		Mwacharo *et al.* (2011)	Kenya	Modern
50	D		EU095156		Mwacharo *et al.* (2011)	Kenya	Modern
51	**D**	**[1]**	**EU095045**	**PN09**	**Mwacharo *et al.* (2011)**	**Kenya**	**Modern**
52	**D06**	**[4]**	**AB009429**		**Liu *et al.* (2006)**	**Indonesia (2), India (1), Japan (1)**	**Modern**
52	D		EU095053	TT18	Mwacharo *et al.* (2011)	Kenya	Modern
52	D		AB268525		Oka *et al.* (2007)	Japan (1), Indonesian Bantam	Modern
53	**E**	**[1]**	**EU095163**		**Mwacharo *et al.* (2011)**	**Kenya**	**Modern**
54	**E**	**[1]**	**EU095159**		**Mwacharo *et al.* (2011)**	**Kenya**	**Modern**
55	**E**	**[4]**	**EU095161**		**Mwacharo *et al.* (2011)**	**Kenya**	**Modern**
55	E		EU095141		Mwacharo *et al.* (2011)	Kenya	Modern
55	E		EU095158		Mwacharo *et al.* (2011)	Kenya	Modern
55	E		EU095177		Mwacharo *et al.* (2011)	Kenya	Modern
56	**E**	**[1]**	**EU095153**		**Mwacharo *et al.* (2011)**	**Kenya**	**Modern**
57	**E**	**[3]**	**EU095038**	**KF17**	**Mwacharo *et al.* (2011)**	**Kenya**	**Modern**
57	E		EU095061	TL02	Mwacharo *et al.* (2011)	Kenya	Modern
57	E		EU095120		Mwacharo *et al.* (2011)	Kenya	Modern
58	**E**	**[1]**	**EU095036**	**KF09**	**Mwacharo *et al.* (2011)**	**Kenya**	**Modern**
59	**E11**	**[8]**	**AY645014**		**Liu *et al.* (2006)**	**Europe (4)**	**Modern**
59	E		EF190858	CHL32	Gongora *et al.* (2008)	Chile	Modern
59	E		EU095066	MG04	Mwacharo *et al.* (2011)	Kenya	Modern
59	E		AB268506		Oka *et al.* (2007)	Japan (4)	Modern
59	E		EU199922		Silva *et al.* (2009)	Sri Lanka (1)	Modern
60	**E03**	**[4]**	**AY704728**		**Liu *et al.* (2006)**	**China (1), Turkmenistan (1)**	**Modern**

60	E		FJ914365	GUAM004	Dancause *et al.* (2011)	Guam	Modern
60	E		EU095060	DC10	Mwacharo *et al.* (2011)	Kenya	Modern
60	E		AB268515		Oka *et al.* (2007)	Japan (3)	Modern
60	E		EU199921		Silva *et al.* (2009)	Sri Lanka (3)	Modern
61	**C01**	**[32]**	**AB007743**		**Liu *et al.* (2006)**	**China (19), Japan (8)**	**Modern**
61	C16		AB098690		Liu *et al.* (2006)	Japan (1)	Modern
61	C		AB268533		Oka *et al.* (2007)	Japan (1)	Modern
61	C		AB268534		Oka *et al.* (2007)	Japan (6)	Modern
61	C		AB268536		Oka *et al.* (2007)	Japan (12)	Modern
61	C		GQ227995		Zhou *et al.* (2010)	China (1)	Modern
62	**E**	**[19]**	**GQ227998**		**Zhou *et al.* (2010)**	**China (1)**	**Modern**
62	E		EF190849	CHL23	Gongora *et al.* (2008)	Chile (1)	Modern
62	E		EF190852	CHL26	Gongora *et al.* (2008)	Chile (1)	Modern
62	E		AB268509		Oka *et al.* (2007)	Japan (5)	Modern
62	E		AB268514		Oka *et al.* (2007)	Japan (2)	Modern
62	E		EU199913		Silva *et al.* (2009)	Sri Lanka (2)	Modern
62	E		EU199914		Silva *et al.* (2009)	Sri Lanka (1)	Modern
62	E		GQ228010		Zhou *et al.* (2010)	China (1)	Modern
62	E		GQ228012		Zhou *et al.* (2010)	China (1)	Modern
62	E			HWMD04	This study	Hawaii (1)	Modern
62	E			HWMD06	This study	Hawaii (1)	Modern
62	E			HWMD08	This study	Hawaii (1)	Modern
62	E			HWMD09	This study	Hawaii (1)	Modern
63	**B01**	**[137]**	**AB009442**		**Liu *et al.* (2006)**	**China (97), Indonesia (3), Laos (1), Thailand (1), India (1), Japan (8), Iran (1), Turkmenistan (1), Europe (1), Unknown (3)**	**Modern**
63	B		EF190830	CHL1	Gongora *et al.* (2008)	Chile (1)	Modern

63	B		EF190843	CHL17	Gongora *et al.* (2008)	Chile (1)	Modern
63	B		EF190862	CHL36	Gongora *et al.* (2008)	Chile (1)	Modern
63	B		EF190863	CHL37	Gongora *et al.* (2008)	Chile (1)	Modern
63	B		EF190833	CHL6	Gongora *et al.* (2008)	Chile (1)	Modern
63	B		AB268539		Oka *et al.* (2007)	Japan (4)	Modern
63	B		AB268540		Oka *et al.* (2007)	Japan (1)	Modern
63	B		AB268541		Oka *et al.* (2007)	Japan (2)	Modern
63	B		AB268542		Oka *et al.* (2007)	Japan (1)	Modern
63	B		EU199938		Silva *et al.* (2009)	Sri Lanka (4)	Modern
63	B		EU199943		Silva *et al.* (2009)	Sri Lanka (1)	Modern
63	B		GQ228000		Zhou *et al.* (2010)	China (1)	Modern
63	B		GQ228007		Zhou *et al.* (2010)	China (1)	Modern
64	**A01**	**[163]**	**AB007749**		**Liu *et al.* (2006)**	**China (93), Laos (1), Japan (15), Iran (1), Europe (2), Unknown (1)**	**Modern**
64	A		FJ914337	VAN21	Dancause *et al.* (2011)	Vanuatu (1)	Modern
64	A		FJ914361	THAI02	Dancause *et al.* (2011)	Thailand (1)	Modern
64	A28		AB114073		Liu *et al.* (2006)	Japan (1)	Modern
64	A31		AB114084		Liu *et al.* (2006)	Japan (1)	Modern
64	A24		AY465960		Liu *et al.* (2006)	China (16)	Modern
64	A		AB268516		Oka *et al.* (2007)	Japan (12)	Modern
64	A		AB268517		Oka *et al.* (2007)	Japan (7)	Modern
64	A		AB294233		Oka *et al.* (2007)	Japan (2)	Modern
64	A		EU199941		Silva *et al.* (2009)	Sri Lanka (8)	Modern
64	A		GQ228011		Zhou *et al.* (2010)	China (1)	Modern
65	**E**	**[4]**	**EF190848**	**CHL22**	**Gongora *et al.* (2008)**	**Chile (1)**	**Modern**
65	E		EF190837	CHL10	Gongora *et al.* (2008)	Chile (1)	Modern
65	E		EF190846	CHL20	Gongora *et al.* (2008)	Chile (1)	Modern
65	E		EF190855	CHL29	Gongora *et al.* (2008)	Chile (1)	Modern
66	**A02**	**[15]**	**AB007758**		**Liu *et al.* (2006)**	**China (1), Japan (1), Iran (1), Europe (1)**	**Modern**

66	A		EF190857	CHL31	Gongora *et al.* (2008)	Chile (1)	Modern
66	A		EF190832	CHL5	Gongora *et al.* (2008)	Chile (1)	Modern
66	A		EU847801		Kanginakudru *et al.* (2008)	India (1)	Modern
66	A		AB268521		Oka *et al.* (2007)	Japan (3)	Modern
66	A		EU199940		Silva *et al.* (2009)	Sri Lanka (7)	Modern
66	A		X52392		Desjardins *et al.* (1990)	-	Modern
67	**B**	**[1]**	**EF190870**	**CHL44**	**Gongora *et al.* (2008)**	**Chile**	**Modern**
68	**A**	**[1]**	**EF535249**		**Gongora *et al.* (2008)**	**Chile**	**Modern**
69	**E**	**[1]**	**EF535248**		**Gongora *et al.* (2008)**	**Chile**	**Modern**
70	**B**	**[14]**	**EU199937**		**Silva *et al.* (2009)**	**Sri Lanka (17)**	**Modern**
71	**B**	**[1]**	**EU199939**		**Silva *et al.* (2009)**	**Sri Lanka**	**Modern**
72	**C06**	**[15]**	**AB009446**		**Liu *et al.* (2006)**	**China (1), Japan (10)**	**Modern**
72	C		AB268531		Oka *et al.* (2007)	Japan (1)	Modern
72	C		AB268532		Oka *et al.* (2007)	Japan (1)	Modern
72	C		EU199947		Silva *et al.* (2009)	Sri Lanka (2)	Modern
73	**D23**	**[2]**	**AY704709**		**Liu *et al.* (2006)**	**India (1)**	**Modern**
73	D		EU199945		Silva *et al.* (2009)	Sri Lanka (1)	Modern
74	**D24**	**[4]**	**AY704717**		**Liu *et al.* (2006)**	**India (5)**	**Modern**
74	D		EU199946		Silva *et al.* (2009)	Sri Lanka (1)	Modern
75	**D**	**[1]**	**EU199942**		**Silva *et al.* (2009)**	**Sri Lanka**	**Modern**
76	**D**	**[1]**	**EU199944**		**Silva *et al.* (2009)**	**Sri Lanka**	**Modern**
77	**E09**	**[11]**	**AY392336**		**Liu *et al.* (2006)**	**Europe (4), China (3)**	**Modern**
77	E		AB268507		Oka *et al.* (2007)	Japan (2)	Modern
77	E		EU199923		Silva *et al.* (2009)	Sri Lanka (2)	Modern
78	**E17**	**[4]**	**AY704708**		**Liu *et al.* (2006)**	**India (1)**	**Modern**
78	E		EU847813		Kanginakudru *et al.* (2008)	India (1)	Modern
78	E		EU847815		Kanginakudru *et al.* (2008)	India (1)	Modern
78	E		EU199916		Silva *et al.* (2009)	Sri Lanka (2)	Modern
78	E		EU199917		Silva *et al.* (2009)	Sri Lanka (1)	Modern
79	**E**	**[1]**	**EU199910**		**Silva *et al.* (2009)**	**Sri Lanka (1)**	**Modern**
80	**E**	**[1]**	**EU199915**		**Silva *et al.* (2009)**	**Sri Lanka**	**Modern**

						(1)	
81	**E**	**[3]**	**EU199918**		**Silva *et al.* (2009)**	**Sri Lanka (2)**	**Modern**
81	E		EU199919		Silva *et al.* (2009)	Sri Lanka (1)	Modern
82	**E**	**[1]**	**EU199920**		**Silva *et al.* (2009)**	**Sri Lanka**	**Modern**
83	**E**	**[1]**	**EU199924**		**Silva *et al.* (2009)**	**Sri Lanka**	**Modern**
84	**E**	**[2]**	**EU199925**		**Silva *et al.* (2009)**	**Sri Lanka**	**Modern**
85		**[1]**	**EU199926**		**Silva *et al.* (2009)**	**Sri Lanka**	**Modern**
86	**G05**	**[10]**	**EU199933**		**Liu *et al.* (2006)**	**China (1)**	**Modern**
86	G		EU199931		Silva *et al.* (2009)	Sri Lanka (8)	Modern
86	G		EU199933		Silva *et al.* (2009)	Sri Lanka (1)	Modern
87	**G09**	**[3]**		**G09**	**Liu *et al.* (2006)**	**China (1)**	**Modern**
87	G		EU199936		Silva *et al.* (2009)	Sri Lanka (2)	Modern
88	**G**	**[2]**	**EU199927**		**Silva *et al.* (2009)**	**Sri Lanka (1)**	**Modern**
88	G		EU199928		Silva *et al.* (2009)	Sri Lanka (1)	Modern
89	**G**	**[19]**	**EU199929**		**Silva *et al.* (2009)**	**Sri Lanka (17)**	**Modern**
89	G		EU199930		Silva *et al.* (2009)	Sri Lanka (1)	Modern
89	G		EU199934		Silva *et al.* (2009)	Sri Lanka (1)	Modern
90	**G**	**[1]**	**EU199932**		**Silva *et al.* (2009)**	**Sri Lanka**	**Modern**
91	**G**	**[10]**	**EU199935**		**Silva *et al.* (2009)**	**Sri Lanka (10)**	**Modern**
92	**G.g. bankiva**	**[1]**	**CHKMTB03**		**Fumihito *et al.* (1996)**		**Modern**
93	**G.g. bankiva**	**[1]**	**CHKMTB02**		**Fumihito *et al.* (1996)**		**Modern**
94	**G.g. bankiva**	**[1]**	**CHKMTB01**		**Fumihito *et al.* (1996)**		**Modern**
95	**G.g. bankiva**	**[2]**	**AP003323**		**Nishibori *et al.* (2005)**		**Modern**
95	G.g. bankiva		HM462178		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
96	**G.g. bankiva**	**[1]**	**AB007718**				**Modern**
97	**I**	**[1]**	**HM462202**		**Berthouly-Salazar et al. (2010)**	**Vietnam**	**Modern**

98	**I**	**[2]**	**HM462203**		**Berthouly-Salazar et al. (2010)**	**Vietnam (1)**	**Modern**
98	I		HM462205		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
99	**I**	**[1]**	**HM462204**		**Berthouly-Salazar et al. (2010)**	**Vietnam**	**Modern**
100	**A**	**[12]**	**HM462082**		**Berthouly-Salazar et al. (2010)**	**Vietnam (1)**	**Modern**
100	A		HM462087		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
100	A		HM462089		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
100	A		HM462093		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
100	A		HM462105		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
100	A		HM462107		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
100	A		HM462146		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
100	A		HM462156		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
100	A		HM462158		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
100	A		HM462163		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
100	A		HM462173		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
100	A		HM462176		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
101	**A**	**[1]**	**HM462201**		**Berthouly-Salazar et al. (2010)**	**Vietnam**	**Modern**
102	**A**	**[1]**	**HM462194**		**Berthouly-Salazar et al. (2010)**	**Vietnam**	**Modern**
103	**A**	**[3]**	**HM462196**		**Berthouly-Salazar et al. (2010)**	**Vietnam (1)**	**Modern**
103	A		HM462197		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
103	A		HM462198		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
104	**A**	**[1]**	**HM462199**		**Berthouly-Salazar et al. (2010)**	**Vietnam**	**Modern**
105	**A**	**[1]**	**HM462200**		**Berthouly-Salazar et al. (2010)**	**Vietnam**	**Modern**
106	**A**	**[4]**	**HM462098**		**Berthouly-Salazar et al. (2010)**	**Vietnam (1)**	**Modern**
106	A		HM462101		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
106	A		HM462151		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
106	A		HM462174		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
107	**A**	**[1]**	**HM462097**		**Berthouly-Salazar et al. (2010)**	**Vietnam**	**Modern**
108	**A**	**[1]**	**HM462195**		**Berthouly-Salazar et al. (2010)**	**Vietnam**	**Modern**
109	**A**	**[1]**	**HM462123**		**Berthouly-Salazar et al. (2010)**	**Vietnam**	**Modern**
110	**A**	**[3]**	**HM462124**		Berthouly-Salazar et al. (2010)	**Vietnam (1)**	**Modern**
110	A		HM462133		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
110	A		HM462183		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
111	**A**	**[1]**	**HM462140**		**Berthouly-Salazar et al. (2010)**	**Vietnam**	**Modern**
112	**B**	**[35]**	**HM462084**		**Berthouly-Salazar et al. (2010)**	**Vietnam (1)**	**Modern**

112	B		HM462086		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
112	B		HM462091		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
112	B		HM462092		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
112	B		HM462094		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
112	B		HM462096		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
112	B		HM462099		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
112	B		HM462104		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
112	B		HM462106		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
112	B		HM462117		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
112	B		HM462125		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
112	B		HM462126		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
112	B		HM462127		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
112	B		HM462128		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
112	B		HM462129		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
112	B		HM462130		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
112	B		HM462132		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
112	B		HM462138		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
112	B		HM462141		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
112	B		HM462145		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
112	B		HM462147		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
112	B		HM462148		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
112	B		HM462153		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
112	B		HM462155		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
112	B		HM462159		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
112	B		HM462160		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
112	B		HM462161		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
112	B		HM462167		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
112	B		HM462169		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
112	B		HM462170		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
112	B		HM462179		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
112	B		HM462184		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
112	B		HM462185		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
112	B		HM462186		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
112	B		HM462187		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern

113	**B**	**[1]**	**HM462209**		**Berthouly-Salazar et al. (2010)**	**Vietnam**	**Modern**
114	**B**	**[2]**	**HM462212**		**Berthouly-Salazar et al. (2010)**	**Vietnam (1)**	**Modern**
114	B		HM462217		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
115	**B**	**[1]**	**HM462213**		**Berthouly-Salazar et al. (2010)**	**Vietnam**	**Modern**
116	**B**	**[1]**	**HM462214**		**Berthouly-Salazar et al. (2010)**	**Vietnam**	**Modern**
117	**B**	**[1]**	**HM462215**		**Berthouly-Salazar et al. (2010)**	**Vietnam**	**Modern**
118	**B**	**[1]**	**HM462216**		**Berthouly-Salazar et al. (2010)**	**Vietnam**	**Modern**
119	**B**	**[2]**	**HM462136**		**Berthouly-Salazar et al. (2010)**	**Vietnam (1)**	**Modern**
119	B		HM462118		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
120	**B**	**[1]**	**HM462143**		**Berthouly-Salazar et al. (2010)**	**Vietnam**	**Modern**
121	**B**	**[1]**	**HM462157**		**Berthouly-Salazar et al. (2010)**	**Vietnam**	**Modern**
122	**B**	**[8]**	**HM462108**		**Berthouly-Salazar et al. (2010)**	**Vietnam (1)**	**Modern**
122	B		HM462088		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
122	B		HM462131		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
122	B		HM462134		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
122	B		HM462137		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
122	B		HM462139		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
122	B		HM462152		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
122	B		HM462180		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
123	**B**	**[1]**	**HM462122**		**Berthouly-Salazar et al. (2010)**	**Vietnam**	**Modern**
124	**B**	**[1]**	**HM462085**		**Berthouly-Salazar et al. (2010)**	**Vietnam**	**Modern**
125	**B**	**[1]**	**HM462095**		**Berthouly-Salazar et al. (2010)**	**Vietnam**	**Modern**
126	**B**	**[2]**	**HM462100**		**Berthouly-Salazar et al. (2010)**	**Vietnam (1)**	**Modern**
126	B		HM462115		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
127	**B**	**[4]**	**HM462165**		**Berthouly-Salazar et al. (2010)**	**Vietnam (1)**	**Modern**
127	B		HM462166		**Berthouly-Salazar et al. (2010)**	Vietnam (1)	Modern
127	B		HM462172		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
127	B		HM462177		**Berthouly-Salazar et al. (2010)**	Vietnam (1)	Modern
128	**B**	**[1]**	**HM462168**		**Berthouly-Salazar et al. (2010)**	**Vietnam**	**Modern**
129	**B**	**[1]**	**HM462210**		**Berthouly-Salazar et al. (2010)**	**Vietnam**	**Modern**
130	**B**	**[1]**	**HM462211**		**Berthouly-Salazar et al. (2010)**	**Vietnam**	**Modern**
131	**C**	**[5]**	**HM462113**		**Berthouly-Salazar et al. (2010)**	**Vietnam (1)**	**Modern**
131	C		HM462116		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
131	C		HM462119		**Berthouly-Salazar et al. (2010)**	Vietnam (1)	Modern

131	C		HM462149		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
131	C		HM462171		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
132	**D**	**[1]**	**HM462103**		**Berthouly-Salazar et al. (2010)**	**Vietnam**	**Modern**
133	**F**	**[4]**	**HM462083**		**Berthouly-Salazar et al. (2010)**	**Vietnam (1)**	**Modern**
133	F		HM462102		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
133	F		HM462109		**Berthouly-Salazar et al. (2010)**	Vietnam (1)	Modern
133	F		HM462114		**Berthouly-Salazar et al. (2010)**	Vietnam (1)	Modern
133	F		HM462120		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
133	F		HM462121		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
134	**F**	**[4]**	**HM462090**		**Berthouly-Salazar et al. (2010)**	**Vietnam (1)**	**Modern**
134	F		HM462110		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
134	F		HM462111		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
134	F		HM462150		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
134	F		HM462154		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
134	F		HM462182		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
135	**F**	**[1]**	**HM462142**		**Berthouly-Salazar et al. (2010)**	**Vietnam**	**Modern**
136	**F**	**[1]**	**HM462162**		**Berthouly-Salazar et al. (2010)**	**Vietnam**	**Modern**
137	**G**	**[5]**	**HM462112**		**Berthouly-Salazar et al. (2010)**	**Vietnam (1)**	**Modern**
137	G		HM462135		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
137	G		HM462144		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
137	G		HM462164		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
137	G		HM462181		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
138	**G**	**[1]**	**HM462188**		**Berthouly-Salazar et al. (2010)**	**Vietnam**	**Modern**
139	**G**	**[1]**	**HM462189**		**Berthouly-Salazar et al. (2010)**	**Vietnam**	**Modern**
140	**G**	**[1]**	**HM462190**		**Berthouly-Salazar et al. (2010)**	**Vietnam**	**Modern**
141	**G**	**[2]**	**HM462191**		**Berthouly-Salazar et al. (2010)**	**Vietnam (1)**	**Modern**
141	G		HM462193		Berthouly-Salazar et al. (2010)	Vietnam (1)	Modern
142	**G**	**[1]**	**HM462192**		**Berthouly-Salazar et al. (2010)**	**Vietnam**	**Modern**
143	**G**	**[1]**	**HM462175**		**Berthouly-Salazar et al. (2010)**	**Vietnam**	**Modern**
144	**I**	**[1]**	**HM462206**		**Berthouly-Salazar et al. (2010)**	**Vietnam**	**Modern**
145	**I**	**[1]**	**HM462207**		**Berthouly-Salazar et al. (2010)**	**Vietnam**	**Modern**
146	**I**	**[1]**	**HM462208**		**Berthouly-Salazar et al. (2010)**	**Vietnam**	**Modern**
147	**D**	**[2]**	**AB268527**		**Oka *et al.* (2007)**	**Japan (1), Indonesian Fighting (1)**	**Modern**

148	**A05**	**[20]**	**AB009447**		**Liu *et al.* (2006)**	**Japan (5), Unknown (1)**	**Modern**
148	A		AB268518		Oka *et al.* (2007)	Japan (3)	Modern
148	A		AB268519		Oka *et al.* (2007)	Japan (11)	Modern
149	**A06**	**[8]**	**AB007736**		**Liu *et al.* (2006)**	**China (4), Japan (2)**	**Modern**
149	A		AB294232		Oka *et al.* (2007)	Japan (2)	Modern
150	**A35**	**[4]**	**AY465994**		**Liu *et al.* (2006)**	**China (2)**	**Modern**
150	A		AB268520		Oka *et al.* (2007)	Japan (4)	Modern
151	**C07**	**[27]**	**AB007745**		**Liu *et al.* (2006)**	**Japan (13)**	**Modern**
151	C		AB268535		Oka *et al.* (2007)	Japan (13)	Modern
151	C		AB268537		Oka *et al.* (2007)	Japan (1)	Modern
152	**C**	**[1]**	**AB268530**		**Oka *et al.* (2007)**	**Japan**	**Modern**
153	**D04**	**[3]**	**AB007741**		**Liu *et al.* (2006)**	**Japan (1)**	**Modern**
153	D		AB268529		Oka *et al.* (2007)	Japan (2)	Modern
154	**C**	**[1]**	**AB268538**		**Oka *et al.* (2007)**	**Japan**	**Modern**
155	**D**	**[12]**	**AB268522**		**Oka *et al.* (2007)**	**Japan (9)**	**Modern**
155	D		FJ914360	THAI01	Dancause *et al.* (2011)	Thailand (1)	Modern
155	D		FJ914364	GUAM003	Dancause *et al.* (2011)	Guam (1)	Modern
155	D		AB268523		Oka *et al.* (2007)	Japan (1)	Modern
156	**D**	**[2]**	**AB268543**		**Oka *et al.* (2007)**	**Japan (1)**	**Modern**
156	D		AB268544		Oka *et al.* (2007)	Japan (1)	Modern
157	**D**	**[4]**	**AB268524**		**Oka *et al.* (2007)**	**Japan (1)**	**Modern**
157	D		FJ914321	VAN05	Dancause *et al.* (2011)	Vanuatu (1)	Modern
157	D		FJ914324	VAN08	Dancause *et al.* (2011)	Vanuatu (1)	Modern
157	D		FJ914339	VAN23	Dancause *et al.* (2011)	Vanuatu (1)	Modern
157	D		FJ914345	VAN29	Dancause *et al.* (2011)	Vanuatu (1)	Modern
157	D		FJ914353	VAN37	Dancause *et al.* (2011)	Vanuatu (1)	Modern
158	**D**	**[3]**	**AB268526**		**Oka *et al.* (2007)**	**Japan (2), Indonesian Bantam (1)**	**Modern**
159	**D**	**[4]**	**AB268528**		**Oka *et al.* (2007)**	**Japan (2), Indonesian Fighting (1)**	**Modern**
159	D		FJ914336	VAN20	Dancause *et al.* (2011)	Vanuatu (1)	Modern

