Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

PATTERNS AND PROCESSES IN ANIMAL EVOLUTION: MOLECULAR PHYLOGENETICS OF SOUTHERN HEMISPHERE FAUNA

A thesis presented in partial fulfilment of the requirements for the degree of
Doctor of Philosophy in Genetics

At the Allan Wilson Centre for Molecular Ecology and Evolution,
Institute of Molecular BioSciences, Massey University, Palmerston North, New Zealand

Renae Pratt

2008

There is no better high than discovery

-Edward O. Wilson.

ABSTRACT

Three kinds of processes are known to modify the geographical spatial arrangement of organisms: dispersal, extinction and vicariance. The Southern Hemisphere has an intriguing and complicated geological history that provides an ideal backdrop to study these processes. This thesis focuses on three historical events that illustrate these processes: the proposed marine inundation of New Zealand in the Oligocene, the asteroid impact at the K – Pg boundary, and the continental breakup of Gondwana. It investigates what impact these events had on species diversification by studying the phylogenetic relationships of two groups of taxa – the family Anostostomatidae (insects), and Neoaves (birds). Anostostomatidae were studied in relation to the Oligocene drowning and the break up of Gondwana as they have a wide southern distribution, found on all "Gondwanan" fragments with the exception of Antarctica, and are thought represent an ancient lineage that predates the Gondwanan breakup. Birds, in particular Neoaves, were studied in relation to the asteroid impact at the K – Pg boundary. Although birds are mobile and many circumnavigate the globe between seasons, they are suggested to have originated in the Southern Hemisphere in Gondwanan times, and subsequently undergone range expansion and diversification around the world. In order to address the relationship (if any) between modern biotic diversity and historical geological events, phylogenetic relationships were determined and where possible, molecular clock analysis carried out. Timing information provided by molecular clock analysis is important as it enables distinction between opposing hypotheses such as vicariance and dispersal.

In *Chapter Two*, the phylogenetic relationships within the family Anostostomatidae are investigated. One of the most controversial times in New Zealand's geological history is during the Oligocene. Some suggest that the lack of fossils and evidence for recent dispersal of numerous taxa support the notion that all modern biota reached the region during the last 25 million years. Anostostomatidae were chosen as they represent a group of insects that are thought to be ancient and there is little published data in the literature. Previous studies focused on the relationships within *Hemideina* and *Deinacrida* suggesting that these groups diversified in the early Miocene. The data presented here are from mitochondrial (COI and 12S) and nuclear (18S and 28S) sequences.

Molecular dating using a relaxed clock as implemented in BEAST suggest that in fact some lineages were present at or shortly after continental breakup and could have survived throughout this turbulent time. As there were no definitive fossils to use for calibration points, geological events were used as calibration points for the molecular clock. Mutation rates obtained from the different analyses were compared to those published for other insects in an attempt to identify the most likely model. Both maximum likelihood and Bayesian analyses support the presence of three distinct ecological groups in New Zealand; *Hemiandrus* (ground weta), *Anisoura/Motuweta* (tusked weta) and *Hemideina–Deinacrida* (tree–giant weta). With regards to their Australasian relatives (taxa from Australia and New Caledonia) it appears that the family is divided with the most northern New Zealand taxa (tusked weta) more closely related to New Caledonian taxa while all other New Zealand taxa are more closely related to Australian taxa. There does not appear to be any link between the Australian and New Caledonian taxa studied here. Results should be viewed with caution however as an increased mutation rate was observed in the New Caledonian-tusked weta lineage, something future studies will have to address.

Chapter Three presents new sequence data and phylogenetic analyses that go towards resolving the apparent basal polytomy of neoavian birds. This chapter includes analyses carried out on previously published data with the addition of nine new mitochondrial genomes. My contribution to this larger project was to perform the phylogenetic analysis and to sequence three of the nine mitochondrial genomes. The genomes I sequenced were the Southern Hemisphere species: dollar bird (*Eurystomus orientalis*), Owlet nightjar (*Aegotheles cristatus cristatus*) and great potoo (*Nyctibius grandis*). The inclusion of these nine new genomes allows assessment and comparison of the six hypothesised groups reported in Cracraft (2001). First an improved conditional down-weighting technique is described reducing noise relative to signal, which is important for resolving deeper divergences. Second, a formula is presented for calculating probabilities of finding predefined groupings in the optimal tree. Maximum likelihood and Bayesian based phylogenetic analyses were carried out and in addition, dating using a relaxed molecular clock was performed in BEAST. Results suggested that the six groups suggested by Cracraft (2001) represent robust lineages.

The results suggested that one group, the owls, are more closely related to other raptors, particularly accipitrids (buzzards/eagles) and the osprey rather than the Caprimulgiformes, which could indicate morphological convergent evolution.

In addition, a group termed shorebirds appears to be distinct from the large group referred to as 'Conglomerati' to which previous publications have suggested they belong. The 'Conglomerati' is the least well studied group and may actually comprise of at least three subgroups (as suggested by Cracraft). Within the three suggested groups, Cracraft grouped shorebirds with pigeons and sandgrouse, neither of which (pigeons or sandgrouse) were analysed here. So although the shorebirds are at least close to the 'Conglomerati' and may be within that group, their exact position is still not clear. The molecular dating reported here utilised two fossil calibrations (*Vegavis* and *Waimanu*), for which there is relatively little dispute as to age or the lineage to which they belong. Calibrations resulting from BEAST analyses suggest that at least 12 distinct lineages were present prior to the K – Pg boundary, a finding supported by previous studies. Robust phylogenies will allow future studies to investigate not only the relationships within Neoaves, but look more closely at the biological and ecological evolution of the group.

Chapter Four for the first time investigates whether the phylogenetic relationships within the family Anostostomatidae follow the conventionally accepted order and timing of Gondwanan breakup. Following the initial restults for taxa studied in Australasia (*Chapter Two*) an attempt to resolve family relationships in a wider spatial (geographic) context was carried out to determine if Australasian taxa are monophyletic when other members of the family are included. Again both maximum likelihood and Bayesian phylogenetic analyses were carried out on both mitochondrial (COI and 12S) and nuclear (18S and 28S) sequences. In this chapter, datasets included samples from across the geographic range of Anostostomatidae (South Africa, Madagascar, South America, Australia, New Caledonia and New Zealand), and two clades were observed, congruent with earlier findings. Sequence divergence within geographic regions was found to be relatively high in the mitochondrial genes (COI and 12S) while low in the nuclear ribosomal RNA genes (18S and 28S) as expected given their relative mutation rates. Under the vicariance paradigm, phylogenetic relationships should follow the order of continental breakup, but this was not found.

Further, if dispersal and colonisation were continuous, no geographic substructure is expected, however distinct geographic substructure within clades was consistently observed. This interesting phylogenetic pattern may be a case of convergent evolution or paraphyletic sampling which highlights taxonomic issues of the group.

Future studies need to include not only molecular data but information on morphology, ecology and behaviour along with the implementation of biogeographic programs that can test alternative hypotheses (such as dispersal and vicariance) directly. Also, the inclusion of the recently reported fossil from the subfamily Euclydesinae (Martins-Neto 2007) should allow for more accurate date estimates within the family.

Taken as a whole the results presented in this thesis suggest that microevolutionary processes are sufficient to explain modern diversity without the need to invoke abiotic events. The three cases investigated here - marine inundation, asteroid impact and continental drift - all appear to have had only a limited effect on the diversity of taxa studied. To reach even stronger conclusions future studies should incorporate different data (for instance nuclear genes, intron position, and genome structure) and use biogeographic software capable of including ecological, morphological and habitat information.

ACKNOWLEDGMENTS

The important thing is never to stop questioning

- Albert Einstein

No thesis can be finished without the support of friends and family no matter how far away they are. For this, I thank my family and friends ‘across the ditch’ for their continuous support throughout this stage in my life. Thanks for picking up the pieces when all seemed to be going wrong and for the ‘motivational’ pep talks when needed. An extra special thank you to Carol Scarpaci. You are my mentor, my friend, an inspiration.

A big thanks to; Tim White, Warwick Allen, Jing Wang and Napat Kantanon (Boffins) for assistance with anything and everything computer related. To members of the Farside lab; Angela McGaughran, Simon Hills, Gillian Gibb, Olga Kardialsky, and Steve Pilkington and from the Pleb lab; Michael Knapp and Richard Carter. I thank you all for advice and direction to varying degrees and to those who showed interest, regardless of whether you were actually interested! Ultimately, I thank you all for your friendship and support.

I am indebted to Andrew Clarke, Barbara Holland, Trish “It’ll be fine!” McLenachan, Melanie Pierson, Sarah Bray, Gail Ashton, Yasmin Pearce, Chris Lane and Corina Jordan for the many counselling sessions and positive encouragement. Also to Anouck de Bonth and Judith Stepper and Dora for their friendship and support and the ability to endure at times a grumpy, stressed-out housemate. To Gemma Henderson, Sheralee Cleland, Gerry Ryan, Melle Steringa, Michelle Turner, Marijn Kool, Bouke de Jong, Nikki Kenters, Tobias Van Den Broek and Lindsay King for the entertainment at pub quizzes, Singstar and Buzz nights, and New Zealand adventures. Laughter really is the best medicine.

Kim McBreen, Matt Phillips and Barbara Holland, I will forever be grateful for the time you spent answering my endless questions about analyses. Thank you for explaining things to me like a 2 year old when necessary and making the time even when you didn’t have it. I would still be reading an introduction to phylogenetic analysis if it weren’t for your patience.

To our AWC administrators Joy, Karen and Susan for all your assistance with financial matters, booking flights and sorting accounts and in general making things run smoothly. I know at times, this can be a thankless job.

Thank you to Ang for proof reading the thesis in one day! To Judith for reading something so far out of your expertise and to Corina for giving up her weekend to proof one last time. Thanks to everyone who has helped over the years with writing and presentation preparation.

Lastly I would like to thank my supervisors without who I would never have had the opportunity to achieve a dream. To my co-supervisors, Steve Trewick and Mary Morgan-Richards, without your treasure trove of samples, this work would not have been possible. I also thank you for you for advice, assistance with writing, and thought provoking discussions. I know I have learnt a lot from this experience. To my chief supervisor, David Penny, I am enormously appreciative for advice and guidance throughout this process. All your stories, sidetracks and motivational support…'you're on the downhill slide, just one more little bump to overcome first!' and of course the financial support that no student can do without.

TABLE OF CONTENTS

ABSTRACT .... ***i***

ACKNOWLEDGMENTS .... ***v***

LIST OF FIGURES .... ***xii***

LIST OF TABLES .... ***xiii***

1. CHAPTER ONE **1**

THESIS STRUCTURE AND SCOPE ***2***

1.1 FACTORS THAT INFLUENCE OUR THINKING (BACKGROUND) ***4***

1.1.1 Drift theory 4

1.1.2 Biogeography 5

1.1.3 Getting a date in the modern era - Fossils Vs. DNA 6

1.1.4 Problems with comparisons 8

1.2 CASE STUDIES ***9***

1.2.1 Testing the depths of the Oligocene (Chapter Two) 10

1.2.2 What impact did the asteroid have? (Chapter Three) 16

1.2.3 Dispersal or drift? Getting around on Gondwana (Chapter Four) 18

1.2.4 Patterns and predictions 21

1.3 CONTRIBUTION OF OTHERS TO THE THESIS ***23***

1.4. REFERENCES ***24***

2. CHAPTER TWO **31**

Diversification of New Zealand weta (Orthoptera: Ensifera: Anostostomatidae) and their relationships in Australasia ***32***

2.1 ABSTRACT ***32***

2.2 INTRODUCTION ***32***

2.3 MATERIALS AND METHODS ***36***

(a) Sampling 36

(b) Molecular Methods 37

(c) Phylogenetic Analysis 38

(d) Tree comparisons 39

(e) Divergence-time estimation 40

2.4 RESULTS ***41***

(a) Phylogeny 41

(b) Summary of phylogenetics 47
(c) Divergence estimation 48
2.5 Discussion ***49***
2.6 Acknowledgements ***51***
2.7 References ***52***
2.8 Appendix ***56***

3. CHAPTER THREE **58**

Towards resolving deep Neoaves phylogeny: data, signal enhancement and priors ... ***59***
3.1 Abstract ***59***
3.2 Introduction ***59***
3.3 Materials and Methods ***66***
(a) Taxon Sampling 66
(b) Molecular Methods 67
(c) Phylogenetic analysis 69
(d) Noise reduction by down-weighting (Site-Stripping) 69
(e) Probabilities of observing predefined clades 71
3.4 Results ***72***
3.5 Discussion ***83***
3.6 Conclusion ***86***
3.7 Acknowledgements ***86***
3.8. References ***87***
3.9 Appendices ***91***
3.9.1 Fossil calibrations and molecular dating with BEAST 91
3.9.2 The probability of observing a pre-specified clade 95
3.9.3 Tabulated results from Bayesian (BPP) and bootstrap (bs) analyses using RAxML and PhyML 97
3.9.4 Consensus tree including all outgroup taxa used for alignment before site stripping 98

4. CHAPTER FOUR **100**

Return to Gondwana: exploring biogeographical patterns of the family Anostostomatidae (Orthoptera: Ensifera) ***101***
4.1 Abstract ***101***
4.2 Introduction ***102***

4.3 MATERIALS AND METHODS **105**
(a) Sampling 105
(b) Molecular Methods 105
(c) Phylogenetic analysis 106
(d) Tree comparisons 107
4.4 RESULTS **108**
(a) Sequence divergence 108
(b) Phylogenetic analysis 110
(c) Stemminess and Relative Compositional Variability 115
(d) Summary phylogenetics 116
4.5 DISCUSSION **117**
4.6 ACKNOWLEDGEMENTS **121**
4.7 REFERENCES **121**
4.8 APPENDICES **123**
4.8.1 Orthoptera Species File Online: taxa hierarchy as of 250908 (Eades and Otte) 124
4.8.2 Sample data including name, location, accession number, and genes used. 126
5. CHAPTER FIVE **128**
Summation **129**
5.1 EVIDENCE FOR ABIOTIC EFFECTS? **129**
5.2 CONCLUDING REMARKS **136**
5.3 FUTURE WORK **136**
5.4 REFERENCES **138**
A. APPENDIX A **141**
Mitogenomics – exploring mitochondrial genomes from New Zealand weta (Orthoptera: Ensifera: Anostostomatidae) **142**
A.1 ABSTRACT **142**
A.2 INTRODUCTION **142**
A.3 MATERIALS AND METHODS **145**
(a) Sampling 145
(b) DNA extraction 145
(c) Long Range PCR 145
(d) Short Range PCR 146

(e)Restriction digestion 147
(f) Solexa sequencing 147
(g) Sequence alignment and annotation 147
A.4 RESULTS ***148***
(a) Genome organisation 148
(b) True sequence Vs. NUMTS 148
(c) Solexa results 149
(d) Heteroplasmy or NUMTS 150
A.5 DISCUSSION ***150***
A.6 ACKNOWLEDGEMENTS ***152***
A.7 REFERENCES ***152***
A.8 RAW DATA ***155***
A.8.1 Contigs that translate 155
A.8.2 Contigs that don't translate - possible pseudogenes 159

B. APPENDIX B **165**

General Materials ***166***
B.1 LABORATROY EQUIPMENT ***166***
B.1.1 General 166
B.1.2 Glassware 166
B .1.3 Consumables 166
B.2 CHEMICALS AND REAGENTS ***167***
B.2.1 Laboratory chemicals and reagents 167
B.2.2 Enzymes 168
B.2.3 DNA Markers 168
B.3 KITS AND READY-TO-USE PRODUCTS ***168***
B.4 SOLUTIONS AND DYES ***169***
B.5 SOFTWARE AND PHYLOGENETIC PROGRAMS ***170***

C. APPENDIX C **171**

C.1 PRIMERS ***172***
C.2 EXTRACTION PROTOCOLS ***172***
C.2.1 Salting Out 172
C.2.2 Rapid mtDNA Extraction (Alkaline Lysis) 173
C.2.3 Bird Kits 174

C.3 PCR PROTOCOL ***175***
C.3.1 Long Range 175
C.3.2 Short Range 178
C4 CLONING PROTOCOL ***184***
C.4.1 Restriction Enzyme Digestion and Cloning 184
C.4.2 Precipitation of digestion and vector 185
C.4.3 Shotgun cloning of digested long range fragments 185
C.4.4 Transformation of Competent Cells 186
C.4.5 Selection of Transformed Cells 186
C.5 SEQUENCING PROTOCOL ***187***
C.5.1 Purification 187
C.5.2 Quantification 188
C.5.3 Sequencing 188
D. APPENDIX D **189**

LIST OF FIGURES

Figure 1.1 Schematic diagram to illustrate different changes being compared 9
Figure 1.2 Topography of Zealandian continent.. 11
Figure 1.3.Was there land in New Zealand?.. 12
Figure 1.4. Riding the West Wind Drift 20
Figure 1.5. The breakup of Gondwana 22
Figure 2.1 Timing of Gondwanan breakup and Australasian distribution map 33
Figure 2.2 New Zealand taxa distribution map... 36
Figure 2.3 Ensifera cladogram.. 42
Figure 2.4 Bayesian phylogram using nuclear DNA 43
Figure 2.5 Maximum likelihood phylogram using mitochondrial DNA... 45
Figure 2.6 Bayesian phylogram for mitochondrial DNA for the genus *Hemiandrus* 46
Figure 3.1 Unrooted Bayesian consensus network with minimal down-weighting....................................... 73
Figure 3.2 Unrooted Bayesian consensus network with maximum down-weighting..................................... 75
Figure 3.3 Rooted Bayesian phylogram with maximum down-weighting 77
Figure 3.4 Probability of finding predefined clades on a tree... 81
Figure 3.5 BEAST output tree for Neoaves... 94
Figure 3.6 Consensus tree of Neoaves (modern birds) before site-stripping 98
Figure 4.1 Rooted Bayesian phylogram using nuclear DNA... 112
Figure 4.2 Rooted Bayesian phylogram using mitochondrial DNA 114
Figure 4.3 Distribution map of Anostostomatidae used in this study 115
Figure A.1 Complete mitochondrial genomes 143
Figure A.2 SEQUENCHER alignments for *Hemiandrus pallitarsis*. 149

List Of Tables

Table 2-1 Sample information 56
Table 3-1 Probabilities of a predefined clade joining the tree as a single group, or as two subclades........ 66
Table 3-2 Support values 97
Table 4-1 Amino acid changes 110
Table 4-2 RCV and stemminess values 116
Table 4-3 OSF online Anostostomatidae heirarcy 124
Table 4-4 Sample information 126
Table A-1 *Hemiandrus pallitarsis* fragments 148
Table B-1 General laboratory equipment used 166
Table B-2 General glassware used 166
Table B-3 List of general consumables used 166
Table B-4 List of chemicals and reagents used 167
Table B-5 List of Enzymes used 168
Table B-6 DNA markers used 168
Table B-7 Kits and ready-to-use products 168
Table B-8 Software and programs used 170
Table C-1 Long range PCR mix 176
Table C-2 Long range primers used for Anostostomatidae 176
Table C-3 Long range primers used for the dollar bird 177
Table C-4 Long range primers used for the great potoo 177
Table C-5 Long range primers used for the owlet nightjar 178
Table C-6 Short range primers used for Anostostomatidae 179
Table C-7 Short range primers used for the dollar bird 181
Table C-8 Short range primers used for the great potoo 182
Table C-9 Short range primers used for the owlet nightjar 183

CHAPTER ONE

Science never pursues the illusory aim of making its answers final, or even probable. Its advance is, rather, towards the infinite yet attainable aim of ever discovering new, deeper, and more general problems, and of subjecting its ever tentative answers to ever renewed and ever more rigorous tests.

-Karl Popper

1959, p. 281

THESIS STRUCTURE AND SCOPE

In the following chapters I use molecular approaches to determine phylogenetic relationships. These relationships are then related to geologically dated events in Earth history and evaluated for potential mechanisms which underlie modern biogeographic patterns in chronological order. The combination of molecular approaches with biogeographic explanation, were used in an attempt to understand the evolutionary importance of large scale abiotic events (e.g. marine inundation, extraterrestrial impacts or continental dispersion) on biotic diversity.

This thesis reports on three geological events known to have occurred in the Southern hemisphere and suggested to be potential driving forces behind modern biotic diversity patterns; Oligocene marine inundation of Zealandia (*ca* 30 Ma[1]); extraterrestrial asteroid impact (65 Ma) and the break up of Gondwana (160 – 70 Ma). The extent to which these events influenced diversity is less well understood and may have facilitated diversification indirectly rather than directly. By understanding the phylogenetic relationships within each group, estimates of divergence time can be made and compared to the timing of the hypothesised stimulus.

This chapter gives a general background on the geographical scope and dating (fossils Vs. DNA), along with a brief summary of the three events (above) thought to have influenced contemporary biogeographic patterns. These address hierarchically, three time scales of evolutionary process as separate, self contained manuscripts;

(i) Australasian Anostostomatidae (*Chapter Two*),

(ii) Resolving deep Neoaves (*Chapter Three*) and

(iii) Determining Anostostomatidae relationships – Gondweta? (*Chapter Four*)

[1] Note: Throughout the thesis million years ago (Ma) denotes either the age of a particular geological event, or indicates a time span or duration.

The marine inundation of Zealandia and the break up of Gondwana are events that are specific to the Southern Hemisphere. The Anostomatidae were chosen as a case study to look at the impact of these events due to their predominantly southern distribution and to fossil evidence suggesting that representatives of the family date prior to the Gondwanan break up (Meads 1990; Martins-Neto 2007). Obviously the extraterrestrial asteroid impact was an event with world-wide ramifications, however, the radiation of the Neoaves was thought to occur in the Southern Hemisphere, giving both of our case studies a Southern Hemisphere flavour (Cracraft 1981).

The main thrust of the thesis is to gather and analyse new molecular data to enable more accurate estimation of phylogeny. To then interpret these phylogenies we need to know something about the biogeography of the organisms, the theory of continental drift, and also the theory of molecular clocks. The coverage of these topics is necessarily brief, as it is beyond the scope of this thesis to give a comprehensive background to all these factors.

The methods used to detect the affects on diversification (of target groups) will be dealt with via two case studies; relationships within the cricket (insect) family Anostostomatidae (*Chapter Two* and *Chapter Four*) and relationships within Neoavian birds (*Chapter Three*). An attempt to identify patterns and understand the effects these events had on diversification (if any) is made and future research directions are included in *Chapter Five*. In addition, appendices are included at the end of the thesis. *Appendix A* contains the preliminary results from the first study of mitochondrial genomes of two anostostomatids, *Hemiandrus pallitarsis* and *Hemideina crassidens*. In addition to basic information on gene organisation and composition, results will enable family specific primers to be designed for future work. *Appendix B* and *C* provide more detailed materials and methods respectively, used throughout this thesis but were omitted from chapters due to journal limitations. Lastly, *Appendix D* includes publication reprints for *Chapter Two* and *Chapter Three* with an additional publication carried out during this time (Phillips, M. J. and Pratt, R. C. 2008. Family-level relationships among the Australasian marsupial “herbivores” (Diprotodontia: Koala, wombats, kangaroos and possums), Molecular Phylogenetics and Evolution, 46: 594-605).

1.1 FACTORS THAT INFLUENCE OUR THINKING (BACKGROUND)

We need to improve our understanding of factors that influence phylogeny. To understand the what, how and when, one needs to consider factors such as dispersal, colonisation, adaptation and extinction. The Southern Hemisphere provides an excellent back-drop to investigate these issues.

1.1.1 Drift theory

Alfred Wegener, a German climatologist and geophysicist, suggested continental fragments could “drift” thus proposing an explanation for disjunct assemblages of related organisms. Wegener based his 1915 book (The Origin of Continents and Oceans) on the observable similarities between continental coastlines, such as the eastern coast of South America fitting neatly into the western coast of Africa. Further evidence from geological study showed that the African and South American continents (fit together) can also be aligned by comparing rock type and age with the same fossil species found on complementary continents (Macdonald 2006).

Until there was an understanding of continental drift, biological patterns seen throughout the Southern Hemisphere were explained by means of now sunken land bridges and dispersal. In 1928, thermal convection was suggested by Arthur Holmes to act like a conveyor belt in the mantle, carrying continents through the sea floor. In 1962, Harry Hess, a petrologist, proposed theories similar to Holmes’s, with some significant differences. Crucially, Holmes suggested continents moved through a rigid seafloor, while Hess proposed that continents move with the seafloor proposing magma as the lubricant. Robert Dietz, after discussion with Hess, published his theory which became known as ‘seafloor spreading’ (Dietz 1961). The process of continental drift may have lead to the observable patterns (biogeographic pattern) we see today.

1.1.2 Biogeography

Biogeography in a broad sense is defined as 'the study of variation in any biological feature (genetic, morphological, behavioural, physiological, demographic or ecological) across geographical dimensions including elevation and depth' (Lomolino 2000). Of particular interest in biogeographic studies are islands (of which, the Southern Hemisphere has an abundance). An island does not necessarily have to be an isolated landmass (continental [e.g. New Zealand] or volcanic [e.g. Hawai'i] in origin) as any population surrounded by barriers that limits dispersal can be considered an island population (for example an isolated mountain ridge surrounded by lowlands). Island equilibrium theory predicts that the number of species on an island should increase with island area and decrease with island distance to the mainland source population (MacArthur & Wilson 1963, 2001). Therefore in theory, populations on large islands exhibit lower extinction rates and islands far from the mainland will have lower colonization rates (MacArthur & Wilson 1963, 2001).

Island taxa therefore play important roles in our understanding of evolutionary processes such as colonization, invasion, and dispersal (MacArthur and Wilson 2001). As islands are 'isolated' from their mainland counterparts, the island populations will diverge over time from their respective mainland relatives due to genetic drift and changes in selective pressures, or both (Johnson et al. 2000). The use of phylogeography (biogeography and molecular systematics) enables reconstruction of a sequence of events, leading to the isolation of co-distributed groups (of ancestral species). Isolation can result in subsequent allopatric speciation and biotic diversification (Riddle 2005). Contemporary species' distribution patterns are influenced by a combination of historical factors. By assessing phylogeographic patterns (using molecular diversity, comparing geographic distributions) and by applying molecular clocks, hypothesis of species origins can be examined.

In addition to geological barriers, climatic fluctuations will be important factors affecting a population's ability to survive. Throughout the earth's history organisms have had to; either adapt to environmental changes, move to more hospitable areas, or face extinction.

Climatic changes not only include climate oscillations (e.g. tropical, temperate, polar) but also factors such as glaciations, flooding, aridification, or volcanic eruptions, which create new niches and change the selective pressures acting on organisms.Consequences of climate change will be both latitudinal and elevational, isolating populations in refugia (such as mountain tops for alpine species) and potentially leading to speciation or extinction due to natural selection. However, the timing of the proposed event has to coincide with the divergence event, if it is to be shown as an important factor in creating new lineages.

1.1.3 Getting a date in the modern era - Fossils Vs. DNA

Fossils were once the only way to gather information about the time when ancestors (of extant organisms) lived, and for a long time it was essentially impossible to determine the geological age of many evolutionary lineages. Limitations of taking the fossil record at face value are illustrated by ratites in New Zealand; where male and female moa were described as separate species based on what is now recognised as extreme sexual dimorphism, with females much larger than their conspecific partners (e.g. Baker et al. 2005; Macdonald 2006). With the advent of molecular genetics in the 1960s the molecular clock hypothesis was proposed (Zuckerkandl and Pauling 1965). It was found that the rate of molecular evolution for each protein studied had a relatively constant, characteristic rate over time. By counting the number of accumulated changes (throughout the sequence) between lineages, it was concluded that these molecules could be used to estimate the evolutionary distance (time) between organisms (Hoy 1994) and thus serve as a rough molecular clock (for review see Bromham and Penny 2003).

The clock can then be calibrated on the prediction that basic processes (DNA replication, transcription, protein synthesis and structure, and metabolism) are remarkably similar in organisms, and thus the proteins and RNA’s that carry out these functions will be highly conserved (Hoy 1994).

However, studies of different protein sequences show that rates of change vary among different proteins (McLaughlin and Dayhoff 1972), and lineages (Hoy 1994), demonstrating that there is not ‘one universal clock’.

There have been suggestions that some aspects of "explosive radiations" may speed up molecular clocks (Conway Morris 1998; Lee 1998; Valentine et al. 1999; Pagel et al. 2006), though no mechanism has been proposed. Furthermore, the accuracy of a clock is dependant on the appropriateness of the substitution model used, the calibration date and the use of multiple, independent data points (Smith and Peterson 2002; Bromham 2003; Bromham and Penny 2003).

Several reasons for incongruence between fossils and molecular dating are apparent (Near and Sanderson 2004). Clearly the fossil record will commonly be incomplete which generally leads to an underestimation of any given lineage (Marshall 1990a). The error associated with the estimate will be dependent on the difference between the fossil age and the actual lineage age (Springer 1995). In addition, inaccurate estimations may be caused by erroneous placement of fossils, incorrect phylogenies or inaccurate geological age estimates (Lee 1999; Benton and Ayala 2003; Conroy and van Tuinen 2003). Lastly, if minimum ages are applied to the crown group that the fossil belongs to, instead of the stem group from the phylogenetic tree, errors will occur (see figure 1.1; Doyle and Donoghue 1993; Magallón 2004).

Estimates of dates provided by molecular clocks may also be biased or imprecise due to the difficulties in adequately modelling sequence evolution (Shields 2004; Heads 2005). Mathematical models of sequence evolution attempt to correct for multiple substitutions at a site and rate variation across sites. Welch and Bromham (2005) give a review of how different phylogenetic dating software packages cope (or don't cope) with rate variation across sites. The currently used models may be more or less adequate for modelling particular genes (Stevens and Schofield 2003). A recent paper by Phillips (2008) has shown that poorly fitting models can lead to bias in branch length estimation and hence poor estimates of dates.

Despite these limiting factors, application of molecular clocks to analyses, enabled hypothesis testing and elucidation of timescales for biodiversification (Arbogast et al. 2002; Bromham 2003; Burridge et al. 2008). Only through the use of multi-disciplinary approaches will we advance our knowledge further.

1.1.4 Problems with comparisons

To date there is still much contention over the adequacies (or inadequacies) of our current knowledge of the fossil record, and the importance of fossils. Absence may not be evidence, but simply a 'result' of factors that inhibited or degraded beyond recognition the fossil evidence critical to a particular species, region or time. The apparent absence of lineages from the fossil record could simply be an absence of identifiable morphological structure (eg. Archibald et al. 2001). As suggested by Darwin (1859 p. 424) '…*organic remains will probably first appear and disappear at different levels, owing to the migrations of species and to geographical changes*.'

By not taking the deficiencies of the fossil record into account, researchers have compared different events using incompatible data sets (Valentine et al. 1999; i.e. stem group fossils, crown group fossils or molecular data e.g. Wray 2001). That is, sequence data reflect the time since the two taxa last shared a common ancestor (their divergence time) while fossils reflect the appearance of anatomical structures that define a specific group at a particular time (its origin) (Wray 2001). The modern diagnostic morphological features (e.g. crown groups) we use may have arisen much later than the actual divergence event, and some lineages could have lead to more than one extant order (see figure 1.1). Thus, it is crucial to draw distinction between the time of initial divergence (of a given group) and the age of the last common ancestor (of all known diagnosable members of a group). On a phylogenetic tree, the origin (for a given taxon) is indicated by the point of divergence between sister taxa (node 1, figure 1.1). The initial divergence may not include the characteristic morphological features of later descendants, acquiring those much later on (i.e. diverged phylogenetically before diverging morphologically see node 2, figure 1.1) (Soligo et al. 2007).

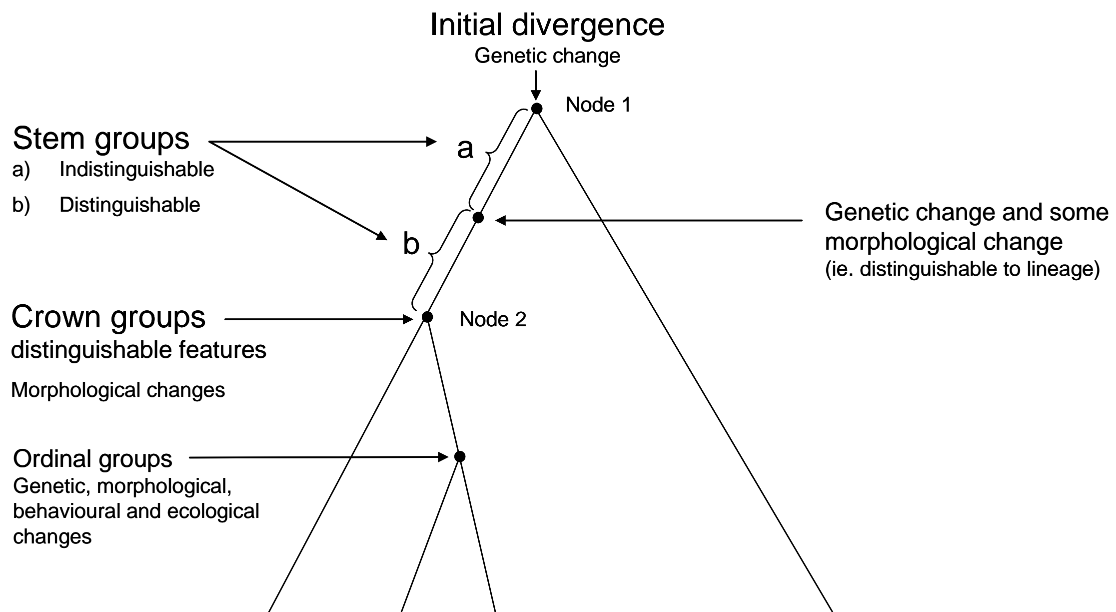


Figure 1.1 Schematic diagram to illustrate different changes being compared

Stem groups may be indistinguishable morphologically yet be genetically divergent. Crown groups are distinguishable both morphologically and genetically. When studies try to compare these two separate 'divergences' incongruence between datasets is apparent.

Fossils record morphological differences of the better adapted biota, in response to ecological and climatic fluctuations, at a particular time and place. This may have little to do with the physical "origins" of the species (i.e. genetic stem group). These different events (morphological and genetic change) will occur at different times which may lead to the 'apparent absence' of lineages from the fossil record, even though they existed at that time (Marshall 1990b; Archibald et al. 2001). Acknowledging the imprecision of the fossil record and molecular clock estimates does not deny their usefulness in evolutionary biology (Bromham et al. 1999). Instead, their use allows the assessment of their compatibility with a given geographic or geologic hypothesis.

1.2 CASE STUDIES

This thesis reports phylogenetic analyses of two case studies (Anostostomatidae weta and Neoavian birds) with regard to external physical events that are assumed to have influenced biotic diversity.

These studies evaluate the relative importance of such events for explaining contemporary diversification by application of molecular clocks to assess the congruence between timing and hypothesised stimulus. In each instance, there is little doubt the physical event has occurred, however whether it has had a major affect on diversity is still debatable. If we find a strong correlation between these known physical events and patterns of diversification, this would provide supporting evidence that microevolutionary processes (mutation, genetic drift, selection) are insufficient to explain current patterns of diversity.

1.2.1 Testing the depths of the Oligocene (Chapter Two)

Since the time of Darwin, the origin of New Zealand's biota has sparked intense debate. Darwin suggested that '…*although New Zealand is here spoken of as an oceanic island, it is in some degree doubtful whether is should be so ranked; it is of large size, and is not separated from Australia by a profoundly deep sea; from its geological character and the direction of its mountain-ranges, the Rev. W. B. Clarke has lately maintained that this island, as well as New Caledonia, should be considered as appurtenances of Australia*' (Darwin 1859 p.531). McGlone (2005) went further and referred to New Zealand as the 'fly paper of the Pacific' suggesting, like many others, that the contemporary biota has arrived via dispersal (e.g. Fleming 1979; Pole 1994). Alternatively, New Zealand has been described as a living icon of Gondwanan heritage, referred to by some, as a 'Moa's ark' (Bellamy et al. 1990) making a case for at least some relictual survivors dating back over 80 Ma (e.g. Stöckler et al. 2002; Knapp et al. 2007). The question of which parts of the biota travelled here with the land, and which reached its shores long after the continents fragmented (Gibbs 2006) should be possible to answer with the use of molecular studies.

Although continental drift has occurred, the biotic origins in parts of the South Pacific, including New Zealand, are accredited to the result of more recent geological activity than Gondwanan breakup. Contemporary New Zealand is very different from the landmass that broke away from Gondwana approximately 80 Ma. The relatively large (3,500,000 km^2) continental fragment Zealandia is now mostly submerged and stretches to New Caledonia in the north, Chatham Islands in the east and beyond sub-Antarctic islands to the south, stretching from approximately latitude 18° S to 56° S (see figure 1.2; Macdonald 2006).

The majority of the submerged block reaches a depth of no more than 1,000 m while the ocean floor reaches depths of 5,000 m (Macdonald 2006).

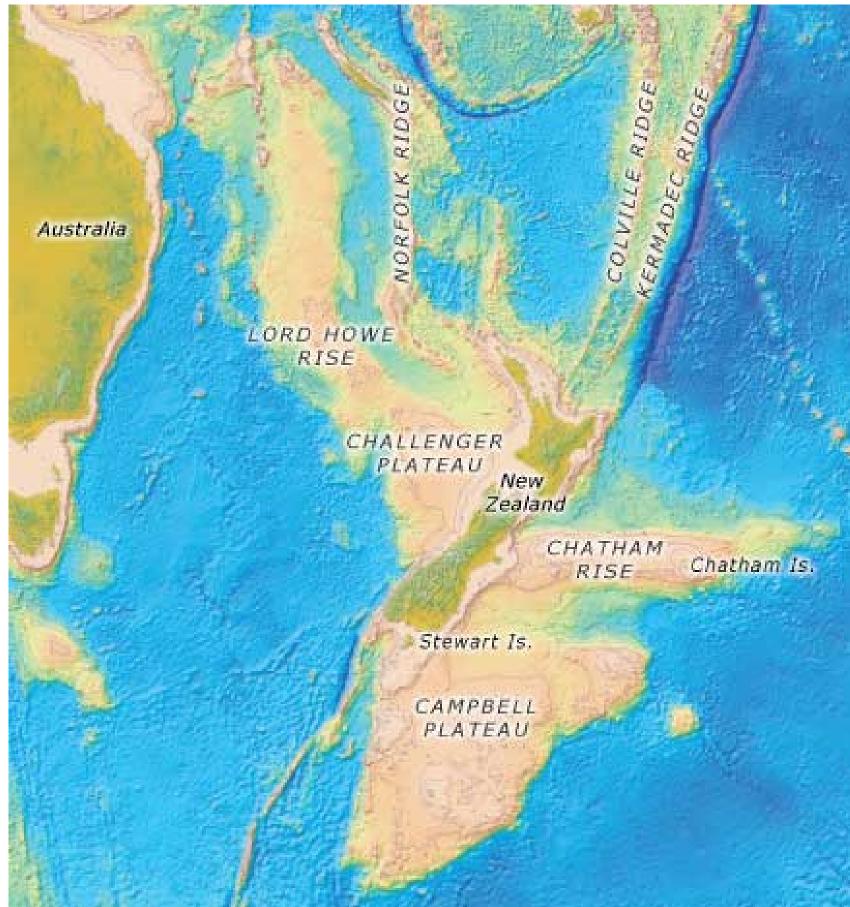


Figure 1.2 Topography of Zealandian continent

Approximately 80 Ma Zealandia began to break away from Gondwana. Today New Zealand makes up the largest land area along with surrounding islands of New Caledonia, Chatham Islands and New Zealand's sub-Antarctic islands. Source: GNS Science.

One of the most controversial epochs in New Zealand's history is the Oligocene (*ca* 35 - 24 Ma) and it has been well established that Zealandia underwent substantial marine inundation (Landis et al. 2008) during this time. Uncertainty remains over the extent of this submergence (e.g. Stöckler et al. 2002; Knapp et al. 2007; Pratt et al. 2008), with hypotheses ranging from; complete submergence, to approximately 18% (50,000 km^2) of the current land above sea level (Cooper and Cooper 1995). This begs the question, how much land does a particular taxon need, and do area models of biogeography adequately explain diversity changes (Fordyce 1991)?

Even if the submersion was not complete (figure 1.3), the severe reduction in land surface would have caused a taxonomic bottleneck, affecting the availability of different niches, thus potentially influencing the evolution of Zealandian biota (Fleming 1979; Stevens 1980).

Cooper and Cooper (1995) testing the Oligocene bottleneck hypothesis, predicted that many modern terrestrial endemics (genera and families) would originate from the Oligocene or early Miocene. Many New Zealand lineages have been shown to have diversified in the Miocene (23 - 5 Ma) rather than the late Cretaceous (100 – 65 Ma) as predicted from Oligocene survival (Cooper and Cooper 1995; Waters et al. 2000; Baker et al. 2005; Knapp et al. 2005; Shepherd and Lambert 2007; Harshmann et al. 2008).

.

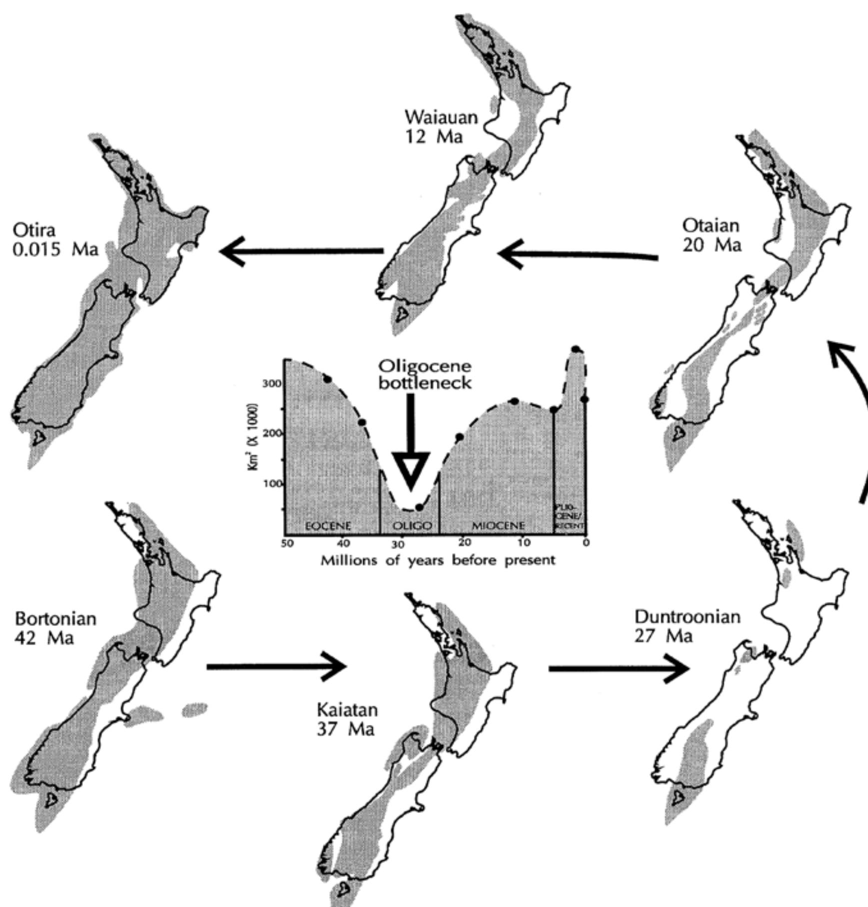


Figure 1.3 Was there land in New Zealand?

Hypothetical changes in palaeography in the New Zealand region through the Cenozoic based on present day outline. Shown in the centre, generalised curve showing the change in land area through the Cenozoic (taken from Cooper & Cooper 1995).

Much contention arises when addressing the degree of submersion. Most recently Landis et al. (2008) reported compelling evidence for total marine inundation in the southern South Island, finding no evidence for Cretaceous (or Paleocene) terrestrial sediment. In addition, the non-marine sediments found in coastal Otago are coarse-grained deposits; the fine-grained mud, expected from old age, meandering rivers, were absent. Given their findings, Landis et al. (2008) suggested a revised approach, proposing that the equally valid assumption of complete marine inundation, be taken as the default hypothesis.

New Zealand has undergone extensive volcanism and tectonic activity, the most striking and recent of which resulted in uplift of the Southern Alps (*ca* 2 - 5 Ma) over the Alpine fault in New Zealand's south island. Clearly, the New Zealand region has been a chaotic one and as such it is argued that most, if not all, of New Zealand's current fauna and flora represent 'recent' (post Miocene) arrivals (Stevens 1980; Pole 1994; Cooper and Cooper 1995; Landis et al. 2008). This view is not shared by those advocating that New Zealand represents a snap-shot of Gondwanan life, isolated for at least 80 Ma (Gibbs 2006). Certainly there are examples of genera that appear to have a vicariant history, for example Agathis (Stöckler et al. 2002; Gibbs 2006; Knapp et al. 2007).

Unfortunately New Zealand's turbulent history is not amenable to the preservation of land animal fossils. New Zealand has a relatively good marine fossil record (e.g. Hornibrook 1978; Beu and Maxwell 1990), and good pollen records (*Nothofagus* (e.g. Hornibrook 1978); podocarps and *Araucaria* (Bowen 1978)). However, many land animal fossils will have been obliterated by the numerous geological events that have occurred throughout time (erosion, sea level changes, mountain building, volcanic activity). The discovery of terrestrial animal fossils older than 1 million years has been relatively rare. Recent excavations from three locations in St Bathans, Central Otago, (South Island), have uncovered a sphenodontid, a crocodilian, geckos, skinks, bats, a minimum of 24 avian taxa (Molnar and Pole 1997; Worthy et al. 2007; Landis et al. 2008), and a terrestrial mammal (Worthy et al. 2006), thus going some way towards improving New Zealand's land animal fossil deficit.

In the absence of good fossils, sister taxa occurring on opposite sides of an oceanic barrier are useful for studying competing explanations concerning a relatively old vicariance event or more recent oceanic dispersal in a region. New Zealand has been suggested as an endemism hot spot (Daugherty et al. 1993), forming a haven for distinctive relict taxa like tuatara (*Sphenodon*), leiopelmatid frogs, peripatus (Onychophora), southern beech (*Nothofagus*) and kauri (*Agathis australis*) (Gibbs 2006). Among the most interesting components of New Zealand's terrestrial fauna, are insects of the family Anostostomatidae (Orthoptera), known in New Zealand by their Maori name, weta.

This family is useful for testing whether New Zealand lineages survived the Oligocene drowning. If the origins and radiations within New Zealand pre-date the Oligocene, dating endemic species and genera should result in ages older than 25 Ma. If New Zealand's taxa represent recent arrivals, it is expected that their nearest relatives should be found on a nearby landmass, and their diversity within New Zealand would post-date the Oligocene. Anostostomatidae are represented in all three major Australasian landmasses (Australia, New Caledonia and New Zealand) therefore predicted origins can be tested.

Recognition of long distance dispersal (LDD) as an explanation for broad distributions of sister taxa throughout much of the Southern Hemisphere has increased in recent years (Pole 1994; Hurr et al. 1999a; Muñoz et al. 2004; de Queiroz 2005). Evidence of this process for New Zealand is abundant (Gaskin 1970; Pole 1994; Cook and Crisp 2005a; Cook and Crisp 2005b; Knapp et al. 2005), including examples from many recent molecular studies (Waters and White 1997; Winkworth et al. 2002; Sanmartin and Ronquist 2004; Knapp et al. 2005). Numerous studies of taxa in the Pacific region have been conducted and there are good historical records showing that birds and insects are common migrants across the Tasman Sea from Australia to New Zealand (Fox 1973; Watt 1975; Michaux 1991). The silvereye (*Zosterops lateralis*) is an example of successful dispersal, first seen in Southland in 1856 and since spreading throughout New Zealand. More recently, the royal spoonbill (*Platalea leucorodia*), swallows (*Hirundo tahitica*), the gray teal (*Anas gracilis*), white-faced heron (*Ardea novaehollandiae*) and Australian coot (*Fulica atra*) (Macdonald 2006) have arrived and colonised successfully. Fewer examples have been documented of taxonomic affinities between New Zealand and New Caledonia, but a number exist (e.g. Chambers et al. 2001; Lee et al. 2001; e.g. Buckley et al. 2002).

There is even some evidence for dispersal against the prevailing winds, with taxa speciating in New Zealand before dispersing around the south Pacific (e.g. Wright et al. 2000; Boon et al. 2001; Winkworth et al. 2002).

Although increasing evidence is being found that dispersal plays an important role for many taxa in New Zealand there are a number of examples including the Kauri, leipelmatid frogs, peripatus, and tuatara whose relationships suggest a vicariant origin. It only takes one solid example of vicariance to show that the Oligocene drowning could not have been complete. Relationships between biological diversity and habitat area as well as between the evolution of species and geographic change (Fordyce 1991) are important and complex. The timing of events such as dispersal and geological change are crucial to understand the underlying processes at work in the region.

This chapter will investigate the phylogenetic relationships within the Australasian Anostostomatidae with a focus on the relationships of New Zealand taxa to potential source populations. Standard molecular phylogenetic methods were used to test monophyly of the three ecological groups previously described in Trewick and Morgan-Richards (2005). Dating methods include dates both prior to the Oligocene (in this case marking Zealandian fragmentation at *ca* 80 Ma) and after the Oligocene (less than 35 Ma). In addition to tree topology, mutation rates will be examined and compared to published data. In simplest terms, molecular data and dating in *Chapter Two* predicts that Anostostomatid lineage diversity in New Zealand should increase after the Oligocene marine inundation if submergence was complete. If the inundation was only partial, there should be evidence in tree structure (ie. lineages older than 35 Ma). Therefore, tree topology and mutation rates will be used to determine timing and origin of diversity in New Zealand.

1.2.2 What impact did the asteroid have? (*Chapter Three*)

Throughout the earth's history numerous extinction events have played an important role in the evolution of new species and their ability to inhabit new niches. Several major extinction events have been described; Late Ordovician (450 Ma), Late Devonian (364 Ma), Permian-Triassic [P - T] (250 Ma), Late Triassic (200 Ma) and Cretaceous-Tertiary [K - Pg formerly K - T] (65 Ma) (Benton 1995)**.** By far the most studied extinction event is the K - Pg extinction. Şengör et al (2008) classified three classes of extinctions: (i) those that take place because of environmental change resulting from physical processes indigenous to the planet [i.e. volcanic eruptions]; (ii) those that take place because of extraterrestrial interference [i.e. asteroid impacts]; and (iii) those that take place because of events within the biosphere [i.e. climate change](Şengör et al. 2008). In addition, purely biological changes (such as a new predatory group interacting with a naïve biota), could also look like a "mass extinction" but this option is rarely considered. Nevertheless, it is probably the best hypotheis for the current "mass extinction" of megafauna over the last 10,000 years.

In general, it is assumed that common and widespread genera and species are more resilient and therefore less affected by extinction events than species that have restricted distributions or special requirements. These widespread groups are more likely to find refuges and be able to recolonise effected areas, so local events may only wipe out local populations. By inhabiting a wider range of geographical areas organisms are adapted to more conditions as a whole, increasing the potential for at least some organisms to further adapt to changing conditions. It has been suggested that environmental changes may even facilitate speciation events (Hoffmann and Parsons 1997) by enabling expansion into newly vacant niches (e.g. Krakatoa), or niches not previously available due to competitors.

The most cited example of external forces (processes) "driving" diversification (patterns) would arguably be the asteroid impact on the Yucatán Peninsula in Mexico 65 Ma**.** The devastation of the asteroid impact was to be a worldwide phenomenon. Alvarez et al. (1980) reported iridium increases in deep-sea limestone exposed in Italy, Denmark and New Zealand. New Zealand also shows a distinct fern spike, replacing a diverse pollen fossil flora shortly after this time (Vajda et al. 2001).

This event marks a contentious era in the evolution of birds and mammals. The extent to which the asteroid impact affected the evolution of biota is still hotly debated. Obviously organisms near the impact would have been obliterated but there is evidence for its effects being global (eg. see Alvarez et al. 1980; Orth et al. 1981; Bohor et al. 1984; Wolfe and Upchurch 1987; Vajda et al. 2001; Vajda and McLoughlin 2004; Vajda and McLoughlin 2007).

Alvarez et al. (1980) suggest (without conclusive evidence) that the likely biological effects of the impact acted like a chain reaction stemming from the temporary absence of sunlight leading to photosynthesis shut-down. The open ocean food chain is based on microscopic floating plants, which show nearly complete extinction. Higher levels in the food chain were affected too, with the nearly total extinction of foraminifera and the complete disappearance of the belemnites, ammonites and marine reptiles (Alvarez et al. 1980). It is suggested that the land based plants were severely affected also, which in turn lead to the extinction of large herbivourous and subsequently carnivourous animals dependent on this vegetation.

Russell (1979) states that; "no terrestrial vertebrate heavier than *ca* 25kg is known to have survived the extinctions". Smaller terrestrial vertebrates such as ancestors of mammals and birds however did survive. An alternative hypothesis proposed by Pond (1977) suggested that the difference in parental care between reptiles and mammals may partially explain the radiation of mammals and birds 65 Ma. She argued that underdeveloped juvenile dinosaurs might have had the greatest difficulty in obtaining sufficient food while juvenile mammals and birds (with undeveloped masticatory and digestive tracts) are dependent on parental care (e.g. suckling or regurgitation) thus less likely to have been affected.

However, this does not mean that without this event (and others like it), that birds and mammals (in this case), would not have competed with dinosaurs and pterosaurs whether directly or indirectly. A modern example to illustrate the type of competition occurring between populations is shown by the effect of small rodents (such as the Polynesian food rat, or kiore, *Rattus exulans*) on *Spenodon* (tuatara) although similar effects are found with small lizards such as skinks. These small rodents, average 60 - 80g, are shown to displace tuatara (up to 1.5kg in some populations).

Research has shown that adult tuatara survive in the presence of kiore, but recruitment of juveniles ceases (Cree et al. 1995). It is still not certain whether there is direct predation on small tuatara, and/or indirect competition for resources that is effecting the decline in tuatara populations, however removing rodents allows juvenile tuatara to survive (Hurr et al. 1999b; Towns et al. 2006).

This idea, that organisms need an external catastrophic event to enable one group to surpass another downplays the importance of competition, natural selection and the processes of normal evolution (i.e. microevolutionary processes). The fossil record shows that mammals, birds and dinosaurs coexisted for at least 100 million years before the K - Pg extinction event so it is unclear why they did not replace the dinosaurs earlier if they were better adapted! At this stage, we are unable to answer questions on the speed of extinctions (either globally or locally), what impact climate change would have had during this time and how much the changes in environment are associated with these evolutionary events. These questions need to be answered to understand the processes involved in evolution.

Of particular concern here is whether members of Cracraft's six groups come together. Members of the six groups are represented with the addition of the nine mitochondrial genomes and development of weighting methods is used to retain phylogenetic information (and resolve the basal polytomy). Tree topology will be used to assess the six main groups (ie. do they form these groups?) and probability calculations used to calculate their likelihood of coming together. Molecular data and dating methods examined in *Chapter Three* predicts an increase in lineage diversity occurring after the asteroid impact 65 Ma, if it (impact) was necessary to produce the diversity of Neoaves (modern birds) seen today.

1.2.3 Dispersal or drift? Getting around on Gondwana (*Chapter Four*)

For over 150 years, biologists have been speculating about the biogeographic mechanisms for patterns observed in Southern Hemisphere taxa. Historically many biogeographers relied on *ad hoc* dispersalist scenarios invoking landbridges and repeated migration from the north to explain widespread biota under the assumption of continental stability (Waters 2008 and references therein).

However following the development and acceptance of plate tectonic theory and continental drift, a paradigm shift occurred and explanations for southern taxa shifted dramatically from dispersal to vicariance (Croizat et al. 1974; Ball 1975). From then on, advocates of dispersal received ongoing criticism (Croizat et al. 1974; Rosen 1978; Craw 1979; Nelson and Ladiges 2001; Ebach and Humphries 2002) as dispersal was seen as an untestable hypothesis, and so “not real science”.

West Wind Drift (WWD) was implicated in a number of studies during the 1960s and 1970s (figure 1.4; Fell 1962; Fell 1967; McDowall 1978; Fleming 1979). WWD links all southern continents providing a backdrop for comparative analyses and hypothesis testing, patterns of which should be strong on oceanic islands (in particular) demonstrating that dispersal occurs in an eastwards (down-wind) direction from the point of entry. Indeed many molecular studies have provided strong evidence for the significant role dispersal has played in biotic patterns around the southern Hemisphere (Jordan 2001; Winkworth et al. 2002; Chriswell et al. 2003; Sanmartin and Ronquist 2004; Waters and Roy 2004; de Queiroz 2005). By contrast, little evidence for the biogeographic role of vicariance has been demonstrated and tends to fail to withstand detailed study with molecular markers (Trewick 2000; Trewick and Wallis 2001).

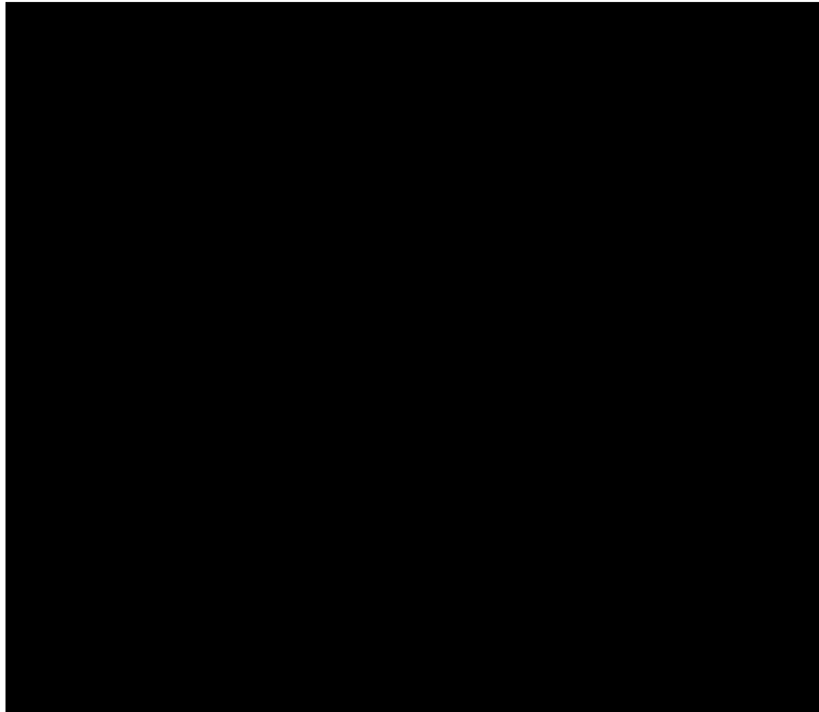


Figure 1.4 Riding the West Wind Drift
West Wind Drift (WWD) dispersal of echinoderms throughout the Southern Hemisphere showing species trail in an eastwards direction from the point of entry (From Fell 1962)

Whether in support of vicariance or dispersal, many hypotheses tend to be too simplistic yet can be useful as false models. If distributions are "driven" by geological rather than biological processes, we simply need to examine the earth's history in order to reveal biological patterns.

Vicariance theory (for all its faults) facilitates the development of testable phylogenetic hypotheses whereby dating becomes extremely important (Buckley et al. 2001; Cracraft 2001; Waters et al. 2001; Givnish and Renner 2004; Boyer and Giribet 2007; Burridge et al. 2008). For instance, a rifted fragment of Gondwana may lose its biota entirely, however be recolonised down-wind from another fragment of Gondwana, therefore still being "Gondwanan" in origin. The timing of divergence in such cases should distinguish between ancient lineages and more recent dispersal events.

1.2.4 Patterns and predictions

During the Palaeozoic era (542 - 251 Ma) (ICS 2004) the position of Gondwana was very different from where continental fragments are today. Gondwana was then situated in the Northern Hemisphere while Laurasia (consisting of Europe, North America and Asia) was in the south. These supercontinents progressively rotated and Gondwana moved into the Southern Hemisphere (Stöckler 2001). McLoughlin (2001) gives an overview of the breakup of Gondwana which began over 160 Ma between Africa and the rest of Gondwana (figure 1.5). By 80 Ma the Tasman Sea opened between Australia and New Zealand and New Zealand was completely separated from Gondwana (*ca* 65 Ma). South America broke away from Antarctica/Australia (*ca* 32 - 28 Ma) with the opening of the Drake Passage and drifted northward before colliding with North America. The separation of Gondwana was complete shortly thereafter, with the fragmentation of Antarctica and Australia giving rise to the most important current in the Southern Ocean, the Antarctic circumpolar current (ACC).

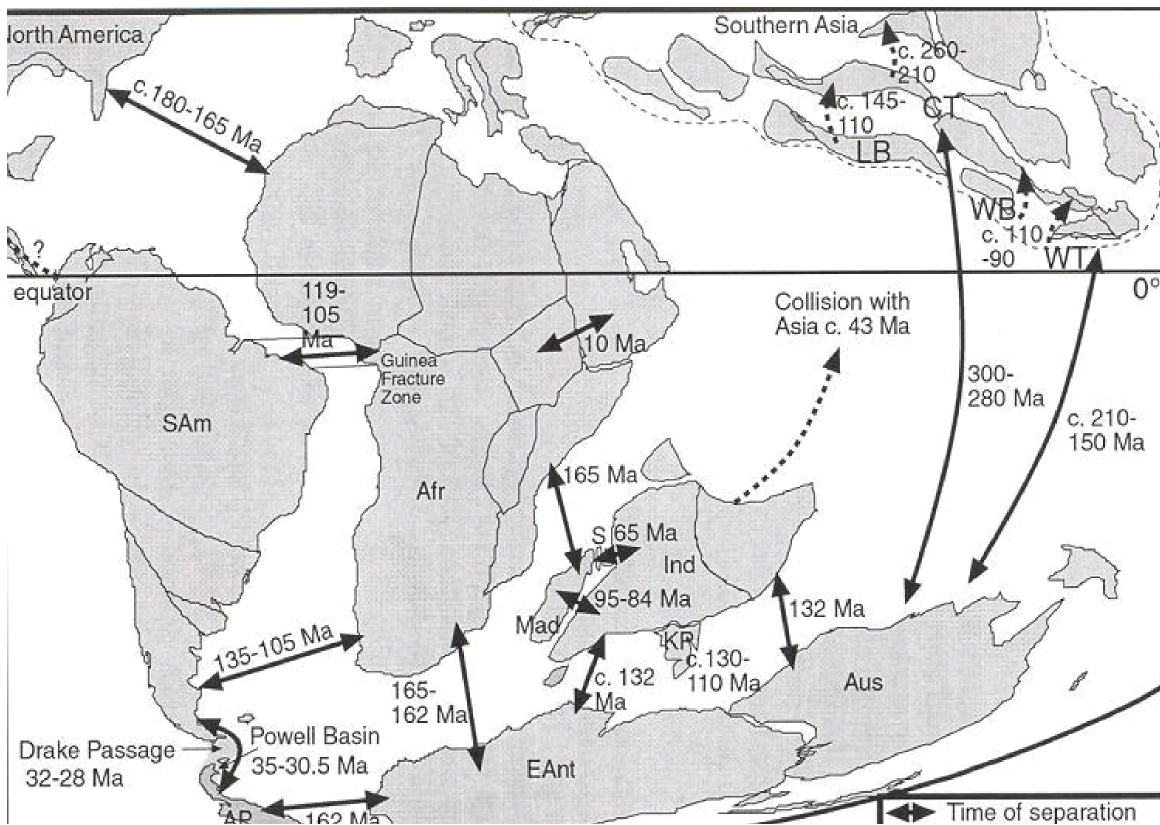


Figure 1.5. The breakup of Gondwana

The fragmentation of Gondwana with estimated dates of separation. (Taken from McLoughlin 2001).

The ACC, encircles the Antarctic continent, flowing eastward through the southern parts of the Atlantic, Indian, and Pacific Oceans. New Zealand is positioned straddling the 'Roaring forties' - a characteristic, strong westerly wind which drives surface water in an easterly direction (i.e. WWD). This ACC along with the WWD, has played an important role in dispersal for organisms that have recently colonised sub-Antarctic islands (e.g. Wagstaff et al. 2007). If taxa represent more recent dispersal events, the direction and likely origins should be distinguishable using the known patterns occurring today (figure 1.4), which are most likely to have occurred in an easterly direction around the southern continents**.**

Using the conventional chronology of Gondwanan breakup (figure 1.5), patterns can be predicted. If the ancestral species occupied the ancient land and subsequently speciated in association with the sequential geological fragmentation of the landmass, we expect to find closer affinities between South American and Australian taxa, then New Zealand/New Caledonian taxa with Madagascan and South African taxa being sister to all the others.

In this case, biotic patterns would be consistent with geology. If however, in the course of this same history an extinction event took place (e.g. New Zealand), and recolonisation occurred from a nearby landmass (e.g. Australia), there would be conflict between the biotic pattern and the geology. In addition, speciation could have occurred in the ancestral land, pre-dating any geological events and resulting in a pattern incongruent with geological history. Similar biogeographic patterns may have very different underlying causes and may be obscured by events such as extinctions so caution should be taken not to assume that similar patterns means similar processes.

In order to assess the Australasian findings from *Chapter Two* in a family context, representatives from other Gondwanan fragments were included. The systematics of the family has historically been problematic so tree topology was used to comment on the current taxonomic hierarchy. By invoking the two most extreme (and simplistic) scenarios for vicariance and dispersal, tree topology was used to investigate the relationships within Anostostomatidae in the absence of a molecular clock. With the extreme scenario of vicariance, it is predicted that tree topology will follow the conventional order and timing of Gondwanan breakup. That is in this case reciprocal monophyly of South African taxa (basal), followed by Madagascan taxa, Zealandian taxa and finally Australian and South American taxa. Divergence from this order would support a contribution from dispersal. Alternatively, no geographic structure would indicate dispersal was occurring throughout the region and origins should be possible to determine.

1.3 CONTRIBUTION OF OTHERS TO THE THESIS

Except where acknowledgment is given, the work contained in this thesis is my own. Contributions to chapters that were published with co-authors are outlined as follows: my academic supervisors David Penny, Mary Morgan-Richards and Steve Trewick are co-authors in recognition of their contribution to the research conducted. David Penny is acknowledged for his contributions for valuable insight and discussion on drafts and providing essential funding (Marsden Fund) and resources.
Mary Morgan-Richards and Steve Trewick are acknowledged for contributions to initial project design for Anostostomatid research, and for useful comments and discussion on drafts. Assistance with phylogenetic methods including site stripping (*Chapter Three*) and

the application of BEAST (*Chapter Two* and *Chapter Three*) along with useful comments on drafts was given by Matt Phillips. Gillian Gibb contributed sequence data and comments on a manuscript (*Chapter Three*), and Mike Hendy is responsible for the calculations in Appendix 3.9.2 (*Chapter Three*).

I was responsible for the project design, molecular lab work, alignments and the bulk of the phylogenetic analysis. Of all authors I contributed the most to the writing of submitted manuscripts although valuable discussion and contributions were made by my co-authors as outlined above.

1.4. REFERENCES

Alvarez, L. W., W. Alvarez, F. Asaro, and H. V. Michel. 1980. Extraterrestrial cause for the Cretaceous-Tertiary extinction: experimental results and theoretical interpretation. Science **208**:1095-1108.

Arbogast, B. S., S. V. Edwards, J. Wakeley, P. Beerli, and J. B. Slowinski. 2002. Estimating divergence times from molecular data on phylogenetic and population genetic timescales. Annual Review of Ecology and systematics **33**:707-740.

Archibald, J. D., A. O. Averianov, and E. G. Ekdale. 2001. Late Cretaceous relatives of rabbits, rodents and other extand eutherian mammals. Nature **414**:62-65.

Baker, A. J., L. J. Huynen, O. Haddrath, C. D. Millar, and D. M. Lambert. 2005. Reconstructing the tempo and mode of evolution in an extinct clade of birds with ancient DNA: the giant moas of New Zealand. Proceeding of the National Academy of Science, USA **102**:8257-8262.

Ball, I. R. 1975. Nature and formulation of biogeographical hypotheses. Systematic Zoology **24**:407–430.

Bellamy, D., B. Springett, and P. Hayden. 1990. Moa's Ark, the voyage of New Zealand. NY: Viking, New York.

Benton, M. J. 1995. Diversification and extinction in the history of life. Science **268**:52-58.

Benton, M. J., and F. J. Ayala. 2003. Dating the tree of life. Science **300**:1698.

Beu, A. G., and P. A. Maxwell. 1990. Cenozoic Mollusca of New Zealand. Pp. 518 in G. Gregory, ed. New Zealand Geological Survey Paleontological Bulletin 58, Wright & Carman, Ltd., Upper Hutt, New Zealand.

Bohor, B. F., E. E. Foord, P. J. Modreski, and D. M. Triplehorn. 1984. Mineralogic evidence for an impact event at the Cretaceous-Teriary boundary. Science **224**:867-869.

Boon, W. M., C. H. Daugherty, and G. K. Chambers. 2001. The Norfolk Island Green Parrot and New Caledonian Red-crowned Parakeet are distinct species. Emu **101**:113-121.

Bowen, F. E. 1978. Ohai Coalfield, Southland. Pp. 379-398 *in* R. P. Suggate, ed. The geology of New Zealand. E. C. Keating Government Printer, Wellington.

Boyer, S. L., and G. Giribet. 2007. A new model Gondwanan taxon: systematics and biogeography of the harvestma family Pettalidae (Arachnida, Opiliones,

Cyphophthalmi), with a taxonomic revision of genera from Australia and New Zealand. Cladistics **23**:337-361.

Bromham, L. 2003. Molecular clocks and explosive radiations. Journal of Molecular Evolution **57**:S13-S20.

Bromham, L., and D. Penny. 2003. The modern molecular clock. Nature - Reviews **March 2003**:216-224.

Bromham, L., M. J. Phillips, and D. Penny. 1999. Growing up with dinosaurs: molecular dates and the mammalian radiation. TREE **14**:113-118.

Buckley, T. R., P. Arensburger, C. Simon, and G. K. Chambers. 2002. Combined data, Bayesian phylogenetics and the origin of the New Zealand Cicada genera. Systematic Biology **51**:4-18.

Buckley, T. R., C. Simon, H. Shimodaira, and G. K. Chambers. 2001. Evaluating hypotheses on the origin and evolution of the New Zealand alpine cicadas (Maoricicada) using multiple-comparison tests of tree topology. Molecular Biology and Evolution **18**:223-234.

Burridge, C. P., D. Craw, D. Fletcher, and J. M. Waters. 2008. Geological dates and molecular rates: fish DNA sheds light on time dependency. Molecular Biology and Evolution **25**:624-633.

Chambers, G. K., W. M. Boon, T. R. Buckley, and R. A. Hichmough. 2001. Using molecular methods to understand the Gondwanan affinities of the New Zealand biota: three case studies. Australian Journal of Botany **49**:377-387.

Chriswell, S. M., J. Wilkin, J. D. Booth, and B. Santon. 2003. Trans-Tasman Sea larval transport: Is Australia a source for New Zealand rock lobsters? Marine Ecology Progress Series **214**:173-182.

Conroy, C. J., and M. van Tuinen. 2003. Extracting time from phylogenies: positive interplay between fossil and genetic data. Journal of Mammalogy **84**:444-455.

Conway Morris, S. 1998. Early metazoan evolution: Reconciling paleontology and molecular biology. American Zoologist **38**:867-877.

Cook, L. G., and M. D. Crisp. 2005a. Not so ancient: the extant crown group of *Nothofagus* represents a post-Gondwanan radiation. Proceedings of the Royal Society of London Series B-Biological Sciences **272**:2535-2544.

Cook, L. G., and M. D. Crisp. 2005b. Directional asymmetry of long-distance dispersal and colonization could mislead reconstructions of biogeography. Journal of Biogeography **32**:741-754.

Cooper, A., and R. A. Cooper. 1995. The Oligocene bottleneck and New Zealand biota: Genetic record of a past environmental crisis. Proceedings of the Royal Society of London Series B-Biological Sciences **261**:293-302.

Cracraft, J. 2001. Avian evolution, Gondwana biogeography and the Cretaceous-Tertiary mass extinction event. The Royal Society of London, B **268**:459-469.

Cracraft, J. 1981. Toward a phylogenetic classification of the recent birds of the world (Class Aves). Auk **98**:681-714.

Craw, R. C. 1979. Generalised tracks and dispersal in biogeography: a response to R. M. McDowall. Systematic Zoology **28**:99–107.

Cree, A., C. H. Daugherty, and J. M. Hay. 1995. Reproduction of a Rare New Zealand Reptile, the Tuatara *Sphenodon punctatus*, on Rat-Free and Rat-Inhabited Islands. Conservation Biology **9**:373-383.

Croizat, L., G. J. Nelson, and D. E. Rosen. 1974. Centers of origin and related concepts. Systematic Zoology **23**:265-287.

Darwin, C. 1859. The Origin of species by means of natural selection or the preservation of favoured races in the struggle for life. The Modern Library, New York.

Daugherty, C. H., G. W. Gibbs, and R. A. Hitchmough. 1993. Mega-island or micro-continent? New Zealand and its fauna. Trends in Ecology & Evolution **8**:437-441.
de Queiroz, A. 2005. The resurrection of oceanic dispersal in historical biogeography. Trends in Ecology & Evolution **20**:68-73.
Dietz, R. S. 1961. Continent and ocean basin evolution by spreading of the sea floor. Nature **190**:854-857.
Douady, C. J., F. Catzeflis, J. Raman, M. S. Springer, and M. Stanhope. 2003. The Sahara as a vicariant agent, and the role of Miocene climatic events, in the diversification of the mammalian order Macroscelidea (elephant shrews). Proceeding of the National Academy of Science, USA **100**:8325-8330.
Doyle, J. A., and M. J. Donoghue. 1993. phylogenies and angiosperm diversification. Paleobiology **19**:141-167.
Ebach, M. C., and C. J. Humphries. 2002. Cladistic biogeography and the art of discovery. Journal of Biogeography **29**:427-444.
Fell, H. B. 1967. Cretaceous and Tertiary surface currents of the oceans. Oceanography and Marine Biology Annual Reviews **5**:317–341.
Fell, H. B. 1962. West-wind drift dispersal of echinoderms in the southern hemisphere. Nature **193**:759–761.
Fleming, C. A. 1979. The geological history of New Zealand and its life. Oxford University Press, Auckland.
Fordyce, R. E. 1991. A new look at the fossil vertebrate record of New Zealand. Pp. 1437 *in* P. Vickers-Rich, J. M. Monaghan, R. F. Baird, and T. H. Rich, eds. Vertebrate Palaeontology of Australasia. Monash University Publications Committee, Melbourne.
Fox, K. J. 1973. Trans-oceanic dispersal of insects to New Zealand. New Zealand Entomologist **5**:240-243.
Gaskin, D. E. 1970. The origins of the New Zealand fauna and flora: A review. Geographical Review **60**:414-434.
Gibbs, G. 2006. Ghosts of Gondwana: The History of Life in New Zealand. Craig Potton Publishing, Nelson, New Zealand.
Givnish, T. J., and S. S. Renner. 2004. Tropical intercontinental disjunctions: Gondwana breakup, immigration from the boreotropics, and transoceanic dispersal. Int. J. Plant Sci. **165**:S1-S6.
Harshmann, J., E. L. Braun, M. J. Braun, C. J. Huddleston, R. C. K. Bowie, J. L. Chojnowski, S. J. Hackett, K.-L. Han, R. T. Kimball, B. D. Marks, K. J. Miglia, W. S. Moore, S. Reddy, F. H. Sheldon, D. W. Steadman, S. J. Steppan, C. C. Witt, and T. Yuri. 2008. Phylogenomic evidence for multiple losses of flight in ratite birds. PNAS **105**:13462-13467.
Heads, M. 2005. Dating nodes on molecular phylogenies: a critique of molecular biogeography. Cladistics **21**:62-78.
Hoffmann, A. A., and P. A. Parsons. 1997. Extreme environmental change and evolution. Cambridge University Press.
Hornibrook, N. d. B. 1978. Paleontology. Pp. 407-425 *in* R. P. Suggate, ed. The geology of New Zealand. E. C. Keating Government Printer, Wellington.
Hoy, M. A. 1994. Insect molecular systematics and evolution. Pp. 337-387. Insect molecular genetics - An introduction to principles and applications. Academic Press, Inc., San Diego.
Hurr, K. A., P. J. Lockhart, P. B. Heenan, and D. Penny. 1999. Evidence for the recent dispersal of *Sophora* (Leguminosae) around the wouthern oceans: Molecular data. Journal of Biogeography **26**:565-577.