159	D		FJ914344	VAN28	Dancause *et al.* (2011)	Vanuatu (1)	Modern
159	D		FJ914352	VAN36	Dancause *et al.* (2011)	Vanuatu (1)	Modern
160	**E07**	**[13]**	**AB007738**		**Liu *et al.* (2006)**	**Japan (2)**	**Modern**
160	E		AB268512		Oka *et al.* (2007)	Japan (11)	Modern
161	**I**	**[1]**	**AB268545**		**Oka *et al.* (2007)**	**Japan**	**Modern**
162	**E**	**[2]**	**AM746050**	**C11**	**Muchadeyi *et al.* (2008)**	**Zimbabwe**	**Modern**
163	**E**	**[50]**	**AM746047**	**C8**	**Muchadeyi *et al.* (2008)**	**Zimbabwe**	**Modern**
164		**[2]**	**AM746055**	**C16**	**Muchadeyi *et al.* (2008)**	**Zimbabwe**	**Modern**
165	**D**	**[2]**	**AM746044**	**C5**	**Muchadeyi *et al.* (2008)**	**Zimbabwe**	**Modern**
166	**D**	**[2]**	**AM746027**	**A4**	**Muchadeyi *et al.* (2008)**	**Zimbabwe**	**Modern**
167	**D**	**[4]**	**AM746025**	**A2**	**Muchadeyi *et al.* (2008)**	**Zimbabwe**	**Modern**
168	**D**	**[124]**	**AM746024**	**A1**	**Muchadeyi *et al.* (2008)**	**Zimbabwe (46), Malawi (16), Sudan (62)**	**Modern**
169	**D**	**[2]**	**AM746026**	**A3**	**Muchadeyi *et al.* (2008)**	**Zimbabwe**	**Modern**
170	**D**	**[10]**	**AM746028**	**A5**	**Muchadeyi *et al.* (2008)**	**Zimbabwe**	**Modern**
171	**D**	**[2]**	**AM746030**	**A7**	**Muchadeyi *et al.* (2008)**	**Zimbabwe**	**Modern**
172	**E**	**[2]**	**AM746043**	**C4**	**Muchadeyi *et al.* (2008)**	**Zimbabwe**	**Modern**
173	**E**	**[3]**	**AM886306**	**C17**	**Muchadeyi *et al.* (2008)**	**Zimbabwe**	**Modern**
174	**E**	**[2]**	**AM746049**	**C10**	**Muchadeyi *et al.* (2008)**	**Zimbabwe**	**Modern**
175	**E**	**[2]**	**AM886307**	**C18**	**Muchadeyi *et al.* (2008)**	**Zimbabwe**	**Modern**
176	**E**	**[4]**	**AM746045**	**C6**	**Muchadeyi *et al.* (2008)**	**Zimbabwe**	**Modern**
177	**E**	**[2]**	**AM746046**	**C7**	**Muchadeyi *et al.* (2008)**	**Zimbabwe**	**Modern**
178	**E**	**[53]**	**AM746042**	**C3**	**Muchadeyi *et al.* (2008)**	**Zimbabwe (39), Malawi (9), Sudan (5)**	**Modern**
179	**E**	**[28]**	**AM746040**	**C1**	**Muchadeyi *et al.* (2008)**	**Zimbabwe**	**Modern**
180	**E**	**[2]**	**AM746054**	**C15**	**Muchadeyi *et al.* (2008)**	**Zimbabwe**	**Modern**
181	**E**	**[14]**	**AM746051**	**C12**	**Muchadeyi *et al.* (2008)**	**Zimbabwe**	**Modern**
182	**E**	**[2]**	**AM746048**	**C9**	**Muchadeyi *et al.* (2008)**	**Zimbabwe**	**Modern**
183	**E**	**[19]**	**AM746041**	**C2**	**Muchadeyi *et al.* (2008)**	**Zimbabwe (7), Malawi (8), Sudan (4)**	**Modern**

184	E	[8]	AM746053	C14	Muchadeyi *et al.* (2008)	Zimbabwe	Modern
185		[2]	AM746052	C13	Muchadeyi *et al.* (2008)	Zimbabwe	Modern
186		[2]	AM746029	A6	Muchadeyi *et al.* (2008)	Zimbabwe	Modern
187	A	[2]	AM746031	C19	Muchadeyi *et al.* (2008)	Zimbabwe	Modern
188	A	[2]	AM746032	C20	Muchadeyi *et al.* (2008)	Zimbabwe	Modern
189	A	[2]	AM746039	B7	Muchadeyi *et al.* (2008)	Zimbabwe	Modern
190	A	[21]	AM746035	B3	Muchadeyi *et al.* (2008)	Zimbabwe	Modern
191	A	[2]	AM746034	B2	Muchadeyi *et al.* (2008)	Zimbabwe	Modern
192	A	[36]	AM746033	B1	Muchadeyi *et al.* (2008)	Zimbabwe	Modern
193	A	[2]	AM746037	B5	Muchadeyi *et al.* (2008)	Zimbabwe	Modern
194	B	[42]	AM746036	B4	Muchadeyi *et al.* (2008)	Zimbabwe	Modern
195	A	[2]	AM746038	B6	Muchadeyi *et al.* (2008)	Zimbabwe	Modern
196	A03	[2]	AB007730		Liu *et al.* (2006)	Japan	Modern
197	A04	[1]	D82921		Liu *et al.* (2006)	Japan	Modern
198	A07	[1]	AF512201		Liu *et al.* (2006)	China	Modern
199	A08	[3]	AF512322		Liu *et al.* (2006)	China	Modern
200	A09	[1]	AF512151		Liu *et al.* (2006)	China	Modern
201	A10	[2]	AF512128		Liu *et al.* (2006)	China	Modern
202	A11	[4]	AF512074		Liu *et al.* (2006)	China	Modern
203	A12	[1]	AF512098		Liu *et al.* (2006)	China	Modern
204	A13	[10]	AF512328		Liu *et al.* (2006)	China	Modern
205	A14	[1]	AF512077		Liu *et al.* (2006)	China	Modern
206	A15	[9]	AF512149		Liu *et al.* (2006)	China	Modern
207	A16	[1]	AF512232		Liu *et al.* (2006)	China	Modern
208	A17	[9]	AF512103		Liu *et al.* (2006)	China	Modern
209	A18	[11]	AF512283		Liu *et al.* (2006)	China (10), Myanmar (1)	Modern
210	A19	[1]	AF512101		Liu *et al.* (2006)	China	Modern
211	A20	[2]	AY465981		Liu *et al.* (2006)	China	Modern
212	A21	[2]	AY392347		Liu *et al.* (2006)	China	Modern
213	A22	[1]	AY392248		Liu *et al.* (2006)	China	Modern
214	A232	[1]	AY392301		Liu *et al.* (2006)	China	Modern
215	A25	[1]	AB098666		Liu *et al.* (2006)	China	Modern
216	A26	[1]	AB098670		Liu *et al.* (2006)	Japan	Modern

217	A27	[1]	AY645000		Liu *et al.* (2006)	Europe	Modern
218	A29	[1]	AB114078		Liu *et al.* (2006)	Japan	Modern
219	A30	[1]	AB114081		Liu *et al.* (2006)	Japan	Modern
220	A32	[1]	AB114085		Liu *et al.* (2006)	Japan	Modern
221	A33	[1]	AY465977		Liu *et al.* (2006)	China	Modern
222	A34	[1]	AY588608		Liu *et al.* (2006)	China	Modern
223	A36	[1]	AY588626		Liu *et al.* (2006)	China	Modern
224	A37	[1]	AY588624		Liu *et al.* (2006)	China	Modern
225	B02	[1]	AB268517		Liu *et al.* (2006)	China	Modern
226	B03	[1]	AF512309		Liu *et al.* (2006)	China	Modern
227	B04	[1]	AF512245		Liu *et al.* (2006)	China	Modern
228	B05	[4]	AF512250		Liu *et al.* (2006)	China	Modern
229	B06	[1]	AF512085		Liu *et al.* (2006)	China	Modern
230	B07	[2]	AF512299		Liu *et al.* (2006)	China	Modern
231	B08	[4]	AF128322		Liu *et al.* (2006)	China	Modern
232	B09	[1]	AF128335		Liu *et al.* (2006)	China	Modern
233	B10	[13]	AF512154		Liu *et al.* (2006)	China (12), Myanmar (4)	Modern
234	B11	[27]	AF512254		Liu *et al.* (2006)	China (26), Myanmar (1)	Modern
235	B12	[1]	AF128342		Liu *et al.* (2006)	China	Modern
236	B13	[13]	AF512190		Liu *et al.* (2006)	China	Modern
237	B14	[2]	AY392362		Liu *et al.* (2006)	China	Modern
238	B15	[1]	AY392229		Liu *et al.* (2006)	China	Modern
239	B16	[1]	AB098665		Liu *et al.* (2006)	China	Modern
240	B17	[2]	AY588635		Liu *et al.* (2006)	China (1), Myanmar (1)	Modern
241	B18	[1]	AY465985		Liu *et al.* (2006)	China	Modern
242	B19	[2]	AY466000		Liu *et al.* (2006)	China	Modern
243	B20	[1]	AY588616		Liu *et al.* (2006)	China	Modern
244	B21	[1]	AY588629		Liu *et al.* (2006)	China	Modern
245	B22	[1]	AY588633		Liu *et al.* (2006)	China	Modern

246	**C02**	**[7]**	**AF512215**		**Liu *et al.* (2006)**	**China**	**Modern**
247	**C03**	**[4]**	**AF512255**		**Liu *et al.* (2006)**	**China**	**Modern**
248	**C04**	**[1]**	**AF512302**		**Liu *et al.* (2006)**	**China**	**Modern**
249	**C05**	**[3]**	**AF512264**		**Liu *et al.* (2006)**	**China**	**Modern**
250	**C08**	**[1]**	**AF512329**		**Liu *et al.* (2006)**	**China**	**Modern**
251	**C09**	**[1]**	**AB098680**		**Liu *et al.* (2006)**	**Japan**	**Modern**
252	**C10**	**[1]**	**AB098699**		**Liu *et al.* (2006)**	**Japan**	**Modern**
253	**C11**	**[1]**	**AB098697**		**Liu *et al.* (2006)**	**Japan**	**Modern**
254	**C12**	**[1]**	**AB098644**		**Liu *et al.* (2006)**	**Japan**	**Modern**
255	**C13**	**[1]**	**AB098649**		**Liu *et al.* (2006)**	**Japan**	**Modern**
256	**C14**	**[1]**	**AB098637**		**Liu *et al.* (2006)**	**Japan**	**Modern**
257	**C15**	**[1]**		**C15**	**Liu *et al.* (2006)**	**Japan**	**Modern**
258	**C17**	**[1]**		**C17**	**Liu *et al.* (2006)**	**Japan**	**Modern**
259	**C18**	**[1]**	**AB114072**		**Liu *et al.* (2006)**	**Japan**	**Modern**
260	**C19**	**[1]**	**AB114075**		**Liu *et al.* (2006)**	**Japan**	**Modern**
261	**D01**	**[14]**	**AY588636**		**Liu *et al.* (2006)**	**China (9), Myanmar (2), India (2)**	**Modern**
261	D		FJ914349	VAN33	Dancause *et al.* (2011)	Vanuatu	Modern
262	**D03**	**[3]**	**AB268532**		**Liu *et al.* (2006)**	**Thailand**	**Modern**
263	**D05**	**[1]**	**AB268534**		**Liu *et al.* (2006)**	**Unknown**	**Modern**
264	**D07**	**[1]**	**AB268536**		**Liu *et al.* (2006)**	**Indonesia**	**Modern**
265	**D08**	**[1]**	**AB268537**		**Liu *et al.* (2006)**	**Indonesia**	**Modern**
266	**D09**	**[1]**	**AB268538**		**Liu *et al.* (2006)**	**Indonesia**	**Modern**
267	**D10**	**[1]**	**AF128317**		**Liu *et al.* (2006)**	**China**	**Modern**
268	**D11**	**[1]**	**AB009432**		**Liu *et al.* (2006)**	**Thailand**	**Modern**
269	**D12**	**[1]**	**CHKMTG04**		**Liu *et al.* (2006)**	**Unknown**	**Modern**
270	**D13**	**[5]**	**D82916**		**Liu *et al.* (2006)**	**China (1), Indonesia (2), Philippines (1), Japan (1)**	**Modern**
271	**D14**	**[1]**	**D82919**		**Liu *et al.* (2006)**	**Indonesia**	**Modern**
272	**D15**	**[3]**	**AF512152**		**Liu *et al.* (2006)**	**China**	**Modern**

273	**D16**	**[1]**	**AY392185**		**Liu *et al.* (2006)**	**China**	**Modern**
274	**D17**	**[2]**	**AY392176**		**Liu *et al.* (2006)**	**China**	**Modern**
275	**D21**	**[8]**	**AB098646**		**Liu *et al.* (2006)**	**Japan (7)**	**Modern**
275	D18		AB098639		Liu *et al.* (2006)	Japan (1)	Modern
276	**D19**	**[1]**		**D19**	**Liu *et al.* (2006)**	**Japan**	**Modern**
277	**D20**	**[1]**		**D20**	**Liu *et al.* (2006)**	**Japan**	**Modern**
278	**D22**	**[1]**	**AY644971**		**Liu *et al.* (2006)**	**India**	**Modern**
279	**D25**	**[1]**	**AY704704**		**Liu *et al.* (2006)**	**India**	**Modern**
280	**D26**	**[4]**	**AY588637**		**Liu *et al.* (2006)**	**China**	**Modern**
281	**E02**	**[1]**	**AF512301**		**Liu *et al.* (2006)**	**China**	**Modern**
282	**E05**	**[4]**	**AF512141**		**Liu *et al.* (2006)**	**China (4)**	**Modern**
283	**E06**	**[20]**	**AY642134**		**Liu *et al.* (2006)**	**China (10), Indonesia (4), Malaysia (1), Japan (2), Europe (6)**	**Modern**
284	**E10**	**[1]**	**AY644973**		**Liu *et al.* (2006)**	**India**	**Modern**
285	**E13**	**[2]**	**AY644967**		**Liu *et al.* (2006)**	**India**	**Modern**
286	**E15**	**[1]**	**AY704712**		**Liu *et al.* (2006)**	**India**	**Modern**
287	**E18**	**[4]**	**AY588631**		**Liu *et al.* (2006)**	**China (2)**	**Modern**
287	E		EU847808		Kanginakudru *et al.* (2008)	India	Modern
287	E		EU847810		Kanginakudru *et al.* (2008)	India	Modern
288	**F01**	**[30]**	**AF512285**		**Liu *et al.* (2006)**	**China**	**Modern**
289	**F02**	**[1]**	**AF512161**		**Liu *et al.* (2006)**	**China**	**Modern**
290	**F03**	**[1]**	**AF512185**		**Liu *et al.* (2006)**	**Myanmar**	**Modern**
291	**F04**	**[15]**	**AY392262**		**Liu *et al.* (2006)**	**China (17), Myanmar (1)**	**Modern**
292	**F05**	**[1]**	**AF512153**		**Liu *et al.* (2006)**	**China**	**Modern**
293	**F06**	**[1]**	**AF512157**		**Liu *et al.* (2006)**	**China**	**Modern**
294	**F07**	**[2]**	**AF512057**		**Liu *et al.* (2006)**	**China**	**Modern**
295	**F08**	**[1]**	**CHKMTS 11**		**Liu *et al.* (2006)**	**Unknown**	**Modern**
296	**F09**	**[1]**	**AB007721**		**Liu *et al.* (2006)**	**Thailand (1), Unknown (3)**	**Modern**

297	**F10**	**[7]**	**AY392312**		**Liu *et al.* (2006)**	**China**	**Modern**
298	**F11**	**[1]**	**AY392328**		**Liu *et al.* (2006)**	**China**	**Modern**
299	**F12**	**[2]**	**AY392305**		**Liu *et al.* (2006)**	**China**	**Modern**
300	**F13**	**[2]**	**AY392311**		**Liu *et al.* (2006)**	**China**	**Modern**
301	**F14**	**[2]**	**AY392309**		**Liu *et al.* (2006)**	**China**	**Modern**
302	**F15**	**[1]**	**AY392354**		**Liu *et al.* (2006)**	**China**	**Modern**
303	**F16**	**[1]**	**AY392173**		**Liu *et al.* (2006)**	**China**	**Modern**
304	**F17**	**[1]**	**AY392314**		**Liu *et al.* (2006)**	**China**	**Modern**
305	**G01**	**[32]**	**AY392293**		**Liu *et al.* (2006)**	**China**	**Modern**
306	**G02**	**[2]**	**AF512112**		**Liu *et al.* (2006)**	**China**	**Modern**
307	**G03**	**[5]**	**AF512067**		**Liu *et al.* (2006)**	**China**	**Modern**
308	**G04**	**[2]**	**AF512292**		**Liu *et al.* (2006)**	**China**	**Modern**
309	**G06**	**[1]**	**AB009435**		**Liu *et al.* (2006)**	**Vietnam**	**Modern**
310	**G07**	**[3]**	**AY392225**		**Liu *et al.* (2006)**	**China**	**Modern**
311	**G08**	**[2]**	**AY392348**		**Liu *et al.* (2006)**	**China**	**Modern**
312	**G10**	**[4]**	**AY392279**		**Liu *et al.* (2006)**	**China**	**Modern**
313	**G11**	**[1]**	**AY392241**		**Liu *et al.* (2006)**	**China**	**Modern**
314	**G12**	**[2]**	**AY392251**		**Liu *et al.* (2006)**	**China**	**Modern**
315	**G13**	**[1]**	**AY392260**		**Liu *et al.* (2006)**	**China**	**Modern**
316	**G14**	**[2]**	**AY392216**		**Liu *et al.* (2006)**	**China**	**Modern**
317	**G15**	**[1]**	**AY392239**		**Liu *et al.* (2006)**	**China**	**Modern**
318	**G16**	**[1]**		**G16**	**Liu *et al.* (2006)**	**China**	**Modern**
319	**G17**	**[1]**	**AY392351**		**Liu *et al.* (2006)**	**China**	**Modern**
320	**G18**	**[1]**	**AY392282**		**Liu *et al.* (2006)**	**China**	**Modern**
321	**G19**	**[3]**	**AY392303**		**Liu *et al.* (2006)**	**China**	**Modern**
322	**G20**	**[1]**	**AY392290**		**Liu *et al.* (2006)**	**China**	**Modern**
323	**G21**	**[1]**	**AY392237**		**Liu *et al.* (2006)**	**China**	**Modern**
324	**G22**	**[1]**	**AY392335**		**Liu *et al.* (2006)**	**China**	**Modern**
325	**G23**	**[1]**	**AY465996**		**Liu *et al.* (2006)**	**China**	**Modern**
326	**G24**	**[1]**	**AY465997**		**Liu *et al.* (2006)**	**China**	**Modern**
327	**H01**	**[1]**	**AB009439**		**Liu *et al.* (2006)**	**Indonesia (1), Unknown (1)**	**Modern**
328	**H02**	**[2]**	**AB007720**		**Liu *et al.* (2006)**	**Unknown**	**Modern**
329	**H03**	**[1]**	**AB009440**		**Liu *et al.* (2006)**	**Indonesia**	**Modern**

						(1), Unknown (1)	
330	**I01**	**[1]**	**AB009434**		**Liu *et al.* (2006)**	**Vietnam**	**Modern**
331	**I02**	**[1]**	**AB007756**		**Liu *et al.* (2006)**	**Unknown**	**Modern**
332	**I03**	**[1]**	**AB009449**		**Liu *et al.* (2006)**	**Vietnam**	**Modern**
333	**G. g. murghi**	**[1]**	**EU847745**		**Kanginakudru *et al.* (2008)**	**India**	**Modern**
334	**G. g. murghi**	**[1]**	**EU847747**		**Kanginakudru *et al.* (2008)**	**India**	**Modern**
334	**G. g. murghi**	**[1]**	**EU847795**		**Kanginakudru *et al.* (2008)**	**India**	**Modern**
334	**G. g. murghi**	**[15]**	**EU847746**		**Kanginakudru *et al.* (2008)**	**India**	**Modern**
334	G. g. murghi		EU847749		Kanginakudru *et al.* (2008)	India	Modern
334	G. g. murghi		EU847751		Kanginakudru *et al.* (2008)	India	Modern
334	G. g. murghi		EU847783		Kanginakudru *et al.* (2008)	India	Modern
334	G. g. murghi		EU847784		Kanginakudru *et al.* (2008)	India	Modern
334	G. g. murghi		EU847785		Kanginakudru *et al.* (2008)	India	Modern
334	G. g. murghi		EU847786		Kanginakudru *et al.* (2008)	India	Modern
334	G. g. murghi		EU847787		Kanginakudru *et al.* (2008)	India	Modern
334	G. g. murghi		EU847788		Kanginakudru *et al.* (2008)	India	Modern
334	G. g. murghi		EU847789		Kanginakudru *et al.* (2008)	India	Modern
334	G. g. murghi		EU847790		Kanginakudru *et al.* (2008)	India	Modern
334	G. g. murghi		EU847797		Kanginakudru *et al.* (2008)	India	Modern
334	G. g. murghi		EU847798		Kanginakudru *et al.* (2008)	India	Modern

334	G. g. murghi		EU847799		Kanginakudru *et al.* (2008)	India	Modern
334	G. g. murghi		EU847800		Kanginakudru *et al.* (2008)	India	Modern
335	**G. g. murghi**	**[1]**	**EU847748**		**Kanginakudru *et al.* (2008)**	**India**	**Modern**
336	**G. g. murghi**	**[1]**	**EU847750**		**Kanginakudru *et al.* (2008)**	**India**	**Modern**
337	**G. g. murghi**	**[1]**	**EU847752**		**Kanginakudru *et al.* (2008)**	**India**	**Modern**
338	**G. g. murghi**	**[1]**	**EU847771**		**Kanginakudru *et al.* (2008)**	**India**	**Modern**
339	**G. g. murghi**	**[1]**	**EU847772**		**Kanginakudru *et al.* (2008)**	**India**	**Modern**
340	**G. g. murghi**	**[1]**	**EU847773**		**Kanginakudru *et al.* (2008)**	**India**	**Modern**
341	**G. g. murghi**	**[1]**	**EU847774**		**Kanginakudru *et al.* (2008)**	**India**	**Modern**
342	**G. g. murghi**	**[1]**	**EU847775**		**Kanginakudru *et al.* (2008)**	**India**	**Modern**
343	**G. g. murghi**	**[1]**	**EU847776**		**Kanginakudru *et al.* (2008)**	**India**	**Modern**
344	**G. g. murghi**	**[1]**	**EU847777**		**Kanginakudru *et al.* (2008)**	**India**	**Modern**
345	**G. g. murghi**	**[1]**	**EU847778**		**Kanginakudru *et al.* (2008)**	**India**	**Modern**
346	**G. g. murghi**	**[1]**	**EU847779**		**Kanginakudru *et al.* (2008)**	**India**	**Modern**
347	**G. g. murghi**	**[1]**	**EU847780**		**Kanginakudru *et al.* (2008)**	**India**	**Modern**
348	**G. g. murghi**	**[1]**	**EU847781**		**Kanginakudru *et al.* (2008)**	**India**	**Modern**
349	**G. g. murghi**	**[1]**	**EU847782**		**Kanginakudru *et al.* (2008)**	**India**	**Modern**
350	**G. g. murghi**	**[1]**	**EU847791**		**Kanginakudru *et al.* (2008)**	**India**	**Modern**
351	**G. g.**	**[1]**	**EU847792**		**Kanginakudru *et al.* (2008)**	**India**	**Modern**

	murghi						
352	**G. g. murghi**	**[1]**	**EU847793**		**Kanginakudru *et al.* (2008)**	**India**	**Modern**
353	**G. g. murghi**	**[1]**	**EU847794**		**Kanginakudru *et al.* (2008)**	**India**	**Modern**
354	**G. g. murghi**	**[1]**	**EU847796**		**Kanginakudru *et al.* (2008)**	**India**	**Modern**
355	**G. g. murghi**	**[1]**	**EU847802**		**Kanginakudru *et al.* (2008)**	**India**	**Modern**
356	**G. g. murghi**	**[1]**	**EU847803**		**Kanginakudru *et al.* (2008)**	**India**	**Modern**
357	**G. g. murghi**	**[1]**	**EU847804**		**Kanginakudru *et al.* (2008)**	**India**	**Modern**
358	**G. g. murghi**	**[1]**	**EU847805**		**Kanginakudru *et al.* (2008)**	**India**	**Modern**
359	**G. g. murghi**	**[1]**	**EU847806**		**Kanginakudru *et al.* (2008)**	**India**	**Modern**
360	**G. g. murghi**	**[1]**	**EU847807**		**Kanginakudru *et al.* (2008)**	**India**	**Modern**
361	**G. g. murghi**	**[1]**	**EU847809**		**Kanginakudru *et al.* (2008)**	**India**	**Modern**
362	**E**	**[2]**	**EU847811**		**Kanginakudru *et al.* (2008)**	**India**	**Modern**
362	E		EU847812		Kanginakudru *et al.* (2008)	India	Modern
363	**E**	**[1]**	**EU847814**		**Kanginakudru *et al.* (2008)**	**India**	**Modern**
364	**D**	**[1]**	**3895A**	**NUAN02**	**This study**	**Niue**	**Ancient**
364	**D**	**[1]**	**3896A**	**NUAN01**	**This study**	**Niue**	**Ancient**
365	**D**	**[36]**	**9030A**	**MQMD05**	**This study**	**Marquesas**	**Modern**
365	D		9065A	RNAN03	This study	Rapa Nui	Ancient
365	D		9073A	RNAN11	This study	Rapa Nui	Ancient
365	D		9066A	RNAN13	This study	Rapa Nui	Ancient
365	D		FJ914319	VAN03	Dancause *et al.* (2011)	Vanuatu	Modern
365	D		FJ914320	VAN04	Dancause *et al.* (2011)	Vanuatu	Modern
365	D		FJ914322	VAN06	Dancause *et al.* (2011)	Vanuatu	Modern
365	D		FJ914329	VAN13	Dancause *et al.* (2011)	Vanuatu	Modern
365	D		FJ914334	VAN18	Dancause *et al.* (2011)	Vanuatu	Modern
365	D		FJ914338	VAN22	Dancause *et al.* (2011)	Vanuatu	Modern

365	D		FJ914340	VAN24	Dancause *et al.* (2011)	Vanuatu	Modern
365	D		FJ914341	VAN25	Dancause *et al.* (2011)	Vanuatu	Modern
365	D		FJ914342	VAN26	Dancause *et al.* (2011)	Vanuatu	Modern
365	D		FJ914343	VAN27	Dancause *et al.* (2011)	Vanuatu	Modern
365	D		FJ914350	VAN34	Dancause *et al.* (2011)	Vanuatu	Modern
365	D		FJ914351	VAN35	Dancause *et al.* (2011)	Vanuatu	Modern
365	D		FJ914354	VAN38	Dancause *et al.* (2011)	Vanuatu	Modern
365	D		FJ914355	VAN39	Dancause *et al.* (2011)	Vanuatu	Modern
365	D		FJ914356	VAN40	Dancause *et al.* (2011)	Vanuatu	Modern
365	D		FJ914358	VAN42	Dancause *et al.* (2011)	Vanuatu	Modern
365	D		8136B	HWAN08	This study	Hawaii	Ancient
365	D		8670A	HWAN02	This study	Hawaii	Ancient
365	D		8671A	HWAN03	This study	Hawaii	Ancient
365	D		8672A	HWAN04	This study	Hawaii	Ancient
365	D		8674A	HWAN05	This study	Hawaii	Ancient
365	D		8675A	HWAN07	This study	Hawaii	Ancient
365	D		9031A	MQMD04	This study	Marquesas	Modern
365	D		9034A	MQMD02	This study	Marquesas	Modern
365	D		9057A	RNAN04	This study	Rapa Nui	Ancient
365	D		9067A	RNAN02	This study	Rapa Nui	Ancient
365	D		9068A	RNAN01	This study	Rapa Nui	Ancient
365	D		9070A	RNAN08	This study	Rapa Nui	Ancient
365	D		9072A	RNAN10	This study	Rapa Nui	Ancient
365	D		9074A	RNAN12	This study	Rapa Nui	Ancient
365	D		9075A	RNAN07	This study	Rapa Nui	Ancient
365	D		9076A	RNAN06	This study	Rapa Nui	Ancient
366	**D**	**[1]**	**9069A**	**RNAN05**	**This study**	**Rapa Nui**	**Ancient**
367	**D**	**[1]**	**8668A**	**HWAN01**	**This study**	**Hawaii**	**Ancient**
368		*[3]*	*9071A*	*RNAN09*		*Rapa Nui*	*Ancient*
368			*EF535242*	*PAQANA004*	*Storey et al. (2007)*	*Rapa Nui*	*Ancient*
368			*EF535243*	*PAQANA006*	*Storey et al. (2007)*	*Rapa Nui*	*Ancient*
369	**E**	**[1]**		**HWMD10**	**This study**	**Hawaii**	**Modern**
370	**E**	**[1]**		**HWMD03**	**This study**	**Hawaii**	**Modern**
371	**E**	**[1]**		**HWMD02**	**This study**	**Hawaii**	**Modern**

372		*[1]*		*AMSA01*	*Storey et al. (2007)*	*American Samoa*	*Ancient*
373		*[2]*	*EF535244*	*PAQANA009*	*Storey et al. (2007)*	*Rapa Nui*	*Ancient*
373			*EF535245*	*PAQANA010*	*Storey et al. (2007)*	*Rapa Nui*	*Ancient*
374		*[1]*	*EF535246*	*PAQANA011*	*Storey et al. (2007)*	*Rapa Nui*	*Ancient*
375		*[1]*	*EF535247*	*PAQHAN001*	*Storey et al. (2007)*	*Rapa Nui*	*Ancient*
376		*[1]*		*ELAR01*	*Storey et al. (2007)*	*El Arenal*	*Ancient*
377		*[2]*		*HWKU01*	*Storey et al. (2007)*	*Hawaii*	*Ancient*
377				*NUIE01*	*Storey et al. (2007)*	*Niue*	*Ancient*
378		*[1]*		*TONG01*	*Storey et al. (2007)*	*Tonga*	*Ancient*
379		*[1]*		*TONG02*	*Storey et al. (2007)*	*Tonga*	*Ancient*
380	**D**	**[1]**	**FJ914362**	**GUAM001**	**Dancause *et al.* (2011)**	**Guam**	**Modern**
381	**D**	**[1]**	**FJ914363**	**GUAM002**	**Dancause *et al.* (2011)**	**Guam**	**Modern**
382	**E**	**[1]**	**FJ914317**	**VAN01**	**Dancause *et al.* (2011)**	**Vanuatu**	**Modern**
383	**E**	**[1]**	**FJ914318**	**VAN02**	**Dancause *et al.* (2011)**	**Vanuatu**	**Modern**
384	**D**	**[9]**	**FJ914323**	**VAN07**	**Dancause *et al.* (2011)**	**Vanuatu**	**Modern**
384	D		FJ914325	VAN09	Dancause *et al.* (2011)	Vanuatu	Modern
384	D		FJ914326	VAN10	Dancause *et al.* (2011)	Vanuatu	Modern
384	D		FJ914327	VAN11	Dancause *et al.* (2011)	Vanuatu	Modern
384	D		FJ914328	VAN12	Dancause *et al.* (2011)	Vanuatu	Modern
384	D		FJ914330	VAN14	Dancause *et al.* (2011)	Vanuatu	Modern
384	D		FJ914331	VAN15	Dancause *et al.* (2011)	Vanuatu	Modern
384	D		FJ914332	VAN16	Dancause *et al.* (2011)	Vanuatu	Modern
384	D		FJ914359	VAN43	Dancause *et al.* (2011)	Vanuatu	Modern
385	**D**	**[2]**	**FJ914333**	**VAN17**	**Dancause *et al.* (2011)**	**Vanuatu**	**Modern**
385	D		FJ914335	VAN19	Dancause *et al.* (2011)	Vanuatu	Modern
386	**D**	**[1]**	**FJ914346**	**VAN30**	**Dancause *et al.* (2011)**	**Vanuatu**	**Modern**
387	**D**	**[1]**	**FJ914347**	**VAN31**	**Dancause *et al.* (2011)**	**Vanuatu**	**Modern**
388	**D**	**[1]**	**FJ914348**	**VAN32**	**Dancause *et al.* (2011)**	**Vanuatu**	**Modern**
389	**D**	**[1]**	**FJ914357**	**VAN41**	**Dancause *et al.* (2011)**	**Vanuatu**	**Modern**
390	**G. g. murghi**	**[1]**	**EU847758**	**INDB11**	**Kanginakudru *et al.* (2008)**	**India**	**Modern**
391	**G. g. murghi**	**[2]**	**EU847764**	**INDB12**	**Kanginakudru *et al.* (2008)**	**India**	**Modern**
391	G. g.		EU847767	INDB7	Kanginakudru *et al.* (2008)	India	Modern

	murghi						
392	**G. g. murghi**	**[1]**	**EU847757**	**INDB14**	**Kanginakudru *et al.* (2008)**	**India**	**Modern**
393	**G. g. murghi**	**[1]**	**EU847765**	**INDB16**	**Kanginakudru *et al.* (2008)**	**India**	**Modern**
394	**G. g. murghi**	**[1]**	**EU847759**	**INDB17**	**Kanginakudru *et al.* (2008)**	**India**	**Modern**
395	**G. g. murghi**	**[1]**	**EU847761**	**INDB20**	**Kanginakudru *et al.* (2008)**	**India**	**Modern**
396	**G. g. murghi**	**[1]**	**EU847756**	**INDB21**	**Kanginakudru *et al.* (2008)**	**India**	**Modern**
397	**G. g. murghi**	**[1]**	**EU847766**	**INDB4**	**Kanginakudru *et al.* (2008)**	**India**	**Modern**
398	**G. g. murghi**	**[1]**	**EU847760**	**INDB8**	**Kanginakudru *et al.* (2008)**	**India**	**Modern**
399	**G. g. murghi**	**[1]**	**EU847769**	**INDRJ16**	**Kanginakudru *et al.* (2008)**	**India**	**Modern**
399	**G. g. murghi**	**[1]**	**EU847770**	**INDK5**	**Kanginakudru *et al.* (2008)**	**India**	**Modern**
400	**G. g. murghi**	**[1]**	**EU847768**	**INDRJ15**	**Kanginakudru *et al.* (2008)**	**India**	**Modern**
401	**G. g. murghi**	**[1]**	**EU847763**	**INDRJ25**	**Kanginakudru *et al.* (2008)**	**India**	**Modern**
402	**G. g. murghi**	**[2]**	**EU847754**	**INDK6**	**Kanginakudru *et al.* (2008)**	**India**	**Modern**
402	G. g. murghi		EU847755	INDB2	Kanginakudru *et al.* (2008)	India	Modern
403	**G. g. murghi**	**[1]**	**EU847762**	**INDRJ2**	**Kanginakudru *et al.* (2008)**	**India**	**Modern**
404	**G. g. murghi**	**[1]**	**EU847753**	**INDB6**	**Kanginakudru *et al.* (2008)**	**India**	**Modern**

Bold – Representative sequence used in collapsed dataset.

Italic underlined – Storey *et al.* (2007) sequences used in phylogenetic trees to identify to haplogroup level, but not used in Network.

References

Abdulla MA, Ahmed I, Assawamakin A, Bhak J, Brahmachari SK, Calacal GC, Chaurasia A, Chen CH, Chen J, Chen YT, Chu J, Cutiongco-de la Paz EM, De Ungria MC, Delfin FC, Edo J, Fuchareon S, Ghang H, Gojobori T, Han J, Ho SF, Hoh BP, Huang W, Inoko H, Jha P, Jinam TA, Jin L, Jung J, Kangwanpong D, Kampuansai J, Kennedy GC, Khurana P, Kim HL, Kim K, Kim S, Kim WY, Kimm K, Kimura R, Koike T, Kulawonganunchai S, Kumar V, Lai PS, Lee JY, Lee S, Liu ET, Majumder PP, Mandapati KK, Marzuki S, Mitchell W, Mukerji M, Naritomi K, Ngamphiw C, Niikawa N, Nishida N, Oh B, Oh S, Ohashi J, Oka A, Ong R, Padilla CD, Palittapongarnpim P, Perdigon HB, Phipps ME, Png E, Sakaki Y, Salvador JM, Sandraling Y, Scaria V, Seielstad M, Sidek MR, Sinha A, Srikummool M, Sudoyo H, Sugano S, Suryadi H, Suzuki Y, Tabbada KA, Tan A, Tokunaga K, Tongsima S, Villamor LP, Wang E, Wang Y, Wang H, Wu JY, Xiao H, Xu S, Yang JO, Shugart YY, Yoo HS, Yuan W, Zhao G and Zilfalil BA (2009). Mapping human genetic diversity in Asia. *Science* 326(5959): 1541-1545.

Addison DJ and Matisoo-Smith E (2010). Rethinking Polynesians origins: a West-Polynesia Triple-I model. *Molecular Ecology* 45: 1-12.

Adebambo A, Mobegi V, Mwacharo J, Oladejo B, Adewale R, Ilori L, Makanjuola B, Afolayan O, Bjornstad G, Jianlin H and Hanotte O (2010). Lack of phylogeographic structure in Nigerian village chickens revealed by mitochondrial DNA D-loop sequence analysis. *International Journal of Poultry Science* 9(5): 503-507.

Allen MS, Matisoo-Smith E and Horsburgh A (2001). Pacific 'Babes': Issues in the origins and dispersal of Pacific pigs and the potential of mitochondrial DNA analysis. *International Journal of Osteoarchaeology* 11(1-2): 4-13.

Anderson A (2009). The rat and the octopus: initial human colonization and the prehistoric introduction of domestic animals to Remote Oceania. *Biological Invasions* 11: 1503-1519.

Bandelt HJ, Forster P and Rohl A (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* 16(1): 37-48.

Bedford S, Spriggs M and Regenvanu R (2006). The Teouma Lapita site and the early human settlement of the Pacific Islands. 80(March): 812-828.

Bellwood P (1991). The Austronesian dispersal and the origin of languages. *Scientific American* 265(1): 88-93.

Berthouly-Salazar C, Rognon X, Van TN, Gely M, Chi CV, Tixier-Boichard M, Bed'Hom B, Bruneau N, Verrier E, Maillard JC and Michaux JR (2010). Vietnamese chickens: a gate towards Asian genetic diversity. *BMC Genetics* 11: 53.

Blust R (2000). Chamorro historical phonology (Mariana Islands, Proto-Austronesian). *Oceanic Linguistics* 39(1): 83-122.

Bulbeck D (2008). An integrated perspective on the Austronesian diaspora: The switch from cereal agriculture to maritime foraging in the colonisation of Island Southeast Asia. *Australian Archaeology* 67(2008): 31-51.

Capelli C, Wilson JF, Richards M, Stumpf MPH, Gratrix F, Oppenheimer S, Underhill P, Pascali VL, Ko TM and Goldstein DB (2001). A predominantly indigenous paternal heritage for the Austronesian-speaking peoples of insular Southeast Asia and Oceania. *American Journal of Human Genetics* 68(2): 432-443.

Carter G (1971). Pre-Columbian chickens in America. Man across the sea: Problems of pre-Columbian contacts. Riley C, Kelley J, Pennington C and Rands R. Austin, University of Texas Press: 178-218.

Champlot S, Berthelot C, Pruvost M, Bennett EA, Grange T and Geigl EM (2010). An efficient multistrategy DNA decontamination procedure of PCR reagents for hypersensitive PCR applications. *PLOS ONE* 5(9): e13042.

Clark G and Anderson A (2009). *The early prehistory of Fiji*. Canberra, ANU E Press.

Clarke AC, Burtenshaw MK, McLenachan Pa, Erickson DL and Penny D (2006). Proceedings of the SMBE Tri-National Young Investigators' Workshop 2005. Reconstructing the origins and dispersal of the Polynesian bottle gourd (*Lagenaria siceraria*). *Molecular Biology and Evolution* 23(5): 893-900.

Cooper A and Poinar HN (2000). Ancient DNA: Do it right or not at all. *Science* 289(5482): 1139.

Dancause KN, Vilar MG, Steffy R and Lum JK (2011). Characterizing genetic diversity of contemporary Pacific chickens using mitochondrial DNA analyses. *PLOS ONE* 6(2): e16843-e16843.

Desjardins P and Morais R (1990). Sequence and gene organization of the chicken mitochondrial genome - A novel gene order in higher vertebrates. *Journal of Molecular Biology* 212(4): 599-634.

Diamond J (2001). Polynesian origins - Slow boat to melanesia? Reply. *Nature* 410(6825): 167-167.

Diamond J and Bellwood P (2003). Farmers and their languages: The first expansions. *Science* 300(5619): 597-603.

Dobney K, Cucchi T and Larson G (2008). The pigs of Island Southeast Asia and the Pacific: New evidence for taxonomic status and human-mediated dispersal. *Asian Perspectives* 47(1): 59-74.

Donohue M and Denham T (2010). Island Southeast Asia during the mid-Holocene: reframing Austronesian history. *Current Anthropology* 51(2): 223-256.

Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A, Duran C, Field M, Heled J, Kearse M, Markowitz S, Moir R, Stones-Havas S, Sturrock S, Thierer T and Wilson AC (2011). Geneious v5.4. *Available from* http://www.geneious.com/.

Fumihito A, Miyake T, Takada M, Shingu R, Endo T, Gojobori T, Kondo N and Ohno S (1996). Monophyletic origin and unique dispersal patterns of domestic fowls. *Proceedings of the National Academy of Sciences of the United States of America* 93(13): 6792-6795.

Gongora J, Rawlence NJ, Mobegi VA, Jianlin H, Alcalde JA, Matus JT, Hanotte O, Moran C, Austin JJ, Ulm S, Anderson AJ, Larson G and Cooper A (2008). Indo-European and Asian origins for Chilean and Pacific chickens revealed by mtDNA. *Proceedings of the National Academy of Sciences of the United States of America* 105(30): 10308-10313.

Green RC (1991). The Lapita Cultural Complex: Current evidence and proposed models. *Indo-Pacific Prehistory Association Bulletin* 11: 295-305.

Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W and Gascuel O (2010). New algorithms and methods to estimate Maximum-Likelihood phylogenies: Assessing the performance of PhyML 3.0. *Systematic Biology* 59(3): 307-321.

Hather J and Kirch PV (1991). Prehistoric sweet potato (*Ipomoea batatas*) from Mangaia Island, Central Polynesia. *Antiquity* 65: 887-893.

Hill C, Soares P, Mormina M, Macaulay V, Clarke D, Blumbach PB, Vizuete-Forster M, Forster P, Bulbeck D, Oppenheimer S and Richards M (2007). A mitochondrial stratigraphy for Island Southeast Asia. *American Journal of Human Genetics* 80(1): 29-43.

Huson DH and Bryant D (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* 23(2): 254-267.

Johannessen CL (1982). Melanotic chicken use and Chinese traits in Guatemala. *Revista de Historia de America* 93: 73-89.

Kanginakudru S, Metta M, Jakati RD and Nagaraju J (2008). Genetic evidence from Indian red jungle fowl corroborates multiple domestication of modern day chicken. *BMC Evolutionary Biology* 8: 174-174.

Keane TM, Creevey CJ, Pentony MM, Naughton TJ and McInerney JO (2006). Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evolutionary Biology* 6(29): 1-17.

Kirch P (1997). *The Lapita peoples: Ancestors of the oceanic world*. Oxford, Blackwell Publishers.

Lansing JS, Cox MP, Downey SS, Gabler BM, Hallmark B, Karafet TM, Norquest P, Schoenfelder JW, Sudoyo H, Watkins JC and Hammer MF (2007). Coevolution of languages and genes on the island of Sumba, eastern Indonesia. *Proceedings of the National Academy of Sciences of the United States of America* 104(41): 16022-16026.

Larson G, Cucchi T, Fujita M, Matisoo-Smith E, Robins J, Anderson A, Rolett B, Spriggs M, Dolman G, Kim TH, Thuy NT, Randi E, Doherty M, Due RA, Bollt R, Djubiantono T, Griffin B, Intoh M, Keane E, Kirch P, Li KT, Morwood M, Pedrina LM, Piper PJ, Rabett RJ, Shooter P, Van den Bergh G, West E, Wickler S, Yuan J, Cooper A and Dobney K (2007). Phylogeny and ancient DNA of *Sus* provides insights into neolithic expansion in Island Southeast Asia and Oceania. *Proceedings of the National Academy of Sciences of the United States of America* 104(12): 4834-4839.

Larson G, Dobney K, Albarella U, Fang MY, Matisoo-Smith E, Robins J, Lowden S, Finlayson H, Brand T, Willerslev E, Rowley-Conwy P, Andersson L and Cooper A (2005). Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science* 307(5715): 1618-1621.

Lebot V (1999). Biomolecular evidence for plant domestication in Sahul. *Genetic Resources and Crop Evolution* 46(6): 619-628.

Lemey P, Rambaut A, Drummond AJ and Suchard MA (2009). Bayesian phylogeography finds its roots. *PLoS Computational Biology* 5(9): e1000520.

Leonard J, Shanks O, Hofreiter M, Kreuz E, Hodges L, Ream W, Wayne R and Fleischer R (2007). Animal DNA in PCR reagents plagues ancient DNA research. *Journal of Archaeological Science* 34(9): 1361-1366.

Librado P and Rozas J (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25(11): 1451-1452.

Liu YP, Wu GS, Yao YG, Miao YW, Luikart G, Baig M, Beja-Pereira A, Ding ZL, Palanichamy MG and Zhang YP (2006). Multiple maternal origins of chickens: out of the Asian jungles. *Molecular Phylogenetics and Evolution* 38(1): 12-19.

Matisoo-Smith E (2002). Something old, something new: Do genetic studies of contemporary populations reliably represent prehistoric populations of Pacific *Rattus exulans*? *Human Biology* 74(3): 489-496.

Matisoo-Smith E, Roberts RM, Irwin GJ, Allen JS, Penny D and Lambert DM (1998). Patterns of prehistoric human mobility in Polynesia indicated by mtDNA from the Pacific rat. *Proceedings of the National Academy of Sciences of the United States of America* 95(25): 15145-15150.

Matisoo-Smith E and Robins JH (2004). Origins and dispersals of Pacific peoples: evidence from mtDNA phylogenies of the Pacific rat. *Proceedings of the National Academy of Sciences of the United States of America* 101(24): 9167-9172.

Muchadeyi FC, Eding H, Simianer H, Wollny CB, Groeneveld E and Weigend S (2008). Mitochondrial DNA D-loop sequences suggest a Southeast Asian and Indian origin of Zimbabwean village chickens. *Animal Genetics* 39(6): 615-622.

Mwacharo JM, Bjornstad G, Mobegi V, Nomura K, Hanada H, Amano T, Jianlin H and Hanotte O (2011). Mitochondrial DNA reveals multiple introductions of domestic chicken in East Africa. *Molecular Phylogenetics and Evolution* 58(2): 374-382.

Nishibori M, Shimogiri T, Hayashi T and Yasue H (2005). Molecular evidence for hybridization of species in the genus *Gallus* except for *Gallus varius*. *Animal Genetics* 36(5): 367-375.