International commission on Stratigraphy (ICS). 2004. accessed Sept. 2008. http://www.stratigraphy.org/cheu.pdf

Johnson, K. P., F. R. Adler, and J. L. Cherry. 2000. Genetic and phylogenetic consequences of island biogeography. Evolution **54**:387-396.

Jordan, G. J. 2001. An investigation of long-distance dispersal based on species native to both Tasmania and New Zealand. Australian Journal of Botany **49**:333-340.

Knapp, M., R. Mudaliar, D. Havell, S. J. Wagstaff, and P. J. Lockhart. 2007. The drowning of New Zealand and the problem of Agathis. Systematic Biology **56**:862-870.

Knapp, M., K. Stöckler, D. Havell, F. Delsuc, F. Sebastiani, and P. J. Lockhart. 2005. Relaxed molecular clock provides evidence for long distance dispersal of *Nothofagus* (Southern Beech). PLoS Biology **3**:e14.

Landis, C. A., H. J. Campbell, J. G. Begg, D. C. Mildenhall, A. M. Paterson, and S. A. Trewick. 2008. The Waipounamu erosion surface:questioning the antiquity of the New Zealand land surface and terrestrial fauna and flora. Geological Magazine:1-25.

Lee, D. E., W. G. Lee, and N. Mortimer. 2001. Where and why have all the flowers gone? Depletion and turnover in the New Zealand Cenozic angiosperm flora in relation to palaeogeography and climate. Australian Journal of Botany **49**:341-356.

Lee, M. S. Y. 1998. Shortening the phylogenetic fuse. Trends in Ecology & Evolution **13**:323.

Lee, M. S. Y. 1999. Molecular clock calibrations and Metazoan divergence dates. Journal of Molecular Evolution **49**:385-391.

Lomolino, M. V. 2000. A call for a new paradigm of island biogeography. Global Ecology and Biogeography **9**:1-6.

MacArthur, R. H., and A. C. Wilson. 2001. The theory of island biogeography. Princeton University Press, Princeton, NJ.

MacArthur, R. H., and A. C. Wilson. 1963. An equilibrium theory of insular zoogeography. Evolution **17**.

Macdonald, J. 2006. In the beginning - geological history and origins of New Zealand biota. Pp. 2-29 *in* S. Parsons, ed. Biology Aotearoa Unique flora, fauna and fungi. Pearson Education New Zealand, Auckland.

Magallón, S. A. 2004. Dating lineages: molecular and paleontological approaches to the temporal framework of clades. Int. J. Plant Sci. **165**:S7-S21.

Marshall, C. R. 1990. Confidence intervals on stratigraphic ranges. Paleobiology **16**:1-10.

Martins-Neto. 2007. New Orthoptera Stenopelmatoidea and Hagloidea (Ensifera) from the Santana Formation (Lower Cretaceous, Northeast Brazil) with description of new taxa. . GAEA **3**:3-8.

McDowall, R. M. 1978. Generalized tracks and dispersal in biogeography. Systematic Zoology **27**:88–104.

McGlone, M. S. 2005. Goodbye Gondwana. Journal of Biogeography **32**:739-740.

McLaughlin, P. J., and M. O. Dayhoff. 1972. Evolution of species and proteins: a time scale. National Biomedical Research Foundation, Washington.

McLoughlin, S. 2001. The breakup history of Gondwana and its impact on pre-Cenozoic floristic provincialism. Australian Journal of Botany **49**:271-300.

Meads, M. J. 1990. The Weta Book: A guide to the identification of wetas. DSIR Land Resources, Lower Hutt.

Michaux, B. 1991. Distributional patterns and tectonic development in Indonesia: Wallace reinterpreted. Australian Systematic Botany **4**:25-36.

Molnar, R. E., and M. Pole. 1997. A Miocene crocodilian from New Zealand. Alcheringa **21**:65-70.

Muñoz, J., A. M. Felicisimo, F. Cabezas, A. R. Burgaz, and I. Martinez. 2004. Wind as a long-distance dispersal vehicle in the Southern Hemisphere. Science **304**.

Near, T. J., and M. J. Sanderson. 2004. Assessing the quality of molecular divergence time estimates by fossil calibrations and fossil-based model selection. Philosophical Transactions: Royal Society of London, Series B **359**:1477-1483.

Nelson, G., and P. Y. Ladiges. 2001. Gondwana, vicariance biogeography and the New York School revisited. Australian Journal of Botany **49**:389–409.

Orth, C. J., J. S. Gilmore, J. D. Knight, C. L. Pillmore, T. H. Tschudy, and J. E. Fassett. 1981. An iridium abundance anomaly at the palynological Cretaceous-Tertiary boundary in Northern New Mexico. Science **214**:1341-1343.

Pagel, M., C. Venditti, and A. Meade. 2006. Large punctuational contribution of speciation to evolutionary divergence at the molecular level. Science **314**:119-121.

Phillips, M. J. 2008. Branch-length estimation bias misleads molecular dating for a vertebrate mitochonrial phylogeny. Gene.

Pole, M. 1994. The New Zealand flora - entirely long-distance dispersal? Journal of Biogeography **21**:625-635.

Pond, C. M. 1977. The significance of lactation in the evolution of mammals. Evolution **31**:177-199.

Pratt, R. C., M. Morgan-Richards, and S. A. Trewick. 2008. Diversification of New Zealand weta (Orthoptera: Ensifera: Anostostomatidae) and their relationships in Australasia. Philosophical Transactions of the Royal Society of London B Biological Sciences.

Riddle, B. R. 2005. Is biogeography emerging from its identity crisis? Journal of Biogeography **32**:185-186.

Rosen, D. E. 1978. Vicariant patterns and historical explanation in biogeography. Systematic Biology **27**:159-188.

Russell, D. A. 1979. The enigma of the extinction of the dinosaurs. Annual Reviews in Earth and Planetary Sciences **7**:163-182.

Sanmartin, I., and F. Ronquist. 2004. Southern Hemisphere biogeography inferred by event-based models: plant versus animal patterns. Systematic Biology **53**:216-243.

Şengör, A. M. C., S. Atayman, and S. Özeren. 2008. A scale of greatness and causal classification of mass extinctions: Implications for mechanisms. PNAS **105**:13736-13740.

Shepherd, L. D., and D. M. Lambert. 2007. The relationships and origins of the New Zealand wattlebirds (Passerifromes, Callaeatidae) from DNA sequence analyses. Molecular Phylogenetics and Evolution **43**:480-492.

Shields, R. 2004. Pushing the envelope on molecular dating. Trends in Genetics **20**:221-222.

Smith, A. B., and K. J. Peterson. 2002. Dating the time of origin of major clades: Molecular clocks and the fossil record. Annual Review of Earth and Planetary Sciences **30**:65-88.

Soligo, C., O. A. Will, S. Tavaré, C. R. Marshall, and R. D. Martin. 2007. New light on the dates of primate origins and divergence. Pp. 29-49 *in* M. J. Ravosa, and M. Dagosto, eds. Primate origins: adaptations and evolution. Springer US.

Stevens, G. R. 1980. New Zealand Adrift: The theory of continental drift in a New Zealand setting. A.H. & A. W. Reed Ltd., Wellington.

Stevens, J. R., and C. J. Schofield. 2003. Phylogenetics and sequence analysis - some problems for the unwary. Trends in Parasitology **19**:582-588.

Stöckler, K. 2001. Origins and evolution of the New Zealand forest flora. Pp. 234. Plant Biology, Institute of Molecular BioSciences. Massey University, Palmerston North.

Stöckler, K., I. L. Daniel, and P. J. Lockhart. 2002. New Zealand Kauri (*Agathis australis* (D.Don) Lindl., Araucariaceae) survives Oligocene drowning. Systematic Biology **51**:827-832.

Towns, D. R., I. A. E. Atkinson, and C. H. Daugherty. 2006. Have the harmful effects of introduced rats on islands been exaggerated? Biological Invasions **8**:863-891.

Trewick, S. A. 2000. Molecular evidence for dispersal rather than vicariance as the origin of flightless insect species on the Chatham Islands, New Zealand. Journal of Biogrography. **27**:1189 - 1200.

Trewick, S. A., and G. P. Wallis. 2001. Bridging the "beech-gap": New Zealand invertebrate phylogeography implicates pleistocene glaciation and pliocene isolation. Evolution **55**:2170-2180.

Vajda, V., and S. McLoughlin. 2007. Extinction and recovery patterns of the vegetation across the Cretaceous-Palaeogene boundary - a tool for unravelling the causes of the end-Permian mass-extinction. Review of Palaeobotany & Palynology **144**:99-112.

Vajda, V., and S. McLoughlin. 2004. Fungal proliferation at the Cretaceous-Tertiary boundary. Science **303**:1489.

Vajda, V., J. I. Raine, and C. J. Hollis. 2001. Indication of global deforestation at the Cretaceous-Tertiary boundary by New Zealand fern spike. Science **294**:1700-1702.

Valentine, J. W., D. Jablonski, and D. H. Erwin. 1999. Fossils, molecules and embryos: new perspectives on the Cambrian explosion. Development **126**:851-859.

Wagstaff, S. J., I. Breitwieser, C. Quinn, and M. Ito. 2007. Age and origin of enigmatic megaherbs from the subantarctic islands. Nature Precedings <http://hdl.handle.net/10101/npre.2007.1272.1>.

Waters, J. M. 2008. Driven by the West Wind Drift? A synthesis of southern temperate marine biogeography, with new directions for dispersalism. Journal of Biogeography **35**:417-427.

Waters, J. M., D. Craw, J. H. Youngson, and G. P. Wallis. 2001. Genes meet geology: fish phylogeographic pattern reflects ancient, rather than modern, drainage connections. Evolution **55**:1844–1851.

Waters, J. M., L. H. Dijkstra, and G. P. Wallis. 2000. Biogeography of a southern hemisphere freshwater fish: how important is marine dispersal? Molecular Ecology **9**:1815-1821.

Waters, J. M., and M. S. Roy. 2004. Out of Africa: the slow train to Australasia. Systematic Biology **53**:18-24.

Waters, J. M., and R. W. G. White. 1997. Molecular phylogeny and biogeography of the Tasmanian and New Zealand Mudfishes (Salmoniformes: Galaxiidae). Australian Journal of zoology **45**:39-48.

Watt, J. C. 1975. The terrestrial insects. Pp. 689 *in* G. Kuschel, ed. Biogeography and Ecology in New Zealand. Dr. W. Junk b.v. Publishers the Hague, Auckland.

Welch, J. J., and L. Bromham. 2005. Molecular dating when rates vary. Trends in Ecology and Evolution **20**:320-327.

Winkworth, R. C., S. J. Wagstaff, D. Glenny, and P. J. Lockhart. 2002. Plant dispersal N.E.W.S from New Zealand. Trends in Ecology & Evolution **17**:514-520.

Wolfe, J. A., and G. R. J. Upchurch. 1987. Leaf assemblages across the Cretaceous-Tertiary boundary in the Raton Basin, New Mexico and Colorado. Proceeding of the National Academy of Science, USA **84**:5096-5100.

Worthy, T. H., A. J. D. Tennyson, M. Archer, A. M. Musser, S. J. Hand, C. Jones, B. J. Douglas, J. A. McNamara, and R. M. D. Beck. 2006. Miocene mammal reveals a

Mesozoic ghost lineage on insular New Zeland, southwest Pacific. PNAS **103**:19419-19423.

Worthy, T. H., A. J. D. Tennyson, C. Jones, J. A. McNamara, and B. J. Douglas. 2007. Miocene waterfowl and other birds from central Otago, New Zealand. Journal of Systematic Palaeontology **5**:1-39.

Wray, G. A. 2001. Dating branches on the tree of life using DNA. Genome Biology **3**:1-7.

Wright, S. D., C. G. Yong, J. W. Dawson, D. J. Whittaker, and R. C. Gardner. 2000. Riding the ice age El Niño? Pacific biogeography and evolution of *Metrosideros* subg. *Metrosideros* (Myrtaceae) inferred from nuclear ribosomal DNA. Proceeding of the National Academy of Science, USA **97**:4118-4123.

Zuckerkandl, E., and L. Pauling. 1965. Evolutionary divergence and convergence in proteins. Pp. 97-166 *in* V. Bryson, and H. J. Vogel, eds. Evolving Genes and Proteins. Academic Press, New York.

CHAPTER TWO

***B**ut if rivers come into being and perish and if the same parts of the earth are not always moist, the sea also must necessarily change correspondingly. And if in places the sea recedes while in others it encroaches, then evidently the same parts of the earth as a while are not always sea, nor always mainland, but in process of time all change.*

-Aristotle, Meteorologica,
ca. 335 B.C.

This chapter is published as;

Pratt, R. C.**, Morgan-Richards, M. and Trewick, S. A. 2008. Diversification of New Zealand weta (Orthoptera: Ensifera: Anostostomatidae) and their relationships in Australasia, Phil. Trans. R. Soc. B,* ***363 *(1508):3427-3437 doi:10.1098/rstb.2008.0112. See Appendix D for reprint*

Diversification of New Zealand weta (Orthoptera: Ensifera: Anostostomatidae) and their relationships in Australasia

2.1 ABSTRACT

New Zealand taxa from the Orthopteran family Anostostomatidae have been shown to consist of three broad groups; *Hemiandrus* (ground weta), *Anisoura/Motuweta* (tusked weta) and *Hemideina–Deinacrida* (tree–giant weta). The family is also present in Australia and New Caledonia, the nearest large land masses to New Zealand. All genera are endemic to their respective countries except *Hemiandrus* which occurs in New Zealand and Australia. We used nuclear and mitochondrial DNA sequence data to study within genera and among species level genetic diversity within New Zealand and to examine phylogenetic relationships of taxa in Australasia. We found the Anostostomatidae to be monophyletic within Ensifera, and justifiably distinguished from the Stenopelmatidae among which they were formerly placed. However, the New Zealand Anostostomatide are not monophyletic with respect to Australian and New Caledonian species in our analyses. Two of the New Zealand groups have closer allies in Australia and one in New Caledonia. We carried out Maximum Likelihood and Bayesian analyses to reveal several well supported subgroupings. Our analysis included the most extensive sampling to date of *Hemiandrus* species and indicate that Australian and New Zealand *Hemiandrus* are not monophyletic. We used molecular dating approaches to test the plausibility of alternative biogeographic hypotheses for the origination of the New Zealand anostostomatid fauna and found support for divergence of the main clades at, or shortly after Gondwanan breakup and dispersal across the Tasman much more recently.

2.2 INTRODUCTION

The biology of New Zealand is, unlike that of most Pacific islands, viewed as continental in nature (Cowie and Holland 2006). This is justified geologically because New Zealand is formed from continental rather than oceanic crust (Neall and Trewick 2008). Consequently the biota of New Zealand is considered to be predominantly “Gondwanan”, having its principle affinities in Australasia (Fleming 1979), and the southern hemisphere in general (Gibbs 2006).

Although the Gondwanan nature of the New Zealand biota is often attributed to the continental (vicariant) history of the land, this explanation has not always been pre-eminent (eg. Fleming 1962; Caughley 1964). The importance of dispersal is widely recognised, and it is more generally accepted that biogeographical pattern alone does not reveal the process(es) of origination of biota (Waters and Craw 2006). Indeed, the distribution of one former icon of vicariance biogeography in New Zealand (and the Southern Hemisphere), southern beech (*Nothofagus*) has recently been shown to be best explained by dispersal to New Zealand (Cook and Crisp 2005a; Knapp et al. 2005). Numerous other molecular studies of a range of taxa, demonstrate that a substantial (if not predominant) part of the New Zealand biota are products of long distance dispersal (Waters and Craw 2006 and references therein; Trewick et al. 2007).

An improving, though far from perfect understanding of the tectonic history in the New Zealand region (Mortimer 2004) is helping reveal why this continental land does not in fact have a predominantly continental biota. The rifting of the continent, Zealandia from Gondwana (including Australia) commenced *ca* 82 Ma (figure 2.1*a*) and the Tasman Sea reached its current width between 63.5 - 55.5 Ma (Veevers and Li 1991; McLoughlin 2001). This was just the start of New Zealand and New Caledonia's story. Zealandia (figure 2.1*b*) subsequently sank so that today, about 93% of the continent is below the surface of the sea (Landis et al. 2008).

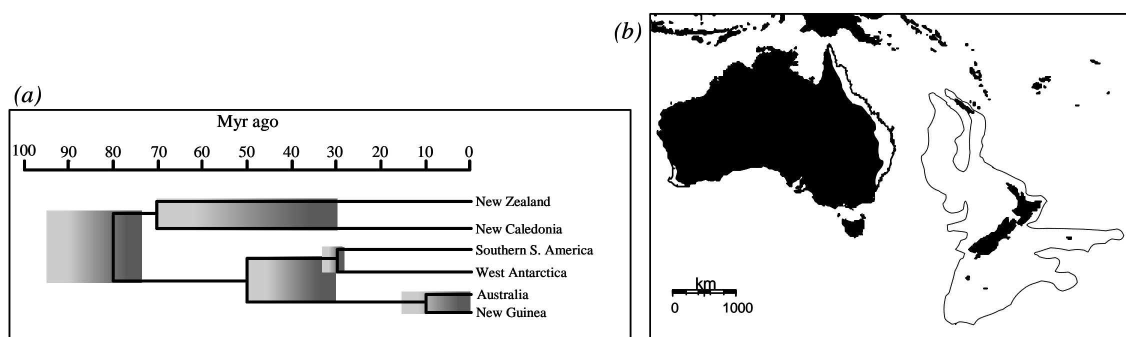


Figure 2.1 Timing of Gondwanan breakup and Australasian distribution map

(a) Geological area cladogram, after Cook & Crisp (2005b) showing the accepted Gondwanan breakup sequence. Shaded boxes represent uncertainty about timing of vicariant events. (b) Australasia sampled for anostostomatid weta (Insecta: Orthoptera) in this study (Australia, New Caledonia and New Zealand). Australian species are limited to the east coast of Queensland and New South Wales and one genus in Western Australia (white fill) whereas most parts of New Zealand and New Caledonian support one or more weta species. The dashed line indicates the approximate boundaries of the submerged continent Zealandia.

This active geological history contrasts with the prolonged geological stability of Australia, the nearest remnant of continental Gondwana. With respect to land area, climate and biotic assemblage, New Zealand (and New Caledonia) has few continental attributes and ample evidence suggests these islands are biologically more like oceanic islands than southern continents (Goldberg et al. 2008; Grandcolas et al. 2008; Trewick and Morgan-Richards in press).

Evidence for the persistence of land in the New Zealand region throughout the Oligocene has been obscured by the extensive tectonic activity initiated in the early Miocene (Landis et al. 2008). The tectonic upheaval that resulted in the formation of New Zealand (as we know it today) began *ca* 24 Ma and still continues (Trewick et al. 2007). For example, the major mountain ranges of New Zealand started forming *ca* 5 Ma. This, and other local geophysical events, may have been more important in the development of the modern biota than ancient vicariant processes. New Caledonia has a similar geological history with tectonic activity forcing a submerged section of Zealandia (and obducted oceanic ultramafic strata) to the sea surface in the late Eocene (*ca* 40 Ma) (Chardon and Chevillotte 2006; Mortimer et al. 2006; Grandcolas et al. 2008; Neall and Trewick 2008)

One of the most interesting components of New Zealand's terrestrial fauna, with both taxonomic and ecological diversity, are insects of the orthopteran family Anostostomatidae, known in New Zealand by their Maori name, weta. Of particular biogeographic interest is the presence of the family on all three major Australasian landmasses; Australia, New Caledonia and New Zealand. The group consists of relatively large insects (20 – 80 mm) that are nocturnal, predominantly flightless and predatory, with a Gondwanan distribution (also found in Central and South America, South Africa, Madagascar and India). In New Zealand, the family is represented by five genera and approximately 56 species. These five genera fall into three distinct groups: (1) nine (plus approx. 30 undescribed) species of *Hemiandrus* Ander 1938 (ground weta); (2) one species of *Anisoura* Ander 1938 and two species of *Motuweta* Johns 1997 (tusked weta) and (3) seven *Hemideina* White 1846 (tree weta) and eleven *Deinacrida* White 1842 (giant weta) (Trewick and Morgan-Richards 2004; Trewick and Morgan-Richards 2005).

The *Hemideina* and *Deinacrida* are unusual among Anostostomatidae in that all species are primarily herbivorous. The diversification of *Hemideina–Deinacrida* dates to the Miocene, with adaptation to diverse habitats following mountain uplift (ca 5 Myr ago Trewick and Morgan-Richards 2005). The three tusked weta species (*Anisoura/Motuweta*), so named because of the impressive tusk-like structures on the mandibles of mature males, form a monophyletic group among the New Zealand taxa (Trewick and Morgan-Richards 2004), although analogous ornamentations are found in some South African species (i.e. *Libanasidus vittatus*, Field and Deans 2001). Within the Australasian anostostomatid genera, *Hemiandrus* is the only genus not endemic to a single landmass, being recorded in both Australia and New Zealand (Johns 1997). Of the *ca* 40 species from New Zealand, (personal communication P. M. Johns) only nine are described (Johns 1997; Jewell 2007), making them the least well characterised weta group in this country. Ovipositor length, which appears to be correlated with degree of maternal care (Gwynne 1995; Johns 1997; Gwynne 2004), was in the past the key morphological character distinguishing the genus *Zealandosandrus* Salmon 1950 from *Hemiandrus* Ander 1838. However, Johns (1997) synonomised *Zealandosandrus*, retaining *Hemiandrus* by precedent. The Australian anostostomatid fauna is poorly characterised with just 13 described species but probably comprises nine genera with approximately 60 species (Johns 1997; personal communication P. M. Johns, G. Monteith). Australia's fauna includes three genera with winged species, one of which (*Transaevum*) is considered to be the most "ancestral" extant member of the group (Monteith and Field 2001). Intriguingly, fossil Orthoptera that are putatively "weta" are reported from 190 Ma old deposits in Queensland, Australia (Meads 1990) but have not been formally described. The New Caledonian weta fauna consists of two genera, *Aistus* (three species with several undescribed) and *Carcinopsis* (six species) (Johns 1997).

Anostostomatids occupy a variety of environments across their geographic range in Australasia. In New Caledonia they are tropical forest inhabitants; in Australia, almost all are found in the wet tropical forests of Queensland with one genus endemic to coastal Western Australia (Monteith and Field 2001) (figure 2.1*b*). New Zealand weta live mostly in temperate forest and sub-alpine environments, with the majority of species in South Island (figure 2.2). Curiously, anostostomatids are absent from comparable temperate forest in southern Australia (Victoria and Tasmania).

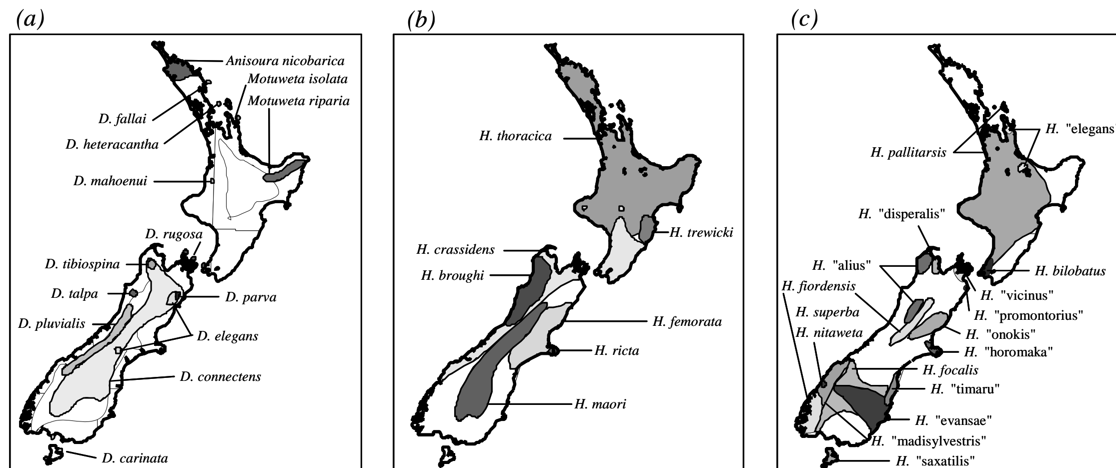


Figure 2.2 New Zealand taxa distribution map

Approximate distributions of New Zealand Anostostomatidae taxa; (a) *Deinacrida* (giant weta) and *Anisoura/Motuweta* (tusked weta) species; (b) *Hemideina* (tree weta) (based on Trewick & Morgan-Richards in press); (c) *Hemiandrus* (ground weta) based on Johns (2001), Jewell (2007), and P.M. Johns (2005, personal communication). In addition to the species shown, *H. maculifrons* is widely distributed throughout both the North and South Islands. Additional undescribed taxa appear to have local distributions and predominate in South Island (personal communication P. M. Johns 2005).

We have undertaken sampling across New Zealand, New Caledonia and Australia to explore the evolution of the New Zealand Anostostomatidae. Representatives of all nine Australian genera, both New Caledonian genera and the five New Zealand genera were included in the present study. We used molecular phylogenetics to recover support for the relationships of the three New Zealand anostostomatid groups. We included representatives of New Zealand *Hemiandrus* species diversity, which have been absent from previous work, in order to explore diversity within this group and test support for monophyly of the genus. We applied relaxed molecular clock methods, to estimate the likely age and origin of New Zealand weta lineages.

2.3 Materials and Methods

(a) Sampling

The majority of sampling was undertaken by the authors in New Zealand and New Caledonia. In addition, samples of *Hemiandrus* were supplied by Darryl Gwynne (University of Toronto, Canada).

New Zealand sampling included all three tusked weta species (*Anisoura/Motuweta*), representatives of the *Hemideina* and *Deinacrida* (previously shown to be monophyletic; Trewick & Morgan-Richards 2005) and representatives within the taxonomic diversity of *Hemiandrus* including putative and new species (Appendix 2.8 table 2-1). Assistance with identification and sampling of undescribed species was provided by P. M. Johns. Material from Australia was supplied by Geoff Monteith (Brisbane Museum, Australia) and Dave Rentz (CSIRO, Australia). This sampling, although not exhaustive, includes at least one representative of each genus in the region and is the most complete dataset to date.

(b) Molecular Methods

Whole genomic DNA was extracted from hind leg muscle following the salting-out method (Sunnucks and Hale 1996) and re-suspended in 50µL TE buffer (0.1mM EDTA, 10mM Tris) or water. PCR reactions were performed in 10µL volumes using ABgene Red Hot *Taq*. Products were visualised on 1% agarose gels stained with SYBRSafe (Invitrogen). Thermal cycling PCR was carried out on an MJ Research PTC-200 thermal cycler and consisted of initial denaturation of 94°C for 2min, followed by 35 cycles of 94°C for 30sec, 48 - 50°C for 30sec and 72°C for 1min 30sec with a final extension of 72°C for 3mins. PCR products were purified with Shrimp Alkaline Phosphatase (SAP) and Exonuclease I (EXOI) following manufacturer's recommendations (USB Corporation). Sequencing used BigDye Terminator v3.1 chemistry and an ABI3730XL Genetic Analyser (Applied Biosystems, Foster City, CA). DNA sequences were deposited on NCBI GenBank (Appendix 2.8 table 2-1; EU676657-EU676800, EU713453-EU13461). Primers used for PCR were; mtDNA COI: LCO1490, HCO2198 (Folmer et al. 1994), CI-J-2195 and L2-N-3014 and CI-J-2195 and L2-N-3014 (Simon et al. 1994); mtDNA third domain 12S rRNA: SR-N-14588 and LR-J-13417 (Simon et al. 1994); nuclear rRNA 18S: 18S-S22, 18S-A1984 (Vawter 1991), 18S_1F (gac gaa aaa taa cga tac ggg), 18S_1R (ctc aat ctg tca atc ctt cca) (this study); and 28S: 28SrD1.2a, 28SrD3.2a, 28SrD4.8a, 28SA, 28SrD4.2b, 28SrD5b, 28SrD7bl, 28SB (Whiting 2002b).

(c) Phylogenetic Analysis

Individual sequence reads were checked against ABI trace files using SEQUENCHER V.4.70 (Gene Codes Corp. Ann Arbour, MI) and aligned using SE-AL v2.0a11 (Rambaut 1996). The protein coding gene COI was translated into amino acids to ensure correct reading frame and to detect evidence of nuclear copies (which when found were subsequently removed). Ribosomal RNA genes were checked for indels. In cases where missing data were included they were coded as N in analyses. In order to evaluate individual genes and concatenated data, we divided the datasets into the following; I – 18S Ensifera; II – 18S Australasian Anostostomatidae; III –combined 18S and 28S Australasian Anostostomatidae; IV – COI-RY coded Australasian Anostostomatidae; V – combined COI-RY and 12S Australasian Anostostomatidae; and VI – COI *Hemiandrus* only.

The COI data were partitioned into three character sets according to codon position, first, second and third. In order to maximise third codon information we treated it in three different ways; as four nucleotides (A, G, T, C), Y coded (Y, A, G) or RY-coded (A and G = R, T and C = Y). In order to avoid potential tree estimation bias due to nucleotide composition or saturation, we used Y- or RY-coding on the third codon position nucleotides for COI sequences in datasets IV and V. Re-coding of this sort has been shown to greatly improve consistency in phylogenetic resolution by reducing bias from differences in nucleotide composition (Phillips and Penny 2003) which is useful when looking at deeper divergences. To assist with tree-rooting and thus confirm ingroup status of our sample we used published Ensifera DNA sequences from both EMBL and NCBI GenBank (Appendix 2.8. table 2-1).

Models of DNA evolution were optimized separately for each data set using MODELTEST 3.7 (Posada and Crandall 1998) and Akaike Information Criterion (AIC) was preferred to the Hierarchical likelihood ratio test (hLRTs) (Posada and Buckley 2004). Maximum likelihood (ML) analyses were implemented using the programs PAUP* (Swofford 2003), GARLI version 0.951 (Zwickl 2006) and PHYML (Guindon and Gascuel 2003).

Model parameters from MODELTEST were implemented using a general time-reversible model with invariable sites and a gamma distribution for variable rate sites (GTR + I + G) model with a Heuristic search under the likelihood criterion with trees obtained from stepwise addition.

Bayesian analyses were implemented using MRBAYES 3.1 (Huelsenbeck and Ronquist 2001). We specified nst = 2 (HKY) and nst = 6 (GTR) with a proportion of invariant sites and gamma distribution of rate variation. Analyses of datasets III (18S+28S), IV (COI) and V (COI+12S) data were undertaken with (parameters unlinked) and without character-set partitions. We used two runs of four Markov chains (each with one cold chain) with 1-10 $\times$ 10^6 generations and default priors, sampling every thousandth tree. A "burn-in" of 10% was removed after examination of log likelihood scores and average standard deviation of the split frequencies. Trees saved below the "burn-in" generation were discarded and a majority rule consensus of the remaining trees was calculated. Multiple replicates of the Bayesian runs were carried out to insure convergence of the posteriors.

(d) Tree comparisons

We assessed the degree of conflict between our phylogenetic estimates by using tree comparison tests, to see if one topology was significantly better at explaining the molecular data than alternative phylogenies. We used the SH-tests (Shimodaira and Hasegawa 1999) implementing a RELL distribution derived from 1,000 bootstrap replicates as executed in PAUP*. For dataset IV (COI), we carried out multiple analyses manipulating the 3rd codon position so that it was; four states, Y-coded and RY-coded. To observe the effect of this simple noise reduction technique, we compared ML topologies obtained from PHYML for each state using either a simple model (HKY85) or a parameter rich model (GTR + I + G). We also used constraint analysis to test the likelihood of alternative tree topologies for monophyly of New Zealand taxa and for monophyly of the genus *Hemiandrus* (New Zealand and Australia).

(e) Divergence-time estimation

We compared the likelihood scores obtained from ML analysis both with and without the implementation of a molecular clock in PAUP* for dataset II (18S Australasia) and dataset IV (COI-RY coded Australasia) datasets. This was carried out both with and without the inclusion of taxa we suspected of having a rate shift due to long branches observed in initial analyses.

SH-tests were applied to resulting trees to determine if there was rate heterogeneity and therefore if the data were acting in a clock-like manner to determine whether to use a strict or relaxed molecular clock in BEAST version 1.4.6 (Drummond and Rambaut 2007).

As there are no suitable fossils for molecular dating, we used geological events as points of reference to test the plausibility of vicariant versus dispersal explanations for New Zealand weta diversification (figure 2.1*a*). In order to explicitly examine the alternative hypotheses for patterns of diversity, we calibrated trees using initial separation of Zealandia from Gondwana (less than 82 Ma as applied by avian evolutionists, see Ericson et al. 2002; Baker et al. 2005) and emergence of New Caledonian (less than 40 Ma). The two dating constraints were separately applied to the nuclear dataset II (18S Australasia) and mitochondrial dataset IV (COI-RY coded Australasia). We removed a clade of five taxa (clade A plus New Caledonian taxa; figure 4 (18S), figure 5 (COI)), shown by initial analyses to have long branches and a substantially elevated rate of molecular evolution (indicated by BEAST rates). First, if Zealandia and Australia parted *ca* 82 Ma, we assumed vicariance and constrained the most basal split of Anostostomatidae to more than 82 Ma (BEAST parameters; relaxed lognormal clock, lognormal distribution, mean =4.74; s.d = 0.2, run for 100 million repetitions sampling trees every 10,000). Second, we constrained the same point on the tree as above, but assuming Zealandia submerged completely and land resurfaced *ca* 40 Ma (BEAST parameters; normal distribution, mean = 36, s.d = 0.2, initial value = 35). We used a root value for the trees (BEAST parameters; uniform distribution, lower limit = 85, upper limit = 250, initial = 100) as unpublished fossils have been dated from 190 – 200 Ma from Queensland that have been attributed to the Anostostomatidae and fossil Ensifera are dated to *ca* 250 Ma (Fig. 7.18 p202, Grimaldi and Engel 2005).

Calibration points were implemented in BEAST using a relaxed uncorrelated lognormal molecular clock (Drummond et al. 2006). Resulting trees were analysed using software distributed with BEAST - TREEANNOTATOR v1.4.6 where the burn-in (1000 trees) was removed and a maximum credibility tree constructed. Trees were then viewed in FIGTREE v1.1.2. Details of the XML files are available on request from the authors.

2.4 RESULTS

A summary of all sequence data collected is given in table 2-1 along with locality information and accession numbers. NCBI BLAST searches returned matches for previously published Orthopteran sequences. The alignment used in each analysis is available as a NEXUS file from the authors on request.

(a) Phylogeny

Dataset I (18S Ensifera) consisted of 1,863 bp after removal of a 37 bp hypervariable indel region between bp 719 - 756 from our original alignment to accommodate the diverse range of taxa, which included seven ingroup taxa and nine outgroup Ensifera (Flook et al. 1999; Terry and Whiting 2005; Jost and Shaw 2006). We confirm the monophyly of Anostostomatidae in our sample and found the Gryllacrididae to be sister to Anostostomatidae with Stenopelmatidae sister to the Anostostomatidae–Gryllacrididae clade. Both of these families have previously been suggested as close relatives to Anostostomatidae (figure 2.3).

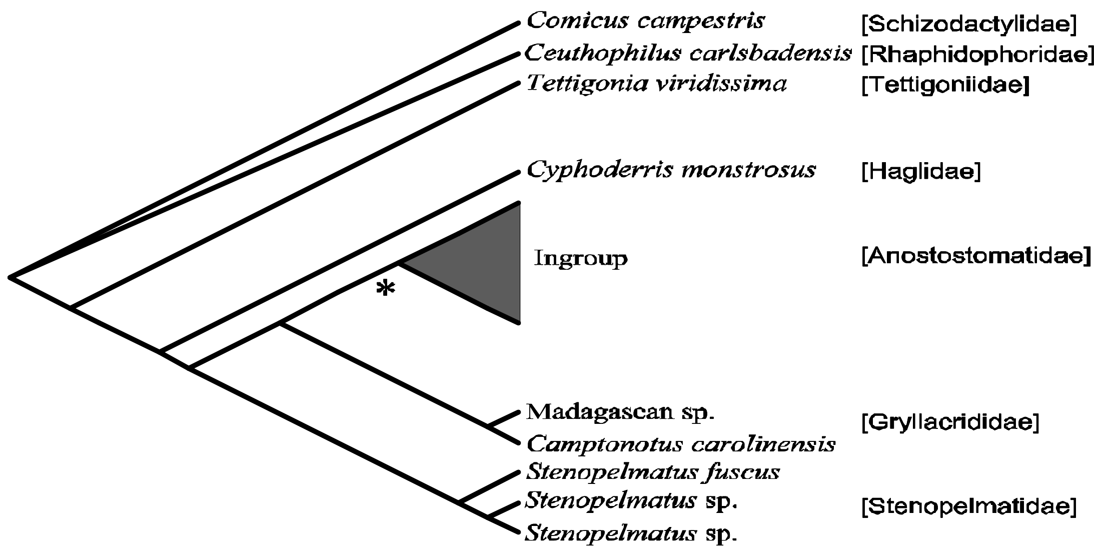


Figure 2.3 Ensifera cladogram

Bayesian cladogram for 1863 bp of nuclear 18S Ensifera data including seven ingroup taxa (EU676721, EU713453, EU713454, EU713455, EU713458, EU713459, Z97570) and published taxa (Flook et al. 1999; Terry & Whiting 2005; Jost & Shaw 2006) used to determine appropriate outgroup (see Appendix 2.8 table 2-1). The hypervariable indel region between bp 719 - 756 was removed due to alignment uncertainty. The asterisk node indicates Bayesian posterior probability (BPP) of 0.96.

Dataset II -18S Australasian Anostostomatidae: After establishing support for monophyly of the Anostostomatidae, we turned our focus to the relationships within the family. We included more representatives from the Australasian region and a slightly shorter fragment of 18S (29 taxa, 1,746 bp), again excluding the problematic indel region. Bayesian and ML analyses yielded similar topologies (figure 2.4). We observed that the New Zealand tusked weta (*Anisoura/Motuweta*; clade A) and New Caledonian taxa (*Aistus* and *Carcinopsis*) formed long branches in the phylogeny. Long branches such as these can result in misleading results even without rate differences (Hendy and Penny 1989) that affect all further tree selection criteria. We explored the effect of these long branches by subjecting the dataset to identical analyses with the inclusion or exclusion of either or both, the New Caledonian taxa (*Aistus* and *Carcinopsis*), or the tusked weta (*Anisoura/Motuweta*) sequences. The exclusion of *Anisoura/Motuweta* resulted in *Aistus* and *Carcinopsis* together being placed as sister to the rest of the Anostostomatidae, from which we infer long branch attraction, resulting from lineage-specific rate increases. When *Aistus* and *Carcinopsis* were removed, the *Anisoura/Motuweta* (clade A; figure 4) were sister to the *Hemideina–Deinacrida* (clade B; figure 4).

In no instances were *Hemiandrus* (New Zealand and Australian) found to be monophyletic, a finding consistent in the following analyses with nuclear and mitochondrial sequences. We identified two clades within the New Zealand *Hemiandrus* (CI and CII; figure 2.4) although not monophyletic in every analysis, both were separate from the Australian *Hemiandrus* species. Dataset III (concatenated 18S + 28S) Australasian anostostomatids (35 taxa, 1,549 bp) drew on data published by Jost & Shaw (2006) and to allow for direct comparison with that study, we cut our 18S sequences to 1,204 bp and included 344 bp of 28S data for a subset of our taxa. We again failed to find monophyly of the Australian taxa or the New Zealand taxa but the grouping of *Anisoura/Motuweta* (clade A) with *Aistus* and *Carcinopsis* as sister to *Hemideina–Deinacrida* (clade B) was supported (tree not shown).

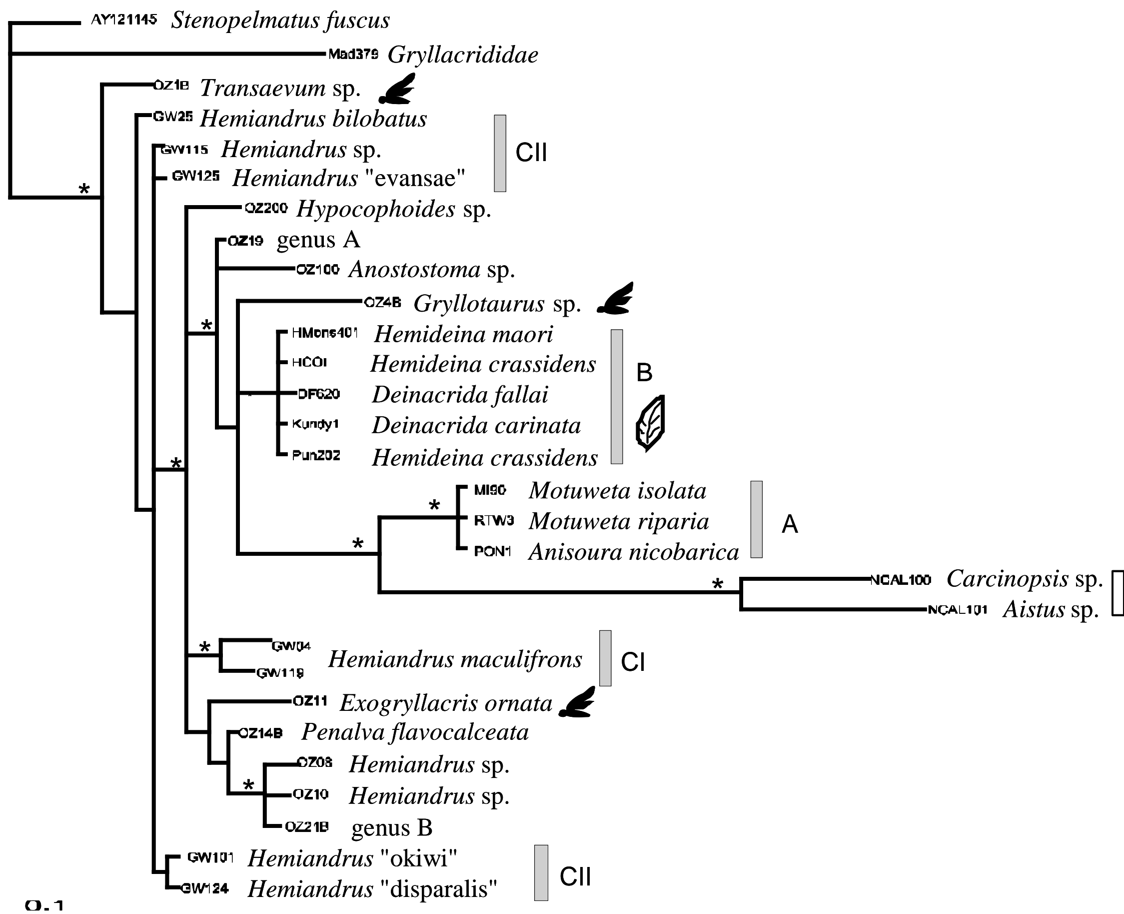


Figure 2.4 Bayesian phylogram using nuclear DNA

Bayesian phylogram for 1,746 bp of nuclear genes 18S including representatives of the Australasian anostostomatids. Winged silhouette indicates winged species and the leaf symbol indicates herbivorous species. The grey and open bars indicate New Zealand and New Caledonian taxa respectively. Clades of New Zealand taxa are indicated as A, B, CI and CII. The single asterisks at nodes indicate BPP above 0.9.

Dataset IV (COI sequences from Australasian taxa); consisted of 1,225 bp aligned DNA sequence data for 28 ingroup taxa, plus two outgroup taxa (one Madagascan gryllacridid [Mad 379] and one South African stenopelmatid [F234] see table 2-1). With RY-coded mitochondrial COI data our phylogenetic analyses returned clades consistent with those observed in analysis of 18S data (figure 4), including strong support for multiple, distinct weta lineages in New Zealand (clades A, B, CI, CII; figure 5). This result was strongly supported by the SH-test constraining New Zealand taxa to form a monophyletic lineage ($P < 0.0001$). *Anisoura/Motuweta* (clade A) was again found to be sister to *Aistus* and *Carcinopsis* and collectively were sister to the *Hemideina/Deinacrida* (clade B) (figure 2.5).

We examined the translated amino acid sequence and observed 10 changes shared by *Aistus, Carcinopsis* and *Anisoura/Motuweta* (out of 409) clearly separating them from the rest of Anostostomatidae. Analysis of dataset V, a reduced set of taxa with concatenated sequences, consisted of COI (1,225 bp) and 12S (435 bp) for Australasian Anostostomatids (25 taxa, 1860 bp). Analyses returned a similar topology with the same taxon subgroupings and no support for New Zealand monophyly as with previous datasets (not shown).

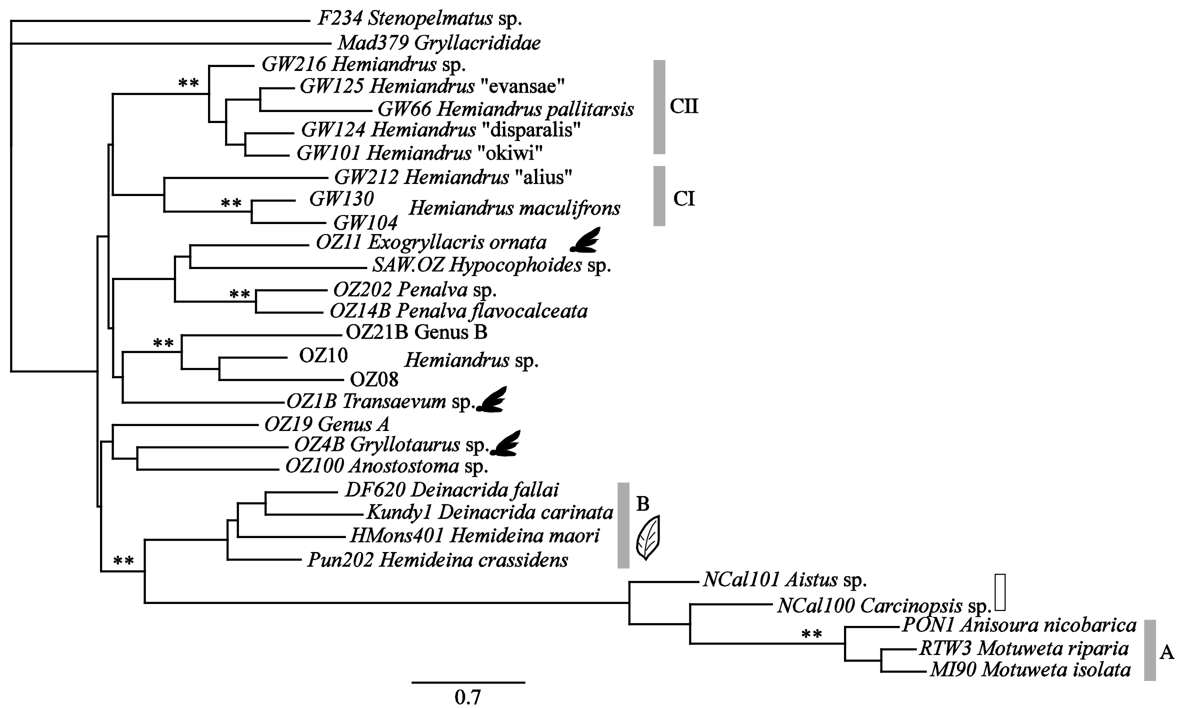


Figure 2.5 Maximum likelihood phylogram using mitochondrial DNA

ML phylogram for 1225 bp of the mitochondrial COI gene representing Australasian anostostomatids. Sequences were R-Y-coded on the third codon position to reduce noise while retaining phylogenetic signal. Winged silhouette indicates winged species and the leaf symbol indicates herbivorous species. The grey and open bars indicate New Zealand and New Caledonian taxa respectively. Clades of New Zealand taxa are indicated as A, B, CI and CII. The double asterisks at nodes indicate BPP 0.99 and above.

Analysis of dataset VI (COI *Hemiandrus* only); for 46 individuals from at least 15 New Zealand species, and two Australian *Hemiandrus* species (848 bp), revealed high genetic diversity (mean genetic distance 6.5 and 7.2% for the North and South Islands respectively) among samples of the widespread New Zealand taxon *H. maculifrons*.

Samples of *H. maculifrons* formed two main clades, one from South Island and one from North Island (clade CI, figure 2.6), with a specimen from northern South Island (sample GW93A from Pelorus Bridge) sister to the North Island clade. In contrast, most other New Zealand *Hemiandrus* species were closely related to one another (e.g. 3.9% between *H. bilobatus* and H. “promontorius”), and these species tend to have narrow geographic ranges. All ground weta (whether North Island or South Island) species with extremely reduced/short ovipositors form a monophyletic group with long ovipositor species paraphyletic to this.

Two taxa with medium ovipositors (*H*. “okiwi’ and *H*. “evensae”) each are sister to long ovipositor species, and not together. Other anostostomatids, including the sister Australian “*Hemiandrus*” have long ovipositors and therefore the reduction of ovipositor length (along with the unusual maternal care that appears to correlate with these conditions), appears to have evolved at least three times in New Zealand. Investigation into monophyly of the genus using the SH-test on dataset IV revealed that constraining all *Hemiandrus* (Australian and New Zealand) to be monophyletic resulted in a tree with a significantly worse likelihood score (P < 0.0001).

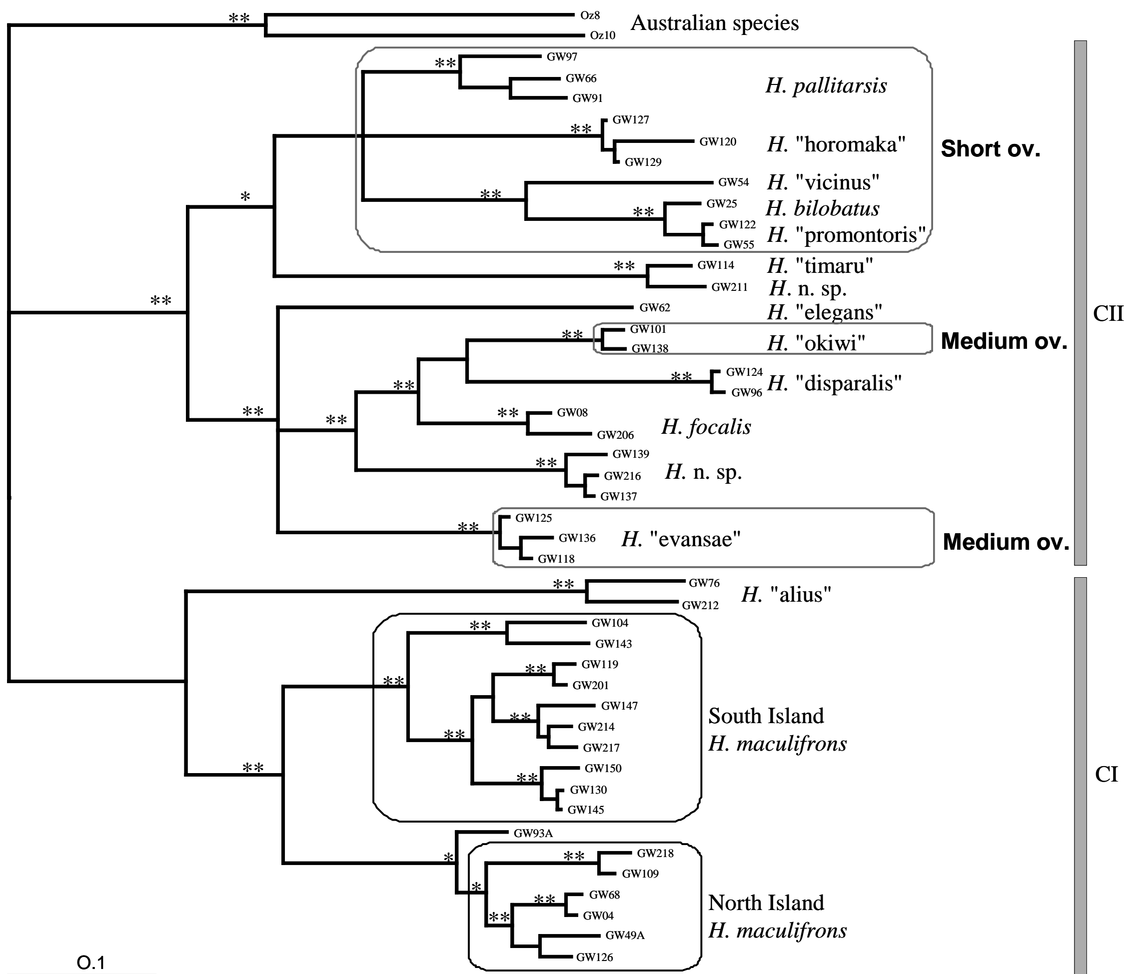


Figure 2.6 Bayesian phylogram for mitochondrial DNA for the genus *Hemiandrus*

Bayesian phylogram of the genus *Hemiandrus* for 848 bp of the mitochondrial COI gene. Species with short or medium-length ovipositors are indicated; all other species have long ovipositors. The two New Zealand clades CI and CII are indicated. The single and double asterisks at nodes indicate BPP above 0.9 and 0.99 respectively.

(b) Summary of phylogenetics

New Zealand weta are not monophyletic, instead forming three (or four) clades. New Zealand tusked weta (clade A) genera are more closely related to the New Caledonian taxa (*Aistus* sp. and *Carcinopsis* sp.; figure 2.4 and 2.5) with no apparent close Australian relative. New Zealand's large herbivorous tree and giant weta (*Hemideina* and *Deinacrida)* form a distinct clade which is strongly supported with less than 1% sequence divergence in the nuclear rDNA gene 18S (clade B, figure 2.4).

All our analyses returned a clade consisting of *Aistus, Carcinopsis*, *Anisoura/Motuweta* (clade A) and *Hemideina–Deinacrida* (clade B). However, long branches due to an apparent rate shift in the *Aistus, Carcinopsis* and *Anisoura/Motuweta* (clade A) lead to us removing these taxa from further analyses so this relationship is not fully understood. The third New Zealand clade, the *Hemiandrus,* consists of two lineages (clade CI and CII, figures 2.4, 2.5 and 2.6) which may not be sisters. One lineage consists of only *H. maculifrons* and *H*. "alius" while the other includes 11+ species sampled (figure 2.6). Lastly, the nine Australian taxa are not monophyletic but form four clades throughout the tree, sister to New Zealand clades, however there is little BPP support for these nodes. This is consistent with the short branch lengths obtained at the base of the tree.

Dataset IV (COI-RY coded Australasia; figure 2.5) returned three Australian clades of interest; (1) the winged *Transaevum*, and non-winged Australian *Hemiandrus* and Genus B, (2) the winged genus *Exogryllacris* sister to the non-winged *Hypocophoides* and *Penalva* species, (3) the third winged species from Australia, *Gryllotaurus*, *Anostostoma* and Genus A species. These three Australian clades were not resolved in analysis of dataset II (18S Australasia; figure 4) but the three winged species were never monophyletic and *Transaevum* was sister to all the other Anostostomatidae.

(c) Divergence estimation

Edge lengths on branches suggest *Aistus* and *Carcinopsis* along with *Anisoura/Motuweta* (clade A, figure 2.4 and 2.5) may have an increased rate of mutation. In order to test for this, we compared ML trees with and without a clock enforced and found evidence in both 18S and COI datasets for a significant deviation from a clock-like model of evolution. By removing *Aistus, Carcinopsis* and *Anisoura/Motuweta* and running the ML analysis again, the significance of the SH-test decreased.

We carried out relaxed clock analyses on 18S (dataset II) and COI (dataset IV). Our calibration points assumed either existence of all anostostomatid weta lineages prior to the separation of Zealandia (82 Ma) or emergence of New Caledonia (40 Ma). However, due to the observed rate change in *Aistus, Carcinopsis* and *Anisoura/Motuweta* we removed this clade (five sequences) from the dating analysis.

Estimates of the rate of evolution of COI vary a great deal due primarily to the disparity between population and species level rates (time dependent rates or J-shape curve: Ho et al. 2005; Penny 2005). Here we restrict our comparisons to recent studies that examine family level divergences of insects which estimate the rate of evolution for COI and COII of insects at about 0.007 – 0.012 substitutions per site per million years (Zakharov et al. 2004 and references therein). In our relaxed molecular clock phylogenetic analysis, constraining the base of the Anostostomatidae clade to 82 Ma, our estimated mutation rates for COI were found to be between 0.0097 – 0.0376 (substitutions per site per million years) [95% CI], compared to a mutation rate of 0.0223 – 0.1219 when applying the 40 Ma constraint. The mutation rate derived using the 82 Ma constraint is more similar to published data for other insects (Zakharov et al. 2004), and therefore supports divergences within Anostostomatidae of more than 80 Ma. Our relaxed clock analyses of COI data suggest that both New Zealand *Hemiandrus* clades and the *Hemideina–Deinacrida* clade may have diverged from Australian relatives more than 82 Ma. Analysis of our 18S sequences with the 82 Ma constraint, estimates divergence for the New Zealand *Hemiandrus* clades that extend to before continental breakup (45 - 120 Ma and 11 - 103 Ma), but the *Hemideina–Deinacrida* clade was estimated to have diverged post-Gondwanan breakup (3 - 38 Ma).

Thus our sequence data supports the idea that some weta lineages may have diverged before the separation of Zealandia from Australia but also that dispersal has since occurred. The phylogenetic placement of *Aistus, Carcinopsis* and *Anisoura/Motuweta* as sister to *Hemideina–Deinacrida* suggests genetic exchange between New Zealand and New Caledonia after separation of Zealandia.

2.5 DISCUSSION

Despite comprehensive morphological studies, phylogenetic relationships within the Ensifera are poorly understood (Gwynne 1995; Whiting 2002a; Desutter-Grandcolas 2003). Johns (1997) removed taxa from Stenoplematidae to form Anostostomatidae, a separation subsequently supported by molecular analyses (Jost and Shaw 2006). Although we are not concerned here with deeper Ensiferan relationships, it is important to know that our taxon set comprises a true ingroup. We found support for monophyly of Anostostomatidae in our analysis (0.96 BPP) and for the close relationship with the Gryllacrididae and Stenopelmatidae (figure 2.3), supporting previous inferences (Jost and Shaw 2006; personal communication P. M. Johns). However, we did not find evidence of a sister relationship of Deinacridinae (*Hemideina* and *Deinacrida*) and Anostostomatinae (rest of the family Johns 1997; Gorochov 2001).

For the first time we have shown that members of the family Anostostomatidae are not monophyletic in New Zealand or Australia. To explain the phylogenetic diversity of New Zealand weta by vicariance requires that at least four distinct clades of Anostostomatidae were already present in Gondwana before Zealandia split from Australia, and that some of these subsequently went extinct in Australia. On the face of it, this seems an unlikely scenario given the small size and geological activity of New Zealand compared to Australia, and indeed this has been shown to be a poor explanation for the distribution of *Nothofagus* beech in the region (Cook and Crisp 2005a). Although we found some variation in node dates inferred from COI and 18S data, we have to reject the hypothesis that all New Zealand lineages arose before continental breakup (*ca* 82 Ma). However, relaxed molecular clock calibrated phylogenies do suggest that some New Zealand clades may have formed before continental separation.

These inferred early splits are consistent with a vicariant origin and survival of some Anostostomatidae lineages on Zealandia throughout the Oligocene marine transgression. Taxa missing from analyses (because of extinction) will always result in long unbroken branches in phylogenetic trees and thus the inference of great age since common ancestors (Cook and Crisp 2005b) whereas recent splits (short branches), cannot be made older by the inclusion of “missing taxa”.

Colonisation of New Zealand from the Australian biota, which includes three separate winged lineages, might have been facilitated by increasing land area after the Oligocene (less than 22 Ma). Dispersal events continue today, and include the establishment of an Australian Gryllacridid in recent years (Green and Ramsay 2003). The current study suggests the two New Caledonia genera are more closely related to one of the New Zealand lineages but not to any Australian taxa. This is despite the comparatively close physical proximity and more similar climate of New Caledonia and Queensland, Australia. Despite evidence of an elevated substitution rate in both nuclear and mitochondrial genes, we observed clear phylogenetic evidence, supported by amino acid substitutions, for the sister relationship of *Aistus, Carcinopsis* and *Anisoura/Motuweta*.

Weta must have colonised New Caledonia after it emerged from the sea *c*a 40 Ma (Grandcolas et al. 2008), but our current taxon sampling is not sufficient to prove whether those weta ancestors came to or from New Zealand. *Hemideina* and *Deinacrida* were found to form a monophyletic group corresponding to the subfamily Deinacridinae Karny 1932. The lack of variation at the 18S gene among eighteen species in these two genera is indicative of recent radiation. Interestingly, Johns (1997) suggested one of the tusked weta genera, *Anisoura,* should be included in the Deinacridinae, however if Deinacridinae is to be valid, it should include all tusked weta genera plus the New Caledonian genera. Further morphological study is required to test this.

Our analysis of Australian and New Zealand *Hemiandrus* shows clearly that they represent separate lineages (figure 2.4). Furthermore *Hemiandrus* in New Zealand consists of two distinct clades that may not be monophyletic. It is clear that taxonomic revision of the genus is required, with *Hemiandrus* being retained only for New Zealand taxa.

Our data are the first to illustrate the depth of diversity within the New Zealand *Hemiandrus* which may amount to some 40 species (Johns 2001), utilising a broad range of habitats from alpine to lowland forest. We note relatively high genetic diversity in the most widespread species (*H. maculifrons*) which might justify separation as two or more allopatric species, but similar levels of sequence divergence are reported from other widespread weta with wide geographic ranges, e.g. *Deinacrida connectens* (Trewick et al. 2000); *Hemideina thoracica* (Morgan-Richards et al. 2001). There is a stark contrast between the wide geographic range of *H. maculifrons* and the small ranges of the numerous, genetically similar, local endemics of South Island (figure 2.2, 2.6). This pattern implies recent diversification in South Island, perhaps in response to habitat diversification since the Pliocene which is also observed in the giant weta (figure 2.2) and many other taxa (Wagstaff and Garnock-Jones 1998; Lockhart et al. 2001; Trewick 2001; Chinn and Gemmell 2004; Trewick 2008) In contrast, much of North Island is young (less than 1 Ma) and relatively homogenous in terms of habitat (Trewick and Morgan-Richards in press).

In terms of geological history, Zealandia bears little resemblance to other larger, more stable southern continents that together originated from Gondwana. Instead, its history is more like that of many oceanic islands, undergoing extensive geological transformations which marked significant changes in climate and topology, and ultimately shaped the biota. The lack of monophyly within the Ansostosomatidae fauna is not entirely unexpected, but the idea that some lineages represent more ancient links between Australia and New Zealand is exciting, as is the apparent exchange between New Zealand and the other Zealandian island New Caledonia. This pattern might be indicative of the New Zealand biota as a whole, where some old lineages surviving from Zealandia, are all but overshadowed by more recent biotic exchange.

2.6 ACKNOWLEDGEMENTS

A special thank you to all those involved in collecting over the years; Geoff Monteith, Peter Johns, George Gibbs, Darryl Gwynne, Scott Dunavan. Special thanks to Matt Phillips and Simon Hills for guidance and assistance with the inner workings of BEAST. Thanks to Andrew Clarke, Sheralee Cleland, Barbara Holland, Manda Jost, David Penny and Michael Whiting for constructive comments on the manuscript.

This research was supported by funding from Marsden Fund granted to David Penny and the Allan Wilson Centre for Molecular Ecology and Evolution. The Orthoptera Society grant awarded to RCP facilitating collaboration with Peter Johns. New Zealand Department of Conservation provided one *Motuweta* and two *Deinacrida* species.

2.7 REFERENCES

Baker, A. J., L. J. Huynen, O. Haddrath, C. D. Millar, and D. M. Lambert. 2005. Reconstructing the tempo and mode of evolution in an extinct clade of birds with ancient DNA: the giant moas of New Zealand. Proceeding of the National Academy of Science, USA **102**:8257-8262.

Caughley, G. 1964. Does the New Zealand vertebrate fauna conform to zoogeographic principles? Tuatara **12**:50-57.

Chardon, D., and V. Chevillotte. 2006. Morphotectonic evolution of the New Caledonia ridge (Pacific Southwest) from post-obduction tectonosedimentary record. Tectonophysics **420**:473-491.

Chinn, W. G., and N. J. Gemmell. 2004. Adaptive radiation within New Zealand endemic spcies of the cockroach genus *Celatoblatta* Johns (Blattidae): a response to Plio-Pleistocene mountain building and climate change. Molecular Ecology **13**:1507-1518.

Cook, L. G., and M. D. Crisp. 2005a. Not so ancient: the extant crown group of *Nothofagus* represents a post-Gondwanan radiation. Proceedings of the Royal Society of London Series B-Biological Sciences **272**:2535-2544.

Cook, L. G., and M. D. Crisp. 2005b. Directional asymmetry of long-distance dispersal and colonization could mislead reconstructions of biogeography. Journal of Biogeography **32**:741-754.

Cowie, R. H., and B. S. Holland. 2006. Dispersal is fundamental to biogeography and the evolution of biodiversity on oceanic islands. Journal of Biogeography **33**:193-198.

Desutter-Grandcolas, L. 2003. Phylogeny and evolution of acoustic communication in extant Ensifera (Insecta, Orthoptera). Zoologica Scripta **32**:525-561.

Drummond, A. J., S. Y. W. Ho, M. J. Phillips, and A. Rambaut. 2006. Relaxed Phylogenetics and Dating with Confidence. PLoS Biology **4**:e88.

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

Ericson, P. G. P., L. Christidis, A. Cooper, M. Irestedt, J. Jackson, U. S. Johansson, and J. A. Norman. 2002. A Gondwanan origin of passerine birds supported by DNA sequences of the endemic New Zealand wrens. Proceedings of the Royal Society of London B Biological Sciences **269**:235-241.

Field, L. H., and N. A. Deans. 2001. Sexual selection and secondary sexual characters of wetas and king crickets. Pp. 179-204 *in* L. H. Field, ed. The biology of wetas, king crickets and their allies. CABI Publishing, New York.

Fleming, C. A. 1979. The geological history of New Zealand and its life. Oxford University Press, Auckland.

Fleming, C. A. 1962. New Zealand biogeography - a palaeontologist's approach. Tuatara **10**:53-108.

Flook, P. K., S. Klee, and C. H. F. Rowell. 1999. Combined molecular phylogenetic analysis of the Orthoptera (Arthropoda, Insecta) and implications for their higher systematics. Systematic Biology **48**:233-253.

Folmer, O., M. Black, W. Hoeh, R. Lutz, and R. Vrijenhoek. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Molecular Marine Biology and Biotechnology **3**:294-299.

Gibbs, G. 2006. Ghosts of Gondwana: The History of Life in New Zealand. Craig Potton Publishing, Nelson, New Zealand.

Goldberg, J., S. A. Trewick, and A. M. Paterson. 2008. Evolution of New Zealand's terrestrial fauna: a review of molecular evicence. Philosophical Transactions of the Royal Society of London B Biological Sciences **363**:3319-3334.

Gorochov, A. V. 2001. The most interesting finds of orthopteroid insects at the end of the 20th century and a new recent genus and species. Journal of Orthoptera Research **10**:353-367.

Grandcolas, P., J. Murienne, T. Robillard, L. Desutter-Grandcolas, H. Jourdan, E. Guilbert, and L. Deharveng. 2008. New Caledonia: a very old Darwinian island? Philosophical Transactions of the Royal Society of London B Biological Sciences **363**:3309-3317.

Green, O. R., and G. W. Ramsay. 2003. A winged weta, *Pterapotrechus* (Orthoptera: Gryllacrididae), established in New Zealand. New Zealand Entomologist **26**:75-77.

Grimaldi, D., and M. S. Engel. 2005. Evolution of Insects. Cambridge University Press, New York.

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

Gwynne, D. T. 1995. Phylogeny of the Ensifera (Orthoptera): A hypothesis supporting multiple origins of acoustical signalling, complex spermatophores and maternal care in crickets, katydids, and weta. Journal of Orthoptera Research **4**:203-218.

Gwynne, D. T. 2004. Reproductive behaviour of ground weta (Orthoptera: Anostostomatidae): Drumming behaviour, nuptial feeding, post-copulatory guarding and maternal care. Journal of the Kansas Entomological Society **77**:414-428.

Hendy, M. D., and D. Penny. 1989. A framework for quantitaive study of evolutionary trees. Systematic Biology **38**:297-309.

Ho, S. Y. W., M. J. Phillips, A. Cooper, and A. J. Drummond. 2005. Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. Molecular Biology and Evolution **22**:1561-1568.

Huelsenbeck, J. P., and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics **17**:754-755.

Jewell, T. 2007. Two new species of *Hemiandrus* (Orthoptera: Anostostomatidae) from Fiordland National Park, New Zealand. Zootaxa **1542**:49-57.

Johns, P. M. 2001. Distribution and conservation status of ground weta, *Hemiandrus* species (Orthoptera: Anostostomatidae). Department of Conservation, Wellington.

Johns, P. M. 1997. The Gondwanaland Weta: Family Anostostomatidae (formerly in Stenopelmatidae, Henicidae or Mimmermidae): Nomenclatural Problems, World Checklist, New Genera and species. Journal of Orthoptera Research **6**:125-138.

Jost, M. C., and K. L. Shaw. 2006. Phylogeny of Ensifera (Hexapoda: Orthoptera) using three ribosomal loci, with implications for the evolution of acoustic communication. Molecular Phylogenetics and Evolution **38**:510-530.

Knapp, M., K. Stöckler, D. Havell, F. Delsuc, F. Sebastiani, and P. J. Lockhart. 2005. Relaxed molecular clock provides evidence for long distance dispersal of *Nothofagus* (Southern Beech). PLoS Biology **3**:e14.

Landis, C. A., H. J. Campbell, J. G. Begg, D. C. Mildenhall, A. M. Paterson, and S. A. Trewick. 2008. The Waipounamu erosion surface:questioning the antiquity of the New Zealand land surface and terrestrial fauna and flora. Geological Magazine:1-25.

Lockhart, P. J., P. A. McLenachan, D. Havell, D. Glenny, D. Huson, and U. Jensen. 2001. Phylogeny, radiation and transoceanic dispersal of New Zealand alpine buttercups: molecular evidence under split decomposition. Annals of the Missouri Botanical Garden **88**:458-477.

McLoughlin, S. 2001. The breakup history of Gondwana and its impact on pre-Cenozoic floristic provincialism. Australian Journal of Botany **49**:271-300.

Meads, M. J. 1990. The Weta Book: A guide to the identification of wetas. DSIR Land Resources, Lower Hutt.

Monteith, G. B., and L. H. Field. 2001. Australian king crickets: distribution, habitats and biology (Orthoptera:Anostostomatidae). Pp. 79-94 *in* L. H. Field, ed. The biology of wetas, king crickets and their allies. CABI Publishing, Wallingford.

Morgan-Richards, M., S. A. Trewick, and G. P. Wallis. 2001. Chromosome races with Pliocene origins: evidence from mtDNA. Heredity **86**:303-312.

Mortimer, N. 2004. New Zealand's Geological Foundations. Gondwana Research **7**:261-272.

Mortimer, N., K. Hoernle, F. Hauff, J. M. Palin, W. J. Dunlap, R. Werner, and K. Faure. 2006. New constraints on the age and evolution of the Wishbone Ridge, southwest Pacific Cretaceous microplates, and Zealandia-West Antarctica breakup. Geology **34**:185-188.

Neall, V. E., and S. A. Trewick. 2008. The age and origin of the Pacific islands: a geological overview. Philosophical Transactions of the Royal Society of London B Biological Sciences **363**:3293-3308.

Penny, D. 2005. Relativity for molecular clocks. Nature **436**:183-184.

Phillips, M. J., and D. Penny. 2003. The root of the mammalian tree inferred from whole mitochondrial genomes. Molecular Phylogenetics and Evolution **28**:171-185.

Posada, D., and T. R. Buckley. 2004. Model selection and model averaging in phylogenetics: Advantages of Akaike Information Criterion and Bayesian approaches over Likelihood Ratio tests. Systematic Biology **53**:793-808.

Posada, D., and K. A. Crandall. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics **14**:817-818.

Rambaut, A. 1996. Se-Al: Sequence Alignment Editor. Available at http://evolve.zoo.ox.ac.uk.

Shimodaira, H., and M. Hasegawa. 1999. Multiple comparisons of Log-likelihoods with applications to phylogenetic inference. Molecular Biology and Evolution **16**:1114-1116.

Simon, C., F. Frati, A. Bechenbach, B. Crespi, H. Liu, and P. Flook. 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Annals of the Entomological Society of America **87**:651-701.

Sunnucks, P., and D. F. Hale. 1996. Numerous transposed sequences of mitochondrial cytochrome oxidase I-II in aphids of the genus Sitobion (Hemiptera: Aphididae). Molecular Biology and Evolution **13**:510-524.

Swofford, D. L. 2003. PAUP*: Phylogenetic Analysis Using Parsimony (and Other Methods), Version 4. Sinauer Associates, Sunderland, Massachusetts.

Terry, M. D., and M. F. Whiting. 2005. Mantophasmatodea and phylogeny of the lower neopterous insects. Cladistics **21**:240-257.

Trewick, S. A. 2008. DNA Barcoding is not enough: mismatch of taxonomy and genealogy in New Zealand grasshoppers (Orthoptera: Acrididae). Cladistics **24**:240–254.

Trewick, S. A. 2001. Identity of an endangered grasshopper (Acididae: Brachaspis): Taxonomy, molecules and conservation. Conservation Genetics **2**:233-243.