O'Brien PJ (1972). The sweet potato: Its origin and dispersal. *American Anthropologist* 74(3): 342-365.

Oka T, Ino Y, Nomura K, Kawashima S, Kuwayama T, Hanada H, Amano T, Takada M, Takahata N, Hayashi Y and Akishinonomiya F (2007). Analysis of mtDNA sequences shows Japanese native chickens have multiple origins. *Animal genetics* 38(3): 287-293.

Oppenheimer S (2004). The 'Express Train from Taiwan to Polynesia': on the congruence of proxy lines of evidence. *World Archaeology* 36(4): 591-600.

Oppenheimer SJ and Richards M (2001). Polynesian origins - Slow boat to melanesia? *Nature* 410(6825): 166-167.

Oskarsson MC, Klutsch CF, Boonyaprakob U, Wilton A, Tanabe Y and Savolainen P (2011). Mitochondrial DNA data indicate an introduction through Mainland Southeast Asia for Australian dingoes and Polynesian domestic dogs. *Proceedings of the Royal Society of London Series B-Biological Sciences*.

Paabo S, Poinar H, Serre D, Jaenicke-Despres V, Hebler J, Rohland N, Kuch M, Krause J, Vigilant L and Hofreiter M (2004). Genetic analyses from ancient DNA. *Annual review of genetics* 38: 645-679.

Perrier X, De Langhe E, Donohue M, Lentfer C, Vrydaghs L, Bakry F, Carreel F, Hippolyte I, Horry J-P, Jenny C, Lebot V, Risterucci A-M, Tomekpe K, Doutrelepont H, Ball T, Manwaring J, de Maret P and Denham T (2011). Multidisciplinary perspectives on banana (*Musa spp.*) domestication. *Proceedings of the National Academy of Sciences of the United States of America* 108(28): 11311-11318.

R development core team (2011). R: A language and environment for statistical computing. R foundation for statistical computing

Razafindrazaka H, Ricaut FX, Cox MP, Mormina M, Dugoujon JM, Randriamarolaza LP, Guitard E, Tonasso L, Ludes B and Crubezy E (2010). Complete mitochondrial DNA sequences provide new insights into the Polynesian motif and the peopling of Madagascar. *European Journal of Human Genetics* 18(5): 575-581.

Robins JH, Matisoo-Smith EA and Furey L (2001). Hit or miss? Factors affecting DNA preservation in Pacific archaeological material. Australasian Connections and New Directions: Proceedings of the 7th Australasian Archaeometry Conference, Auckland, University of Auckland.

Ross M (2005). The Batanic languages in relation to the early history of the Malayo-Polynesian subgroup of Austronesian. *Journal of Austronesian Studies* 1(2): 1-23.

Savolainen P, Leitner T, Wilton AN, Matisoo-Smith E and Lundeberg J (2004). A detailed picture of the origin of the Australian dingo, obtained from the study of mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America* 101: 12387-12390.

Shutler J and Marck J (1975). On the dispersal of the Austronesian horticulturalists. *Archaeology and Physical Anthropology in Oceania* 10: 81-113.

Silva P, Guan X, Ho-Shing O, Jones J, Xu J, Hui D, Notter D and Smith E (2009). Mitochondrial DNA-based analysis of genetic variation and relatedness among Sri Lankan indigenous chickens and the Ceylon junglefowl (*Gallus lafayetti*). *Animal Genetics* 40(1): 1-9.

Soares P, Rito T, Trejaut J, Mormina M, Hill C, Tinkler-Hundal E, Braid M, Clarke DJ, Loo JH, Thomson N, Denham T, Donohue M, Macaulay V, Lin M, Oppenheimer S and Richards MB (2011). Ancient voyaging and Polynesian origins. *American Journal of Human Genetics* 88(2): 239-247.

Specht J (2007). Small islands in the big picture: the formative period of Lapita in the Bismarck Archipelago. Oceanic Explorations: Lapita and Western Pacific Settlement. Bedford S, Sand C and Connaughton SP. Canberra, ANU E Press. 26: 51-70.

Spriggs M (1997). *The Island Melanesians*. Oxford, Blackwell Publishers.

Storey AA, Quiroz D, Ramirez JM, Beavan-Athfield N, Addison DJ, Walter R, Hunt T, Athens JS, Huynen L and Matisoo-Smith EA (2008). Pre-Columbian chickens, dates, isotopes, and mtDNA. *Proceedings of the National Academy of Sciences of the United States of America* 105(48): E99.

Storey AA, Ramirez JM, Quiroz D, Burley DV, Addison DJ, Walter R, Anderson AJ, Hunt TL, Athens JS, Huynen L and Matisoo-smith EA (2007). Radiocarbon and DNA evidence for a pre-Columbian introduction of Polynesian chickens to Chile. *Proceedings of the National Academy of Sciences of the United States of America* 104(25): 10335-10339.

Storey AA, Spriggs M, Bedford S, Hawkins SC, Robins JH, Huynen L and Matisoo-Smith E (2010). Mitochondrial DNA from 3000-year old chickens at the Teouma site, Vanuatu. *Journal of Archaeological Science* 37(10): 2459-2468.

Summerhayes G (2000). *Lapita Interaction*. Canberra, ANU E Press.

Terrell J (1988). History as a family tree, history as an entangled bank: constructing images and interpretations of prehistory in the South Pacific. *Americas, The* 62: 642-657.

USB (2006). Shrimp DNase, Recombinant. Corporation U. Cleveland, USB Corporation: 1-2.

West B and Zhou B-X (1988). Did chickens go north? New evidence for domestication. *Journal of Archaeological Science* 15: 515-533.

Whitaker T and Carter G (1954). Oceanic drift of gourds-experimental observations. *American Journal of Botany* 41: 697-700.

Wickler S (2004). Modelling colonisation and migration in Micronesia from a zooarchaeological perspective. Colonisation, Migration, and Marginal Areas: A Zooarchaeological Approach. Mondini M, Munoz S and Wickler S. Oxford, Oxbow: 28-40.

Wilmshurst JM, Hunt TL, Lipo CP and Anderson AJ (2011). High-precision radiocarbon dating shows recent and rapid initial human colonization of East Polynesia. *Proceedings of the National Academy of Sciences of the United States of America* 108(5): 1815-1820.

Zhou B, Chen SY, Zhu Q, Yao YG and Liu YP (2010). Matrilineal components and genetic relationship of Silkies from China and Japan. *Journal of Poultry Science* 47(1): 22-27.

Chapter 6

Discussion and Conclusion

Overall summary of this thesis

This thesis aims to answer questions about the evolutionary and demographic history of species using the various tools available to evolutionary biologists. One common theme is the ability of ancient DNA (aDNA) to provide insight into past (and potentially lost) genetic diversity so that the evolutionary history of individual species can be reconstructed. This is especially important in the case of the Australian species, as a paucity of aDNA studies has resulted in the description of evolutionary histories of many Australian species from a limited viewpoint: that of the modern mitochondrial genome. The view from purely modern mitochondrial DNA data is limited in temporal depth, as species have often had genetic signal potentially erased by the most recent genetic bottleneck (Mourier *et al.* 2012). Adding aDNA to datasets of extant populations still remains an important way to extend the view of past (and potentially lost) genetic diversity past the last bottleneck. However, ancient DNA is just one technique that can be combined with other analysis and hypothesis-building methods to answer fundamental questions about evolutionary history. Where the questions are about human migration, such archaeological techniques as cultural reconstructions of pottery styles, linguistics, and modern human genetics can be added to aDNA studies of commensal species. If questions about taxonomy and speciation are being addressed, the standard skills of a palaeontologist for morphology and morphometrics are required in addition to aDNA to revise current species designations. And where species' responses to past climatic cycles are needed to investigate possible responses to future climate change, then ecological niche modelling (ENM) can provide insights that aDNA cannot currently access. All these methods have been utilised in this thesis to answer questions about a diverse range of evolutionary and demographic histories.

Chapters 3 and 4 of my thesis provide examples of the first studies that combine ancient DNA, museum samples, and modern specimens/sequences in the Australasian context to investigate population genetic and evolutionary biology questions. Ethanol preserved museum tissues have been used previously in studies of Australian species, *e.g.,* the Numbat, Pademelons, and the Australian Dragon (Fumagalli *et al.* 1999; Melville *et al.* 2007; Macqueen *et al.* 2009; Macqueen *et al.* 2010), and more difficult museum specimen types, such as feathers, dried toepads, and insect legs have been targeted (Eastwood and Hughes 2003; Rollins *et al.* 2009; Seabrook-Davison *et al.* 2009). However, few studies have used bone or teeth samples to investigate population level issues in Australia, and the two studies that do use skeletal material

are from museum collections of less than 250 years old (Macqueen *et al.* 2011; Paplinska *et al.* 2011). The few truly aDNA studies in Australia (*i.e.*, using sub-fossil samples older than 250 years) have not been used to answer population level questions. Heupink *et al.* (2011) examined purely taxonomic issues about the Emu, and used only farmed individuals to generate the mainland Emu dataset for comparison to the King Island Emu, so no population level work on wild populations was attempted. Similarly, the focus of work by Oskam *et al.* (2010) was on optimisation of DNA extraction from eggshell in the Australian context, with only four ancient Emu samples analysed. Therefore, chapter 3 is the first population level study in Australia using ancient samples thousands of years old, and chapter 4 is one of the first studies using ancient DNA techniques on freshly collected subfossil bone and teeth material.

In contrast, chapter 5 examines how genuine patterns of diversity can arise out of datasets where conclusions were previously drawn from erroneous ancient sequence data, and highlights how careful use of contamination controls can separate authentic from flawed data. Although the study in chapter 5 is not the first to use the genetic diversity of commensal animals to track human movements in the Pacific, the study is one of the first to use anti-contamination measures to ensure modern contaminating DNA is not obscuring our ability to reconstruct the migration routes of the Polynesians. This chapter also highlights the differences in temporal and spatial scales perpetuated by human-mediated dispersal across sea barriers compared to the Emu, which naturally dispersed to the southern offshore islands of Australia. Much greater distances can be traversed as a human-commensal dyad, and with fewer restrictions temporally and spatially.

Chapter 2: Methodology of this thesis

Even though the field of ancient DNA is progressing and new methods are being developed (Anderson *et al.* 2004; Drummond *et al.* 2005; Drummond and Rambaut 2007), many of the traditional analysis techniques still have a purpose, even if it is only to confirm the results of other, newer methods. However, the traditional techniques will require alternate forms for many formulas and statistics so that the additional information held in heterochronous datasets can be utilised to the maximum. As ancient DNA researchers acquire more bioinformatics skills or more statisticians/mathematicians apply their skills to answer evolutionary questions, this issue of a mismatch between heterochronous datasets and synchronous statistical tests will need to be resolved, *viz.* Depaulis *et al.* (2009), as the creation and use of programs using heterochronous datasets with population genetics statistics generated by synchronous tools are increasing (Excoffier and Foll 2011).

Chapter 3: the Emu (Dromaius spp.)

Issues encountered

The main problems encountered in this study were methodological issues related to the limitations of ancient DNA work in the Australian context, especially from the sub-optimal DNA preservation environments of King Island sand dunes in comparison to the cave deposits on Kangaroo Island, but also for the eggshell samples. The success rate of 36% for eggshell samples that span 12,500-16,500 ya in this study suggest it may not be possible to extend the ancient genetic record of Australia much past about 20,000 years old (e.g., the oldest Australian eggshell samples from Oskam *et al.* (2010) were dated from ~19,000 ya BP), although the limiting factors to the temporal depth that eggshell can span include the use of caves by Aborigines with Emu eggshells (see below) and the deposition environment of cave deposits to preserve the eggshell. The temporal span of Aboriginal occupation varies across different parts of the Australian landscape, such that cave occupation sites in the southwest corner may be more limited temporally compared to the east or north coasts, or vice versa (Flood 2004). All the Pleistocene eggshell samples from Oskam *et al.* (2010) and chapter 3 came from the same site, Tunnel Cave in western Australia. The study in chapter 3 had addition samples from Devil's Lair and WRS cave, however neither cave deposit yielded any amplifiable DNA, while Oskam *et al.* (2010) did obtain DNA from eggshell from WRS cave up to 700 years old. The successful DNA yield from Pleistocene-age samples suggests the cave conditions from Tunnel Cave offer the greatest chance of adding increased temporal depth to aDNA studies in Australia so far.

<u>Sample size</u>

Although only five of the 14 eggshell samples yielded DNA (a success rate of 36%) the amount of sample available was substantially less than those of the bone samples. Moreover, the age range that the eggshell samples spanned: 12,500-16,500 ya BP, provided the genetic dataset with considerable additional temporal information (including temporal information for two unique haplotypes, H17 and H18). In comparison, the island calibrations only span 8,800-14,000 ya (although dating the occurrence of two unique haplotypes, H4 and H5). Many of the island samples were also considerably damaged (both physically and genetically), with only eight of the 21 King Island Emu samples (38%) and four of the six Tasmanian Emu samples (67%) successfully amplifying and sequencing. Yet all 11 of the Kangaroo Island Emu samples produced sequence data. These 28 ancient samples represent one quarter of the total dataset, yet the low level of genetic diversity spanning the last 16.5 ky suggests that it may not be possible to improve resolution into the Emu's evolutionary history by simply increasing the sample size within the same time-span. Increasing the temporal depth of the eggshell sampling may be one of

the few ways in which greater resolution can be gained. Access to additional eggshell samples from cave deposits in the south-west (e.g., Tunnel Cave), will give greater power to analysis of the ancient western population, however exploration of cave deposits in the temperate east and Tasmania is also required to add to the sampling of the ancient eastern population. The level of emu eggshell in cave deposits is directly related to the level of human habitation, as Aborigines are known to have consumed emu eggs (Flood 2004; Richards *et al.* 2007). Therefore, archaeologists who specialise in eastern and southern cave sites will be critical contacts for further developing this project. Additionally, the large collection of emu eggshells dating back to ~140 kya from Lake Eyre, Port Augusta and the Murray-Darling held by Miller *et al.* (2005) would be an important resource to expand this project even further back in time.

Sample preservation

I believe that the relatively low success rate of King Island Emu ancient DNA extractions, PCRs and the problems associated with gaining the full sequence length reflects the poor state of preservation in the Australian context. As DNA starts degrading immediately upon death of the individual, and several factors influence DNA preservation (including high temperatures, pH and degree of water movement through the surrounding matrix), the sandy deposition environment on King Island may have caused higher DNA decay in those samples (see Figures 1 & 2 for contrasting pictures of a relatively well preserved and poorly preserved King Island Emu femora bones, respectively). In contrast, the limestone caves of Kangaroo Island appear to have preserved DNA to a greater degree (Hope *et al.* 1977).

The DNA yield of the King Island Emu (38%) vs. Kangaroo Island Emu (100%) bone samples also indicates that the more southerly latitude (and cooler temperatures) of King Island does not appear to have overcome the damage done by the deposition environment. However, it is also possible that the inferred age differences between the King Island (12,000 ya) and Kangaroo Island Emus (8,800 ya) populations, actually does reflect the relative ages of the samples (i.e., King Island Emu bones are actually older than the Kangaroo Island Emu bones). Carbon dating of the bones would be required before this can be investigated. Storage of the ancient samples at ambient temperatures in museums since excavation in the late 19th century may also have exacerbated their poor DNA preservation. Given the age of the eggshell samples and the amount of sample available for extraction, the success rate of the eggshell samples reinforces previously published results (Oskam *et al.* 2010) indicating the calcite structure provides greater protection for endogenous DNA from degradation.

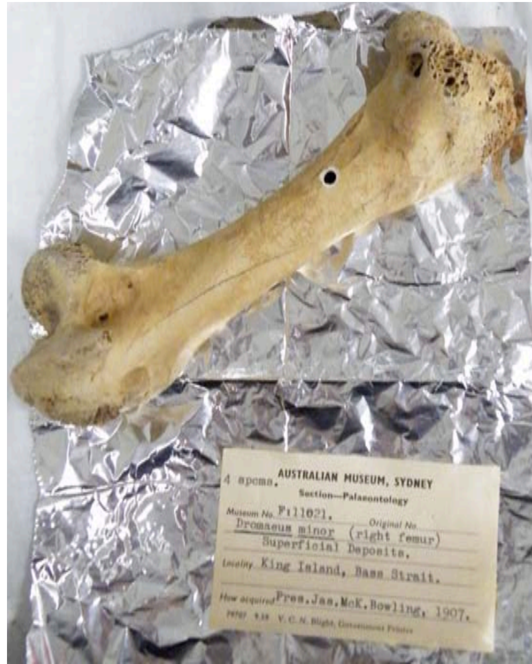


Figure 1. Example of King Island Emu femur bone from superficial deposit from sand blown site on King Island, in the Bass Strait between the Australian mainland and Tasmania. Note although there is some damage to the bone architecture, there were still enough features for all the morphological measurements to be taken. However, this was not always the case.

Figure 2. Example of King Island Emu femora bones from superficial deposit from sand blown site on King Island, in the Bass Strait between the Australian mainland and Tasmania. Most of these femora bones are too degraded for aDNA to be successfully retrieved.

<u>**Independence of sampling**</u>

In order to ensure that no Emu individual was sampled more than once, only right (R) bones were sampled for DNA (R femora for King Island Emus and R tibia from Kangaroo Island

Emus), and where articulated skeletons were recovered these were noted. Ideally, the same specimens would have been sampled for both DNA and morphometrics, so that direct correlations could be drawn between DNA sequence and morphological variation. However, specimens suited to morphological measurements were not always the best preserved for DNA purposes (supported by the lack of DNA amplification from many of the King Island Emu bones). The samples too damaged for measuring were roughly the same size and proportions as those bones with full morphological features suitable for measuring, therefore the assumption that all small emu bones collected from King Island actually represented the King Island Emu is likely valid.

Island formation as calibration dates

The technique of using geological events to calibrate divergence events has been used in other evolutionary biology studies, including the use of human entry into New Guinea and Australia to calibrate human mtDNA control region haplogroups (Endicott and Ho 2008); and the age of volcanic island formations to date insect clades, birds and plant species in Hawaii (Fleischer *et al.* 1998; Percy *et al.* 2008; Lerner *et al.* 2011), and birds in the Galapagos islands (Johnson and Weckstein 2011). Fleischer *et al.* (1998) and Percy *et al.* (2008) used the Hawaiian island formation dates to calibrate clades and estimate evolutionary rates via genetic distance measurements. However, in general, island calibration dates have been used as prior distributions on nodes of phylogenetic trees, for example (Endicott and Ho 2008) used the time to most recent common ancestor of haplogroup P in New Guinea/Australia in BEAST calculations as a lognormal distribution with minimum age of 40 kya, mean age of 45 kya and 95% of distribution between 40-55 kya. These calibrations are placed on nodes because the events being investigated are divergence events between clades, i.e., based on purely modern genetic data where two clades inhabit regions on either side of a barrier.

In contrast, I have used island separation ages as tip calibration ages to date the presence of haplotypes on islands. Therefore, the dates of isolation of island populations from those on the mainland are neither maximum, nor minimum ages for the lineages, but actual point estimates of the presence of the lineage on the landscape. It is necessary to assume that the lineages found on islands were also present on the adjacent mainland just prior to the rising sea completely isolated the islands – excluding mutations that occurred in the island populations since separation from the mainland.

Although the assumption of no mutations allows the date of separation of the island from the mainland to be used as an estimate of the date at which the particular lineages were present in southern Australia (see Figure 3 panel A), this assumption may or may not be valid. The

evolutionary rate estimated in chapter 3 of 10.6% substitutions-per-site-per-Million-years, actually equates to just under 0.52 substitutions over the 350 bp fragment of mtDNA *CR* over the 14,000 years since Tasmania separated from the Victorian mainland (Lambeck and Chappell 2001). Therefore, it is possible that mutations have occurred in the mtDNA *CR* of the island Emus since separation from the mainland. Within the reticulation on the network in Figure 2 (chapter 3), spontaneous back-mutations may have occurred to result in a matching haplotype to a mainland sequence (see Figure 3 panel B, below), with the original island Emu having a different haplotype. In this case the invalid assumption of no mutation underestimates the level of diversity present on the island since separation from the mainland. Although the scenario outlined above sounds unlikely, mutational hotspots that were highlighted in the reticulated 'central cube' in the phylogenetic network in 2 of chapter 3 may in fact display this sort of behaviour (Bandelt *et al.* 2002), such that, island haplotypes may mutate backward and forth within the 'central cube' of related haplotypes with little cost. However, the most parsimonious explanation is still that no new mutations have occurred in the intervening time period (as per panel A of Figure 3, below).

One assumption not explicitly stated is that the island Emu populations have only been isolated from mainland Emu populations since the most recent rise in sea level, and that the island Emu lineages have not been restricted for multiple climatic cycles to what were southern peninsulas when sea levels were lower. Given the high level of migration and long distance dispersal of the Emu, it is highly unlikely that no gene flow occurred between the Emu populations on what were these southern peninsulas and those further north on what is now the mainland, especially when the temperate conditions of the south eastern corner would have provided a refugium from the aridity of central Australia during the Last Glacial Maximum (~ 20,000 ya). For example, Tasmania was joined to the mainland via a landbridge from 43-14 kya and gene flow was extremely likely across the plain that was later inundated to become Bass Strait.

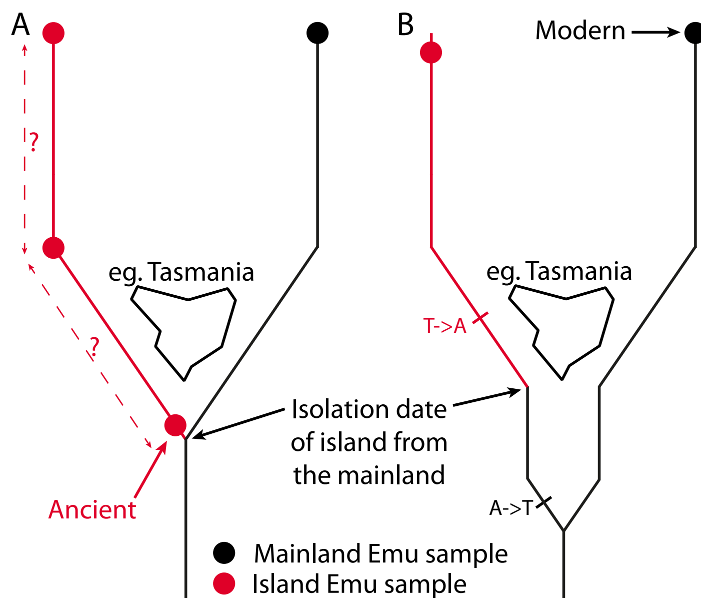


Figure 3. Diagram representing two possible scenarios by which the same mtDNA *CR* haplotype is found in both modern mainland Emu populations and ancient island Emu populations (e.g., Tasmanian Emu). A) The most parsimonious explanation is that the same haplotype was present on the (adjacent) mainland prior to the rise in sea level separating (at least) two Emus with the same haplotype. Consequently, when the descendants of the original Emus were sampled we obtain identical mtDNA *CR* sequences. B) Two similar haplotypes (differing by only one nucleotide) were present on the (adjacent) mainland and became separated when the sea level rose. However in the intervening period of time between the island separation and sampling the descendent Emus, the island Emu underwent a random mutation in the same nucleotide site at which the two Emus differ, mutating this nucleotide back to match the descendent of the other original Emu.

Expansion tests in Arlequin

The expansion tests implemented in Arlequin use mismatch distributions and the coalescent to 1) evaluate the evidence for spatial/demographic expansions; and 2) estimate the timing of such expansions (in units of mutation). There are several issues when using such expansion tests with heterochronous data, as the tests were designed for synchronous data, and do not deal with any potential evolution that may have occurred between sampling points. Previous studies (Depaulis *et al.* (2009) have highlighted this issue, but have only been concerned with instances where population genetic statistics were calculated from combined ancient and modern datasets (*i.e.,* as fully interbreeding populations). Depaulis *et al.* (2009) found that where the temporal sampling across the tree is substantial compared to the time to most recent common ancestor (tMRCA) (*e.g.,* where ancient samples are approximately 1Ne generations) and where the ancient and modern sample sizes are roughly equal in number, the heterochrony has a substantial effect on conclusions about demographic events (specifically erroneous conclusions about population structure or contraction). Even where the ancient sampling comprises only 0.1 Ne generations and with a lower proportion of the total sample size, spurious results predicting

population expansions and bottlenecks can occur (Depaulis *et al.* 2009). However this issue should not impact the studies in this thesis overly, as the datasets have been partitioned into time bins to reduce the impact of the heterochrony on the population genetic statistics.

In chapter 3, the eastern samples were separated from the western samples and each was pooled into time bins to remove the potentially confounding effects of geography. Within these time bins the ancient Emu 'populations' still spanned a few thousand years (*e.g.,* the ancient western population spanned 4,000 years – from 12,500-16,500 ya – and the ancient eastern population spanned 5,200 years – from 8,800-14,000 ya). However, as the Emu dataset had the same haplotypes present in both ancient and modern populations, it was assumed that even over these 4-5.2 thousand years not enough mutations had occurred (0.19 subs/350bp/5,200 years) for there to be issues with the population genetic statistics being estimated.

To test whether the Emu dataset as a whole was measurably evolving, the Path-O-Gen software (Rambaut 2009) was used to perform a Pearson's correlation test (r) between the tip-to-root distance and the sampling time of the data. The correlation was found to be very weak (correlation coefficient of 0.27, see Figure 4 for an example of the strength of a coefficient of 0.27 compared to 0.9). However, when 10 tip date randomisations were done to assess the significance of this correlation, all the 10 replicates had lower correlation values (p-value 0.0; Figure 5, below). As none of the replicates had a more extreme correlation than the empirical data, the Emu dataset appears to be measurably evolving, although at a very slow rate over the 16,500 years that the dataset spans. Only 10 randomisations were done, as no software program has yet been developed to automate the process of randomising the tip dates, generating a Maximum Likelihood tree, and inputting this data into Path-O-Gen. Therefore, the limited number of randomisations may be suggesting a significance of this correlation between tip-to-root distance and sampling time that is not actually represented in the data. As Navascues *et al.* (2010) suggests the Path-O-Gen test for 'measurable evolution' as a way of determining whether heterochronous datasets can be pooled for further tests using isochronous tools, this low level of 'measurable evolution' suggests the Emu dataset may not be violating the assumptions of isochronous analysis techniques too severely.

The effect of the sampling times on the genetic diversity of the Emu was further tested using AMOVA (implemented in Arlequin), to estimate the proportion of genetic variation explained by the sampling times, as suggested by Navascues *et al.* (2010). Although not originally intended to test temporal structure, AMOVA can also be applied to heterochronous data to apportion the genetic variation stemming from geographical vs. temporal structure (Navascues *et al.* 2010). To avoid limitations in testing for heterochrony with a test designed to

detect only geographical population structure, Navascues *et al.* (2010) note that both temporal and spatial information should be included as a nested analysis. When sampling time is nested inside geographic location for AMOVA in the Emu dataset (see Table 1, below), only 6% of the genetic variation was explained by geography and 3% of the genetic variation was explained by sampling times. The majority of the genetic variation (91%) was within population variation. Therefore, the time bins used to separate the Emu dataset into 'populations', revealed that the 'populations' showed little evidence of heterochrony (see Table 1). There were significant levels of substructure (F_{ST} = 0.091, p-value = 0.006) within 'populations', but no significant temporal divergence of populations by drift (F_{CT} = 0.061, p-value = 0.309), nor any geographic differentiation of populations (F_{SC} = 0.034, p-value = 0.094).

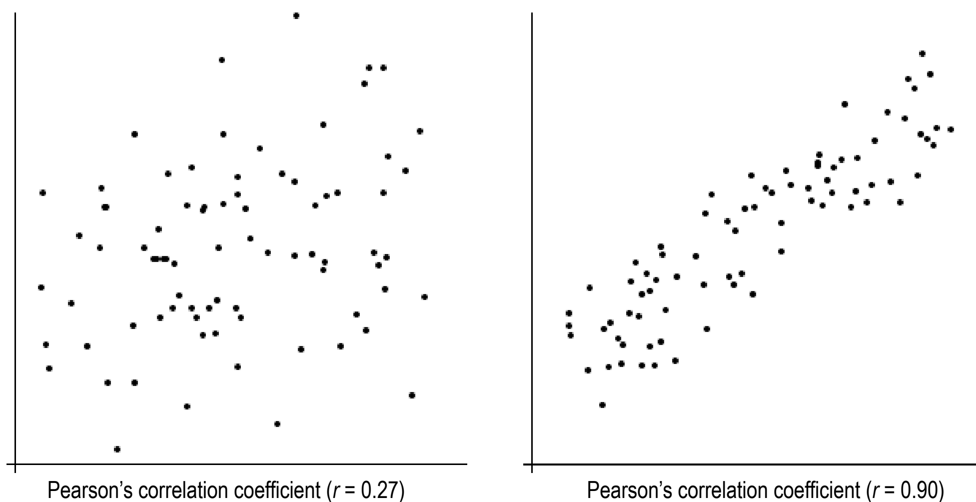


Figure 4. Examples of the strength of a Pearson's correlation coefficient *r* of 0.27 (weak) in comparison to a value of 0.9 (strong correlation). The correlation between the tip-to-root distance and sampling time of 0.27 for the Emu dataset suggested weak evidence of a measurably evolving dataset.

Table 1. Analysis of Molecular Variance (AMOVA) of the western vs. eastern Emu populations, with further divisions of ancient vs. modern nested within the geographic groupings.

Source of variation	*d.f.*	*Sum of squares*	*Variance components*	*Percentage of variation*
Among groups (geographic)	1	41.871	0.60110	6.05
Among populations within groups (temporal)	2	31.203	0.31375	3.16
Within populations	101	910.621	9.01605	90.79
Total	104	983.695	9.93090	

	Fixation values	P-value
F_{SC} (among groups)	0.034	0.094
F_{CT} (among populations within groups)	0.061	0.309
F_{ST} (within populations)	0.091	0.006

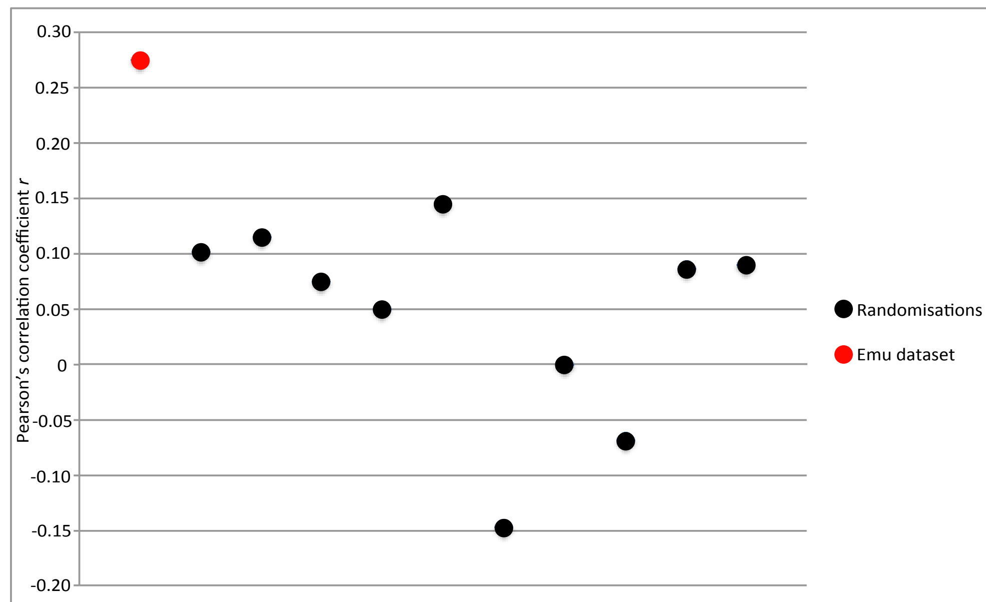


Figure 5. Pearson's correlation coefficient *r* between the tip-to-root distance and sampling time for the Emu dataset (red) and the 10 tip date randomisations (black). Although the Emu dataset only showed weak correlation between the tip-to-root distance and the sampling time (*i.e.*, weak evidence as a measurably evolving dataset), it was significantly different from the 10 random datasets (p-value = 0.0, as only 10 randomisations were performed).

The Arlequin expansion tests must be treated carefully when transforming coalescent time to absolute time, in order to estimate the years since expansion. This issue is the main restriction to the usefulness of these tests. The correct usage of molecular rates and generation times is essential to transforming *tau* to time in years accurately (Schenekar and Weiss 2011), with errors in this process stemming from both the estimation of the molecular rate and the generation times. In the case of the Emu, the generation times were calculated from age at first and last breeding obtained from the literature, which resulted from banding studies on wild birds (Davies 2002). For chapters 3 and 4, the molecular rates were translated from the BEAST analyses (which have an associated 95% HPD error).

In chapter 3, the estimate of molecular rate does not appear too inaccurate, as even with low diversity levels the Emu dataset passed the date randomisation test of Ho et al. (2011b), ensuring that the primary calibration dates contain enough temporal signal for the molecular rate to be estimated with some confidence. The margin of error for the Emu mutation rate (10.6% per Million years, 95% HPD of 2.46-21.7% per Million years) also overlaps with that of the Moa mutation rate from mtDNA *CR* data (8.7% per Million years, 95% HPD of 2.34%-20.4% per Million years; Bunce *et al.* 2009). Bunce *et al.* (2009) caution against relying too heavily on this rate as the topology used to perform the calculation was incorrect (Moa inter-species relationships were not accurate). However, the mtDNA *CR* dataset in Bunce *et al.* (2009) is the more appropriate of the ratite mutation rates published, as samples from the mtDNA *CR* dataset covered a similar temporal period to the Emu dataset in chapter 3 (radiocarbon dated samples spanned 1-19 kya). Furthermore, both the Bunce *et al.* (2009) rate and the Emu rate from chapter 3 were based on the much faster mutating hypervariable regions of the mtDNA *CR* and used tip dates to calibrate the tree, thereby avoiding the rates curve issue (Ho *et al.* 2007; Ho *et al.* 2011a). All other ratite mutation rate calculations available use the slower evolving conserved domain of the *CR* (0.49-0.54% per Million years) and protein coding genes (0.27% per Million years), and were calibrated using geological dates (Cooper *et al.* 2001; Baker *et al.* 2005). The Baker *et al.* (2005) and Cooper *et al.* (2001) studies used the problematic geological date of the separation of New Zealand from Gondwana (82 Mya) as a calibration for the divergence of Moa from other ratite and tinamou taxa. However, as recent studies postulate multiple losses of flight in ratites based on nuclear genes showing ratites are actually polyphyletic (Harshman *et al.* 2008; Phillips *et al.* 2010), the use of this geological event to calibrate ratite phylogenetic trees has been dropped recently as it does not rule out the dispersal of ratites by flying. Therefore, although the tMRCA of the Moa and Emu is 70.6 Mya (46.6-92.5 Mya, 95% HPD) (Phillips *et al.* 2010), the parallel life history traits that the Moa and Emu share, such as being large bodied, cursorial birds

that cannot fly, suggests the overlap in mutation rate 95% HPD is not unexpected. Few other ratite mutation rates have been published for comparison, with a kiwi (*Apteryx spp.*) phylogeny being reconstructed using modern sequence data with secondary calibration dates, but with no rates reported (Burbidge *et al.* 2003). The mutation rate from neognathe birds is also not appropriate for comparison to the Emu due to the higher basal metabolic rate in neognathes compared to the ratites (possibly due to lack of the large pectoral muscle mass; Maloney 2008). Therefore, the Moa mtDNA *CR* mutation rate of Bunce *et al.* (2009) is the most appropriate for comparison to the Emu.

However, even taking the sources of error involved in generation times and evolutionary rate into account, the timing of the range expansion in the Emu Eastern Ancient population (349 kya, with 210-434 kya 95% quartiles) was found to match that of the Eastern Modern population relatively well (348 kya, with 226-439 kya 95% quartiles; see chapter 3, Table 3 and Figure 6). The similarity in expansion times suggests that the signal in the ancient Eastern population reflects the same events as the signal in the modern Eastern population, providing support that the expansion event is accurate. The significance of the concordance of signals is a question that cannot be evaluated without more specimens from each of Eastern Ancient populations. As all known King and Kangaroo Island Emu bone remains have been sampled from Australian museums (either as part of the study in Chapter 3, or in Heupink *et al.* 2011) and the few Tasmanian Emu remains held by the British Museum of Natural History in Tring have been sampled in chapter 3, specimens would be needed from museums elsewhere in the world.

In conclusion, the substantial errors associated with calculating the timing of expansions means that this test should not be used to draw conclusions in isolation. The combination of results from the Arlequin expansion tests with the BayeSSC analysis, the node ages from BEAST, and the landscape genetics plot of association between genetic diversity and longitude all suggest the same story of Emu range expansions out of the southwestern corner of Australia, in response to multiple climate ameliorations after glacial periods.

Linking dates to climate events

In chapter 3, the evidence of spatial expansions out of a southwestern refugium roughly coincides with interglacials, however the point estimate for the older expansion event actually occurs during a cold/arid period prior to the start of the MIS 9 interglacial (the error margins involved in calculating the expansion date is discussed above). However, there is a level of uncertainty involved in reconstructing the marine isotope stages (MIS) climatic cycles, which is done from Antarctic ice core isotope stratigraphy and planktonic foraminifera abundance data (EPICA *et al.* 2004; Spooner *et al.* 2011). These sources of error in ice core stratigraphy include

assumptions about snow accumulation rate and movement of ice over time in the models (Dreyfus *et al.* 2007), and in the lag between northern hemisphere insolation and Antarctic cooling (Kawamura *et al.* 2007), while the planktonic foraminifera abundance data is subject to errors in species identification and assumptions about conservation of the ecological niche of the foraminifera over time (Spooner *et al.* 2011). However, even given these minor uncertainties, the combined Antarctic ice core and planktonic foraminifera data for the southern hemisphere provides evidence that both interglacial periods in question (MIS 9 and MIS 5) were warmer than other interglacials, were associated with high rainfall and monsoon activity, and had more extreme interglacial-glacial transitions than experienced during any of the last 5 glacial cycles (EPICA *et al.* 2004; Spooner *et al.* 2011).

These climatic optima appear to have provided suitable conditions for the Emu to colonise previously uninhabitable landscapes. However, doubt exists as to how unsuitable the habitat of the Nullarbor Plain was to the Emu. Although most current Emu populations are concentrated in the more temperate south-west and south-east regions, this high dispersing species is currently found in dry desert regions, albeit only in very low numbers and usually following rainfall events (Marchant and Higgins 1990).

Other demographic events

Although the Bayesian Skyline Plots (BSP) shown in chapter 3 for the entire Emu dataset showed interesting signs of a demographic expansion ~ 5 ya, the wide error margins (1-10 kya) confounded potential hypotheses about events promoting the large demographic expansion in the Emu. A horizontal line drawn across the BSP remained within the margins of the demographic expansion, further supporting a constant population size. A Bayes Factor comparison for the demographic expansion against a constant population size showed strong evidence in support of a constant population size. Bayes Factors are a harmonic mean estimator of the marginal likelihoods of different models. The reliability and repeatability of model classification from Bayes Factors has recently been subject to debate, with a constant population size often the model erroneously selected as a better fit to the data (Baele *et al.* 2012). Two new marginal likelihood estimators, path sampling (PS) and stepping-stone sampling (SS) have been found to be consistently more reliable, but are computationally intensive (Baele *et al.* 2012). The version of BEAST released in the last few months (v1.7.1) does apparently allow the PS and SS marginal likelihood estimators to be compared between models, but the datasets in this thesis have yet to be repeated using the new release of BEAST.

If this demographic expansion does prove to be reliable, then it is interesting to speculate about possible causes. The timing of the demographic expansion roughly coincides with the

introduction of the dingo (*Canis familiaris*) to Australia (c. 5 kya; Savolainen *et al.* 2004), and slightly predates the extinction of the Thylacine and Tasmanian Devil from the mainland (3 kya; Gale 2009). The dingoes arrival is unlikely to have played a direct role in the Emu expansion however, as this predator has been found to regulate Emu population size quite significantly (Pople *et al.* 2000; Gale 2009; Fillios *et al.* 2010). If the Thylacine was also a predator of the Emu, its extinction may have precipitated a demographic expansion in the Emu. However, studies of Thylacine jaw dynamics suggest it likely hunted smaller prey as an ambush predator (Jones and Stoddart 1998; Wroe *et al.* 2007; Figueirido and Janis 2011), although Thylacines may have preyed on Emu chicks. Evidence demonstrates that the two species hunted with different predator styles, suggesting dietary niches may not have completely overlapped (Figueirido and Janis 2011).

The impetus behind the Emu demographic expansion may be a more complex scenario involving a type of meso-predator release. An example of meso-predator release occurred when an apex predator was substantially reduced in the Australian ecosystem after European colonisation (*e.g.* the dingo), which allowed meso-predators (middle tier predators such as the foxes and cats) to increase in abundance, and predate to a larger degree on small prey species (including small mammals of a critical-weight range 1-5 kg) (Short and Smith 1994; Johnson *et al.* 2007). However historically, as the Thylacine was declining and subsequently became extinct, the dingo stepped into the role of apex predator (Figueirido and Janis 2011). Although roughly similar in size, the two species have slightly different niches: the Thylacine consumed terrestrial animals 1-5 kg in size, *e.g.* small mammals like dasyurids, bandicoots, bettongs, possums *etc* (Jones and Stoddart 1998), in contrast to the dingo that today hunts large-medium sized macropods (Brook and Kutt 2011). In the short term, the decline and extinction of the Thylacine likely resulted in an increase in small mammal prey, which may have allowed small early dingo populations to shift dietary niche slightly and take advantage of the abundant prey. A surge in dingo numbers over time may have subsequently suppressed small mammals to basal levels, continuing until European colonisation when dingos faced its own apex predator. If these small mammals (mostly insectivores) were being suppressed by dingo predation over the last few thousand years, insect numbers may have exploded allowing Emu population sizes to expand in response. The top-down approach of the meso-predator release theory is likely not the only factor involved in the Emu expansion though. Bottom-up theories about anthropogenic and natural ecosystem changes may also have played a role in the Emu expansion, *e.g.* fire-stick farming and increased monsoonal weather patterns.

The timing of the Emu expansion roughly coincided with sea level rise (8.8–6.5 kya), the creation of substantial mangrove swamps along the coasts, and major rivers experienced increased water levels (8 kya) (Smith *et al.* 2008; Bird *et al.* 2010). It was also a time of temperature stability and homogeneous climatic conditions (warmer and wetter than modern) with increased monsoonal activity in the western Simpson Desert at 7–5 kya, and the Flinders Ranges at 8.8–5.3 kya (McCarthy *et al.* 1996; Shulmeister 1999; Quigley *et al.* 2010). By around 5 kya the El-Nino Southern Oscillation (ENSO) was reaching modern variability patterns with major periods of below-average precipitation (at ~5.5 kya), and increased dust production (~4.8kya), with Lake Mega-Frome drying out and again being dominated by playa conditions (~5kya) (Shulmeister 1999; Marx *et al.* 2009; Cohen *et al.* 2011). It therefore appears that during periods of higher effective precipitation (prior to 5kya), the Emu is starting to increase to its large late Holocene population size. However, it persists at this large population size into modern times even during the major periods of below-average rainfall (associated with the ENSO cycles).

A combination of top-down and bottom-up factors probably led to the late Holocene increase in Emu population size.

Significance of the work

The main significance of the Emu study is two-fold: this work provides the only study of current genetic diversity for this species; and the inclusion of the ancient samples offers a temporal aspect for detecting likely events responsible for creating this lack of genetic diversity. If purely modern Emu samples were used in this analysis, the last glacial maximum would be considered the more likely period during which the Emu experienced a population bottleneck responsible for the species' current reduced genetic diversity. In addition to the island calibrations, the ancient bone samples provide additional temporal calibrations for the coalescent analyses to date the post-bottleneck expansion events. These ancient island samples also provided another view of the taxonomic status for the island species. Without the ancient DNA from the Kangaroo Island Emu, the morphological analysis of this island dwarf would stand as the definitive evidence for the taxonomic classification of *Dromaius baudinianus* as a separate species.

The genetic potential within ancient eggshell is one of the few ways that past (and potentially lost) genetic diversity can be used to investigate the response of Australian species to past climatic cycles. Although only five Emu eggshell samples yielded DNA, eggshell still offers the greatest time depth for examining the past responses of Australian species to climatic cycles: 12,500-16,500 ya BP in chapter 3, but also 19,000 ya BP in Oskam *et al.* (2010). The lack of both genetic diversity and strong phylogeographic patterning in the Emu is symptomatic of many

Australian species (Houlden *et al.* 1999; Zenger *et al.* 2003; Brown *et al.* 2006). The results from the Emu study (chapter 3) may also narrow the focus on areas and time periods where other Australian species have experienced loss in diversity. For example, the phylogeographic barriers and particular interglacials highlighted in chapter 3 as significantly influencing the Emu's evolutionary history may also have provided opportunities for other species to expand after glacial-associated-bottlenecks. Further investigation of the evolutionary history of Australian species through a genetic survey of the fossil record, would provide critical information on the ability of the unique Australian biota to respond to past and future climatic change. However, no other material than eggshell is likely to provide preserved DNA over long time-frames, and inference methods *(e.g.,* based on the coalescent) would need to be implemented to reconstruct the evolutionary history of non-bird species. Furthermore, comparison of the evolutionary histories amongst different Australian species can be made to obtain an overall picture of those aspects of population dynamics that are shared and which differ between Australian taxa, like which barriers to gene flow have impacted high vs. low dispersing taxa, and which regions have provided refugia for range restricted vs. widespread taxa.

Furthermore, the low genetic diversity found within the Emu appears to stem from genetic bottlenecks caused by contraction back to a refugial zone in the southwestern corner of the continent during arid glacial cycles, prior to range expansions out of this refugia during interglacials (MIS 5 and MIS 9). One potential consequence of this low genetic diversity is a potential susceptibility to disease, such as seen in Tasmanian Devils (*Sarcophilus harrisii*) with facial tumour disease (Hawkins *et al.* 2006; Miller *et al.* 2011). In the case of the emu, the apparent lack of genetic variation (especially if the mitochondrial control region's lack of diversity is mirrored in immune response genes as seen in the Tasmanian Devil; Siddle *et al.* 2007) might suggest that despite being a common and widespread species, the Emu is at risk of similar strong environmental challenges, such as readily communicable disease.

Future directions

Although a lack of genetic diversity was found in the mtDNA *CR* of the Emu, it is likely that there is more diversity within the nuclear genome. The only study of microsatellites in the Emu found relatively high levels of diversity, however the degree to which this stems from purely post-expansion expansion or also reflects pre-bottleneck diversity is unknown (Hammond *et al.* 2002). With the development of NGS technology and advances in ancient library preparation methods (Knapp *et al.* 2012; Meyer *et al.* 2012), further analysis of nuclear diversity from the modern, museum and ancient samples obtained for this study will be able to build upon the

mitochondrial work reported here. There is potential to use nuclear single-nucleotide polymorphism (SNP) chips designed for modern domesticated species to hybridise aDNA libraries, and examine ancient diversity at these SNP sites to look further back beyond more than the last bottleneck. Where this has been done previously with the BovineSNP50 BeadChip by Illumina (designed for domestic cattle), species as distantly related as giraffes have been successfully targeted (Decker *et al.* 2009). Although the giraffe and bovids diverged up to 29 Mya, the drop-off in genotype call rate by using the bovineSNP50 chip for the giraffe was only about 35%. Luckily, the same USDA funding that financed the cattle SNP chip also partially funded the production of a 60K chicken SNP chip by Illumina (Illumina Inc., San Diego, CA) (Fulton 2012). The Illumina 60K chicken SNP chip could be used to generate Emu nuclear data relatively easily and cheaply, although with some anticipated loss of genotype calls due to the more distant evolutionary relationship between chickens and Emus (diverged 73-98 Mya; Harlid *et al.* 1997). By transferring this technique to the 60K chicken SNP chip, even greater resolution to the timings of the Emu post-bottleneck expansions of the MIS 5 and MIS 9 interglacials could be obtained.