Trewick, S. A., and M. Morgan-Richards. 2004. Phylogenetics of New Zealand's tree, giant and tusked weta (Orthoptera: Anostostomatidae): evidence from mitochondrial DNA. Journal of Orthoptera Research **13**:185-196.

Trewick, S. A., and M. Morgan-Richards. 2005. After the deluge: mitochondrial DNA indicates Miocene radiation and Pliocene adaptation of tree and giant weta (Orthoptera: Anostostomatidae). Journal of Biogeography **31**:295-309.

Trewick, S. A., and M. Morgan-Richards. in press. Biology of New Zealand. University of California Press, Berkeley, CA.

Trewick, S. A., A. M. Paterson, and H. J. Campbell. 2007. Hello New Zealand. Journal of Biogeography **34**:1-6.

Trewick, S. A., G. P. Wallis, and M. Morgan-Richards. 2000. Phylogeographical pattern correlates with Pliocene mountain building in the alpine scree weta (Orthoptera, Anostostomatidae). Molecular Ecology **9**:657-666.

Vawter, L. 1991. Evolution of blattoid insects and of the small subunit ribosomal gene. University of Michigan, Ann Arbor.

Veevers, J. J., and Z. X. Li. 1991. Review of seafloor spreading around Australia. II. Marine magnetic anomaly modelling. Australian Journal of Earth Sciences **38**:391-408.

Wagstaff, S. J., and P. J. Garnock-Jones. 1998. Evolution and biogeography of the Hebe complex (Scrophulariaceae) inferred from ITS sequences. New Zealand Journal of Botany **36**:425-437.

Waters, J. M., and D. Craw. 2006. Goodbye Gondwana? New Zealand biogeography, geology and the problem of circularity. Systematic Biology **55**:351-356.

Whiting, M. F. 2002a. Phylogeny of the holometabolous insect orders: molecular evidence. Zoologica Scripta **31**:3-15.

Whiting, M. F. 2002b. Mecoptera is paraphyletic: multiple genes and phylogeny of Mecoptera and Siphonaptera. Zoologica Scripta **31**:93-104.

Zakharov, E., M. Caterino, and F. Sperling. 2004. Molecular phylogeny, historical biogeography and divergence time estimates for swallowtail butterflies of the genus *Papilio* (Lepidoptera: Papilionidae). Sys. Biol. **54**:193-215.

Zwickl, D. J. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. The University of Texas, Austin.

2.8 APPENDIX

Table 2-1 Sample information

All specimens are housed at the AWC Massey University, Palmerston North, New Zealand.

species	specimen ID	country	locality	18S	28S	COI	12S	I	II	III	IV	V	VI
					GeneBank/EMBL accession numbers					dataset			
Carcinopsis sp.	NCal 100	New Caledonia		EU676733	EU676696	EU676751	EU676669		✓	✓	✓	✓	
Aistus sp.	NCal 101	New Caledonia		EU676734	EU676695	EU676752	EU676670		✓	✓	✓	✓	
Aistus sp.	NCal7	New Caledonia		EU713454				✓					
Anostostoma sp.	OZ1	Australia		EU713458				✓		✓			
Transaevum sp.	OZ1B	Australia		EU676723	EU676699	EU676754	EU676672		✓	✓	✓	✓	
Penalva lateralis	OZ2	Australia		EU713459				✓		✓			
Transaevum sp.	OZ2B	Australia			EU676700				✓				
Gryllotaurus sp.	OZ4B	Australia		EU676724	EU676701	EU676755	EU676673		✓	✓	✓	✓	
Hemiandrus sp.	OZ08	Australia		EU676725	EU676697	EU676756	EU676674		✓	✓	✓	✓	✓
Hemiandrus sp.	OZ10	Australia		EU676726	EU676702	EU676757	EU676675		✓	✓	✓	✓	✓
Exogryllacris ornata	OZ11	Australia		EU676727	EU676703	EU676758	EU676676		✓	✓	✓	✓	
Penalva flavocalceata	OZ14B	Australia		EU676728	EU676704	EU676759	EU676677		✓	✓	✓	✓	
Genus A	OZ19	Australia		EU676729	EU676698	EU676760	EU676678		✓	✓	✓	✓	
Genus B	OZ21B	Australia		EU676730	EU676705	EU676761	EU676679		✓	✓	✓	✓	
Anostostoma sp.	OZ100	Australia		EU676731	EU676706	EU676762	EU676680		✓	✓	✓	✓	
Hypocophoides sp.	OZ200	Australia		EU676732	EU676707				✓	✓			
Penalva sp.	OZ202	Australia				EU676763					✓		
Hypocophoides sp.	SAW.OZ	Australia				EU676753	EU676671				✓	✓	
Deinacrida carinata	Kundy1	New Zealand	SI, Kundy Island	EU676711	EU676684	EU676737	EU676658		✓	✓	✓	✓	
Hemideina maori	HMons401	New Zealand	SI, Porters Pass	EU676708	EU676685	EU676736	EU676657		✓	✓	✓	✓	
Hemideina crassidens	HCOI	New Zealand	NI, Palmerston North	EU676709					✓				
Hemideina crassidens	Pun202	New Zealand	SI, Punakaikai	EU676712		EU676738	EU676659		✓		✓	✓	
Deinacrida fallai	DF620	New Zealand	NI, Poor Knights	EU676710		EU676739	EU676660		✓		✓	✓	
Deinacrida connectens	Math1	New Zealand	SI, Matheson	EU713453				✓					
Hemideina crassidens	Hemideina	New Zealand		Z97570				✓					
Motuweta isolata	MI90	New Zealand	NI, Middle Island, Mecury Islands	EU676720	EU676689	EU676748	EU676666		✓	✓	✓	✓	
Motuweta riparia	RTW3	New Zealand	NI, Motu River	EU676721	EU676687	EU676750	EU676668	✓	✓	✓	✓	✓	
Anisoura nicobarica	PON1	New Zealand	NI, Hokianga	EU676722	EU676688	EU676749	EU676667		✓	✓	✓	✓	
Hemiandrus focalis	GW08	New Zealand	SI, Conical Hill			EU676773							✓
Hemiandrus focalis	GW206	New Zealand	SI, Obelisk			EU676774							✓
Hemiandrus bilobatus	GW25	New Zealand	NI, Wellington	EU676714		EU676794				✓			✓
Hemiandrus pallitarsis	GW66	New Zealand	NI, Lake Waikaremoana			EU676740	EU676661				✓	✓	✓
Hemiandrus pallitarsis	GW91	New Zealand	NI, Mt. Holdworth			EU676768							✓
Hemiandrus pallitarsis	GW97	New Zealand	NI, Wellington			EU676797							✓
Hemiandrus subantarcticus	Snare2	New Zealand	SI, Snare Island	EU713455	EU676686			✓	✓	✓			
Hemiandrus "vicinus"	GW54	New Zealand	SI, Whites Bay			EU676788							✓
Hemiandrus "elegans"	GW62	New Zealand	NI, Moehau			EU676790							✓
Hemiandrus "disparalis"	GW96	New Zealand	NI, Karamea river			EU676800							✓
Hemiandrus "disparalis"	GW124	New Zealand	SI, near Dobson	EU676718		EU676743			✓	✓	✓		✓
Hemiandrus "okiwi"	GW101	New Zealand	SI, Takaka	EU676715		EU676741			✓		✓		✓
Hemiandrus "okiwi"	GW138	New Zealand	SI, St. Arnaud			EU676783							✓
Hemiandrus "timaru"	GW114	New Zealand	SI, Kurinui		EU676693	EU676769			✓				✓
Hemiandrus "evansae"	GW118	New Zealand	SI, Kurinui			EU676795							✓
Hemiandrus "evansae"	GW125	New Zealand	SI, Portabello	EU676719	EU676694	EU676744			✓	✓	✓		✓

				GeneBank/EMBL accession numbers				dataset					
species	specimen ID	country	locality	18S	28S	COI	12S	I	II	III	IV	V	VI
Hemiandrus "evansae"	GW136	New Zealand	SI, Blue Mountains			EU676780							✓
Hemiandrus "horomaka"	GW120	New Zealand	SI, Banks Peninsular			EU676771							✓
Hemiandrus "promontoris"	GW122	New Zealand	SI, Cape Campbell			EU676777							✓
Hemiandrus new sp. (Johns pers. comm.)	GW211	New Zealand	SI, Bourkes Pass			EU676792							✓
Hemiandrus sp.	GW55	New Zealand	SI, Marfell's Beach			EU676789							✓
Hemiandrus sp.	GW115	New Zealand	SI, Kurinui	EU676716					✓				
Hemiandrus sp.	GW127	New Zealand	SI, Porters Pass			EU676778							✓
Hemiandrus sp.	GW129	New Zealand	SI, Porters Pass		EU676692	EU676779			✓				✓
Hemiandrus sp.	GW137	New Zealand	SI, Kahurangi NP			EU676781							✓
Hemiandrus sp.	GW139	New Zealand	SI, St. Arnaud			EU676782							✓
Hemiandrus sp.	GW216	New Zealand	SI, Mt. Haidinger			EU676747	EU676665				✓	✓	✓
Hemiandrus "alius"	GW76	New Zealand	SI, Asakiri Valley			EU676766							✓
Hemiandrus "alius"	GW212	New Zealand	SI, Mt. Haidinger			EU676746	EU676664				✓	✓	✓
Hemiandrus maculifrons	GW04	New Zealand	NI, Gisborne	EU676713		EU676798			✓				✓
Hemiandrus maculifrons	GW49A	New Zealand	NI, Puketi Forest			EU676765							✓
Hemiandrus maculifrons	GW68	New Zealand	NI, Wellington		EU676690	EU676767				✓			✓
Hemiandrus maculifrons	GW93A	New Zealand	SI, Pelorus			EU676791							✓
Hemiandrus maculifrons	GW104	New Zealand	SI, Takaka			EU676742	EU676662				✓	✓	✓
Hemiandrus maculifrons	GW109	New Zealand	NI, Opepe			EU676799							✓
Hemiandrus maculifrons	GW119	New Zealand	SI, Catlins	EU676717		EU676770			✓				✓
Hemiandrus maculifrons	GW126	New Zealand	NI, Pureora			EU676793							✓
Hemiandrus maculifrons	GW130	New Zealand	SI, Craigieburn Range		EU676691	EU676745	EU676663			✓	✓	✓	✓
Hemiandrus maculifrons	GW143	New Zealand	SI, Lewis Pass			EU676784							✓
Hemiandrus maculifrons	GW145	New Zealand	SI, Arthur's Pass			EU676785							✓
Hemiandrus maculifrons	GW147	New Zealand	SI, Paparoa			EU676787							✓
Hemiandrus maculifrons	GW150	New Zealand	SI, Franz Josef			EU676786							✓
Hemiandrus maculifrons	GW201	New Zealand	SI, Takitimu Range			EU676772							✓
Hemiandrus maculifrons	GW214	New Zealand	SI, Mt. Owen			EU676775							✓
Hemiandrus maculifrons	GW217	New Zealand	SI, Tasman wilderness area			EU676776							✓
Hemiandrus maculifrons	GW218	New Zealand	NI, Raurimu			EU676796							✓
Outgroup													
Stenopelmatus sp.	F79	South Africa		EU713456				✓					
Stenopelmatus sp.	F234	South Africa		EU713457		EU713461		✓			✓		
Gryllacrididae	Mad379	Madagascar		EU676735	EU676683	EU676764	EU676681	✓	✓	✓	✓	✓	
Gryllidae	Mad373	Madagascar		EU713460	EU676682					✓			
Tettigonia viridissima				Z97587				✓					
Comicus campestris				Z97564				✓					
Ceuthophilus carlsbadensis				Z97563				✓					
Cyphoderris monstrosus				Z97566				✓					
Camptonotus carolinenesis				AY521876				✓		✓			
Australostoma sp.				AF514559	AF514437 ‡					✓			
Australostoma sp.				AF514525						✓			
Sia sp.				AF514561	AF514440 ‡					✓			
Sia sp.				AF514527						✓			
Stenopelmatus sp.				AF514562	AF514441 ‡					✓			
Stenopelmatus sp.				AF514528						✓			
Stenopelmatus fuscus				AY121145				✓	✓	✓			

CHAPTER THREE

The great tragedy of Science — the slaying of a beautiful hypothesis by an ugly fact

- T. H. Huxley

This chapter published as;

Pratt, R. C., Gibb, G. C., Morgan-Richards, M., Phillips, M. J., Hendy, M. D. and Penny, D. in press. Towards resolving deep Neoaves phylogeny: data, signal enhancement and priors. Molecular Biology and Evolution.

See Appendix D for reprint

Towards resolving deep Neoaves phylogeny: data, signal enhancement and priors

3.1 ABSTRACT

We report three developments towards resolving the challenge of the apparent basal polytomy of neoavian birds. First we describe improved conditional down-weighting techniques to reduce noise relative to signal for deeper divergences and find increased agreement between datasets. Second, we present formulae for calculating the probabilities of finding predefined groupings in the optimal tree. Finally, we report a significant increase in data: nine new mitochondrial genomes (the dollarbird, New Zealand kingfisher, great potoo, Australian owlet-nightjar, white-tailed trogon, barn owl, a roadrunner (a ground cuckoo), New Zealand long-tailed cuckoo and the peach-faced lovebird) and together they provide data for each of the six main groups of Neoaves proposed by Cracraft in 2001. We use his six main groups of modern birds as priors for evaluation of results. These include passerines; cuckoos; parrots and three other groups termed 'WoodKing' (woodpeckers/rollers/kingfishers), 'SCA' (owls/potoos/owlet-nightjars/hummingbird/swift) and 'Conglomerati'. In general the support is highly significant with just two exceptions, the owls move from the 'SCA' group to the raptors, particularly accipitrids (buzzards/eagles) and the osprey, and the shorebirds may be an independent group from the rest of the 'Conglomerati'. Molecular dating of mitochondrial genomes, support a major diversification of at least 12 Neoavian lineages in the Late Cretaceous. Our results form a basis for further testing with both nuclear coding sequences and rare genomic changes.

3.2 INTRODUCTION

Perhaps the greatest current challenge of avian systematics for molecular evolutionists and systematists alike is the resolution of the polytomy at the base of the Neoaves. The basic paleognath (tinamous and ratites)/neognath (all other modern birds) division is supported by studies of morphology (Cracraft 2001), nuclear coding DNA (García-Moreno and Mindell 2000; Barrowclough et al. 2006), and mitochondrial (mt) genomes (Mindell et al. 1999; Harrison et al. 2004; Slack et al. 2007).

Within the Neognathae, the Galloanserae (chickens, ducks and their relatives) represent the earliest divergence, leaving the large majority (all remaining orders) of birds in the Neoaves. Again, coding regions of both mitochondrial genomes and nuclear DNA, together with morphological data, agree with the Galloanserae division. However, resolving the relationships within Neoaves is still elusive.

Thus resolution of the basal Neoavian polytomy could be seen as the 'last frontier' for resolving deep-level systematics among modern birds. There are a range of views best illustrated by the two ends of a spectrum – firstly the theory that the basal polytomy is due to an 'explosive radiation' after the Cretaceous-Paleogene (K - Pg, formerly K - T) boundary. That is, birds and mammals 'inherited the earth' only after the demise of the dinosaurs and pterosaurs (Feduccia 2003; Chubb 2004; Poe and Chubb 2004; Ericson et al. 2006). The other end of the spectrum are hypotheses that basal avian lineages were diversifying in an 'adaptive radiation' long before the asteroid impact that marks the K - Pg boundary (Cooper and Penny 1997; Cracraft 2001; van Tuinen and Hedges 2001; Penny and Phillips 2004; Pereira and Baker 2006; van Tuinen et al. 2006; Brown et al. 2007; Brown et al. 2008). This latter approach represents mainstream evolutionary theory in that it attempts to explain the past by reference to known mechanisms – to 'causes now in operation' (Penny and Phillips 2004).

Poe and Chubb (2004) suggested that the large polytomy at the base of Neoaves represents a rapid radiation that 'might be considered essentially simultaneous'. If a lack of resolution is not caused by truly short times between divergences, then ultimately relationships should be resolvable (Whitfield and Lockhart 2007). We have already commented (Gibb et al. 2007) that an 'explosive radiation' implies both short divergence times between avian orders, and also that the ecological and morphological differences that identify the crown groups of orders within Neoaves must have occurred over the same short time-scale. It would scarcely be an 'explosive radiation' if the lineages diverged quickly but it then took tens of millions of years for genetic changes to occur leading to the ecological and morphological characters that distinguish crown-group Neoavian orders today. Apart from being real, short branch lengths in phylogenies can result for a number of reasons.

For example different characters or data sources which provide support for conflicting trees (rather than from the absence of support) can result in short branches, under these conditions even standard maximum likelihood (ML) can seriously underestimate branch lengths (Penny et al. 2008). Additionally, use of inappropriate genes or analysis methods can return short branches. To improve divergence time estimates, Brown et al. (2008) recommend longer sequences (they used 4 594 base pairs of mtDNA) and the use of multiple independent nuclear loci. In addition to incorporating longer sequences, improved analytic methods such as networks, allows visualization of conflict between essentially equally good phylogenies (Holland et al. 2004).

A greater understanding of the evolutionary history of Neoaves is still needed. We have found Cracraft's (2001) six groupings within Neoaves to be useful as 'informal priors' for recent studies on passerines and for the group termed 'Conglomerati' (e.g. Gibb et al. 2007). With additional data presented here, we have representatives (and test) his six prior groupings within Neoaves:

(i) Passeriformes (passerines, perching birds);

(ii) Psittaciformes (parrots);

(iii) Cuculiformes (cuckoos);

(iv) Coliiformes, Coraciiformes, Trogoniformes and Piciformes (mousebirds [not included in the present study], rollers/bee-eaters/ kingfishers, trogons, woodpeckers/toucan, jacamars). Informally, and for ease of communication, we call this the 'WoodKing' grouping;

(v) Strigiformes, Caprimulgiformes, Apodiformes and Musophagiformes (owls, nightjars, potoos, owlet-nightjars, hummingbirds, swifts and turacos). We have sampled three of the four groups, so we will refer to these as 'SCA' (or 'CA' when omitting the owls [Strigiformes]);

(vi) A diverse group dubbed the 'Conglomerati' (Slack et al. 2007) which includes three main subgroups: a) Falconiformes, Pelecaniformes, Ciconiiformes, Procellariiformes, Gaviiformes Sphenisciformes, and Podicipediformes (raptors, pelicans and relatives, storks, seabirds, loons, penguins and grebes); b) Charadriiformes, Columbiformes, Pteroclidae (shorebirds, pigeons and sandgrouse); c) Gruidae, Rallidae, Otididae, Eurypygidae and Turnices (cranes, rails, bustards, kagu, sun bitterns and button quails).

In recent papers (Slack et al. 2006; Slack et al. 2007) we concentrated on the relationships within (i) and (vi), but have recently extended this to include members of groups (iv) and (v) (Gibb et al. 2007; Morgan-Richards et al. 2008). Because, we previously only had a single representative for the orders Psittaciformes (parrots) and Strigiformes (owls) (Harrison et al. 2004) we omitted them from recent analyses because they are long branches which are known to be problematic in phylogeny generally (Hendy and Penny 1989), including birds (Harrison et al. 2004).

At this stage we are particularly concerned as to whether the members within each of the above six groups come together – this will help evaluate whether the deep Neoavian lineages are resolvable. As such, we are not especially concerned if some of Cracraft's six groups have paraphyletic lineages within them – that is a taxonomic question, not a question about the resolvability of the deepest lineages. For example, we are interested in whether the combined Coliiformes/Coraciiformes/Piciformes ('WoodKing') form a natural group within Neoaves – even if one of the subgroups turns out to be paraphyletic within this 'WoodKing' grouping.

After this work was submitted, Hackett et al. (2008) published what is probably the most comprehensive report on bird evolution since Cracraft (2001). While the scope of data is impressive, it is largely based on non-coding intron sequences. Nuclear intron data, in combination with current mitochondrial and fossil data has the potential to be extremely useful as long as we can be confident in alignments that span the phylogenetic depth of the avian clade. Some authors have suggested that intron sequences are not appropriate for deeper divergences (Shapiro and Dumbacher 2001) due to alignment ambiguities resulting from multiple insertions and deletions (for general comments on alignments see Löytynoja and Goldman 2008). For example, Morgan-Richards et al. (2008) showed the alignment of β-*fibrinogen* intron 7 (which supports the controversial metaves/coronaves split) has no constant sites across the wide taxon sampling required for determining deep avian divergences. Introns are potentially well suited to resolving rapid radiations as they evolve fast enough to accumulate changes during this time (of divergence), while being slow enough to not become random and therefore lose signal (Daniels et al. 2002).

In general, the main avian orders found in both Hackett et al. (2008) and Cracraft (2001) are the same (e.g. Passeriformes, Psittaciformes, Cuculiformes etc.). However the relationships among the orders is different, with the deep branches of Neoaves receiving low support (less than 80% bootstrap support) in Hackett et al. (2008). One difference is that within "land birds". Within "land birds" Hackett et al. (2008) found Passeriformes (passerines) sister to Psittaciformes (parrots) and suggested a sister relationship between these and Falconidae. These groupings only have support when the intron data is included. Ericson et al. (2006) also inferred this relationship but only when all genes were combined, including β-*fibrinogen* intron 7 (see their Supplementary Material figures ESM-1-8). If correct, this placement would be very interesting, however intron alignment and/or long branch attraction may be a factor here. The long internal branch to passerines may be attracting the long internal branch of parrots (see Hackett et al. 2008, figure 3). At this point, support for such a grouping has not been found with mitochondrial DNA (e.g. Gibb et al. 2007; Brown et al. 2008). Cracraft (2001), by comparison included morphological, geographical and early molecular data in support of his groupings. We therefore feel it appropriate to use Cracraft (2001) as the basis for testing deep divergences within Neoavian birds rather than any one molecular dataset.

As a step towards increasing the taxon sampling of coding sequences, we add nine new mitochondrial genomes: the dollarbird (*Eurystomus orientalis*) and New Zealand kingfisher (*Halcyon sancta vagans*) as representatives of the Coraciiformes, together with the white-tailed trogon (*Trogon viridis*) from Trogoniformes are suggested to group with Piciformes; the great potoo (*Nyctibius grandis*) and Australian owlet-nightjar (*Aegotheles cristatus cristatus*) as representatives of Caprimulgiformes; barn owl (*Tyto alba*), expected to pair with the New Zealand owl (morepork, *Ninox novaeseelandiae*) to form Strigiformes; the roadrunner (a ground cuckoo, *Geococcyx california*) expected to pair with the New Zealand long-tailed cuckoo (*Eudynamys taitensis*) to form Cuculiformes; and the peach-faced lovebird (*Agapornis roseicollis*) from Psittaciformes expected to join with the budgerigar (*Melopsittacus undulate*s) and the ground parrot (kakapo, *Strigops habroptilus*). Thus we have reduced the number of long branches in our dataset by the addition of representatives from each of the six Neoavian lineages described in Cracraft (2001).

For the third subgroup of the 'Conglomerati' only a rail (takahe) and the kagu are published. It is unclear whether these really are a natural group, and this again leaves us with two long isolated branches. In accord with our previous practice, (Morgan-Richards et al. 2008) these species have temporarily been omitted until sequences from more closely related species are available for each of them.

We now change from the birds to the analysis. Perhaps the most fundamental problem occurring while reconstructing deep-level phylogeny is substitution saturation (Curole and Kocher 1999; Phillips et al. 2006). Phylogenetic signal can be eroded by factors including superimposed substitutions, and 'non historical' biases (such as from compositional non-stationarity) - which accumulate more rapidly at faster evolving sites. Attempts to limit these problems have been made in recent studies by identifying fast evolving sites at which signal erosion is expected to be high (Morgan-Richards et al. 2008).

In previous work (Delsuc et al. 2003; Phillips and Penny 2003; Phillips et al. 2004) we found standard RY-coding (Honeycutt and Adkins 1993), especially the third codon positions, to be advantageous for the most variable partitions of nucleotide data. This recoding both increases the proportion of changes on internal branches of the tree (that is, a 'treeness' measure) and decreases the differences in nucleotide composition (relative compositional variability, RCV). This latter is important in reducing nucleotide composition effects because they have been long known to bias tree reconstruction (Lockhart et al. 1992). Because of the better fit of the data to the model (higher treeness, and less variability in nucleotide composition) this has been our preferred method of analysis for vertebrate mitochondrial data.

Down-weighting the faster evolving sites, or grouping faster evolving nucleotides (or amino acids) into a single category has been quite widely used (see Honeycutt and Adkins 1993; Philippe et al. 2000; Jeffroy et al. 2006) - although the theoretical aspects have not been well developed in phylogenetics. Rodriguez-Ezpeleta et al. (2007) report that omitting the fastest evolving sites, grouping amino acids into functional categories, and some mixture models, all enhanced the phylogenetic signal for deeper divergences. Susko and Roger (2003) similarly report improvements from down-weighting.

However, some approaches may not be optimal if valuable sites are excluded simply because they are grouped under some prior definition (e.g. codon positions that have many fast evolving sites). Conversely, some saturated sites may be retained because they are in a category that, on average, does not have site saturation. Thus we can also group the justification for down-weighting into those using; '*a priori*' categories (such as third position, stems versus loops in RNA or amino acid groups), and 'conditional' categories (down-weighting of each site independently).

We have used both in the past; the RY-coding (*a priori* weighting category) (Phillips et al. 2004), and also a conditional down-weighting (Penny and Hendy 1986), based on the numbers of observed and expected incompatibilities. In general, all the methods mentioned here are examples of a standard statistical approach of 'noise reduction/signal enhancement' (Proakis et al. 2002). Here we implement a 'conditional' noise reduction technique in which the information retained from the sequence is determined on a site-by-site basis. The Materials and Methods section has more detail on this conditional recoding (down-weighting) of sites, an approach that we call site-stripping (Morgan-Richards et al. 2008).

Along with data partitioning/down-weighting and fossil calibrations (see Appendix 3.9.1) our additional sequences mean we can, in principle, calculate the probabilities that prior hypotheses are supported. In other words, we can calculate the proportion of trees that will have a split (or clade) that has been predicted. Unfortunately there appears to be little use in phylogenetics for specifying *priori* hypotheses, and then testing the probability of finding them with new data. Rather, results are treated somewhat '*post hoc*', looking at the trees after they are built, and then trying to explain the results. In principle, a Bayesian approach allows alternative hypotheses to be given different weightings, but it appears that a 'flat prior' is the norm; and this does not really differentiate between trees or hypotheses. Based on nuclear coding sequences, Lin et al. (2002) took the four-way split within eutherian mammals and calculated the probability of finding the same split from mitochondrial data. The result was certainly very highly significant; $P \approx 2.1 \times 10^{-7}$. It is very important in phylogeny, though perhaps seldom carried out, to give quantitative estimates of the increase in information from a phylogeny.

Penny et al. (1991) demonstrated that it is simple to calculate probabilities (see Appendix 3.9.2, and table 3-1) of a single pair of taxa predicted to come together on the tree. That is, there is one chance in $2n$-5 of them coming together 'by chance' on an unrooted tree, and one in $2n$-3 for a rooted tree (given n taxa). It is even more improbable that a predefined grouping of three or more taxa will come together, and here we develop measures in order to evaluate quantitatively the priors from Cracraft (2001).

Table 3-1 Probabilities of a predefined clade joining the tree as a single group, or as two subclades.

k Taxa in clade	Unrooted Trees Minimal	Rooted Trees Minimal
2	$1/(2n-5)$	$1/(2n-3)$
3	$3/(2n-5)(2n-7)$	$3/(2n-3)(2n-5)$
4	$15/(2n-5)(2n-7)(2n-9)$	$15/(2n-3)(2n-5)(2n-7)$
5	$105/(2n-5)(2n-7)(2n-9)(2n-11)$	$105/(2n-3)(2n-5)(2n-7)(2n-9)$

Our approach here is to infer evolutionary trees from complete mitochondrial genomes as a start on resolving the deep Neoavian splits. We investigate the robustness of the priors set out by Cracraft (2001) by using a novel site-stripping method and exploring networks (see Appendix 3.9.1. for fossil calibrations). In addition we assess quantitatively the usefulness of these priors by calculating clade probabilities. By finding resolution in the basal node of the Neoaves, hypotheses regarding the number of lineages present before the K – Pg boundary can be tested.

3.3 Materials and Methods

(a) Taxon Sampling

The dollarbird (*Eurystomus orientalis*) and barn owl (*Tyto alba*), were supplied from the Australian Museum, Sydney, under sample numbers EBU 11118 and EBU 2564 respectively. The roadrunner (*Geococcyx california*), great potoo (*Nyctibius grandis*) and the white-tailed trogon (*Trogon viridis*) were provided by the Louisiana State University Museum of Natural Science Collection of Genetic Resources under sample numbers LSUMZ B-8504, LSUMZ B-8954 and LSUMZ B-28495. The Australian owlet-nightjar (*Aegotheles cristatus cristatus*) was provided by Fritz Geiser, University of New England, Armidale, Australia.

The New Zealand long-tailed cuckoo (*Eudynamys taitensis*), and the New Zealand kingfisher (*Halcyon sancta vagans*) were obtained from Dick Gill, NZ Department of Conservation, Waikanae, New Zealand. The peach-faced lovebird (*Agapornis roseicollis*) was obtained locally from commercial breeders.

(b) Molecular Methods

Extractions of genomic DNA from each of the newly sampled birds were performed at the Allan Wilson Centre from 25 - 50 mg of liver tissue using the High Pure PCR Template Preparation Kit (Roche Applied Science, Mannheim, Germany) according to the manufacturer's instructions. To minimize the chance of obtaining nuclear copies of mitochondrial (mt) genes (NUMTS), 2 - 4 overlapping long range PCR fragments (3.5 - 12 kb in length) were first amplified using the Expand Long template PCR System (Roche Applied Science, Mannheim, Germany). The products were excised from 1% agarose gels and purified using QIAquick Gel Extraction Kit (Qiagen GmbH, Hilden, Germany) as per manufacturer's instructions. These long range products were subsequently used as template DNA for following short range PCR's (overlapping fragments 0.5 – 3 Kb in length). Short range primer combinations were found using our laboratory database as described in Slack et al. (2006) and any new primers required were designed using Oligo 4.03 (National Biosciences, Inc., Plymouth, MN). Sequencing was performed using BigDye Terminator Cycle Sequencing reagents according to the manufacturer's instructions (Applied Biosystems, Foster City, CA), then sequenced on an ABI 3730 automated sequencer (Applied Biosystems, Foster City, CA). Sequences were aligned using SEQUENCHER 4.7 (Gene Codes Corp. Ann Arbour, MI) then manually edited and checked for complete concurrence between overlapping sequences.

Where necessary (for example, with length heteroplasmy in control regions from microsatellite repeats) PCR products were cloned using the TOPO TA cloning kit for sequencing (Invitrogen). For each region at least three clones were sequenced to safeguard against PCR errors. In all cases, overlaps between sequences were sufficient to ensure synonymy and sequence identity was confirmed through BLAST searches (http://www.ncbi.nlm.nih.gov/blast/), confirmation of amino acid translation in coding regions, and alignment with other species.

In addition to the nine new bird mitochondrial genomes reported in this paper, 36 other complete avian mt genomes from NCBI GenBank were included in the analyses: 31 Neoaves and five Galloanserae. Paleognath taxa were not included in this dataset; although their overall placement is now well established (Gibb et al. 2007; Slack et al. 2007), there are still important but unresolved issues around the placement of tinamous (Harshmann et al. 2008; Phillips et al. in prep.). Instead we rooted our Neoaves trees with the Galloanserae sequences (Gibb et al. 2007; Morgan-Richards et al. 2008). The full data set is available from authors on request.

The Galloanserae taxa are: Japanese quail (*Coturnix japonica*; AP003195), magpie goose (*Anseranas semipalmata*; AY309455), redhead duck (*Aythya americana*; AF090337), greater white-fronted goose (*Anser albifrons*; AF363031), Australian brush-turkey (*Alectura lathami*; AY346091). The 31 Neoaves taxa (modern birds) are: rifleman (New Zealand wren, *Acanthisitta chloris*; AY325307), gray-headed broadbill (*Smithornis sharpei*; AF090340), fuscous flycatcher (*Cnemotriccus fuscatus*; AY596278), superb lyre bird (*Menura novaehollandiae*; AY542313), rook (*Corvus frugilegus*; Y18522), ivory billed toucan (*Pteroglossus azara*; DQ780882), pileated woodpecker (*Dryocopus pileatus*; DQ780879), morepork (a New Zealand owl, *Ninox novaeseelandiae*; AY309457), kakapo (flightless parrot *Strigops habroptilus*; AY309456), budgerigar (*Melopsittacus undulates*; EF450826), ruby-throated hummingbird (*Archilochus colubris*; EF532935), common swift (*Apus apus*; AM237310), peregrine falcon (*Falco peregrinus*; AF090338), forest falcon (*Micrastur gilvicollis*; DQ780881), Eurasian buzzard (*Buteo buteo*; AF380305), osprey (*Pandion haliaetus*; DQ780884), Blyth's hawk eagle (*Spizaetus alboniger*; AP008239), blackish oystercatcher (*Haematopus ater*; AY074886), ruddy turnstone (*Arenaria interpres*; AY074885), southern black-backed gull (*Larus dominicanus*; AY293619), red-throated loon (*Gavia stellata*; AY293618), little blue penguin (*Eudyptula minor*; AF362763), rockhopper penguin (*Eudyptes chrysocome*; AP009189), black-browed albatross (*Diomedea melanophris*; AY158677), Kerguelen petrel (*Pterodroma brevirostris*; AY158678), frigatebird (*Fregate sp;* AP009192), Australian pelican (*Pelecanus conspicillatus*; DQ780883), Australasian little grebe (*Tachybaptus novaehollandiae*; EF532936), greater flamingo (*Phoenicopterus ruber roseus*; EF532932), great crested grebe (*Podiceps cristatus*; AP009194) and the Oriental white stork (*Ciconia boyciana*; AB026193).

(c) Phylogenetic analysis

Sequences were aligned in SE-AL v2.0a11 at the amino acid level for protein-coding genes and based on secondary structure for RNA genes. The dataset has 12 protein-coding genes, two ribosomal RNAs (rRNA) and 22 transfer RNAs (tRNA). Gaps, ambiguous sites adjacent to gaps, the NADH6 (light-strand encoded), and stop codons (often incomplete in the DNA sequence), were excluded from the alignment. The 12 protein coding genes were separated into first, second and third codon positions while rRNA and tRNA genes were partitioned into stems (S) and loops (L). Protein-coding genes were checked for NUMTS by translating into amino acids.

Previous studies from birds (Slack et al. 2003), mammals (Lin et al. 2002) and simulations (Holland et al. 2003) have all shown the addition of outgroups can disrupt the ingroup tree. However in such cases (from theory and with simulated data) the ingroup tree (that is, with the outgroup omitted) is more likely to be correct (Holland et al. 2003). We therefore ran separate analyses either including or excluding the outgroup (five birds from the Galloanserae). A combined total of 13,412 nucleotides (excluding gaps) were used for the basis of further analyses (see below for number of characters per dataset). As mentioned earlier, we partitioned the data; codons 1 and 2; codon 3; RNA stems; and RNA loops (Slack et al. 2007) for site-stripping. Maximum likelihood (ML) analyses were carried out using standard programs including PAUP* 4.0b10 (Swofford 2001) and GARLI v0.95 (Zwickl 2006). Bayesian analysis was carried out in MRBAYES (Huelsenbeck and Ronquist 2001) and consensus networks were implemented in SPLITSTREE version 4 (Holland et al. 2004; Huson and Bryant 2006) and BEAST (Drummond et al. 2006; Drummond and Rambaut 2007). Optimal parameters for the ML models were determined using MODELTEST 3.7 (Posada and Crandall 1998) and the AIC values used. The hierarchical and AIC tests were in agreement for MODELTEST. Initial results from ML analyses were consistent with Bayesian analysis and were not used for site-stripping. The best model for the full dataset was GTR + I + G. Bayesian analyses used default parameters and were run for 10 million generations or until convergence was obtained. In addition to Bayesian posterior probabilities, we ran analyses in PHYML (Guindon and Gascuel 2003) and RAXML (Stamatakis 2006) to carry out 100 bootstrap replicates on the datasets both with and without the outgroup (see Appendix 3.9.3 table 3-2 for results).

(d) Noise reduction by down-weighting (Site-Stripping)

Site-stripping compares sites based on the actual number of mutations required on the tree ('tree steps') versus the maximum possible number of mutations for that site (Max). The calculation is the limit (L) = Tree steps*Tree steps/Max steps. If the threshold 'strictness' (s) is, for example, 4 then sites for which $L \leq 4$ remain unchanged, and sites for which $L > 4$ are RY-coded. If after RY-coding L is still > 4 the site is excluded. The higher the threshold (the larger the value of s), the more sites that are included and fewer are RY-coded. Conversely, the lower the threshold (lower values of s) the more sites RY-coded or excluded. Therefore, the weighting of each site is a function of that site, and is not predetermined by being a member of a class. This allows for example, some hypervariable sites from first and second positions to be RY-coded, or omitted. A range of s values were used, and resulted in the following:

$s = \infty$, 0 sites RY coded, 0 sites excluded (all 13 412 sites included),
$s = 6.0$, 20 sites RY coded, 0 sites excluded,
$s = 4.5$, 91 sites RY coded, 1 site excluded,
$s = 3.5$, 254 sites RY coded, 8 sites excluded,
$s = 2.28$, 578 sites RY coded, 63 sites excluded,
$s = 2.0$, 891 sites RY coded, 159 sites excluded.

Bayesian inference analyses were carried out on each of the data matrices, including the fully weighted data. Note that RY-coding increases the ML scores, as it amalgamates some nucleotide categories, thus the data is now different and it is not valid to compare directly ML scores from RY and nucleotide coding (Steel MA, personal communication). Similarly, consistency index (CI) values are not directly comparable between nucleotide and RY-coded datasets. That is, for an unrooted 4-taxon tree, for which A, C, T, G states are random (i.e. no signal remains), the expected average CI for nucleotide data is 0.949, while for the RY-coded data it is 0.778.

In order to identify fast-evolving sites and so facilitate noise reduction by site-stripping, nine additional close relatives were added to the alignment, but were removed before phylogenetic analysis as they don't break up long branches or add phylogenetic signal and would only increase analysis times.

The nine taxa included were: chicken (*Gallus gallus*; AP003317), green junglefowl (*Gallus varius*; AP003324), gray junglefowl (*Gallus sonneratii*; AP006741), white stork (*Ciconia ciconia*; AB026818), Canadian goose (*Branta canadensis*; DQ019124), tundra swan (*Cygnus columbianus*; DQ083161), mountain hawkeagle (*Nisaetus nipalensis*; AP008238), Pacific loon (*Gavia pacifica*; AP009190), American kestrel (*Falco sparverius*; DQ780880) (see Appendix 3.9.4 figure3.6 for consensus tree of the above nine taxa and their close relatives).

(e) Probabilities of observing predefined clades

It is straightforward to calculate the probability of observing a pre-specified clade in a tree from a new dataset, that is, a tree using additional data not used to predict the clade. In general, for n taxa there are $B(n) = (2n-5)!!$ unrooted binary trees, where the double factorial notation (!!) is the product of every second number, that is, $1 \times 3 \times 5 \times \ldots \times 2n-5$ (table 3-1). Thus, for example, there are $\approx 3 \times 10^{20}$ possible unrooted binary trees for 20 taxa. In addition, even if a predicted clade for three or more taxa ends up as two subclades on the tree, we can calculate the probability of observing this and it can indicate that there is still high information content in the data. The calculation can be extended further to three or more subclades, but here we concentrate mainly on our prior clades being found. We need to define the composition of the prior clades carefully; they alter with the question being considered. For example, we may be interested in whether the three parrots do form a natural group, in which case we would calculate the probabilities of three parrots coming together with $n = 40$ taxa in the data set – this is close to a ‘flat prior’. Alternatively, we may accept the grouping within orders and consider just $n = 21$ deep Neoavian groups (figure 3.2, Appendix 3.9.2 for details); this is testing the groupings (‘priors’) of Cracraft (2001) but other tests are also possible.

3.4 RESULTS

The first of our three approaches to improving the Neoavian tree was the inclusion of more sequences. The nine new mitochondrial genome sequences are deposited in GenBank under the following accession numbers: dollarbird (EU344978, 17,774bp), barn owl (EU410491, more than 16,148bp, incomplete ND6 and control region [CR]), roadrunner (EU410488, 17,091bp), great potoo (EU344977, more than 14,396bp, incomplete ND6 and CR), white-tailed trogon (EU410490, 17,751bp), Australian owlet-nightjar (EU344979, 18,607bp), peach-faced lovebird (EU410486, 16,732bp), New Zealand long-tailed cuckoo (EU410487, 17,559bp), and the New Zealand kingfisher (EU410489, 17,549bp). Following the gene-order nomenclature from Gibb et al. (2007), the cuckoo and roadrunner were found to have the remnant CR(2) gene order first described in the falcon (Mindell et al. 1999). All other birds have the standard avian gene order as in the chicken (Desjardins and Morais 1990) with the possible exception of the great potoo and barn owl; because their sequences are currently incomplete in the CR and adjacent genes, their gene order is unknown.

Site-stripping (noise reduction) was our second approach to improving the tree. Bayesian analyses were carried out on a range of down-weighting values, initially excluding the outgroup, then including it. Figure 3.1 shows the result for the unrooted Neoavian dataset with a high threshold (minimum down-weighting, strictness $s = 6.0$), while figure 3.2 shows results from the maximum down-weighting used (a low threshold, strictness $s = 2.0$). Both figures are networks showing splits occurring in at least 25% of Bayesian phylogenies (Holland et al. 2004).

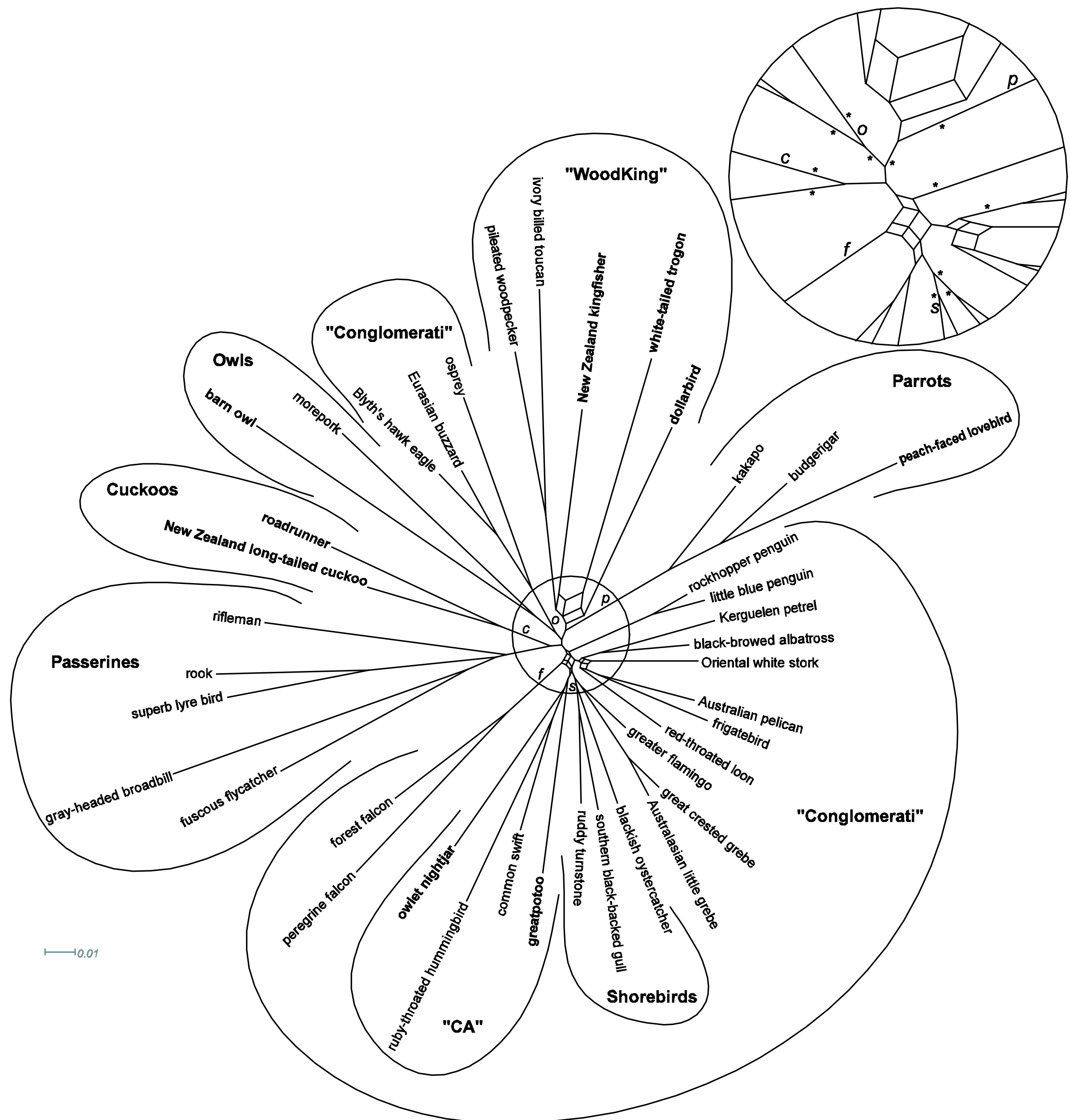


Figure 3.1 Unrooted Bayesian consensus network with minimal down-weighting

Consensus network of Neoaves (modern birds), based on whole mtDNA genomes with <u>minimal</u> down-weighting (threshold s = 6.0; 20 sites RY-coded, 0 sites excluded). Only splits occurring in more than 25% of trees are included in the network. Insert shows the central portion (indicated by the circle) expanded. Selected branches are labelled to ease comparison; c = cuckoos, f = falcons; o = owls, p = parrots and s = shorebirds. New taxa included in are highlighted in bold. Splits indicated by asterisk have 99+ Bayesian posterior support (BPP).

Our third approach was a quantitative test of predefined groupings. There are several clades that were predicted and were returned both with different down-weightings and with or without the outgroup. Perhaps the most straightforward example is that of the three parrots; this is just a trial calculation because we really had no doubt that the parrots would come together, as predicted from previous DNA sequence analyses (de Kloet and de Kloet 2005). The New Zealand ground parrot (kakapo) was the first to diverge. This is a trial calculation with 'flat priors', and the probability of three taxa coming together on a tree is P = 0.0005 ($P = 3/(2n-5)(2n-7)$, with $n = 40$ (see Appendix 3.9.2.). There is considerable rate variation within parrots, with the kakapo being slower than the others, while the peach-faced lovebird is the fastest (which is evident by the long edge in both figures 3.1 and 3.2).

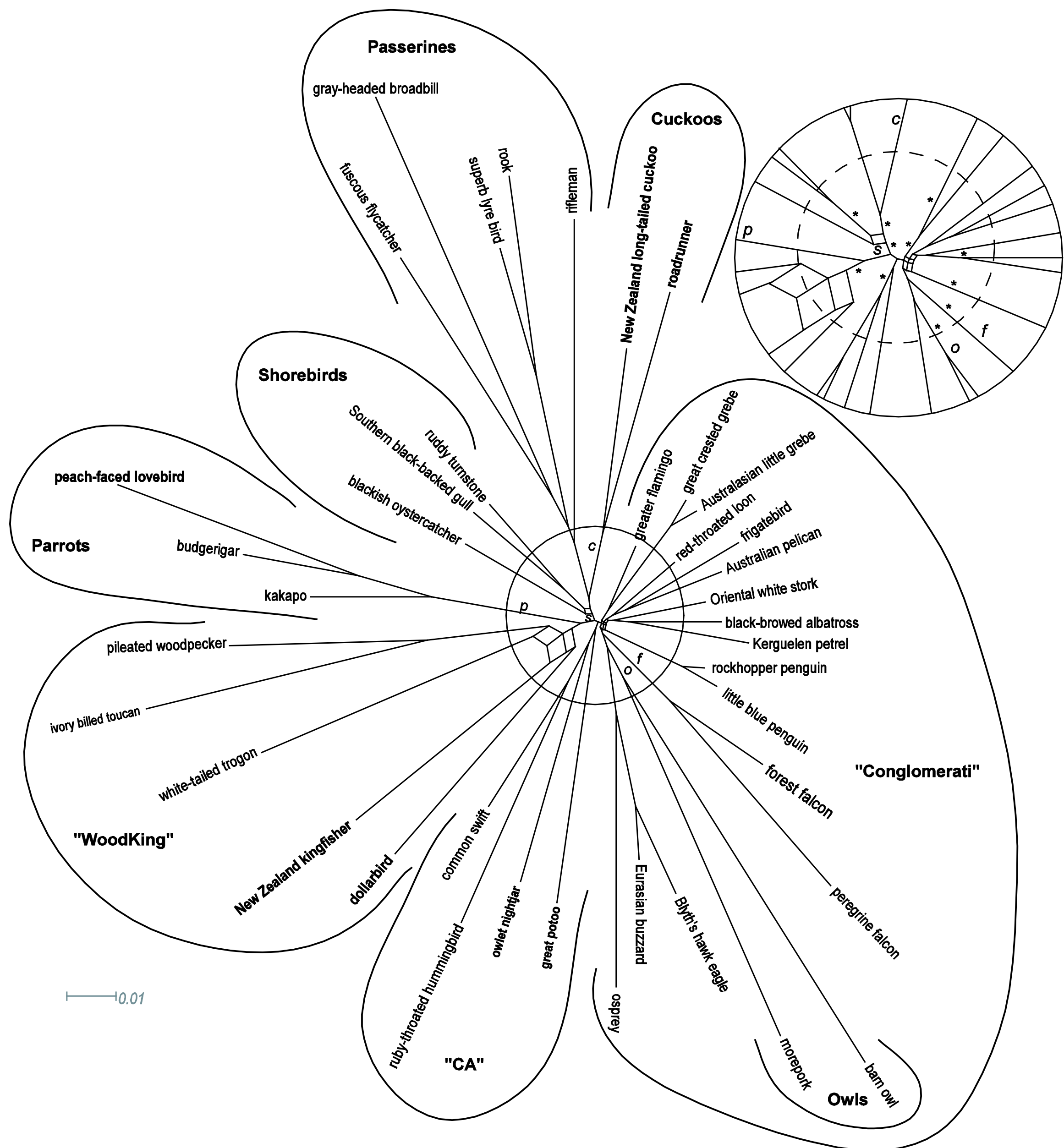


Figure 3.2 Unrooted Bayesian consensus network with maximum down-weighting

Consensus network of Neoaves with the maximum down-weighting used (threshold s = 2.0, 891 sites RY-coded, 159 sites excluded). Only splits occurring in more than 25% of trees are included in the network. Insert shows the central portion (indicated by the circle) expanded. Dashed line indicates 21 independent groups (see later and Appendix 3.9.2 for detail). Selected branches are labelled to ease comparison; c = cuckoos, f = falcons; o = owls, p = parrots and s = shorebirds. New taxa included in are highlighted in bold. Splits indicated by asterisk have 99+ BPP.

Again as expected from our informal priors, the New Zealand long-tailed cuckoo and the roadrunner (a ground cuckoo) always paired and do in prior morphological and molecular studies, thus we considered them to form one independent 'taxonomic group'. Of the 40 taxa analyzed, we consider only 21 groups to be independent (see later and Appendix 3.9.2 for details of the groupings), consequently the probability of two taxa forming a clade in the tree is P = 0.027 (P = 1/(2n-5), with n = 21 see Appendix 3.9.2). Perhaps unexpectedly, cuckoos then group as sister to the five passerines in our dataset, a result also observed by Mayr et al. (2003) using combined molecular and morphological data however, bootstrap support was relatively low (see their figure 5 and 7). As in previous analyses the passerines always group together, with the New Zealand wrens (rifleman in this case), basal to the oscines and suboscines. The cuckoo/passerine pairing was found with all down-weightings, both with the ingroup alone and with the outgroup included (figure 3.3). This grouping of the cuckoo/passerine clades is interesting and requires testing with both nuclear coding sequences and rare genomic changes (Boore 2006).

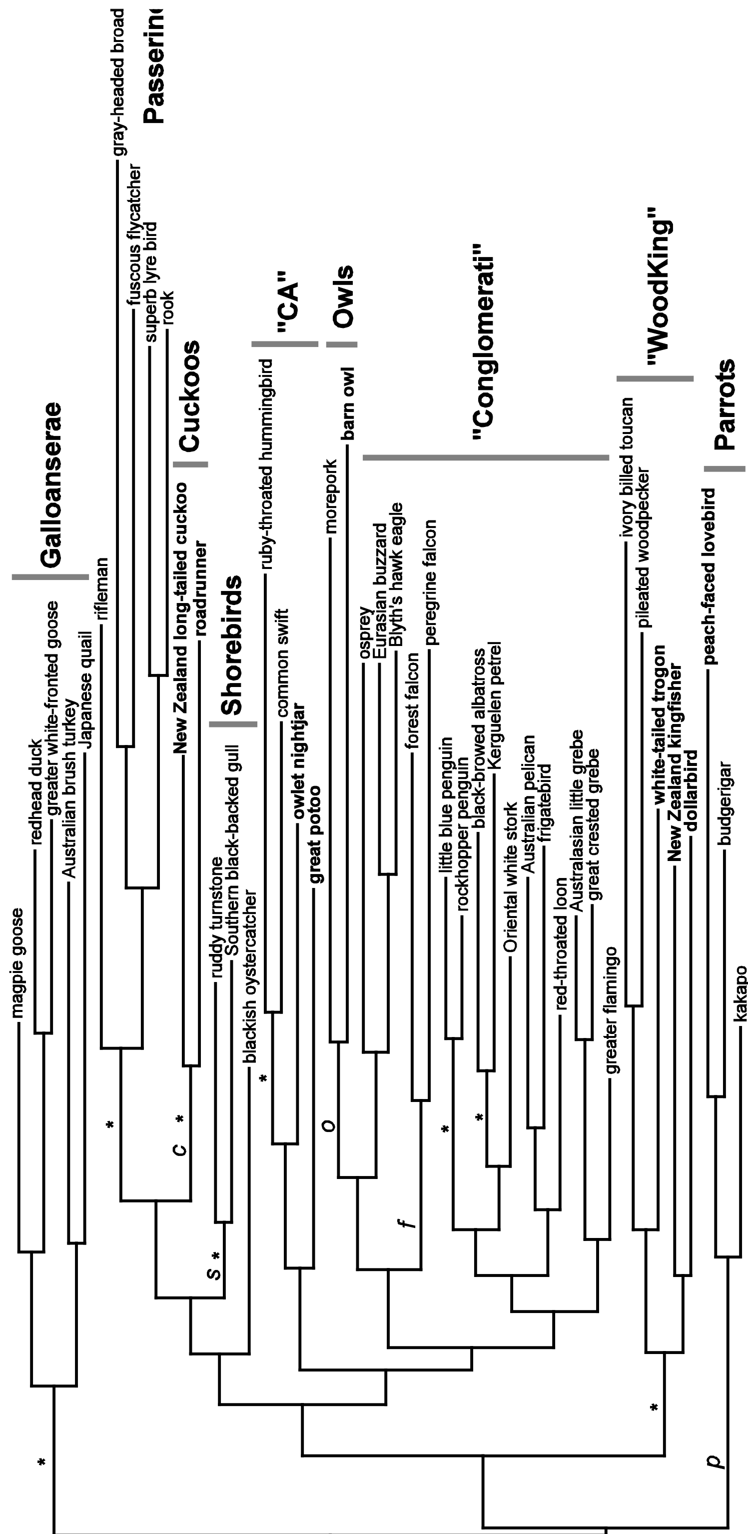


Figure 3.3 Rooted Bayesian phylogram with maximum down-weighting

Bayesian phylogram of Neoaves with maximum down-weighting and including the five Galloanserae as the outgroup. Only splits occurring in more than 25% of trees are included in the network. Selected branches are labelled to ease comparison; c = cuckoos, f = falcons; o = owls, p = parrots and s = shorebirds. New taxa included in are highlighted in bold. Splits indicated by asterisk have 99+ BPP.

The first real test of Cracraft's (2001) priors, that is, testing groups above the Order level, stems from the clade we refer to as 'WoodKing'. In this case, all five taxa were always found as a clade, irrespective of the down-weighting, and with or without the outgroup (that is, with both unrooted and rooted trees). As expected, the Piciformes (pileated woodpecker and ivory billed toucan), were always paired. Thus if it is assumed that the woodpecker and toucan are sufficiently close, then really there are only four independent taxa, and $P = 7.12 \times 10^{-6}$, is a highly significant result (see Appendix 3.9.2. for details). Note that the calculation of the probabilities allows all possible ways of observing the four taxa on the tree, including any paraphyletic groups within it. As shown in figures 3.1 and 3.2, there is conflicting signal linking the kingfisher with either the white-tailed trogon or with the dollarbird. Being able to show both signals is a major advantage of networks (Holland et al. 2004) because it helps to prevent premature conclusions. With increased down-weighting of the faster sites (a stricter threshold), we again observed variation in the position of the dollarbird which tended to be deeper in the clade, though still within the 'WoodKing' group). Additional taxon sampling should resolve the splits fully, but our main conclusion is that the predicted grouping of Piciformes, Trogoniformes and Coraciiformes (Cracraft 2001) is found (though not necessarily reciprocally monophyletic). Hackett et al. (2008) sampled more widely and found strong support for Coraciiformes + Piciformes (see their figure 2, clade C), although Trogoniformes fell outside this and had less support (see Appendix 3.9.3 table 3-2).

However, the next result was not in our informal priors. We found that the parrots and the 'WoodKing' group, irrespective of down-weighting extent, are always adjacent clades on the unrooted tree (supported by high BPP values but not by bootstrapping see Appendix 3.9.3 table 3-2). This result will need further investigation as the parrot lineage has considerable rate variation and there is a long internal branch from the three parrots to the rest of the tree. Because the grouping was not part of our priors we cannot calculate the increase in support, but if additional data types support this relationship then the probabilities could be calculated.

Next we consider the 'SCA' group (Strigiformes, Caprimulgiformes, and Apodiformes) predicted by Cracraft (2001). Four of the six taxa available form a monophyletic clade, to the exclusion of the two owls. The great potoo, Australian owlet-nightjar, common swift and ruby-throated hummingbird, formed a group of four ('CA'). The swift and hummingbird pairing was highly supported by BPP and bootstrapping (see Appendix 3.9.3 table 3-2) as predicted from previous studies (Johansson et al. 2001; van Tuinen and Hedges 2001; Barker et al. 2004; Harshmann et al. 2008; Morgan-Richards et al. 2008); the Australian owlet-nightjar came deeper, and finally the great potoo, which was always basal. It should be noted that the potoo did move slightly with bootstrapping and we found more support for the Apodiformes and the owlet nightjar to the exclusion of the potoo (see Appendix 3.9.3 table 3-2)(see also Mayr and Provine 1980; Barrowclough et al. 2006). Hackett et al. (2008) also observed high support (98%) for Apodiformes + *Aegotheles,* as have previous nuclear and morphological studies (see also Mayr and Provine 1980; Barrowclough et al. 2006). The position of this group of four was variable in the tree. With lesser down-weighting (a higher threshold, s = 6.0) the group was found within the informal 'Conglomerati' group (figure 3.1), but with the maximum down-weighting it was outside this group (apart from the shorebirds). Similarly, the barn owl and the morepork always paired, joining together quite deep in the tree (that is, although both are 'owls' they represent old divergences; indeed some short preliminary runs did not even join them together; see Appendix 3.9.3 table 3-2 for support values). We need to be cautious here as both owls have some of the highest rates of sequence evolution among the Neoaves. Although the owls did not group with the other members of the 'SCA' clade, they were always found to group with the buzzard, hawk eagle (Accipitridae) and osprey. Only with the highest down-weighting used (figure 3.2 and 3.3) did the falcons unite with the owls/Accipitridae/osprey; with the falcons basal. However we found no bootstrap support and low BPP support for this grouping (see Appendix 3.9.3 table 3-2). Grouping of the Accipitridae, osprey and owls is certainly interesting. Relationships within the birds of prey are controversial as speculation over convergence and raptorial specialisation have been raised (e.g. Livezey and Zusi 2007). This relationship was not found by Hackett et al (2008) and needs to be tested with nuclear coding data. We return to the raptors and owl question later.

Even though the proposed 'SCA' clade came out as two groups in the tree, this in itself still has high information content. The simplest calculation is to assume the alternative prediction of the two owls being independent of the other four, and in this case the probability of observing the "2 plus 4" grouping is simply the product of the probability of observing a pair ($P = 1/(2n\text{-}5) = 0.013$ for $n = 40$) and the probability of observing a group of four ($P = 2.5 \times 10^{-6}$, see Appendix 3.9.2). This gives the combined probability of 3.4×10^{-8} (about 1 chance in 300 million of observing the pattern). Strictly speaking, we need not make the assumption that the two owls (in particular) separate from the other four taxa, there are a total of 21 combinations (${}^6C_2$) of pairs from six taxa; only one has the two owls paired. In general we would multiply the probability by 21 (see Appendix 3.9.2), still giving around one chance in ten million of observing these two groups on the tree. Similarly, it could have been just a single taxon (six choices) that separated from the other five, or two groups of 3 taxa (10, or ${}^6C_3$ combinations), but halved because each triplet of taxa is found twice. Appendix 3.9.2 and figure 3.4 show the general calculation, but in this case (because the owls join with the raptors as an alternative prediction see Mayr and Provine 1980), it is reasonable to use the probability of finding just the "two plus four" grouping.

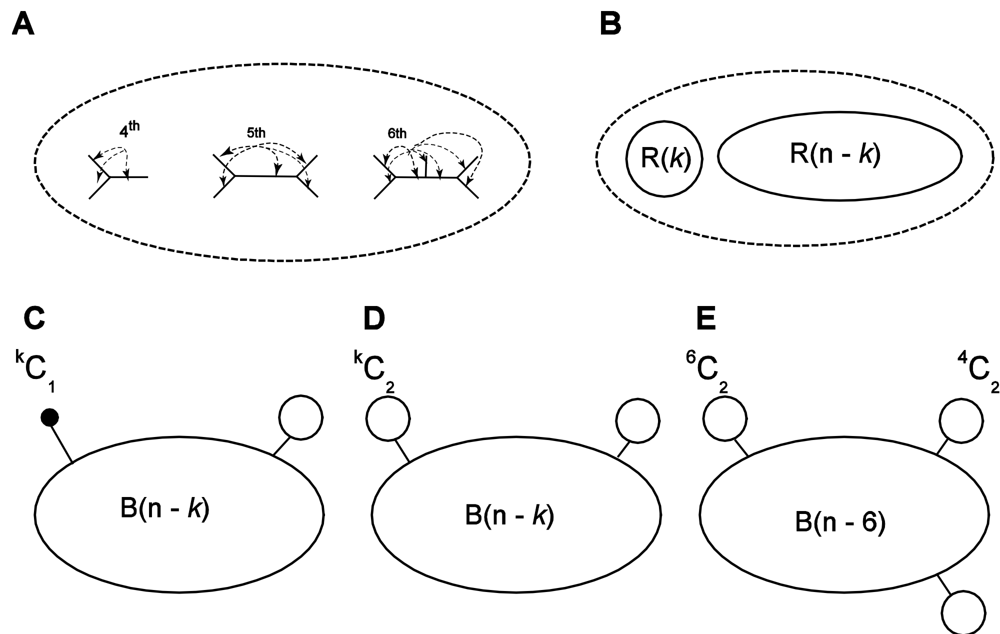


Figure 3.4 Probability of finding predefined clades on a tree

The basis for calculations for the probability of finding predefined clades (or subclades) on a tree. A. A procedure for counting the number of trees. There is only one unrooted tree for three taxa, and there are three edges ('branches') to add the fourth taxon – giving three trees for four taxa. Each of these three trees has 5 edges where the 5th taxon can be added giving $1 \times 3 \times 5 = 15$ trees. Similarly, each of these 15 trees has 7 edges for the 6th taxon to be added, leading to the formula $B(n) = (2n - 5)!!$, for the number of unrooted binary trees. B. Calculating the probability of pre-specified clade of k taxa on a new tree. There are $R(k)$ rooted trees for the clade of k taxa and $B(n - k)$ for the remaining n-k taxa, leading to the calculation for the probability of observing a pre-specified clade of k taxa forming a clade in each tree with n taxa. C. The proposed grouping of *k* taxa with two subgroups; a single taxon in one and k-1 in the other. D. A similar case with 2 taxa on one group and *k*-2 in the other. There are ${}^{k}C_2$ (*k* choose 2) ways of selecting the two taxa. E. An example where the group of $k = 6$ ends up as m = 3 subgroups of 2 taxa each. There are ${}^{6}C_2$ for selecting the first pair of taxa, and ${}^{4}C_2$ for the second pair, and 3! ways for ordering the three pairs on a given tree.

Our conclusion at this point is that excellent progress is being made in understanding the deeper levels of phylogeny of the Neoaves. If the predictions from Cracraft (2001) are, in general, being well supported with new data, then this implies that the basal polytomy is resolvable. However, our next step is to check that there are no major changes when the outgroup is added; this has been a major problem when a smaller number (24 ingroup) of taxa were sequenced (see Harrison et al. 2004).

.

Figure 3.3 shows our tree rooted with five Galloanserae taxa for the maximum down-weighting value (threshold s = 2.0). The two main points from this figure are that

(1) there is only one local change to the unrooted tree when the root is added, and

(2) the root comes between parrots and all other Neoaves.

The first point refers to the difference in the position of the flamingo/grebe clade between the unrooted and rooted trees. In the unrooted tree they are basal to the loon/albatross/Pelecaniformes/stork grouping. In the rooted tree we find them basal to the same group as before, plus the Accipitridae/osprey/owl/falcon clade. In other words, the tree is 'locally stable' in the terminology of Cooper and Penny (1997). Finding just a single edge (branch) different between two trees is very highly significant, about 5.7×10^{-54} for 40 taxa in the ingroup (Penny et al. 1982).

On the surface, this latter finding (the root between parrots and other Neoaves) could be suspicious because the branch at the base of the parrots is the longest internal branch on the tree! However, with lesser down-weighting (s = 6.0), the root joins one step away on the short branch at the base of the parrot/'WoodKing' group. The movement of the root from a shorter branch is not expected unless parrots are really the first subdivision of Neoaves. Without any prior information as to the root of the Neoaves, it is not possible to give any quantitative statement of confidence in this rooting.

Other aspects of the rooted tree are also interesting and noteworthy. We again find the same owl/raptor clade appearing only with the strongest down-weighting – that is, only the Accipitridae/osprey (but not falcons) unite with the owls with lesser down-weighting. The tree produced from BEAST resulted in the falcons joining the 'CA' group (Appendix 3.9.1 figure 3.5). This latter observation, if real, is interesting as both falcons and the 'CA' taxa are inflight foragers typically specialized for bill-capture of prey; in comparison, the owls are specialized for inflight talon capture (like hawks). However there was only conflicting support for this relationship shown in the network figure 3.1 and therefore we are unable to comment further without additional data.

It is worth noting that some of the deeper groupings within this large raptor/shore bird/water bird ('Conglomerati') still vary somewhat depending on taxon sampling (Morgan-Richards et al. 2008), and it is not clear yet whether further taxon sampling will resolve these issues. It may well be that the three subgroups in the 'Conglomerati' (group (vi) of Cracraft 2001; see chapter introduction) should be considered independently.

3.5 DISCUSSION

Resolving the evolutionary relationships within the modern birds (Neoaves) has been both problematic and controversial, with some suggesting that it will never be fully resolved (Poe and Chubb 2004). Here we have shown that with more and longer DNA coding sequences, along with improved noise-reduction techniques, relationships within Neoaves are expected to be resolvable. This should occur relatively quickly with the addition of data from nuclear coding and rare genomic changes as they become available. Our approach to resolving this issue has been threefold: presenting additional data; improving noise reduction/signal enhancement techniques; and getting beyond 'flat priors', where it is assumed (sometimes correctly!) that there is no useful prior knowledge.

We consider it important that prior hypotheses can be evaluated quantitatively, and thus the formulae developed in Appendix 3.9.2 will be useful for a wide range of studies. However, there is still more work required developing these analyses. For example, the calculation for pre-specified groups is for the optimal placement of that clade on the tree (even if the bootstrap or Bayesian posterior priors are less than 100% support). If these support values are indeed higher, then this gives even more confidence in the clade(s), so in that respect our probabilities are conservative. Thus more thought is required on how to combine the calculations developed here with the strength of support for branches in the tree from new data. Similarly, the calculation allows for any subtree within the clades (or subclades). However, if we pre-specified that a particular grouping and subtree is expected, then the probability of finding this arrangement on the tree is even lower. For example, the calculation allows 15 ways (5!!) that a group of four could join the larger tree. This is because there are three unrooted trees for four taxa, each with 5 edges for joining to the rest of the tree – and the prediction does not specify which of the 3 × 5 (15) trees would be observed.

In contrast, if we predict precisely how the group will join (forming a clade) then the number of possible trees is reduced 15-fold. Overall, it is important that we make better use of well-considered prior hypotheses when studying trees based on new data. Even though we are a long way from having the 'one-tree' (in this case) for Neoaves, we can be confident that the issue is resolving, and that the datasets have, in a formal sense, high information content.

A related question is estimating how many trees within Neoaves are still likely – a 'confidence set' of trees. At this higher taxonomic level, it is not yet clear which groupings are stable, and which may be subject to change. For our Neoaves dataset, there appear to be around 21 major groupings (shown as crossing into the inner dashed line in the insert of figure 3.2 and in Appendix 3.9.2). These groupings include; Passeriformes, cuckoos, parrots, two shore bird lineages, three raptor clades (falcons, buzzard/osprey and owls), flamingoes, grebes, Pelecaniformes, tube-noses, storks, penguins, loon rollers, kingfishers, woodpecker/toucan/trogon, potoos, owlet-nightjars, and hummingbirds/swifts.

In principle, there are $R(21) \approx 3 \times 10^{23}$ possible rooted binary trees, of which only a vanishingly small proportion are realistic. With eutherian mammals, there were initially 19 orders identified and therefore $R(19) \approx 2 \times 10^{20}$ possible rooted binary trees. But it quickly became apparent that no more than about 10^2 trees were likely (Lin et al. 2002) - an improvement of 18 orders of magnitude. The next step for birds is an equivalent analysis for Neoaves, and thus getting beyond the debilitating view of the flat priors – that all trees were equally likely, that there is no information in previous studies.

The results from down-weighting the faster evolving sites are interesting, and these techniques need to be developed and tested further. With increased down-weighting we find closer agreement between earlier predictions and the actual tree found. From first principles we expect that reducing the influence of the saturated sites will help, and in general it appears that the predefined groups are found more strongly. It is for this type of reason that we would like to see further development and evaluation of the noise-reduction techniques including their application to nuclear coding data. Although it is outside the range of this study, a simulation study is now an important next step.

Turning now from the more general issues, to the Neoaves in particular, our current study assesses the stability and probability of the six groups proposed by Cracraft (2001) using a novel analysis method to down-weight sites of whole mitochondrial genome sequences. These Cracraftian priors were found to be robust with four of the six groups within Neoaves being recovered, the other two having relatively small changes; the owls moving to the raptors, and possibily that the raptor/water carnivores ('Conglomerati') may be diphyletic. In this latter respect, it appears preferable, at least in the short term, to treat the three subgroups of Cracraft's group (vi) independently. Our resulting phylogenies appear relatively stable, differing little in overall topology both with and without the addition of the outgroup (Galloanserae).

The placement of the root of Neoaves needs additional support. Our analyses put the root in one of two possible locations; either with parrots (which have a higher mutation rate in mtDNA) as the most basal lineage, or, with lesser down-weighting, the parrots plus the 'WoodKing' grouping as the basal clade. Because the separation of parrots from the rest of the Neoaves occurred with the highest down-weighting, we cannot easily dismiss this possible rooting. Morphologically, parrots are distinct (see Dyke and van Tuinen 2004; Waterhouse 2006 for review) and a fragment of a mandible from the Maastrichtian, latest Cretaceous (65 - 70 Ma Lance Formation, North America) has been described (Stidham 1998). The identification of this fossil is contentious (Mayr and Provine 1980; Dyke and van Tuinen 2004), though previous molecular work suggests a Cretaceous diversification for each of the African, Australian and South American parrots (Miyaki et al. 1998). Dating carried out on our current dataset suggests that the most basal parrot in our analysis, the Kakapo, split from the other parrots sometime after the K - Pg boundary. The lineage as a whole however pre-dates the K - Pg boundary with a mean date of *ca* 85 Ma (see Appendix 3.9.1 figure 3.5). Hackett et al. (2008) suggested that the root of Neoaves be placed with the sister grouping of the Gruiformes/Caprimulgiformes/Apodiformes to Podicipediformes/Phoenicopteriformes/Phaethontidae/Pteroclididae/Mesitornithidae/Colum -biformes (see their figure 2, clades K, L, M and N). However they do not give support values for this lineage, and they state that their rooted tree only occurs when the β–*fibrinogen* intron 7 data is included.

We have already demonstrated (Morgan-Richards et al. 2008) that by our more rigorous standards (deleting columns around gaps back to a constant column) the intron sequences of this locus are not informative for deep divergences. Clearly the root of Neoaves is still under debate however we now have a number of possibilities to be tested by future analyses.

3.6 Conclusion

The basal split within Neoaves and its timing are resolvable issues. If modern birds radiated over a short period (say 2 - 5 Ma) after the K - Pg extinction then it will be very difficult to resolve the polytomy at the base of modern birds. However, using whole mitochondrial DNA coding sequences gives us a solid point from which to build. By the addition of more taxa, nuclear coding sequences and rare genomic changes, we expect resolution at the ordinal level to be achievable. In addition, the further development of noise reduction techniques for coding sequences (both organellar and nuclear) will enable more robust trees to be produced. We estimate at least 12 Neoavian lineages had evolved prior to the K - Pg boundary, similarly, van Tuinen, Stidham, and Hadly (2006) and Brown et al. (2008) support pre-K - Pg origins for multiple modern lineages. In addition, Clarke et al. (2004) estimate a minimum of five Anseriformes lineages (duck, chicken and ratite bird relatives) before this time, supporting the presence of a diverse array of modern bird lineages prior to the extinction event. Lastly, with regards to the search for the ‘one tree’, we feel our data has made significant progress with support for four of the six Cracraftian groups. Given the very low probability of observing groupings by chance, the data is highly informative and should stimulate future work incorporating data from all facets of avian evolution.

3.7 Acknowledgements

We thank the Australian Museum, Sydney; the Louisiana State University Museum of Natural Science Collection of Genetic Resources; Fritz Geiser, University of New England, Armidale, Australia; and Dick Gill, Department of Conservation, Waikanae, for supplying tissue samples. Thanks to Trish McLenachan for help with alignments and sequences, to Olga Kardailsky and Steve Trewick for additional mitochondrial genomes.

Thank you to Simon Hills and Patrick Biggs for computer time and assistance with BEAST analyses. Work carried out at the Allan Wilson Centre for Molecular Ecology and Evolution was supported by the Marsden Fund (Royal Society of New Zealand) to D.P. The manuscript was improved by constructive comments from Associate Editor M. Gouy and three anonymous referees.

3.8. REFERENCES

Barker, F. K., A. Cibois, P. Schikler, J. Feinstein, and J. Cracraft. 2004. Phylogeney and diversification of the largest avian radiation. PNAS **101**:11040-11045.

Barrowclough, G. F., J. G. Groth, and L. A. Mertz. 2006. The RAG-1 exon in the avian order Caprimulgiformes: Phylogeny, heterozygosity and base composition. Molecular Phylogenetic and Evolution **41**:238-248.

Boore, J. L. 2006. The use of genome-level characters for phylogenetic reconstruction. Trends in Ecology & Evolution **21**:439-446.

Brown, J. W., R. B. Payne, and D. P. Mindell. 2007. Nuclear DNA does not reconcile 'rocks' and 'clocks' in Neoaves: a comment on Ericson *et al.* Biology Letters **5**.

Brown, J. W., J. S. Rest, J. García-Moreno, M. D. Sorenson, and D. P. Mindell. 2008. Strong mitochondrial DNA support for a Cretaceous origin of modern avian lineages. BMC Biology.

Chubb, A. L. 2004. New nuclear evidence for the oldest divergence among neognath birds: the phylogenetic utility of ZENK (i). Mol. Phyl. Evol. **30**:140-151.

Clarke, J. A., C. P. Tambussi, J. I. Noriega, G. M. Erikson, and R. A. Ketcham. 2004. Definitive fossil evidence for the extant avian radiation in the Cretaceous. Nature **433**:305-308.

Cooper, A., and D. Penny. 1997. Mass survival of birds across the Cretaceous-Tertiary bounday: Molecular evidence. Science **275**:1109-1113.

Cracraft, J. 2001. Avian evolution, Gondwana biogeography and the Cretaceous-Tertiary mass extinction event. The Royal Society of London, B **268**:459-469.

Curole, J. P., and T. D. Kocher. 1999. Mitogenomics: digging deeper with complete mitochondrial genomes. Trends in Ecology & Evolution **14**:394-398.

Daniels, S. R., B. A. Stewart, G. Gouws, M. Cunningham, and C. A. Matthee. 2002. Phylogenetic relationships of the southern African freshwater crab fauna (Decapoda: Potamonautidae: *Potamonautes*) derived from multiple data sets reveal biogeographic patterning. Molecular Phylogenetics and Evolution **25**:511-523.

de Kloet, R. S., and S. R. de Kloet. 2005. The evolution of the spindlin gene in birds: Sequence analysis of an intron of the spindlin W and Z gene reveals four major divisions of the Psittaciformes. Mol. Phylog. Evol. **36**:706-721.

Delsuc, F., M. J. Phillips, and D. Penny. 2003. Comment on "Hexapod Origins: monophyletic or paraphyletic?" Science **301**:1482d-e.

Desjardins, P., and R. Morais. 1990. Sequence and gene organization of the chicken mitochondrial genome. Journal of Molecular Evolution **32**:153-161.

Drummond, A. J., S. Y. W. Ho, M. J. Phillips, and A. Rambaut. 2006. Relaxed Phylogenetics and Dating with Confidence. PLoS Biology **4**:699-710.

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

Dyke, G. J., and M. van Tuinen. 2004. The evolutionary radiation of modern birds (Neornithes): reconciling molecules, morphology and the fossil record. Zoological Journal of the Linnean Society **141**:153-177.

Ericson, P. G. P., C. L. Anderson, T. Britton, A. Elzanowski, U. S. Johansson, M. Källersjö, J. I. Ohlson, T. J. Parsons, D. Zuccon, and G. Mayr. 2006. Diversification of Neoaves: integration of molecular sequence data and fossils. Biol. Lett. 5: (in press) **4**:543-547.

Feduccia, A. 2003. "Big bang" for Tertiary birds? Trends in Ecology & Evolution **8**:172-176.

García-Moreno, J., and D. P. Mindell. 2000. Using homologous genes on opposite sex chromosomes (gametologs) in phylogenetic analysis: a case study with avian CHD. Mol. Biol. Evol **17**:1826-1832.

Gibb, G. C., O. Kardailsky, R. T. Kimball, E. Braun, and D. Penny. 2007. Mitochondrial genomes and avian phylogeny: complex characters and resolvability without explosive radiations. Mol. Biol. Evol **24**:269-280.

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

Harrison, G. L., P. A. McLenachan, M. J. Phillips, K. E. Slack, A. Cooper, and D. Penny. 2004. Four new avian mitochondrial genomes help get to basic evolutionary questions in the Late Cretaceous. Molecular Biology and Evolution **21**:974-983.

Harshmann, J., E. L. Braun, M. J. Braun, C. J. Huddleston, R. C. K. Bowie, J. L. Chojnowski, S. J. Hackett, K.-L. Han, R. T. Kimball, B. D. Marks, K. J. Miglia, W. S. Moore, S. Reddy, F. H. Sheldon, D. W. Steadman, S. J. Steppan, C. C. Witt, and T. Yuri. 2008. Phylogenomic evidence for multiple losses of flight in ratite birds. PNAS **105**:13462-13467.

Hendy, M. D., and D. Penny. 1989. A framework for quantitaive study of evolutionary trees. Systematic Biology **38**:297-309.

Holland, B. R., K. T. Huber, V. Moulton, and P. J. Lockhart. 2004. Using consensus networks to visualize contradictory evidence for species phylogeny. Mol. Biol. Evol **21**:1459-1461.

Holland, B. R., D. Penny, and M. D. Hendy. 2003. Outgroup misplacement and phylogenetic inaccuracy under a molecular clock - a simulation study. Systematic Biology **52**:229-238.

Honeycutt, R. L., and R. M. Adkins. 1993. Higher-level systematics of eutherian mammals - an assessment of molecular characters and phylogenetic hypotheses. Annu. Rev. Ecol. System. **24**:279-305.

Huelsenbeck, J. P., and F. Ronquist. 2001. MrBayes: Bayesian inference on phylogenetic trees. Bioinformatics **17**:754-755.

Huson, D. H., and D. Bryant. 2006. Application of Phylogenetic Networks in Evolutionary Studies. Molecular Biology and Evolution **23**:254-267.

Jeffroy, O., H. Brinkmann, F. Delsuc, and H. Philippe. 2006. Phylogenomics: the beginning of incongruence? Trends in Genetics **22**:225-231.

Johansson, U. S., T. J. Parsons, M. Irestedt, and P. G. P. Ericson. 2001. Clades within the 'higher land birds', evaluated by nuclear DNA sequences. J. Zool. Syst. Evol. Research **39**:37-51.

Lin, Y.-H., P. A. McLenachan, A. R. Gore, M. J. Phillips, R. Ota, M. D. Hendy, and D. Penny. 2002. Four new mitochondrial genomes and the increased stability of

evolutionary trees of mammals from improved taxon sampling. Molecular Biology and Evolution **19**:2060-2070.

Livezey, B. C., and R. L. Zusi. 2007. Higher-order phylogeny of modern birds (Theropoda, Aves: Neornithes) based on comparative anatomy. II. Analysis and discussion. Zoological Journal of the Linnean Society **149**:1-95.

Lockhart, P. J., D. Penny, M. D. Hendy, C. J. Howe, T. J. Beanland, and A. W. D. Larkum. 1992. Controversy on chloroplast origins. FEBS Letters **301**:127-131.

Löytynoja, A., and N. Goldman. 2008. Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. Science **320**:1632-1635.

Mayr, E., and W. B. e. Provine. 1980. The Evolutionary Synthesis: Perspectives on the unification of biology. Harvard University Press, Cambridge.

Mayr, G., A. Manegold, and U. S. Johansson. 2003. Monophyletic groups within 'higher land birds' - comparison of morphological and molecular data. Jouran of Zoological Systematics and Evolutionary Research **41**:233-248.

Mindell, D. P., D. Sorenson, D. E. Dimcheff, M. Hasegawa, J. C. Ast, and T. Yuri. 1999. Interordinal relationships of birds and other reptiles based on whole mitochondrial genomes. Systematic Biology **48**:138-152.

Miyaki, C. Y., S. R. Matioli, T. Burke, and A. Wajntal. 1998. Parrot evolution and paleogeographical events: mitochondrial DNA evidence. Molecular Biology and Evolution **15**:544-551.

Morgan-Richards, M., S. A. Trewick, A. Bartosch-Härlid, O. Kardailsky, M. J. Phillips, P. A. McLenachan, and D. Penny. 2008. Bird evolution: testing the Metaves clade with six new mitochondrial genomes. BMC Evolutionary Biology **8**:20.

Penny, D., L. R. Foulds, and M. D. Hendy. 1982. Testing the theory of evolution by comparing phylogenetic trees constructed from five different protein sequences. Nature **297**:197-200.

Penny, D., and M. D. Hendy. 1986. Estimating the reliability of evolutionary trees. Mol. Biol. Evol **3**:403-417.

Penny, D., M. D. Hendy, and M. A. Steel. 1991. Testing the theory of descent. Pp. 155-183 *in* M. M. Miyamoto, and J. Cracraft, eds. Phylogenetic Analysis of DNA Sequences. Oxford University Press.

Penny, D., and M. J. Phillips. 2004. The rise of birds and mammals: are microevolutionary processes sufficient for macroevolution? Trends Ecol. Evol **19**:516-522.

Penny, D., W. T. White, M. D. Hendy, and M. J. Phillips. 2008. A bias in ML estimates of branch lengths in the presence of multiple signals. Molecular Biology and Evolution **25**:239-242.

Pereira, S. L., and A. J. Baker. 2006. A mitogenomic timescale for birds detects variable phylogenetic rates of molecular evolution and refutes the standard molecular clock. Molecular Biology and Evolution **23**:1731-1740.

Philippe, H., P. Lopez, H. Brinkmann, K. Budin, A. Germot, J. Laurent, D. Moreira, M. Muller, and H. Le Guyader. 2000. Early-branching or fast-evolving eukaryotes? An answer based on slowly evolving positions. Proc. Roy. Soc. Lond. B-Biol. Sci. **267**:1213-1221.

Phillips, M. J., F. Delsuc, and D. Penny. 2004. Genome-scale phylogeny: sampling and systematic errors are both important. Molecular Biology and Evolution **21**:1455-1458.

Phillips, M. J., G. C. Gibb, E. A. Crimp, and D. Penny. in prep. Tinamous and moa flock together: mitochondrial genome sequence analysis reveals independent losses of flight among ratites.

Phillips, M. J., P. A. McLenachan, C. Down, G. C. Gibb, and D. Penny. 2006. Combined mitochondrial and nuclear DNA sequences resolve the interrelations of the major Australasian marsupial radiations. Systematic Biology **55**:122-137.

Phillips, M. J., and D. Penny. 2003. The root of the mammalian tree inferred from whole mitochondrial genomes. Molecular Phylogenetics and Evolution **28**:171-185.

Poe, S., and A. L. Chubb. 2004. Birds in a bush: five genes indicate explosive evolution of avian orders. Evolution **58**:404-415.

Posada, D., and K. A. Crandall. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics **14**:817-818.

Proakis, J. G., C. M. Rader, F. Ling, C. L. Nikias, M. Moonen, and I. K. Proudler. 2002. Algorithms for statistical signal processing. Prentice Hall, Upper Saddle River, NJ.

Rodriguez-Ezpeleta, N., H. Brinkmann, B. Roure, N. Lartillot, B. F. Lang, and H. Phipippe. 2007. Detecting and overcoming systematic errors in genome-scale phylogenies. Systematic Biology **56**:389-399.

Shapiro, L. H., and J. P. Dumbacher. 2001. Adenylate kinase intron 5: A new nuclear locus for avian systematics. The Auk **118**:248-255.

Slack, K. E., F. Delsuc, P. A. McLenachan, U. Arnason, and D. Penny. 2007. Resolving the root of the avian mitogenomic tree by breaking up long branches. Molecular Phylogenetics and Evolution **42**:1-13.

Slack, K. E., A. Janke, D. Penny, and U. Arnason. 2003. Two new avian mitochondrial genomes (penguin and goose) and a summary of bird and reptile mitogenomic features. Gene **302**:43-52.

Slack, K. E., C. M. Jones, T. Ando, G. L. Harrison, E. Fordyce, U. Arnason, and D. Penny. 2006. Early penguin fossils, plus mitochondrial genomes, calibrate avian evolution. Mol. Biol. Evol **23**:1144-1155.

Stamatakis, A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyese with thousands of taxa and mixed models. Bioinformatics **22**:2688-2690.

Stidham, T. A. 1998. A lower jaw from a Cretaceous parrot. Nature **396**:29-30.

Susko, E., C. Field, C. Blouin, and A. J. Roger. 2003. Estimation of rates-across-sites distributions in phylogeneitc substitution models. Systematic Biology **52**:594-603.

Swofford, D. L. 2001. PAUP* Phylogenetic analysis using parsimony (*and other methods). Sinauer Associates, Sunderland.

van Tuinen, M., and S. B. Hedges. 2001. Calibration of avian molecular clocks. Mol. Biol. Evol **18**:206-213.

van Tuinen, M., T. A. Stidham, and E. A. Hadly. 2006. Tempo and mode of modern bird evolution observed with large-scale taxonomic sampling. Historical Biology **18**:209-225.

Waterhouse, D. M. 2006. Parrots in a nutshell: The fossil record of Psittaciformes (Aves). Historical Biology **18**:223-234.

Whitfield, J. B., and P. J. Lockhart. 2007. Deciphering ancient rapid radiations. TREE **22**:258-265.

Zwickl, D. J. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. The University of Texas, Austin.

3.9 APPENDICES

3.9.1 Fossil calibrations and molecular dating with BEAST

INTRODUCTION

Fossil calibrations

The integration of molecular phylogenies with reliable fossil calibrations is required, even though the fragmentary nature of avian fossils is a limiting factor. Whether modern birds were diversifying well before the K - Pg boundary is central to resolution of the Neoavian polytomy. One hypothesis, based on time estimates from fossil calibration points and molecular clocks, infers that the majority, if not all, of the major clades of Neornithes diverged during the Mesozoic. That is, divergence occurred within the Cretaceous, which is partially at odds with a literal reading of the fossil record (Dyke and van Tuinen 2004). An alternative hypothesis is that a specific modern bird 'morphotype' was present at low diversity during the Mesozoic – the so-called 'transitional shorebirds' or 'waterbirds'. These are proposed to have survived the end-Cretaceous extinction (Unwin 1993; Feduccia 1999; Littlewood et al. 2001; Douady et al. 2003) and given rise to all modern orders. However, aquatic animals are expected (on average) to have a better fossil record than that of land animals, and thus to provide the best calibration points. Therefore the conclusion that 'waterbirds' were the only modern birds present in the Late Cretaceous could be misleading in that these fossils will not represent the full diversity present in the Late Cretaceous (just those birds most likely to fossilize).

To help reconcile these opposing views, two particularly important bird fossils; an Anseriform, *Vegavis* (very Late Cretaceous [Maastrichtian]) reported in Clarke et al. (2005) and an early penguin fossil, *Waimanu*, from Slack et al. (2006) provide prior distributions for primary calibrations. *Waimanu* is expected to be particularly useful because penguins are aquatic, relatively large (for birds), have solid bones (not hollow) and as they are not flying birds, they are relatively distinctive (and hence are easier to identify). For at least these four reasons penguins are expected to have a better fossil record than most birds. This minimizes the problem, discussed quantitatively by Soligo et al. (2007), that the oldest known fossils can severely under-estimate the actual time of appearance of a group.

In contrast, a study by Ericson et al. (2006) relied chiefly on the split between hummingbirds and swifts (47.5 Ma) as a fixed time of divergence. The hummingbird fossil, on all criteria mentioned above for penguins, is expected to be a severe underestimate of the actual divergence. As a result, and not surprisingly, Ericson et al. (2006) obtained relatively recent dates for phylogenetic splits in Neoaves. This result was later challenged by Brown et al. (2007), using the same data but more appropriate fossil calibrations.

METHODS

Dating with BEAST

Brown et al. (2008) have shown that a wide variety of molecular dating methods and a variety of assumptions lead to similar results (in this case, support for Neoaves [and other avian groups] diversifying in the Cretaceous). Because of their results we have used the program BEAST (Drummond et al. 2006). Brown et al. (2008) point out two important ways that BEAST differs from other programs in their analyses, and which are important for our study. It does not require a fixed tree topology, instead allowing for simultaneous estimation of branch lengths, topology, substitution model parameters and dates. This is particularly important for datasets where there are still some uncertain relationships among clades. Second, BEAST does not assume autocorrelation of substitution rates across the tree (Drummond et al. 2006).

In order to test estimates of divergence between particular clades we used two well accepted calibrations; *Vegavis* (Clarke et al. 2005) which is outside the Neoaves and the penguin *Waimanu* (Slack et al. 2006) which is within. The *Vegavis* calibration is within the outgroup taxa (Galloanserae) at the base of the Anseriformes and the *Waimanu* calibration is at the base of the marine seabirds lineage, including the penguins. For calibration of *Vegavis*, the mean estimate of 68 Ma with a normal distribution (95% +/- 6 or s.d = 3.06) makes the bulk of the prior older than *Vegavis* while allowing for the possibility that *Vegavis* is an Anseriform outside of the Anatidae-*Anseranas* split.

For calibration of *Waimanu*, the mean estimate of 66 Ma with a normal distribution (95% +/- 6 or s.d = 3.06) was used so that the minimum of 60 Ma covers any potential geological dating error.

The upper 95% of each parameter allows for the possibility of early Maastrichtian members within the crown groups (e.g. loon in the case of marine birds) and another putative Anseriform to have existed. In addition, a conservative (uniform) bound of 66 - 124 Ma was used on the root – given that *Vegavis* definitely belongs to Galloanserae at 66 Ma and the discovery of now well sampled sites back to 124 Ma (e.g. Jehol Group comprising the Yixian and Jiufotang formations; Zhou et al. 2007) containing many ecological equivalents (but no modern birds). Two independent BEAST runs of 10 million generations each were carried out on the dataset to ensure convergence. We used the program TREEANNOTATOR to exclude the first 10% of trees as burnin and generated the maximum credibility tree with the following results.

RESULTS

Based on the extensive analyses of Brown et al. (2008) we used BEAST and two main calibration points, one inside the ingroup (*Waimanu*, *ca* 62 Ma) and one in the outgroup (*Vegavis*, *ca* 66 Ma). In addition, the Neoaves (or Neognaths) were constrained not to be older than 124 Ma, but unlike Brown et al. (2008) we did not constrain the tree topology. Using mean dates obtained from BEAST we estimate that 12 - 15 Neoavian lineages were present prior to the K - Pg boundary (more than 65 Ma) as shown in figure 3.10.1 below.

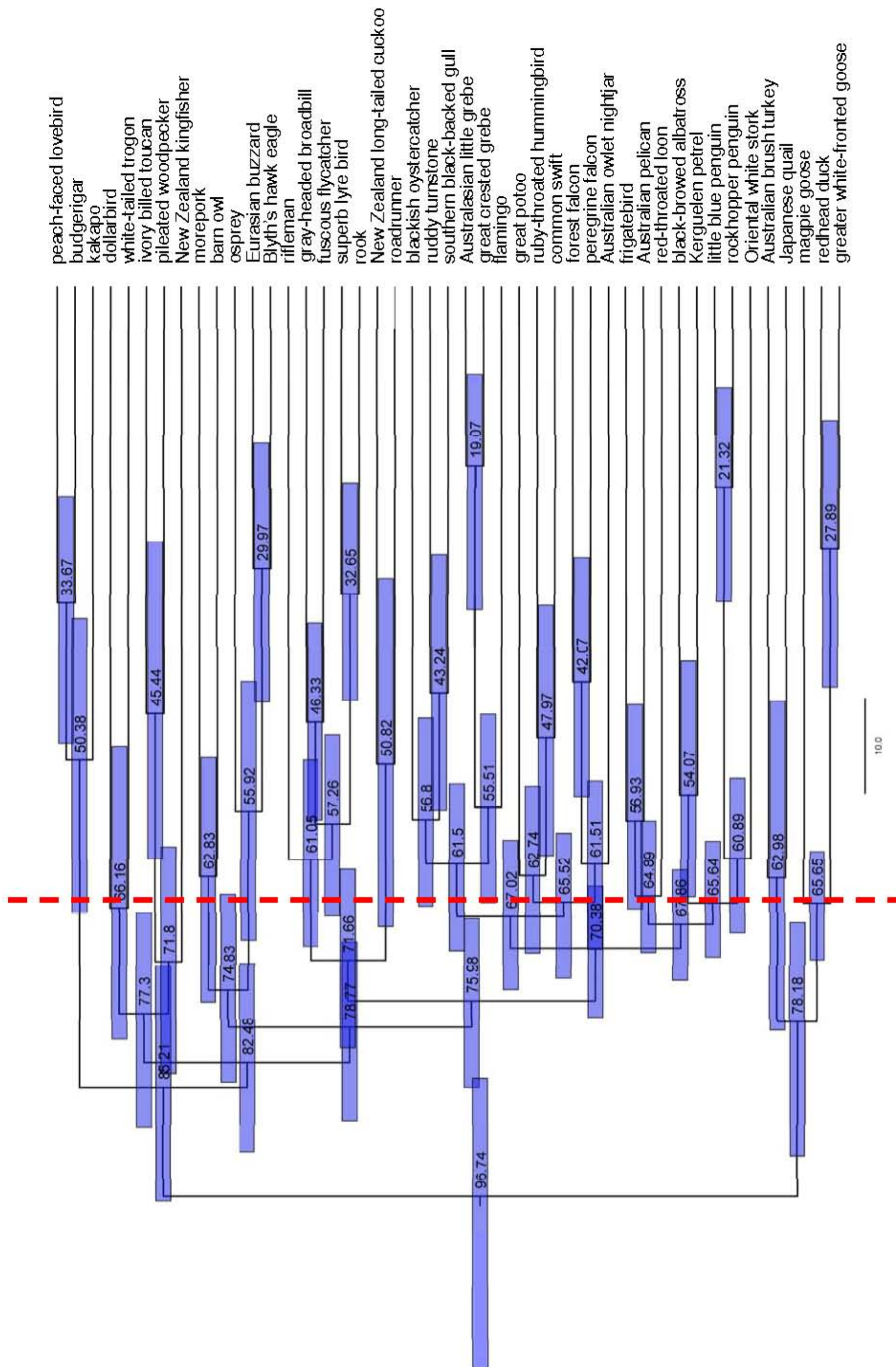


Figure 3.5 BEAST output tree for Neoaves

Max credibility tree from BEAST output for 45 birds. Our data indicate the presence of *ca* 12 - 15 Neoavian lineages present prior to/at the K - Pg boundary. Numbers are mean date estimates (Ma), blue bars show distribution/error on each date estimate and red dashed line indicates the K - Pg boundary (65 Ma).

3.9.2 The probability of observing a pre-specified clade

We calculate the probability of observing a pre-specified clade of k taxa in a binary tree on n taxa, as the proportion of all binary trees containing that clade. This can be extended to the probabilities of the clade being found as m = two, three or even more subclades. We see below (that for all but small values of n and k) these probabilities are very low, so that finding a pre-specified clade on a tree formed from new data is highly significant.

It is well known that for n taxa there are;

$$B(n) = (2n-5)!! = 1 \times 3 \times 5 \times \ldots \times (2n-5) \qquad - (1)$$

unrooted binary trees where each tip (leaf) of the tree is labeled by a unique taxon (see Penny et al. 1991). Similarly, the number of rooted binary trees is;

$$R(n) = (2n-3)!! = B(n+1). \qquad - (2)$$

A simplified approach to deriving the formulae is indicated in figure 3.4A; and the calculations are straightforward in an Excel spreadsheet.

Probability of a specific subset forming a single clade (that is, m = 1)

The probability (P) of observing a predefined clade of k taxa in a binary tree of n taxa is;

$$P(n,k) = R(k) \times B(n-k+1) / B(n) \qquad - (3)$$

where the numerator is the number of rooted subtrees for the clade (R(k)), multiplied by the number of trees on the remaining taxa (B(n-k+1), including a leaf for the clade). Dividing by the number of unrooted trees (B(n)) gives the proportion of trees having that clade of k taxa.

For rooted trees the probability (P_R) of observing a predefined clade of k taxa in a rooted binary tree of n taxa is similarly;

$$P_R(n,k) = R(k) \times R(n-k+1) / R(n) = P(n+1,k) \qquad - (4)$$

For this question we consider all binary trees as equally likely; the trees are derived from a Markov model where there is no prior information about the distribution of tree shapes (Steel and Penny 1993). For two taxa in the predefined clade, the equation simplifies (see table 3-1) to;

$$P(n,2) = 1/(2n-5) \text{ for unrooted trees, and}$$

$$P_R(n,k) = 1/(2n-3) \text{ for rooted trees}$$

And for three taxa it simplifies to;

$$P(n,3) = 3/(2n\text{-}5)\ (2n\text{-}7) \text{ for unrooted trees, and}$$

$$P_R(n, 3) = 3/(2n\text{-}3)\ (2n\text{-}5) \text{ for rooted trees}$$

In our analyses we consider the 40 taxa to account for 20 independent taxonomic groupings; passerines (5 taxa), cuckoos (2 taxa), parrots (3 taxa), shorebirds (3 taxa) owls (2 taxa), dollarbird, kingfisher, trogon, woodpecker + toucan (2 taxa), potoo, owlet nightjar, Apodiformes (2 taxa), Accipitriformes (3 taxa), falcons (2 taxa), flamingo + grebes (3 taxa), Pelecaniformes (2 taxa), tube-noses (2 taxa), stork, penguins (2 taxa) and the loon. For rooted trees, there are 21 independent groupings, the above plus Galloanserae (5 taxa). If, for example, $k = 5$ and $n = 40$, and the probability of observing a pre-specified clade on new data where all taxa are included is

$$P_R(40, 5) = \mathrm{R}(5) \times \mathrm{R}(36)/\ \mathrm{R}(40) \approx 3.5 \times 10^{-6}$$

or, where only the independent taxonomic groups are included,

$$P_R(21, 5) = \mathrm{R}(5) \times \mathrm{R}(17)/\ \mathrm{R}(21) \approx 6.3 \times 10^{-5}$$

Predefined clade found as *m* ≥ 2 subgroups

The calculations can be extended to cases where the predicted clade is partitioned into $m = 2$ or more subclades on a tree. Figures 3.4C and 3.4D show two cases where a predefined clade appears in two separate areas of the tree ($m = 2$). For a large number of taxa, it is still most unlikely that a predefined clade will be in just two locations on a new tree. In the case shown here, there are $k = 4$ taxa in the clade, and for $m = 2$ they can occur as either; a single taxon and a group of three (figure 3.4C); or as two groups, each with 2 taxa (figure 3.4D). When the clade of k taxa is split into m subclades, with $k_1, k_2, \ldots, k_m$ taxa respectively then we must consider each combination of the m subclades separately, we find the probability is,

$$\mathrm{P(n,k,m)} = \frac{\mathrm{k!\ B(n\text{-}k\text{+}m)}}{\mathrm{m!\ B(n)}} \sum_{\Sigma k_i = k} \prod_i \frac{\mathrm{R}(k_i)}{k_i!} \qquad – (5)$$

3.9.3 Tabulated results from Bayesian (BPP) and bootstrap (bs) analyses using RAxML and PhyML

Table 3-2 Support values

Comparison for Bayesian Posterior Probablilities (MrBayes) and bootstrap values (PhyML and RAxML) for the main groupings found in the current study and those of Hackett et al. (2008).

		Hackett (bs)		Our BPP			Our PhyML (bs)				Our RAxML (bs)			
Our groups	Hackett et al. groups	ML	ML + outgp	s=2	s=6	s=2 + outgroup	s=2	s=6	s=2 + outgroup	s=6 + outgroup	s=2	s=6	s=2 + outgroup	s=6 + outgroup
accipitridae + osprey	Accipitridae + Cathartidae (E)	70	61	100	100	100	100	100	100	100	100	100	100	100
CAM'	Caprimulgiformes + Apodiformes (L)	99	98	61.2	45.2	64.8	–	–	33.3	–	–	51(box)	–	46 (box)
Cuckoos	Cuculiformes	100	100	100	100	100	100	100	100	100	100	100	100	100
falcons	Falconidae		100	100	100	100	100	100	100	100	100	100	100	100
grebes				100	100	100	100	100	100	100	100	100	100	100
Owls	Strigiformes	100	100	99.6	99.8	99.8	66.6	66.6	75	91.6	74	68	70	64
Parrots	Psittaciformes	100	100	100	100	100	100	100	100	100	100	100	100	100
Passerines	Passeriformes	100	100	100	100	100	100	100	100	100	100	100	100	100
penguins				100	100	100	100	100	100	100	100	100	100	100
Shorebirds	Charadriformes	100	100	70.2	100	88	75	50	50	50	56	71	56	72
WoodKing'				99.2	97.4	99.9	–	–	–	33.3	56	37	44	30
Outgroup	Galloanserae (P)	100	100	N/A	N/A	100	N/A	N/A	100	100	N/A	N/A	100	100
albatross + petrel				100	100	100	91.6	58.3	50	83.3	96	96	94	94
flamingo + grebe	Phoenicopteriformes + Podicipediformes (N)	100	100	99.9	100	98.7	66.6	33.3	–	58.3	58	77	47	66
gull + turnstone	*Larus* + *Arenaria*		100	100	100	100	100	100	100	100	99	100	100	100
Owls + accipitridae/osprey				96.6	99.9	96.9	–	–	–	–	54	55	58	52
parrots + 'WoodKing'				98.2	99.9	99.9 + outgp	–	–	–	–	29	43	27+outgp	–
Passerines + Cuckoos				99.9	95.4	99.9	33.3	–	33.3	–	57	38	54	41
shorebirds + flamingo/grebe				–	99.6	–	–	41.6	–	58.3	–	50	–	42
pelican + frigatebird	Pelecaniformes		88	96.2	95.4	96.8	–	–	–	–	38	27	37	44
swift + hummingbird	Apodiformes		98	100	100	100	100	100	100	100	100	99	100	98
WoodKing' minus Trogoniformes	Coraciformes + Piciformes (C)	99	98	–	–	–	–	–	–		–	–	–	–
Woodpecker + aracari				100	100	100	100	100	100	100	100	100	100	100
pelican + frigatebird + loon				96.3	48.5	95.9								
swift + hummingbird + owletnightjar	Apodiformes + *Aegotheles* (K)	99	98	98.2	77.8	97.1	91.6	100	83.3	91.6	61	51	55	46
falcons + owls + accipitridae/osprey				73.7	–	68.7	–	–	–	–	–	–	–	–
passerines + parrots	Passeriformes + Psittaciformes (A)	65	77	–	–	–	–	–	–	–	–	–	–	–
passerines + parrots + falcons	Passeriformes + Psittaciformes + Falconidae (B)	64	73	–	–	–	–	–	–	–	–	–	–	–
– denotes not found in our analyses														

3.9.4 Consensus tree including all outgroup taxa used for alignment before site stripping

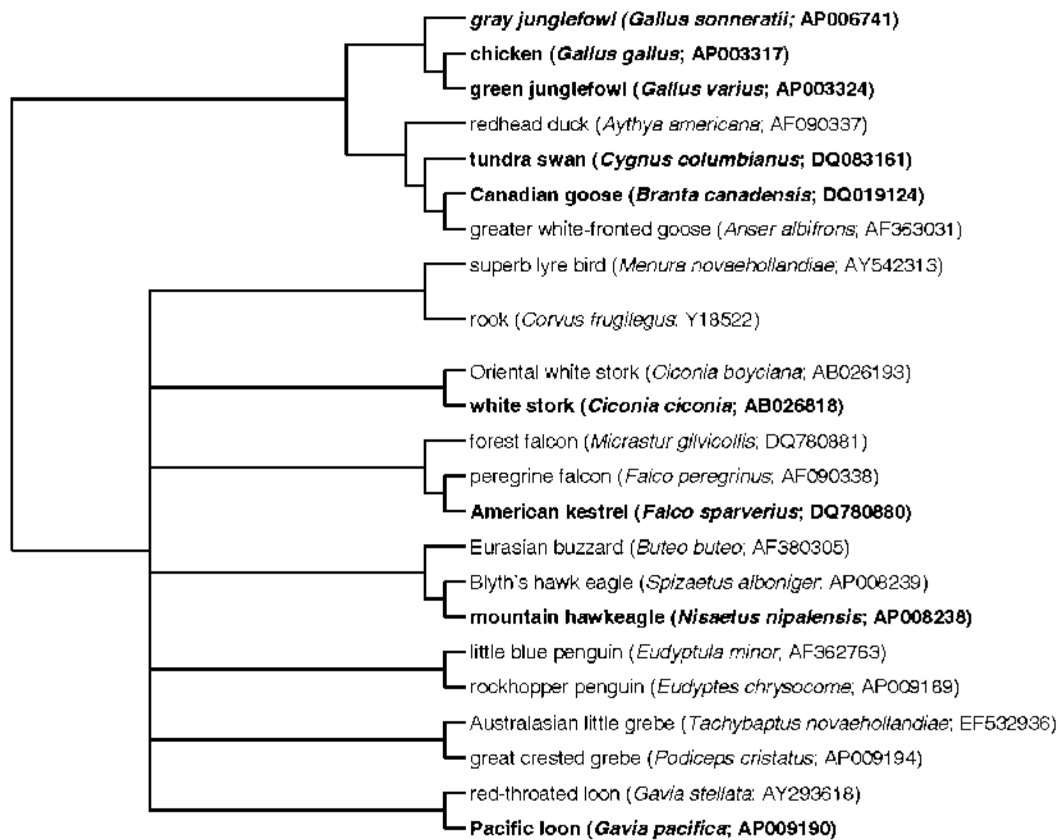


Figure 3.6 Consensus tree of Neoaves (modern birds) before site-stripping

Consensus tree before the implementation of the site-stripping noise reduction method. Included are the additional, closely related taxa not used in further analyses (highlighted in bold).

REFERENCES

Brown, J. W., R. B. Payne, and D. P. Mindell. 2007. Nuclear DNA does not reconcile 'rocks' and 'clocks' in Neoaves: a comment on Ericson *et al.* Biology Letters **5**.

Brown, J. W., J. S. Rest, J. García-Moreno, M. D. Sorenson, and D. P. Mindell. 2008. Strong mitochondrial DNA support for a Cretaceous origin of modern avian lineages. BMC Biology.

Clarke, J. A., C. P. Tambussi, J. I. Noriega, G. M. Erikson, and R. A. Ketcham. 2005. Definitive fossil evidence for the extant avian radiation in the Cretaceous. Nature **433**:305-308.

Douady, C. J., F. Catzeflis, J. Raman, M. S. Springer, and M. Stanhope. 2003. The Sahara as a vicariant agent, and the role of Miocene climatic events, in the diversification of the mammalian order Macroscelidea (elephant shrews). Proceeding of the National Academy of Science, USA **100**:8325-8330.

Drummond, A. J., S. Y. W. Ho, M. J. Phillips, and A. Rambaut. 2006. Relaxed Phylogenetics and Dating with Confidence. PLoS Biology **4**:699-710.

Dyke, G. J., and M. van Tuinen. 2004. The evolutionary radiation of modern birds (Neornithes): reconciling molecules, morphology and the fossil record. Zoological Journal of the Linnean Society **141**:153-177.

Ericson, P. G. P., C. L. Anderson, T. Britton, A. Elzanowski, U. S. Johansson, M. Källersjö, J. I. Ohlson, T. J. Parsons, D. Zuccon, and G. Mayr. 2006. Diversification of Neoaves: integration of molecular sequence data and fossils. Biol. Lett. 5: (in press) **4**:543-547.

Feduccia, A. 1999. The origin and evolution of birds. Yale University Press, New Haven.

Littlewood, D. T. J., P. D. Olson, M. J. Telford, E. A. Herniou, and M. Riutort. 2001. Elongation Factor 1-Alpha sequences alone do not assist in resolving the position of the Acoela within the Metazoa. Molecular Biology and Evolution **18**:437-442.

Slack, K. E., C. M. Jones, T. Ando, G. L. Harrison, E. Fordyce, U. Arnason, and D. Penny. 2006. Early penguin fossils, plus mitochondrial genomes, calibrate avian evolution. Mol. Biol. Evol **23**:1144-1155.

Soligo, C., O. A. Will, S. Tavaré, C. R. Marshall, and R. D. Martin. 2007. New Light on the dates of primate origins and divergence. Pp. 29-49 *in* M. J. Ravosa, and M. Dagosto, eds. Primate Origins: Adaptaions and Evolution. Springer US.

Steel, M. A., and D. Penny. 1993. Distributions of tree comparison metrics - some new results. Systematic Biology **42**:126-141.

Unwin, D. M. 1993. Aves. Pp. 717-737 *in* M. J. Benton, ed. The Fossil Record II. Chapman & Hall, London.

Zhou, Z., Y. Huang, and F. Shi. 2007. The mitochondrial genome of *Ruspolia dubia* (Orthoptera: Concephalidae) contains a short A+T-rich region of 70bp in length. Genome **50**:855-866.

CHAPTER FOUR

***T**he modern map holds the record of "modern" life but it does not explain how this life was gotten together*

-Leon Croizat

Return to Gondwana: exploring biogeographical patterns of the family Anostostomatidae (Orthoptera: Ensifera)

4.1 ABSTRACT

Classification within the Orthopteran family Anostostomatidae is known to be complicated and problematic owing to much confusion about the taxonomic hierarchy. In *Chapter Two*, Australasian Anostostomatidae, were found to be paraphyletic within New Zealand and Australia. Interestingly, New Zealand taxa appear diverse with representatives of both clades, one with affinities to Australia and the other with affinities to New Caledonia. Here we return to the story of weta in Gondwana and try to resolve family relationships in a wider spatial (geographic) context to determine if Australasian taxa are monophyletic when other members of the family are included. The phylogenetic relationships within the Anostostomatidae were investigated using maximum likelihood and Bayesian analyses carried out on both mitochondrial and nuclear sequences. Samples from across the geographic range were included (South Africa, Madagascar, South America, Australia, New Caledonia and New Zealand), and two clades were observed, congruent with earlier findings. Interestingly, New Zealand, Madagascan and South African faunas were all paraphyletic with lineages found throughout the tree (in both major clades) while Australian and Brazilian taxa were paraphyletic however confined to one clade each. New Caledonian taxa sampled were only found in one clade and represented a monophyletic lineage. Not surprisingly, sequence divergence within geographic regions was found to be relatively high in the mitochondrial genes (COI and 12S) while low in the nuclear ribosomal RNA genes (18S and 28S). Under the vicariance paradigm, phylogenetic relationships should follow the order of continental breakup, but this was not found. Further, if dispersal and colonisation were continuous, no geographic substructure is expected, however distinct geographic substructure within clades was consistently observed. This interesting phylogenetic pattern may be a case of convergent evolution or paraphyletic sampling which highlights the taxonomic issues of the group. Lastly, taxonomic revision of Anostostomatidae is suggested as two clades were consistently found that may correspond to just two subfamilies (which subsume the seven currently recognised).

4.2 INTRODUCTION

Previously, an unexpected pattern was found in the Australasian region (see *Chapter Two*) whereby two distinct lineages were observed among samples from Australia, New Caledonia and New Zealand. Interestingly, both lineages (clades) are represented in the New Zealand fauna while only one each is present in Australia and New Caledonia. Specifically, one lineage consisted of taxa from New Caledonia and the tusked weta (*Anisoura/Motuweta*) from the far north of New Zealand, while the second lineage included taxa from a diverse array of genera from Australia and New Zealand. One explanation for this split is convergence of molecular mutation rates resulting in a false grouping of New Caledonian and New Zealand tusked weta lineages. Dating methods reveal such a rate shift has occurred at least once. Molecular dating also indicated the possibility of dispersal between the landmasses, however direction was not determined. This observation, (i.e. two distinct lineages in Australasia), is unexpected based on the geological history of the region. In order to investigate if this split is a consequence of regional history, a broader approach investigating family relationships throughout the Southern Hemisphere was undertaken.

Unlike continents in the north, the land in the Southern Hemisphere is widely dispersed and separated by large oceanic gaps. Although the spatial arrangement of land is well explained by continental drift, the influence of vicariance (via continental drift) on taxon distribution is less well understood. Historical vicariant patterns may be overshadowed or even masked by more recent events (e.g. dispersal and extinction), which is especially problematic in the absence of a good fossil record. The Southern Hemisphere contains a mixture of large continents and many islands. Large, stable landmasses tend to be the source of migration for nearby oceanic islands (Johnson et al. 2000) created through volcanic activity, while the origins of continental island biota (i.e. islands made of continental crust) are more difficult to determine.

A number of competing hypotheses are invoked to explain the wide distribution of modern biota throughout the Southern Hemisphere, however two extremes of the continuum underlie most studies; either the biota represent vicariant (ancient) links to Gondwana or more modern dispersal events are responsible for the spread of taxa.

It should be noted that these ideas do not have to represent mutually exclusive hypotheses. "Gondwanan" biota can mean taxa of ancient origins or taxa found on southern lands either by historical processes or recent dispersal (e.g. see Goldberg et al. 2008 for Gondwanan definitions). That is, congruence between timing of geographical isolation (on landmasses) and genetic divergence of sister taxa (on those landmasses) need to be similar for vicariance to be supported. By contrast, incongruence between geographic isolation and genetic divergence might support long-distance dispersal as a probable explanation. Of course there are a number of possibilities that are a mixture of the two extremes but for simplicity are not considered at this time. In addition to the above, inferences about distributions are affected by sampling error, extinction and range expansion (Ronquist 1997; Hunn and Upchurch 2001; Lieberman 2002; Donoghue and Moore 2003).

By invoking two extreme scenarios an evaluation of the plausibility of classical hypotheses and predictions about observed patterns can be made. The breakup of Gondwana not only provides a prediction of regional monophyly but also the order of lineage relationships based on the order of separation. That is, ancestral lineages must pre-date Gondwanan breakup. Alternatively, a lack of regional monophyly could indicate origins via range expansion (post breakup), with basal lineages indicative of dispersal origin (source). In an earlier study on Australasian taxa (Pratt et al. 2008) we tested timing of divergence by comparing two alternative calibration points; (1) the Oligocene marine inundation of New Zealand and (2) the separation of Zealandia from Australia (see *Chapter Two*). Comparison of accepted mutation rates (for insects), to those obtained via dating methods, supported divergence of the main clades at, or shortly after, Australia (Gondwana) and Zealandia (New Zealand and New Caledonia) separated 80 Ma (see *Chapter Two*).

The majority of taxonomic work on the Anostostomatidae was undertaken between 1803 and 1937 (see Johns 1997 for review). Historically, anostostomatid taxa were variously placed in different families and subfamilies including Mimnermidae Stål (1878); the Gryllacididae Stål 1876; Deinacridinae Karny 1932; and Henicinae Karny 1928.

These earlier works, however, are rife with synonymies, misspellings and incorrect identifications, which resulted in taxonomic confusion for the group (see Johns 1997; Brettschneider 2006). **To** date, most research on this family has been highly localised and there are no recent reviews on taxonomic levels lower than the superfamily (Brettschneider 2006). The Orthoptera Species File (OSF) online database lists the current taxonomic hierarchy for members of Orthoptera along with distribution maps (http://orthoptera.speciesfile.org/Common/basic/Taxa.aspx) (Eades and Otte see also Appendix 4.8.1 table 4-3). Representatives found in the Australasian (AUS) region are amongst the most well studied (Trewick and Morgan-Richards 2004; Pratt et al. 2008). Recently members of the South African fauna received attention (Toms 2001; Brettschneider 2006), however most other regions lack formal descriptions and little or no molecular systematic analyses. A fossil dating to 190 - 200 Ma from Queensland, Australia has been attributed to Anostostomatidae (Meads 1990), although no formal identification has been published. More recently, a new extinct subfamily (Euclydesinae) was described from a fossil considered to be early Cretaceous in age (*ca* 112 Ma) in the municipal district of Nova Olinda, Northeast Brazil (Family Mimnermidae, a synonym of Anostostomatidae) (Martins-Neto 2007), confirming the existence of the family in Gondwana during the breakup (although after the separation of Africa).

No previous phylogenetic study has attempted to evaluate the family on a global scale. This study includes representatives from South Africa (SAF), Madagascar (Mad), Australia (OZ), New Zealand (NZ), New Caledonia (NC), Chile (Chile) and Brazil (BZ) (see Appendix 4.8.2 table 4-4). Australia, New Zealand, and New Caledonia together will be referred to as Australasia (AUS). The aim is to: (1) determine if the two clades identified in Australasia are represented throughout the families' current range (and associated process behind biogeographic patterns), and (2) consider the implications (of this pattern) for family level systematics.

4.3 MATERIALS AND METHODS

(a) Sampling

The majority of sampling was undertaken by the authors in New Zealand and New Caledonia. In addition, samples of *Hemiandrus* were supplied by Darryl Gwynne (University of Toronto, Canada). New Zealand sampling included all three tusked weta species (*Anisoura/Motuweta*), representatives of the *Hemideina* and *Deinacrida* (previously shown to be monophyletic; Trewick and Morgan-Richards 2005; Pratt et al. 2008) and representatives within the taxonomic diversity of *Hemiandrus*. Sampling from other regions was less extensive, but still representative of the known generic diversity; Australia included representatives from the nine known (or postulated) genera including two recently proposed by Peter Johns (personal communication); Brazilian sampling includes ten individuals from the five proposed genera (four currently described, of which two are sampled here), Madagascan sampling includes seven specimens from two genera (four currently described) and South African sampling includes 15 samples from four genera (nine described). Taxa included are listed in Appendix 4.8.1 table 4-3 and Appendix 4.8.2 table 4-4. Material from Australia, Brazil, Madagascar, and South Africa, was supplied by Geoff Monteith (Brisbane Museum, Australia), Dave Rentz (CSIRO, Australia), Francisco de A. G. de Mello (Instiuto de Biocijncias, UNESP, Brazil), Mohsen Mofidi-Neyestanak (Imperial College London, UK), Helene Brettschneider and Phillip Bateman (University of Pretoria, South Africa). Assistance with identification and sampling of undescribed species was provided by P. M. Johns. This sampling, although not exhaustive, includes representatives of 47 genera and is the most complete dataset to date for the entire family. In addition, the following were used as outgroup taxa; Acrididae; Tettigoniidae; Stenopelmatidae; Gryllacrididae; Rhaphidophoridae (see Appendix 4.8.2 table 4-4 for specimen information).

(b) Molecular Methods

Whole genomic DNA was extracted from hind leg muscle following a salting-out method (Sunnucks & Hale 1996) and re-suspended in 50µL TE buffer (0.1mM EDTA, 10mM Tris) or water. PCR reactions were performed in 10µL volumes using ABgene Red Hot *Taq*. Products were visualised on 1% agarose gels stained with SYBRSafe (Invitrogen).

Thermal cycling PCR was carried out on an MJ Research PTC-200 thermal cycler and consisted of initial denaturation of 94°C for 2min, followed by 35 cycles of 94°C for 30sec, 48 - 50°C for 30sec and 72°C for 1min 30sec with a final extension of 72°C for 3min. PCR products were purified with Shrimp Alkaline Phosphatase (SAP) and Exonuclease I (EXOI) following manufacturer's recommendations (USB Corporation). Sequencing was carried out using BigDye Terminator v3.1 chemistry and an ABI3730XL Genetic Analyser (Applied Biosystems, Foster City, CA). DNA sequences were deposited on NCBI GenBank (Appendix 4.8.2 table 4-4). Primers used for PCR were; mtDNA COI: LCO1490, HCO2198 (Folmer et al. 1994), CI-J-1718 and CI-N-2191, CI-J-2195 and L2-N-3014 (Simon et al. 1994); mtDNA third domain 12S rRNA: SR-N-14588 and LR-J-13417 (Simon et al. 1994); nuclear rRNA 18S: 18S-S22, 18S-A1984 (Vawter 1991), 18S_1F and 18S_1R (Pratt et al. 2008); and 28S: 28SrD1.2a, 28SrD3.2a, 28SrD4.8a, 28SA, 28SrD4.2b, 28SrD5b, 28SrD7bl, 28SB (Whiting 2002b).

(c) Phylogenetic analysis

Individual sequence reads were checked against ABI trace files using SEQUENCHER V.4.70 (Gene Codes Corp. Ann Arbour, MI) and aligned using SE-AL v2.0a11 (Rambaut 1996). The protein coding gene COI was translated into amino acids to ensure correct reading frame and to detect evidence of nuclear copies (which were subsequently removed). Ribosomal RNA genes were checked for indels. In cases where missing data were included they were coded as N in analyses. In order to evaluate individual genes and concatenated data, we divided the datasets into the following; I – 18S Anostostomatidae; II – combined 18S and 28S Anostostomatidae; III – COI-RY-coded Anostostomatidae; IV – combined COI-RY and 12S Anostostomatidae.

The COI data were partitioned by codon position. In order to maximise third codon information we treated it in two different ways; as four nucleotides (A, G, T, C) or RY-coded (A and G = R, T and C = Y). In order to avoid potential tree estimation bias due to nucleotide composition or saturation, we used RY-coding on the third codon position nucleotides for COI sequences in dataset III and IV.

Re-coding of this sort has been shown to greatly improve consistency in phylogenetic resolution by reducing bias from differences in nucleotide composition (Phillips and Penny 2003) which is useful when looking at deeper divergences. Models of DNA evolution were optimized separately for each data set using MODELTEST 3.7 (Posada and Crandall 1998). Following Posada and Buckley (2004) the Akaike Information Criterion (AIC) was preferred to the Hierarchical likelihood ratio test (hLRTs) (Posada and Buckley 2004). Maximum likelihood (ML) analyses were implemented using the programs PAUP* (Swofford 2003) and GARLI version 0.951 (Zwickl 2006). Model parameters from MODELTEST were implemented using a general time-reversible model with invariable sites and a gamma distribution for variable rate sites (GTR + I + G), with a heuristic search under the likelihood criterion with trees obtained from stepwise addition. Phylogenetic confidence in topologies was assessed by bootstrapping (Felsenstein 1985) with heuristic analysis of 100 replicate data sets as implemented in GARLI (Zwickl 2006).

Bayesian analyses were implemented using MRBAYES 3.1 (Huelsenbeck and Ronquist 2001b). The parameter rich GTR model was specified (nst = 6) with a proportion of invariant sites and gamma distribution of rate variation. Analyses of datasets I (18S), II (18S + 28S), III (COI), and IV (COI + 12S), data were undertaken with (parameters unlinked) and without character-set partitions. Two runs of four Markov chains (each with one cold chain) with 1 - 10 × 10^6 generations and default priors were used, sampling every thousandth tree. A “burn-in” of 10% was removed after examination of log likelihood scores and average standard deviation of the split frequencies. Trees saved below the “burn-in” generation were discarded and a majority rule consensus of the remaining trees was calculated. Multiple replicates of the Bayesian runs were carried out to ensure convergence of the posteriors.

(d) Tree comparisons

In order to assess the degree of signal erosion and compositional non-stationarity the stemminess and relative compositional variability was examined.

Stemminess

Stemminess is defined as the proportion of overall tree-length contributed by internal branches (Fiala and Sokal 1985). Following Phillips et al. (2001) analysis using minimum evolution (ME) trees (produced from uncorrected distances in PAUP*) was carried out. Given the same topology, greater phylogenetic signal erosion results in shorter internal branches relative to their attendant external branches (hence, lower stemminess) (see *Appendix D* reprints Phillips and Pratt 2008).

Relative compositional variability (RCV)

The average variability in base composition between taxa is referred to as RCV. For nucleotides this can be calculated using;

$$RCV = \sum_{i=1}^{n} \left|A_i - A^*\right| + \left|T_i - T^*\right| + \left|C_i - C^*\right| + \left|G_i - G^*\right| / n.t$$

Where A_i, T_i, C_i and G_i are the frequencies of each nucleotide for the *i*th taxon and A*, T*, C* and G* are average nucleotide frequencies across the *n* taxa, and *t* is the number of sites. Uninformative sites were excluded from RCV calculations as they are known to dilute apparent non-stationarity.

4.4 RESULTS

A summary of sequence data collected, along with sample information and accession numbers is given in the Appendix 4.8.2 table 4-4. NCBI nucleotide BLAST searches were carried out and returned matches for previously published Orthopteran sequences. The alignment used in each analysis is available as a NEXUS file from RCP on request.

(a) Sequence divergence

Pairwise sequence divergences among anostostomatid taxa were compared across the entire dataset for COI-RY-coded (less than 19.5%) and 18S sequences (less than 3.9%) as they represented the most complete datasets.

The clades termed "Capricorn" and "Austral" were observed in both nuclear (figure 4.1) and mitochondrial (figure 4.2 and see figure 4.3 for geographic representation[2]) datasets and the clades are discussed separately later.

Relatively large divergences were detected within dataset III - COI-RY. The "Capricorn" clade has a maximum of 16.7% (New Zealand: *Anisoura-Motuweta* 7%; New Caledonia: *Carcinopsis* and *Aistus* less than 9.6%; Brazil: *Lutosa*, *Apotetamenus* plus three new genera 0.4 - 13%; and Madagascar: *Spizaphilus* A, *Spizaphilus* B and *Brachyporus* less than 16.7%) while the "Austral" clade was equally diverse ranging up to 14.8% (New Zealand: *Hemideina*, *Deinacrida* 2.3 - 10.4%, *Hemiandrus* 4.9 - 13.6%; Australia: *Anostostoma*, *Transaevum*, *Penalva*, *Gryllotaurus*, *Hemiandrus*, *Exogryllacris*, *Hypocophoides,* genus A and genus B 9.2 - 14.8%; and South Africa: *Libanasa*, *Nasidius*, *Libanasidus* and *Onosandrus*, *Spizaphilus* C 0.2 - 10.3%.

In contrast, and as expected, relatively small divergences were observed within dataset I – 18S. "Capricorn" clade: less than 1.3% (New Zealand: *Anisoura-Motuweta* 0.0%; New Caledonia: *Carcinopsis* and *Aistus* 1.0%; Brazil: *Lutosa*, *Apotetamenus* plus three new genera 0.1 - 2.5%; and Madagascar: *Spizaphilus* A, *Spizaphilus* B and *Brachyporus* 0.7 - 1.3%) and "Austral" clade: 2.8% (New Zealand: *Hemideina*, *Deinacrida* 0.0%, *Hemiandrus* 1%; Australia: *Anostostoma*, *Transaevum*, *Penalva*, *Gryllotaurus*, *Hemiandrus*, *Exogryllacris*, *Hypocophoides,* genus A and genus B 1.3%; and South Africa: *Libanasa*, *Nasidius*, *Libanasidus* and *Onosandrus*, *Spizaphilus* C 0.8%).

Coding sequences (dataset III - COI) were translated into amino acid sequences to assess the number of changes observed for each clade ("Capricorn" and "Austral"). Out of 496 amino acid positions we found 20 changes distinguished the "Capricorn" clade from the "Austral" clade (see table 4-1). A further four amino acids from taxa characteristic of the "Capricorn" clade show homoplasy. Deterministic amino acid changes were spread throughout the coding region sequenced, although certain areas were more variable (shown by close position number).

[2] The "Capricorn" clade refers to the distribution of taxa near the Tropic of Capricorn while the "Austral" clade is used as most taxa (sampled here), are from this region.

Table 4-1 Amino acid changes

Amino acid changes distinguishing “Capricorn” and “Austral” clades for protein coding COI gene.

Position	“Capricorn”	“Austral”	homoplasic sites
107	Ser	Ala/Thr	
126	Ala	Ser	
169	Ile	Met	
170	Asn	Thr/Ser	
173	Arg	Gln	
182	Leu	Ala/Thr	
186	Ile	Leu	NCal101 Val
251	Val	Ile	
258	Ile	Lys/Thr	
260	Thr	Ala/Thr	
268	Phe	Tyr	
271	Met	Leu	
313	Val	Ile	
320	Ile	Leu	
321	Tyr	His	
324	Arg	Gln	NCal101 Gln
328	Asn	Ser	PON1 Ser
331	Met	Leu	
334	Ser	Ala/Thr	
338	Ile	Val	
350	Ile	Val	
478	Asn	Pro	BZ04 Ser
482	Asn	Val/Thr/Ile/Met	
490	Leu	Asn/Lys/Ala/Val/His/Gln/Thr	

(b) Phylogenetic analysis

Nuclear data (Dataset I - 18S), consisted of 949 bp after removal of a 37 bp hypervariable indel region (between bp 719 and 756) from the original alignment for 61 taxa (55 ingroup and 6 outgroup taxa; Mad379 - Gryllacrididae, Mad380 - Gryllacrididae AF514554 - Gryllacridae AY121145 - Stenopelmatidae, AY521870 - Rhaphidophoridae and AY037173 - Acrididae). Bayesian and ML analyses yielded similar topologies (figure 4.1). Support was recovered for two sister clades within Anostostomatidae, each consisting of taxa from several geographic regions.

The "Capricorn" clade (figure 4.1), comprised of species from New Zealand (NZI: *Anisoura-Motuweta*); New Caledonia (NC: *Aistus* and *Carcinopsis*); Brazil (BZI: new genus and *Apotetamenus*; and BZII: *Lutosa*, and two new genera); Madagascar (MadI: *Spizaphilus* A and B and MadII: *Brachyporus*), and South Africa (SAFI: *Libanasa*). While the "Austral" clade (figure 4.1) included all remaining New Zealand taxa (NZII: *Hemideina* and *Deinacrida,* NZIII and NZ IV: *Hemiandrus*), Madagascan taxa (MadIII: *Spizaphilus* C); and South Africa taxa (SAFII: *Nasidius*, *Libanasidus* and *Onosandrus*); plus all Australian taxa (OZI: *Penalva* and *Hypocophoides*; OZII: *Hemiandrus* and Genus B; and OZIII: *Anostostoma*, *Gryllotaurus*, and Genus A); and the Chilean sample (Chile: *Cratomelus*). The positions of two Australian taxa, *Exogryllacris* and *Transaevum* (floating branches in this analysis) are incongruent with previous work (Pratt et al. 2008) and may result from a low signal to noise ratio which may also account for the split within NZIV (see figure 4.2).

In previous work NC and NZI were removed from analyses because of an observed rate shift. The current study indicates that it is most likely NC with a rate shift, as indicated by the long-branch length. The "Austral" clade, although less well resolved, also shows similar patterns observed in previous work (Pratt et al. 2008). The Australian taxa are not monophyletic, instead splitting into three lineages (OZI, OZII and OZIII). Similarly, the NZ taxa are not monophyletic, represented by NZII, NZIII and NZIV. Interestingly, *Hemiandrus* is not monophyletic in the current dataset with Australian *Hemiandrus* (OZII) and two New Zealand *Hemiandrus* (NZIII and NZIV) lineages. Taxa from Madagascar (MadIII) and South Africa (SAFII), however do represent monophyletic lineages within the "Australasia" clade. Dataset II (concatenated 18S + 28S) included a subset of the taxa in dataset I (38 taxa, 1,512 bp) and again the same general pattern, splitting the Anostostomatidae into two clades, was observed (data not shown).

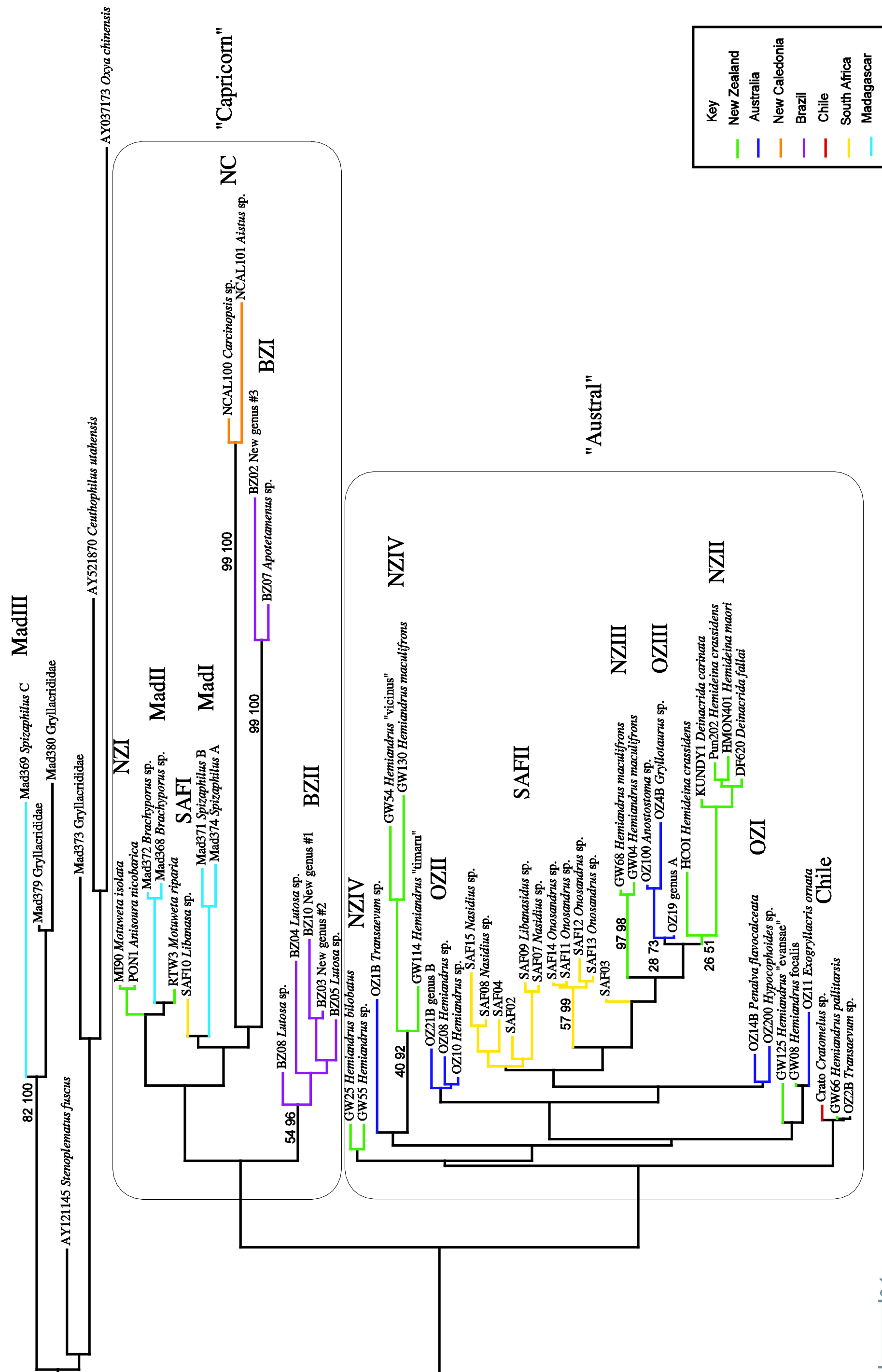


Figure 4.1 Rooted Bayesian phylogram using nuclear DNA

Rooted Bayesian phylogram for 18S. Within "Capricorn" the geographic regions are represented by New Zealand (*Anisoura-Motuweta*), New Caledonia (*Aistus*, *Carcinopsis*), Brazil (*Lutosa*, *Apotetamenus* and three new genera) and Madagascar (*Spizaphilus* A and B and *Brachyporus*). While "Austral" represents New Zealand (*Hemideina*, *Deinacrida*, *Hemiandrus*), Australia (*Anostostoma*, *Transaevum*, *Penalva*, *Gryllotaurus*, *Hemiandrus*, *Exogryllacris*, *Hypocophoides*, genus A and genus B), South Africa (*Libanasa*, Nasidius, Libanasidus and Onosandrus) and Madagascar (Spizaphilus C). Bootstrap support and Bayesian Posterior Probabilities (BPP) are listed on branches respectively.

Mitochondrial data (Dataset III – COI-RY) consisting of 1,228 bp of aligned DNA sequence data included 53 ingroup taxa plus one outgroup (NC_010219 - Acrididae). The RY-coded COI data returned clades consistent with those observed in analysis of 18S data (figure 4.2), including strong support for the two clades, "Capricorn" and "Austral". Strong bootstrap support was found for the same lineages (see support values) however, some conflict (low support) in the tree was again observed (networks not shown). The sister relationship between MadI and MadII is strongly supported with mitochondrial data as is the monophyletic grouping of NZI and BZII in the "Capricorn" clade. The "Austral" clade again supported paraphyletic lineages from New Zealand and Australia, splitting into three lineages each (NZII, NZIII, NZIV and OZI, OZII, OZIII respectively). The long-branch associated with MadIII placed it sister to OZI in this dataset. As MadIII represents a long edge associated with outgroup taxa (not included here see figure 4.1) it will not be included in further detail. Analysis of dataset IV, a reduced set of taxa with concatenated sequences, consisted of COI (1 228 bp) and 12S (451 bp) for 48 taxa (total 1 679 bp) across the geographic distribution of Anostostomatidae, and returned a topology similar to figure 4.2 with the same taxon subgroupings, and with no support for geographic monophyly, as with previous datasets (not shown).

Figure 4.2 Rooted Bayesian phylogram using mitochondrial DNA

Rooted Bayesian phylogram for COI-RY coded. As with figure 4.1 within the "Capricorn" clade the geographic regions are represented by New Zealand (*Anisoura-Motuweta*), New Caledonia (*Aistus*, *Carcinopsis*), Brazil (*Lutosa*, *Apotetamenus* and three new genera) and Madagascar (*Spizaphilus* A and B and Brachyporus). While the "Austral" clade represents New Zealand (*Hemideina*, *Deinacrida*, *Hemiandrus*), Australia (*Anostostoma*, *Transaevum*, *Penalva*, *Gryllotaurus*, *Hemiandrus*, *Exogryllacris*, *Hypocophoides* and genus A and genus B), South Africa (*Libanasa*, *Nasidius*, *Libanasidus* and *Onosandrus*) and Madagascar (*Spizaphilus* C). Bootstrap support and Bayesian Posterior Probabilities are listed on branches respectively.

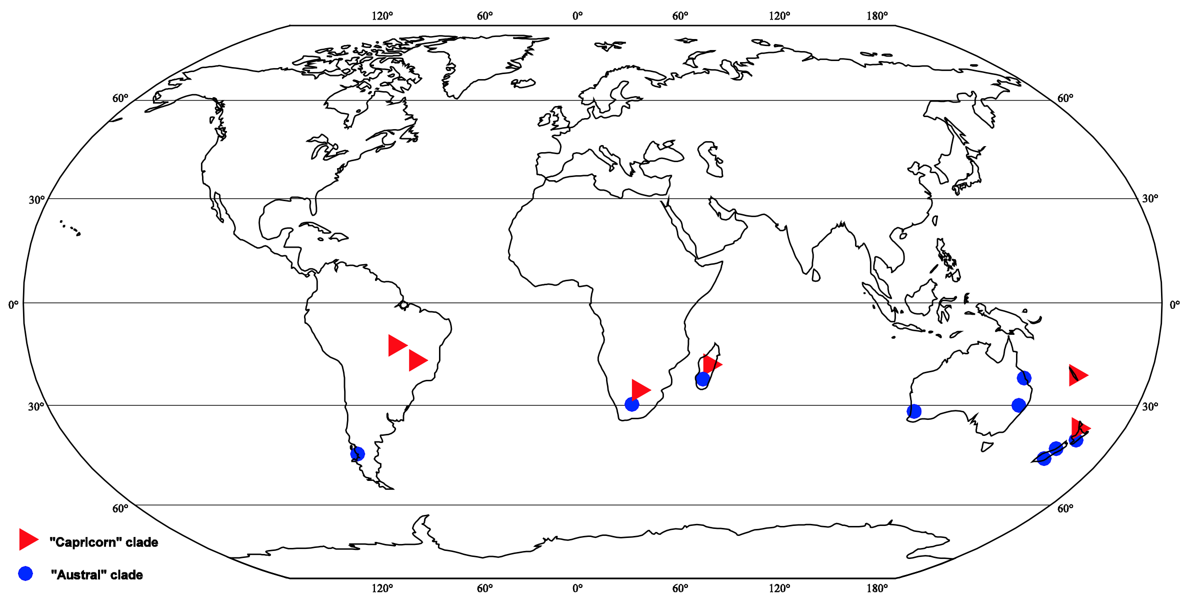


Figure 4.3 Distribution map of Anostostomatidae used in this study

Distribution map of the world highlighing the phylogenetic relationships within the Anostostomatidae. Red triangles represents the "Capricorn" clade, the northern lineage, while the "Austral" clade (blue circles) represents the more southern lineage. Note: symbols do not equate to specific collection location, merely general distribution.

(c) Stemminess and Relative Compositional Variability

The stemminess and RCV results (table 4-2) suggest that RY-coding at the third codon position was warranted for COI. While codon two also had a high RCV value, it also had a high Stemminess value and therefore it is inappropriate to RY-code it here. Other coding methods (such as site-stripping in *Chapter Three*) could be more appropriate here. Without RY-coding the third codon position sites had a higher relative compositional variability. Low stemminess compounds the influence of strong compositional bias by reducing the amount of "historical" signal competing with bias signal. By excluding the source of compositional bias in the data, the likelihood of producing a misleading phylogenetic reconstruction is reduced. Therefore improving the results, with both RCV and stemminess values, comparable to the other partitions (table 4-2).

Table 4-2 RCV and stemminess values

Relative compositional variability and stemminess among the Anostostomatidae COI sequences

Parition	RCV	Stemminess
codon 1	0.229	0.264
codon 2	0.854	0.470
codon 3	0.445	0.154
codon 3 (RY coded)	0.245	0.224

Thirty-nine trees were returned for the stemminess analyses for codon two. Calculation of stemminess values was automated using a python script (personal communication B. R. Holland) and were found to be very similar to one another (small changes at the fourth decimal place), and therefore not reported individually.

(d) Summary phylogenetics

High bootstrap support (98%) and high Bayesian posterior probability (99.8 BPP) for the separation of Anostostomatidae into two clades for dataset III - COI-RY was observed. Dataset I – 18S had low signal shown by low support (31% bootstrap, 51.4 BPP), which is not unexpected due to the slow rate of evolution for nuclear genes compared to mitochondrial genes. Datasets II and IV (18S + 28S and COI-RY + 12S respectively) returned similar results, with separation into two clades for the concatenated mitochondrial data, while the much reduced (taxa wise) combined nuclear data was less well resolved (due to long edges introduced because of reduced taxon sampling). It should be noted that the root of the tree was found between the two clades of Anostostomatidae. With a consistent split between taxa in both mitochondrial and nuclear datasets we therefore consider the two clades as probable subfamilies. In addition, the monophyly of geographic regions is rejected, with the exception of New Caledonia (see figures 4.1 and 4.2 lineage NC).

4.5 DISCUSSION

Molecular phylogenetic studies are important tools that are widely used to both generate and test biogeographic hypotheses. Important in this endeavour is the systematic framework (on which to begin testing) to ensure spatial (geographic) coverage of the group being tested and to avoid paralogous sampling. Taxonomic classification within Anostostomatidae has not been reliable, with a number of poorly resolved or contradictory classifications (Gwynne 1995; Johns 1997; Gorochov 2001b; Desutter-Grandcolas 2003). Designation at the tribal and even subfamilial level varies, with several attempts at classification being carried out, with noticeable differences in their conclusions (eg. OSF online resource Eades and Otte; but see Johns 1997; Gorochov 2001b) making it difficult to know if analogous taxa are being compared. Results from the current research consistently supported Anostostomatidae being separated into two widely distributed subgroups (clades) that subsume the seven existing subfamilies (OSF online Eades and Otte) as indicated by the position of the root in between "Capricorn" and "Austral" clades. This finding, supporting two clades throughout the family covering a wide geographic distribution, is evident from both nuclear and mitochondrial datasets (including amino acids). Although there was an emphasis on sampling in the Australasian region, the pattern is supported by previous findings (see Brettschneider 2006; Pratt et al. 2008) where Brettschneider (2006) reported a split within South African taxa making the genus *Libanasa* sister to all other South African Anostostomatidae fauna (corresponding to "Capricorn" and "Austral" here). Current analysis supports this split (see figure 4.1 SAFI) however the scale of separation (and its implications) as shown here, could not be appreciated without the perspective of broader spatial and taxonomic sampling.

At the outset, improving the understanding of systematics and biogeographic patterns within Anostostomatidae was the goal. By using molecular data, previous findings could be tested on a larger (family) scale. Analysis of slowly evolving nuclear genes (18S and 28S), and more rapidly evolving mitochondrial genes (COI and 12S), revealed concordant patterns. Most striking was the paraphyly revealed within regions, and the recognition that geographic proximity (of landmasses) does not necessarily equate to phylogenetic proximity. As suggested above, sampling bias may play a role in the patterns observed.

This is not to say the pattern is insignificant, simply that until further sampling is incorporated (ie. breaking up long branches), the patterns shown here may result from paralagous sampling.

To explain the phylogenetic relationships in a biogeographic context, regions will be considered separately. Australasia has previously been shown to harbour a diverse array of anostostomatid taxa. The inclusion of taxa from a broader geographic range doesn't change these findings (see Pratt et al. 2008), indicating that both New Zealand and Australia house distinct, and diverse anostostomatid fauna. This finding is intriguing and unexpected considering the geological history of these two regions. At the generic level, this study included all of the proposed diversity of Australian genera, which all fell into the "Austral" clade (although not as a monophyletic group). Additional species level sampling within Australian taxa will be required to fully resolve these patterns. In contrast, New Zealand has representatives in both major clades. New Zealand taxa of the "Austral" clade are paraphyletic with a lineage including all representatives of *Deinacrida-Hemideina* (NZII) and two *Hemiandrus* lineages, NZIII and NZIV (see Pratt et al. 2008 for discussion on *Hemiandrus*), that do not appear to be sisters.

Africa and Madagascar are geographic neighbours but share no genera sampled here (although see Appendix 4.8.1 table 4-3). Accordingly, no sister relationships were found in the current analysis between these geographic regions. South African taxa occurred in both major clades (as stated above) with most generic diversity observed in the "Austral" clade. Madagascan taxa were also found in both clades, though the majority of taxa currently sampled fell into the "Capricorn" clade. Interestingly, one genus, *Spizaphilus* appears in both clades. This is most likely a taxonomic misidentification which needs to be investigated further.

Lastly, two lineages from South America were included; encompassing five Brazilian genera and one genus from Chile. The Chilean *Cratomelus* was consistently placed in the "Austral" clade while Brazilian taxa were found only in the "Capricorn" clade. Within "Capricorn", two lineages were observed and appear sister to the rest of the "Capricorn" taxa. Additional species and genera from Central and South America will be useful for future studies to help alleviate potential long-branch issues.

Although many landmasses do not have monophyletic Anostostomatidae faunas, phylogeographic structure within each clade was found. This pattern may be interpreted in a number of ways: the observed pattern may suggest that dispersal has occurred (a sporadic phenomenon); it may be indicative of sampling error due to local extinction or missing data in general (as sister taxa may not have been included in our sampling); or it might indicate old links pre-dating Gondwana breakup. Without dating estimates, differentiation between these hypotheses is not possible, although comment can be made on the predicted patterns if the two extreme hypotheses, as suggested earlier, are considered. Even if we consider the two clades independently, there appears to be no support for a vicariant history due to Gondwanan breakup (which would lead to the expectation of South African taxa being basal to all other lineages). Even so, dispersal along known paths (such as the WWD and ACC) does not appear to explain the current taxonomic links either.