The study in chapter 3 highlights how aDNA can be incorporated with contemporary sampling of museum tissue collections to investigate taxonomic issues and the past responses of species to climate change. Many other taxa can also be examined in this way, although few have the DNA preservation potential of eggshell. Other species with the potential for genetic analysis of eggshell preserved in the Australian archaeological record include the Wedge-tailed Eagle and the Malleefowl. However, in order to sample ancient eggshell from the Wedge-tailed Eagle or Malleefowl accurate identification of the eggshell to species level would be required, which is not a simple task from cave deposits as the eggshell is often found as very small fragments, making morphological identification difficult. *A posteriori* identification using DNA is another possibility, however this requires processing a lot more samples from non-target species than is necessarily. If studying eggshell from other Australian species is not feasible, then bone remains from large animals are the next most viable sample type for animals, *e.g.,* kangaroos, wallabies, echidnas or possums. The large bones of these animals would have a higher chance of preserving aDNA, and as their modern ranges cover the more temperate east coast and Tasmania, caves in this region may have preserved enough bone remains for aDNA study. Furthermore, many of these species are also lacking in modern genetic diversity, providing an opportunity for aDNA to reveal the processes involved beyond the last bottleneck.

Chapter 4: the Ghost Bat (Macroderma gigas)

Issues Encountered

Although one of the problems encountered in this study was a methodological issue related to the preservation of ancient DNA in the Australian context, the lack of radiocarbon-dated samples has also limited the accuracy of conclusions drawn about the timescale of gene flow in the Ghost Bat. These and other issues encountered in this study are detailed below.

Sample Type

One of the initial aims of this study was to analyse scat remains of the Ghost Bat to examine both the predator (Ghost Bat) population dynamics, but also the prey (rodent) species identification. When pilot work to establish the feasibility of successfully extracting either Ghost Bat or prey DNA from dried scats failed to yield anything other than bacterial DNA, this part of the study was abandoned.

Calibrations

In contrast to chapter 3, the Ghost Bat dataset was lacking radiocarbon-dated samples completely as the external funding for this study covered fieldwork only. The ancient samples consisted of subfossil bone/teeth remains from the surface of cave deposits and museum specimens collected as surface finds in the 1900's, with the principal of superposition suggesting these surface specimens were the last representatives of late Holocene populations. However, without directly dated samples, either fossil calibration points or a molecular rate estimate were required to establish a temporal aspect to the Ghost Bat evolutionary history.

Many other studies still use the average 2%-per-million-years divergence rate in animals and birds to calibrate molecular phylogenies (Eberhard and Bermingham 2005; Gill *et al.* 2005; Bollmer *et al.* 2006) without thought to the fact that this is a rough estimate calculated over 30 years ago, and based on an average across many species (Brown *et al.* 1979). More recent research has now established that the mammalian mutation rate varies across species (by as much as 20-fold for mtDNA third codon position) in accordance with many life history traits, including body mass, age at sexual maturity, and longevity (Nabholz *et al.* 2008). These inter-specific differences have substantial impacts on the fields of molecular phylogeny, molecular dating and population genetics. In fact, many papers follow a circular yet convoluted argument for using mutation rates from the literature or for comparing their own calculated evolutionary rate to others in the literature across species with substantial variation in life history traits. For example, a study on bats by Petit *et al.* (1999) uses a rate from Lopez *et al.* (1997) on felids that uses calculations from an earlier paper by Lopez *et al.* (1994), which in turn uses a divergence rate calculated from an estimate of cytoplasmic mutation rates by Hasegawa *et al.* (1985) compared to

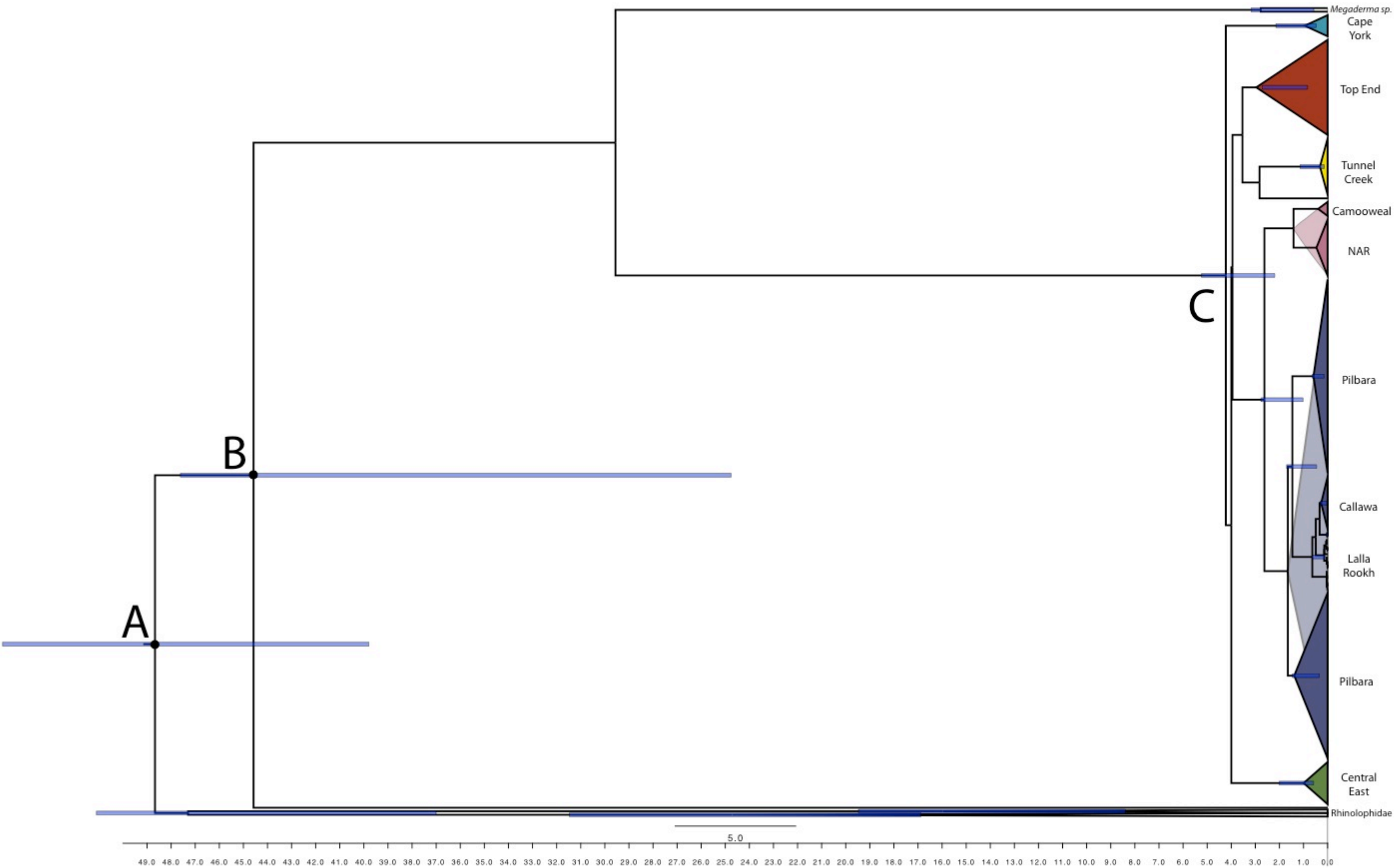
an estimate of nuclear pseudogene mutation rates by Li *et al.* (1981). Petit *et al.* (1999) then go on to say that the rate calculated in their analysis is in general agreement with other *CR* rates in humans, *Mus musculus,* Snow Goose, and Shrews, yet the recent paper by Nabholz *et al.* (2008) find that Chiroptera (bat) species evolve twice as slow as Rodentia (rodent) species. Now that we are gaining an understanding about the error involved in assuming constant molecular rates across species (*i.e.,* 'strict molecular clocks'), this entire chain of transferring rate estimates across divergent species and between different genes is highly suspect.

Therefore, two different approaches to calibrating the Ghost Bat mtDNA *CR* dataset were compared: 1) using the 225 Ghost Bat (*Macroderma gigas*) sequences and 2 outgroup sequences (*Megaderma lyra*), with a secondary calibration point of 16.5 Mya (± 2.75 Mya) on the split between the *Macroderma* and *Megaderma* genera calculated from Teeling *et al.* (2005); and 2) using an additional 228 sequences of *Hipposideros* and *Rhinolophus* sequences so that the minimum divergence date of 37 Mya for the divergence between *Hipposideros* and *Rhinolophus* genera, and a maximum divergence date of 55 Mya for the root of the Rhinolophoidea tree. The first approach is detailed in chapter 4, but the second approach is described below.

BEAST vs. *BEAST: the Ghost Bat

The recent development of the *BEAST (pronounced StarBeast) analysis within BEAST allows the inclusion of multiple loci across multiple individuals per species to co-estimate many gene trees within a shared species tree in combination with the demography of both extant and ancestral species (Heled and Drummond 2010). However, there is no reason why one locus for a large number of individuals cannot be used in *BEAST to improve resolution, as an increase in the number of samples increases the number of coalescence events being used to estimate population size for each species. A trial of *BEAST on the Ghost Bat dataset included increasing the resolution of the Yinpterochiroptera suborder by including *Rhinolophus* (Horseshoe bats) and *Hipposideros* (Roundleaf bats) sequences, which allowed two primary calibration points to be used to date the tree instead of the single secondary calibration point in the BEAST analysis (see Figure 6 for the *BEAST tree). Primary calibrations use fossil dates directly, while secondary calibrations are those derived from the nodes on a tree that was calibrated using fossil calibration points. Therefore, primary calibrations are considered better than secondary calibration points to calibrate a phylogenetic tree, as secondary calibrations include any errors associated with the primary calibration points plus those methodological errors associated with reconstructing the tree and assumptions about the accuracy of the primary analysis. However, the *BEAST analysis yielded a much slower evolutionary rate (1.47% per Million years) than the BEAST analysis (4.22% per Million years), which was more consistent with studies of other bat species (4.6% per

Million years in *Mystacina tuberculata*) (Lloyd 2003). The slower rate was assumed to stem from the two primary calibration points being much deeper within the tree (37 and 55 Mya) than the secondary calibration point (16.5 Mya), in accordance with the time dependence of molecular rates (Ho *et al.* 2007; Ho *et al.* 2011a). Although the secondary calibration date used in the BEAST analysis is not ideal it appears to give a more realistic molecular rate of the fast evolving mtDNA *CR*, and the divergence dates are therefore assumed to be more accurate too. Therefore, the only way to ensure reliable calibration dates for the Ghost Bat phylogenetic tree and estimate the clade divergence dates accurately, is to radiocarbon date the NAR bone/teeth samples. At the present time this is prohibitively expensive, but there is the possibility of future collaboration with external funding bodies, such as mining companies and the Australian Institute of Nuclear Science and Engineering (AINSE) research awards.



Megaderma sp.
Cape York
Top End
Tunnel Creek
Camooweal
NAR
Pilbara
Callawa
Lalla Rookh
Pilbara
Central East
Rhinolophidae
A
B
C
5.0
49.0 48.0 47.0 46.0 45.0 44.0 43.0 42.0 41.0 40.0 39.0 38.0 37.0 36.0 35.0 34.0 33.0 32.0 31.0 30.0 29.0 28.0 27.0 26.0 25.0 24.0 23.0 22.0 21.0 20.0 19.0 18.0 17.0 16.0 15.0 14.0 13.0 12.0 11.0 10.0 9.0 8.0 7.0 6.0 5.0 4.0 3.0 2.0 1.0 0.0

Figure 6. *BEAST phylogenetic tree with 225 sequences of *Macroderma gigas,* 2 sequences of *Megaderma lyra,* and 228 sequences of outgroup (Rhinolophus and Hipposideros). The calibration points used in this *BEAST analysis are those shown at A: maximum of 55 Mya (calculated at 47.27 Mya, 39.79-55.00 Mya 95% HPD) and B: minimum of 37 Mya (calculated at 35.62 Mya, 24.76-47.62 Mya 95% HPD). Other node dates calculated in this *BEAST analysis include C: the tMRCA *Macroderma gigas* at 3.58 Mya (2.20-5.23 Mya 95% HPD).

Assumptions

One key assumption in chapter 4 is that the ancient samples collected from the surface of cave floors represents the most recent (historical) populations of Ghost Bat that inhabited the cave. This assumption is based to a large degree on the speed at which cave surface material gets turned over and buried, which in turn depends on the type of cave it is (open mouthed, pitfall, *etc.*), how accessible the entrance is, and therefore the degree of animal/human presence in the caves (see Figure 7 for example of inaccessible cave in the Flinders Ranges). All of the NAR caves were readily accessible by animals and humans, yet had abundant Ghost Bat (and other) subfossil material around the edges of the cave (*i.e.,* slightly protected from damage by footfalls of larger animals). Another indication that these were relatively recent remains is the fact that aDNA was successfully extracted from even the long thin wing bones (metacarpal or phalanges, see Figure 8), in addition to the teeth that are generally a better source of aDNA (see Figure 9).

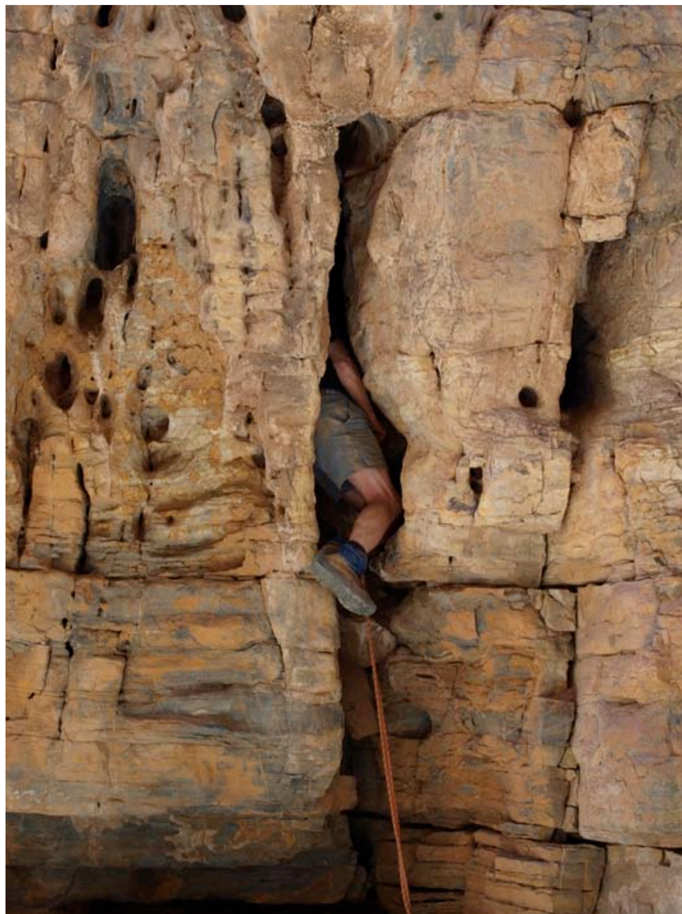


Figure 7. Example of a highly inaccessible cave in Chambers Gorge, Flinders Ranges, South Australia where subfossil remains are visible on the cave floor surface with little disturbance by animals (photo credit: Aaron Camens).

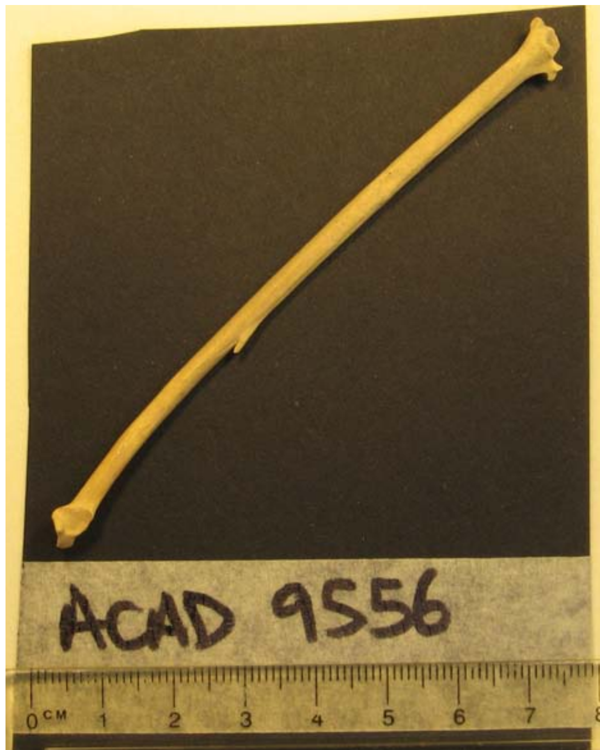


Figure 8. Ghost Bat wing bone (metacarpal or phalange) from individual E08 from the NAR population that yielded DNA in 158bp and 217bp fragments. The ability to amplify these sized DNA fragments (in particular the 217bp fragment) from such a thin bone suggests the Ghost Bat bone and teeth remains found on the surface of the NAR caves most likely represent the most recent (historical) populations inhabiting the E24 and J03 caves.

Figure 9. Ghost Bat RH partial mandible with teeth from individual J04 from the NAR population that yielded DNA in 158bp and 217bp fragments.

Sample size

Only three extinct populations were sampled in chapter 4: the NAR caves, the central Australian population, and the Flinders Ranges population, with the last two populations having only n=5 and n=3 samples, respectively. Although the extraction of ancient DNA was attempted from more samples, the poor preservation condition of the mummified remains from the Flinders Ranges (33%) were likely responsible for the low success rate. The central Australian population had a higher success rate (63%), largely because the samples were mummified cartilage from inside the nasal cavity of museum-prepared skulls, and were therefore protected from the external environment prior to collection and the ambient museum environment after collection. The subfossil remains from the NAR populations had a slightly higher success rate (65%), suggesting teeth and bone remains are the ideal sample type for small mammals in the Australian context. However, museums are less likely to allow destructive sampling of bone and teeth remains over scraping nasal cartilage out of the inside of the rostrum where any damage cannot be seen.

At many of the caves where populations have become extinct, live Ghost Bats have been seen within the last 200 years, highlighting the recent nature of the extinctions in these regions. However, other extinct populations have been collected as subfossil material and are held in various museums, such as the type specimen from Mt Margaret (Dobson 1880) and other remains from caves in eastern NSW. Uncertainty does exist about whether Ghost Bat remains have been collected from caves in the Rawlinson and Kathleen Ranges in central WA; from Yanchep-Wanneroo, south of Perth; and from south of the Flinders Ranges in South Australia (Molnar *et al.* 1984), and possibly held in museum collections overseas. Some of these localities are critical for resolving questions raised in chapter 4 about the timing of gene flow that links populations within the central East and within the central West regions (discussed in chapter 4).

Expansion tests from Arlequin

With the Ghost Bat dataset, each extinct population was assumed to represent the same population (*i.e.,* the most recent population inhabiting the NAR caves, the CA caves, and the FR caves). Therefore, similar to the Emu study, no evolution is assumed to have taken place between the deaths of each of the individual Ghost Bats within the set of samples designated as a 'population'. If these assumptions are not valid, and some of the sampled NAR individuals were from 10,000 ya and others from only 100 ya, then there will potentially be issues with the accumulation of mutations in the intervening 9,900 years, leading to an overestimation of the likelihood of expansions following a drastic bottleneck (Depaulis *et al.* 2009). However, given the well-preserved nature of the DNA in these samples and the fact that mummified skin and nasal cartilage yielded aDNA from central Australia, it is unlikely that the samples span a sufficient temporal depth to cause substantial problems with the calculation of the population

genetic statistics used in the Arlequin expansion tests (*i.e.,* number of segregating sites, S). Another point to note is that the tMRCA (origin date) calculated for each ancient population using BEAST is always prior to the predicted expansion events, further supporting the idea that populations were experiencing demographic or range shifts around these times. Furthermore, without money to fund carbon dating of these samples, this issue cannot be resolved one way or the other.

Further errors were also potentially introduced into the transformation of the coalescent time since expansion to real time by the estimate used for Ghost Bat generation times. As no ecological studies of Ghost Bat populations have been undertaken due to the sensitivity of the species to disturbance, the age at first and last breeding had to be estimated from captive populations (Jones 2008). As captive bats are not subject to the same selective pressures (e.g., predation, access to food, etc.) as wild populations, the ages at first and last breeding may be inaccurate representations of generation time in wild populations. However, a more accurate estimate of generation time cannot be gained without significant disturbance to this vulnerable species.

Significance

Chapter 4: the Ghost Bat highlights the ability of ancient DNA to substantially increase resolution of studies investigating the evolutionary history of vulnerable species. Previous work on the population genetics of extant *M. gigas* populations found higher genetic diversity at lower latitudes (Pine Creek/Claravale vs. Rockhampton/Cape Hillsborough) and monophyly of regional clades, however without the additional information from the extinct populations little would be known about gene flow at the continental scale or the divergence times of clades. Previous studies of modern populations of the Ghost Bat found a polytomy at the base of the phylogenetic tree that resulted in a lack of resolution in the relationships between the major colonies (Worthington Wilmer *et al.* 1994; Worthington Wilmer *et al.* 1999). With the inclusion of ancient DNA from extinct populations, the relationship between these major colonies can now be resolved. Furthermore, the placement of the extinct populations in this phylogenetic tree provide insight into potential gene flow events, which in turn could be aligned with the ENM predictions about the timing of climatically 'suitable' corridors across the landscape. If only modern DNA samples were available for inferring patterns of gene flow, one would have to surmise that the extinct populations would have been more closely related to the more geographically proximate colony. However, by including the ancient samples a more accurate (and yet surprising) representation of gene flow between extant and extinct populations has emerged.

With respect to the modelling of the Ghost Bat's response to past and future climate change, this work has the potential to influence management decisions made by both mining

companies and government departments. However, this would be premature given the extent to which the climatic models used in the Ghost Bat study are based on many unverified assumptions linking ambient climate to cave microclimates.

Future work

In the short-term

Additional sampling from these un-sampled extinct populations at Mount Margaret in Queensland and in eastern NSW would be useful in resolving the question about gene flow within the central Eastern region – between the Rockhampton/Cape Hillsborough colonies in Queensland and those in central Australia/the Flinders Ranges (Figure 10). Similarly, sampling from the Rawlinson and Kathleen Ranges in central WA, and from the Yanchep-Wanneroo colonies south of Perth would be useful in resolving questions about gene flow in the central Western region – between the NAR populations in WA and those in Camooweal, Queensland (Figure 10). The inclusion of these populations in the ecological niche modelling (ENM) may also add resolution to the hindcasting of suitable Ghost Bat habitats into the last interglacial and forecasting to 2030 and 2080 to predict the Ghost Bat's ability to respond to future climate change.

In the long-term

Like the Ghost Bat, there are other similarly vulnerable bat species that roost in caves which may lend themselves to ancient DNA study of past populations. The Orange leaf-nosed bat (*Rhinonicteris aurantius*) is one example of a cave dwelling bat that is listed as vulnerable, and where analysis of its evolutionary history and ecological niche may provide knowledge about patterns of past gene flow and the effect that changing climates may have had on past, and will have on future, distributions.

With more nuclear and multi-locus datasets being developed, and as Next Generation Sequencing (NGS) technology is used more routinely, clarity back even further in time will be available for many species. NGS has not yet been developed to its full potential though, with many early studies using the power of NGS to generate data for other purposes, such as the transcriptomic dataset for the Glanville fritillary butterfly, *Melitaea cinxia*, sequenced using 454 pyrosequencing that will be the basis for further ecological studies using microarrays to examine levels of gene expression for key functional genes (Vera *et al.* 2008). However, the application of NGS to microarray development has potential benefits to the Ghost Bat, as examination of functional genes associated with thermoregulation could allow investigation of the assumption of constancy in ecological niche that is fundamental to the ENM used in chapter 5.