That said, it is reasoned that an organism with strong migratory ability will have the potential to undergo major range expansions (e.g. long distance dispersal), and as such would tend to show higher levels of gene flow, preventing genetic divergence among populations. Given the genetic diversity detected between South Africa and Australia (less than 17.88% COI and less than 3.27% 18S), between Madagascar and South Africa (less than 19.51% COI and less than 3.17% 18S), New Zealand and New Caledonia (less than 18.69% COI and less than 3.14% 18S), Australia and New Zealand (less than 18.74% COI and less than 1.6% 18S), New Zealand and Brazil (less than 18.74% COI and less than 3% 18S), it seems likely that dispersal has occurred at some point during the history of the two clades, yet it is not frequent enough to cause a loss in geographic regionalism. Increased sampling could reveal additional mtDNA lineages and thus decrease the divergence estimate (Waters and Roy 2004).

Finally, a comment on the implications these findings have on the family level systematics. Currently, there are seven recognised subfamilies including 42 genera (see Appendix 3.8.1. Eades and Otte). This study includes more than half of these genera (28) with representatives from the five most diverse subfamilies. The two subfamilies not included were Euclydesinae (one genus with one species; *Euclydes* [extinct] from Brazil), and Leiomelinae (one genus with four species; *Leiomelus* from Chile).

Of the included subfamilies; Anostostomatinae (14 genera, ten included), four genera were found in the "Capricorn" clade and six in the "Austral" clade. In the Lutosinae (seven genera, three included), two genera were found in the "Capricorn" clade and one in the "Austral" clade. Anabropsinae (five genera, one included), Cratomelinae (one genus, one included), and Deinacridinae (two genera, two included), were all placed in the "Austral" clade. In addition, five genera (*Aistus*, *Anisoura* and New genus #1, #2 and #3) were found in the "Capricorn" clade, while five other genera (*Hemiandrus*, *Hypocophoides*, *Transaevum* and genus A and genus B) were in the "Austral" clade. With this in mind, the molecular results indicate that a reduction in the number of subfamilies within Anostostomatidae would be appropriate. Determining which of the subfamily names to retain will need to be addressed with both taxonomic and molecular expertise and will be addressed in a future publication.

Clearly, more taxonomic and molecular work needs to be carried out to fully resolve these relationships as our study is by no means conclusive. However, it forms a solid foundation which is hoped will encourage further research into this fascinating group. In addition to more thorough sampling from South America, future studies should include samples from India and a more complete sampling of African and Madagascan taxa, so questions regarding the direction of movement of proto India can be investigated (Briggs 2003).

The current understanding of southern hemisphere geology does not predict the distribution of the two major clades within Anostostomatidae. Neither continental fragmentation nor recent dispersal, appear to be useful explanations for the evolutionary patterns observed here. The current analyses suggest that the two clades represent two subfamilies, and should be considered separate datasets. This hypothesis needs to be considered carefully as the impact of sampling error (distributional noise) and/or multiple overlying signals (produced by different adaptive responses and ecologies) may be influencing the results. If the dynamic history of the southern lands is to be untangled a taxonomic revision of the family will be required. It is important to remember that absence of taxa from areas is not as informative as their presence.

4.6 ACKNOWLEDGEMENTS

This study would not have been possible without the donation of specimens from around the world. In particular a special thank you to Mohsen Mofidi-Neyestanak and George Beccaloni (Madagascan genera), Helene Brettschneider and Bill Bateman (South African genera) and Franciso de A. G. de Mello (Brazilian genera) for providing samples. Thank you also to all those involved over the years closer to home, namely Geoff Monteith, Peter Johns, George Gibbs, Darryl Gwynne, Scott Dunavan and the Department of Conservation, New Zealand. This research was supported by funding from the Marsden Fund granted to David Penny and the Allan Wilson Centre for Molecular Ecology and Evolution.

4.7 REFERENCES

Brettschneider, H. 2006. Systematics of southern African Anostostomatidae (Orthoptera) based on morphological and molecular data. Pp. 147. Natural and Agricultural Sciences. University of Pretoria, Pretoria.

Briggs, J. C. 2003. The biogeographic and tectonic history of India. Journal of Biogeography **30**:381-388.

Desutter-Grandcolas, L. 2003. Phylogeny and evolution of acoustic communication in extant Ensifera (Insecta, Orthoptera). Zoologica Scripta **32**:525-561.

Donoghue, M. J., and B. R. Moore. 2003. Toward an integrative historical biogeography. Integr. Comp. Biol. **43**:261-270.

Eades, D. C., and D. Otte. Orthoptera Species File Online.version. retrieval date. http://Orthoptera.SpeciesFile.org

Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39**:783-791.

Fiala, K. L., and R. R. Sokal. 1985. Factors determining the accuracy of cladogram estimation: evaluation using computer simulation. Evolution **39**:609-622.

Folmer, O., M. Black, W. Hoeh, R. Lutz, and R. Vrijenhoek. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Molecular Marine Biology and Biotechnology **3**:294-299.

Goldberg, J., S. A. Trewick, and A. M. Paterson. 2008. Evolution of New Zealand's terrestrial fauna: a review of molecular evicence. Philosophical Transactions of the Royal Society of London B Biological Sciences **363**:3319-3334.

Gorochov, A. V. 2001. The higher classification, phylogeny and evolution of the superfamily Stenopelmatiodea. Pp. 3-33 *in* L. H. Field, ed. The biology of wetas, king crickets and their allies. CABI Publishing, Wallingford.

Gwynne, D. T. 1995. Phylogeny of the Ensifera (Orthoptera): A hypothesis supporting multiple origins of acoustical signalling, complex spermatophores and maternal care in crickets, katydids, and weta. Journal of Orthoptera Research **4**:203-218.

Huelsenbeck, J. P., and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics **17**:754-755.

Hunn, C. A., and P. Upchurch. 2001. The importance of time/space in diagnosing the causality of phylogenetic events: towards a "chronobiogeographical" paradigm? Systematic Biology **50**:391-407.

Johns, P. M. 1997. The Gondwanaland Weta: Family Anostostomatidae (formerly in Stenopelmatidae, Henicidae or Mimmermidae): Nomenclatural Problems, World Checklist, New Genera and species. Journal of Orthoptera Research **6**:125-138.

Johnson, K. P., F. R. Adler, and J. L. Cherry. 2000. Genetic and phylogenetic consequences of island biogeography. Evolution **54**:387-396.

Lieberman, B. S. 2002. Phylogenetic biogeography with and without the fossil record: guaging the effects of extinction and paleontological incompleteness. Palaegeography, Palaeoclimatology, Palaeoecology **178**:39-52.

Martins-Neto. 2007. New Orthoptera Stenopelmatoidea and Hagloidea (Ensifera) from the Santana Formation (Lower Cretaceous, Northeast Brazil) with description of new taxa. . GAEA **3**:3-8.

Meads, M. J. 1990. The Weta Book: A guide to the identification of wetas. DSIR Land Resources, Lower Hutt.

Phillips, M. J., Y.-H. Lin, G. L. Harrison, and D. Penny. 2001. Mitochondrial genomes of a bandicoot and a brushtail possum confirm the monophyly of australiadelphian marsupials. Proceedings of the Royal Society of London B Biological Sciences **268**:1533-1538.

Phillips, M. J., and D. Penny. 2003. The root of the mammalian tree inferred from whole mitochondrial genomes. Molecular Phylogenetics and Evolution **28**:171-185.

Phillips, M. J., and R. C. Pratt. 2008. Family-level relationships among the Australasian marsupial "herbivores" (Diprotodontia: Koala, wombats, kangaroos and possums). Molecular Phylogenetics and Evolution **46**:594-605.

Posada, D., and T. R. Buckley. 2004. Model selection and model averaging in phylogenetics: Advantages of Akaike Information Criterion and Bayesian approaches over Likelihood Ratio tests. Systematic Biology **53**:793-808.

Posada, D., and K. A. Crandall. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics **14**:817-818.

Pratt, R. C., M. Morgan-Richards, and S. A. Trewick. 2008. Diversification of New Zealand weta (Orthoptera: Ensifera: Anostostomatidae) and their relationships in Australasia. Philosophical Transactions of the Royal Society of London B Biological Sciences.

Rambaut, A. 1996. Se-Al: Sequence Alignment Editor. Available at http://evolve.zoo.ox.ac.uk.

Ronquist, F. 1997. Dispersal-vicariance analysis: A new approach to the quantification of historical biogeography. Systematic Biology **46**:195-203.

Simon, C., F. Frati, A. Bechenbach, B. Crespi, H. Liu, and P. Flook. 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Annals of the Entomological Society of America **87**:651-701.

Swofford, D. L. 2003. PAUP*: Phylogenetic Analysis Using Parsimony (and Other Methods), Version 4. Sinauer Associates, Sunderland, Massachusetts.

Toms, R. B. 2001. South African King Crickets (Anostostomatidae). Pp. 73-78 *in* L. H. Field, ed. The biology of wetas, King crickets and their allies. CABI Publishing, New York.

Trewick, S. A., and M. Morgan-Richards. 2004. Phylogenetics of New Zealand's tree, giant and tusked weta (Orthoptera: Anostostomatidae): evidence from mitochondrial DNA. Journal of Orthoptera Research **13**:185-196.

Trewick, S. A., and M. Morgan-Richards. 2005. After the deluge: mitochondrial DNA indicates Miocene radiation and Pliocene adaptation of tree and giant weta (Orthoptera: Anostostomatidae). Journal of Biogeography **31**:295-309.
Vawter, L. 1991. Evolution of blattoid insects and of the small subunit ribosomal gene. University of Michigan, Ann Arbor.
Waters, J. M., and M. S. Roy. 2004. Out of Africa: the slow train to Australasia. Systematic Biology **53**:18-24.
Whiting, M. F. 2002. Mecoptera is paraphyletic: multiple genes and phylogeny of Mecoptera and Siphonaptera. Zoologica Scripta **31**:93-104.
Zwickl, D. J. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. The University of Texas, Austin.

4.8 Appendices

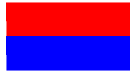
4.8.1 Orthoptera Species File Online: taxa hierarchy as of 250908 (Eades and Otte)

Table 4-3 OSF online Anostostomatidae hierarchy

Order:	Orthoptera		
Suborder:	Ensifera		
Superfamily:	Stenopelmatioidea	Burmeister, 1838	
Family:	Anostostomatidae	Saussure, 1859	

	Genus		Suggested Distribution
Subfamily:	Anabropsinae	Rentz, D. C. F. & Weissman, 1973	
Genus:	*Anabropsis*	Rehn, 1901	Congo, Equador, Mexico, Costa Rica, Guatemala
	Apteranabropsis	Gorochov, 1988	Vietname, China, Nepal
	Exogryllacris	Willemse, 1963	Australia
	Leponosandrus	Gorochov, 2001	Australia?
	Paterdecolyus	Griffini, 1913	Madagascar, India, Vietnam, Japan
Subfamily:	Anostostomatinae	Saussure, 1859	
Genus:	*Anostostoma*	Gray, 1837	Australia, Mexico
	Bochus	Péringuey, 1916	South Africa
	Borborothis	Brunner von Wattenwyl, 1888	South Africa, Angola
	Brachyporus	Brunner von Wattenwyl, 1888	South Africa, Madagascar
	Carcinopsis	Brunner von Wattenwyl, 1888	New Caledonia
	Gryllotaurus	Karny, 1929	Australia
	Henicus	Gray, 1837	South Africa, Angola, Zimbabwe, Tanzania
	Libanasidus	Péringuey, 1916	South Africa
	Motuweta	Johns, 1997	New Zealand
	Nasidius	Stål, 1878	South Africa, Angola, Zimbabwe, Tanzania
	Onosandridus	Péringuey, 1916	South Africa, Angola, Congo, Zimbabwe
	Onosandrus	Stål, 1878	South Africa
	Penalva	Walker, 1870	Australia
	Spizaphilus	Kirby, 1906	Madagascar
Subfamily:	Cratomelinae	Brunner von Wattenwyl, 1888	
Genus:	*Cratomelus*	Blanchard, 1851	Chile
Subfamily:	Deinacridinae	Karny, 1932	
Genus:	*Deinacrida*	White, 1842	New Zealand
	Hemideina	Walker, 1869	New Zealand, Australia?, Chile?
Subfamily:	Euclydesinae	Martins-Neto, 2007	
Genus:	*Euclydes*	Martins-Neto, 2007	Brazil
Subfamily:	Leiomelinae	Gorochov, 2001	
Genus:	*Leiomelus*	Ander, 1936	Chile, Argentina, Uraguay, Paraguay

	Genus		**Suggested Distribution**
Subfamily:	Lutosinae	Gorochov, 1988	
Genus:	*Apotetamenus*	Brunner von Wattenwyl, 1888	Colombia, Ecuador, Peru, Bolivia, Brazil
	Hydrolutos	Issa & Jaffe, 1999	Venezuela
	Libanasa	Walker, 1869	South Africa, Tanzania
	Licodia	Walker, 1869	Cuba, Haiti, Dominican Republic
	Lutosa	Walker, 1869	Mexico, Cuba, Brazil, Paraguay
	Neolutosa	Gorochov, 2001	Brazil, Panama, Costa Rica, Nicaragua, Hinduras, El Salvador, Guatemala, Belize
	Papuaistus	Griffini, 1911	New Guinea, Indonesia
Tribe:	Glaphyrosomatini	Rentz, D. C. F. & Weissman, 1973	
Genus:	*Cnemotettix*	Caudell, 1916	Mexico, USA
	Glaphyrosoma	Brunner von Wattenwyl, 1888	Guatemala, Mexico
	Aistus	Brunner von Wattenwyl, 1888	New Caledonia
	Anisoura	Ander, 1932	New Zealand
	Coccinellomima	Karny, 1932	unknown
	Dolichochaeta	Philippi, 1863	Chile, Argentina, Paraguay, Uraguay
	Gryllacropsis	Brunner von Wattenwyl, 1888	India
	Hemiandrus	Ander, 1938	New Zealand, Philippines
	Hypocophoides	Karny, 1930	India
	Hypocophus	Brunner von Wattenwyl, 1888	Madagascar
	Transaevum	Johns, 1997	Australia



"Capricorn" clade

"Austral" clade

4.8.2 Sample data including name, location, accession number, and genes used.

Table 4-4 Sample information

All specimens except those from Madagascar are housed at the AWC, Palmerston North, Massey University (Contact S. Trewick). Madagascan samples are held at the Imperial College London (Contact M. Mofidi-Neyestanak).

				GeneBank/EMBL accession numbers				dataset			
species	specimen ID	country	locality	18S	28S	COI	12S	I	II	III	IV
Carcinopsis sp.	NCal 100	New Caledonia	Aoupinie	EU676733	EU676696	EU676751	EU676669	✓	✓	✓	✓
Aistus sp.	NCal 101	New Caledonia	Table Unio road	EU676734	EU676695	EU676752	EU676670	✓	✓	✓	✓
Transaevum sp.	OZ1B	Australia	Mt Finnigan	EU676723	EU676699	EU676754	EU676672	✓	✓	✓	✓
Transaevum sp.	OZ2B	Australia	The Crater	FJ481474	EU676700			✓	✓		
Gryllotaurus sp.	OZ4B	Australia	Cape Tribulation	EU676724	EU676701	EU676755	EU676673	✓	✓	✓	✓
Hemiandrus sp.	OZ08	Australia	Bellenden Ker summit	EU676725	EU676697	EU676756	EU676674	✓	✓	✓	✓
Hemiandrus sp.	OZ10	Australia	Mt. Finnigan summit	EU676726	EU676702	EU676757	EU676675	✓	✓	✓	✓
Exogryllacris ornata	OZ11	Australia	Bartle Frere	EU676727	EU676703	EU676758	EU676676	✓	✓	✓	✓
Penalva flavocalceata	OZ14B	Australia	Bartle Frere	EU676728	EU676704	EU676759	EU676677	✓		✓	✓
Genus A	OZ19	Australia	Mt. Kooroomool	EU676729	EU676698	EU676760	EU676678	✓	✓	✓	✓
Genus B	OZ21B	Australia	Bartle Frere	EU676730	EU676705	EU676761	EU676679	✓	✓	✓	✓
Anostostoma sp.	OZ100	Australia	Peachester	EU676731	EU676706	EU676762	EU676680	✓	✓	✓	✓
Hypocophoides sp.	OZ200	Australia	Burekup, Western Australia	EU676732	EU676707			✓	✓		
Penalva sp.	OZ202	Australia	Bawley Point, New South Wales			EU676763				✓	
Hypocophoides sp.	SAW.OZ	Australia	Perth, Western Australia			EU676753	EU676671			✓	✓
Deinacrida carinata	Kundy1	New Zealand	SI, Kundy Island	EU676711	EU676684	EU676737	EU676658	✓	✓	✓	✓
Hemideina maori	HMons401	New Zealand	SI, Porters Pass	EU676708	EU676685	EU676736	EU676657	✓	✓	✓	✓
Hemideina crassidens	HCOI	New Zealand	NI, Palmerston North	EU676709		FJ481427	FJ481450	✓		✓	✓
Hemideina crassidens	Pun202	New Zealand	SI, Punakaikai	EU676712	FJ481507	EU676738	EU676659	✓	✓	✓	✓
Deinacrida fallai	DF620	New Zealand	NI, Poor Knights	EU676710		EU676739	EU676660	✓		✓	✓
Motuweta isolata	MI90	New Zealand	NI, Middle Island, Mecury Islands	EU676720	EU676689	EU676748	EU676666	✓	✓	✓	✓
Motuweta riparia	RTW3	New Zealand	NI, Motu River	EU676721	EU676687	EU676750	EU676668	✓	✓	✓	✓
Anisoura nicobarica	PON1	New Zealand	NI, Hokianga	EU676722	EU676688	EU676749	EU676667	✓	✓	✓	✓
Hemiandrus sp.	GW55	New Zealand	SI, Marfell's Beach	fj481470	FJ481506	EU676789	FJ481449	✓	✓	✓	✓
Hemiandrus focalis	GW08	New Zealand	SI, Conical Hill	FJ481468		EU676773	FJ481446	✓		✓	✓
Hemiandrus bilobatus	GW25	New Zealand	NI, Wellington	EU676714	FJ481505	EU676794	FJ481447	✓	✓	✓	✓
Hemiandrus "vicinus"	GW54	New Zealand	SI, Whites Bay	FJ481469		EU676788	FJ481448	✓		✓	✓
Hemiandrus pallitarsis	GW66	New Zealand	NI, Lake Waikaremoana	FJ481471		EU676740	EU676661	✓		✓	✓
Hemiandrus "timaru"	GW114	New Zealand	SI, Kurinui	FJ481472	EU676693	EU676769		✓	✓	✓	
Hemiandrus "disparalis"	GW124	New Zealand	SI, near Dobson	EU676718				✓			
Hemiandrus "evansae"	GW125	New Zealand	SI, Portabello	EU676719	EU676694			✓	✓		
Hemiandrus maculifrons	GW04	New Zealand	NI, Gisborne	EU676713	FJ481504	EU676798	FJ481445	✓	✓	✓	✓
Hemiandrus maculifrons	GW93A	New Zealand	SI, Pelorus			EU676791				✓	
Hemiandrus maculifrons	GW104	New Zealand	SI, Takaka			EU676742	EU676662			✓	✓
Hemiandrus maculifrons	GW130	New Zealand	SI, Craigieburn Range	FJ481473	EU676691	EU676745	EU676663	✓	✓	✓	✓

species	specimen ID	country	locality	18S	GeneBank/EMBL accession numbers 28S	COI	12S	I	dataset II	III	IV
Brachyporus	Mad368	Madagascar	Ranomafana National Park	FJ481475		FJ481428	FJ481451	✓		✓	✓
Brachyporus	Mad370	Madagascar	Ranomafana National Park			FJ481430	FJ481453			✓	✓
Brachyporus	Mad372	Madagascar	Ranomafana National Park	FJ481478		FJ481432	FJ481455	✓		✓	✓
Brachyporus	Mad378	Madagascar		FJ481481				✓			
Spizaphilus A	Mad374	Madagascar	Ranomafana National Park	FJ481480		FJ481433	FJ481456	✓		✓	✓
Spizaphilus B	Mad371	Madagascar	Ranomafana National Park	FJ481477	FJ481508	FJ481431	FJ481454	✓	✓	✓	✓
Spizaphilus C	Mad369	Madagascar	Ranomafana National Park	FJ481476		FJ481429	FJ481452	✓		✓	✓
Gryllacrididae?	Mad373	Madagascar	Ranomafana National Park	FJ481479				✓			
Gryllacrididae?	Mad380	Madagascar	Ranomafana National Park	FJ481482				✓			
Libanasidus?	SAF01	South Africa	Johannesburg			FJ481439	FJ481462			✓	✓
	SAF02	South Africa	Johannesburg	FJ481490				✓			
Libanasidus?	SAF03	South Africa	Johannesburg	FJ481491	FJ481510	FJ481440	FJ481463	✓	✓	✓	✓
	SAF04	South Africa	Johannesburg	FJ481492	FJ481511			✓	✓		
Libanasidus	SAF06	South Africa	Limpopo Province, Tzaneen			FJ481441	FJ481464			✓	✓
Libanasidus	SAF09	South Africa	Gauteng Province, Klipreviersberg Nat. Res	FJ481495				✓			
Nasidius	SAF07	South Africa	Gauteng Province, Klipreviersberg Nat. Res	FJ481493	FJ481512			✓	✓		
Nasidius	SAF08	South Africa	Cape Bosbokstrand	FJ481494	FJ481513	FJ481442	FJ481465	✓	✓	✓	✓
Nasidius	SAF15	South Africa	Gauteng Province, Klipreviersberg Nat. Res	FJ481501	FJ481518			✓	✓		
Onosandrus	SAF11	South Africa	Mpumalanga Province, Bridal veil falls	FJ481497	FJ481514			✓	✓		
Onosandrus	SAF12	South Africa		FJ481498	FJ481515			✓	✓		
Onosandrus	SAF13	South Africa	KwaZulu-Natal Province, Ngome Forest	FJ481499	FJ481516	FJ481443	FJ481466	✓	✓	✓	✓
Onosandrus	SAF14	South Africa	Mpumalanga Province, Bridal veil falls	FJ481500	FJ481517			✓	✓		
Libanasa	SAF10	South Africa	KwaZulu-Natal Province, Ngome Forest	FJ481496				✓			
Libanasa	SAF16	South Africa	Cape Province, Constantia			FJ481444	FJ481467			✓	✓
New genus #1	BZ01	Brazil	State of S-O Paulo, Sales-polis			FJ481434	FJ481457			✓	✓
New genus #3	BZ02	Brazil	State of Esp'rito Santo, Linhares	FJ481483		FJ481435	FJ481458	✓		✓	✓
New genus #2	BZ03	Brazil	State of S-O Paulo, Ubatuba	FJ481484		FJ481436	FJ481459	✓		✓	✓
Lutosa sp. #1	BZ04	Brazil	State of S-O Paulo, Sales-polis	FJ481485	FJ481509	FJ481437	FJ481460	✓	✓	✓	✓
Lutosa sp. #2	BZ05	Brazil	State of S-O Paulo, Sales-polis	FJ481486				✓			
Lutosa sp. #3	BZ06	Brazil	State of S-O Paulo, Campos do Jord-o							✓	
Apotetamenus sp.	BZ07	Brazil	State of S-O Paulo, Piracicaba	FJ481487				✓			
Lutosa sp. #3	BZ08	Brazil	State of Esp'rito Santo, Linhares	FJ481488				✓			
New genus #1	BZ10	Brazil	State of S-O Paulo, Sales-polis	FJ481489		FJ481438	FJ481461	✓		✓	✓
Cratomelus sp.	Crato	Chile		FJ481502				✓		✓	
Outgroup											
Gryllacrididae	Mad379	Madagascar	Ranomafana National Park	EU676735				✓			
Stenoplematus fuscus (Stenopelmatidae)				AY121145				✓			
Conocephalus sp. (Tettigoniidae)				AF514568				✓			
Ceuthophilus utahensis (Rhaphidophoridae)				AY521870				✓			
Oxya chinensis (Acrididae)				AY037173	AF416876	NC_010219	NC_010219	✓	✓	✓	✓

CHAPTER FIVE

One never notices what has been done; one can only see what remains to be done

–Marie Curie

Summation

5.1 EVIDENCE FOR ABIOTIC EFFECTS?

The aim of this thesis was to use molecular data to determine the phylogenetic relationships of Anostostomatid crickets and Neoavian birds in order to assess the relationship (if any) between biotic diversity and three notable geological events: (1) the Oligocene marine inundation *ca* 30 Ma, (2) the asteroid impact 65 Ma, and (3) the break up of Gondwana *ca* 160 – 70 Ma. Three different kinds of process are known to modify the geographical spatial arrangement of organisms: dispersal, extinction and vicariance. The thesis examines the relative importance of dispersal versus vicariance in the Anostostomatid crickets in relation to events (1) and (3) (*Chapter Two* and *Four* respectively), and the importance of the proposed extinction at the K – Pg boundary to diversification in Neoaves (*Chapter Three*). An estimate of the timing of diversification is crucial to testing whether these processes (dispersal, extinction and vicariance) provide an explanation for species diversification, therefore we included molecular dating methods in the analyses whereever possible.

It is clear that the underlying processes acting on biotic diversity in the Southern Hemisphere are complicated. The taxa investigated in this thesis were chosen for their broad southern links and for their potential to test the impact of three abiotic events (Oligocene marine inundation, K - Pg asteroid impact and Gondwanan breakup). Within the taxa studied here, molecular results do not support a clear link between the suggested 'stimuli' provided by the events and the observed timing of diversification. For simplicity, the two case studies involving Anostostomatidae (*Chapter Two* and *Chapter Four*) will be discussed first, followed by a discussion of Neoaves (*Chapter Three*).

This thesis is the first molecular study looking at the Anostostomatidae family. *Chapters Two* and *Four* tested the importance of two different geological events (Oligocene marine inundation and Gondwanan breakup) that occurred across very different timescales, which meant that different taxon sampling was required in each chapter. The two phylogenetic tree topologies found were largely congruent with each other despite the broader taxon sampling in *Chapter Four*, with both trees showing two main clades.

One of the clades contained the New Zealand tusked weta and New Caledonian, Brazilian and some Madagascan and South African taxa – this was termed the "Capricorn" clade. The scond clade, termed "Austral", contained all other New Zealand taxa, all Australian taxa, the remaining Madagascan and South African taxa, and a sample from Chile.

The patterns found are not easily explained by vicariance or dispersal alone, and might be indicative of broader taxonomic discrepancies of the family such as historic mis-identifications due to convergent evolution. These findings are important for a number of reasons. Anostostomatids are an ancient lineage pre-dating the breakup of Gondwana. An intriguing fossil found in 190 million year old deposits in Queensland, Australia (Meads 1990), and more recently an extinct subfamily from Brazil *ca* 112 million years old (Martins-Neto 2007), certainly indicate the possibility of the family being passively transported throughout the region along with continental plates. However, along with climatic changes seen worldwide, some regions in the Southern Hemisphere (most notably New Zealand and New Caledonia) have undergone extensive geological change which is thought to have caused massive extinction events (Campbell and Hutching 2007; Trewick, Paterson, and Campbell 2007).

It has been suggested (see Landis et al. 2008) that New Zealand's geological history resulted in complete (or almost complete) biotic turnover, therefore any lineages present (and current diversity) should be no older than Miocene in age (due to the Oligocene marine inundation). However current research does not support this, instead finding unexpected (and unpredicted) high genetic diversity within anostostomatid taxa (*Chapter Two*). Molecular dating suggested a number of anostostomatid lineages were present before or shortly after the breakup between Australia (and the rest of Gondwana) and Zealandia (New Zealand and New Caledonia), with additional recent dispersal and colonisation in the last few million years. This finding, that both vicariance and dispersal have played a role in the lineage diversity of Anostostomatid crickets, is in contrast to many recent molecular studies on New Zealand taxa (Cook and Crisp 2005; Knapp et al. 2005; Goldberg, Trewick, and Paterson 2008 for review) which have only found a role for dispersal.

These findings are consistent with previous molecular studies on Kauri (Stöckler, Daniel, and Lockhart 2002; Knapp et al. 2007) supporting the idea that some low lying land was present at the height of the Oligocene marine inundation in the Zealandia region.

Recent research conducted in the South Island of New Zealand has reported new fossils that support a diverse array of taxa present after the Oligocene. For example, Worthy et al. (2009) report findings from the Early Miocene (19 – 16 Ma) which represent a diverse array of fauna, in many ways similar to the present one (in New Zealand). The St Bathan's fossil collection includes birds, skinks, geckoes, bats, frogs, fish and a crocodilian. Three fragmentary specimens indistinguishable from modern tuatara are also found. Fossilised frog bones are rare but two vertebrae from a leiopelmatid frog have been found. The diversity discovered so far supports endemism at the species and generic level which Worthy et al. (2009) interpret to reflect a long *in situ* faunal evolution resultant from an initial vicariant event, supplemented by ongoing dispersal.

Similarly, fossils recently discovered in Otago and Southland by Lee et al. (2009) support the idea of a diverse biota. They report on fossils found in deposits spanning key time intervals representing the Late Eocene, Oligocene and Early Miocene in age. They conclude that there is evidence for a long continuity in numerous New Zealand species (plants, fish, insects). Along with these fossils, important evidence for past climatic conditions (found in the sediments) allow paleoenvironments to be determined, enabling flora and fauna to be placed in an ecological context. Prior to these reports there were few fossils older than a million years reported from New Zealand. With such a diverse array of identifiable fauna and flora there is a good chance that other habitat types and biota were present (although perhaps not as likely to fossilize).

Fossil finds such as these along with new techniques that allow collection of fossils from now sunken land (ie. seamounts and the sunken Zealandian plateau in general) could provide definitive evidence for continuous land in this region. Extensive erosion of old surfaces (along with tectonic activity) has led to the opinion amoungst geologists that it was unlikely that there was a continued presence of land from the Gondwanan breakup to the present (e.g. Landis et al. 2008).

With the Zealandian region being so active (volcanically), it is not unreasonable to believe that there was some land or an archipelago of islands (such as Hawai'i is now) that was able to support some endemic biota surviving from Gondwanan times.

Complimentary to further fossil finds, more could be learnt about whether land persisted from molecular phylogenetic studies that include dating information. In this regard, and due to New Zealand's lack of endemics such as mammals, I think that the most promising taxa to look at will be endemic insects. Insects represent an ancient group, they tend to be small (compared to say mammals), meaning they require less area to survive and can be easily transported by wind between land as would be necessary if they were to move between islands of a changing archipelago (although this is not always the case as just because they have the ability doesn't mean they do). Their short generation time, ability to undergo metamorphosis and in some cases ability to go into a sort of hibernation (diapause), may make them less susceptible to changes in environment. Unlike iconic New Zealand species such as tuatara and leiopelmatid frogs, many insect groups have close relatives on other landmasses making it easier to calibrate divergence times.

Another area where further molecular phylogenetic work will be required is in relation to the observed increase in mutation rate in the clade that included the tusked weta from New Zealand and New Caledonia (see *Chapter Two*), a subset of the "Capricorn" clade (*Chapter Four*). This finding is very interesting as a lineage rate change had never been suggested for this group. To determine if this is simply an issue with the genes used here, or if it represents an overall increase in mutation rate in these taxa, more genes will need to be studied.

The taxa studied in *Chapter Two* represent the complete generic diversity from New Zealand, New Caledonia and Australia. The grouping of the New Caledonian taxa with the New Zealand tusked weta without an associated Australian taxon was surprising. New Caledonia, like New Zealand, has had an active geological history (Gondwanan breakup of Zealandia, Oligocene marine inundation and extensive uplifting of ultramafic rocks in the Late Eocene) and is therefore not necessarily predicted to harbour old lineages (Murienne et al. 2005; Grandcolas et al. 2008).

Chapter Four looked at this group again but with much broader taxon sampling, the group of New Zealand tusked weta and New Caledonian taxa appears again but with the addition of taxa from Madagascar, Brazil and South Africa – dubbed the “Capricorn” clade. This fundamental split was also found in the study by Brettschneider (2006) of South African taxa, but its significance wasn’t recognised until it was included in the broader taxon sampling provided here. One possibility is that it represents a northern versus southern range expansion. If this was the case then it predicts that taxa from New Guinea and Asia will also be in the “Capricorn” clade. Further sampling from South Africa and Central and South America in combination with sampling from Asia and in particular India (given its Gondwanan heritage) will be needed to fully understand the biogeographic history of these taxa.

When considering the Anostostomatidae over a wide geographic distribution (*Chapter Four*), in order to assess the effect of Gondwanan breakup, again unexpected relationships were observed that are not easily explained. The conventional breakup sequence (see McLoughlin 2001) did not correspond with overall tree topology. For example, South African taxa were predicted to be sister to the rest of the anostostomatid taxa, however this was not found. Alternatively, if dispersal was common and ongoing throughout the Southern Hemisphere for this family, there should be little, if any, geographic structuring. However, results here did support geographic substructure within clades. Without invoking multiple (and extensive) extinctions throughout the Southern Hemisphere, a possible explanation for the pattern observed is paralagous sampling. As the family is well known for taxonomic discrepancies (e.g. Johns 1997; Gorochov 2001), this is a real possibility. Indeed, the results reported here suggest that taxonomic revision is desperately needed. Even taking the possibility of paralagous sampling into account and assessing the two clades individually, Gondwanan breakup does not appear to explain the relationships observed, nor does dispersal alone. It might be that taxa missing from the current analysis will resolve the relationships within the family, or it may turn out to be an example of convergent evolution (of taxonomic characters used to define the family), whereby the current comparisons are inappropriate.

The molecular phylogenetic results presented here show that the taxonomy of the Anostostomatidae family urgently needs to be revised. Behavioural, ecological, physiological and reproductive elements need to be investigated and mapped onto the current phylogenies. With a combination of molecular data and other traits, hypotheses about the evolution of traits such as flightlessness, loss of ears (*Hemiandrus*), reduction in ovipositor length (*Hemiandrus*), change from predatory to herbivourous diet (*Hemideina* and *Deinacrida*) and large size can be tested. Further to the developments within the family, general questions about the evolution of traits found in Orthoptera (to which Anostostomatidae belong and are suggested to represent an early lineage) can be addressed and may help to develop ideas about what sort of biota was present throughout time.

Along with members of the Anostostomatidae, I sequenced avian lineages from the Southern Hemisphere as part of a larger study to determine if the K – Pg asteroid impact 65 Ma is the driving force behind modern avian radiations (Alvarez et al. 1980; Ericson et al. 2006; Harshmann et al. 2008). Numerous studies suggest that the asteroid impact did not result in the diversification of Neoavian lineages (e.g. Cracraft 2001; Penny and Phillips 2004; Brown et al. 2008). The findings presented here (*Chapter Three*) suggest that there were at least 12 distinct lineages present before the asteroid impact, supporting a diverse array of modern bird lineages prior to the extinction of the dinosaurs. Indeed the fossil record suggests that birds, mammals and dinosaurs coexisted for some 100 million years before the K – Pg extinction.

The asteroid impact no doubt had an effect world wide (Alvarez et al. 1980), but extinction events (with the exception of those at/near the impact site) that occurred after the impact may have been caused indirectly. Competition between populations is not a new phenomenon and plays a significant role in evolution. For example, instead of the asteroid impact enabling the birds (and mammals) to diversify, the biological differences between birds and dinosaurs (e.g. regurgitation and parental care in general; Pond 1977) may have enabled the survival of avian young and therefore avian lineages.

Results reported here do not resolve the Neoavian polytomy. They do however make a valuable contribution and will provide useful platform for future research. Increasing amounts of mitochondrial data along with nuclear genes (other than introns) and rare genomic changes should resolve the Neoavian polytomy. However, resolution of the polytomy alone will not be enough. Support for branch length will be equally important as well as resolving the branching order, although for different reasons. Branch length tells you something about the time over which the genetic changes took place. They may indeed have occurred over a short time (geologically speaking) but the associated morphological changes may have taken much longer to occur.

Along with branch length, branching order is important. Organisms in general found near water (lakes, rivers or ocean) are more likely to fossilize. Owing to their fragile nature (due to a hollow construction adapted to flight), the avian fossil record is somewhat biased towards heavier, more robust aquatic species. Due to this, some have hypothesised that modern birds have arisen from 'transitional shorebirds' or 'waterbirds' (Unwin 1993; Feduccia 1999; Littlewood et al. 2001) Results reported here do not support this placement of the Neoavian root. Instead, the parrot lineage was found to be basal with the shorebirds found deeper within the tree. As discussed in *Chapter Three*, placing the root with parrots should be viewed with caution. By establishing well supported phylogenies, future work will be able to address more biological questions such as: What morphotype did the first true birds have? What ecological niche did they inhabit? Is it likely that they were in competition with dinosours and pterosaurs?

Along with good fossils, development of better analytical methods to account for biases in the sequence data, along with rates of change across lineages, is extremely important not only to avian systematics, but phylogenetic studies in general. Development of models to better fit evolutionary patterns should go a long way towards reconciling rates and dates. The fossil record will never be absolute and only through a multidisciplinary approach will an understanding of past events be reached.

5.2 CONCLUDING REMARKS

Debate will no doubt continue over the importance of large scale abiotic events in taxon history throughout the Southern Hemisphere. Clearly, dispersal has been, and continues to be, very important throughout the world and especially in the south. Understanding the physical history (geology) of regions is extremely important. Without a clear understanding of the land, there won't be an understanding of biological change. Individual groups need to be assessed in detail so a basic understanding of life history is reached, in association with wider taxonomic sampling, to fully resolve patterns and processes across regions.

Molecular phylogenetic studies, such as those reported here, indicate that rare physical events (such as asteroid impacts) and slow geological events (such as continental drift) may have an effect on species diversification, however, biogeographic processes and evolutionary processes are equally if not more important for understanding species diversity. The extreme viewpoint taken by Alvarez et al. (1980), for example, that a cataclysmic asteroid impact was a nessecary precursor for the radiation of mammals and birds is not supported by a growing body of results to which this thesis belongs. An important first step for any study should be the production of a robust phylogeny on which to base and test hypotheses. Until this is obtained, little will be known, or understood about the patterns and processes of evolution under which lineages have changed over time. The molecular phylogenetic results presented here indicate that the suggested abiotic stimuli of marine inundation, asteroid impacts or continental drift do not appear to have been the "driving force" for diversification of Anostostomatidae or modern birds (Neoaves).

5.3 FUTURE WORK

With the advent of sequencing technologies such as 454 and Solexa, there is now the opportunity to sequence larger fragments quicker and more cheaply than ever before. Future phylogenetic studies will have to incorporate not only whole mitochondrial, but also whole nuclear genomes and rare genomic changes, along with ecological and biological aspects. With an increase in sequence data, basic questions (for example regarding secondary structures, protein interactions, and identifying genes of interest), will become

easier and instead of studying a single group, Order-wide or higher comparisons will be possible.

Many groups of researchers around the world are devoting computing and sequencing resources to trying to resolve the phylogeny of modern birds. However, not so much is being done to study less charasmatic taxa that are more specific to the New Zealand region such as weta. More specifically for this thesis, there is much to be done both taxonomically and phylogenetically within the family Anostostomatidae. Future sampling will have to include taxa from throughout the global distribution with particular emphasis on those from Africa, India, Japan, Indonesia, New Guinea and Central and South America (Eades and Otte see also Appendix 4.8.1 table 4-3). Broad taxonomic sampling from the geographic distribution will enable more vigorous testing of the current findings to be undertaken.

In addition to general taxonomic sampling, the development of new primers will enable larger fragments of mtDNA to be sequenced, will be important for future studies. Progress made here (see *Appendix A*) will enable marker development to be carried out for New Zealand taxa initially, with the hope that family specific markers will subsequently be identified. A large proportion of work still needs to be done on *Hemiandrus* species both in Australia and New Zealand. Very little is known about them and even less has been published (with the exception of work carried out in this thesis). It appears clear that the Australian and New Zealand species are not closely related and future work should start with a taxonomic revision of the group. More specifically from this research, *Hemiandrus maculifrons* looks particularly interesting as it is a widespread species throughout New Zealand (found in both the North and South Islands) and appears to be distinct from the other *Hemiandrus* species (refer to figure 2.6). More detailed studies into the correlation between maternal care and ovipositor length would also be useful. The reduction in ovipositor length appears to have occurred more than once in New Zealand *Hemiandrus* species and is not the ancestral state within the Orthopteran insects.

Utilising new sequencing technologies will also allow for more detailed studies into nuclear regions which could identify genes under selection in this family but could be applied to the wider Orthopteran group. Differences observed in the *Hemideina* and *Deinacrida* group should also be investigated further. Since both size and metabolic rate have been shown in other organisms to affect the mutation rate, development of metabolic techniques may provide useful information on this lineage as they are generally large and predominantly herbivorous, which is unusual in Orthoptera.

5.4 REFERENCES

Alvarez, L. W., W. Alvarez, F. Asaro, and H. V. Michel. 1980. Extraterrestrial cause for the Cretaceous-Tertiary extinction: experimental results and theoretical interpretation. Science **208**:1095-1108.

Brettschneider, H. 2006. Systematics of southern African Anostostomatidae (Orthoptera) based on morphological and molecular data. Pp. 147. Natural and Agricultural Sciences. University of Pretoria, Pretoria.

Brown, J. W., J. S. Rest, J. García-Moreno, M. D. Sorenson, and D. P. Mindell. 2008. Strong mitochondrial DNA support for a Cretaceous origin of modern avian lineages. BMC Biology **6**:6.

Campbell, H. J., and G. Hutching. 2007. In search of ancient New Zealand. Penguin, Auckland & GNS Science, Lower Hutt, New Zealand.

Cook, L. G., and M. D. Crisp. 2005. Not so ancient: the extant crown group of *Nothofagus* represents a post-Gondwanan radiation. Proceedings of the Royal Society of London Series B-Biological Sciences **272**:2535-2544.

Cracraft, J. 2001. Avian evolution, Gondwana biogeography and the Cretaceous-Tertiary mass extinction event. The Royal Society of London, B **268**:459-469.

Eades, D. C., and D. Otte. Orthoptera Species File Online. Version 2.0/3.4.

Ericson, P. G. P., C. L. Anderson, T. Britton, A. Elzanowski, U. S. Johansson, M. Källersjö, J. I. Ohlson, T. J. Parsons, D. Zuccon, and G. Mayr. 2006. Diversification of Neoaves: integration of molecular sequence data and fossils. Biological Letters **2**:543-547.

Feduccia, A. 1999. The origin and evolution of birds. Yale University Press, New Haven.

Goldberg, J., S. A. Trewick, and A. M. Paterson. 2008. Evolution of New Zealand's terrestrial fauna: a review of molecular evicence. Philosophical Transactions of the Royal Society of London B Biological Sciences **363**:3319-3334.

Gorochov, A. V. 2001. The higher classification, phylogeny and evolution of the superfamily Stenopelmatiodea. Pp. 3-33 *in* L. H. Field, ed. The biology of wetas, king crickets and their allies. CABI Publishing, Wallingford.

Grandcolas, P., J. Murienne, T. Robillard, L. Desutter-Grandcolas, H. Jourdan, E. Guilbert, and L. Deharveng. 2008. New Caledonia: a very old Darwinian island? Philosophical Transactions of the Royal Society of London B Biological Sciences **363**:3309-3317.

Harshmann, J., E. L. Braun, M. J. Braun, C. J. Huddleston, R. C. K. Bowie, J. L. Chojnowski, S. J. Hackett, K.-L. Han, R. T. Kimball, B. D. Marks, K. J. Miglia, W.

S. Moore, S. Reddy, F. H. Sheldon, D. W. Steadman, S. J. Steppan, C. C. Witt, and T. Yuri. 2008. Phylogenomic evidence for multiple losses of flight in ratite birds. PNAS **105**:13462-13467.

Johns, P. M. 1997. The Gondwanaland Weta: Family Anostostomatidae (formerly in Stenopelmatidae, Henicidae or Mimmermidae): Nomenclatural Problems, World Checklist, New Genera and species. Journal of Orthoptera Research **6**:125-138.

Knapp, M., R. Mudaliar, D. Havell, S. J. Wagstaff, and P. J. Lockhart. 2007. The drowning of New Zealand and the problem of Agathis. Systematic Biology **56**:862-870.

Knapp, M., K. Stöckler, D. Havell, F. Delsuc, F. Sebastiani, and P. J. Lockhart. 2005. Relaxed molecular clock provides evidence for long distance dispersal of *Nothofagus* (Southern Beech). PLoS Biology **3**:e14.

Landis, C. A., H. J. Campbell, J. G. Begg, D. C. Mildenhall, A. M. Paterson, and S. A. Trewick. 2008. The Waipounamu erosion surface:questioning the antiquity of the New Zealand land surface and terrestrial fauna and flora. Geological Magazine:1-25.

Lee, D. E., J. M. Bannister, U. Kaulfuss, J. G. Conran, and J. K. Lindqvist. 2009. Late Eocene palms, ferns and fungi, Late Oligocene podocarps, proteaceans and epicrids, Earlz Miocene orchids, fish and insects: How exceptional fossil deposits in Otago and Southland are expanding what we know about the history of the New Zealand terrestrial biota. Pp. 28-30. Geology and Genes IV. Geological Society of New Zealand Miscellaneous Publication, Christchurch.

Littlewood, D. T. J., P. D. Olson, M. J. Telford, E. A. Herniou, and M. Riutort. 2001. Elongation Factor 1-Alpha sequences alone do not assist in resolving the position of the Acoela within the Metazoa. Molecular Biology and Evolution **18**:437-442.

Martins-Neto. 2007. New Orthoptera Stenopelmatoidea and Hagloidea (Ensifera) from the Santana Formation (Lower Cretaceous, Northeast Brazil) with description of new taxa. GAEA **3**:3-8.

McLoughlin, S. 2001. The breakup history of Gondwana and its impact on pre-Cenozoic floristic provincialism. Australian Journal of Botany **49**:271-300.

Meads, M. J. 1990. The Weta Book: A guide to the identification of wetas. DSIR Land Resources, Lower Hutt.

Murienne, J., P. Grandcolas, M. D. Piulachs, Z. Bellés, C. D'Haese, F. Legendr, R. Pellens, and E. Guilbert. 2005. Evolution on a shaky piece of Gondwana: is local endemisim recent in New Caledonia? Cladistics **21**:2-7.

Penny, D., and M. J. Phillips. 2004. The rise of birds and mammals: are microevolutionary processes sufficient for macroevolution? Trends in Ecology & Evolution **19**:516-522.

Pond, C. M. 1977. The significance of lactation in the evolution of mammals. Evolution **31**:177-199.

Stöckler, K., I. L. Daniel, and P. J. Lockhart. 2002. New Zealand Kauri (*Agathis australis* (D.Don) Lindl., Araucariaceae) survives Oligocene drowning. Systematic Biology **51**:827-832.

Trewick, S. A., A. M. Paterson, and H. J. Campbell. 2007. Hello New Zealand. Journal of Biogeography **34**:1-6.

Unwin, D. M. 1993. Aves. Pp. 717-737 *in* M. J. Benton, ed. The Fossil Record II. Chapman & Hall, London.

Worthy, T. H., S. J. Hand, M. S. Y. Lee, M. Hutchinson, A. J. D. Tennyson, R. P. Scofield, B. Marshall, J. P. Worthy, J. M. T. Nguyen, W. E. Boles, and M. Archer. 2009. New Zealand`s St Bathans fauna: an update on its composition and relationships.

Pp. 40-43. Geology & Genes IV. Geological Society of New Zealand Miscellaneous Publication, Christchurch.

APPENDIX A

The most exciting phrase to hear in science, the one that heralds the most discoveries, is not "Eureka!" (I found it!) but "That's funny..."

-Isaac Asimov

Mitogenomics – exploring mitochondrial genomes from New Zealand weta (Orthoptera: Ensifera: Anostostomatidae)

A.1 ABSTRACT

The sequencing of complete mitochondrial genomes is becoming increasingly common. Their use in phylogenetic reconstruction enables the study of genome evolution including gene order and structure. Here I describe the protocols used for extraction, PCR, cloning, sequencing, sequence assembly of the mitochondrial genome of the ground weta (*Hemiandrus pallitarsis*). While attempts to complete the genome were confounded by the amplification of apparently multiple NUMTS (fragments of mtDNA that have moved to the nucleus; these are often pseudogenes), new data presented here will enable the resolution of these issues.

A.2 INTRODUCTION

Whole mitochondrial genome sequencing is increasing in popularity with comparisons being used successfully to address many phylogenetic questions for vertebrates (eg. Curole and Kocher 1999; Phillips and Penny 2003; Gibb et al. 2007; Slack et al. 2007). Recently the number of complete Orthopteran insect mitochondrial genomes has increased (Flook et al. 1995; Kim et al. 2005; Fenn et al. 2007; Zhou et al. 2007; Zhang and Huang 2008) and phylogenetic utility is being thoroughly investigated in a number of related groups (Cameron et al. 2004; Cameron et al. 2007; Carapelli et al. 2007; Kjer and Honeycutt 2007). Mitochondrial genomes have the potential to provide better resolution for deep relationships as has been shown for vertebrates (eg. see above references) and have been shown to be useful for resolving intraordinal relationships within Diptera (Cameron et al. 2007) and Hymenoptera (Castro and Dowton 2007). In addition to the sequence information, insight into gene rearrangements (Boore and Brown 1998), gene insertions/deletions (Rokas and Holland 2000) and length variability in genic or intergenic regions (Schneider and Ebert 2004) can be investigated.

Mitochondria are membrane-enclosed organelles found in most eukaryotic cells. Described as the “cellular power plants” they are responsible for the generation of adenosine triphosphate (ATP) which is used by the cells as a source of chemical energy. Along with this energy function, mitochondria are known to play a role in processes such as signalling, cellular differentiation, cell death and growth. Generally, animal mitochondrial genomes are double-stranded circular DNA molecules typically *ca*16 Kb in length (for exceptions in Cnidaria see Bridge et al. 1992; in a crustacean see Raimond et al. 1992). The mitochondrial genome usually consists 37 genes: 13 protein coding genes, 22 transfer RNAs and two ribosomal RNA genes plus a control region (Boore 1999). The arrangement of these genes and RNAs can be quite variable throughout the animal kingdom however the closer the taxa are, the more similar their order will be.

Figure A.1 Complete mitochondrial genomes

Complete annotated mitochondrial genome of; a) *Drosohpila yakuba* (Clary and Wolstenholme 1985), b) *Homo sapiens* (Anderson et al. 1981) and c) *Locusta migratoria* (Flook et al. 1995). Gene rearrangements are not uncommon, however generally, the same complement of genes and RNAs are present.

There are several characteristics that make the small, compact mitochondria useful for studying evolutionary biology. Comparisons can be made across orders as there are homologous genes found in the mtDNA of animals, plants, protists, fungi and prokaryotes (Lang et al. 1997; Gray et al. 1998; Gray 1999). The lack of recombination, maternal inheritance and rapid evolution (when compared to nuclear genes) makes mtDNA useful for genome rearrangements and effects on amino acid substitution patterns (Boore 2000; Helfenbein et al. 2004).

Anostostomatidae belong to the order Orthoptera (crickets, katydids, grasshoppers, weta and allies). Orthoptera is comprised of over 20,000 species including some of the largest living insects (body length over 11.5cm and wing spans more than 22cm)(Rentz 1991). The order is split into two suborders; Ensifera (long horned; crickets and their allies) and Caelifera (short horned; grasshoppers). Fossil Orthoptera date back *ca* 250 Ma for Ensifera and *ca* 225 Ma for Caelifera although the lineage as a whole would be much older (300+ Ma) (Grimaldi and Engel 2005 p. 202, Fig. 7.18). Orthopteran mitochondrial genomes have been shown to have a useful tRNA gene rearrangement. Flook et al. (1995) reported the rearrangement of $tRNA^{Lys}$ and $tRNA^{Asp}$ in *Locusta migratoria*; however gene rearrangements in most insect orders appears limited. Until recently, only two Orthopteran mitochondrial genomes had been sequenced, one Caelifera and one Ensifera. Currently, there are nine sequences available on GenBank; *Locusta migratoria*, X80245 (15,722bp) (Flook et al. 1995); *Gryllotalpa orientalis* AY660929 (15,521bp) (Kim et al. 2005); *Anabrus simplex* EF373911 (15,766bp)(Fenn et al. 2007); *Ruspolia dubia* EF583824 (14,971bp) (Zhou et al. 2007); *Oxya chinensis* EF437157 (15,443bp)(Zhang and Huang 2008); *Locusta migratoria* EU287446 (16,053bp) (Xiao, Zhou and Huang unpublished); *Oedaleus decorus asiaticus* EU513374 (16,259bp) (Ma and Kang unpublished); *Gastrimargus marmoratus* EU513373 (15,924bp) (Ma and Kang unpublished) and *Chorthippus chinensis* EU029161 (15,599bp) (Liu and Huang unpublished). Of these, three are from Ensifera (*Gryllotalpa orientalis*, *Anabrus simplex* and *Ruspolia dubia*).

Here I present data from the first mitochondrial genomes to be sequenced from the Anostostomatidae. Two species were attempted (*Hemiandrus pallitarsis* and *Hemideina crassidens*) however results focus on *Hemiandrus pallitarsis* due to time constraints. However preliminary analysis of *Hemideina crassidens* supports the findings from *Hemiandrus pallitarsis*. I was unable to suggest a definitive gene order for the family at this time, but instead provide insight for future work on the group.

A.3 MATERIALS AND METHODS

(a) Sampling

The two samples chosen for mitochondrial DNA sequencing represent two widely distributed and abundant species, *Hemiandrus pallitarsis* (ground weta) from the Manawatu region and *Hemideina crassidens* (tree weta) from the Wellington region. Permits for collection were obtained and Iwi consultation was undertaken prior to work being carried out [Protocol No. GMO 04/MU/09 Steve Trewick (invertebrates)].

(b) DNA extraction

Initial trials to obtain mitochondrial enriched extracts were carried out as per *Appendix C 2.2*. However it was soon apparent that there was insufficient starting material to obtain the quantities of the DNA needed. Instead total genomic DNA was extracted from hind leg muscle following the salting-out method (Appendix C 2.1; Sunnucks and Hale 1996) and re-suspended in 50µL TE buffer (0.1mM EDTA, 10mM Tris, pH 8.0) or water.

(c) Long Range PCR

Specific 16S primers (see *Appendix C* table C-2) were designed using Oligo 4.03 (National Biosciences, Inc., Plymouth, MN) from sequence data obtained using LR-J-13417 and SR-N-14588 (Simon et al. 1994) and used to successfully amplify the mitochondrial genome (*ca* 15Kb) with the Expand Long template PCR System (Roche Applied Science, Mannheim, Germany). A second set of primers were designed, identical to the first, but with restriction sites attached (see *Appendix C* table C-2).

Thermal cycling PCR was carried out on an MJ Research PTC-200 thermal cycler and consisted of initial denaturation of 94°C for 2min, followed by 10 cycles of 94°C for 30sec and 68°C for 14min followed by 25 cycles of 94°C for 30sec and 68°C for 14min with a final extension of 68°C for 7min. The products were visualized on a 1% (w:v) agarose gel. Attempts to use long range products as template DNA for short range PCR (overlapping fragments 0.5 – 3 kb in length) were variable. The universal mitochondrial primers developed by Simon et al. (1994) were inadequate so restriction digestion and shotgun cloning the long range fragment with the aim of primer design was employed.

Restriction digestion using EcoRI (Roche) and HindIII (Roche) were carried out on long range PCR products and cloned into PUC118 using the TOPO TA cloning kit for sequencing (Invitrogen; see *Appendix C*.4 for cloning detail). For each region at least three clones were sequenced to safeguard against PCR errors. In all cases, overlaps between sequences were sufficient to ensure synonymy and sequence identity was confirmed through BLAST searches (http://www.ncbi.nlm.nih.gov/blast/), confirmation of amino acid translation in coding regions, and alignment with other species.

(d) Short Range PCR

PCR reactions were performed in 10μL volumes using ABgene Red Hot *Taq*. Products were visualised on 1% (w:v) agarose gels stained with SYBRSafe (Invitrogen). Thermal cycling PCR was carried out on an MJ Research PTC-200 thermal cycler and consisted of initial denaturation of 94°C for 2min, followed by 35 cycles of 94°C for 30sec, 48 - 50°C for 30sec and 72°C for 1min 30sec with a final extension of 72°C for 3min. PCR products were purified with Shrimp Alkaline Phosphatase (SAP) and Exonuclease I (EXOI) following manufacturer's recommendations (USB Corporation). Sequencing was performed using BigDye Terminator Cycle Sequencing reagents according to the manufacturer's instructions (Applied Biosystems, Foster City, CA), then sequenced on an ABI 3730 automated sequencer (Applied Biosystems, Foster City, CA). BLAST searches were carried out and sequences aligned using SEQUENCHER 4.7 (Gene Codes Corp. Ann Arbour, MI). Sequences were manually edited and checked for complete concurrence between overlapping sequences. Primers used for PCR were; mtDNA COI: LCO1490, HCO2198 (Folmer et al. 1994), CI-J-2195 and L2-N-3014 (Simon et al. 1994); mtDNA third domain 12S rRNA: SR-N-14588 and LR-J-13417 (Simon et al. 1994).

(e)Restriction digestion

To complement sequence data obtained through short range PCR, long range PCR products were digested with three different restriction enzymes (EcoRI and HindIII) and in combination (EcoRI/HindIII). Digestions were carried out following manufacturers' instructions and run on 1% (w:v) agarose gels to visualise fragment size. Cloning was carried out as per manufacturers' recommendation (see TOPO TA cloning kit for sequencing Invitrogen).

(f) Solexa sequencing

Long range fragments were prepared for sequencing using Repli-G (Qiagen) according to the manufacturer's protocol. Following the Repli-G amplification, a total of 400ng was run on a Solexa sequencing machine (AWCGS) mixed with long range templates from five other species (bird, human, rat, mollusc and a frog). Solexa technology requires a 'scaffold' to align new sequences to. Here the Oriental mole cricket (NC_006678) and cave cricket (EU938374) were used (personal communication B. McCormish). Successful retrieval of the many short reads produced from this technology, is somewhat dependent on having a close match to scaffold sequences (especially from a mixed run such as this). Neither of the reference sequences used here were particularly close to the weta however retrieval was aided by the comparatively high A-T content of insect mt genomes.

(g) Sequence alignment and annotation

Sequences from PCR and cloning were initially aligned in SEQUENCHER to identify identical clone sequences. To avoid sequence variation introduced by possible cloning, three identical clones were required. Overlapping sequences (contigs) assembled and blasted in GenBank. Alignments were exported to SE-AL v2.0a11 and translated at the amino acid level for protein-coding genes to identify andidentify possible NUMT sequences.

A.4 RESULTS

A summary of sequences obtained from *Hemiandrus pallitarsis* can be seen in figure A.2. Sequences were aligned in SEQUENCHER to either the Oriental mole cricket or cave cricket and protein coding regions translated in SEAL. Similar results were obtained for *Hemideina crassidens* however due to the difficult assembly and time constraints, will not be shown here.

(a) Genome organisation

The complete mitochondrial genome sequence was not determined in the current study so the gene order could not be determined. However coding sequences of *Hemiandrus pallitarsis* (8,542 bp) were obtained (figure A.2 A). Of the 13 protein coding genes, six were identified (although not all complete). Blast searches were carried out and although the gene regions were identifiable, the database produced few close matches with an average match of 65 - 73% (which was not unexpected as the Anostostomatidae are a group with little prior information).

(b) True sequence Vs. NUMTS

In the case of protein coding genes, sequences were translated and matched to reference mt genomes. Fragments that did not translate were deemed NUMTS or pseudogenes. Fragments identified can be seen in table A-1 below and raw sequences can be found in A.8.1 (coding) and A.8.2 (NUMT).

Table A-1 *Hemiandrus pallitarsis* fragments

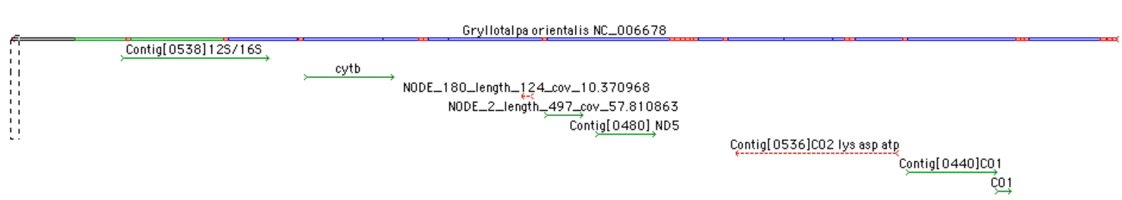
coding	NUMT
tVal-16S-tLeu-ND1 (2,034 bp)	Cytb-ND1 pseudo (1,508bp)
Cytb-ND6 (1,258 bp)	ND4-tHis-ND5 pseudo (1,479 bp)
ND4 (150 bp)	ND3-CO2 pseudo (2,396 bp)
tHis-ND5 (523 bp)	COI-ND2 pseudo (1,102 bp)
ND5 (825 bp)	
tLeu-CO3-ATP8-ATP6-tLys-CO2 (2,275 bp)	

(c) Solexa results

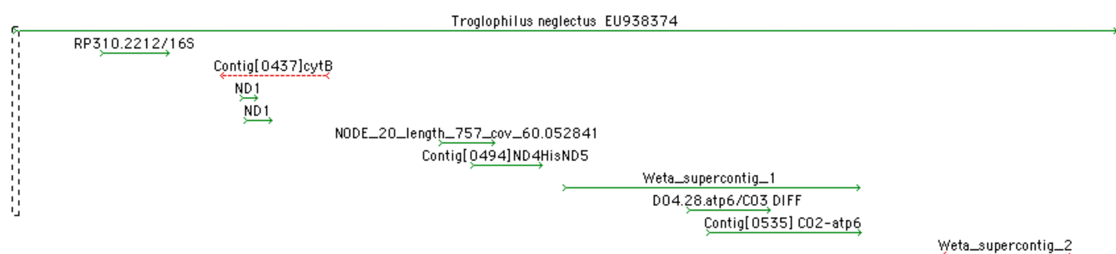
Three supercontigs were assembled; two subsequently turned out to be putative NUMTS (see figure A.2 B). The third "mystery" contig (*ca* 4.9Kb in length; figure A.2 C), contained long repeat regions, and did not match any sequences on the search database. However it did match identically other sequences obtained via shotgun cloning. There are three possibilities for this repeat region; (1) the fragment is mitochondrial control region sequence, (2) the fragment is NUMT control region or (3) it is an artefact from the Repli-G amplification. Until more data is available, this fragment will remain unknown. In addition, smaller fragments (less than 800bp) were retrieved from the Solexa run and incorporated into the total dataset where possible.

Hemiandrus pallitarsis **(New Zealand ground weta)**

A. Coding regions of complete mitochondrial genome aligned to *Gryllotalpa orientalis* NC_006678



B. Non-coding (NUMTS) or pseudogenes aligned to *Troglophilus neglectus* EU938374



C. Unidentified fragments

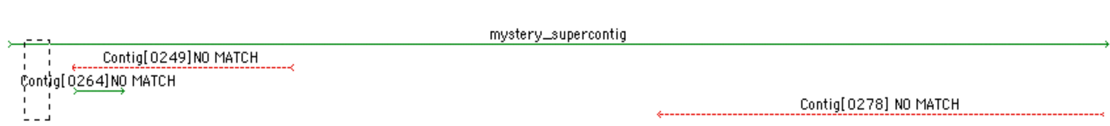


Figure A.2 SEQUENCHER alignments for *Hemiandrus pallitarsis*.

A. Identified mitochondrial coding genes and rRNAs generated from shotgun cloning and Solexa data, B. NUMT sequences generated from shotgun cloning and Solexa data and C. unidentified supercontig sequence from Solexa data and unidentified contigs generated from shotgun cloning.

(d) Heteroplasmy or NUMTS

The possibility of cross-contamination was thoroughly controlled for by working with only one individual at a time along with negative controls (in case of reagent contamination). Comparing results obtained via standard PCR with those obtained using cloning techniques showed a higher than expected rate of mutation. Alignments were assessed by eye and those determined to be NUMTS were removed. The addition of sequences obtained via Solexa sequencing also contained variants and were assessed in the same way. Due to the shear number of differences, it is more likely that there are a multiple NUMT sequences rather than simply heteroplasmy within individuals, and may represent multiple events.

A.5 DISCUSSION

Using multiple methods including mtDNA enrichment (extraction method), long range template DNA, restriction digests and shotgun cloning enabled partial mitochondrial genomes of the two Anostostomatidae taxa (*Hemiandrus pallitarsis* and *Hemideina crassidens)* to be sequenced although only the results from *Hemiandrus pallitarsis* are reported here.

Heteroplasmic sites within individuals were consistently identified and accounted for more error than expected by cloning alone. This strongly suggested that presence of NUMT sequences which were later confirmed with Solexa sequencing. NUMTS evolve under different constraints to functional copies and therefore have the potential to mislead phylogenetic reconstruction if included. Mitochondrial genomes of many organisms have been shown to harbour NUMTS (Zhang and Hewitt 1996; Bensasson et al. 2001; Richly and Leister 2004) the most notable example being the 7.9Kb NUMT found in 38 - 79 tandemly repeated copies in the cat (Lopez et al. 1994). NUMTS appear to be common in Orthoptera as shown by Sword et al (2007) who found a more than 6,000 bp NUMT region in grasshoppers. NUMT sequences identified here span the entire mt genome and further work is needed to determine its characteristics. It may well be that there is an entire pseudo mitochondrial genome.

NUMTS may arise in a number of ways; from many independent transfers as well as through the amplification (multiplication) of single event. NUMTS are not equally abundant across species (Blanchard and Schmidt 1996) and have been shown to potentially be useful in evolutionary studies. They can be used to infer ancestral states, mitochondrial phylogeny root and in some cases, may be useful to study spontaneous mutation in nuclear genomes (Bensasson et al. 2001). If unrecognised, NUMTS can lead to robust, believable but incorrect phylogenies.

The use of ‘universal’ primers (Kocher et al. 1989) allows comparison of many sequences from different taxa but also increases the chance of amplifying NUMTS. Also, gene rearrangements might make priming sites defective in the mitochondria leading to preferential amplification of NUMT sequences. Identification of coding sequences (as apposed to NUMTS) will enable specific primers to be designed which should alleviate the pseudogene issues found here. Alternatively, heteroplasmy is possible (Morgan-Richards and Wallis 2003). This results in two (or more) mt DNA haplotypes within an individual. However we have no prior evidence for hybridization in *Hemiandrus pallitarsis* (or *Hemideina crassidens*) and it is more likely to be the result of NUMT sequences.

In addition to completing the mt genomes reported here (using specific primers), the NUMTS themselves should be investigated further. NUMTS provide a window on the dynamics of genome evolution and intergenomic interactions (Bensasson et al. 2000). Using Fluorescent in situ hybridization (FISH) could potentially show which regions of the NUMT sequence map in the nuclear genome.

De novo amplification of mitochondrial sequences for the two Anostostomatidae reported here, are an important foundation for future work on this family. Primer design from sequenced fragments and sequencing missing regions should now be a reasonably straight forward process.

A.6 Acknowledgements

I would like to thank Andrew Clarke and Trish McLenachan for all their discussion, ideas and assistance throughout this project. This work was supported by the Marsden Fund awarded to D. Penny.

A.7 References

Anderson, S., A. Bankier, B. Barrell, M. DeBruijn, A. Coulson, J. Drouin, I. Speron, D. Nierlich, B. Roe, F. Sanger, P. Schreier, A. B. Smith, R. Staden, and I. Young. 1981. Sequence and organization of the human mitochondrial genome. Nature **290**:457-465.

Bensasson, D., D.-X. Zhang, D. L. Hartl, and G. M. Hewitt. 2001. Mitochondrial pseudogenes: evolution's misplaced witnesses. Trends in Ecology & Evolution **16**:314-321.

Bensasson, D., D.-X. Zhang, and G. M. Hewitt. 2000. Frequent assimilation of mitochondrial DNA by grasshopper nuclear genomes. Molecular Biology and Evolution **17**:406-415.

Blanchard, J. L., and G. W. Schmidt. 1996. Mitochondrial DNA migration events in yeast and humans: Integration by a common end-joining mechanism and alternative perspectives on nucleotide substitution pattern. Journal of Molecular Evolution **13**:537-548.

Boore, J. L. 1999. Animal mitochondrial genomes. Nucleic Acids Research **27**:1767-1780.

Boore, J. L. 2000. The duplication/random loss model for gene rearrangement exemplified by mitochondrial genomes of deuterostome animals *in* D. Sankoff, and J. Nadeau, eds. Comparative Genomics. Kluwer Academic Publishers, Dodrecht, Netherlands.

Boore, J. L., and W. M. Brown. 1998. Big trees from little genomes: mitochondrial gene order as a phylogenetic tool. Current Opinion in Genetics & Development **8**:668-674.

Bridge, D., C. W. Cunningham, R. DeSalle, and L. Buss, W. 1992. Class-level relationships in the phylum Cnidaria: Evidence from mitochondrial genome structure. Proceeding of the National Academy of Science, USA **89**:8750-8753.

Cameron, S. L., C. L. Lambkin, S. C. Barker, and M. F. Whiting. 2007. A mitochondrial genome phylogeny of Diptera: whole genome sequence data accurately resolve relationships over broad timescales with high precision. Systematic Entomology **32**:40-59.

Cameron, S. L., K. B. Miller, C. A. D'Haese, M. F. Whiting, and S. C. Barker. 2004. Mitochondrial genome data alone are not enough to unambiguously resolve the relationships of Entognatha, Insecta and Crustacea *sensu lato* (Arthropoda). Cladistics **20**:534-557.

Carapelli, A., P. Lio, F. Nardi, E. van der Wath, and F. Frati. 2007. Phylogenetic analysis of mitochondrial protein coding genes confrims the reciprocal paraphyly of Hexapoda and Crustacea. BMC Evolutionary Biology **7**:S8.

Castro, L. R., and M. Dowton. 2007. Mitochondrial genomes in the Hymenoptera and their utility as phylogenetic markers. Systematic Entomology **32**.

Clary, D. O., and D. R. Wolstenholme. 1985. The mitochondrial DNA molecule of Drosophila yakuba: nuclotide sequence, gene organization, and genetic code. Journal of Molecular Evolution **22**:252-271.

Curole, J. P., and T. D. Kocher. 1999. Mitogenomics: digging deeper with complete mitochondrial genomes. Trends in Ecology & Evolution **14**:394-398.

Fenn, J. D., S. L. Cameron, and M. Whiting. 2007. The complete mitochondrial genome sequence of the Mormon cricket (Anabrus simplex: Tettigoniidae: Orthoptera) and an analysis of control region variability. Insect Molecular Biology **16**:239-252.

Flook, P. K., C. H. Rowell, and G. Gellissen. 1995. The sequence, organization, and evolution of the *Locusta migratoria* mitochondrial genome. Journal of Molecular Evolution **41**:928-941.

Gibb, G. C., O. Kardailsky, R. T. Kimball, E. Braun, and D. Penny. 2007. Mitochondrial genomes and avian phylogeny: complex characters and resolvability without explosive radiations. Mol. Biol. Evol **24**:269-280.

Gray, M. W. 1999. Evolution of organellar genomes. Current Opinion in Genetics & Development **9**:678-687.

Gray, M. W., B. F. Lang, R. Cedergren, G. B. Golding, C. Lemieux, D. Sankoff, M. Turmel, N. Brossard, E. Delage, T. G. Littlejohn, I. Plante, P. Rioux, D. Saint-Louis, Y. Zhu, and G. Burger. 1998. Genome structure and gene content in protist mitochondrial DNAs. Nucleic Acids Research **26**:865-878.

Grimaldi, D., and M. S. Engel. 2005. Evolution of Insects. Cambridge University Press, New York.

Helfenbein, K. G., H. M. Fourcade, R. G. Vanjani, and J. L. Boore. 2004. The mitochondrial genome of Paraspadella gotoi is highly reduced and reveals that chaetognaths are sister-group to protostomes. Proceeding of the National Academy of Science, USA **101**:10639-10643.

Kim, I., S. Y. Cha, M. H. Yoon, J. S. Hwang, S. M. Lee, H. D. Sohn, and B. R. Jin. 2005. The complete nucleotide sequence and gene organization of the mitochondrial genome of the oriental mole cricket, *Gryllotalpa orientalis* (Orthoptera: Gryllotalpidae). Gene **353**:155-168.

Kjer, K. M., and R. L. Honeycutt. 2007. Site specific rates of mitochondrial genome and phylogeny of eutheria. BMC Evolutionary Biology **7**:8.

Kocher, T. D., W. K. Thomas, A. Meyer, S. V. Edwards, S. Pääbo, F. X. Villablanca, and A. C. Wilson. 1989. Dynamics of mitochondrial DNA evolution in animals: Amplification and sequencing with conserved primers. Proceedings of the National Academy of Science, USA **86**:6196-6200.

Lang, B. F., G. Burger, C. J. O'Kelly, R. Cedergren, G. B. Golding, C. Lemieux, D. Sankoff, M. Turmel, and M. W. Gray. 1997. An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. Nature **387**:493-497.

Lopez, J. V., N. Yuhki, R. Masuda, W. Modi, and S. J. O. O'Brien. 1994. *Numt*, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. Journal of Molecular Evolution **39**:174-190.

Morgan-Richards, M., and G. P. Wallis. 2003. A comparison of five hybrid zones of the weta *Hemideina thoracica* (Orthoptera: Anostostomatidae): degree of cytogenetic differentiation fails to predict zone with. Evolution **57**:849-861.

Phillips, M. J., and D. Penny. 2003. The root of the mammalian tree inferred from whole mitochondrial genomes. Molecular Phylogenetics and Evolution **28**:171-185.

Raimond, R., I. Marcade, D. Bouchon, T. Rigaud, J.-P. Borry, and C. Souty-Grosset. 1992. Organiszation of the large mitochondrial genome in the isopod *Armadillidium vulgare*. Genetics **151**:203-210.

Rentz, D. C. 1991. Orthoptera. Pp. 369-393 *in* I. D. Naumann, ed. The insects of Australia: a textbook for students and research workers. Melbourne University Press and Cornell University Press, Melbourne, Australia.

Richly, E., and D. Leister. 2004. NUMTs in sequenced eukaryotic genomes. Molecular Biology and Evolution **21**:1081-1084.

Rokas, A., and P. W. Holland. 2000. Rare genomic changes as a tool for phylogenetics. Trends in Ecology and Evolution **15**:454-459.

Schneider, A., and D. Ebert. 2004. Covariation of mitochondrial genome size with gene lengths: evidence for gene length reduction during mitochondrial evolution. Journal of Molecular Evolution **59**:90-96.

Simon, C., F. Frati, A. Bechenbach, B. Crespi, H. Liu, and P. Flook. 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Annals of the Entomological Society of America **87**:651-701.

Slack, K. E., F. Delsuc, P. A. McLenachan, U. Arnason, and D. Penny. 2007. Resolving the root of the avian mitogenomic tree by breaking up long branches. Molecular Phylogenetics and Evolution **42**:1-13.

Sunnucks, P., and D. F. Hale. 1996. Numerous transposed sequences of mitochondrial cytochrome oxidase I-II in aphids of the genus Sitobion (Hemiptera: Aphididae). Molecular Biology and Evolution **13**:510-524.

Sword, G. A., L. B. Senior, J. F. Gaskin, and A. Joern. 2007. Double trouble for grasshopper molecular systematics: intra-individual heterogeneity of both mitochondrial 12S-valine-16S and nuclear internal transcribed spacer ribosomal DNA sequences in *Hesperotettix viridis* (Orthoptera: Acididae). Systematic Entomology **32**:420-428.

Zhang, C., and Y. Huang. 2008. Complete mitochondrial genome of Oxya chinensis (Orthoptera, Acridoidea). Acta Biochimica et Biophysica Sinica **40**:7-18.

Zhang, D.-X., and G. M. Hewitt. 1996. Nuclear integrations: challenges for mitochondrial DNA markers. Trends in Ecology and Evolution **11**:247-251.

Zhou, Z., Y. Huang, and F. Shi. 2007. The mitochondrial genome of *Ruspolia dubia* (Orthoptera: Concephalidae) contains a short A+T-rich region of 70bp in length. Genome **50**:855-866.

A.8 RAW DATA

A.8.1 Contigs that translate

12S (about 100bp) tVal- 16S- tLeu-ND1

ACACATCGCCCGTCGCTCTCATTTTTTTAAAGGTGAGATAAGTCGTAACATAGTAGATGTACTGGAAAGTGTA
TCTAGAAAGTTTAACGAAATAAAGCTTAAAGTGAAGTGTTCCATTTACACTGGAGTGAAATTTGTGCAATTCA
GATTGTTTTGAAATAAGAAATTTAATTTTATTTGGTGTTGAYTTTACATGTTYAATAAAAGTAATTTTATTAA
TGGTTGTTTAAGTATAGTTTATAGAATGAATTTTATGATTGATAGTTTATTAGTAYTGTGAAAGATTAATTGA
AAATTGTGAAATATGGGATTAAAATATAGTAGATTTTTATTGTACCTTGTGTATCAGTGTTGATTTAATAATG
ATTAAAAATTTGGTATTCCCGATTTAAAGAGAGTTAATGTAGTAGTTTAGTTAATGTATTAAAATTATTAGTG
ATATTATATTAGAAGTGAGACGTTAGGCGTTCTTTAGGATATCTGGTTTTCTAAGAATTGAATTTAATTCAGT
TGTTAAGAGTATCTTTTTATGTCAAGTAAGAGATATAAATTAACATAAGAAGTTTGGGGGATAAGCTCCGAAT
TTATGTTTATAAGTCGAAGATTCAGTAAAGGATAAGTAGGTTTTAAATTAGCTATCTTTAGGAGTGTGTTATA
ATTTATTTATTGGTGTTTATGATTTTATAAGGGTTTTAGATAAGTGTATTAGAATTATTTCAAAAATTGTATT
TGATTTGAAATAATGATTGAATTAGTATAAGAAAATTGTGTAAATTGAATTGATTTTGTTAATTTATAGTAAG
GAATTAGGCAAATTTAAATGCTCGCCTGTTTATCAAAAACATGTCTTTTTGACAATGAAATTAAAAAGTCTGG
CCTACCCACTGAAGAATTGAAGGGCCGCGGTATTTTGACCGTGCAAAGGTAGCWTAATCATTAGTCTTTTAAT
TGAAGGCTGGTATGAATGGTTTAACGAGGTATTAACTGTCTCATTATAATTAAATTTGAATTTAACTTTTTAG
TCTTAAGGCTAGAATGATTTTAGAGGATGAGAAGACCCTATAGAGCTTTATAAATATAATTAATTGATTAGTT
ATGGGTGTTTATATTAATTTATAGATTAATTTATTTGGTTGGGGTGACTTGAAGATTAAATAAACTCTTTATT
TATAGTAACATAAATATATGGGATATTGATCCATTTTTAGTGATTAAAAGATTAAGTTACCTTAGGGATAACA
GCGTAATTTTTTTTAAGAGTTCTTATCGACAAAAAAGATTGCGACCTCGATGTTGGATTAAGGAATATTATTA
GGTGTAGAAGTTTGATGAATAGGTATGTTCGACCTTTAAATCCTTACATGATCTGAGTTTAGATCGGCGTAAG
CCAGGTCGGTTTCTATCCTTAATTTGTTAAAATATTTTAGTACGAAAGGACCAAATATTTAAAATAATTTTAT
TAAATTGAATATTAATAATGACTGCTTTGACAGATAAGTGTATTGGATTTAGAATTCATTAATGTAAATATGA
ATTTACAGGCAGTATTGAATATTTTAGATGTAGTTTTGTTATTATTAAGATCCATTTTATTATTAATTTGTGT
TTTGGTAGGGGTGGCTTTTTTGACGTTATTGGAGCGTAAGGTATTAGGTTATATTCAGATTCGCAAGGGCCCA
AATAAGGTTGGGATTGTTGGTATTCCTCAACCTTTTGCCGATGCAATTAAGTTATTTACTAAGGAACAAACTT
ATCCTAGTTTATCAAATTATCTTTTATATTATTTTTCGCCTATTATGAGATTGTTTTTGTCTTTATTAGTATG
AATAATTTTTCGTTATTTAACTTTTTTCTTTTCTTTTAATTTTGGGTTGTTATTTTTTTTGTGTTGTACGAGT
ATATGAGTATACACTGTAATGGTAGCGGGGTGGGCTTCAAATTCGAGTTATGCTTTATTGGGGGGATTGCGGG
CTGTTGCTCAAACGATTTCTTATGAAGTGAGATTGGCTTTGATTTTATTTTCTTTCATTTTTT

cytb-ND6

AAATTATCTCATGATATTAGAAGAAGGGGGTTAAATATGTAGTAGGTGAAATAAAGAATTGTTAATAATTGGC
CAGTTAGGATATAAGGGTCTTCTACAGGTCGGGCYCCRATTCATGTTAAAAGGATWACAATTACGGTTATAGA
TCARAATAGAATTTGATTTARGGGGTAGAATTGAAGACCYCGRAATTTATTRKTATTATAGAATGGGAGRATA
AATAAGATTGCGATGGATAARACTARTGCAATGACACCTCCTAATTTATTAGGGATTGAACGGAGAATTGCAT
ACGCAAATAAGAAATATCATTCTGGTTGGATGTGTACRGGGGTAACAAGGGGATTGGCTGGGATGAAGTTATC
RGGGTCYCCYARYATGTAAGGTTCGAGRAGGGTTAATAGGGTTARAAGAATAATRAGGATTARMAATCCTACA
ATGTCYTTGAATGARARGTAGGGRTGRAAGGGGATTTTATCAATATTTCTGTCTAAGCCTAGGGGATTATTTG
AACCTGTTTGATGAAGGAAAAGTAGATGAATTATTGTAGCTGCAGCAACAATAAAGGGTAGAACAAAATGGAA
GGTGAAGAATCGGGTTAATGTTGCATTATCAACAGCGAATCCCCCCCAAACTCATTGAACTAAGTCAATTCCT
AGATAGGGGATAGCTGATAGGAGGTTGGTAATAACTGTTGCCCCTCAGAATGATATTTGTCCTCAGGGAAGAA
CATATCCTATAAAAGCTGTAGCCATTATTAAGAAAAGAATGATTACACCTACTATTCAAGTATGGAGGAAGTT
GAAAGAGCCATAGTATATGCCTCGACCAATGTGCAGATAAATGCAAATGAAGAAGAATGATGCTCCATTAGCG
TGAAGGGTTCGTAAGAATCAGCCATAATTTACATCCCGGCAAATGTGGGCAATTCTGTTAAATGCTAGATCAA
TTTGGGCAGTGTAGTGTATAGCGAGGAATAAGCCAGTGGCAATTTGAATAATTAAGCATAGGCCTAATAGGGA
CCCAAAGTTCCATCAGGTTGARATATTGGATGGGGCTGGAAGGTCAACGAGGGCATTGTTGGCAATTTTGAAT
ARTGGGTGACTAGTTCGTATAGGTTTGTTCATTAATTTTTTTGGCGTAGAGGTCCTTGGAAAATATTGGTGAT
TTTTACGATAGCAATTAATGTCAAGAATAAATATAGTACTAGTATGAGTGTAATAATTTCGGTGGGGTTATTA
TACAGTTTTATTAGAGT

node 180 ND4

TACCTAGAATAGCTTTATGGTGATTTTTATTGAGATCCTCAAATATGGCTGCTCCTCCCTCTTTAAATTTGCT
GGGGGAAATTGGTTTGTTGAATAGAATTGTTTCGTGAACGTGAGTATCAATGATTACGTTTATGCTATTATCT
TTTT

node 2 tHis-ND5

AGTTTAATTAAAATATTGACTTGTGGGGTCAGAGATATGATATGATTTGTTTTAAGTTTATGATGTGGATATT
ATCAATTTGTTTGTTAAGATTTATTGTCTTATTTATTTTAAGATTGTTTTTAGTGAGTGTTGGATTTTATTTT
ATTATAAATGATTTAGTTTATTTTATTGAATGGGAGGTTTTAAGCGTTAATTCAAATTCAATTGTGTTGACTT
TTTTATTTGATTGAATATCCTTAATTTTTATGGGGTTTGTTTTATTTATTTCTTCTTTGGTTATTTTTTATAG
AGATGAATATATAAGAGGGGATGTAACAATTAATCGATTTATTGTTCTAGTATTAATGTTTGTTTTATCAATG
ATGTTTCTGATTATTAGACCTAATTTGATTAGAATTTTATTAGGGTGAGATGGGTTGGGATTAGTTTCCTATT
GTCTAGTAATTTATTATCAGAATGTAAAGTCATATAATGCGGGGATGTTAACGGCTTTATCTAACCGAATTGG
TGATGTGGCTTT

contig 480 ND5

AACATTAGTAACGGCAGGGGTTTATTTATTGATTCGATTTAATGCAATTATTTTTGATAATGGCTGGGGGCAA
ATTTTGTTGATGATTGCTGGGCTAACAATGTTTATAGCTGGATTAGGGGCAAATTTTGAATTTGATTTGAAAA
AAATTATTGCTTTGTCCACACTTAGACAGTTGGGGCTAATAATAAGAATTTTGGCGATAGGGTTTCCTAAGTT
GGCTTATTTTCATTTGTTGACTCATGCTTTATTTAAGGCGTTATTATTTATGTGTGCTGGGGCAATTATTCAT
AATATAAAAAATTCCCAGGATATTCGTAACATGGGGGCGTTAGTTATTCAAATACCTTTGACAGCTGCCTGTT
TTAATGTTTCTAATTTTGCTTTGTGTGGGATGCCTTTTTTAGCGGGGTTTTATTCTAAGGATTTGCTTTTAGA
AGTGGTTTCTTTATCTTATGTTAATGTATTTAGTTTTTTCCTTTACTTTTTCTCTACTGGATTGACTGTATGT
TATTCTTTTCGGTTGGCTTACTATTCAATAACTGGGGACTTGAATTGTTCTTCTTTTCATCCTTTAAATGATG
AAGGGTGGGTAATGTTGCGGGGGATGACTACATTAATAATTATGGCTGTGTTAGGGGGAAGAATGTTAAGATG
GGTATTGTTCCCCACACCCAGAATGATTTGTTTACCTTTGGGGCTAAAAACTATGGCATTAACGGTGAGATTA
TTAGGGGGTTGATTGGGTTATGAGTTGTCTAAATTTTCTTTAGTTTTTAGTTTACAATCTTTAAGATATCATT
TATCTTTAAGTTTTATAGGATC

contig 0536 tleu-CO3-ATP8-ATP6-tlys-CO2

GTTTGGGGAAAGTGGAAGAATACATGGCGTATTAAGCAGGTTAATAGGAAGGTAGTTCCAATAATTACATGGA
GTCCGTGGAACCCTGTGGCTACAAAAAAGGTCGAACCATACACAGCATCCGCAATAGTGAAGGGGGCTTCAAT
ATATTCGTAGGCCTGAAGAATAGAGAAATAGACCCCAAGTAAGACAGTAAAGAATAAACCTTGAGTTCTTTGT
CTAAAATTACCTTCTATAAGACTGTGGTGGGCCCAAGTTACTGTTACTCCTGATGCCAAAAGAATAGCTGTAT
TTAATAAAGGAATTTGGAATGGGTTAAAGGGTTGAATGCCTGCTGGGGGCCACATTGCACCTAATTCAATAGC
TGGGGATAGACTTCTATGGAAAAAGGCTCAGAAGAAGGAGATAAAGAAAAAAACTTCTGAGATAATAAATAAG
ATTATACCTCAGCGTAGCCCAAGATTAACTGGTAATGTGTGAAGGCCTTGATATGTACCCTCTCGAGTAATAT
CTCGTCATCATTGAATTATAGTTAAAATAGTGATTAAACTACCGAGTAAGATAAGAGAGTTATTATATTGATG
GAATCATTTGGCGAGGCCTGCGACAATAATTAGAGCGCCAATTGCCCCGGTAAGGGGTCAGGGTCTGGGATTG
ACTAAATGATAGGGATGATTGGCATGTGTGGACATAAGTTTATTAGTTTACTTCTCTAGAATATAAGGTGGCT
AGTACAGCAAAAACATATGATTGAATTATTGCAACTGCGGCTTCAAGAGTCAATAGGGCGATTTGAGCAACTA
CTAGGATTCTTAAGATAGAAATTGAAAGTGAAGGGCCAGTGTTACCAAGGAGGGTTAGTAGAAGATGTCCCGC
GATTATATTAGCTGCTAGCCGTACGGCTAGGGTACCTGGTCGGATAATGTTACTGATTGTTTCGATACAAACT
ATAAATGGTATTAGAATAGAAGGTGTTCCTTGGGGGACAAGATGAGCAAATATATGTTGAGTGTGGTTGATTC
AGCCAAAGATTATAAATCTTAATCATAGAGGAAGGGATAGAGTTAATGTTAAAGTCAAATGGCTAGTTCTTGT
AAAAACGTAGGGAAATAATCCTAGGAAGTTATTAAATAAGATTAGAGAAAATAGGGAGATAAAGATGAAGGTT
CTTCCATTTAAGCCTGTGGGGCCAATTAAAGTTTTGAATTCTTTGTGTAAAGTGTTGTTAATATTAGCTCAAA
TAATATTGAAGCGAGAGGGCAGGAATCAAAAAGCTATGGGGATGATGAGTAGGCCAGTAAAGGTTCTTAATCA
GTTTAATGACAGATTTATAATACTTGTTGAGGGGTCAAATACTGAGAATAAATTTGCTATCATTTTCAGTTAA
GAGATTTAAGGGTAATAGTATTTTCTTGAATTTTTGGGGTGTTAAGAAAGGGGGTGAAATAGTTTATAGTAGC
AAAGAGAATAAGAGTTAATGAAAATGTTATAAATAAAATTAATCAACTAATAGGGGCTRTTTGAGGAATAGAA
GGATATCAGACAAATCTAATATTTTTAGTATGACATACTAACGTTAATTTAACTAATCTTTCGGATATAGTGC
GTTTTATCACCAGAAGTAAATGTTAATTTACTATAGGATGGCTTAAGAGACCAGTACTTACTTTCAGTCATCT
GATGAGGCTTCAAATGCTCTTAATATTCAGGAAATAAAAGTTTTTGTATTTACTCTTTCGATAACAATAGGTA
TAAATCTATGATTTGCTCCACAGATTTCTGAACATTGGCCATAGAATAAACCGGGCCGGTTCATAAAAAATCT
GGTTTGATTAAGACGACCAGGGGTTGCGTCTACTTTTACACCTAGGGCTGGGACTGTTCAGGAATGAAGAACA

TCAGCGGCTGTGATTAGGATACGAATTTGGGTATTTACGGGGAGAATAGTCCGGTTATCTACATCCAGTAAAC
GAAATCCGTCTTTGCTTATTTCATTGTAAGGAATTATGTATGAGTCAAATTCATGGGGGGTTACAAAATCTAT
ATATTCATAGCTTCAGTATCATTGATGCCCTACAGTTTTAACTGTTACTAGGGGGTCTAAGGTTTCGTCTAGA
AGGTATAGAAGTCGGAGTGAAGGGAGTGCAATAAAAATAAGTGTAACGGCAGGTAGAATAGTTCAGATTACTT
CAATAGTTTGACCTTCAAGTAGGTAGCGGTGGGTAAATTTATTGAAGAATAGTGAACTTATAATATAGGCAAC
TAAAACAGTAAT

contig 440 tleu-CO1

AAGCTTACTAATCAGAGCCGAATTAGGTCAGCCTGGCTACCTAATTGGAGACGATCAAATTTACAACGTTATT
GTTACTGCCCATGCATTTGTTATAATTTTTTTTATAGTAATACCCATCATAATCGGAGGCTTCGGAAATTGAC
TCGTCCCCTTAATGCTAGGGGCCCCTGATATAGCTTTCCCCCGAATAAATAACATAAGCTTCTGACTTTTACC
CCCCTCTTTAACCCTACTGTTGGCCAGTAGTCTTGTCGAAAATGGTGCCGGCACAGGTTGAACTGTCTACCCC
CCTCTTTCAGCAGGAATTGCCCACGCTGGTGCTTCTGTAGACTTAGCAATCTTCTCCCTTCATTTAGCCGGAG
TCTCATCAATCCTGGGCGCAGTAAATTTTATTACTACTACCATTAATATACGGTCCCCAGGAATATCTCTAGA
CCAAACACCTTTATTTGTTTGAGCCGTAGCAATCACAGCCCTACTTTTATTACTATCTCTCCCTGTATTAGCA
GGGGCTATTACTATACTTCTAACCGATCGAAATCTTAATACTTCATTCTTTGACCCTGCAGGAGGAGGTGACC
CCATTCTTTACCAGCACCTATTCTGGTTTTTTGGCCATCCAGAAGTATACATTTTAATTTTACCCGGATTTGG
CATAATTTCTCATATTATTAGCCAAGAAAGTGGGAAAAAGGAAGCCTTCGGAACTTTAGGAATAATCTACGCT
ATACTAGCAATTGGCCTTCTAGGATTTGTTGTCTGAGCCCACCATATRTTTACGGTAGGTATAGACGTTGATA
CCCGAGCCTACTTTACATCAGCTACTATAATTATTGCCGTACCTACTGGAATTAAAATTTTCAGATGATTAGC
TACCTTACACGGTACCCAATTAAATTATAGACCTGCCTTACTTTGAGCCTTAGGATTTGTCTTCTTATTTACT
ATTGGGGGCTTAACTGGTTTTGTTCTAGCAAATTCCTCCTTAGACATCATTCTTCACGACACATATTATGTAG
TAGCCAATTTTCATTATGTTCTTTCTATAGGAGCTGTTTTTGCTATTATAGCTGGATTTATCCAATGATACCC
CCCATTCACAGGATTAACCCTCAACCCCATATGACTAAAAATTCAATTTTCCGTTATATTCATTGGGGTAAAT
TTAACCTTTTTCCCCCAACACTTCCTGGGCCTAGCAGGAATACCTCGCCATTACTCTGACTACCCTGATGCAT
TTACATCATGAAATATCGTTTCAAGACTAGGATCAACAATCTCTTTTATTGGAATCCTAATGCTAATTTTTAT
TATTTGAGAAAGCATAATTACTAACCGAACAGTTCTATTTCCCCTTAGCATGGTAAACTCACTAGAATGATAC
CAAAGCCTACCCCCAGCTGAACATAGTTACAATGAATTACCTATATTAACTAATTAATAACCTAATATGGCAG
ATAAGTGCAATGAACTT

A.8.2 Contigs that don't translate - possible pseudogenes

RP310.22 12/16S

12/16S

GTCGTAACATAGTAGATGTACTGGAAAGTGTATCTAGAAAGTTTAACGAAATAAAGCTTAAAGTGAAGCGTTC
CATTTACACTGGGTAGAAATTTGTGCGATTCGGATTGTTTTGAAATAAAAAATTTAAGTTTATTTGATTGTTG
ACTTTTTTATATGTTTAATAGAAGCAATTTTATTAATAGTTGTTTATGTATAGTTTATAGAATAAATTTTATA
ATTGATAGCGTATTAGTATTGTGAAAGATTAATTGAAATATTATGAAATATAAAATTGAAGAATAGTAGATTT
TTTATTGTACCTTGTGTATCAGGGTTGATTTAAGTATAGTTAAGAATTTAATATTCCCGATTTAAAGAGAGTT
AATGTAGTAAGTTTAGTTAATGTATTAAAATTATTGGAAATACTACATTAGAAGTGAGACGTTAGGCGTTCTT
TAAGATATCTGGTTTTCTAAGAATTGAATTTAATTCAGTTGTTAAGAGTATCTTTTTATATTAAGTAAGGGAT
ATAAATTAACATAAGAAGTTTGGGGGATAAGCTCTGAATTTATGTTTATAAATTAAAGGAATTATTAAGGATA
AGTAGGCTTTAAATTAGCTATCTTTAGGAGTATGTTATAATTTATTTATTGGTGTTTATGATTTTATAAGGGT
TTTAAATAAGTGTATTAGAATTATTTCAGAAATTGTGCTTGATTTGAAATAATGGTTGAATTAGTATAATAAA
ATTGTGTAGATTGAATCGATTTTGTTAATTTATAGTAAGGAATTAGGCAAATTTAAATGCTCGCCTGTTTATC
AAAAACATGTCTTTTTGACAATGAGATTAAAAAGTCTGGCCTGCCCACTGAAGAATTGAAGGGCCGCGGTATT
TTGACCGTGCAAAGGTAGCATAATCATTAGTCTTTTAATTGAAGGCTGGAATGAATGGTTTAACGAGGTATTA

Contig 437 cytB - ND1 pseudo

ATTTCATTAATAACTTATTATCAGCAATTCTGTACCTTGGAATTGACAGTCCAATGAGTTTGAGATGGATTTT
CAGTTGATAATGCTACCTTAACCCGATTCTTCACCTTCCCCTTTATTCTACCATTTATTGTAGCCGCGGCTAC
AATAATTCATTTATTATTCCTCGACCAAACCGGATCTAGGATTAGGTAGTAAGGTAGATAAAATTCCATTTCA
CCCCAACTTTACAAATAAAGACATTGTTGGATTTCTAATCTTAATCATCCTTCTTACCATGAACCTTACTTGA
ACCCTACATATTAGAAGACCCCGATAATCTTATTCCCGTCAACCCCATTGTTACCCCTGTTTATATTCTACCC
GAATGATACTTTCTCTTTGCTTATGCAATTATACGGTCTATCCCTAATGAATTAGGAGGCGCAATCCTCGTAA
TACAGTATATTGCCATCTTATTTATCTGACCATTTTATAAAAGAAATAAATTTCGAGATTTACAATTTTACCC
TTTAATTCAAACCTTATTTTGATTCATAACTATCATGGTCATCCTTTTAACATGAATTCGAGCCCGATCAGTT
TAAGATCCTTATATTTTAACAGGTCAACTATTAACTATTCTATATTTCACATATTATCTATTAAATCCTACGA
TATTAAACCTATGAGGCAATTTAGTTTCTTAAATTAGTTATTTAACATTCAAGAAGTATTTTGTTTTGAAAAC
ATAGGAAAGAAGTTCAACTCTTCTAAAACTTTTACTGAGTTAAATTCACCGAATTAACAAGGATAATTGTTTT
AAAGCCAACAAAAAAATTAATGATAATGGTAAAATACTCTTTTAAGCTAAGTACATCAATTTATCGAGACGAA
AACGAGGTAATGTCCCCCAAACACAAATAAATATAAAAGACAAAAAAGTCAACTTAATAAATAATATGAAAGA
ATTTACATCGCATCCTAAAATGATTACACAAAATAATATACTTATAAAAAGAATACTTGCACGCTCTGCCAAG
AAAATTAATGCAAATTCCCCTCTTCTATATTGTACGTATATCCTGATACTAACTCCGATTCCTCTCTCAGCAA
AATAAAATGGCGTACGATTTGTTTCTGCCAAGCAAGAAGCAAGTCATGCCAAATATAAAGGAAAAGTAATTAT
AAAAATCACAAATACCTGTGATAATAAATAAATACTCCCCAAAAATAACTATCAATTAAGTTATAAATGAAAA
TAAAATTAAAACTATTCTTATTTCATGTGAAATATGGGGATTTACGGGCTGTAGCTCAGACTAATAAAGCATT
ATTAGAATTTCAAGCACATCCAGCTACTATTACAGTATGTACACCTATTTTTGTACAACACAAAAAATAATAA
ACCAAAATTAAAATAAAATAAAAAAGTCAAATAAGGAAAAATTATTCACACCAATTAAGATAAACATAATTCA
ACATTTATTTTCCTATATTTGTCTAATTTTAGCTTCTTTTTTAGTAAA

ND4 His ND5 pseudo

AGGTCTTTTTATTTTAATAGTGTTGATTTTACTATTAATATTATATTGCACATTTAGAAGATTGAGATTATTT
TCTTTTTATTTGTTTTTTGAGGGTAGATTAATCCCAACTTTATTTTTGATTTTGGGTTGGGGGTATCAGCCTG
AGCGATTACAGGCGGGGGTTTATTTATTATTTTATACTTTGTTGGCTTCTTTACCTTTATTAGTGGGGTTATT
CAAGGTATACAATATATTAGGGAGTTTATATATGCCGGTGTTGGGGTTGCATATTGATCATATATTTTTTTAT
GTTTGTTTAATTTTAGCCTTTCTAGTAAAAATACCTATGTTTATATTTCATTTGTGACTTCCTAAGGCACATG
TTGAGGCTCCTGTTTCAGGGTCGATGATTTTGGCGGGGGTTTTATTAAAATTGGGGGGATATGGATTGTTGCG
CATTTATAAGATGTTGTCAATATCTGGATTAAAATATAATTATATTTCGGTAGGAATTAGTCTGGTTGGGTGA
GTGTTGGTGAGTTTAATTTGTTTACGGCAGACTGATTTGAAGTCTATAATTGCGTATTCTTCAGTGGCTCATA
TAGGAATTGCTTTAGGGGGTTGATGACAATAACATATTGGGGGTTTTGTAGAGCTTACACTTTAATAATTGCT
CAAGGACTTTGTTCTTCAGGGATGTTTTGTTTAGCTAACATTTCTTATGAGCGATTAGGGAGACGGAGTTTAT
TAATTAATAGGGGTTTGATAAATTTTATACCTAGAATAGCTTTATGGTGATTTTTATTGAGATCCTCAAATAT
GGCTGCTCCTCCCTCCTTAAATTTGCTGGGGGAAATTGGTTTGTTGAATAGAATTGTTTCGTGAACGTGAGTA
TCAATGATTACGTTGATGCTAATCTTTTTTTAGAGCGGCTTATTCATTGTATTTGTATTCCTACAGACAGCAT
GGGCAGATTTATTCGGGAGTTTATTCTTGTTCCATGGGGTATTCTCGGGAGTATTTATTGTTGTTTTTACATT
AATTCCCTTTAAATTTATTAATTGTGAAGGGTGATGTTTGTATGCTTTGGTTGTAGAGAGCTAGTATTACTTA
AATAGTTTAATTAAAATACTGACTTGTGGTGTCAAGGATATGATATGATTTGTTTTAGGTTTATGTTGTGGGT
ATTATCTATTTGTTTATTAAGATTTATTGTCCTATTTATTTTAAGATTGTTTTTAGTGGGTGTTGGGTTTTAT
TTTATTATAAATGATTTAGTTTACTTTATTGAATGGGAGGTTTTAAGCGTCAATTCAAGTTCAATTGTGATAA
CTTTTTTATTTGATTGAATATCTTTAATTTTTATGGGGTTTATTTTGTTTATTTCTTCTTTAGTTATTTTTTA
TAGAGATGAGTATATAAGAGGGGATGTAACAATTAATCGATTTATTATTTTAGTGTTAATATTTGTTTTATCA
ATGATATTCTTAATTATTA

Weta supercontig 1

AGCGGCGCCTACCCCCGTATCTGCATTGGTGCATTCTTCAACATTAGTAACGGCAGGGGTTTATTTATTGATT
CGGTTTAATGTAATTATTTTTGGTAATGGGTGGGGCCAAATTTTGTTGATGATTGCTGGGCTAACAATGTTTA
TAGCTGGATTAGGGGCAAATTTTGAATTTGATTTTTAAAAAATTATTGCTTTATCTACACTTAGACAGTTGGG
GTTAATAATGAGAA:TTTGGCGATAGGTTTCCCTAAATTGGCTTATTTTCATCTGTTGACTCACGCTTTATTT
AAGGCATTATTGTTTATATGTGCTGGGGCAGTTATTCATAATATAAAGAATTCTCAGGATATTCGTAATATAG
GGGCGTTAGTAACTCAAATGCCTTTGACAGCTGCTTGTTTTAATGTCTCTAATTTTGCTTTATGTGGGATGCC
TTTTTTAGCGGGGTTTTATTCTAAGGATTTACTTTTGGAAGTGGTTTCTTTATCCTATGTTAATATACTTAGA
TTTTTTCTT:TATTTTTTCTCTACCGGATTGACTGTATGCTATTCTCTTCGGTTGGCTTATTATTCAATAACT
GGGGATTTAAATTGTACATCTTTTCATCCTTTAAATGATGAAGGGTGGATTATGTTACGGGGAATAGTGACAT
TAATAATTATGGCTGTATTAGGGGGGAGAATGCTAAGTTGAGTATTGTTTCCTACACCTAGAATGATTTGTTT
ACCTTTGGGGTTAAAAACTCTGGCATTAACAGTGAGATTATTAGGGGGTTGGTTGGGCTATGAGTTGTCTAAA
TTTTCTTTAGTTTTTAGTTTGCAATCTTTAAAGTATTATTTATATTTAAGTTTTGTGGGGTCCATGTGATTTA
TGCCTTTTATTTCAACTTACGGGGTGAATTACGTTCCTTTAAAATTAGGCAATTATATAACTAAAAGAG:TAG
ATCAGGGTTGAAGAGAAGATTTCGGGGGGCAGTATTTGCATAATAATATTAG:AGAATACTCATTATTGGTGC
AATGATGGCAGAATAATAATGTAAAGATTTATTTAATTAGTTTTGTTTTATGAGTTATTATTTTAATAATATT
AGTTATTTTTTAAAGATTTAAGTAGCTTATAGGAAGAGTTTAACACTGAAGATGTTAAGGAGA:TAGAAGTAT
CCTTGAATAGTT:CAGATTTAATTTTGGATTAGTATTTATAAATATTATAATATTGTCTATACAATGAAAATG

TATTGTTTTTATTAAACTATA:TAAATGATTGAAGAAATGTAAACGGAATTGAACCGTTAAAGTTAAAAGTTA
GCAGCTTTTACTTG:A:CT:TTCACATTTCCTTAATTGGAAATAAATTTCAATAATATTATTAACAGTAATAT
ACCTCTTTTTGGTTTCAATTA:ATGTATTACTAAAATCCTTTCAGTAAATAAGGATACTAAGTTATCTAA::T
::GACCAGGTCGAAGTT:GGTTGCAATTGATTGCTTCTTATTTAAGATAAATTGATAGAACAATACTTT:TTG
AGTGCAAATCA:AATGTTATAATTAACTATAATCCTGTTAATGAGCTCATTCTAGGGCACCTTGATTTCATTC
ATGGAAT::AGCCCTACG:ATT:AGGATT:AGCAAGAAGAAT:ATTCTTACAA:TAGCCCATAGAGCGATATT
GGATCAATGTATAATAATAATGATGGGAAGAAGGAGGGCAATTTCTACATCAAAAATGAGGAAGATTACTGCA
ATTAGGAAGAATCGTAGGGAGAAGGGGAGGCGGGCTGATCCTTGGGGATCGAACCCACATTCAAAGGGGGAAT
TTTTTCTCGATCATTAATTATTTTTTTTGAAAGAATAGAGGCGAGAGTTATAACGATTATACAAATTGAAATA
:ATA:ATTG:AAGCTGTCAA:GTATAT:AGT:TAAA:ATTATTTGCT::TTTGGGAATCAAAT:CT:TTTTGA
TTGGAAGTCA:AACGTACTT:ATATACTAAGCAAATAATAAGATTTTAAATTAACTTCCTCATCAGTAAATGG
ATATGTAGAGGAATAATCATACT:A::CATCTACAAAGTGTCAGTATCAAGCTG:CAGCCTCAA:ATCCAAAA
TGATGGTTTGGGGAAAAGTGGAATAATACGTGACGTATGAGACR:GGTTAATAGGAAGGTAGTTCCAATAATT
ACATGGAGTCCG:TGGAAACCTGTGGCTACAAAAAAGGTCGAACCATATACGGCGTCAGCAATAGTGAAGGGG
GCTTCGATATATTCATAGGCTTGAAGAATAGAAAAATAGATTCCAAGTAAAACAGTAAAGAATAAACCCTGGG
TTC:T:TT:GTCTAAAATTACCTTCTATAAGACTGTGGTGGGCTCAAGTTACTGTCACTCCT:GA:TGCT:AA
AAGAATAGCT:GTATT:CA:ATAAAGGGATTTGAAATGGGTTAAAGGGTTGAATGCCTGCCGGGGGCCATATT
GCCCCTAATTCAATAGCTGGGGATAGACTTCTATGGAAAAAGGCTCAGAAGAAAGAGATAAAG:AAA::AAAA
TTT:CTGAGATAATAAACAAGATTATGCCTCAGCG:TAGCCCGAG:GTTAACTGGTAATGTGTGAAGACCTTG
ATATGTTCCTTCTCGGATGATATCTCGTCATCATTGGATTAT:A:GTTAAAATAGTGATTAAACCG:CCGAGT
AAGACGAGAGAGTTATTATATTGATGGAATCATTTGACGAGACCTAAGA:CAA:TAATTAG:AGCGCCAATTG
CCCCGGTGAGGGGTTAGGGTCTGGGATTGACTAAATGATAGGGATGATTGGCGTGTGTAGACATAAGTTTGTT
AGTTTACTTCTCTAGAATATAAGGTGGCTAGTACAGCAAAAACATA:T:GATTG:AATTATTGCAACTGCAGC
TTCGAGGGTTAATAGGGCGATTTG:GGCAATTACT:AGGATTCTTAGGATAGAAATTGAGAGTGAAGGGCCTG
TGTTACCAAGGAGAGTTAATAGGAGATGTCCTGCGATTATGTTAGCTGCTA:GTCGAACGGCTAGGGTACCTG
GCCGGATA:ATGTTACTGATTGTTTCGATGCAAACTATAAATGGTATTA::GAATAGAAGGGGTTCCTTGAGG
GACAAGGTGGGCAAAT:ATATGTTGAGTGTG:A::TTAATC:::CAGCCAAATATTAT:AAA:TCTTAGCCAA
AGAGGGAGGGAAAGGGT:TAG:TGT::TAAAGTTA:AATGGCTAGTTCTTGTAAAAACGTAGGG:AAAGAGTC
CTAGGAA:ATTATTGAATAGGATTAGA:GAGAATAAGGA::TATAA:AGATGAAGGTTCTTCCATTTAAGCCT
GTGGGGCCAATTAAAGTTTTGAATTCTTTATGTAAAGTGTTGGTAATATTGGCTCAAATGATATTGAAGCGAG
AGGGCAGGAATCAAAAAGTTATAGGGATGATGAGTAGGCCGATAAAGGTTCTTA:ATCAGTTTATTGATAGGT
TTATAATACTTGTTGAGGGGTCAAATAC:TGAGAATAGATTTGTTATCATTTTCAATTAAGAGAT:TTAAG:G
AT:AATGT:TATTTTCTTGAATT:TT:TGGGGCGTTGGGGAAGAG:GGTAAAGTAGTTTATTGTAGCAAAGAG
AATGAGAGTTAATGAAAAGA:TTAT:AAAT:AAAATTAATCAACTAATAGGAGCTATTTGAGGAATAGAAGGA
TATTAGATAATCTAA:TATTTTTAGTATGAC:ATACTAAT:GTTAAT:TTAACTAATCTTTCG:GGCACATTG
CGTTTATCACCAGAAGTAAATGTTAATTTACTATAGGA:::T:GGCTT:AAGAGA:CCAGTACTT:ACTTTCA
GTCATCTGATGAGGCTTCAAATGCTCTTAATATTCAGGAAATAAAAGTTTTTGTATTTACTCTTTCGATAACG
ATAG:GTATGAATCTATGATTTGCTCCGCAGATTTCTGAACACTGGCCGTAGAATAAACCAGGCCGGTTTATA
AAAAATCTGGT:TTGATTGAGACGGCCAGGGGTTGCGTCCACCTTTACACCTAGAGCTGGGACTGTTCAGGAA
TGAGGACATCAGCGGCTGTA:ATTAGGATG:CGAATTTGGGTATTTACAGGAAGAATAGTTCCGTTATCTACA

TCCAGTAGACGAAATCCGTCTTCGTTTATTTCATTGTAGGGAATTAT:GTATGAGTCAAATT::CGT:GAGGA
GTTACAA:AGT:CTA:TGTATTCATAACTTCAATATCATTGATGGCCTACAGTTTTAACTGTTACTAGGGG

Contig 0535 ND3-CO2

AATTAGCGGCGCGATTCGCCCTTGATCATTAATTATTTTTTTGAAAGAATAGAGGCCAGGGTTATAACAATT
ATACAAATTAAGATAATAATGGAAGCTGTCAAGTGTATAGTTAAGATTATTTACTTTTGAGAATCAAATCTTT
TTGATTGGAAGTCAAATGTACTTAATATACTAAGCAAATAATAAGATTTTAAATTAACTTCCTCATCAGTAAA
TGGAGATGTAGAGGAATAACCATACTACATCTACAAAGTGTCAGTATCAAGCTGCAGCCTCAAATCCAAAATG
ATGGTTTGGGGAAAAGTGGAATAATACGTGACGTATGAGACAGGTTAATAGGAAGGTAGTTCCAATAATTACA
TGGAGTCCGTGGAAACCTGTGGCTACAAAAAAGGTCGAACCATATACGGCGTCAGCAATAGTGAAGGGGGCTT
CGATATATTCATAGGCTTGAAGAATAGAAAAATAGATTCCAAGTAAAACAGTAAAGAATAAACCCTGGGTTCT
TTGTCTAAAATTACCTTCTATAAGACTGTGGTGGGCTCAAGTTACTGTCACTCCTGATGCTAAAAGAATAGCT
GTATTCAATAAAGGGATTTGAAATGGGTTAAAGGGTTGAATGCCTGCCGGGGGCCATATTGCCCCTAATTCAA
TAGCTGGGGATAGACTTCTATGGAAAAAGGCTCAGAAGAAAGAGATAAAGAAAAAAATTTCTGAGATAATAAA
CAAGATTATGCCTCAGCGTAGCCCGAGGTTAACTGGTAATGTGTGAAGACCTTGATATGTTCCTTCTCGGATG
ATATCTCGTCATCATTGGATTATAGTTAAAATAGTGATTAAACYRCCGAGTAAGAYGAGAGAGTTATTATATT
GATGGAATCATTTGACGAGACCTAAGACAATAATTAGAGCGCCAATTGCCCCGGTGAGGGGTYAGGGTCTGGG
ATTGACTAAATGATAGGGATGATTGGCGTGTGTAGACATAAGTTTGTTAGTTTACTTCTCTAGAATATAAGGT
GGCTAGTACAGCAAAAACATATGATTGAATTATTGCAACTGCAGCTTCGAGGGTTAATAGGGCGATTTGGGCA
ATTACTAGGATTCTTAGGATAGAAATTGAGAGTGAAGGGCCTGTGTTACCAAGGAGAGTTAATAGGAGATGTC
CTGCGATTATGTTAGCTGCTAGTCGAACGGCTAGGGTACCTGGCCGGATAATGTTACTGATTGTTTCGATGCA
AACTATAAATGGTATTAGAATAGAAGGGGTTCCTTGAGGGACAAGGTGGGCAAATATATGTTGAGTGTGATTA
ATCCAGCCAAATATTATAAATCTTAGCCAAAGAGGGAGGGAAAGGGTTAGTGTTAAAGTTAAATGGCTAGTTC
TTGTAAAAACGTAGGGAAAGAGTCCTAGGAAATTATTGAATAGGATTAGAGAGAATAAGGAGATAAAGATGAA
GGTTCTTCCATTTAAGCCTGTGGGGCCAATTAAAGTTTTGAATTCTTTATGTAAAGTGTTGGTAATATTGGCT
CAAATGATATTGAAGCGAGAGGGCAGGAATCAAAAAGTTATAGGGATGATGAGTAGGCCGATAAAGGTTCTTA
ATCAGTTTATTGATAGGTTTATAATACTTGTTGAGGGGTCAAATACTGAGAATAGATTTGTTATCATTTTCAA
TTAAGAGATTTAAGGATAATGTTATTTTCTTGAATTTTTGGGGCGTTGGGGAAGAGGGTAAAGTAGTTTATTG
TAGCAAAGAGAATGAGAGTTAATGAAAAGATTATAAATAAAATTAATCAACTAATAGGATCTATTTGAGGAAT
AGAAGGATATTAGATAATCTAATATTTTTAGTATGACATACTAATGTTAATTTAACTAATCTTTCGGGCACAT
TGCGTTTATCACCAGAAGTAAATGTTAATTTACTATAGGATGGCTTAAGAGACCAGTACTTACTTTCAGTCAT
CTGATGAGGCTTCAAATGCTCTTAATATTCAGGAAATAAAAGTTTTTGTATTTACTCTTTCGATAACGATAGG
TATGAATCTATGATTTGCTCCGCAGATTTCTGAACACTGGCCGTAGAATAAACCAGGCCGGTTTATAAAAAAT
CTGGTTTGATTGAGACGGCCAGGGGTTGCGTCCACCTTTACACCTAGAGCTGGGACTGTTCAGGAATGAGGAC
ATCAGCGGCTGTAATTAGGATGCGAATTTGGGTATTTACAGGAAGAATAGTTCGGTTATCTACATCCAGTAGA
CGAAATCCGTCTTCGTTTATTTCATTGTAGGGAATTATGTATGAGTCAAATTCGTGAGGAGTTACAAAGTCTA
TGTATTCATAACTTCAATATCATTGATGGCCTACAGTTTTAACTGTTACTAGGGGTCCAG

weta superconig 2

TGTAAGGTAGCTAATCATCTGAAAATTTTAATTCCAGTAGGTACGGCAATAATTATAGTAGCTGATGTAAAGT
AGGCTCGGGTATCAACGTCTATACCTACCGTAAATATATGGTGGGCTCAGACAACAAATCCTAGAAKGCCAAT
TGCTRGTATAGCGTAGATTATTCCTAAAGTTCCGAAGGCTTCCTTTTTCCCACTTTCTTGGCTAATAATATGA
GAAATTATGCCAAATCCGGGTAAAATTAAAATGTATACTTCTGGATGGCCAAAAAACCAGAATAGGTGCTGGT
AAAGAATGGGGTCACCTCCTCCTGCAGGGTCAAAGAATGAAGTATTAAGATTTCGATCGGTTAGAAGTATAGT
AATAGCCCCCTGCTAATACAGGGAGAGATAGTAATAAAAGTAGGGCTGTGATTGCTACGGCTCAAACAAATAA
AGGTGTTTGGTCTAGAGATATTCCTGGGGACCGTATATTAATGGTAGTAGTAATAAAATTTACTGCGCCCAGG
ATTGATGAGACTCCGGCTAAATGAAGGGAGAAGATTGCTAAGTCTACAGAAGCACCAGCGTGGGCAATTCCTG
CTGAAAGAGGGGGGTAGACAGTTCAACCTGTGCCGGCACCATTTTCAACAAGGCTACTGGCCAACAGTAGGGT
TAAAGAGGGGGGTAAAAGTCAGAAGCTTATGTTATTTATTCGGGGGAAAGCTATATCAGG:GCCCCTAGCATT
AAGGGGACGAGTCAATTTCCGAAGCCTCCGATTATGATGGGTATTACTATAAAAAAAATTATAACAAATGCAT
GGGCAGTAACAATAACGTTGTAAATTTGATCGTCTCCAATTAGGTAGCCAGGCTGACCTAATTCGGCTCTGAT
TAGTAAGCTTAGGGAAGTGCCTACCATTCCAGCTCAAGCGCCAAAAATAAAGTAAAGAGTTCCAATATCCTTA
TGATTTGTTGAGAATAATCATTGTTGCGGTAGAATGGCTGAGT:TAAAGGTAATAAATTGTAAATTTATGAAG
GAGAGT:TTTCTCTTCTACCAAAACTAAGGATGAGCTTTATAGTCAAATTATGATGTTAGACTGCAATTCTAA
GGGTGTAAAGTTGAATTTACTAAGGCTT:AAAATCTTGTAATTTTATTGACAGCTTTGAAGGCTATTAGTTT:
:ATTTAACTTAAAACCTTGATTGGAAATCAAGGTTTAAATGATGTTAAATAAAGTTGTGCATATTAGTAACCC
TAGGTTTGATAGTATTG:CGA:GT:AGAACA:A:TGGAAATTAGGGTAGAATT::TT:TATAGC:TTAGGGAG
AAAATTCACTTTAATTCAGTGTAAGATAATAGAAACGCTGAGAAACAAATTCGTAAGTAAAAGAATAG:GGTG
ATTAGGGTTATTACAACTATTACAGTAATAATTGGAAGTATACCTAATTCTGTTATAGATTGGATGATGATTC
ATTTGGGTAAGAAGCCTAGGAAGGGGGGCAAATCACCTAGGGAAAGTAAAAGAGAG:AATAAACAGAATTTTA
ATATGGGGTTCTTGTTTATTATTAGAAAACCTTGACTAATATGAAACACTTGGAAGGTGTTAAAAGTGAAAAT
TAATGAGGTATTTAAAAAGGTATAGATTAAGAAGTAAAGTTCTCACAGATTTTCTCCTGAAATTATGGCAGCA
ATTATTCAACCTAAGTGGTTAATTGAGGAATAAGCTATTAATTTCCGTAGAGATGTTTGGTTGAAGCCCCCAA
GTGAGCCTACTAGCACCGAGGCCAGGATTACTAGGGTGGAGAATAAGC

Contig 493: CO1-ND2 Pseudo

ATTCTCCACCCTAGTAATCCTGGCCTCGGTGCTAGTAGGCTCACTTGGGGGCTTCAACCAAACATCTCTACGG
AAATTAATAGCTTATTCCTCAATTAACCACTTAGGTTGAATAATTGCTGCCATAATTTCAGGAGAAAATCTGT
GAGAACTTTACTTCTTAATCTATACCTTTTTAAATACCTCATTAATTTTCACTTTTAACACCTTCCAAGTGTT
TCATATTAGTCAAGGTTTTCTAATAATAAACAAGAACCCCATATTAAAATTCTGTTTATTCTCTCTTTTACTT
TCCCTAGGTGATTTGCCCCCCTTCCTAGGCTTCTTACCCAAATGAATCATCATCCAATCTATAACAGAATTAG
GTATACTTCCAATTATTACTGTAATAGTTGTAATAACCCTAATCACCCTATTCTCTTACTTACGAATTTGTTT
CTCAGCGTTTCTATTATCTTACACTGAATTAAAGTGAATTTTCTCCCTAAGCTAAAAATTCTACCCTAATTTC
CATTGTTCTACTCGCAATACTATCAAACCTAGGGTTACTAATATGCACAACTTTATTTAACATCATTTAAACC
TTGATTTCCAATCAAGGTTTTAAGTTAAATAAACTAATAGCCTTCAAAGCTGTCAATAAAATTACAAGATTTT
AAGCCTTAGTAAATTCAACTTTACACCCTTAGAATTGCAGTCTAACATCATAATTTGACTATAAAGCTCATCC
TTAGTTTTGGTAGAAGAGAAAACTCTCCTTCATAAATTTACAATTTATTACCTTTAACTCAGCCATTCTACCG
CAACAATGATTATTCTCAACAAATCATAAGGATATTGGAACTCTTTACTTTATTTTTGGCGCTTGAGCTGGAA
TGGTAGGCACTTCCCTAAGCTTACTAATCAGAGCCGAATTAGGTCAGCCTGGCTACCTAATTGGAGACGATCA

AATTTACAACGTTATTGTTACTGCCCATGCATTTGTTATAATTTTTTTTATAGTAATACCCATCATAATCGGA
GGCTTCGGAAATTGACTCGTCCCCTTAATGCTAGGGGCCCTGATATAGCTTTCCCCCGAATAAATAACATAAG
CTTCTGC

Node 20

AGGTCTTTTTATTTTAATAGTGTTGATTTTACTATTAATATTATATTGCACATTTAGAAGATTGAGATTATTT
TCTTTTTATTTGTTTTTTGAGGGTAGATTAATCCCAACTTTATTTTTGATTTTGGGTTGGGGGTATCAGCCTG
AGCGATTACAGGCGGGGGTTTATTTATTATTTTATACTTTGTTGGCTTCTTTACCTTTATTAGTGGGGTTATT
CAAGGTATACAATATATTAGGGAGTTTATATATGCCGGT:GTTG:GGGTTGCA:TATTGATCATATATTTTTT
TATGTTTGTTTAATTTTAGCCTTTCTAGTAAAAATACCTATGTTTATATTTCATTTGTGACTTCCTAAGGCAC
ATGTTGAGGCTCCTGTTTCAGGGTCGATGATTTTGGCGGGGGTTTTATTAAAATTGGGGGGATATGGATTGTT
GCGCATTTATAAGATGTTGTCAATATCTGGATTAAAATATAATTATATTTCGGTAGGAATTAGTCTGGTTGGG
TGAGTGTTGGTGAGTTTAATTTGTTTACGGCAGACTGATTTGAAGTCTATAATTGCGTATTCTTCAGTGGCTC
ATATAGGAATTGCTTTAGGGGG:TTGATGACAATAACATATTGGGGGTTTTGTAGAGCTTACACTTTAATAAT
TGCTCAAGGACTTTGTTCTTCAGGGATGTTTTGTTTAGCTAACATTTCTTATGAGCGATTAGGGAGACGGAGT
TTATTAATTAATAGGGGTTTGATAAATTTTATACCTAGAATAGCTTTATGGTGATTT

APPENDIX B

General Materials

B.1 LABORATROY EQUIPMENT

B.1.1 General

Table B-1 General laboratory equipment used

Item	Source
Microcentrifuge	Eppendorf, Germany
Transilluminator (wavelength: 302nm)	UVP Incorporated
Video Camera	Panasonic
Agarose Gel rig	BioRad, USA
Microwave Oven	Panasonic
Milli-Q System	Milli-Q
Peltier MJ Thermal Cycler PTC-200	Applied Biosystems, USA
Biometra T1 Thermocylcer	Whatman Biometra, Germany

B.1.2 Glassware

Table B-2 General glassware used

Item	Source
screw lid bottles (50ml)	Schott, Germany
screw lid bottles (100ml)	Schott, Germany
screw lid bottles (250ml)	Schott, Germany
screw lid bottles (500ml)	Schott, Germany

B .1.3 Consumables

Table B-3 List of general consumables used

Item	Source
Falcon tubes (50ml)	Greiner bio-one, Germany
Filter pipette tips (1000ml)	Axygen Scientific, U.S.A
Filter pipette tips (200µl)	Axygen Scientific, U.S.A
Filter pipette tips (20µl)	Axygen Scientific, U.S.A
Filter pipette tips (10µl)	Axygen Scientific, U.S.A
Filter pipette tips (2µl)	Axygen Scientific, U.S.A
Kimwipe (paper towel)	Kimberly-Clark, Thailand
Microcentrifuge tubes (1.6ml)	Axygen Scientific, U.S.A
Microcentrifuge tubes (0.6ml)	Axygen Scientific, U.S.A
PCR tubes (0.2ml)	Axygen Scientific, U.S.A
Scaple blade (size 22)	Swann Morton Limited, England

B.2 CHEMICALS AND REAGENTS

B.2.1 Laboratory chemicals and reagents

Table B-4 List of chemicals and reagents used

Item	Source
DNA grade Agarose	Probiogen Biochemicals
bacteriological agar	Oxoid, England
Bromophenol blue	Serva, Germany
cetyl trimethyl ammonium bromide (CTAB)	Sigma, U.S.A.
Chloroform	BDH, England
Chloroform:Isoamyl alcohol 24:1	Sigma, U.S.A.
dNTP	Boehringer Mannheim, Germany
Ethanol p.A. 99.8-100%	BDH, England
Ethidium bromide	Invitrogen, U.S.A
Ethylenediaminetera-acetic acid (EDTA)	BDH, England
Ficoll Ty 400	Pharmarcia
Glacial acetic acid (0.05% (v/v))	BDH, England
Isopropanol	BDH, England
Buffered Phenol	Invitrogen, U.S.A
Potassium acetate (KOAc)	Sigma, U.S.A.
Poluvinyl-pyrrolidone (PVP)	Sigma, U.S.A.
Sodium acetate (NaAc)	BDH, England
Sodium chloride (NaCl)	BDH, England
Sodium hydroxide (NaOH)	Sigma, U.S.A.
Tris-acetate	BDH, England
Tryptone	MERK, Germany
xylene cyanol	Sigma, U.S.A.
Yeast extract	MERK, Germany
X-Gal	Progen
Ampicillin	Sigma, U.S.A.
IPTG	Progen

B.2.2 Enzymes

Table B-5 List of Enzymes used

Item	Source
EcoRI	Roche Applied Science, Mannheim, Germany
Expand™ Long template PCR System	Roche Applied Science, Mannheim, Germany
HindIII	Roche Applied Science, Mannheim, Germany
MSE	
Proteinase K	Sigma, U.S.A.
Red Hot® *Taq*	ABgene, UK
Rnase A	Sigma, U.S.A.
T4 Ligase (2000 U/µl)	New England Biolabs, U.S.A.

B.2.3 DNA Markers

Table B-6 DNA markers used

Item	Source
1Kb plus DNA™ Ladder	Life Technologies, U.S.A.
Low DNA Mass™ Ladder	Life Technologies, U.S.A.
High DNA Mass™ Ladder	Life Technologies, U.S.A.
SybrSAFE™	Invitrogen, U.S.A.

B.3 KITS AND READY-TO-USE PRODUCTS

Table B-7 Kits and ready-to-use products

Item	Source
Dye Terminator Sequencing Mix	Perkin-Elmer Corporation, U.S.A.
High Pure™ PCR Template Preparation Kit	Roche Applied Science, Mannheim, Germany
Max efficiency DH5∞™ competent cells	Life Technologies, U.S.A.
PGEM-T Easy	Promega, U.S.A.
QIAquick® Gel Extraction Kit	Qiagen GmbH, Hilden, Germany
One Shot® TOP 10 Chemically Competent *E.coli*	Invitogen, U.S.A.
TOPO TA cloning® kit (sequencing aves)	Invitrogen, U.S.A.

B.4 SOLUTIONS AND DYES

TE

1M TRIS (pH 8.0), 500mM EDTA (pH 8.0) and Milli-Q water

TAE

1M TRIS (pH 8.0), Acetic acid, 500mM EDTA (pH 8.0) and milli-Q water

TNES (Extraction Buffer for Salting Out Method)

1M TRIS (pH 8.0), 5M NaCl, 500mM EDTA (pH 8.0), 5% SDS, milli-Q water

10× Loading Buffer

Bromophenol blue (0.16% final conc.), Xylene cyanol (0.16% final conc.), Ficoll (25% final conc.), milli-Q water

High Mass Ladder

TE (10mM Tris, 1mM EDTA), High Mass Ladder (Invitrogen), 10× Loading Dye

Low Mass Ladder

TE (10mM Tris, 1mM EDTA), Low Mass Ladder (Invitrogen), 10× Loading Dye

1Kb+ Ladder

TE (10mM Tris, 1mM EDTA), 1Kb+ Mass Ladder (Invitgrogen), 10× Loading Dye

B.5 SOFTWARE AND PHYLOGENETIC PROGRAMS

Table B-8 Software and programs used

Item	Source
BEAST	http://beast.bio.ed.ac.uk/
Garli	http://www.zo.utexas.edu/faculty/antisense/garli/Garli.html
MRBAYES	http://mrbayes.csit.fsu.edu/
Oligo	http://www.oligo.net/
PAUP*	http://paup.csit.fsu.edu/
PhyML	http://atgc.lirmm.fr/phyml/binaries.html
RAxML	http://icwww.epfl.ch/~stamatak/index-Dateien/Page443.htm
Se-Al	http://tree.bio.ed.ac.uk/software/seal/
SEQUENCHER	http://www.sequencher.com/?referrer=Google_AP-D_names&gclid=COeZ2fe6tpQCFQgfswodngbyTw
SplitsTree	www.splitstree.org
Tracer	http://tree.bio.ed.ac.uk/software/tracer/

APPENDIX C

Research is what I'm doing when I don't know what I'm doing.

~Wernher Von Braun

C.1 PRIMERS

Oligonucleotides were ordered from either Invitrogen or Sigma Aldrich. Stocks were rehydrated with milli-Q water to a final concentration of 1 nmol/µl and stored at -80°C. From this, a working stock used in all PCR and sequencing reactions of 10 pmol/µl was kept at -20°C.

In order to amplify certain regions, primers had to be modified and/or redesigned. Oligonucleotide primers were designed using the program Oligo 4.0 (http://www.oligo.net/). For the nuclear gene 18S, internal/nested primers were designed and used for difficult samples (*Chapters Two* and *Three*). For amplification of the whole mitochondrial genome of *Hemiandrus pallitarsis* and *Hemideina crassidens* (*Appendix A*) specific long range primers for the large ribosomal subunit gene (16S) were designed from previously amplified sequences. In addition, fragments generated from restriction enzyme digestion were used as template DNA to design internal primers for ND5, CytB, ATP 6 and ATP 8 (*Appendix A*).

C.2 EXTRACTION PROTOCOLS

C.2.1 Salting Out

1. Muscle tissue was taken from the hind femur and put into a 1.6 mL microcentrifuge tube using clean forceps.

2. 10µL of Proteinase K (Sigma; 10 mg/ml in water) 600µL TNES (1M Tris [pH 8.0], 5M NaCl, 500mM EDTA [pH 8.0], 5% SDS) were added and mixed then incubated either at 37˚C overnight or at 55˚C for 1 - 3 hours.

3. 170µL of 5M NaCl was added after incubation and mixed by repeated inversion for 15sec

4. Samples were then spun at 14,000 rpm (or top speed) for 5min

5. Supernatant was pipetted into new tube and an equal volume of ice cold EtOH was added and mix by repeated inversion (*ca* 780µL)

6. Samples were then spun for 5min at 14,000 rpm

7. EtOH was removed by pouring leaving a DNA pellet

8. The DNA pellet was rinsed with 400µL of 70% (v:v) EtOH and then spun for 5min at 14,000 rpm to remove salt residues

9. EtOH was remove and samples allowed to air dry

Samples were re-suspended in 20 - 50μL 1× TE (T_{10} E_1:10mM Tris-HCl, 1mM EDTA [ph 8.0] or milli-Q water)

Quantification was carried out by running samples 2μL on a 1% (w:v) agarose gel in 1× TAE and then dilute appropriately to a working concentration of 10ng/ μL.

C.2.1.1 References

Aljanabi, S.M. and Martinez, I. (1997). Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Research*. **25**, 4692-4693.

Sunnucks, P., England, P.E., Taylor, A.C. and Hales, D.F. (1996). Microsatellite and chromosome evolution of parthenogenetic *Sitobion* aphids in Australia. *Genetics*, **144**, 747-756.

C.2.2 Rapid mtDNA Extraction (Alkaline Lysis)

Buffer A
0.25M sucrose
10mM EDTA
10mM Tris-HCL [pH 7.5]

Buffer B
0.15M NaCl
10mM EDTA

1. Individual specimens were anesthetised in the -20˚C freezer and tissue taken from the hind femur (due to the size of the organism this was less than 1 g).
2. Tissue was homogenised with cold Buffer A to “suck” apart the tissue in order to maintain whole mitochondria.
3. Homogenised solution was transfered to new chilled tubes with cooled tips and spun in a refrigerated microfuge at 4˚C for 2min at 2,000 rpm

4. The top layer supernatant was removed and the centrifugation repeated 4 - 5 times until the liquid was clear. This process pellets the cell debris at the bottom of the tube while leaving the mitochondria in suspension. Cell debris was kept and used for normal salting out extractions.

5. Suspended samples were spun at 14,000 rpm for 10min to pellet mitochondria at the bottom of the tube

6. The pellet was resuspended in 50µL of Buffer B

7. 100µL of 0.18M NaOH/1% SDS (w:v) was added and gently mixed (by repeated inversion) then placed on ice for 5min to lyse mitochondria

8. 75µL 5M KOAc (pH 5.0 potassium acetate, see Maniatis) was added and samples spun at 14,000 rpm for 5min

9. This was followed by a standard phenol/chloroform (24:1) extraction and DNA precipitation (supernatant plus equal volume [112µL] of phenol and same of chloroform/isoamyl alcohol, shake, spin, and remove supernatant. Repeat with chloroform/isoamyl only. Precipitate by adding 1/10th the volume of 5M NaCl and spin. Add supernatant to 2 volumes of ice cold absolute EtOH, spin and wash pellet with 70% [v:v] EtOH. Remove EtOH and re-suspend DNA in TE or milli-Q water).

C.2.3 Bird Kits

Roche's High Pure PCR Template Preparation Kit was used to extract DNA from tissue samples for all three birds. The manufacturer's instructions were followed below;

1. 25 - 50 mg of tissue was added to 1ml tube

2. 200µL of Tissue Lysis Buffer and 40µL proteinase K were added and mixed immediately. Samples were incubated at 55°C for at least 1hour or until the tissue is digested

3. After incubation, 200µL of Binding Buffer was added and mixed by repeated inversion then incubated at 72°C for 10min

4. 100µL isopropanol was added and mixed well

5. The liquid sample was pipetted into the upper reservoir of a combined High Pure filter tube-collection tube assembly

6. Samples were centrifuged for 1min at 8,000 rpm

7. Flowthrough and the collection tube were discarded. The filter tube was then combined with a new collection tube and 500µL of Inhibitor Removal Buffer added.

8. Samples were again centrifuged for 1 min at 8,000 rpm

9. Discard the flowthrough and collection tube then combine the filter tube with a new collection tube and add 500µL of Wash Buffer to the upper reservoir.

10. Centrifuge for 1 min at 8,000 rpm

11. Flowthrough and the collection tube were discarded. The filter tube was then combined with a new collection tube and 500µL of Wash Buffer was added.

12. Samples were centrifuged for 1 min at 8,000 rpm

13. Flowthrough and the collection tube were discarded. The filter tube was then combined with a new collection tube and 500µL of Wash Buffer was added to the upper reservoir and centrifuge for 1 min at 8,000 rpm

14. Flowthrough was discarded and again combined with the filter tube with the same collection tube then centrifuge again at maximum speed (14,000 rpm) for approximately 10 sec to remove any residual Wash buffer

15. Collection tubes were discarded and the filter inserted into a clean 1.5mL eppi tube

16. 200µL pre-warmed (70°C) Elution Buffer was added to the filter tube and centrifuge for 1 min at 8,000 rpm

17. Steps 15 - 16 were repeated

18. Quantify the eluted DNA by running on a 1% (w:v) agarose gel and store stock at -20°C

C.3 PCR PROTOCOL

C.3.1 Long Range

Long range PCR was carried out for both invertebrate samples (weta) and vertebrate samples (aves). Both reactions used the same kit and enzyme, Expand Long Template *Taq* from Roche. Reactions were carried out in a volume of 50µL with the following per reaction;

Table C-1 Long range PCR mix

Long range PCR reaction mix. All reagent volumes are in μL with the total volume of 50μL.

Long Range PCR	
Reagent	**x1**
H_2O	37.8
10x Buffer	5.0
dNTPs (2mM)	2.5
P1 (10μM)	1.5
P2 (10μM)	1.5
Taq	0.750

Primers used for weta amplification were designed for this project from template DNA that was previously sequenced from the target species and allowed the amplification of the whole mitochondrial genome in one fragment (with the exception of *ca* 800 bp from which the primers were originally designed). PCR reactions were carried out on a Biometra T1 Thermocycler and were optimised for each of the two species. The following format was successful for both *Hemiandrus pallitarsis* and *Hemideina crassidens*; 94°C for 2min, followed by 10 cycles of 94°C for 30sec and 68°C for 14min followed by 25 cycles of 94°C for 30sec and 68°C for 14min with a final extension of 68°C for 7min.

Table C-2 Long range primers used for Anostostomatidae

Long range primers used for DNA amplification for both *Hemiandrus pallitarsis* and *Hemideina crassidens*. The combination 16S-RP1F and 16S-RP1R was used in most instances however the combination 16S-RP1Frs and 16S-RP1Rrs were used for instances were restriction enzyme digestion of the product was to be used. The extra sequence at the start allowed for sequencing of the end fragments.

Name	DNA Sequence	Orientation	Source
16S-RP1F	ACA CAT CGC CCG TCG CTC TCA TTT TT	5' light/forward	This study
16S-RP1R	GTA ATC TTA TTG TAA TCC ACC ACC ACT T	3' heavy/reverse	This study
16S-RP1Frs	GTG AAT TCA AGC TTA CAC ATC GCC CGT CGC TCT CAT TTT T	5' light/forward	This study
16S-RP1Rrs	GTG AAT TCA AGC TTG TAA TCT TAT TGT AAT CCA CCA CCA CTT	3' heavy/reverse	This study

Standard long range PCR primers for aves were used to amplify overlapping fragments of *ca* 10 - 12Kb in length. In the case of the great potoo, additional primer combinations were required for shorter fragments (*ca* 6Kb) to generate the whole genome. The standard PCR cycle followed the same format for all three species of birds; initial denaturation of 93°C for 3 min, followed by 9 cycles of 94°C for 30sec, 48 - 50°C for 30sec and 68°C for 10 min; followed by a further 23 cycles of 94°C for 30sec, 48-50°C for 30sec and 68°C for 10 min plus 20 seconds per cycle, with a final extension of 68°C for 20min with a final pause of 10°C.

Table C-3 Long range primers used for the dollar bird

Long range primers used for the amplification of *Eurystomus orientalis* (dollar bird). The combination of Av1753F12S to Av13026tLeuR-LR and Av10647COIIIF to Av2901R16S or AV16137tPro amplified overlapping fragments from which short range PCR was then carried out.

Long Range PCR Primers	Sequence
Av1753F12S	AAA CTG GGA TTA GAT ACC CCA CTA T
Av13026tLeuR-LR	CTT GGA KTT GCA CCA AGR TDV TTG GTT CCT AAG ACC A
Av10647COIIIF	TTT GAA GCA GCA GCC TGA TAY TG
Av16137tProR	ARA ATR CCA GCT TTG GGA GTT GG
Av2901R16S	GCA CTC TTT GTT GRT GGC TGC TT

Table C-4 Long range primers used for the great potoo

Long range primers used for the amplification of *Nyctibius grandis* (great potoo). The combination of Av1753F12S to Av13026tLeuR-LR, Av10647COIIIF to Av16137tProR and Av13063tLeu-LR to Av3797R16S amplified overlapping fragments from which short range PCR was then carried out.

Long Range PCR Primers	Sequence
Av1753F12S	AAA CTG GGA TTA GAT ACC CCA CTA T
Av13026tLeuR-LR	CTT GGA KTT GCA CCA AGR TDV TTG GTT CCT AAG ACC A
Av5201tMetR	CCA TCA TTT TCG GGG TAT GG
Av10647COIIIF	TTT GAA GCA GCA GCC TGA TAY TG
Av16137tProR	ARA ATR CCA GCT TTG GGA GTT GG
Av13063tLeuF-LR	TGG TCT TAG GAR CCA TCT ATC TTG GTG CAA MTC CAA GT
Av3797R16S	CGA CCT GGA TTT CTC CGG TCT G

Table C-5 Long range primers used for the owlet nightjar

Long range primers used for the amplification of *Aegotheles cristatus* (Australian owlet nightjar). The combination of Av1753F12S to Av13026tLeuR-LR and Av10647COIIIF to Av2901R16S or Av16137tProR amplified overlapping fragments from which short range PCR was then carried out.

Long Range PCR Primers	Sequence
Av1753F12S	AAA CTG GGA TTA GAT ACC CCA CTA T
Av13026tLeuR-LR	CTT GGA KTT GCA CCA AGR TDV TTG GTT CCT AAG ACC A
Av10647COIIIF	TTT GAA GCA GCA GCC TGA TAY TG
Av16137tProR	ARA ATR CCA GCT TTG GGA GTT GG
Av2901R16S	GCA CTC TTT GTT GRT GGC TGC TT

C.3.2 Short Range

Short range PCR was carried out using different kits and enzymes for invertebrates (weta) and vertebrates (aves) and will be mentioned separately.

C.3.2.1 Invertebrate (weta) PCR

After trialling a range of enzymes for invertebrate short range PCR, Red Hot *Taq* (ABgene) was chosen as it consistently returned the best amplification results. All PCR reactions were carried out in a volume of 10µL.

Short Range PCR	
Reagent	**x1**
H_2O	5.825
10x Buffer	1
dNTPs (2mM)	1
Mg++	0.8
P1 (10µM)	0.4
P2 (10µM)	0.4
Taq	0.075

PCR reactions were carried out on either a Biometra T1 Thermocycler or an MJ Research PTC-200 thermal cycler and consisted of the following general format; initial denaturation of 94°C for 2 min, followed by 35 cycles of 94°C for 30sec, 48 - 50°C for 30sec and 68°C for 1 min 30sec with a final extension of 72°C for 1min 30sec.

Table C-6 Short range primers used for Anostostomatidae

Short range PCR primers used throughout this project are listed below. Numerous combinations of primers were used.

Name (gene)	DNA Sequence	Orientation (3' or 5')	Source/Reference
LCO1490	GGT CAA CAA ATC ATA AAG ATA TTG G	5' light/forward	Folmer et al. 1994
HCO2198	TAA ACT TCA GGG TGA CCA AAA AAT CA	3' heavy/reverse	Folmer et al. 1994
C1-J-1718	GGA GGA TTT GGA AAT TGA TTA GTT CC	5' light/forward	Simon et al. 1994
C1-N-2191	CCC GGT AAA ATT AAA ATA TAA ACT TC	3' heavy/reverse	Simon et al. 1994
C1-J-2195	TTG ATT TTT TGG TCA TCC AGA AGT	5' light/forward	Simon et al. 1994
L2-N-3014	TCC AAT GCA CTA ATC TGC CAT ATT A	3' heavy/reverse	Simon et al. 1994
TK-N-3785	GTT TAA GAG ACC AGT ACT TG	3' heavy/reverse	Simon et al. 1994
LR-J-13417	ATG TTT TTG TTA AAC AGG CG	5' light/forward	Simon et al. 1994
LR-N-13398	CGC CTG TTT AAC AAA AAC AT	3' heavy/reverse	Simon et al. 1994
SR-N-14588	AAA CTA GGA TTA GAT ACC CTA TTA T	3' heavy/reverse	Simon et al. 1994
inDNA 42 (18S)	TAA TGA TCC TTC CGC AGG TTC A	5' light/forward	Vawter, 1991
inDNA 43 (18S)	TCC CTG GTT GAT CCT GCC AGT A	3' heavy/reverse	Vawter, 1991
18S_1F	GAC GAA AAA TAA CGA TAC GGG	5' light/forward	This study
18S_IR	CTC AAT CTG TCA ATC CTT CCA	3' heavy/reverse	This study
28S rD1.2a	CCC SSG TAA TTT AAG CAT ATT A	5' light/forward	Whiting, 2002
28S rD3.2a	AGT ACG TGA AAC CGT TCA SGG GT	5' light/forward	Whiting, 2002
28S A	GAC CCG TCT TGA AGC ACG	5' light/forward	Whiting, 2002
28S Rd4.2a	CTA GCA TGT GYG CRA GTC ATT GG	5' light/forward	Whiting, 2002
28S B	TCG GAA GGA ACC AGC TAC	3' heavy/reverse	Whiting, 2002
28S rD5b	CCA CAG CGC CAG TTC TGC TTA C	3' heavy/reverse	Whiting, 2002
28S rD7b1	GAC TTC CCT TAC CTA CAT	3' heavy/reverse	Whiting, 2002

C.3.2.2 Vertebrate (aves) PCR

Numerous primer combinations were needed to amplify the genome for each bird. Listed here are the primers used for both amplification and sequencing. FastStart *Taq* DNA Polymerase (Roche) was used in 20µL reactions.

Short Range PCR	
Reagent	**x1**
H_2O	8.4
10x Buffer	2
dNTPs (2mM)	2.5
Betaine	4
P1 (10µM)	1
P2 (10µM)	1
Taq	0.1

PCR reactions were carried out on Biometra T1 Thermocycler and consisted of the following general format; initial denaturation of 94°C for 3 min, followed by 35 cycles of 94°C for 30sec, 48 - 55°C for 30sec and 72°C for 2 min (*ca* 1 min per 1Kb of sequence length) with a final extension of 72°C for 2 min.

Table C-7 Short range primers used for the dollar bird

Primers used for *Eurystomus orientalis* (dollar bird) are listed below. Various combinations were used to amplify fragments of *ca* 2 Kb at a time.