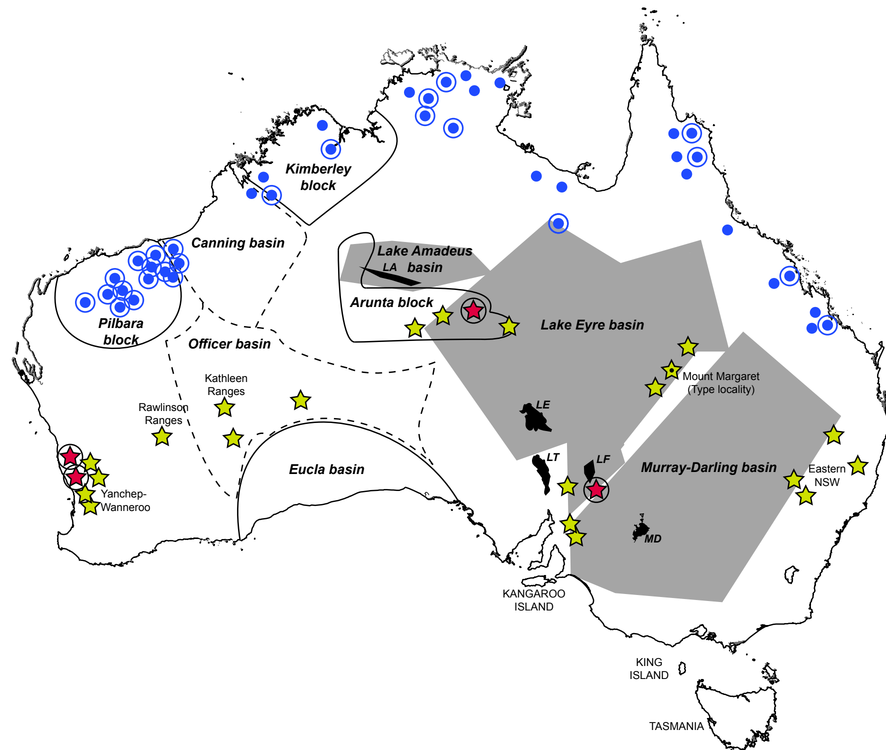


Figure 10. Past and present distributions of the Ghost Bat (*Macroderma gigas*). Blue circles represent extant populations; those with an outer ring are those sampled by Worthington Wilmer (Worthington Wilmer *et al.* 1994; Worthington Wilmer *et al.* 1999) and included in this study. Stars represent extinct populations; circled red stars are the extinct populations sampled in this study; yellow stars are extinct populations not sampled in this study, with a central dot indicating the type locality (Mount Margaret).

Chapter 5: the Polynesian Chicken (Gallus gallus)

Although chapter 5 reconstructs more recent 'patterns' of genetic diversity, the phylogeographic signal in this commensal allows testing of hypotheses about Polynesian contact with South America that would otherwise be obscured by the dispersal of modern European chickens throughout the Pacific, and the risk of DNA contaminating lab reagents (Storey *et al.* 2007; Storey *et al.* 2010). Ancient chicken DNA is currently a key means to investigate this process, due to the lack of resolution in other cultivated/commensal species in the Pacific (Matisoo-Smith *et al.* 1998; Larson *et al.* 2007).

Issues Encountered

Again the main problems encountered in this study were those typical of many ancient DNA studies: methodological issues related to the preservation of ancient DNA (in this case in hot tropical island environments where bones were retrieved from exposed alkaline sand

dunes/middens); and low sample sizes due to poor DNA yield *etc.* These issues (and others) are outlined below.

Sample size

It is well known within the ancient DNA field that sample sizes will always be lower in aDNA studies than those in modern phylogeographic analyses. There will inevitably be a reduction from the numbers of ancient samples analysed (n=37 ancient chicken samples extracted) to those that are successfully amplified and sequenced (n=22). However, by increasing both the sample size from each island and the number of islands sampled, a more representative picture across the Pacific can be reconstructed. Increasing the sampling of chickens in Micronesia is of particular importance. Even though few early commensal remains have been found in the Micronesian archaeological record, it is the comparison of these few early samples with those of later commensals that will provide evidence about whether Micronesia was part of the initial migration route, and about later trade links between Micronesia and other island groups (Addison and Matisoo-Smith 2010).

Replication

The poor DNA preservation in ancient samples has an impact not only on obtaining sequence data, but also with the ease of repeated PCR amplification, cloning, Sanger sequencing, *etc*. In many instances, internal replication (repeated PCR/sequencing within the same lab) was difficult – a good indication that external replication (repeated extraction, PCR, and sequencing from the same sample in another ancient DNA lab) will also require repeated efforts.

Independence of sampling

The issue of independence of sampling is normally assisted by the stratigraphic control across archaeological sites, which ensures that the provenance of each sample is known and the likelihood of multiple bones belonging to the same individual can be assessed. However, the Hawaiian archaeological site at Makauwahi cave is unique. The regular deposition of stratigraphic layers at Makauwahi cave was interrupted by a tsunami (seen as a layer of high-energy sedimentation) that dated to approximately 1430-1665 A.D. Evidence from the archaeological record shows this high-energy water disturbed material remains belonging to Unit V (at that time Unit V was the surface of the cave floor), and deposited a sedimentary layer as Unit VI (Burney *et al.* 2001). Although all of the chicken bones analysed in chapter 5 were excavated from Unit V (which was disturbed by the tsunami), the archaeologists have reconstructed a robust chronology of the site from radiocarbon dated samples (the site shows internal consistency from 33 of the 34 radiocarbon dates; Burney *et al.* 2001). However, the bioturbation within Unit V may result in uncertainty with regards to independence of sampling of the chicken bones. As all seven of the ancient Hawaiian chicken bones fall within the same 330

bp mtDNA *CR* haplotype, theoretically this may represent only one individual. However, most of the focus of the study in chapter 5 is on eastern Polynesia and the link to South America, and on the initial introduction routes from western Polynesia and Micronesia. Therefore, whether or not the seven Hawaiian samples actually represent only one individual is not critical to the focus of this study.

Significance

The study on Polynesian chickens in chapter 5 highlights how a lack of well-preserved DNA, in addition to contamination by modern commensal DNA, can completely alter the conclusions drawn about research.

The chicken shares a long history of domestication with human cultures and is deliberately bred for specific physical traits for cultural or commercial purposes: Chinese and Japanese cultures have used the feathers of chickens, and their ability to sing or fight in cultural practices since at least 2.5 kya (Komiyama *et al.* 2004); and commercial breeders have selected for increased growth rates in broiler breeds (Rauw *et al.* 1998) and earlier maturity and larger egg size in egg laying breeds, such as the White leghorn (Jensen and Andersson 2005). This selection for phenotypic traits during domestication has potentially caused genetic bottlenecks in some breeds (long-crowing chickens, Komiyama *et al.* 2004) but introgression with wild junglefowl in other instances has ensured higher levels of diversity (cock-fighting breeds, Silkies, and the White Leghorn have introgressed with Green Junglefowl, Sawai *et al.* 2010). Due to this highly domesticated history, little is known about the traditional breeds of pre-Colonial era native cultures. The Polynesian chicken is one of the few examples where traditional genetic patterns can still be observed in an otherwise highly domesticated species. If the D_{AP} mitochondrial haplotype can be used to identify chickens on Pacific islands with also purely Polynesian nuclear DNA, then this will be one of the few examples of pre-Colonial wild-type chickens, and can potentially be used as a genetic reservoir of chicken diversity.

Furthermore, as the Polynesian chicken embodies the traditional Polynesian sport of cock-fighting, or *faatitoraamoa* in Tahitian (Halloran and Halloran 1969), a pure Polynesian chicken could become a symbol of traditional Polynesian lifestyle to be preserved. Pre-historically, Tahitians had songs and religious traditions (including *Ruaifaatoa,* the god of cockfighting) connected to *faatitoraamoa,* which was performed for the amusement of the owners (Ellis 1831). If the D_{AP} sub-haplogroup that was identified from the Polynesian chicken in this thesis can be used to locate the original Polynesian strain of chicken on some remote Pacific Island, like the eastern Solomon Islands, which has not experienced major European settlement (Parker n.d.), this iconic bird can potentially be used to reinvigorate interest in traditional, and in turn healthier, Polynesian lifestyles. Encouraging the breeding and raising of chickens will greatly improve the

health of Solomon Island society by increasing dietary protein levels in the predominantly rural, and substantially protein depleted and malnourished, islander populations (Glatz *et al.* 2009). Thirty per cent of Solomon Islander infants are underweight due to poor nutrition and yet only one chicken and a few eggs are eaten on average by Solomon Island families per month (Glatz *et al.* 2009). If the re-introduction of the Polynesian chicken to Solomon Island society can increase enthusiasm for more efficient husbandry efforts (e.g., providing food, water and secure housing to free range chickens vs. requiring the chickens to scavenge to survive) and increase the flock size of the average family, then dietary protein will increase as a by-product of the increased social status and financial gain of owning chickens and fighting roosters (Glatz *et al.* 2009).

Furthermore, by regaining this (potentially) lost genetic diversity of the Polynesian chicken for introduction into the commercial breeding stock of chicken farmers worldwide, an important food source will gain genetic diversity, thereby reducing susceptibility to diseases and significantly increasing its ability to survive future worldwide pandemics, such as the outbreak of Avian influenza virus (AIV) in 2003. Recent genomic work on commercial chicken breeds (*e.g.,* Ross broilers, White Leghorns) has highlighted the loss of 70% of genetic diversity from individual commercial breeding lines, of which up to 25% cannot be regained by crossing commercial breeds as the diversity was lost prior to the development of the commercial chicken industry (Muir *et al.* 2008a). The bottlenecks involved in the generation of commercial chicken breeds include selecting for large body size in broiler lines, and increased egg numbers and size in layers. The development of more efficient broiler and laying breeds has allowed breeders to address the sharp increase in demand over the last three decades (chicken meat production has increased 436% since 1970, and with more than 55 million metric tons of chicken eggs now produced annually; Muir *et al.* 2008b). However, the lack of genetic diversity in such a significant commercial food source is a crucial biosecurity risk, similar to the corn blight in the USA in 1970-1971 that stemmed from widespread use of an inbred Texas male-sterile cytoplasm, and resulted in a loss of food energy many times greater than that lost in the famine-producing potato-blight epidemic of the 1840s that swept across Europe (Ullstrop 1972). With the recent outbreak of AIV in 2003, the issue of a lack of rare alleles that might provide a resistance to certain diseases in important commercial species has again been raised, with Muir *et al.* (2008b) emphasising that non-commercial and native poultry populations represent a reservoir of genetic diversity that potentially pose a possible way to prepare for future biosecurity challenges.

Future work

In the short-term

Hypotheses about links between modern South American chickens, and modern European (H_0) or modern Pacific chickens (H_1) can be tested using Simcoal and/or BayeSSC simulations

(see Figure 11 below; Excoffier *et al.* 2000). Although the significance tests (Binomial test and Fisher's exact test) used in chapter 5 argue for a European-South American link (1400-1500 A.D.), rather than a Pacific-South American link (1200-1500 A.D.), these tests are limited to the groupings input by the user. In effect the significance tests do not incorporate phylogenetic information about how closely related the haplotypes are to each other. However, more robust statistical conclusions can be drawn from Simcoal, as simulations of populations with specific genetic characteristics (*e.g.,* diversity levels) are used that take phylogenetic information into account. Furthermore, as there are at least 40 samples from each of the modern South American (n=41), European (n=56), and Pacific (n=63) populations, the level of phylogenetic information included in the population genetics statistics used by SimCoal should be relatively detailed (*i.e.,* small sample sizes should not confound any signal of shared ancestry).

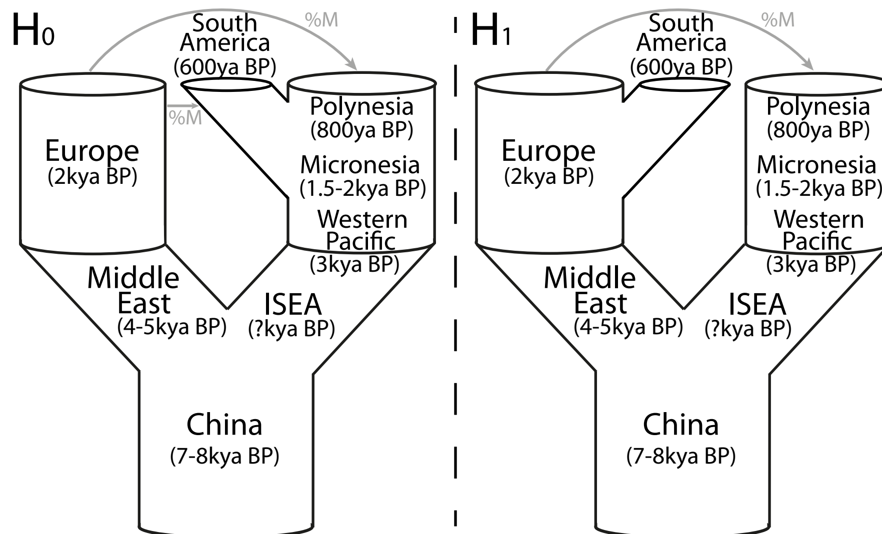


Figure 11. Hypotheses about genetic links between South American compared to either European or Pacific chickens that can be tested in SimCoal or BayeSSC. The null hypothesis is based on the current theories of Storey *et al.* (2008b) that Polynesians brought chickens to pre-Columbian South America (yet taking into account a certain level (%M) of historical gene flow from Europe to both South America and the Pacific); the alternate hypothesis is that modern South American chickens were introduced from Europe (with no gene flow from the Pacific to South America, yet a certain level of historical gene flow (%M) from European chickens into the Pacific). Dates for archaeological chicken remains in each of the localities shown come from the following references: China, Middle East, and Europe (West and Zhou 1988); Western Pacific (Storey *et al.* 2008a; Storey *et al.* 2010); Micronesia (Steadman and Intoh 1994); Polynesia (Storey *et al.* 2008b); South America (Storey *et al.* 2008b). Note that few chicken remains are known from archaeological sites in Island Southeast Asia (ISEA).

<u>In the long-term</u>

The recent revolution in sequencing technology (Next Generation Sequencing) has the ability to shed light on, and allowing finer-scale resolution of, questions about Polynesian history, migration routes and potential South American contact. Genomic scale research on commensals

is already underway in labs around the world, for example the Australian Centre for Ancient DNA received ARC funding to support a project entitled ‘Reconstructing the human colonisation of the Pacific using modern and ancient chicken DNA’; and J. M. Mwacharo at the University of Nottingham is currently sequencing whole mitochondrial genomes (WMG) of chickens from Indonesia, China, and Madagascar. It is hoped that the increased resolution provided by WMG data sets will be sufficient to track the chicken-human dyad, however in essence the mitochondrial genome acts as a single locus. Therefore, even the jump in sequence length from 350 bp of control region data to 16,775 bp in the entire chicken mitochondrial genome (Desjardins and Morais 1990) may not be adequate to resolve anthropological questions about migration routes in the Pacific. This is where the Illumina 60K chicken SNP chip, or other genomics technology, will also be useful in reconstructing the domestication history of the Polynesian chicken. At the genomic scale, not only could migration routes be tracked to a finer scale, but trade routes between individual islands could possibly be detected. Furthermore, the link between genotype and phenotype in the domestication process of the Polynesian chicken could be investigated by targeting those genes likely linked to traits selected for in the Polynesian chicken, such as aggression for cock fighting (*e.g.,* synaptosome, lipid metabolism and memory formation genes), egg colour (*e.g.,* biliverdin pigment gene on chromosome 1), and plumage colouration (*e.g.,* tyrosinase and melanocortin 1 receptor genes) (Bitgood 1985; Buitenhuis *et al.* 2009; Liu *et al.* 2010). The egg colouration genes in particular would be interesting to assess in the Polynesian chicken for comparison to the Araucana chicken breeds that are native to Chile and lay blue eggs. A shared history in genes like the biliverdin pigment gene may offer irrefutable evidence that the ancestors of the Araucana chickens were brought to South America from Polynesia. Another important locus to target with NGS would be the chicken equivalent of the major histocompatibility complex (MHC), which is a 92-kb region of the *B locus* with many genes similar to mammalian MHC (Kaufman *et al.* 1999). Assessing SNPs in the Polynesian MHC genes may identify a potential reservoir of genetic diversity useful in addressing future biosecurity challenges.

Future work in general

If the field of ancient DNA is to expand and capitalise on the genomics era, then data analysis and bioinformatics will need to become a bigger part of the skill set needed by aDNA researchers. The extraction and amplification methods used in ancient DNA have already been optimised substantially, so unless a new technological leap in DNA extraction methodology is seen similar to that in sequencing technologies (*i.e.,* NGS), the limitation now for aDNA studies is on optimal use of the data generated with NGS and the analysis techniques available for use

with the datasets. Future work on each of the organisms studied in this thesis will be able to utilise the vast array of information about individuals, populations and species that nuclear loci from both extant and ancient samples will offer.

The study by Depaulis *et al.* (2009) has highlighted many new possibilities for increasing the resolution of ancient genetic data, and in the future for taking advantage of the large amount of money being spent to generate ancient genomic data. These genomic datasets are being published in the public domain and yet they have had few population level analyses performed on them, therefore there is the potential to mine this huge database of information to answer specific questions related to population genetics, evolutionary biology, and phylogenetics at a deep resolution. The gigabases of data being generated from NGS runs will stimulate an even greater need to be able to incorporate heterochronous data into tests designed for synchronous data.

However, more work needs to be done to eliminate these heterochronous biases in population genetics statistics. For example, one way to eliminate this bias as a possible source of error in the expansion tests in Arlequin is to develop alternate formulas to estimate the number of segregating sites taking into account heterochrony, and derive an independent estimate of the mutation rate (which needs to be calculated from a different dataset otherwise a circular argument arises). Together with the formulas of Depaulis *et al.* (2009) for $\boldsymbol{\pi}$ (diversity estimator of $\boldsymbol{\theta}$), a new estimate of $\boldsymbol{\theta}$ (mutational parameter of the population, $\boldsymbol{\theta = 2Ne\mu}$) needs to be derived that takes into account the heterochrony of the dataset. Such equations, specific to heterochronous data, could be implemented in existing packages like Arlequin, and used to re-assess the expansion tests results obtained form ancient DNA data sets.

However, this is just one simple example of the need for heterochronous techniques, many more will develop as NGS is explored. Therefore, it appears that the future of aDNA is in bioinformatics.

Concluding remarks

Ancient DNA methods were used in this thesis to explore under-studied taxa in Australasia and provide novel information about the evolutionary history of the Emu, the Ghost Bat, and the Polynesian chicken. In chapter 3, I have provided insight into the evolutionary history of taxa (the Emu) lacking any previous knowledge of phylogeographic patterns, and have provided clarity regarding specific and sub-specific divergences. In chapter 4, I have expanded the view of a vulnerable species (the Ghost Bat) previously on analysed with modern data and have provided temporal evidence of the deep splits between populations. Lastly, in chapter 5, I have examined a group (the Polynesian chicken) with a long history of commensalism with humans, but that had doubt cast by previous studies about prehistoric links between populations.

The application of other analysis techniques to the aDNA data allowed similarities in species responses to past climate events to be identified (interglacial expansions in both the Emu and Ghost Bat), and the effects of differences in the mechanisms behind vicariance patterns to be highlighted (natural dispersal between mainland and island vs. human-mediated island hopping).

References

Addison DJ and Matisoo-Smith E (2010). Rethinking Polynesians origins: a West-Polynesia Triple-I model. *Molecular Ecology* 45: 1-12.

Anderson CNK, Ramakrishnan U, Chan YL and Hadly EA (2004). Serial SimCoal: A population genetics model for data from multiple populations and points in time. *Bioinformatics* 21(8): 1733-1734.

Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA and Alekseyenko AV (2012). Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Molecular Biology and Evolution*.

Baker AJ, Huynen LJ, Haddrath O, Millar CD and Lambert DM (2005). Reconstructing the tempo and mode of evolution in an extinct clade of birds with ancient DNA: The giant Moas of New Zealand. *Proceedings of the National Academy of Sciences of the United States of America* 102(23): 8257-8262.

Bandelt H-J, Quintana-Murci L, Salas A and Macaulay V (2002). The fingerprint of phantom mutations in mitochondrial DNA data. *The American Journal of Human Genetics* 71(5): 1150-1160.

Bird MI, Austin WEN, Wurster CM, Fifield LK, Mojtahid M and Sargeant C (2010). Punctuated eustatic sea-level rise in the early mid-Holocene. *Geology* 38(9): 803-806.

Bitgood JJ (1985). Locating pea comb and blue egg in relation to the centromere of chromosome 1 in the chicken. *Poultry Science* 64(8): 1411-1414.

Bollmer JL, Kimball RT, Whiteman NK, Sarasola JH and Parker PG (2006). Phylogeography of the Galapagos hawk (*Buteo galapagoensis*): A recent arrival to the Galapagos Islands. *Molecular Phylogenetics and Evolution* 39(1): 237-247.

Brook LA and Kutt AS (2011). The diet of the dingo (*Canis lupus dingo*) in north-eastern Australia with comments on its conservation implications. *Rangeland Journal* 33(1): 79-85.

Brown M, Cooksley H, Carthew SM and Cooper SJB (2006). Conservation units and phylogeographic structure of an arboreal marsupial, the yellow-bellied glider (*Petaurus australis*). *Australian Journal of Zoology* 54(5): 305-317.

Brown WM, George M and Wilson AC (1979). Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America* 76(4): 1967-1971.

Buitenhuis B, Hedegaard J, Janss L and Sorensen P (2009). Differentially expressed genes for aggressive pecking behaviour in laying hens. *Bmc Genomics* 10.

Bunce M, Worthy TH, Phillips MJ, Holdaway RN, Willerslev E, Haile J, Shapiro B, Scofield RP, Drummond AJ, Kamp PJJ and Cooper A (2009). The evolutionary history of the extinct ratite Moa and New Zealand Neogene paleogeography. *Proceedings of the National Academy of Sciences of the United States of America* 106(49): 20646-20651.

Burbidge ML, Colbourne RM, Robertson HA and Baker AJ (2003). Molecular and other biological evidence supports the recognition of at least three species of Brown Kiwi. *Conservation Genetics* 4: 167-177.

Burney DA, James HF, Burney LP, Olson SL, Kikuchi W, Wagner WL, Burney M, McCloskey D, Kikuchi D, Grady FV, Gage R and Nishek R (2001). Fossil evidence for a diverse biota from Kaua'i and its transformation since human arrival. *Ecological Monographs* 71(4): 615-641.

Cohen TJ, Nanson GC, Jansen JD, Jones BG, Jacobs Z, Treble P, Price DM, May J-H, Smith AM, Ayliffe LK and Hellstrom JC (2011). Continental aridification and the vanishing of Australia's megalakes. *Geology* 39(2): 167-170.

Cooper A, Lalueza-Fox C, Anderson S, Rambaut A, Austin J and Ward R (2001). Complete mitochondrial genome sequences of two extinct Moas clarify ratite evolution. *Nature* 409(6821): 704-707.

Davies SJJF (2002). *Ratites and Tinamous: Tinamidae, Rheidae, Dromaiidae, Casuaiidae, Apterygidae, Struthionidae*. Oxford, Oxford University Press.

Decker JE, Pires JC, Conant GC, McKay SD, Heaton MP, Chen KF, Cooper A, Vilkki J, Seabury CM, Caetano AR, Johnson GS, Brenneman RA, Hanotte O, Eggert LS, Wiener P, Kim JJ, Kim KS, Sonstegard TS, Van Tassell CP, Neibergs HL, McEwan JC, Brauning R, Coutinho LL, Babar ME, Wilson GA, McClure MC, Rolf MM, Kim J, Schnabel RD and Taylor JF (2009). Resolving the evolution of extant and extinct ruminants with high-throughput phylogenomics. *Proceedings of the National Academy of Sciences of the United States of America* 106(44): 18644-18649.

Depaulis F, Orlando L and Hanni C (2009). Using classical population genetics tools with heterochroneous data: time matters! *PLOS ONE* 4(5): e5541-e5541.

Desjardins P and Morais R (1990). Sequence and gene organization of the chicken mitochondrial genome. A novel gene order in higher vertebrates. *Journal of molecular biology* 212(4): 599-634.

Dobson GE (1880). On some new or rare species of Chiroptera in the collection of the Gottingen Museum. *Proceedings of the Zoological Society of London* 48(3): 461-465.

Dreyfus GB, Parrenin F, Lemieux-Dudon B, Durand G, Masson-Delmotte V, Jouzel J, Barnola JM, Panno L, Spahni R, Tisserand A, Siegenthaler U and Leuenberger M (2007).

Anomalous flow below 2700 m in the EPICA Dome C ice core detected using delta O-18 of atmospheric oxygen measurements. *Climate of the Past* 3(2): 341-353.

Drummond A and Rambaut A (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* 7(1): 214.

Drummond AJ, Rambaut A, Shapiro B and Pybus OG (2005). Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution* 22(5): 1185-1192.

Eastwood R and Hughes JM (2003). Phylogeography of the rare myrmecophagous butterfly *Acrodipsas cuprea* (Lepidoptera : Lycaenidae) from pinned museum specimens. *Australian Journal of Zoology* 51(4): 331-340.