Primers	Sequence
Av1861R12S	TCG ATT ATA GAA CAG GCT CCT C
Av1753F12S	AAA CTG GGA TTA GAT ACC CCA CTA T
Av2901R16S	GCA CTC TTT GTT GRT GGC TGC TT
Av2672F16S	GTG GGA TGA CTT GTT AGT
Av4015Rtleu	GGA GAG GAT TTG AAC CTC TG
Av4903NDIR	GTT GGT CAT ATC GGA ATC GTG
Av4165FNDI	CGA AAA RTC CTA AGC TAC ATR CA
Av4747ND1F	CCA TTC GCC CTA TTC TTC CTA GC
Av5583RND2	CCT TGG AGG ACT TCT GGG AA
Av5199tMetF	GGT CAG CTA AAT AAG CTA TCG GG
Av5600FND2	TGA TTY CCA GAA GTV TTN CAA GG
Av7062CO1R	AGG TCT ACA GAG GCT CCA GC
Av7635COIR	GGC GGG TCT CAT TTG ATT GT
Av6838FCOI	CGT TAC CGC CCA TGC CTT CGT
tTyr6598F	GAA GAG GAA TTG AAC CTC TG
Av7856COIR	GCC TGT GAA TAG TGG GAA TC
Av6921COIF	ACT CGT ACC GCT CAT AAT CG
Av8377RCOII	ATA GGA GAT GAG GCG TCT TG
Av8013FCOI	GGC TCM TTM ATT CAA TAA CAG CCG
tSer8227FH-LR	AAG GAA TCG AAC CCC CTT TAR CTG CTT TCA AGC CA
Av9241ATP6R	TGG TCG AAG AAG CTT AGG TTC A
Av8872COIIF	ATC CCA GGA CGA CTA AAY CAA AC
Av10218COIIIR-LR	GGG GCT AGG CTT GAG TGG AAG AAG GCT C
Av9942FCOIII	ATG GCH CAC CAA GCA CAC TC
Av10884ND3R	GGG TCR AAG CCR CAT TCG TAG GG
Av10647COIIIF	TTT GAA GCA GCA GCC TGA TAY TG
Av12138ND4R	ATT GGA GCT TCT ACG TGG GCT T
Av11832ND4F	GTA ATA ATC CAA CCG TTC ATC
Av12788FND4	CTC AAA CAC ACG AGA ACA CC
Av13063tLeuF-LR	TGG TCT TAG GAR CCA TCT ATC TTG GTG CAA MTC CAA GT
Av14739ND5R	CTT CKG GKC CTA TTT TTT TGT ATC A
Av15266CytbR	TAT CCT ACG AAG GCA GTT GCT A
Av16065tThrR	GYG RTC TTC YGT CTT TGG TTT ACA AGA C
Av15107CytbF	CAT CCG TTG CCC ACA CAT GYC G
Av15671FCytb	CCC AGA AAA CTT CAC ACC AGC
AvFlam900ND6	GAA AAC ATA CCC CCA CCA TCA AC
Av522DloopR	TGG CCC TGA CTT AGG AAC CAG
Av807DloopR	CTA GKT GTG GGT CAA AGT GCA TCA GTG
FwdM13	CCC AGT CAC GAC GTT GTA AAA CG
RevM13	AGC GGA TAA CAA TTT CAC ACA GG

Table C-8 Short range primers used for the great potoo

Primers used for *Nyctibius grandis* (great potoo) are listed below. Various combinations were used to amplify fragments of *ca* 2Kb at a time.

Primers	Sequence
Av1753F12S	AAA CTG GGA TTA GAT ACC CCA CTA T
Av2901R16S	GCA CTC TTT GTT GRT GGC TGC TT
Av2035F12S	GAG ATG GAA GAA ATG GGC TAC
Av2703F16S	GAC TTG TTA GTA GAG GTG AAA AGC C
16S3348F	GGG ACT TGT ATG AAT GGC TA
Av5201tMetR	CCA TCA TTT TCG GGG TAT GG
Av5423ND2R	CGG AGG CGG TTG CTT GTA CT
Av5583RND2	CCT TGG AGG ACT TCT GGG AA
Av4747ND1F	CCA TTC GCC CTA TTC TTC CTA GC
tMet5197F	GGT CAG CTA ATT AAG CTA TCG
COI6642Rb	CGG GTA ATG AAG GTC ACA GG
Av6335tTrpF	GGC CTT CAA AGC CTT AAA TAA GAG
Av7662COIR	AGG AAG ATG AAG CCY AGA GCT CA
Av7318COIF	ACA TTC TTT GAY CCW GCR GGA GG
COII8323RtAsp	GGT TGG CCA CAT GAG ATG
Av8171tSerF-LR	CGA ACC TCC ACA CGC TGG TTT CAA GCC
Av8872COIIF	ATC CCA GGA CGA CTA AAY CAA AC
Av9043FtLys	GCT ATG GAA CAG CAC TAG C
Av9881ATP6R	GGC TTA GTA GGA GGA CGA AG
Av10218COIIIR-LR	GGG GCT AGG CTT GAG TGG AAG AAG GCT C
Av9874FATP6	GCA GTA GCC ATA ATC CAA GC
Av9942FCOIII	ATG GCH CAC CAA GCA CAC TC
Av10647COIIIF	TTT GAA GCA GCA GCC TGA TAY TG
ND311120Fb	CAA GGA GGC CTA GAA TGA GCA
HS12131RND4	CGG CTG TGA GTG CGT TCR TA
Av11492FND4	AAC YTN AAT CTH CTA CAA TGC TAA
ND411869FH	TTC GAA GCC ACC CTA ATC CC
t-Leu13052RH-LR	CTT GGA TTT GCA CCA AGR ARR TAG GTT CCT AAG ACC A
Av12217FND4	CTA GGM GGM TAT GGC ATT ATA CG
Av13734ND5R	AGG CCA AAT TGR GCT GAT TTT CC
Av13525ND5F	GMT GAG ARG GMG TAG GAA TCA TRT CA
Av13840FND5	AGC ACH ATA GTH GTA GCC GGA A
Av15266CytbR	TAT CCT ACG AAG GCA GTT GCT A
Av15425RCytb	GGA AGT GAA GGG CGA AGA ATC
Av15107CytbF	CAT CCG TTG CCC ACA CAT GYC G
FwdM13	CCC AGT CAC GAC GTT GTA AAA CG
RevM13	AGC GGA TAA CAA TTT CAC ACA GG

Table C-9 Short range primers used for the owlet nightjar

Primers used for *Aegotheles cristatus* (Australian owlet nightjar) are listed below. Various combinations were used to amplify fragments of *ca* 2Kb at a time.

Short Range PCR Primers	Sequence
Av438FDloopB	TCA CGT GAA ATS AGC AAC CC
Av510Fdloop	ATT CTT TCC CCC TAM ACC
Av1861R12S	TCG ATT ATA GAA CAG GCT CCT C
Av1753F12S	AAA CTG GGA TTA GAT ACC CCA CTA T
Av2901R16S	GCA CTC TTT GTT GRT GGC TGC TT
Av2703F16S	GAC TTG TTA GTA GAG GTG AAA AGC C
Av4015RtLeu	GGA GAG GAT TTG AAC CTC TG
1.4UP	AAT CCA GGT CGG TTT CTA TCT
Av4903ND1R	GTT GGT CAT ATC GGA ATC GTG
Av5201tMetR	CCA TCA TTT TCG GGG TAT GG
Av4921ND1F	CCC ACG ATT TCG MTA YGA CCA
Av5199tMetF	GGT CAG CTA AAT AAG CTA TCG GG
Av7062COIR	AGG TCT ACA GAG GCT CCA GC
Av6838FCOI	CGT TAC CGC CCA TGC CTT CGT
Av8377RCOII	ATA GGA GAT GAG GCG TCT TG
Av8171tSerF-LR	CGA ACC TCC ACA CGC TGG TTT CAA GCC
Av9881ATP6	GGC TTA GTA GGA GGA CGA AG
Av9043FtLys	GCT ATG GAA CAG CAC TAG C
Av10884ND3R	GGG TCR AAG CCR CAT TCG TAG GG
Av10387COIIF	CCA TAG CAT CAC AGA AGG AAA C
Av10647COIIIF	TTT GAA GCA GCA GCC TGA TAY TG
ND4L11410Fb	TCT CTG CAT GCG AAG CAG G
ND311120Fb	CAA GGA GGC CTA GAA TGA GCA
Av12912tHisR	CGG CAG GTA AGA AGA GTC TAA C
Av11832ND4F	GTA ATA ATC CAA CCG TTC ATC
Av13026tLeuR-LR	CTT GGA KTT GCA CCA AGR TDV TTG GTT CCT AAG ACC A
Av12788FND4	CTC AAA CAC ACG AGA ACA CC
Av13734ND5R	AGG CCA AAT TGR GCT GAT TTT CC
Av13563ND5F	GAT GAC ACG GAC GAG CAG AAG
Av15266CtybR	TAT CCT ACG AAG GCA GTT GCT A
Av15307CtybR	CAG TGG CTC CTC AGA ATG ATA T
Av15107CtybF	CAT CCG TTG CCC ACA CAT GYC G
Av16728tGluR	GGY TTT TCA GGC CGT AGR TCT TGG
Av16531FND6	ACC ACC ARC ATH CCC CCY AAA TA
FwdM13	CCC AGT CAC GAC GTT GTA AAA CG
RevM13	AGC GGA TAA CAA TTT CAC ACA GG

C4 CLONING PROTOCOL

The following cloning protocol was used for the amplification of the mitochondrial genome of the two selected weta species. A similar cloning procedure was carried out for difficult regions such as the control region (D-loop) in the bird species.

C.4.1 Restriction Enzyme Digestion and Cloning

Clean PCR product was digested using either EcoRI (Roche) (GAATTC) or HindIII (Roche) (AAGCTT) restriction enzymes. Fragments were used to create a map of the genome which would later be used to assemble the cloned fragments in the right order.

Each restriction digest consisted of 1× Buffer H (5mM Tris-HCl, 10mM NaCl, 1mM $MgCl_2$, 100µM Dithioerythritol [DTT], pH 7.5), OR Buffer B (10mM Tris-HCl, 10mM NaCl, 5mM $MgCl_2$, 1mM 2-mercaptoethanol, pH 8.0) and 10 U EcoR I or Hind III (Roche) and 5µL plasmid DNA. Restriction digests were incubated at 37ºC for 2 hours.

A 5µL aliquot of each restriction digest and a sample of undigested plasmid DNA were combined with 2µL of 10× loading dye (see B.4 SOLUTIONS AND DYES). The undigested samples were included as a negative restriction digest control. Each aliquot was then loaded onto a 1% (w:v) agarose/1× TAE buffer gel. The standard size maker, 1Kb Plus DNA ladder (see B 1.4 SOLUTIONS AND DYES) and a standard mass ladder (Low DNA Mass Ladder, Invitrogen) were loaded. Samples were electrophoresed at 100 volts in 1× TAE buffer. Following electrophoresis the gels were visualised and photographed using the Gel Doc system (BIORAD).

Digests appeared on the agarose gel as a number of discrete bands ranging in size from 450 bp to 6.3 Kb depending on the restriction enzyme used. The sum of all resulting fragments should have summed to the total size of the original long range PCR product (*ca*16 Kb) but often it was difficult to distinguish the multiple lower fragments. The concentration of the insert DNA was estimated by comparison with the Low DNA Mass Ladder (see B.4 SOLUTIONS AND DYES)

C.4.2 Precipitation of digestion and vector

Samples were mixed with 3M NaOAc (1/10th of the volume of sample) and 100% (v:v) ethanol (2 ½ times the volume of sample) and mixed gently then placed in the freezer overnight (*ca* 16 hours). Centrifugation at 14,000 rpm for 30 min in a cold (4ºC) centrifuge to pellet the DNA was carried out after the incubation. The supernatant was removed (removing excess salt) and a further 200μL of 70% (v:v) ethanol was added to wash to DNA pellet. Samples were spun for a further 10min at 14,000 rpm in a cold (4ºC) centrifuge then the ethanol was removed and the pellet allowed too dry before resuspending in 10μL of milli-Q water. Products were run on a 1% (w:v) agarose gel with a 1Kb Plus Ladder and Low DNA Mass Ladder to check size and DNA concentration.

C.4.3 Shotgun cloning of digested long range fragments

Vector (pUC 118, 300ng) was digested with either EcoRI or HindIII (ie. the same as the sample) and was dephosphorylated to stop the vector from ligating to itself. Shrimp alkaline phosphatase (USB Corporation) was added to cut precipitated vector in 1× reaction buffer (50mM Tris-HCl, pH 9.0 and 10mM $MgCl_2$), in a total volume of 20μL. The mix was incubated at 37°C for 15min followed by an incubation at 65°C for a further 15min to deactivate the enzyme. An aliquot (50ng: 3 - 4μL) of cut dephosphorylated vector was used for ligation.

Ligation reactions consisted of 1× Rapid Ligation Buffer (30mM Tris-HCl [pH 7.8], 10mM MgCl2, 10mM dithiothreitol [DDT], 1mM adenosine triphosphate [ATP], 5% (w:v) polyethylene glycol [PEG] (Promega), 50ng pUC 118 vector, 3 Weiss U T4 DNA ligase (Promega) and 5μL of digested fragment (in a total volume of 10μL. A positive control ligation reaction was prepared, consisting of the above reagents but including 2μL of Control Insert DNA (542 bp) (Promega) in place of the PCR product. A negative control ligation reaction was prepared consisting of the above reagents but no insert DNA. Ligation reactions were incubated at 4ºC overnight (*ca* 16 hours).

C.4.4 Transformation of Competent Cells

Ligation reactions were transformed into One Shot TOP 10 Chemically Competent *E.coli* (Invitogen). One vial per ligation was thawed on wet ice (this kit provides single use 50µL aliquots of competent cells). Ligation reactions were added to the competent cells and stirred gently using a pipette tip and incubated on ice for 30min. Cells were then heat-shocked for 45sec in a 42°C water bath then placed on ice for 2min. To this 250µL of S.O.C Medium (2%[(w:v) tryptone, 0.5% (w:v) yeast extract, 20mM glucose, 10mM NaCl, 10mM $MgCl_2$, 10mM MgSO4, 2.5mM KCl - Invitrogen) was added and the solution shaken for one hour at 37°C at 225 rpm. A positive transformation control reaction was prepared with 1.5µL (1.5pg) pUC118 plasmid provided in the kit.

After incubation with S.O.C medium, a sterilised cell spreader was used to plate 150µL of each transformation (20µL for the pUC 118 positive transformation control) onto Luria-Bertani (LB) agar plates (85mM NaCl, 1% (w:v)] tryptone [Merck], 0.5% (w:v) yeast extract [Merck], 1.5% (w:v) bacteriological agar [Oxoid) containing 100µg mL^{-1} ampicillin (Sigma), 400µg mL^{-1} isoprophyl-β-D-galactoside (IPTG, Progen) and 40µg mL^{-1} 5−Bromo-4-chloro-3-indoxyl-beta-D-galactopyranoside (X-gal, Progen). The stock solutions of ampicillin, IPTG and X-Gal were 100mg mL^{-1} in milli-Q water, 200mg mL^{-1} in Milli-Q water and 20mg mL^{-1} in N,N'-dimethyl-formamide (DMF) respectively. All stock solutions of ampicillin, IPTG and X-Gal were stored at -20ºC, with the X-Gal in a light proof container. Plates were allowed to dry for 10min at room temperature under a sterilising flame before being inverted and incubated at 37ºC overnight. Plates were placed in the fridge for a day to develop the blue colour of the colonies without an insert. It is the insert-containing white colonies which are picked.

C.4.5 Selection of Transformed Cells

Plates typically had in excess of 50 - 100 colonies, *ca* 20 - 30% of which were white and were putatively recombinant. For screening purposes, 40 colonies were picked for colony PCR and LB broth culture. Colony PCR is performed on white colonies to verify they contain insert DNA and that the insert is of the expected size(s).

Each colony PCR reaction consisted of 1× PCR buffer, 250μM of each dNTP, 0.5μM M13F (CCC AGT CAC GAC GTT GTA AAA CG), 0.5μM M13R (AGC GGA TAA CAA TTT CAC ACA GG) primer and 1U Taq DNA polymerase in a total volume of 20μL. Template DNA was added by lightly touching the white colony with a sterile pipette tip and then briefly dipping the pipette tip into the PCR reaction, ensuring there was no carry-over of LB agar which could inhibit the PCR reaction. The pipette tip was then dipped into wells in a plate containing 100μL of LB broth [85mM NaCl, 1% (w:v) tryptone [brand], 0.5% (w:v) yeast extract [brand] as prepared above) inoculated with ampicillin (100μg mL^{-1}). The plates were then incubated at 37°C *ca* 16h. Colony PCR was carried out in a thermal cycler with the following programme; hold 94°C for 2min, 35 cycles of 94°C for 30sec, 50°C for 30sec, 72°C for 45sec, with a final extension of 72°C for 5min.

A 5 μL aliquot of each colony PCR was combined 10× Loading dye and loaded onto a 1% (w:v) agarose gel stained with SYBRSafe (Invitrogen) along with a standard size marker. Following electrophoresis, gels were visualised and photographed. Successful amplifications appeared as discrete bands with a size corresponding to the size of the insert plus 266 bp. Sequencing of successful clones was carried out as described below for at least 3 clones of each PCR product size. Inserts were sequenced using the M13 forward and reverse primers.

C.5 SEQUENCING PROTOCOL

C.5.1 Purification

After amplification, PCR products were purified with Shrimp Alkaline Phosphatase (SAP) and Exonuclease I (EXO)(USB Corportation): 1μL (1U) of SAP and 0.5μL (5U) of EXO were added directly to the PCR reaction and the mixture was incubated at 37°C for 30min followed by 80°C for 15min, to deactivate the enzymes. The PCR sample was then ready to be used in a sequencing reaction.

C.5.2 Quantification

The amount of DNA template required for a sequencing reaction was calculated using the following formula:

$$\text{ng required} = \frac{\text{size of insert (bp)}}{40}$$

C.5.3 Sequencing

An appropriate amount of PCR product was added to a sequencing reaction containing: 3.5µL sequencing buffer (Applied Biosystems; Foster City, CA)., 1µL BigDye Terminator 3.1 (Applied Biosystems; Foster City, CA), 10pmol of primer (M13fwd or M13rev) and water to 20µL with the following programme; 27 cycles of: 94°C for 10sec, 50°C for 5sec, 60°C for 4min; followed by a hold cycle at 10°C.

Following sequencing, the reactions were cleaned using magnetic beads, to remove unincorporated dyes and nucleotides, in preparation for running on the capillary sequencer.

1. 10µL of CleanSeq (Agencourt) and 62µL of 85% (v:v) EtOH were added to each reaction and pipetted up and down 7 times to mix.
2. Reactions were incubated at room temperature and placed on a magnetic plate for 3 mins. The magnet captured the beads as a ring around the inside of the tube.
3. The supernatant was removed and the beads were washed with 100µL of 85% (v:v) EtOH.
4. After 1min the EtOH was removed and the beads were dried.
5. The reactions were removed from the magnetic plate and 40µL of 0.1mM EDTA (pH8.0) was added.
6. After 3min the reactions were placed back on to the magnetic plate. Following bead capture, the samples were removed and transferred into a clean tube in a volume of 20µL.
7. The samples were taken to the AWCGS for running on the AB3730 capillary machine (ABI3730XL Genetic Analyzer ;Applied Biosystems; Foster City, CA).

APPENDIX D

Thesis publication reprints

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B

Phil. Trans. R. Soc. B (2008) **363**, 3427–3437
doi:10.1098/rstb.2008.0112
Published online 4 September 2008

Diversification of New Zealand weta (Orthoptera: Ensifera: Anostostomatidae) and their relationships in Australasia

Renae C. Pratt[1,*], Mary Morgan-Richards[1] and Steve A. Trewick[2]

[1]*Allan Wilson Centre for Molecular Ecology and Evolution, Institute of Molecular Biosciences, Massey University, Private Bag 11-222, Palmerston North, New Zealand*
[2]*Institute of Natural Resources and Management, Massey University, Private Bag 11-222, Palmerston North 4442, New Zealand*

New Zealand taxa from the Orthopteran family Anostostomatidae have been shown to consist of three broad groups, *Hemiandrus* (ground weta), *Anisoura*/*Motuweta* (tusked weta) and *Hemideina–Deinacrida* (tree–giant weta). The family is also present in Australia and New Caledonia, the nearest large land masses to New Zealand. All genera are endemic to their respective countries except *Hemiandrus* that occurs in New Zealand and Australia. We used nuclear and mitochondrial DNA sequence data to study within genera and among species-level genetic diversity within New Zealand and to examine phylogenetic relationships of taxa in Australasia. We found the Anostostomatidae to be monophyletic within Ensifera, and justifiably distinguished from the Stenopelmatidae among which they were formerly placed. However, the New Zealand Anostostomatidae are not monophyletic with respect to Australian and New Caledonian species in our analyses. Two of the New Zealand groups have closer allies in Australia and one in New Caledonia. We carried out maximum-likelihood and Bayesian analyses to reveal several well supported subgroupings. Our analysis included the most extensive sampling to date of *Hemiandrus* species and indicate that Australian and New Zealand *Hemiandrus* are not monophyletic. We used molecular dating approaches to test the plausibility of alternative biogeographic hypotheses for the origin of the New Zealand anostostomatid fauna and found support for divergence of the main clades at, or shortly after, Gondwanan break-up, and dispersal across the Tasman much more recently.

Keywords: Anostostomatidae; biogeography; COI; 18S; Zealandia; New Zealand

1. INTRODUCTION

The biology of New Zealand is, unlike that of most Pacific islands, viewed as continental in nature (Cowie & Holland 2006). This is justified geologically because New Zealand is formed from continental rather than oceanic crust (Neall & Trewick 2008). Consequently, the biota of New Zealand is considered to be predominantly 'Gondwanan', having its principal affinities in Australasia (Fleming 1979) and the Southern Hemisphere in general (Gibbs 2006). Although the Gondwanan nature of the New Zealand biota is often attributed to the continental (vicariant) history of the land, this explanation has not always been pre-eminent (e.g. Fleming 1962; Caughley 1964). The importance of dispersal is widely recognized, and it is more generally accepted that biogeographical pattern alone does not reveal the process(es) of origination of biota (Waters & Craw 2006). Indeed, the distribution of one former icon of vicariance biogeography in New Zealand (and the Southern Hemisphere), southern beech (*Nothofagus*) has recently been shown to be best explained by dispersal to New Zealand (Cook & Crisp 2005*a*; Knapp *et al.* 2005). Numerous other molecular studies of a range of taxa demonstrate that a substantial (if not predominant) part of the New Zealand biota are products of long-distance dispersal (Waters & Craw 2006 and references therein; Trewick *et al.* 2007).

An improving, though far from perfect understanding of the tectonic history in the New Zealand region (Mortimer 2004) is helping to reveal why this continental land does not in fact have a predominantly continental biota. The rifting of the continent Zealandia from Gondwana (including Australia), commenced *ca* 83 Ma (figure 1*a*) and the Tasman Sea reached its current width between 63.5 and 55.5 Ma (Veevers & Li 1991; McLoughlin 2001). This was just the start of New Zealand and New Caledonia's story. Zealandia (figure 1*b*) subsequently sank so that today approximately 93 per cent of the continent is below the surface of the sea (Landis *et al.* 2008).

This active geological history contrasts with the prolonged geological stability of Australia, the nearest remnant of continental Gondwana. With respect to land area, climate and biotic assemblage, New Zealand (and New Caledonia) has few continental attributes and ample evidence suggests that these islands are

* Author for correspondence (r.c.pratt@massey.ac.nz).

Electronic supplementary material is available at http://dx.doi.org/10.1098/rstb.2008.0112 or via http://journals.royalsociety.org.

One contribution of 15 to a Theme Issue 'Evolution on Pacific islands: Darwin's legacy'.

This journal is © 2008 The Royal Society

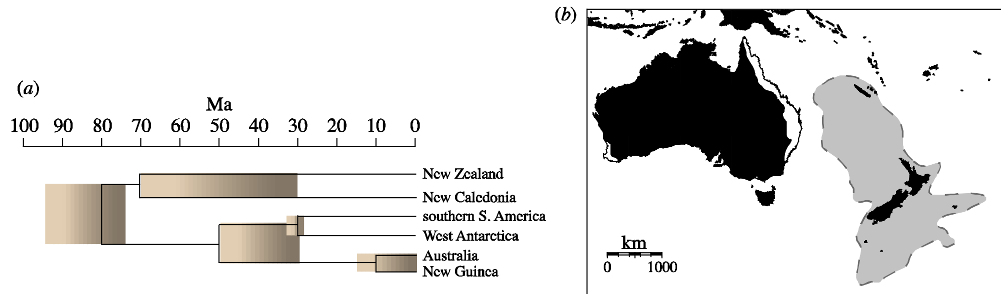


Figure 1. (*a*) Geological area cladogram, after Cook & Crisp (2005*b*) showing the accepted Gondwanan break-up sequence. Shaded boxes represent uncertainty about timing of vicariant events. (*b*) Australasia sampled for anostostomatid weta (Insecta: Orthoptera) in this study (Australia, New Caledonia and New Zealand). Australian species are limited to the east coast of Queensland and New South Wales and one genus in Western Australia (white fill) whereas most parts of New Zealand and New Caledonian support one or more weta species. The dashed outline and light grey shading indicate the approximate boundaries of the submerged continent Zealandia.

biologically more like oceanic islands than southern continents (Goldberg *et al.* 2008; Grandcolas *et al.* 2008; Trewick & Morgan-Richards in press).

Evidence for the persistence of land in the New Zealand region throughout the Oligocene has been obscured by the extensive tectonic activity initiated in the early Miocene (Landis *et al.* 2008). The tectonic upheaval that resulted in the formation of New Zealand (as we know it today) began *ca* 24 Ma and still continues (Trewick *et al.* 2007). For example, the major mountain ranges of New Zealand started forming only *ca* 5 Ma. This, and other local geophysical events, may have been more important in the development of the modern biota than ancient vicariant processes. New Caledonia has a similar geological history with tectonic activity forcing a submerged section of Zealandia (and obducted oceanic ultramafic strata) to the sea surface in the late Eocene (*ca* 40 Ma Chardon & Chevillotte 2006; Mortimer *et al.* 2006; Grandcolas *et al.* 2008; Neall & Trewick 2008).

One of the most interesting components of New Zealand's terrestrial fauna, with both taxonomic and ecological diversity, are insects of the orthopteran family Anostostomatidae, known in New Zealand by their Maori name, weta. Of particular biogeographic interest is the presence of the family on all three major Australasian landmasses: Australia, New Caledonia and New Zealand. The group consists of relatively large insects (20–80 mm) that are nocturnal, predominantly flightless and predatory, with a Gondwanan distribution (also found in Central and South America, South Africa, Madagascar and India). In New Zealand, the family is represented by five genera and approximately 56 species. These five genera fall into three distinct groups: (i) nine (plus approx. 30 undescribed) species of *Hemiandrus* Ander 1938 (ground weta), (ii) one species of *Anisoura* Ander 1938 and two species of *Motuweta* Johns 1997 (tusked weta), and (iii) seven *Hemideina* White 1846 (tree weta) and 11 *Deinacrida* White 1842 (giant weta) (Trewick & Morgan-Richards 2004, 2005).

The *Hemideina* and *Deinacrida* are unusual among Anostostomatidae in that all species are primarily herbivorous. The diversification of *Hemideina–Deinacrida* dates to the Miocene, with adaptation to diverse habitats following mountain uplift (*ca* 5 Ma Trewick & Morgan-Richards 2005). The three tusked weta species (*Anisoura*/*Motuweta*), so named owing to the impressive tusk-like structures on the mandibles of mature males, form a monophyletic group among New Zealand taxa (Trewick & Morgan-Richards 2004), although analogous ornamentations are found in some South African species (i.e. *Libanasidus vittatus;* Field & Deans 2001). Within the Australasian anostostomatid genera, *Hemiandrus* is the only genus not endemic to a single landmass, being recorded in both Australia and New Zealand (Johns 1997). Of the approximately 40 species from New Zealand (P. M. Johns 2005, personal communication) only nine are described (Johns 1997; Jewell 2007), making them the least well-characterized weta group in this country. Ovipositor length, which appears to be correlated with degree of maternal care (Gwynne 1995; Johns 1997; Gwynne 2004), was in the past the key morphological character distinguishing the genus *Zealandosandrus* Salmon 1950 from *Hemiandrus* Ander 1838. However, Johns (1997) synonymized *Zealandosandrus*, retaining *Hemiandrus* by precedent. The Australian anostostomatid fauna is poorly characterized with just 13 described species but probably comprises nine genera with approximately 60 species (Johns 1997; G. Monteith 1999, personal communication). Australia's fauna includes three genera with winged species, one of which (*Transaevum*) is considered to be the most 'ancestral' extant member of the group (Monteith & Field 2001). Intriguingly, fossil Orthoptera that are putatively 'weta' are reported from 190 Ma deposits in Queensland, Australia (Meads 1990) but have not been formally described. The New Caledonian weta fauna consists of two genera, *Aistus* (three species with several undescribed; Johns 1997) and *Carcinopsis* (six species; Johns 1997).

Anostostomatids occupy a variety of environments across their geographical range in Australasia. In New Caledonia, they are tropical forest inhabitants; in Australia, almost all are found in the wet tropical forests of Queensland with one genus endemic to coastal Western Australia (Monteith & Field 2001; figure 1*b*). The New Zealand weta live mostly in temperate forest and subalpine environments, with the majority

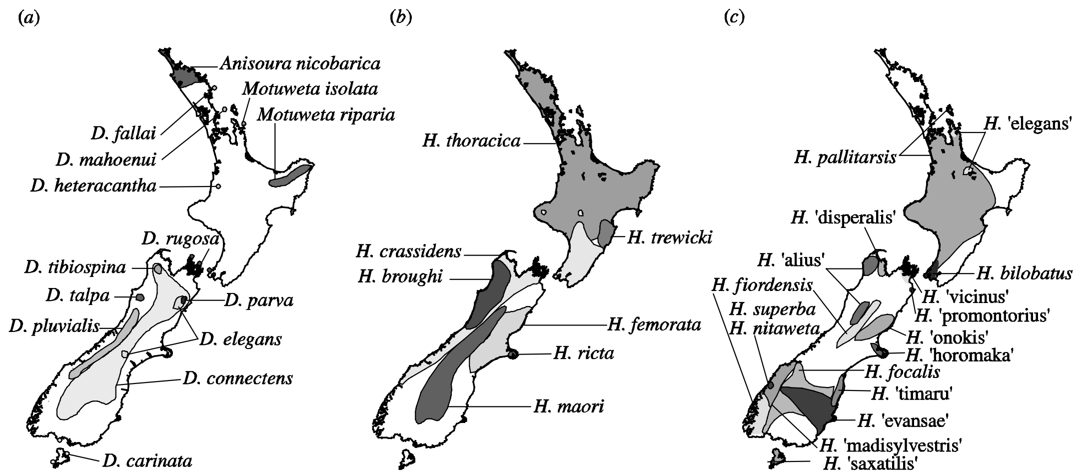


Figure 2. Approximate distributions of New Zealand Anostostomatidae taxa: (*a*) *Deinacrida* (giant weta) and *Anisoura*/*Motuweta* (tusked weta) species; (*b*) *Hemideina* (tree weta) (based on Trewick & Morgan-Richards in press); (*c*) *Hemiandrus* (ground weta) based on Johns (2001), Jewell (2007) and P. M. Johns (2005, personal communication). In addition to the species shown, *H. maculifrons* is widely distributed throughout both the North and South Islands. Additional undescribed taxa appear to have local distributions and predominate in South Island (P. M. Johns 2005, personal communication).

of species in South Island (figure 2). Curiously, anostostomatids are absent from comparable temperate forest in southern Australia (Victoria and Tasmania).

We have undertaken sampling across New Zealand, New Caledonian and Australia to explore the evolution of the New Zealand Anostostomatidae. Representatives of all nine Australian genera, both New Caledonian genera and the five New Zealand genera were included in the present study. We used molecular phylogenetics to recover support for the relationships of the three New Zealand anostostomatid groups. We include representatives of New Zealand *Hemiandrus* species diversity, which have been absent from previous work, in order to explore diversity within this group and test support for monophyly of the genus. We apply relaxed molecular clock methods, to estimate the likely age and origin of New Zealand weta lineages.

2. MATERIAL AND METHODS

(a) *Sampling*

The majority of sampling was undertaken by the authors in New Zealand and New Caledonia. In addition, samples of *Hemiandrus* were supplied by Darryl Gwynne (University of Toronto, Canada). The New Zealand sampling included all three tusked weta species (*Anisoura*/*Motuweta*), representatives of the *Hemideina* and *Deinacrida* (previously shown to be monophyletic; Trewick & Morgan-Richards 2005) and representatives within the taxonomic diversity of *Hemiandrus* including putative and new species (see the electronic supplementary material). Assistance with identification and sampling of undescribed species was provided by P. M. Johns. Material from Australia was supplied by Geoff Monteith (Brisbane Museum, Australia) and Dave Rentz (CSIRO, Australia). This sampling, although not exhaustive, includes at least one representative of each genus in the region and is the most complete dataset to date.

(b) *Molecular methods*

Whole genomic DNA was extracted from hind leg muscle following the salting-out method (Sunnucks & Hale 1996) and resuspended in 50 μl TE buffer (0.1 mM EDTA, 10 mM Tris) or water. Polymerase chain reactions (PCRs) were performed in 10 μl volume using ABgene Red Hot *Taq*. Products were visualized on 1 per cent agarose gels stained with SYBRSafe (Invitrogen). Thermal cycling PCR was carried out on an MJ Research PTC-200 thermal cycler and consisted of initial denaturation of 94°C for 2 min, followed by 35 cycles of 94°C for 30 s, 48–50°C for 30 s and 72°C for 1 min 30 s with a final extension of 72°C for 3 min. PCR products were purified with Shrimp Alkaline Phosphatase and Exonuclease I following manufacturer's recommendations (USB Corporation). Sequencing used BigDye Terminator v. 3.1 chemistry and an ABI3730XL Genetic Analyzer (Applied Biosystems; Foster City, CA). DNA sequences were deposited on NCBI GenBank (see the electronic supplementary material; EU676657–EU676800 and EU713453–EU13461). Primers used for PCR were the following: mtDNA COI: LCO1490, HCO2198 (Folmer *et al.* 1994), CI-J-2195 and L2-N-3014 (Simon *et al.* 1994); mtDNA third domain 12S rRNA: SR-N-14588 and LR-J-13417 (Simon *et al.* 1994); nuclear rRNA 18S: 18S-S22, 18S-A1984 (Vawter 1991), 18S_1F (gac gaa aaa taa cga tac ggg) and 18S_1R (ctc aat ctg tca atc ctt cca) (this study); and 28S: 28SrD1.2a, 28SrD3.2a, 28SrD4.8a, 28SA, 28SrD4.2b, 28SrD5b, 28SrD7bl and 28SB (Whiting 2002*a*).

(c) *Phylogenetic analysis*

Individual sequence reads were checked against ABI trace files using SEQUENCHER v. 4.70 (Gene Codes Corp. Ann Arbour, MI) and aligned using SE-AL v2.0a11 (Rambaut 1996). The protein-coding gene COI was translated into amino acids to ensure correct reading frame and to detect evidence of nuclear copies (which were subsequently removed). Ribosomal RNA genes were checked for indels. In cases where missing data were included, they were coded as N in analyses. In order to evaluate individual genes and concatenated data, we divided the datasets into the following: I—18S Ensifera; II—18S

Australasian Anostostomatidae; III—combined 18S and 28S Australasian Anostostomatidae; IV—COI–RY-coded Australasian Anostostomatidae; V—combined COI–RY and 12S Australasian Anostostomatidae; and VI—COI *Hemiandrus* only.

The COI data were partitioned into three character sets according to the codon position, first, second and third. In order to maximize third codon information, we treated it in three different ways: as four nucleotides (A, G, T, C), Y-coded (Y, A, G) or RY-coded (A and G=R, T and C=Y). In order to avoid potential tree estimation bias due to nucleotide composition or saturation, we used Y or RY coding on the third codon position nucleotides for COI sequences in dataset IV and V. Recoding of this sort has been shown to greatly improve consistency in phylogenetic resolution by reducing bias from differences in nucleotide composition (Phillips & Penny 2003), which is useful when looking at deeper divergences. To assist with tree rooting and thus confirm ingroup status of our sample, we used published Ensifera DNA sequences from both EMBL and NCBI GenBank (see the electronic supplementary material).

Models of DNA evolution were optimized separately for each dataset using ModelTest v. 3.7 (Posada & Crandall 1998) and Akaike Information Criterion was preferred to the hierarchical likelihood ratio test (Posada & Buckley 2004). Maximum-likelihood (ML) analyses were implemented using the programs PAUP* (Swofford 2003), GARLI v. 0.951 (Zwickl 2006) and PhyML (Guindon & Gascuel 2003). Model parameters from ModelTest were implemented using a general time-reversible model with invariable sites and a gamma distribution for variable rate sites (GTR+I+G) model with a heuristic search under the likelihood criterion with trees obtained from stepwise addition.

Bayesian analyses were implemented using MrBayes v. 3.1 (Huelsenbeck & Ronquist 2001). We specified nst=2 (HKY) and nst=6 (GTR) with a proportion of invariant sites and gamma distribution of rate variation. Analyses of datasets III (18S+28S), IV (COI) and V (COI +12S) were undertaken with (parameters unlinked) and without character set partitions. We used two runs of four Markov chains (each with one cold chain) with $1–10\times10^6$ generations and default priors, sampling every thousandth tree. A 'burn-in' of 10 per cent was removed after examination of log-likelihood scores and average standard deviation of the split frequencies. Trees saved below the burn-in generation were discarded and a majority rule consensus of the remaining trees was calculated. Multiple replicates of the Bayesian runs were carried out to insure convergence of the posteriors.

(d) *Tree comparisons*

We assessed the degree of conflict between our phylogenetic estimates by using tree comparison tests, to see if one topology was significantly better at explaining the molecular data than alternative phylogenies. We used the SH tests (Shimodaira & Hasegawa 1999) implementing a RELL distribution derived from 1000 bootstrap replicates as executed in PAUP*. For dataset IV (COI), we carried out multiple analyses manipulating the third codon position so that it was; four states, Y-coded and RY coded. To observe the effect of this simple noise reduction technique, we compared ML topologies obtained from PhyML for each state using either a simple model (HKY85) or a parameter-rich model (GTR +I+G). We also used constraint analysis to test the likelihood of alternative tree topologies for the monophyly of New Zealand taxa and the genus *Hemiandrus* (New Zealand and Australia).

(e) *Divergence time estimation*

We compared the likelihood scores obtained from ML analysis both with and without the implementation of a molecular clock in PAUP* for dataset II (18S Australasia) and dataset IV (COI–RY-coded Australasia). This was carried out both with and without the inclusion of taxa we suspected of having a rate shift due to long branches observed in initial analyses. SH tests were applied to resulting trees to determine whether there was rate heterogeneity and therefore if the data were acting in a clock-like manner to determine whether to use a strict or relaxed molecular clock in BEAST v. 1.4.6 (Drummond & Rambaut 2007).

As there are no suitable fossils for molecular dating, we used geological events as points of reference to test the plausibility of vicariant versus dispersal explanations for the New Zealand weta diversification (figure 1*a*). In order to explicitly examine the alternative hypotheses for patterns of diversity, we calibrated trees using initial separation of Zealandia from Gondwana (less than 82 Ma as applied by avian evolutionists, see Ericson *et al.* 2002; Baker *et al.* 2005) and emergence of New Caledonian (less than 40 Ma). The two dating constraints were separately applied to the nuclear dataset II (18S Australasia) and mitochondrial dataset IV (COI–RY-coded Australasia). We removed a clade of five taxa (clade A plus New Caledonian taxa), shown by initial analyses to have long branches and a substantially elevated rate of molecular evolution (indicated by BEAST rates). First, if Zealandia and Australia parted *ca* 82 Ma, we assumed vicariance and constrained the most basal split of Anostostomatidae to more than 82 Ma (BEAST parameters; relaxed lognormal clock, lognormal distribution, mean=4.74; s.d.=0.2, run for 100 million repetitions sampling trees every 10 000). Second, we constrained the same point on the tree as above, but assuming Zealandia submerged completely and land resurfaced *ca* 40 Ma (BEAST parameters; normal distribution: mean=36, s.d.=0.2, initial value=35). We used a root value for the trees (BEAST parameters; uniform distribution: lower limit=85, upper limit=250, initial=100) as unpublished fossils have been dated from 190–200 Ma from Queensland that have been attributed to the Anostostomatidae and fossil Ensifera are dated back to *ca* 250 Ma (fig. 7.18, p. 202, Grimaldi & Engel 2005). Calibration points were implemented in BEAST using a relaxed uncorrelated lognormal molecular clock (Drummond *et al.* 2006). Resulting trees were analysed using software distributed with BEAST–TreeAnnotator v. 1.4.6 where the burn-in (1000 trees) was removed and a maximum credibility tree constructed. Trees were then viewed in FigTree v. 1.1.2. Details of the XML files are available on request from the authors.

3. RESULTS

A summary of all sequence data collected can be viewed in the electronic supplementary material along with locality information and accession numbers. NCBI BLAST searches returned matches for previously published Orthopteran sequences. The alignment used in each analysis is available as a Nexus file from the authors on request.

(a) *Phylogeny*

Dataset I (18S Ensifera) consisted of 1863 bp after removal of a 37 bp hypervariable indel region between bp 719 and 756 from our original alignment to accommodate the diverse range of taxa, which included seven ingroup taxa and nine outgroup

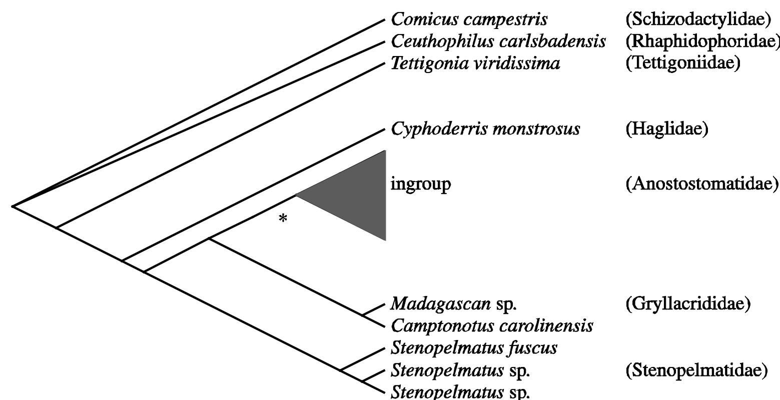


Figure 3. Bayesian cladogram for 1863 bp of nuclear 18S Ensifera data including seven ingroup taxa (EU676721, EU713453, EU713454, EU713455, EU713458, EU713459 and Z97570) and published taxa (Flook *et al.* 1999; Terry & Whiting 2005; Jost & Shaw 2006) used to determine appropriate outgroup (see the electronic supplementary material). The hypervariable indel region between bp 719 and 756 was removed due to alignment uncertainty. The asterisk node indicates Bayesian posterior probability of 0.96.

Ensifera (Flook *et al.* 1999; Terry & Whiting 2005; Jost & Shaw 2006). We confirm the monophyly of Anostostomatidae in our sample and found the Gryllacrididae to be sister to Anostostomatidae with Stenopelmatidae sister to the Anostostomatidae–Gryllacrididae clade. Both of these families have previously been suggested as close relatives to Anostostomatidae (figure 3).

Dataset II-18S Australasian Anostostomatidae: After establishing support for monophyly of the Anostostomatidae, we turned our focus to the relationships within the family. We included more representatives from the Australasian region and a slightly shorter fragment of 18S (29 taxa, 1746 bp), again excluding the problematic indel region. Bayesian and ML analyses yielded similar topologies (figure 4). We observed that the New Zealand tusked weta (*Anisoura*/*Motuweta*; clade A) and New Caledonian taxa (*Aistus* and *Carcinopsis*) formed long branches in the phylogeny. Long branches like these can result in misleading results even without rate differences (Hendy & Penny 1989) that affect all further tree selection criteria. We explored the effect of these long branches by subjecting the dataset to identical analyses with the inclusion or exclusion of either or both the New Caledonian taxa (*Aistus* and *Carcinopsis*), or the tusked weta (*Anisoura*/*Motuweta*) sequences. The exclusion of *Anisoura*/*Motuweta* resulted in *Aistus* and *Carcinopsis* together being placed as sister to the rest of the Anostostomatidae, from which we infer long branch attraction, resulting from lineage-specific rate increases. When *Aistus* and *Carcinopsis* were removed, the *Anisoura*/*Motuweta* (clade A; figure 4) were sister to the *Hemideina–Deinacrida* (clade B; figure 4). In no instances were *Hemiandrus* (New Zealand and Australian) found to be monophyletic, a finding consistent in the following analyses with nuclear and mitochondrial sequences. We identified two clades within the New Zealand *Hemiandrus* (CI and CII; figure 4) although not monophyletic in every analysis, both were separate from the Australian *Hemiandrus* species. Dataset III (concatenated 18S+28S) Australasian anostostomatids (35 taxa, 1549 bp) drew on data published by Jost & Shaw (2006) and to allow for direct comparison with that study, we cut our 18S sequences to 1204 bp and included 344 bp of 28S data for a subset of our taxa. We again failed to find the monophyly of Australian taxa or New Zealand taxa but the grouping of *Anisoura*/*Motuweta* (clade A) with *Aistus* and *Carcinopsis* as sister to *Hemideina–Deinacrida* (clade B) was supported (tree not shown).

Dataset IV (COI sequences from Australasian taxa), 1225 bp aligned DNA sequence data for 28 ingroup taxa, plus two outgroup taxa (one Madagascan gryllacridid (Mad 379) and one South African stenopelmatid (F234); see the electronic supplementary material). With RY-coded mitochondrial COI data, our phylogenetic analyses returned clades consistent with those observed in analysis of 18S data (figure 4), including strong support for multiple, distinct weta lineages in New Zealand (clades A, B, CI, CII; figure 5). This result was strongly supported by the SH test constraining New Zealand taxa to form a monophyletic lineage ($p<0.0001$). *Anisoura*/*Motuweta* (clade A) was again found to be sister to *Aistus* and *Carcinopsis* and collectively were sister to the *Hemideina*/*Deinacrida* (clade B; figure 5). We examined the translated amino acid sequence and observed 10 changes shared by *Aistus*, *Carcinopsis* and *Anisoura*/*Motuweta* (out of 409) clearly separating them from the rest of Anostostomatidae. Analysis of dataset V, a reduced set of taxa with concatenated sequences, consisted of COI (1225 bp) and 12S (435 bp) for Australasian Anostostomatids (25 taxa, 1860 bp). Analyses returned a similar topology with the same taxon subgroupings and no support for the New Zealand monophyly as with previous datasets (not shown).

Analysis of dataset VI (COI *Hemiandrus* only), for 46 individuals from at least 15 New Zealand species and two Australian *Hemiandrus* species (848 bp), revealed high genetic diversity (mean genetic distance, 6.5 and 7.2% for the North and South Islands, respectively) among samples of the widespread New Zealand taxon

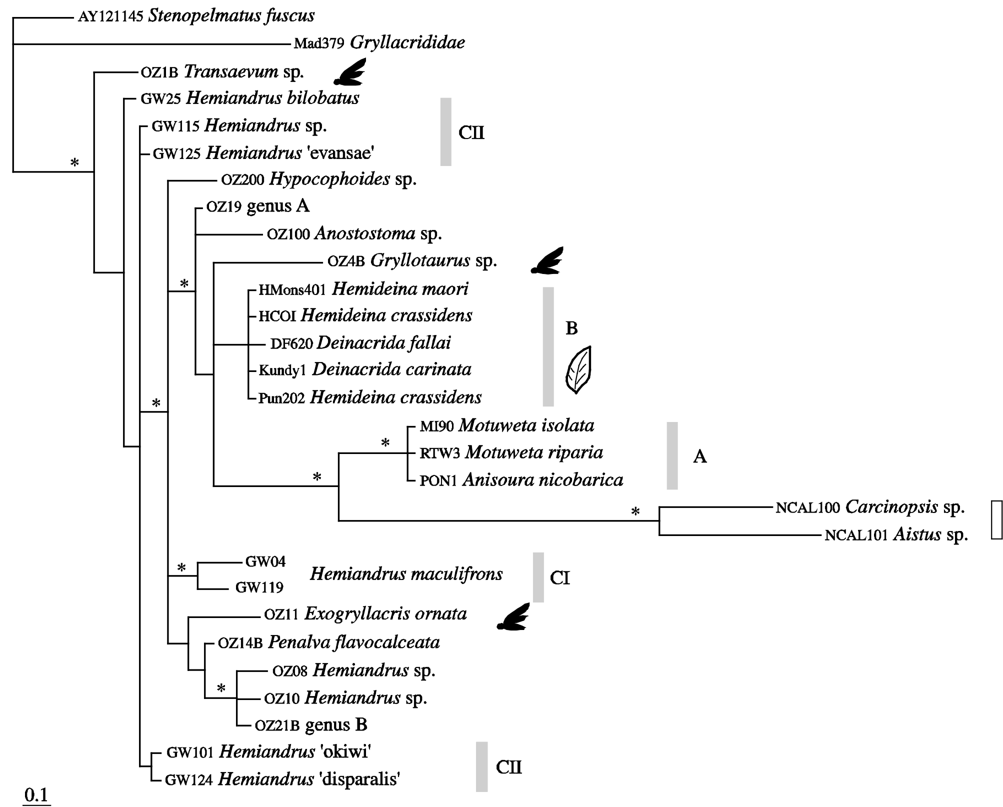


Figure 4. Bayesian phylogram for 1746 bp of nuclear genes 18S including representatives of the Australasian anostostomatids. The grey and white bars indicate New Zealand and New Caledonian taxa, respectively. Clades of New Zealand taxa are indicated as A, B, CI and CII. The single asterisks at nodes indicate Bayesian posterior probabilities above 0.9.

Hemiandrus maculifrons. Samples of *H. maculifrons* formed two main clades, one from South Island and one from North Island (clade CI, figure 6), with a specimen from northern South Island (sample GW93A from Pelorus Bridge) sister to the North Island clade. By contrast, most other New Zealand *Hemiandrus* species were closely related to one another (e.g. 3.9% between *H. bilobatus* and *H.* 'promontorius'), and these species tend to have narrow geographical ranges. All ground weta (whether North Island or South Island) species with extremely reduced/short ovipositors form a monophyletic group with long ovipositor species paraphyletic to this. Two taxa with medium ovipositors (*H.* 'okiwi' and *H.* 'evansae') each are sister to long ovipositor species, and not together. Other anostostomatids, including the sister Australian '*Hemiandrus*', have long ovipositors and therefore the reduction of ovipositor length (along with the unusual maternal care that appears to correlate with these conditions) appears to have evolved at least three times in New Zealand. Investigation into monophyly of the genus using the SH test on dataset IV revealed that constraining all *Hemiandrus* (Australian and New Zealand) to be monophyletic resulted in a tree with a significantly worse likelihood score ($p<0.0001$).

(b) *Summary of phylogenetics*

New Zealand weta are not monophyletic, instead forming three (or four) clades. The New Zealand tusked weta (clade A) genera are more closely related to the New Caledonian taxa (*Aistus* sp. and *Carcinopsis* sp.; figures 4 and 5) with no apparent close Australian relative. New Zealand's large herbivorous tree and giant weta (*Hemideina* and *Deinacrida)* form a distinct clade that is strongly supported with less than 1 per cent sequence divergence in the nuclear rDNA gene 18S (clade B, figure 4). All our analyses returned a clade consisting of *Aistus*, *Carcinopsis*, *Anisoura*/*Motuweta* (clade A) and *Hemideina*–*Deinacrida* (clade B). However, long branches due to an apparent rate shift in the *Aistus*, *Carcinopsis* and *Anisoura*/*Motuweta* (clade A) lead to us remove these taxa from further analyses, so this relationship is not fully understood. The third New Zealand clade, the *Hemiandrus,* consists of two lineages (clades CI and CII; figures 4–6) that may not be sisters. One lineage consists of only *H. maculifrons* and *H.* 'alius' while the other includes 11+ species sampled (figure 6). Finally, the nine Australian taxa are not monophyletic but form four clades throughout the tree, sister to New Zealand clades; however, there is little BPP support for these

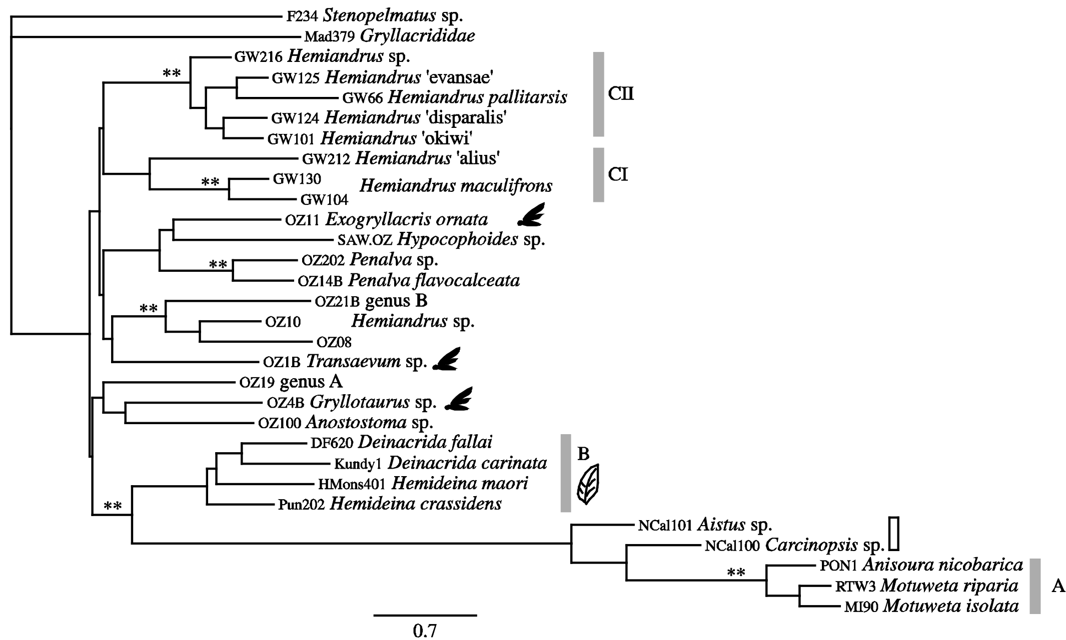


Figure 5. ML phylogram for 1225 bp of the mitochondrial COI gene representing Australasian anostostomatids. Sequences were RY coded on the third codon position to reduce noise while retaining phylogenetic signal. Winged silhouette indicates winged species and the leaf symbol indicates herbivorous species. The grey and open bars indicate New Zealand and New Caledonian taxa, respectively. Clades of New Zealand taxa are indicated as A, B, CI and CII. The double asterisks at nodes indicate Bayesian posterior probabilities of 0.99 and above.

nodes. This is consistent with the short branch lengths obtained at the base of the tree.

Dataset IV (COI–RY-coded Australasia; figure 5) returned three Australian clades of interest: (i) winged *Transaevum* sister to the non-winged Australian *Hemiandrus* and genus B, (ii) winged *Exogryllacris* sister to the non-winged *Hypocophoides* and *Penalva*, and (iii) winged *Gryllotaurus* sister to the non-winged *Anostostoma* and genus A. These three Australian clades were not resolved in the analysis of dataset II (18S Australasia; figure 4) but the three winged species were never monophyletic and *Transaevum* was sister to all the other Anostostomatidae.

(c) *Divergence estimation*

Edge lengths on branches suggest that *Aistus* and *Carcinopsis* along with *Anisoura*/*Motuweta* (clade A; figures 4 and 5) may have an increased rate of mutation. In order to test for this, we compared ML trees with and without a clock enforced and found evidence in both 18S and COI datasets for a significant deviation from a clock-like model of evolution. By removing *Aistus*, *Carcinopsis* and *Anisoura*/*Motuweta* and running the ML analysis again, the significance of the SH test decreased.

We carried out relaxed-clock analyses on 18S (dataset II) and COI (dataset IV). Our calibration points assumed either existence of all anostostomatid weta lineages prior to the separation of Zealandia (82 Ma) or emergence of New Caledonia (40 Ma). However, due to the observed rate change in *Aistus*, *Carcinopsis* and *Anisoura*/*Motuweta*, we removed this clade (five sequences) from the dating analysis.

Estimates of the rate of evolution of COI vary a great deal, primarily due to the disparity between population- and species-level rates (time-dependent rates or J-shape curve; Ho *et al.* 2005; Penny 2005). Here, we restrict our comparisons to recent studies that examine family-level divergences of insects, which estimate the rate of evolution for COI and COII of insects at approximately 0.007–0.012 substitutions per site per million years (Zakharov *et al.* 2004 and references therein). In our relaxed molecular clock phylogenetic analysis, constraining the base of the Anostostomatidae clade to 82 Ma, our estimated mutation rates for COI were between 0.0097 and 0.0376 [95% CI], compared with a mutation rate of 0.0223–0.1219 when applying the 40 Ma constraint. The mutation rate derived using the 82 Ma constraint is more similar to published data for other insects (Zakharov *et al.* 2004), and therefore supports divergences within Anostostomatidae of more than 80 Ma. Our relaxed-clock analyses of COI data suggest that both New Zealand *Hemiandrus* and *Hemideina–Deinacrida* lineages may have diverged from Australian relatives more than 82 Ma. Analysis of our 18S sequences with the 82 Ma constraint estimates divergence for New Zealand *Hemiandrus* clades that extend to before continental break-up (45–120 Ma and 11–103 Ma), but the *Hemideina–Deinacrida* clade was estimated to have diverged post-Gondwanan break-up (3–38 Ma). Thus, our sequence data support the idea that some weta lineages may have

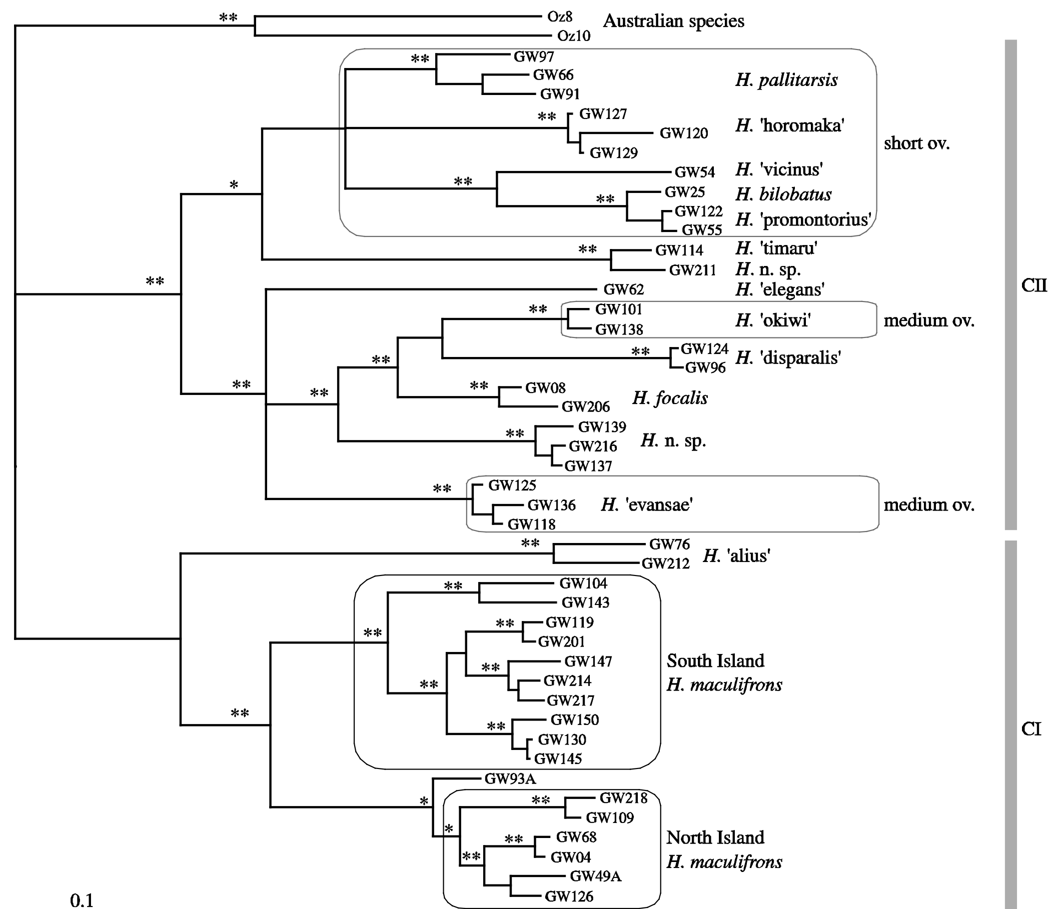


Figure 6. Bayesian phylogram of the genus *Hemiandrus* for 848 bp of the mitochondrial COI gene. Species with short or medium-length ovipositor are indicated; all other species have long ovipositors. The two New Zealand clades CI and CII are indicated. The single and double asterisks at nodes indicate Bayesian posterior probabilities above 0.9 and 0.99, respectively.

diverged before the separation of Zealandia from Australia but also that dispersal has since occurred. The phylogenetic placement of *Aistus, Carcinopsis* and *Anisoura/Motuweta* as sister to *Hemideina–Deinacrida* suggests genetic exchange between New Zealand and New Caledonia after separation of Zealandia.

4. DISCUSSION

Despite comprehensive morphological studies, phylogenetic relationships within the Ensifera are poorly understood (Gwynne 1995; Whiting 2002*b*; Desutter-Grandcolas 2003). Johns (1997) removed taxa from Stenopelmatidae to form Anostostomatidae, a separation subsequently supported by molecular analyses (Jost & Shaw 2006). Although we are not concerned here with deeper Ensiferan relationships, it is important to know that our taxon set comprises a true ingroup. We found support for the monophyly of Anostostomatidae in our analysis (0.96 BPP) and for the close relationship with the Gryllacrididae and Stenopelmatidae (figure 3), supporting previous inferences (Jost & Shaw 2006; P. M. Johns 2007, personal communication). However, we did not find evidence of a sister relationship of Deinacridinae (*Hemideina* and *Deinacrida*) and Anostostomatidae (rest of the family; Johns 1997; Gorochov 2001).

For the first time, we have shown that members of the family Anostostomatidae are not monophyletic in New Zealand or Australia. To explain the phylogenetic diversity of the New Zealand weta by vicariance requires that at least four distinct clades of Anostostomatidae were already present in Gondwana before Zealandia split from Australia, and that some of these subsequently went extinct in Australia. On the face of it, this seems an unlikely scenario, given the small size and geological activity of New Zealand compared with Australia, and indeed this has been shown to be a poor explanation for the distribution of *Nothofagus* beech in the region (Cook & Crisp 2005*a*). Although we found some variation in node dates inferred from COI and 18S data, we have to reject the hypothesis that all

New Zealand lineages arose before continental break-up (*ca* 82 Ma). However, relaxed molecular clock calibrated phylogenies do suggest that some New Zealand clades may have formed before continental separation. These inferred early splits are consistent with a vicariant origin and survival of some Anostostomatidae lineages on Zealandia throughout the Oligocene marine transgression. Taxa missing from analyses (owing to extinction) will always result in long unbroken branches in phylogenetic trees and thus the inference of great age since common ancestors (Cook & Crisp 2005*b*) whereas recent splits (short branches) cannot be made older by the inclusion of 'missing taxa'.

Colonization of New Zealand from the Australian biota, which includes three separate winged lineages, might have been facilitated by increasing land area after the Oligocene (less than 22 Ma). Dispersal events continue today, and include the establishment of an Australian Gryllacridid in recent years (Green & Ramsay 2003). The current study suggests that the two New Caledonia genera are more closely related to one of the New Zealand lineages but not to any Australian taxa. This is despite the comparatively close physical proximity and more similar climate of New Caledonia and Queensland, Australia. Despite evidence of an elevated substitution rate in both nuclear and mitochondrial genes, we observed clear phylogenetic evidence, supported by amino acid substitutions, for the sister relationship of the *Aistus, Carcinopsis* and *Anisoura*/*Motuweta*. Weta must have colonized New Caledonia after it emerged from the sea *ca* 40 Ma (Grandcolas *et al.* 2008), but our current taxon sampling is not sufficient to prove whether those weta ancestors came to or from New Zealand. *Hemideina* and *Deinacrida* were found to form a monophyletic group corresponding to the subfamily Deinacridinae Karny 1932. The lack of variation at the 18S gene among 18 species in these two genera is indicative of a recent radiation. Interestingly, Johns (1997) suggested that one of the tusked weta genera, *Anisoura,* should be included in the Deinacridinae; however, if Deinacridinae is to be valid, it should include all tusked weta genera plus the New Caledonian genera. Further morphological study is required to test this.

Our analysis of Australian and New Zealand *Hemiandrus* shows clearly that they represent separate lineages (figure 4). Furthermore, *Hemiandrus* in New Zealand consists of two distinct clades that may not be monophyletic. It is clear that taxonomic revision of the genus is required, with *Hemiandrus* being retained only for New Zealand taxa. Our data are the first to illustrate the depth of diversity within the New Zealand *Hemiandrus* that may amount to some 40 species (Johns 2001), using a broad range of habitats from alpine to lowland forest. We note relatively high genetic diversity in the most widespread species (*H. maculifrons*) that might justify separation as two or more allopatric species, but similar levels of sequence divergence are reported from other widespread weta with wide geographical ranges, e.g. *Deinacrida connectens* (Trewick *et al.* 2000) and *Hemideina thoracica* (Morgan-Richards *et al.* 2001). There is a stark contrast between the wide geographical range of *H. maculifrons* and the small ranges of the numerous, genetically similar, local endemics of South Island (figures 2 and 6). This pattern implies recent diversification in South Island, perhaps in response to habitat diversification since the Pliocene, which is also observed in the giant weta (figure 2) and many other taxa (Wagstaff & Garnock-Jones 1998; Lockhart *et al.* 2001; Trewick 2001; Chinn & Gemmell 2004; Trewick 2008). By contrast, much of the North Island is young (less than 1 Ma) and relatively homogeneous in terms of habitat (Trewick & Morgan-Richards in press).

In terms of geological history, Zealandia bears little resemblance to other larger, more stable southern continents that together originated from Gondwana. Instead, its history is more like that of many oceanic islands, undergoing extensive geological transformations that mark significant changes in climate and topology, and ultimately shaping the biota. The lack of monophyly within the Anostostomatidae fauna is not entirely unexpected, but the idea that some lineages represent more ancient links between Australia and New Zealand is exciting, as is the apparent exchange between New Zealand and the other Zealandian island, New Caledonia. This pattern might be indicative of the New Zealand biota as a whole, where some old lineages surviving from Zealandia, are all but overshadowed by more recent biotic exchange.

A special thank you to all those involved in collecting over the years: Geoff Monteith, Peter Johns, George Gibbs, Darryl Gwynne, Scott Dunavan. Special thanks to Matt Phillips and Simon Hills for guidance and assistance with the inner workings of BEAST. Thanks to Andrew Clarke, Sheralee Cleland, Barbara Holland, Manda Jost, David Penny and Michael Whiting for their constructive comments on the manuscript. This research was supported by funding from the Marsden Fund granted to David Penny and the Allan Wilson Centre for Molecular Ecology and Evolution. The Orthoptera Society grant was awarded to R.C.P. facilitating collaboration with Peter Johns. New Zealand Department of Conservation provided one *Motuweta* and two *Deinacrida* species.

REFERENCES

Baker, A. J., Huynen, L. J., Haddrath, O., Millar, C. D. & Lambert, D. M. 2005 Reconstructing the tempo and mode of evolution in an extinct clade of birds with ancient DNA: the giant moas of New Zealand. *Proc. Natl Acad. Sci. USA* **102**, 8257–8262. (doi:10.1073/pnas.0409435102)

Caughley, G. 1964 Does the New Zealand vertebrate fauna conform to zoogeographic principles? *Tuatara* **12**, 50–57.

Chardon, D. & Chevillotte, V. 2006 Morphotectonic evolution of the New Caledonia ridge (Pacific Southwest) from post-obduction tectonosedimentary record. *Tectonophysics* **420**, 473–491. (doi:10.1016/j.tecto.2006.04.004)

Chinn, W. G. & Gemmell, N. J. 2004 Adaptive radiation within New Zealand endemic species of the cockroach genus *Celatoblatta* Johns (Blattidae): a response to Plio-Pleistocene mountain building and climate change. *Mol. Ecol.* **13**, 1507–1518. (doi:10.1111/j.1365-294X.2004.02160.x)

Cook, L. G. & Crisp, M. D. 2005*a* Not so ancient: the extant crown group of *Nothofagus* represents a post-Gondwanan radiation. *Proc. R. Soc. B* **272**, 2535–2544. (doi:10.1098/rspb.2005.3219)

Cook, L. G. & Crisp, M. D. 2005*b* Directional asymmetry of long-distance dispersal and colonization could mislead

reconstructions of biogeography. *J. Biogeogr.* **32**, 741–754. (doi:10.1111/j.1365-2699.2005.01261.x)

Cowie, R. H. & Holland, B. S. 2006 Dispersal is fundamental to biogeography and the evolution of biodiversity on oceanic islands. *J. Biogeogr.* **33**, 193–198. (doi:10.1111/j.1365-2699.2005.01383.x)

Desutter-Grandcolas, L. 2003 Phylogeny and evolution of acoustic communication in extant Ensifera (Insecta, Orthoptera). *Zool. Scr.* **32**, 525–561. (doi:10.1046/j.1463-6409.2003.00142.x)

Drummond, A. J. & Rambaut, A. 2007 BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214. (doi:10.1186/1471-2148-7-214)

Drummond, A. J., Ho, S. Y. W., Phillips, M. J. & Rambaut, A. 2006 Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4**, e88. (doi:10.1371/journal.pbio.0040088)

Ericson, P. G. P., Christidis, L., Cooper, A., Irestedt, M., Jackson, J., Johansson, U. S. & Norman, J. A. 2002 A Gondwanan origin of passerine birds supported by DNA sequences of the endemic New Zealand wrens. *Proc. R. Soc. B* **269**, 235–241. (doi:10.1098/rspb.2001.1877)

Field, L. H. & Deans, N. A. 2001 Sexual selection and secondary sexual characters of wetas and king crickets. In *The biology of wetas, king crickets and their allies* (ed. L. H. Field), pp. 179–204. Wallingford, UK: CABI Publishing.

Fleming, C. A. 1962 New Zealand biogeography—a palaeontologist's approach. *Tuatara* **10**, 53–108.

Fleming, C. A. 1979 *The geological history of New Zealand and its life*. Auckland, New Zealand: Oxford University Press.

Flook, P. K., Klee, S. & Rowell, C. H. F. 1999 Combined molecular phylogenetic analysis of the Orthoptera (Arthropoda, Insecta) and implications for their higher systematics. *Syst. Biol.* **48**, 233–253. (doi:10.1080/106351599260274)

Folmer, O., Black, M., Hoeh, W., Lutz, R. & Vrijenhoek, R. 1994 DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotechnol.* **3**, 294–299.

Gibbs, G. 2006 *Ghosts of Gondwana: the history of life in New Zealand*. Nelson, New Zealand: Craig Potton Publishing.

Goldberg, J., Trewick, S. A. & Paterson, A. M. 2008 Evolution of New Zealand's terrestrial fauna: a review of molecular evidence. *Phil. Trans. R. Soc. B* **363**, 3319–3334. (doi:10.1098/rstb.2008.0114)

Gorochov, A. V. 2001 The higher classification, phylogeny and evolution of the superfamily Stenopelmatiodea. In *The biology of wetas, king crickets and their allies* (ed. L. H. Field), pp. 3–33. Wallingford, UK: CABI Publishing.

Grandcolas, P., Murienne, J., Robillard, T., Desutter-Grandcolas, L., Jourdan, H., Guilbert, E. & Deharveng, L. 2008 New Caledonia: a very old Darwinian island? *Phil. Trans. R. Soc. B* **363**, 3309–3317. (doi:10.1098/rstb.2008.0122)

Green, O. R. & Ramsay, G. W. 2003 A winged weta, *Pterapotrechus* (Orthoptera: Gryllacrididae), established in New Zealand. *New Zeal. Entomol.* **26**, 75–77.

Grimaldi, D. & Engel, M. S. 2005 *Evolution of insects*. New York, NY: Cambridge University Press.

Guindon, S. & Gascuel, O. 2003 A simple, fast and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696–704. (doi:10.1080/10635150390235520)

Gwynne, D. T. 1995 Phylogeny of the Ensifera (Orthoptera): a hypothesis supporting multiple origins of acoustical signalling, complex spermatophores and maternal care in crickets, katydids, and weta. *J. Orthopt. Res.* **4**, 203–218. (doi:10.2307/3503478)

Gwynne, D. T. 2004 Reproductive behaviour of ground weta (Orthoptera: Anostostomatidae): drumming behaviour, nuptial feeding, post-copulatory guarding and maternal care. *J. Kans. Entomol. Soc.* **77**, 414–428. (doi:10.2317/E-34.1)

Hendy, M. D. & Penny, D. 1989 A framework for the quantitative study of evolutionary trees. *Syst. Biol.* **38**, 297–309.

Ho, S. Y. W., Phillips, M. J., Cooper, A. & Drummond, A. J. 2005 Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Mol. Biol. Evol.* **22**, 1561–1568. (doi:10.1093/molbev/msi145)

Huelsenbeck, J. P. & Ronquist, F. 2001 MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755. (doi:10.1093/bioinformatics/17.8.754)

Jewell, T. 2007 Two new species of *Hemiandrus* (Orthoptera: Anostostomatidae) from Fiordland National Park. *New Zealand. Zootaxa* **1542**, 49–57.

Johns, P. M. 1997 The Gondwanaland weta: family Anostostomatidae (formerly in Stenopelmatidae, Henicidae or Mimmermidae): nomenclatural problems, world checklist, new genera and species. *J. Orthopt. Res.* **6**, 125–138. (doi:10.2307/3503546)

Johns, P. M. 2001 *Distribution and conservation status of ground weta, Hemiandrus species* (Orthoptera: Anostostomatidae). Wellington, New Zealand: Department of Conservation.

Jost, M. C. & Shaw, K. L. 2006 Phylogeny of Ensifera (Hexapoda: Orthoptera) using three ribosomal loci, with implications for the evolution of acoustic communication. *Mol. Phylogenet. Evol.* **38**, 510–530. (doi:10.1016/j.ympev.2005.10.004)

Knapp, M., Stöckler, K., Havell, D., Delsuc, F., Sebastiani, F. & Lockhart, P. J. 2005 Relaxed molecular clock provides evidence for long distance dispersal of *Nothofagus* (Southern Beech). *PLoS Biol.* **3**, e14. (doi:10.1371/journal.pbio.0030014)

Landis, C. A., Campbell, H. J., Begg, J. G., Paterson, A. M. & Trewick, S. A. 2008 The Waipounamu erosion surface: questioning the antiquity of the New Zealand land surface and terrestrial fauna and flora. *Geol. Mag.* **145**, 1–25. (doi:10.1017/S0016756807004268)

Lockhart, P. J., McLenachan, P. A., Havell, D., Glenny, D., Huson, D. & Jensen, U. 2001 Phylogeny, radiation and transoceanic dispersal of New Zealand alpine buttercups: molecular evidence under split decomposition. *Ann. Miss. Bot. Gard.* **88**, 458–477. (doi:10.2307/3298586)

McLoughlin, S. 2001 The breakup history of Gondwana and its impact on pre-Cenozoic floristic provincialism. *Aust. J. Bot.* **49**, 271–300. (doi:10.1071/BT00023)

Meads, M. J. 1990 *The weta book: a guide to the identification of wetas*. Lower Hutt, New Zealand: DSIR Land Resources.

Monteith, G. B. & Field, L. H. 2001 Australian king crickets: distribution, habitats and biology (Orthoptera: Anostostomatidae). In *The biology of wetas, king crickets and their allies* (ed. L. H. Field), pp. 79–94. Wallingford, UK: CABI Publishing.