Eberhard JR and Bermingham E (2005). Phylogeny and comparative biogeography of *Pionopsitta* parrots and *Pteroglossus* toucans. *Molecular Phylogenetics and Evolution* 36(2): 288-304.

Ellis W (1831). *Polynesian researches: during a residence of nearly eight years in the Society and Sandwich Islands*. London, Henry G. Bohn.

Endicott P and Ho SYW (2008). A Bayesian evaluation of Human mitochondrial substitution rates. *American Journal of Human Genetics* 82(4): 895-902.

EPICA, Barbante C, Barnes PRF, Barnola J-M, Bigler M, Castellano E, Cattani O, Chappellaz J, Dahl-Jensen D, Delmonte B, Dreyfus G, Durand G, Falourd S, Fischer H, Fluckiger J, Hansson ME, Huybrechts P, Jugie G, Johnsen SJ, Jouzel J, Kaufmann P, Kipfstuhl J, Lambert F, Lipenkov VY, Littot GC, Longinelli A, Lorrain R, Maggi V, Masson-Delmotte V, Miller H, Mulvaney R, Oerlemans J, Oerter H, Orombelli G, Parrenin F, Peel DA, Petit J-R, Raynaud D, Ritz C, Ruth U, Schwander J, Siegenthaler U, Souchez R, Stauffer B, Steffensen JP, Stenni B, Stocker TF, Tabacco IE, Udisti R, van de Wal RSW, van den Broeke M, Weiss J, Wilhelms F, Winther J-G, Wolff EW and Zucchelli M (2004). Eight glacial cycles from an Antarctic ice core. *Nature* 429(6992): 623-628.

Excoffier L and Foll M (2011). Fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics* 27(9): 1332-1334.

Excoffier L, Novembre J and Schneider S (2000). SIMCOAL: A general coalescent program for the simulation of molecular data in interconnected populations with arbitrary demography. *Journal of Heredity* 91(6): 506-509.

Figueirido B and Janis CM (2011). The predatory behaviour of the Thylacine: Tasmanian tiger or marsupial wolf? *Biology Letters*.

Fillios M, Gordon C, Koch F and Letnic M (2010). The effect of a top predator on kangaroo abundance in arid Australia and its implications for archaeological faunal assemblages. *Journal of Archaeological Science* 37(5): 986-993.

Fleischer RC, McIntosh CE and Tarr CL (1998). Evolution on a volcanic conveyor belt: using phylogeographic reconstructions and K–Ar-based ages of the Hawaiian Islands to estimate molecular evolutionary rates. *Molecular Ecology* 7(4): 533-545.

Flood J (2004). *Archaeology of the Dreamtime*, J. B. Publishing.

Fulton JE (2012). Genomic selection for poultry breeding. *Animal Frontiers* 2(1): 30-36.

Fumagalli L, Moritz C, Taberlet P and Friend JA (1999). Mitochondrial DNA sequence variation within the remnant populations of the endangered numbat (Marsupialia : Myrmecobiidae : *Myrmecobius fasciatus*). *Molecular Ecology* 8(9): 1545-1549.

Gale SJ (2009). Event chronostratigraphy: A high-resolution tool for dating the recent past. *Quaternary Geochronology* 4(5): 391-399.

Gill FB, Slikas B and Sheldon FH (2005). Phylogeny of titmice (Paridae): II. Species relationships based on sequences of the mitochondrial cytochrome-b gene. *Auk* 122(1): 121-143.

Glatz P, Miao Z, Hughes B, Jansen T, Aleve H and Sandakabatu C (2009). Feeding village poultry in Solomon Islands, Australian Centre for International Agricultural Research: 1-158.

Halloran AF and Halloran AG (1969). The feral chickens (Phasianidae) of French Polynesia. *Proceedings of the Oklahoma Academy of Sciences* 49: 169-170.

Hammond EL, Lymbery AJ, Martin GB, Groth D and Wetherall JD (2002). Microsatellite analysis of genetic diversity in wild and farmed Emus (*Dromaius novaehollandiae*). *The Journal of Heredity* 93(5): 376-380.

Harlid A, Janke A and Arnason U (1997). The mtDNA sequence of the ostrich and the divergence between paleognathous and neognathous birds. *Molecular Biology and Evolution* 14(7): 754-761.

Harshman J, Braun EL, Braun MJ, Huddleston CJ, Bowie RCK, Chojnowski JL, Hackett SJ, Han K-L, Kimball RT, Marks BD, Miglia KJ, Moore WS, Reddy S, Sheldon FH, Steadman DW, Steppan SJ, Witt CC and Yuri T (2008). Phylogenomic evidence for multiple losses of flight in ratite birds. *Proceedings of the National Academy of Sciences of the United States of America* 105(36): 13462-13467.

Hasegawa M, Kishino H and Yano TA (1985). Dating of the Human-Ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* 22(2): 160-174.

Hawkins CE, Baars C, Hesterman H, Hocking GJ, Jones ME, Lazenby B, Mann D, Mooney N, Pemberton D, Pyecroft S, Restani M and Wiersma J (2006). Emerging disease and population decline of an island endemic, the Tasmanian devil *Sarcophilus harrisii*. *Biological Conservation* 131(2): 307-324.

Heled J and Drummond AJ (2010). Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution* 27(3): 570-580.

Heupink TH, Huynen L and Lambert DM (2011). Ancient DNA suggests dwarf and 'Giant' Emu are conspecific. *PLOS ONE* 6(4).

Ho SYW, Lanfear R, Bromham L, Phillips MJ, Soubrier J, Rodrigo AG and Cooper A (2011a). Time-dependent rates of molecular evolution. *Molecular Ecology*.

Ho SYW, Lanfear R, Phillips MJ, Barnes I, Thomas JA, Kolokotronis SO and Shapiro B (2011b). Bayesian estimation of substitution rates from ancient DNA sequences with low information content. *Systematic Biology* 60(3): 366-374.

Ho SYW, Shapiro B, Phillips MJ, Cooper A and Drummond AJ (2007). Evidence for time dependency of molecular rate estimates. *Systematic Biology* 56(3): 515-522.

Hope JH, Lampert RJ, Edmondson E, Smith MJ and Van Tets GF (1977). Late Pleistocene faunal remains from Seton Rock Shelter, Kangaroo Island, South Australia. *Journal of Biogeography* 4(4): 363-385.

Houlden BA, Costello BH, Sharkey D, Fowler EV, Melzer A, Ellis W, Carrick F, Baverstock PR and Elphinstone MS (1999). Phylogeographic differentiation in the mitochondrial control region in the koala, *Phascolarctos cinereus* (Goldfuss 1817). *Molecular Ecology* 8(6): 999-1011.

Jensen P and Andersson L (2005). Genomics meets ethology: a new route to understanding domestication, behavior, and sustainability in animal breeding. *Ambio* 34(4/5): 320-324.

Johnson CN, Isaac JL and Fisher DO (2007). Rarity of a top predator triggers continent-wide collapse of mammal prey: dingoes and marsupials in Australia. *Proceedings of the Royal Society of London Series B-Biological Sciences* 274(1608): 341-346.

Johnson KP and Weckstein JD (2011). The Central American land bridge as an engine of diversification in New World doves. *Journal of Biogeography* 38(6): 1069-1076.

Jones K. (2008). "Husbandry Manual for Ghost bat *Macroderma gigas* Mammalia: Megadermatidae." Western Sydney Institute of TAFE Retrieved 09 December 2011, 2011, from http://www.australianzookeeping.org/Husbandry_Manuals.htm - Order_Chiroptera_(Bats)_.

Jones ME and Stoddart DM (1998). Reconstruction of the predatory behaviour of the extinct marsupial thylacine (*Thylacinus cynocephalus*). *Journal of Zoology* 246: 239-246.

Kaufman J, Milne S, Gobel TWF, Walker BA, Jacob JP, Auffray C, Zoorob R and Beck S (1999). The chicken B locus is a minimal essential major histocompatibility complex. *Nature* 401(6756): 923-925.

Kawamura K, Parrenin F, Lisiecki L, Uemura R, Vimeux F, Severinghaus JP, Hutterli MA, Nakazawa T, Aoki S, Jouzel J, Raymo ME, Matsumoto K, Nakata H, Motoyama H, Fujita S, Goto-Azuma K, Fujii Y and Watanabe O (2007). Northern Hemisphere forcing of climatic cycles in Antarctica over the past 360,000 years. *Nature* 448(7156): 912-916.

Knapp M, Stiller M and Meyer M (2012). Generating Barcoded Libraries for Multiplex High-Throughput Sequencing. Ancient DNA. Shapiro B and Hofreiter M, Humana Press. 840: 155-170.

Komiyama T, Ikeo K and Gojobori T (2004). The evolutionary origin of long-crowing chicken: its evolutionary relationship with fighting cocks disclosed by the mtDNA sequence analysis. *Gene* 333: 91-99.

Lambeck K and Chappell J (2001). Sea level change through the last glacial cycle. *Science* 292(5517): 679-686.

Larson G, Cucchi T, Fujita M, Matisoo-Smith E, Robins J, Anderson A, Rolett B, Spriggs M, Dolman G, Kim TH, Thuy NT, Randi E, Doherty M, Due RA, Bollt R, Djubiantono T, Griffin B, Intoh M, Keane E, Kirch P, Li KT, Morwood M, Pedrina LM, Piper PJ, Rabett RJ, Shooter P, Van den Bergh G, West E, Wickler S, Yuan J, Cooper A and Dobney K (2007). Phylogeny and ancient DNA of *Sus* provides insights into neolithic expansion in Island Southeast Asia and Oceania. *Proceedings of the National Academy of Sciences of the United States of America* 104(12): 4834-4839.

Lerner HRL, Meyer M, James HF, Hofreiter M and Fleischer RC (2011). Multilocus resolution of phylogeny and timescale in the extant adaptive radiation of Hawaiian Honeycreepers. *Current Biology* 21(21): 1838-1844.

Li WH, Gojobori T and Nei M (1981). Pseudogenes as a paradigm of neutral evolution. *Nature* 292(5820): 237-239.

Liu WB, Chen SR, Zheng JX, Qu LJ, Xu GY and Yang N (2010). Developmental phenotypic-genotypic associations of tyrosinase and melanocortin 1 receptor genes with changing profiles in chicken plumage pigmentation. *Poultry Science* 89(6): 1110-1114.

Lloyd BD (2003). Intraspecific phylogeny of the New Zealand short-tailed bat *Mystacina tuberculata* inferred from multiple mitochondrial gene sequences. *Systematic Biology* 52(4): 460-476.

Lopez JV, Culver M, Stephens JC, Johnson WE and O'Brien SJ (1997). Rates of nuclear and cytoplasmic mitochondrial DNA sequence divergence in mammals. *Molecular Biology and Evolution* 14(3): 277-286.

Lopez JV, Yuhki N, Masuda R, Modi W and Obrien SJ (1994). Numt, a recent transfer and tandem amplificiation of mitochondrial DNA to the nuclear genome of the domestic cat. *Journal of Molecular Evolution* 39(2): 174-190.

Macqueen P, Goldizen AW, Austin JJ and Seddon JM (2011). Phylogeography of the pademelons (Marsupialia: Macropodidae: *Thylogale*) in New Guinea reflects both geological and climatic events during the Plio-Pleistocene. *Journal of Biogeography* 38(9): 1732-1747.

Macqueen P, Goldizen AW and Seddon JM (2009). Response of a southern temperate marsupial, the Tasmanian pademelon (*Thylogale billardierii*), to historical and contemporary forest fragmentation. *Molecular Ecology* 18(15): 3291-3306.

Macqueen P, Seddon JM, Austin JJ, Hamilton S and Goldizen AW (2010). Phylogenetics of the pademelons (Macropodidae: *Thylogale*) and historical biogeography of the Australo-Papuan region. *Molecular Phylogenetics and Evolution* 57(3): 1134-1148.

Maloney SK (2008). Thermoregulation in ratites: a review. *Australian Journal of Experimental Agriculture* 48(10): 1293-1301.

Marchant S and Higgins PJ (1990). *Handbook of Australian, New Zealand and Antarctic birds: Volume 1, Ratites to ducks*. Oxford, Oxford University Press.

Marx SK, McGowan HA and Kamber BS (2009). Long-range dust transport from eastern Australia: A proxy for Holocene aridity and ENSO-type climate variability. *Earth and Planetary Science Letters* 282(1-4): 167-177.

Matisoo-Smith E, Roberts RM, Irwin GJ, Allen JS, Penny D and Lambert DM (1998). Patterns of prehistoric human mobility in Polynesia indicated by mtDNA from the Pacific rat. *Proceedings of the National Academy of Sciences of the United States of America* 95(25): 15145-15150.

McCarthy L, Head L and Quade J (1996). Holocene palaeoecology of the northern Flinders Ranges, South Australia, based on stick-nest rat (*Leporillus spp.*) middens: a preliminary overview. *Palaeogeography Palaeoclimatology Palaeoecology* 123: 205-218.

Melville J, Goebel S, Starr C, Keogh JS and Austin JJ (2007). Conservation genetics and species status of an endangered Australian dragon, *Tympanocryptis pinguicolla* (Reptilia : Agamidae). *Conservation Genetics* 8(1): 185-195.

Meyer M, Kircher M, Gansauge M-T, Li H, Racimo F, Mallick S, Schraiber JG, Jay F, Pr√°fer K, de Filippo C, Sudmant PH, Alkan C, Fu Q, Do R, Rohland N, Tandon A, Siebauer M,

Green RE, Bryc K, Briggs AW, Stenzel U, Dabney J, Shendure J, Kitzman J, Hammer MF, Shunkov MV, Derevianko AP, Patterson N, Andr√©s AM, Eichler EE, Slatkin M, Reich D, Kelso J and P√§√§bo S (2012). A High-Coverage Genome Sequence from an Archaic Denisovan Individual. *Science*.

Miller GH, Fogel ML, Magee JW, Gagan MK, Clarke SJ and Johnson BJ (2005). Ecosystem collapse in Pleistocene Australia and a human role in megafaunal extinction. *Science* 309(5732): 287-290.

Miller W, Hayes VM, Ratan A, Petersen DC, Wittekindt NE, Miller J, Walenz B, Knight J, Qi J, Zhao FQ, Wang QY, Bedoya-Reina OC, Katiyar N, Tomsho LP, Kasson LM, Hardie RA, Woodbridge P, Tindall EA, Bertelsen MF, Dixon D, Pyecroft S, Helgen KM, Lesk AM, Pringle TH, Patterson N, Zhang Y, Kreiss A, Woods GM, Jones ME and Schuster SC (2011). Genetic diversity and population structure of the endangered marsupial *Sarcophilus harrisii* (Tasmanian devil). *Proceedings of the National Academy of Sciences of the United States of America* 108(30): 12348-12353.

Molnar RE, Hall LS and Mahoney JH (1984). New fossil localities for *Macroderma* Miller, 1906 (Chiroptera: Megadermatidae) in New South Wales and its past and present distribution in Australia. *Australian Mammalogy* 7: 63-73.

Mourier T, Ho SYW, Gilbert MTP, Willerslev E and Orlando L (2012). Statistical guidelines for detecting past population shifts using ancient DNA. *Molecular Biology and Evolution*.

Muir WM, Wong GK, Zhang Y, Wang J, Groenen MAM, Crooijmans R, Megens HJ, Zhang HM, McKay JC, McLeod S, Okimoto R, Fulton JE, Settar P, O'Sullivan NP, Vereijken A, Jungerius-Rattink A, Albers GAA, Lawley CT, Delany ME and Cheng HH (2008a). Review of the initial validation and characterization of a 3K chicken SNP array. *Worlds Poultry Science Journal* 64(2): 219-225.

Muir WM, Wong GK-S, Zhang Y, Wang J, Groenen MAM, Crooijmans RPMA, Megens H-J, Zhang H, Okimoto R, Vereijken A, Jungerius A, Albers GAA, Lawley CT, Delany ME, MacEachern S and Cheng HH (2008b). Genome-wide assessment of worldwide chicken SNP genetic diversity indicates significant absence of rare alleles in commercial breeds. *Proceedings of the National Academy of Sciences of the United States of America* 105(45): 17312-17317.

Nabholz B, Glemin S and Galtier N (2008). Strong variations of mitochondrial mutation rate across mammals - the Longevity Hypothesis. *Molecular Biology and Evolution* 25(1): 120-130.

Navascues M, Depaulis F and Emerson BC (2010). Combining contemporary and ancient DNA in population genetic and phylogeographical studies. *Molecular Ecology Resources* 10(5): 760-772.

Oskam CL, Haile J, McLay E, Rigby P, Allentoft ME, Olsen ME, Bengtsson C, Miller GH, Schwenninger JL, Jacomb C, Walter R, Baynes A, Dortch J, Parker-Pearson M, Gilbert MTP, Holdaway RN, Willerslev E and Bunce M (2010). Fossil avian eggshell preserves ancient DNA. *Proceedings of the Royal Society of London Series B-Biological Sciences* 277(1690): 1991-2000.

Paplinska JZ, Taggart DA, Corrigan T, Eldridge MDB and Austin JJ (2011). Using DNA from museum specimens to preserve the integrity of evolutionarily significant unit boundaries in threatened species. *Biological Conservation* 144(1): 290-297.

Parker R. (n.d.). "kaikokorako.com."   Retrieved 30-Jan-2012, 2012.

Percy DM, Garver AM, Wagner WL, James HF, Cunningham CW, Miller SE and Fleischer RC (2008). Progressive island colonization and ancient origin of Hawaiian *Metrosideros* (Myrtaceae). *Proceedings of the Royal Society of London Series B-Biological Sciences* 275(1642): 1479-1490.

Petit E, Excoffier L and Mayer F (1999). No evidence of bottleneck in the postglacial recolonization of Europe by the Noctule Bat (*Nyctalus noctula*). *Evolution* 53(4): 1247-1258.

Petit E and Mayer F (1999). Male dispersal in the noctule bat (*Nyctalus noctula*): where are the limits? *Proceedings of the Royal Society of London Series B-Biological Sciences* 266(1430): 1717-1722.

Phillips MJ, Gibb GC, Crimp EA and Penny D (2010). Tinamous and Moa flock together: mitochondrial genome sequence analysis reveals independent losses of flight among Ratites. *Systematic Biology* 59(1): 90-107.

Pople AR, Grigg GC, Cairns SC, Beard LA and Alexander P (2000). Trends in the numbers of red kangaroos and emus on either side of the South Australian dingo fence: evidence for predator regulation? *Wildlife Research* 27: 269-276.

Quigley MC, Horton T, Hellstrom JC, Cupper ML and Sandiford M (2010). Holocene climate change in arid Australia from speleothem and alluvial records. *The Holocene* 20(7): 1093-1104.

Rambaut A (2009). Path-O-Gen. Institute of Evolutionary Biology

Rauw WM, Kanis E, Noordhuizen-Stassen EN and Grommers FJ (1998). Undesirable side effects of selection for high production efficiency in farm animals: a review. *Livestock Production Science* 56: 15-33.

Richards T, Pavlides C, Walshe K, Webber H and Johnston R (2007). Box Gully: new evidence for Aboriginal occupation of Australia south of the Murray River prior to the last glacial maximum. *Archaeology in Oceania* 42(1): 1-11.

Rollins LA, Woolnough AP, Wilton AN, Sinclair R and Sherwin WB (2009). Invasive species can't cover their tracks: using microsatellites to assist management of starling (*Sturnus vulgaris*) populations in Western Australia. *Molecular Ecology* 18(8): 1560-1573.

Savolainen P, Leitner T, Wilton AN, Matisoo-Smith E and Lundeberg J (2004). A detailed picture of the origin of the Australian dingo, obtained from the study of mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America* 101: 12387-12390.

Sawai H, Kim HL, Kuno K, Suzuki S, Gotoh H, Takada M, Takahata N, Satta Y and Akishinonomiya F (2010). The origin and genetic variation of domestic chickens with special reference to Junglefowls *Gallus g. gallus*. *PLOS ONE* 5(5): e10639.

Schenekar T and Weiss S (2011). High rate of calculation errors in mismatch distribution analysis results in numerous false inferences of biological importance. *Heredity* 107(6): 511-512.

Seabrook-Davison M, Huynen L, Lambert DM and Brunton DH (2009). Ancient DNA resolves identity and phylogeny of New Zealand's extinct and living quail (*Coturnix sp.*). *PLOS ONE* 4(7).

Short J and Smith A (1994). Mammal decline and recovery in Australia. *Journal of Mammalogy* 75(2): 288-297.

Shulmeister J (1999). Australasian evidence for mid-holocene climate change implies precessional control of Walker Circulation in the Pacific. *Quaternary International* 57-58(0): 81-91.

Siddle HV, Kreiss A, Eldridge MDB, Noonan E, Clarke CJ, Pyecroft S, Woods GM and 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 of the United States of America* 104(41): 16221-16226.

Smith MA, Williams AN, Turney CSM and Cupper ML (2008). Human-environment interactions in Australian drylands: exploratory time-series analysis of archaeological records. *The Holocene* 18(3): 389-401.

Spooner MI, De Deckker P, Barrows TT and Fifield LK (2011). The behaviour of the Leeuwin Current offshore NW Australia during the last five glacial-interglacial cycles. *Global and Planetary Change* 75(3-4): 119-132.

Steadman DW and Intoh M (1994). Biogeography and prehistoric exploitation of birds from Fais Island, Yap State, Federated States of Micronesia. *Pacific Science* 48(2): 116-135.

Storey AA, Ladefoged T and Matisoo-Smith EA (2008a). Counting your chickens: Density and distribution of chicken remains in archaeological sites of Oceania. *International Journal of Osteoarchaeology* 18(3): 240-261.

Storey AA, Quiroz D, Ramirez JM, Beavan-Athfield N, Addison DJ, Walter R, Hunt T, Athens JS, Huynen L and Matisoo-Smith EA (2008b). Pre-Columbian chickens, dates, isotopes, and mtDNA. *Proceedings of the National Academy of Sciences of the United States of America* 105(48): E99.

Storey AA, Ramirez JM, Quiroz D, Burley DV, Addison DJ, Walter R, Anderson AJ, Hunt TL, Athens JS, Huynen L and Matisoo-smith EA (2007). Radiocarbon and DNA evidence for a pre-Columbian introduction of Polynesian chickens to Chile. *Proceedings of the National Academy of Sciences of the United States of America* 104(25): 10335-10339.

Storey AA, Spriggs M, Bedford S, Hawkins SC, Robins JH, Huynen L and Matisoo-Smith E (2010). Mitochondrial DNA from 3000-year old chickens at the Teouma site, Vanuatu. *Journal of Archaeological Science* 37(10): 2459-2468.

Teeling EC, Springer MS, Madsen O, Bates P, O'Brien SJ and Murphy WJ (2005). A molecular phylogeny for bats illuminates biogeography and the fossil record. *Science* 307(5709): 580-584.

Ullstrop AJ (1972). The impacts of the southern corn leaf blight epidemics of 1970-1971. *Annual Review of Phytopathology* `10: 37-50.

Vera JC, Wheat CW, Fescemyer HW, Frilander MJ, Crawford DL, Hanski I and Marden JH (2008). Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. *Molecular Ecology* 17(7): 1636-1647.

West B and Zhou B-X (1988). Did chickens go north? New evidence for domestication. *Journal of Archaeological Science* 15: 515-533.

Worthington Wilmer J, Hall L, Barratt E and Moritz C (1999). Genetic structure and male-mediated gene flow in the Ghost Bat (*Macroderma gigas*). *Evolution* 53(5): 1582-1591.

Worthington Wilmer J, Moritz C, Hall L and Toop J (1994). Extreme population structuring in the threatened Ghost Bat, *Macroderma gigas* - Evidence from mitochondrial DNA. *Proceedings of the Royal Society of London Series B-Biological Sciences* 257(1349): 193-198.

Wroe S, Clausen P, McHenry C, Moreno K and Cunningham E (2007). Computer simulation of feeding behaviour in the thylacine and dingo as a novel test for convergence and niche

overlap. *Proceedings of the Royal Society of London Series B-Biological Sciences* 274(1627): 2819-2828.

Zenger KR, Eldridge MDB and Cooper DW (2003). Intraspecific variation, sex-biased dispersal and phylogeography of the Eastern Grey Kangaroo (*Macropus giganteus*). *Heredity* 91(2): 153-162.