Morgan-Richards, M., Trewick, S. A. & Wallis, G. 2001 Chromosome races with Pliocene origins: evidence from mtDNA. *Heredity* **86**, 303–312. (doi:10.1046/j.1365-2540.2001.00828.x)

Mortimer, N. 2004 New Zealand's geological foundations. *Gondw. Res.* **7**, 261–272. (doi:10.1016/S1342-937X(05)70324-5)

Mortimer, N., Hoernle, K., Hauff, F., Palin, J. M., Dunlap, W. J., Werner, R. & Faure, K. 2006 New constraints on the age and evolution of the Wishbone Ridge, southwest Pacific Cretaceous microplates, and Zealandia-West Antarctica breakup. *Geology* **34**, 185–188. (doi:10.1130/G22168.1)

Neall, V. E. & Trewick, S. A. 2008 The age and origin of the Pacific Islands: a geological overview. *Phil. Trans. R. Soc. B* **363**, 3293–3308. (doi:10.1098/rstb.2008.0119)

Penny, D. 2005 Relativity for molecular clocks. *Nature* **436**, 183–184. (doi:10.1038/436183a)

Phillips, M. J. & Penny, D. 2003 The root of the mammalian tree inferred from whole mitochondrial genomes. *Mol. Phylogenet. Evol.* **28**, 171–185. (doi:10.1016/S1055-7903(03)00057-5)

Posada, D. & Buckley, T. R. 2004 Model selection and model averaging in phylogenetics: advantages of Akaike Information Criterion and Bayesian approaches over Likelihood Ratio tests. *Syst. Biol.* **53**, 793–808. (doi:10.1080/10635150490522304)

Posada, D. & Crandall, K. A. 1998 Modeltest: testing the model of DNA substitution. *Bioinformatics* **14**, 817–818. (doi:10.1093/bioinformatics/14.9.817)

Rambaut, A. 1996 Se-Al: sequence alignment editor. See http://evolve.zoo.ox.ac.uk.

Shimodaira, H. & Hasegawa, M. 1999 Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* **16**, 1114–1116.

Simon, C., Frati, F., Bechenbach, A., Crespi, B., Liu, H. & Flook, P. 1994 Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Ann. Entomol. Soc. Am.* **87**, 651–701.

Sunnucks, P. & Hale, D. F. 1996 Numerous transposed sequences of mitochondrial cytochrome oxidase I–II in aphids of the genus *Sitobion* (Hemiptera: Aphididae). *Mol. Biol. Evol.* **13**, 510–524.

Swofford, D. L. 2003 *PAUP*: phylogenetic analysis using parsimony (*and other methods)*, v. 4.0.b10. Sunderland, MA: Sinauer Associates.

Terry, M. D. & Whiting, M. F. 2005 Mantophasmatodea and phylogeny of the lower neopterous insects. *Cladistics* **21**, 240–257. (doi:10.1111/j.1096-0031.2005.00062.x)

Trewick, S. A. 2001 Identity of an endangered grasshopper (Acididae: *Brachaspis*): taxonomy, molecules and conservation. *Conserv. Genet.* **2**, 233–243. (doi:10.1023/A:1012263717279)

Trewick, S. A. 2008 DNA barcoding is not enough: mismatch of taxonomy and genealogy in New Zealand grasshoppers (Orthoptera: Acrididae). *Cladistics* **24**, 240–254. (doi:10.1111/j.1096-0031.2007.00174.x)

Trewick, S. A. & Morgan-Richards, M. 2004 Phylogenetics of New Zealand's tree, giant and tusked weta (Orthoptera: Anostostomatidae): evidence from mitochondrial DNA. *J. Orthopt. Res.* **13**, 185–196. (doi:10.1665/1082-6467(2004)013[0185:PONZTG]2.0.CO;2)

Trewick, S. A. & Morgan-Richards, M. 2005 After the deluge: mitochondrial DNA indicates Miocene radiation and Pliocene adaptation of tree and giant weta (Orthoptera: Anostostomatidae). *J. Biogeogr.* **31**, 295–309. (doi:10.1111/j.1365-2699.2004.01179.x)

Trewick, S. A. & Morgan-Richards, M. In press. *Biology of New Zealand. Encyclopedia of Islands*. Berkeley, CA: University of California Press.

Trewick, S. A., Wallis, G. P. & Morgan-Richards, M. 2000 Phylogeographical pattern correlates with Pliocene mountain building in the alpine scree weta (Orthoptera, Anostostomatidae). *Mol. Ecol.* **9**, 657–666. (doi:10.1046/j.1365-294x.2000.00905.x)

Trewick, S. A., Paterson, A. M. & Campbell, H. J. 2007 Hello New Zealand. *J. Biogeogr.* **34**, 1–6. (doi:10.1111/j.1365-2699.2006.01643.x)

Vawter, L. 1991 Evolution of blattoid insects and of the small subunit ribosomal gene. PhD dissertation, University of Michigan, Ann Arbor.

Veevers, J. J. & Li, Z. X. 1991 Review of seafloor spreading around Australia. II. Marine magnetic anomaly modelling. *Aust. J. Earth Sci.* **38**, 391–408. (doi:10.1080/08120099108727980)

Wagstaff, S. J. & Garnock-Jones, P. J. 1998 Evolution and biogeography of the *Hebe* complex (Scrophulariaceae) inferred from ITS sequences. *New Zeal. J. Bot.* **36**, 425–437.

Waters, J. M. & Craw, D. 2006 Goodbye Gondwana? New Zealand biogeography, geology and the problem of circularity. *Syst. Biol.* **55**, 351–356. (doi:10.1080/10635150600681659)

Whiting, M. F. 2002*a* Mecoptera is paraphyletic: multiple genes and phylogeny of Mecoptera and Siphonaptera. *Zool. Scr.* **31**, 93–104. (doi:10.1046/j.0300-3256.2001.00095.x)

Whiting, M. F. 2002*b* Phylogeny of the holometabolous insect orders: molecular evidence. *Zool. Scr.* **31**, 3–15. (doi:10.1046/j.0300-3256.2001.00093.x)

Zakharov, E., Caterino, M. & Sperling, F. 2004 Molecular phylogeny, historical biogeography and divergence time estimates for swallowtail butterflies of the genus *Papilio* (Lepidoptera: Papilionidae). *Syst. Biol.* **54**, 193–215. (doi:10.1080/10635150490423403)

Zwickl, D. J. 2006 Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. PhD dissertation, The University of Texas, Austin.

	MBE	msn248	AP
	Journal Name	Art. No.	CE Code

Toward Resolving Deep Neoaves Phylogeny: Data, Signal Enhancement, and Priors

Renae C. Pratt, Gillian C. Gibb,* Mary Morgan-Richards,* Matthew J. Phillips,† Michael D. Hendy,* and David Penny**

5 *Allan Wilson Centre for Molecular Ecology and Evolution, Massey University, Palmerston North, New Zealand; and †Centre for Macroevolution and Macroecology, School of Botany and Zoology, Australian National University, Canberra ACT, Australia

We report three developments toward resolving the challenge of the apparent basal polytomy of neoavian birds. First, we describe improved conditional down-weighting techniques to reduce noise relative to signal for deeper divergences and find increased agreement between data sets. Second, we present formulae for calculating the probabilities of finding 10 predefined groupings in the optimal tree. Finally, we report a significant increase in data: nine new mitochondrial (mt) genomes (the dollarbird, New Zealand kingfisher, great potoo, Australian owlet-nightjar, white-tailed trogon, barn owl, a roadrunner [a ground cuckoo], New Zealand long-tailed cuckoo, and the peach-faced lovebird) and together they provide data for each of the six main groups of Neoaves proposed by Cracraft J (2001). We use his six main groups of modern birds as priors for evaluation of results. These include passerines, cuckoos, parrots, and three other groups termed 15 "WoodKing" (woodpeckers/rollers/kingfishers), "SCA" (owls/potoos/owlet-nightjars/hummingbirds/swifts), and "Conglomerati." In general, the support is highly significant with just two exceptions, the owls move from the "SCA" group to the raptors, particularly accipitrids (buzzards/eagles) and the osprey, and the shorebirds may be an independent group from the rest of the "Conglomerati". Molecular dating mt genomes support a major diversification of at least 12 neoavian lineages in the Late Cretaceous. Our results form a basis for further testing with both nuclear-coding sequences and rare 20 genomic changes.

Introduction

Perhaps, the greatest current challenge of avian systematics for molecular evolutionists and systematists alike is the resolution of the polytomy at the base of the Neoaves. 25 The basic paleognath (tinamous and ratites)–neognath division (all other modern birds) is supported by studies of morphology (Cracraft and Clarke 2001), nuclear-coding DNA (Groth and Barrowclough 1999; García-Moreno and Mindell 2000), and mitochondrial (mt) genomes (Sorenson 30 et al. 2003; Harrison et al. 2004; Slack et al. 2007). Within the Neognathae, the Galloanserae (chickens, ducks, and their relatives) represent the earliest divergence, leaving the large majority (all remaining orders) of birds in the Neoaves. Again, coding regions of both mt genomes and 35 nuclear DNA, together with morphological data, agree with the Galloanserae division. However, resolving the relationships within Neoaves is still elusive.

Thus, resolution of the basal Neoavian polytomy could be seen as the "last frontier" for resolving deep-level 40 systematics among modern birds. There are a range of views best illustrated by the two ends of a spectrum—first, the theory that the basal polytomy is due to an "explosive radiation" after the Cretaceous–Paleogene (K–Pg, formerly K–T) boundary. That is, birds and mammals "inherited the 45 earth" only after the demise of the dinosaurs and pterosaurs (Feduccia 2003; Chubb 2004; Poe and Chubb 2004; Ericson et al. 2006). The other end of the spectrum are hypotheses that basal avian lineages were diversifying in an "adaptive radiation" long before the asteroid impact that 50 marks the K–Pg boundary (Cooper and Penny 1997; Cracraft 2001; van Tuinen and Hedges 2001; Penny and Phillips 2004; Pereira and Baker 2006; van Tuinen et al. 2006; Brown et al. 2007, 2008). This latter approach represents mainstream evolutionary theory in that it attempts to explain the past by reference to known mechanisms—to 55 "causes now in operation" (Penny and Phillips 2004).

Poe and Chubb (2004) suggested that the large polytomy at the base of Neoaves represents a rapid radiation that "might be considered essentially simultaneous." If a lack of resolution is not caused by truly short times between diver- 60 gences, then ultimately relationships should be resolvable (Whitfield and Lockhart 2007). We have already commented (Gibb et al. 2007) that an explosive radiation implies both short divergence times between avian orders and also that the ecological and morphological differences 65 that identify the crown groups of orders within Neoaves must have occurred over the same short timescale. It would scarcely be an explosive radiation if the lineages diverged quickly, but it then took tens of millions of years for genetic changes to occur leading to the ecological and morpholog- 70 ical characters that distinguish crown group Neoavian orders today. Apart from being real, short branch lengths in phylogenies can result for a number of reasons. For example, different characters or data sources that provide support for conflicting trees (rather than from the absence of 75 support) can result in short branches; under these conditions, even standard maximum likelihood (ML) can seriously underestimate branch lengths (Penny et al. 2008). Additionally, use of inappropriate genes or analysis methods can return short branches. To improve divergence time 80 estimates, Brown et al. (2008) recommend longer sequences (they used 4,594 bp of mtDNA) and the use of multiple independent nuclear loci. In addition to incorporating longer sequences, improved analytic methods such as networks allows visualization of conflict between essen- 85 tially equally good phylogenies (Holland et al. 2004).

A greater understanding of the evolutionary history of Neoaves is still needed. We have found Cracraft's (2001) six groupings within Neoaves to be useful as "informal priors" for recent studies on passerines and for the group 90 termed "Conglomerati" (e.g., Gibb et al. 2007). With

Key words: Neoaves, mitochondrial genomes, site-stripping, down-weighting, hypothesis testing, avian evolution.

E-mail: r.c.pratt@massey.ac.nz.

Mol. Biol. Evol. 25(12):1–14. 2008
doi:10.1093/molbev/msn248
Advance Access publication November 3, 2008

© The Author 2008. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. All rights reserved.
For permissions, please e-mail: journals.permissions@oxfordjournals.org

additional data presented here, we have representatives (and test) his six prior groupings within Neoaves:

(i) Passeriformes (passerines, perching birds);
(ii) Psittaciformes (parrots);
(iii) Cuculiformes (cuckoos);
(iv) Coliiformes, Coraciiformes, Trogoniformes, and Piciformes (mousebirds [not included in the present study], rollers/bee-eaters/kingfishers, trogons, woodpeckers/toucan, and jacamars). Informally, and for ease of communication, we call this the "WoodKing" grouping;
(v) Strigiformes, Caprimulgiformes, Apodiformes, and Musophagiformes (owls, nightjars, potoos, owlet-nightjars, hummingbirds, swifts, and turacos). We have sampled three of the four groups, so we will refer to these as "SCA"(or "CA" when omitting the owls [Strigiformes]); and
(vi) a diverse group dubbed the "Conglomerati" (Slack et al. 2007) that includes three main subgroups: 1) Falconiformes, Pelecaniformes, Ciconiiformes, Procellariiformes, Gaviiformes Sphenisciformes, and Podicipediformes (raptors, pelicans and relatives, storks, seabirds, loons, penguins, and grebes); 2) Charadriiformes, Columbiformes, and Pteroclidae (shorebirds, pigeons, and sandgrouse); and 3) Gruidae, Rallidae, Otididae, Eurypygidae, and Turnices (cranes, rails, bustards, kagu, sun bitterns, and button quails).

In recent papers (Slack et al. 2006, Slack et al. 2007), we concentrated on the relationships within (i) and (vi) but have recently extended this to include members of groups (iv) and (v) (Gibb et al. 2007; Morgan-Richards et al. 2008). Because we previously only had a single representative for the orders Psittaciformes (parrots) and Strigiformes (owls) (Harrison et al. 2004), we omitted them from recent analyses because they are long branches that are known to be problematic in phylogeny generally (Hendy and Penny 1989), including birds (Harrison et al. 2004).

At this stage, we are particularly concerned as to whether the members within each of the above six groups come together—this will help evaluate whether the deep Neoavian lineages are resolvable. As such, we are not especially concerned if some of Cracraft's six groups have paraphyletic lineages within them—that is, a taxonomic question, not a question about the resolvability of the deepest lineages. For example, we are interested in whether the combined Coliiformes/Coraciiformes/Piciformes ("WoodKing") form a natural group within Neoaves—even if one of the subgroups turns out to be paraphyletic within this "WoodKing" grouping.

After this work was submitted, Hackett et al. (2008) published what is probably the most comprehensive report on bird evolution since Cracraft (2001). Although the scope of data is impressive, it is largely based on noncoding intron sequences. Nuclear intron data, in combination with current mt and fossil data, have the potential to be extremely useful as long as we can be confident in alignments that span the phylogenetic depth of the avian clade. Some authors have suggested that intron sequences are not appropriate for deeper divergences (Shapiro and Dumbacher 2001) due to alignment ambiguities resulting from multiple insertions and deletions (for general comments on alignments see Löytynoja and Goldman 2008). For example, Morgan-Richards et al. (2008) showed that the alignment of β-fibrinogen intron 7 (which supports the controversial metaves–coronaves split) has no constant sites across the wide taxon sampling required for determining deep avian divergences. Introns are potentially well suited to resolving rapid radiations as they evolve fast enough to accumulate changes during this time (of divergence), while being slow enough to not become random and therefore lose signal (Matthee et al. 2007).

In general, the main avian orders found in both Hackett et al. (2008) and Cracraft (2001) are the same (e.g., Passeriformes, Psittaciformes, and Cuculiformes etc.). However, the relationships among the orders are different, with the deep branches of Neoaves receiving low support (<80% bootstrap support) in Hackett et al. (2008). One difference is their "land birds." Within land birds, Hackett et al. (2008) found Passeriformes (passerines) sister to Psittaciformes (parrots) and suggested a sister relationship between these and Falconidae. These groupings only have support when the intron data are included. Ericson et al. (2006) also inferred this relationship but only when all genes were combined, including β-fibrinogen intron 7 (see their Supplementary Material figs. ESM-1–8). If correct, this placement would be very interesting; however, intron alignment and/or long-branch attraction may be a factor here. The long internal branch to passerines may be attracting the long internal branch of parrots (see Hackett et al. 2008, fig. 3). At this point, support for such a grouping has not been found with mtDNA (e.g., Gibb et al. 2007; Brown et al. 2008). Cracraft (2001) by comparison included morphological, geographical, and early molecular data in support of his groupings. We therefore feel it is appropriate to use Cracraft (2001) as the basis for testing deep divergences within Neoavian birds rather than any one molecular data set.

As a step toward increasing the taxon sampling of coding sequences, we add nine new mt genomes: the dollarbird (*Eurystomus orientalis*) and New Zealand kingfisher (*Halcyon sancta vagans*) as representatives of the Coraciiformes, together with the white-tailed trogon (*Trogon viridis*) from Trogoniformes, are suggested to group with Piciformes; the great potoo (*Nyctibius grandis*) and Australian owlet-nightjar (*Aegotheles cristatus cristatus*) as representatives of Caprimulgiformes; barn owl (*Tyto alba*), expected to pair with the New Zealand owl (morepork, *Ninox novaeseelandiae*) to form Strigiformes; the roadrunner (a ground cuckoo, *Geococcyx california*) expected to pair with the New Zealand long-tailed cuckoo (*Eudynamys taitensis*) to form Cuculiformes; and the peach-faced lovebird (*Agapornis roseicollis*) from Psittaciformes expected to join with the budgerigar (*Melopsittacus undulatus*) and the ground parrot (kakapo, *Strigops habroptilus*). Thus, we have reduced the number of long branches in our data set by the addition of representatives from each of the six Neoavian lineages described in Cracraft (2001). For the third subgroup of the "Conglomerati", only a rail (takahe) and the kagu are published. It is unclear whether these really are a natural group (Morgan-Richards et al.

2008), and this again leaves us with two long isolated branches. In accord with our previous practice, these species have temporarily been omitted until sequences from more closely related species are available for each of them.

We now change from the birds to the analysis. Perhaps, the most fundamental problem occurring while reconstructing deep-level phylogeny is substitution saturation (Curole and Kocher 1999; Phillips et al. 2006). Phylogenetic signal can be eroded by factors including superimposed substitutions and "nonhistorical" biases (such as from compositional nonstationarity)—which accumulate more rapidly at faster evolving sites. Attempts to limit these problems have been made in recent studies by identifying fast-evolving sites at which signal erosion is expected to be high (Morgan-Richards et al. 2008).

In previous work (Delsuc et al. 2003; Phillips and Penny 2003; Phillips et al. 2004), we found standard RY coding (Honeycutt and Adkins 1993), especially the third-codon positions, to be advantageous for the most variable partitions of nucleotide data. This recoding both increases the proportion of changes on internal branches of the tree (i.e., a "treeness" measure) and decreases the differences in nucleotide composition (relative compositional variability). This latter is important in reducing nucleotide composition effects because they have been long known to bias tree reconstruction (Lockhart et al. 1992). Because of the better fit of the data to the model (higher treeness and less variability in nucleotide composition), this has been our preferred method of analysis for vertebrate mt data.

Down-weighting the faster evolving sites or grouping faster evolving nucleotides (or amino acids) into a single category has been quite widely used (see Honeycutt and Adkins 1993; Philippe et al. 2000; Jeffroy et al. 2006)—although the theoretical aspects have not been well developed in phylogenetics. Rodriguez-Ezpeleta et al. (2007) report that omitting the fastest evolving sites, grouping amino acids into functional categories, and some mixture models, all enhanced the phylogenetic signal for deeper divergences. Susko and Roger (2007) similarly report improvements from down-weighting. However, some approaches may not be optimal if valuable sites are excluded simply because they are grouped under some prior definition (e.g., codon positions that have many fast-evolving sites). Conversely, some saturated sites may be retained because they are in a category that, on average, does not have site saturation. Thus, we can also group the justification for down-weighting into those using "a priori" categories (such as third position, stems vs. loops in RNA or amino acid groups) and "conditional" categories (down-weighting of each site independently).

We have used both in the past, the RY coding (a priori weighting category) (Phillips et al. 2004), and also a conditional down-weighting (Penny and Hendy 1986), based on the numbers of observed and expected incompatibilities. In general, all the methods mentioned here are examples of a standard statistical approach of "noise reduction/signal enhancement" (Proakis et al. 2002). Here, we implement a conditional noise reduction technique in which the information retained from the sequence is determined on a site-by-site basis. The Materials and Methods section has more detail on this conditional recoding (down-weighting) of sites, an approach that we call site-stripping (Morgan-Richards et al. 2008).

Along with data partitioning/down-weighting and fossil calibrations (see Supplementary Material S1 online), our additional sequences mean that we can, in principle, calculate the probabilities that prior hypotheses are supported. In other words, we can calculate the proportion of trees that will have a split (or clade) that has been predicted. Unfortunately, there appears to be little use in phylogenetics for specifying a priori hypotheses and then testing the probability of finding them with new data. Rather, results are treated somewhat "post hoc," looking at the trees after they are built and then trying to explain the results. In principle, a Bayesian approach allows alternative hypotheses to be given different weightings, but it appears that a "flat prior" is the norm; and this does not really differentiate between trees or hypotheses. Based on nuclear-coding sequences, Lin et al. (2002) took the four-way split within eutherian mammals and calculated the probability of finding the same split from mt data. The result was certainly very highly significant; $P \approx 2.1 \times 10^{-7}$. It is very important in phylogeny, though perhaps seldom carried out, to give quantitative estimates of the increase in information from a phylogeny. Penny et al. (1991) demonstrated that it is simple to calculate probabilities (see Appendix and table 1) of a single pair of taxa predicted to come together on the tree. That is, there is one chance in $2n - 5$ of them coming together "by chance" on an unrooted tree and one in $2n - 3$ for a rooted tree (given n taxa). It is even more improbable that a predefined grouping of three or more taxa will come together, and here, we develop measures in order to evaluate quantitatively the priors from Cracraft (2001).

Our approach here is to infer evolutionary trees from complete mt genomes as a start on resolving the deep Neoavian splits. We investigate the robustness of the priors set out by Cracraft (2001) by using a novel site-stripping method and exploring networks (see Supplementary Material S1 online for fossil calibrations). In addition, we assess quantitatively the usefulness of these priors by calculating clade probabilities. By finding resolution in the basal node of the Neoaves, hypotheses regarding the number of lineages present before the K–Pg boundary can be tested.

Materials and Methods

Taxon Sampling

The dollarbird (*E. orientalis*) and barn owl (*T. alba*) were supplied from the Australian Museum, Sydney, Australia, under sample numbers EBU 11118 and EBU 2564, respectively. The roadrunner (*G. california*), great potoo (*N. grandis*), and the white-tailed trogon (*T. viridis*) were provided by the Louisiana State University Museum of Natural Science Collection of Genetic Resources under sample numbers LSUMZ B-8504, LSUMZ B-8954, and LSUMZ B-28495. The Australian owlet-nightjar (*A. c. cristatus*) was provided by Fritz Geiser, University of New England, Armidale, Australia. The New Zealand long-tailed cuckoo (*E. taitensis*) and the New Zealand kingfisher (*H. s. vagans*) were obtained from Dick Gill, NZ Department of Conservation, Waikanae, New Zealand.

Table 1
Probabilities of a Predefined Clade Joining the Tree as a Single Group or as Two Subclades

k Taxa in clade	Unrooted Trees Minimal	Rooted Trees Minimal
2	$1/(2n-5)$	$1/(2n-3)$
3	$3/(2n-5)(2n-7)$	$3/(2n-3)(2n-5)$
4	$15/(2n-5)(2n-7)(2n-9)$	$15/(2n-3)(2n-5)(2n-7)$
5	$105/(2n-5)(2n-7)(2n-9)(2n-11)$	$105/(2n-3)(2n-5)(2n-7)(2n-9)$

The peach-faced lovebird (*A. roseicollis*) was obtained locally from commercial breeders.

Molecular Methods

Extractions of genomic DNA from each of the newly sampled birds were performed at the Allan Wilson Centre from 25 to 50 mg of liver tissue using the High Pure PCR Template Preparation Kit (Roche Applied Science, Mannheim, Germany) according to the manufacturer's instructions. To minimize the chance of obtaining nuclear copies of mt genes (NUMTs), 2–4 overlapping long-range polymerase chain reaction (PCR) fragments (3.5–12 kb in length) were first amplified using the Expand Long template PCR System (Roche Applied Science). The products were excised from 1% agarose gels and purified using a QIAquick Gel extraction kit (Qiagen GmbH, Hilden, Germany) as per the manufacturer's instructions. These long-range products were subsequently used as template DNA for following short-range PCRs (overlapping fragments 0.5–3 kb in length). Short-range primer combinations were found using our laboratory database as described in Slack et al. (2006), and any new primers required were designed using Oligo 4.03 (National Biosciences, Inc., Plymouth, MN). Sequencing was performed using BigDye Terminator Cycle Sequencing reagents according to the manufacturer's instructions (Applied Biosystems, Foster City, CA) and then sequenced on an ABI 3730 automated sequencer (Applied Biosystems). Sequences were aligned using Sequencher 4.7 (Gene Codes Corp., Ann Arbor, MI) and then manually edited and checked for complete concurrence between overlapping sequences.

Where necessary (e.g., with length heteroplasmy in control regions [CRs] from microsatellite repeats), PCR products were cloned using the TOPO TA cloning kit for sequencing (Invitrogen, Carlsbad, CA). For each region, at least three clones were sequenced to safeguard against PCR errors. In all cases, overlaps between sequences were sufficient to ensure synonymy and sequence identity was confirmed through Blast searches (http://www.ncbi.nlm.nih.gov/blast/), confirmation of amino acid translation in coding regions, and alignment with other species.

In addition to the nine new bird mt genomes reported in this paper, 36 other complete avian mt genomes from NCBI GenBank were included in the analyses: 31 Neoaves and 5 Galloanserae. Paleognath taxa were not included in this data set; although their overall placement is now well established (Gibb et al. 2007; Slack et al. 2007), there are still important but unresolved issues around the placement of tinamous (Hackett et al. 2008; Harshmann et al. 2008; Phillips MJ, Gibb GC, Crimp EA, Penny D, in preparation). Instead, we rooted our Neoaves trees with the Galloanserae sequences (Gibb et al. 2007; Morgan-Richards et al. 2008). The full data set is available from the authors on request.

The Galloanserae taxa are Japanese quail (*Coturnix japonica*, AP003195), magpie goose (*Anseranas semipalmata*, AY309455), redhead duck (*Aythya americana*, AF090337), greater white-fronted goose (*Anser albifrons*, AF363031), and Australian brush turkey (*Alectura lathami*, AY346091). The 31 Neoaves taxa (modern birds) are rifleman (New Zealand wren, *Acanthisitta chloris*, AY325307), gray-headed broadbill (*Smithornis sharpei*, AF090340), fuscous flycatcher (*Cnemotriccus fuscatus*, AY596278), superb lyre bird (*Menura novaehollandiae*, AY542313), rook (*Corvus frugilegus*, Y18522), ivory billed toucan (*Pteroglossus azara*, DQ780882), pileated woodpecker (*Dryocopus pileatus*, DQ780879), morepork (a New Zealand owl, *N. novaeseelandiae*, AY309457), kakapo (flightless parrot *S. habroptilus*, AY309456), budgerigar (*M. undulatus*, EF450826), ruby-throated hummingbird (*Archilochus colubris*, EF532935), common swift (*Apus apus*, AM237310), peregrine falcon (*Falco peregrinus*, AF090338), forest falcon (*Micrastur gilvicollis*, DQ780881), Eurasian buzzard (*Buteo buteo*, AF380305), osprey (*Pandion haliaetus*, DQ780884), Blyth's hawk eagle (*Spizaetus alboniger*, AP008239), blackish oystercatcher (*Haematopus ater*, AY074886), ruddy turnstone (*Arenaria interpres*, AY074885), southern black-backed gull (*Larus dominicanus*, AY293619), red-throated loon (*Gavia stellata*, AY293618), little blue penguin (*Eudyptula minor*, AF362763), rockhopper penguin (*Eudyptes chrysocome*, AP009189), black-browed albatross (*Diomedea melanophris*, AY158677), Kerguelen petrel (*Pterodroma brevirostris*, AY158678), frigatebird (*Fregate* sp., AP009192), Australian pelican (*Pelecanus conspicillatus*, DQ780883), Australasian little grebe (*Tachybaptus novaehollandiae*, EF532936), greater flamingo (*Phoenicopterus ruber roseus*, EF532932), great crested grebe (*Podiceps cristatus*, AP009194), and the Oriental white stork (*Ciconia boyciana*, AB026193).

Phylogenetic Analysis

Sequences were aligned in Se-Al v2.0a11 at the amino acid level for protein-coding genes and based on secondary structure for RNA genes. The data set has 12 protein-coding genes, 2 ribosomal RNAs (rRNA), and 22 transfer RNAs (tRNA). Gaps, ambiguous sites adjacent to gaps, NADH6 (light-strand encoded), and stop codons (often incomplete in the DNA sequence) were excluded from the alignment. The 12 protein-coding genes were separated into first-, second-, and third-codon positions, whereas rRNA and tRNA genes were partitioned into stems (S) and loops (L).

Protein-coding genes were checked for NUMTs by translating into amino acids.

435 Previous studies from birds (Slack et al. 2003), mammals (Lin et al. 2002), and simulations (Holland et al. 2003) have all shown that the addition of outgroups can disrupt the ingroup tree. However, in such cases (from theory and with simulated data), the ingroup tree (i.e., with the outgroup omitted) is more likely to be correct (Holland et al. 440 2003). We therefore ran separate analyses either including or excluding the outgroup (five birds from the Galloanserae). A combined total of 13,412 nucleotides (excluding gaps) were used for the basis of further analyses (see below 445 for number of characters per data set). As mentioned earlier, we partitioned the data: codons 1 and 2, codon 3, RNA stems, and RNA loops (Slack et al. 2007) for site-stripping. ML analyses were carried out using standard programs including PAUP* 4.0b10 (Swofford 2001) and GARLI v0.95 450 (Zwickl 2006). Bayesian analysis was carried out in MrBayes (Huelsenbeck and Ronquist 2001), and consensus networks were implemented in SplitsTree version 4 (Holland et al. 2004; Huson and Bryant 2006) and BEAST (Drummond et al. 2006; Drummond and Rambaut 2007). 455 Optimal parameters for the ML models were determined using Modeltest 3.7 (Posada and Crandall 1998) and the AIC values used. The hierarchical and AIC tests were in agreement for Modeltest. Initial results from ML analyses were consistent with Bayesian analysis and were not used 460 for site-stripping. The best model for the full data set was GTR + I + G. Bayesian analyses used default parameters and were run for 10 million generations or until convergence was obtained. In addition to Bayesian posterior probabilities (BPPs), we ran analyses in PhyML (Guindon 465 and Gascuel 2003) and RAxML (Stamatakis 2006) to carry out 100 bootstrap replicates on the data sets both with and without the outgroup (see Supplementary Material S2 online for results).

Noise Reduction by Down-weighting (Site-Stripping)

470 Site-stripping compares sites based on the actual number of mutations required on the tree ("tree steps") versus the maximum possible number of mutations for that site (max). The calculation is the limit (L) = tree steps × tree steps/max steps. If the threshold "strictness" (s) is, for ex- 475 ample, 4, then sites for which $L \leq 4$ remain unchanged and sites for which $L > 4$ are RY coded. If after RY coding L is still > 4, the site is excluded. The higher the threshold (the larger the value of s), the more sites that are included and fewer are RY coded. Conversely, the lower the threshold 480 (lower values of s), the more sites RY coded or excluded. Therefore, the weighting of each site is a function of that site and is not predetermined by being a member of a class. This allows, for example, some hypervariable sites from first and second positions to be RY coded or omitted. A 485 range of s values were used and resulted in the following:

$s = \infty$, 0 sites RY coded, 0 sites excluded (all 13,412 sites included);
$s = 6$, 20 sites RY coded, 0 sites excluded;
$s = 4.5$, 91 sites RY coded, 1 site excluded;
490 $s = 3.5$, 254 sites RY coded, 8 sites excluded;
$s = 2.28$, 578 sites RY coded, 63 sites excluded; and
$s = 2$, 891 sites RY coded, 159 sites excluded.

Bayesian inference analyses were carried out on each of the data matrices, including the fully weighted data. Note that RY coding increases the ML scores as it amalgamates 495 some nucleotide categories; thus, the data are now different, and it is not valid to compare directly ML scores from RY and nucleotide coding (Steel MA, personal communication). Similarly, consistency index (CI) values are not directly comparable between nucleotide and RY coded 500 data sets. That is, for an unrooted four-taxon tree, for which A, C, T, and G states are random (i.e., no signal remains), the expected average CI for nucleotide data is 0.949, whereas for the RY coded data, it is 0.778.

In order to identify fast-evolving sites and so facilitate 505 noise reduction by site-stripping, nine additional close relatives were added to the alignment but were removed before phylogenetic analysis as they do not break up long branches or add phylogenetic signal and would only increase analysis times. The nine taxa included were chicken (*Gallus gallus*, 510 AP003317); green junglefowl (*Gallus varius*, AP003324); gray junglefowl (*Gallus sonneratii*, AP006741); white stork (*Ciconia ciconia*, AB026818); Canadian goose (*Branta canadensis*, DQ019124); tundra swan (*Cygnus columbianus*, DQ083161); mountain hawk eagle (*Nisaetus nipalensis*, 515 AP008238); Pacific loon (*Gavia pacifica*, AP009190); and American kestrel (*Falco sparverius*, DQ780880) (see Supplementary Material S3 online for the consensus tree of the above nine taxa and their close relatives).

Probabilities of Observing Predefined Clades 520

It is straightforward to calculate the probability of observing a prespecified clade in a tree from a new data set, that is, a tree using additional data not used to predict the clade (see Appendix A for details). In general, for n taxa, there are $B(n) = (2n - 5)!!$ unrooted binary trees, where the double factorial 525 notation (!!) is the product of every second number, that is, $1 \times 3 \times 5 \times \ldots \times 2n - 5$ (table 1). Thus, for example, there are $\approx 3 \times 10^{20}$ possible unrooted binary trees for 20 taxa. In addition, even if a predicted clade for three or more taxa ends up as two subclades on the tree, we can calculate the probability of ob- 530 serving this; and it can indicate that there is still high information content in the data. The calculation can be extended further to three or more subclades, but here, we concentrate mainly on our prior clades being found. We need to define the composition of the prior clades carefully; they alter with 535 the question being considered. For example, we may be interested whether the three parrots do form a natural group, in which case we would calculate the probabilities of three parrots coming together with $n = 40$ taxa in the data set—this is close to a flat prior. Alternatively, we may accept the grouping 540 within orders and consider just $n = 21$ deep Neoavian groups (see later, fig. 2); this is testing the groupings ("priors") of Cracraft (2001). Other tests are also possible.

Results

The first of our three approaches to improving the 545 Neoavian tree was the inclusion of more sequences. The

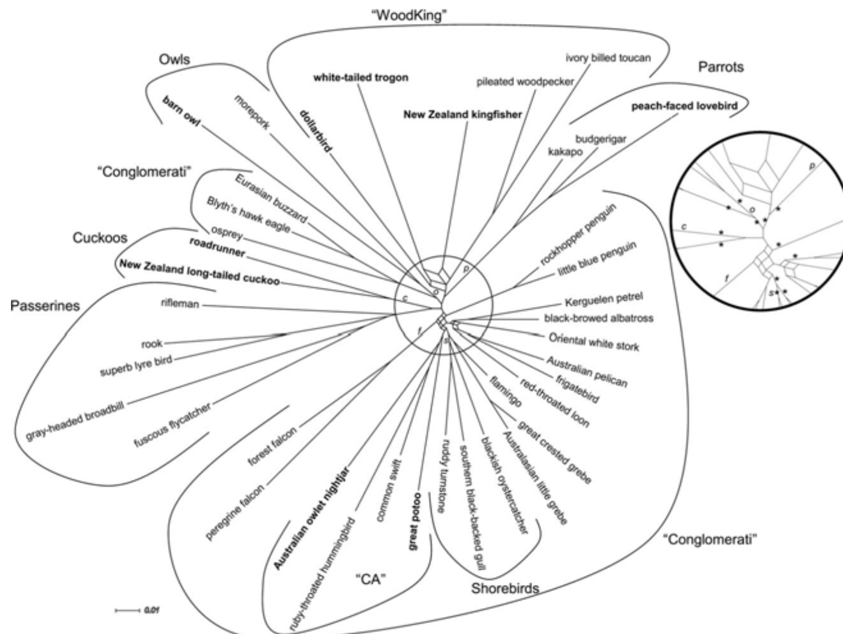


FIG. 1.—Unrooted Bayesian consensus network of Neoaves (modern birds), based on whole mtDNA genomes with "minimal" down-weighting (threshold $s = 6.0$; 20 sites RY coded, 0 sites excluded). Only splits occurring in >25% of trees are included in the network. Inset shows the central portion (indicated by the circle) expanded. Selected branches are labeled to ease comparison: c, cuckoos; f, falcons; o, owls; p, parrots; and s, shorebirds. New taxa included are highlighted in bold. Splits indicated by asterisk have 99+ Bayesian posterior support (BPP).

nine new mt genome sequences are deposited in GenBank under the following accession numbers: dollarbird (EU344978, 17,774 bp); barn owl (EU410491, >16,148 bp, incomplete ND6 and CR); roadrunner (EU410488, 17,091 bp); great potoo (EU344977, >14,396 bp, incomplete ND6 and CR); white-tailed trogon (EU410490, 17,751 bp); Australian owlet-nightjar (EU344979, 18,607 bp); peach-faced lovebird (EU410486, 16,732 bp); New Zealand long-tailed cuckoo (EU410487, 17,559 bp); and the New Zealand kingfisher (EU410489, 17,549 bp). Following the gene-order nomenclature from Gibb et al. (2007), the cuckoo and roadrunner were found to have the remnant CR(2) gene order first described in the falcon (Mindell et al. 1998). All other birds have the standard avian gene order as in the chicken (Desjardins and Morais 1990) with the possible exception of the great potoo and barn owl; because their sequences are currently incomplete in the CR and adjacent genes, their gene order is unknown.

Site-stripping (noise reduction) was our second approach to improving the tree. Bayesian analyses were carried out on a range of down-weighting values, initially excluding the outgroup, then including it. Figure 1 shows the result for the unrooted Neoavian data set with a high threshold (minimum down-weighting, strictness $s = 6.0$), whereas figure 2 shows results from the maximum down-weighting used (a low threshold, strictness $s = 2.0$). Both figures are networks showing splits occurring in at least 25% of Bayesian phylogenies (Holland et al. 2004).

Our third approach was a quantitative test of predefined groupings. There are several clades that were predicted and were returned both with different down-weightings and with or without the outgroup. Perhaps, the most straightforward example is that of the three parrots; this is just a trial calculation because we really had no doubt that the parrots would come together, as predicted from previous DNA sequence analyses (de Kloet RS and de Kloet SR 2005). The New Zealand ground parrot (kakapo) was the first to diverge. This is just a trial calculation with flat priors, and the probability of three taxa coming together on a tree is $P = 0.0005$ ($P = 3/(2n - 5)(2n - 7)$, with $n = 40$ (see Appendix A). There is considerable rate variation within parrots, with the kakapo being slower than the others, whereas the peach-faced lovebird is the

fastest (which is evident by the long edge in both figs. 1 and 2).

Again as expected from our informal priors, the New Zealand long-tailed cuckoo and the roadrunner (a ground cuckoo) always paired and do in prior morphological and molecular studies; thus, we considered them to form one independent "taxonomic group." Of the 40 taxa analyzed, we consider only 21 groups to be independent (see later and Appendix A for details of the groupings); consequently, the probability of two taxa forming a clade in the tree is $P = 0.027$ ($P = 1/(2n - 5)$, with $n = 21$ see Appendix A). Perhaps unexpectedly, cuckoos then group as sister to the five passerines in our data set, a result also observed by Mayr et al. (2003) using combined molecular and morphological data; however, bootstrap support was relatively low (see their figs. 5 and 7). As in previous analyses, the passerines always group together, with the New Zealand wrens (rifleman in this case), basal to the oscines and suboscines. The cuckoo–passerine pairing was found with all down-weightings, both with the ingroup alone and with the outgroup included (fig. 3). This grouping of the cuckoo–passerine clades is an interesting hypothesis and requires testing with both nuclear-coding sequences and rare genomic changes (Boore 2006).

The first real test of Cracraft's (2001) priors, that is, testing groups above the order level, stems from the clade we refer to as "WoodKing". In this case, all five taxa were always found as a clade, irrespective of the down-weighting, and with or without the outgroup (i.e., with both unrooted and rooted trees). As expected, the Piciformes (pileated woodpecker and ivory billed toucan) were always paired. Thus, if it is assumed that the woodpecker and toucan are sufficiently close, then really there are only four independent taxa, and $P = 7.12 \times 10^{-6}$ is a highly significant result (see Appendix A for details). Note that the calculation of the probabilities allows all possible ways of observing the four taxa on the tree, including any paraphyletic groups within it. As shown in figures 1 and 2, there is conflicting signal linking the kingfisher with either the white-tailed trogon or with the dollarbird. Being able to show both signals is a major advantage of networks (Holland et al. 2004) because it helps to prevent premature conclusions. With increased down-weighting of the faster sites (a stricter threshold), we again observed variation in the position of the dollarbird, which tended to be deeper in the clade, though still within the "WoodKing" group. Additional taxon sampling should resolve the splits fully, but our main conclusion is that the predicted grouping of Piciformes, Trogoniformes, and Coraciiformes (Cracraft 2001) is found (though not necessarily reciprocally monophyletic). Hackett et al. (2008) sampled more widely and found strong support for Coraciiformes + Piciformes (see their fig. 2, clade C), although Trogoniformes fell outside this and had less support (see our Supplementary Material S2 online).

However, the next result was not in our informal priors. We found that the parrots and the WoodKing group, irrespective of down-weighting extent, are always adjacent clades on the unrooted tree (supported by high BPP values but not by bootstrapping, see Supplementary Material S2 online). This result will need further investigation as the parrot lineage has considerable rate variation and there is a long internal branch from the three parrots to the rest of the tree. Because the grouping was not part of our priors, we cannot calculate the increase in support, but if additional data types support this relationship, then the probabilities could be calculated.

Next, we consider the "SCA" group (Strigiformes, Caprimulgiformes, and Apodiformes) predicted by Cracraft (2001). Four of the six taxa available form a monophyletic clade, to the exclusion of the two owls. The great potoo, Australian owlet-nightjar, common swift, and ruby-throated hummingbird formed a group of four ("CA"). The swift and hummingbird pairing was highly supported by BPP and bootstrapping (see Supplementary Material S2 online) as predicted from previous studies (Johansson et al. 2001; van Tuinen and Hedges 2001; Cracraft et al. 2004; Hackett et al. 2008; Morgan-Richards et al. 2008), the Australian owlet-nightjar came deeper, and finally the great potoo, which was always basal. It should be noted that the potoo did move slightly with bootstrapping, and we found more support for the Apodiformes and the owlet-nightjar to the exclusion of the potoo (see Supplementary Material S2 online) (see also Mayr 2002a; Barrowclough et al. 2006). Hackett et al. (2008) also observed high support (98%) for Apodiformes + *Aegotheles*, as have previous nuclear and morphological studies (see also Mayr 2002b; Barrowclough et al. 2006). The position of this group of four was variable in the tree. With lesser down-weighting (a higher threshold, $s = 6.0$), the group was found within the informal "Conglomerati" group (fig. 1), but with the maximum down-weighting, it was outside this group (apart from the shorebirds). Similarly, the barn owl and the morepork always paired, joining together quite deep in the tree (i.e., although both are "owls," they represent old divergences; indeed some short preliminary runs did not even join them together, see Supplementary Material S2 online for support values). We need to be cautious here as both owls have some of the highest rates of sequence evolution among the Neoaves. Although the owls did not group with the other members of the "SCA" clade, they were always found to group with the buzzard, hawk eagle (Accipitridae), and osprey. Only with the highest down-weighting used (figs. 2 and 3) did the falcons unite with the owls/Accipitridae/osprey; with the falcons basal, however, we found no bootstrap support and low BPP support for this grouping (see Supplementary Material S2 online). Grouping of the Accipitridae, osprey, and owls is certainly interesting. Relationships within the birds of prey are controversial as speculation over convergence and raptorial specialization has been raised (e.g., Livezey and Zusi 2007). This relationship was not found by Hackett et al. (2008) and needs to be tested with nuclear-coding data. We return to the raptors and owl question later.

Even though the proposed "SCA" clade came out as two groups in the tree, this in itself still has high information content. The simplest calculation is to assume the alternative prediction of the two owls being independent of the other four, and in this case, the probability of observing the "two plus four grouping is simply the product of the probability of observing a pair ($P = 1/(2n - 5) = 0.013$ for $n = 40$) and the probability of observing a group of four ($P = 2.5 \times 10^{-6}$, see Appendix). This gives the combined

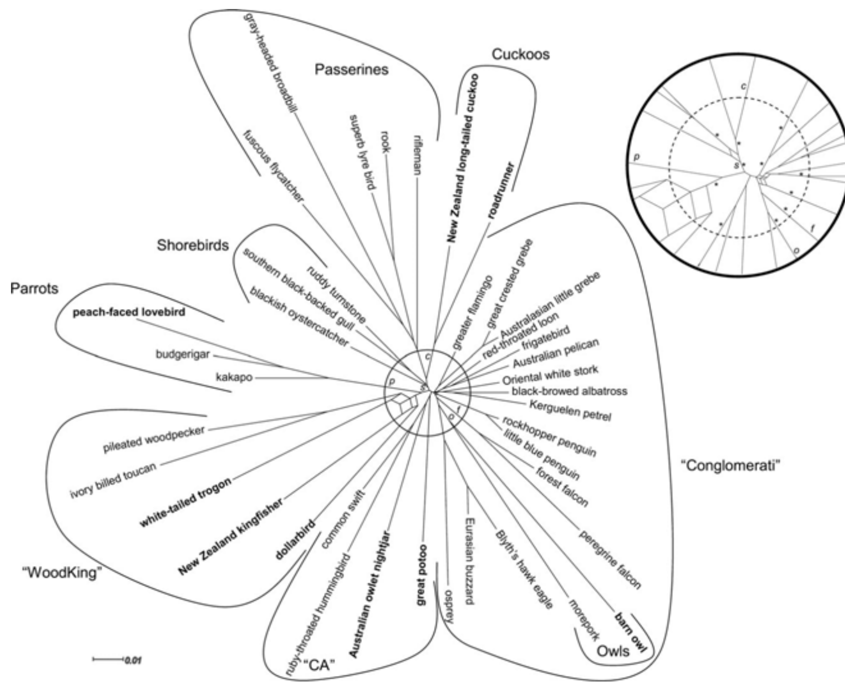


FIG. 2.—Unrooted Bayesian consensus network of Neoaves with the "maximum" down-weighting used (threshold $s = 2.0$; 891 sites RY coded, 159 sites excluded). Otherwise, the conventions are the same as in figure 1.

probability of 3.4×10^{-8} (about 1 chance in 300 million of observing the pattern). Strictly speaking, we need not make the assumption that the two owls (in particular) separate from the other four taxa; there are a total of 21 combinations ($^6C_2$) of pairs from six taxa; only one has the two owls paired. In general, we would multiply the probability by 21 (see Appendix A), still giving around 1 chance in 10 million of observing these two groups on the tree. Similarly, it could have been just a single taxon (six choices) that separated from the other five or two groups of three taxa (10, or $^6C_3$ combinations) but halved because each triplet of taxa is found twice. Appendix A and figure 4 show the general calculation, but in this case (because the owls join with the raptors, as an alternative prediction see Mayr 2005), it is reasonable to use the probability of finding just the two plus four grouping.

Our conclusion at this point is that excellent progress is being made in understanding the deeper levels of phylogeny of the Neoaves. If the predictions from Cracraft (2001) are, in general, being well supported with new data, then this implies that the basal polytomy is resolvable. However, our next step is to check that there are no major changes when the outgroup is added; this has been a major problem when a smaller number (24 ingroup) of taxa were sequenced (see Harrison et al. 2004).

Rooted Tree

Figure 3 shows our tree rooted with five Galloanserae taxa for the maximum down-weighting value (threshold $s = 2$). The two main points from this figure are that

1. there is only one local change to the unrooted tree when the root is added, and
2. the root comes between parrots and all other Neoaves.

The first point refers to the difference in the position of the flamingo/grebe clade between the unrooted and rooted trees. In the unrooted tree, they are basal to the loon/albatross/Pelecaniformes/stork grouping. In the rooted tree, we find them basal to the same group as before plus the Accipitridae/osprey/owl/falcon clade. In other words, the tree is "locally stable" in the terminology of Cooper and Penny (1997). Finding just a single edge (branch) different between two trees is very highly significant, about 5.7×10^{-54} for 40 taxa in the ingroup (Penny et al. 1982).

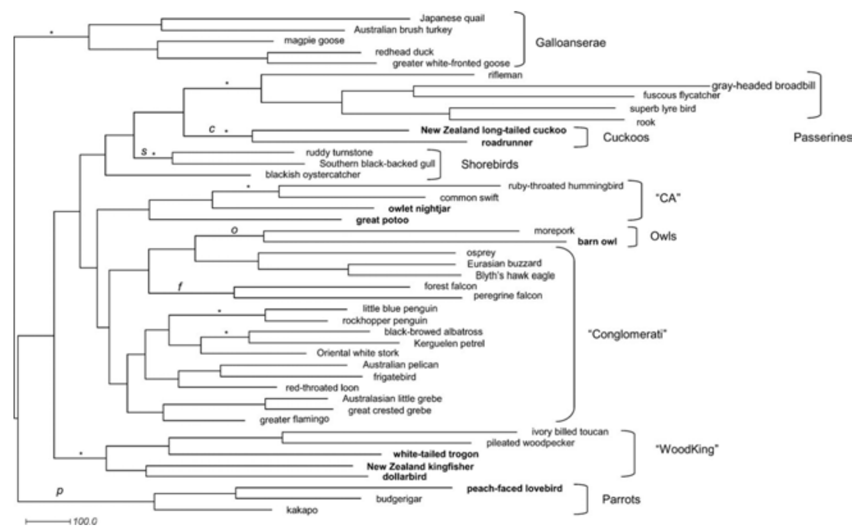


FIG. 3.—Rooted Bayesian phylogram of Neoaves with maximum down-weighting and including the five Galloanserae as the outgroup. Conventions are the same as in figure 1.

On the surface, this latter finding (the root between parrots and other Neoaves) could be suspicious because the branch at the base of the parrots is the longest internal branch on the tree! However, with lesser down-weighting ($s = 6.0$), the root joins one step away on the short branch at the base of the parrot/"WoodKing" group. The movement of the root from a shorter branch is not expected unless parrots are really the first subdivision of Neoaves. Without any prior information as to the root of the Neoaves, it is not possible to give any quantitative statement of confidence in this rooting.

Other aspects of the rooted tree are also interesting and noteworthy. We again find the same owl/raptor clade appearing only with the strongest down-weighting—that is, only the Accipitridae/osprey (but not falcons) unite with the owls with lesser down-weighting. The tree produced from BEAST resulted in the falcons joining the "CA" group (Supplementary Material S1 online). This latter observation, if real, is interesting as both falcons and the "CA" taxa are in-flight foragers typically specialized for bill-capture of prey; in comparison, the owls are specialized for inflight talon capture (like hawks). However, there was only conflicting support for this relationship shown in the network figure 1, and therefore, we are unable to comment further without additional data. It is worth noting that some of the deeper groupings within this large raptor/shore bird/water bird ("Conglomerati") still vary somewhat depending on taxon sampling (Morgan-Richards et al. 2008), and it is not clear yet whether further taxon sampling will resolve these issues. It may well be that the three subgroups in the "Conglomerati" (group (vi) of Cracraft 2001, see Introduction) should be considered independently.

Discussion

Resolving the evolutionary relationships within the modern birds (Neoaves) has been both problematic and controversial, with some suggesting that it will never be fully resolved (Poe and Chubb 2004). Here, we have shown that with more and longer DNA-coding sequences, along with improved noise reduction techniques, relationships within Neoaves are expected to be resolvable. This should occur relatively quickly with the addition of data from nuclear coding and rare genomic changes as they become available. Our approach to resolving this issue has been 3-fold: presenting additional data, improving noise reduction/signal enhancement techniques, and getting beyond flat priors, where it is assumed (sometimes correctly!) that there is no useful prior knowledge.

We consider it important that prior hypotheses can be evaluated quantitatively, and thus, the formulae developed in Appendix A will be useful for a wide range of studies. However, there is still more work required in developing these analyses. For example, the calculation for prespecified groups is for the optimal placement of that clade on the tree (even if the bootstrap or Bayesian posterior priors are less than 100% support). If these support values are indeed higher, then this gives even more confidence in the clades, so in that respect our probabilities are conservative. Thus, more thought is required on how to combine the

calculations developed here with the strength of support for branches in the tree from new data. Similarly, the calculation allows for any subtree within the clades (or subclades). However, if we prespecified that a particular grouping and subtree is expected, then the probability of finding this arrangement on the tree is even lower. For example, the calculation allows 15 ways (5!!) that a group of four could join the larger tree. This is because there are three unrooted trees for four taxa, each with five edges for joining to the rest of the tree—and the prediction does not specify which of the 3 × 5 (15) trees would be observed. In contrast, if we predict precisely how the group will join (forming a clade), then the number of possible trees is reduced 15-fold. Overall, it is important that we make better use of well-considered prior hypotheses when studying trees based on new data. Even though we are a long way from having the "one tree" (in this case) for Neoaves, we can be confident that the issue is resolving and that the data sets have, in a formal sense, high information content.

A related question is estimating how many trees within Neoaves are still likely—a "confidence set" of trees. At this higher taxonomic level, it is not yet clear which groupings are stable and which may be subject to change. For our Neoaves data set, there appears to be around 21 major groupings (shown as crossing into the inner dashed line in the insert of fig. 2 and in Appendix A). These groupings include Passeriformes, cuckoos, parrots, two shore bird lineages, three raptor clades (falcons, buzzard/osprey, and owls), rollers, kingfishers, woodpecker/toucan/trogon, potoos, owlet-nightjars, hummingbirds/swifts, flamingoes, grebes, Pelecaniformes, tubenoses, storks, penguins, and loon.

In principle, there are $R(21) \approx 3 \times 10^{23}$ possible rooted binary trees of which only a vanishingly small proportion are realistic. With eutherian mammals, there were initially 19 orders identified and therefore $R(19) \approx 2 \times 10^{20}$ possible rooted binary trees. But it quickly became apparent that no more than about 10^2 trees were likely (Lin et al. 2002)—an improvement of 18 orders of magnitude. The next step for birds is an equivalent analysis for Neoaves and thus getting beyond the debilitating view of the flat priors—that all trees were equally likely and that there is no information in previous studies.

The results from down-weighting the faster evolving sites are interesting, and these techniques need to be developed and tested further. With increased down-weighting, we find closer agreement between earlier predictions and the actual tree found. From first principles, we expect that reducing the influence of the saturated sites will help, and in general, it appears that the predefined groups are found more strongly. It is for this type of reason that we would like to see further development and evaluation of the noise reduction techniques including their application to nuclear-coding data. Although it is outside the range of this study, a simulation study is now an important next step.

Turning now from the more general issues to the Neoaves in particular, our current study assesses the stability and probability of the six groups proposed by Cracraft (2001) using a novel analysis method to down-weight sites of whole mt genome sequences. These Cracraftian priors were found to be robust with four of the six groups within Neoaves being recovered, the other two having relatively small changes; the owls moving to the raptors and possibility that the raptor/water carnivores ("Conglomerati") may be diphyletic. In this latter respect, it appears preferable, at least in the short term, to treat the three subgroups of Cracraft's group (vi) independently. Our resulting phylogenies appear relatively stable, differing little in overall topology both with and without the addition of the outgroup (Galloanserae).

The placement of the root of Neoaves needs additional support. Our analyses put the root in one of two possible locations: either with parrots (which have a higher mutation rate in mtDNA) as the most basal lineage or, with lesser down-weighting, the parrots plus the "WoodKing" grouping as the basal clade. Because the separation of parrots from the rest of the Neoaves occurred with the highest down-weighting, we cannot easily dismiss this possible rooting. Morphologically, parrots are distinct (for review, see Dyke and Cooper 2000; Waterhouse 2006) and a fragment of a mandible from the Maastrichtian, latest Cretaceous (65–70 Ma Lance Formation, North America) has been described (Stidham 1998). However, the identification of this fossil is contentious (Dyke and Mayr 1999; Mayr 2002a), though previous molecular work suggests a Cretaceous diversification for each of the African, Australian, and South American parrots (Miyaki et al. 1998). Dating carried out on our current data set suggests that the most basal parrot in our analysis, the kakapo, split from the other parrots sometime after the K–Pg boundary. The lineage as a whole however predates the K–Pg boundary with a mean date of ~85 Ma (see Supplementary Material S1 online). Hackett et al. (2008) suggested that the root of Neoaves be placed with the sister grouping of the Podicipediformes/Phoenicopteriformes/Phaethontidae/Pteroclididae/Mesitornithidae/Columbiformes to Gruiformes/Caprimulgiformes/Apodiformes (see their fig. 2, clades K, L, M, and N). However, they do not give support values for this lineage, and they state that their rooted tree only occurs when the β-fibrinogen intron 7 data are included. We have already demonstrated (Morgan-Richards et al. 2008) that by our more rigorous standards (deleting columns around gaps back to a constant column) that the intron sequences of this locus are not informative for deep divergences. Clearly, the root of Neoaves is still under debate; however, we now have a number of possibilities to be tested by future analyses.

Conclusion

The basal split within Neoaves and its timing are resolvable issues. If modern birds radiated over a short period (say 2–5 Ma) after the K–Pg extinction, then it will be very difficult to resolve the polytomy at the base of modern birds. However, using whole mtDNA coding sequences gives us a solid point from which to build. By the addition of more taxa, nuclear-coding sequences, and rare genomic changes, we expect resolution at the ordinal level to be achievable. In addition, the further development of noise reduction techniques for coding sequences (both organellar and nuclear) will enable more robust trees to be produced. We estimate that at least 12 Neoavian lineages had evolved prior to the K–Pg boundary, similarly, van Tuinen et al. (2006) and Brown et al. (2008) support pre-K–Pg origins

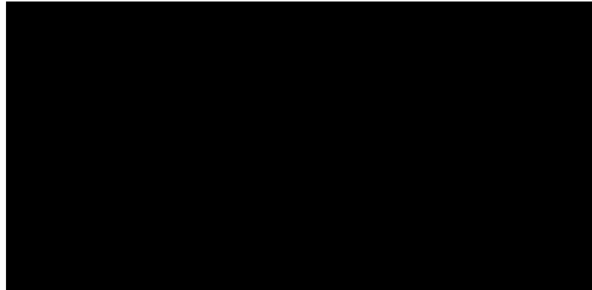


FIG. 4.—The basis for calculations for the probability of finding predefined clades (or subclades) on a tree. (*A*) A procedure for counting the number of trees. There is only one unrooted tree for three taxa, and there are three edges (branches) to add the fourth taxon—giving three trees for four taxa. Each of these three trees has five edges where the fifth taxon can be added giving $1 \times 3 \times 5 = 15$ trees. Similarly, each of these 15 trees has 7 edges for the sixth taxon to be added, leading to the formula $B(n) = (2n - 5)!!$, for the number of unrooted binary trees. (*B*) Calculating the probability of prespecified clade of k taxa on a new tree. There are $R(k)$ rooted trees for the clade of k taxa and $B(n - k)$ for the remaining $n - k$ taxa, leading to the calculation for the probability of observing a prespecified clade of k taxa forming a clade in each tree with n taxa. (*C*) The proposed grouping of k taxa with two subgroups; a single taxon in one and $k - 1$ in the other. (*D*) A similar case with two taxa on one group and $k - 2$ in the other. There are ${}^{k}C_2$ (k choose 2) ways of selecting the two taxa. (*E*) An example where the group of $k = 6$ ends up as $m = 3$ subgroups of 2 taxa each. There are ${}^{6}C_2$ for selecting the first pair of taxa and ${}^{4}C_2$ for the second pair and 3! ways for ordering the three pairs on a given tree.

for multiple modern lineages. In addition, Clarke et al. (2004) estimate a minimum of five Anseriformes lineages (duck, chicken, and ratite bird relatives) before this time supporting the presence of a diverse array of modern bird lineages prior to the extinction event. Lastly, with regards to the search for the one tree, we feel our data have made significant progress with support for four of the six Cracraftian groups. Given the very low probability of observing groupings by chance, the data are highly informative and should stimulate future work incorporating data from all facets of avian evolution.

Supplementary Material

Supplementary Material S1 (fossil calibrations and molecular dating with BEAST), Supplementary Material S2 (tabulated results from Bayesian analysis [BPP] and bootstrap analyses [bs] using RAxML and PhyML), and Supplementary Material S3 (consensus tree including all outgroup taxa used for alignment before for site-stripping) are available at *Molecular Biology and Evolution* online (http://www.mbe.oxforjournals.org/).

Acknowledgments

We thank the Australian Museum, Sydney; the Louisiana State University Museum of Natural Science Collection of Genetic Resources; Fritz Geiser, University of New England, Armidale, Australia; and Dick Gill, Department of Conservation, Waikanae, for supplying tissue samples. Thanks to Trish McLenachan for help with alignments and sequences and to Olga Kardailsky and Steve Trewick for additional mitochondrial genomes. Thanks to Simon Hills and Patrick Biggs for computer time and assistance with BEAST analyses. Work carried out at the Allan Wilson Centre for Molecular Ecology and Evolution was supported by the Marsden Fund (Royal Society of New Zealand) to D.P. The manuscript was improved by constructive comments from Associate Editor M. Gouy and three anonymous referees.

Appendix A

The Probability of Observing a PreSpecified Clade

We calculate the probability of observing a prespecified clade of k taxa in a binary tree on n taxa, as the proportion of all binary trees containing that clade. This can be extended to the probabilities of the clade being found as $m = 2, 3$, or even more subclades. We see below (that for all but small values of n and k) that these probabilities are very low so that finding a prespecified clade on a tree formed from new data is highly significant.

It is well known that for n taxa there are

$$B(n) = (2n - 5)!! = 1 \times 3 \times 5 \times \ldots \times (2n - 5) \quad (1)$$

unrooted binary trees where each tip (leaf) of the tree is labeled by a unique taxon (see Penny et al. 1991). Similarly, the number of rooted binary trees is

$$R(n) = (2n - 3)!! = B(n + 1). \quad (2)$$

A simplified approach to deriving the formulae is indicated in figure 4*A*, and the calculations are straightforward in an Excel spreadsheet.

Probability of a Specific Subset Forming a Single Clade (i.e., $m = 1$)

The probability (P) of observing a predefined clade of k taxa in a binary tree of n taxa is

$$P(n,k) = R(k) \times B(n-k+1)/B(n), \quad (3)$$

where the numerator is the number of rooted subtrees for the clade ($R(k)$), multiplied by the number of trees on the remaining taxa ($B(n-k+1)$, including a leaf for the clade). Dividing by the number of unrooted trees ($B(n)$) gives the proportion of trees having that clade of k taxa.

For rooted trees, the probability (P_R) of observing a predefined clade of k taxa in a rooted binary tree of n taxa is similarly

$$P_R(n,k) = R(k) \times R(n-k+1)/R(n) = P(n+1,k). \quad (4)$$

For this question, we consider all binary trees as equally likely; the trees are derived from a Markov model where there is no prior information about the distribution of tree shapes (Steel and Penny 1993). For two taxa in the predefined clade, the equation simplifies (see table 1) to

$P(n,2) = 1/(2n-5)$ for unrooted trees and

$P_R(n,k) = 1/(2n-3)$ for rooted trees.

And for three taxa, it simplifies to

$P(n,3) = 3/(2n-5)(2n-7)$ for rooted trees and

$P_R(n,3) = 3/(2n-3)(2n-5)$ for rooted trees.

In our analyses, we consider the 40 taxa to account for 20 independent taxonomic groupings: passerines (five taxa), cuckoos (two taxa), parrots (three taxa), shorebirds (three taxa), owls (two taxa), dollarbird, kingfisher, trogon, woodpecker + toucan (two taxa), potoo, owlet-nightjar, Apodiformes (two taxa), Accipitriformes (three taxa), falcons (two taxa), flamingo + grebes (three taxa), Pelecaniformes (two taxa), tubenoses (two taxa), stork, penguins (two taxa), and the loon. For rooted trees, there are 21 independent groupings, the above plus Galloanserae (five taxa). If, for example, $k = 5$ and $n = 40$ and the probability of observing a prespecified clade on new data where all taxa are included is

$$P_R(40,5) = R(5) \times R(36)/R(40) \approx 3.5 \times 10^{-6}$$

or where only the independent taxonomic groups are included,

$$P_R(21,5) = R(5) \times R(17)/R(21) \approx 6.3 \times 10^{-5}.$$

Predefined Clade Found as $m \geq 2$ Subgroups

The calculations can be extended to cases where the predicted clade is partitioned into $m = 2$ or more subclades on a tree. Figure 4*C* and *D* shows two cases where a predefined clade appears in two separate areas of the tree ($m = 2$). For a large number of taxa, it is still most unlikely that a predefined clade will be in just two locations on a new tree. In the case shown here, there are $k = 4$ taxa in the clade, and for $m = 2$, they can occur as either a single taxon and a group of three (fig. 4*C*) or as two groups, each with two taxa (fig. 4*D*). When the clade of k taxa is split into m subclades, with $k_1, k_2, \ldots, k_m$ taxa, respectively, then we must consider each combination of the m subclades separately and we find that the probability is

$$P(n,k,m) = \frac{k!B(n-k+m)}{m!B(n)} \sum_{\Sigma k_i = k} \prod_i \frac{R(k_i)}{k_i!}. \quad (5)$$

Literature Cited

Barrowclough GF, Groth JG, Mertz LA. 2006. The RAG-1 exon in the avian order Caprimulgiformes: phylogeny, heterozygosity and base composition. Mol Phyl Evol. 41:238–248.

Boore JL. 2006. The use of genome-level characters for phylogenetic reconstruction. Trends Ecol Evol. 21:439–446.

Brown JW, Payne RB, Mindell DP. 2007. Nuclear DNA does not reconcile 'rocks' and 'clocks' in Neoaves: a comment on Ericson et al. Biol Lett 3:257–259.

Brown JW, Rest JS, García-Moreno J, Sorenson MD, Mindell DP. 2008. Strong mitochondrial DNA support for a Cretaceous origin of modern avian lineages. BMC Biol. 6:6.

Chubb AL. 2004. New nuclear evidence for the oldest divergence among neognath birds: the phylogenetic utility of ZENK (i). Mol Phyl Evol. 30:140–151.

Clarke JA, Tambussi CP, Noriega JI, Erikson GM, Ketcham RA. 2004. Definitive fossil evidence for the extant avian radiation in the Cretaceous. Nature. 433:305–308.

Cooper A, Penny D. 1997. Mass survival of birds across the Cretaceous-Tertiary boundary: molecular evidence. Science. 275:1109–1113.

Cracraft J. 2001. Avian evolution, Gondwana biogeography and the Cretaceous-Tertiary mass extinction event. Proc Roy Soc Lond B Biol Sci. 268:459–469.

Cracraft J, Barker FK, Braun M, et al. 2004. (13 co-authors). 2004. Phylogenetic relationships among modern birds (Neornithes). In: Cracraft J, Donoghue MJ, editors. Assembling the Tree of Life. New York: Oxford University Press. p. 468–489.

Cracraft J, Clarke J. 2001. The basal clades of modern birds. In: Gauthier J, Gall LF, editors. New perspectives on the origin and early evolution of birds. New Haven (CT): Peabody Museum of Natural History, Yale University. p. 143–156.

Curole JP, Kocher TD. 1999. Mitogenomics: digging deeper with complete mitochondiral genomes. Trends Ecol Evol. 14:394–398.

de Kloet RS, de Kloet SR. 2005. The evolution of the spindlin gene in birds: sequence analysis of an intron of the spindlin W and Z gene reveals four major divisions of the Psittaciformes. Mol Phyl Evol. 36:706–721.

Delsuc F, Phillips MJ, Penny D. 2003. Comment on "Hexapod origins: monophyletic or paraphyletic?" Science. 301:1482.

Desjardins P, Morais R. 1990. Sequence and gene organization of the chicken mitochondrial genome. J Mol Evol. 32:153–161.

Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4:699–710.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol. 7:214.

Dyke GJ, Cooper JH. 2000. A new psittaciform bird from the London clay (Lower Eocene) of England. Palaeontology. 43:271–285.

Dyke GJ, Mayr G. 1999. Did parrots exist in the Cretaceous period? Nature. 399:317–318.

1100 Ericson PGP, Anderson CL, Britton T, Elzanowski A, Johansson US, Källersjö M, Ohlson JI, Parsons TJ, Zuccon D, Mayr G. 2006. Diversification of Neoaves: integration of molecular sequence data and fossils. Biol Lett. 2:543–547.

1105 Feduccia A. 2003. 'Big bang' for Tertiary birds? Trends Ecol Evol. 18:172–176.

García-Moreno J, Mindell DP. 2000. Using homologous genes on opposite sex chromosomes (gametologs) in phylogenetic analysis: a case study with avian CHD. Mol Biol Evol. 1110 17:1826–1832.

Gibb GC, Kardailsky O, Kimball RT, Braun E, Penny D. 2007. Mitochondrial genomes and avian phylogeny: complex characters and resolvability without explosive radiations. Mol Biol Evol. 24:269–280.

1115 Groth JG, Barrowclough GF. 1999. Basal divergences in birds and the phylogenetic utility of the nuclear RAG-1 gene. Mol Phyl Evol. 12:115–123.

Guindon D, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likeli-1120 hood. Syst Biol. 52:696–704.

Hackett SJ, Kimball RT, Reddy S, et al. (17 co-authors). 2008. A phylogenomic study of birds reveals their evolutionary history. Science. 320:1763–1768.

Harrison GL, McLenachan PA, Phillips MJ, Slack KE, 1125 Cooper A, Penny D. 2004. Four new avian mitochondrial genomes help get to basic evolutionary questions in the late Cretacous. Mol Biol Evol. 21:974–983.

Harshmann J, Braun EL, Braun MJ, et al. (17 co-authors). 2008. Phylogenomic evidence for multiple losses of flight in ratite 1130 birds. Proc Natl Acad Sci USA. 105:13462–13467.

Hendy MD, Penny D. 1989. A framework for quantitative study of evolutionary trees. Syst Biol. 38:297–309.

Holland BR, Huber KT, Moulton V, Lockhart PJ. 2004. Using consensus networks to visualize contradictory evidence for 1135 species phylogeny. Mol Biol Evol. 21:1459–1461.

Holland BR, Penny D, Hendy MD. 2003. Outgroup misplacement and phylogenetic inaccuracy under a molecular clock—a simulation study. Syst Biol. 52:229–238.

Honeycutt RL, Adkins RM. 1993. Higher-level systematics of 1140 eutherian mammals—an assessment of molecular characters and phylogenetic hypotheses. Annu Rev Ecol Syst. 24: 279–305.

Huelsenbeck JP, Ronquist F. 2001. MrBayes: Bayesian inference on phylogenetic trees. Bioinformatics. 17:754–755.

1145 Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol. 23: 254–267.

Jeffroy O, Brinkmann H, Delsuc F, Philippe H. 2006. Phylogenomics: the beginning of incongruence? Trends 1150 Genet. 22:225–231.

Johansson US, Parsons TJ, Irestedt M, Ericson PGP. 2001. Clades within the 'higher land birds', evaluated by nuclear DNA sequences. J Zool Syst Evol Res. 39:37–51.

Lin Y-H, McLenachan PA, Gore AR, Phillips MJ, Ota R, 1155 Hendy MD, Penny D. 2002. Four new mitochondrial genomes and the increased stability of evolutionary trees of mammals from improved taxon sampling. Mol Biol Evol. 19:2060–2070.

Livezey BC, Zusi RL. 2007. Higher-order phylogeny of modern 1160 birds (Theropoda, Aves: Neornithes) based on comparative anatomy. II. Analysis and discussion. Zool J Linn Soc. 149:1–95.

Lockhart PJ, Penny D, Hendy MD, Howe CJ, Beanland TJ, Larkum AWD. 1992. Controversy on chloroplast origins. FEBS Lett. 301:127–131.

Löytynoja A, Goldman N. 2008. Phylogeny-aware gap place- 1165 ment prevents errors in sequence alignment and evolutionary analysis. Science. 320:1632–1635.

Matthee CA, Eick G, Willows-Munro S, Montgelard C, Pardini AT, Robinson TJ. 2007. Indel evolution of mammalian introns and the utility of non-coding nuclear markers in 1170 eutherian phylogenetics. Mol Phylogenet Evol. 42:827–837.

Mayr G. 2002a. On the osteology and phylogenetic affinities of the Pseudasturidae—lower Eocene stem-group representatives of parrots (Aves, Psittaciformes). Zool J Linn Soc. 136: 715–729. 1175

Mayr G. 2002b. Osteological evidence for paraphyly of the avian order Caprimulgiforms (nightjars and allies). J Ornithol. 143:82–97.

Mayr G. 2005. The postcranial osteology and phylogenetic position of the Middle Eocene Messelastur gratulator Peters, 1180 1994—a morphological link between owls (Strigiformes) and falconiform birds? J Vertebr Paleontol. 25:635–645.

Mayr G, Manegold A, Johansson US. 2003. Monophyletic groups within 'higher land birds'—comparison of morphological and molecular data. J Zoolog Syst Evol Res. 41: 1185 233–248.

Mindell DP, Sorenson MD, Dimcheff DE. 1998. Multiple independent origins of mitochondrial gene order in birds. Proc Natl Acad Sci USA. 95:10693–10697.

Miyaki CY, Matioli SR, Burke T, Wajntal A. 1998. Parrot 1190 evolution and paleogeographical events: mitochondrial DNA evidence. Mol Biol Evol. 15:544–551.

Morgan-Richards M, Trewick SA, Bartosch-Härlid A, Kardailsky O, Phillips MJ, McLenachan PA, Penny D. 2008. Bird evolution: testing the Metaves clade with six 1195 new mitochondrial genomes. BMC Evol Biol. 8:20.

Penny D, Foulds LR, Hendy MD. 1982. Testing the theory of evolution by comparing phylogenetic trees constructed from five different protein sequences. Nature. 297:197–200.

Penny D, Hendy MD. 1986. Estimating the reliability of 1200 evolutionary trees. Mol Biol Evol. 3:403–417.

Penny D, Hendy MD, Steel MA. 1991. Testing the theory of descent. In: Miyamoto MM, Cracraft J, editors. Phylogenetic analysis of DNA sequences. Oxford: Oxford University Press. p. 155–183.

Penny D, Phillips MJ. 2004. The rise of birds and mammals: are 1205 microevolutionary processes sufficient for macroevolution? Trends Ecol Evol. 19:516–522.

Penny D, White WT, Hendy MD, Phillips MJ. 2008. A bias in ML estimates of branch lengths in the presence of multiple signals. Mol Biol Evol. 25:239–242. 1210

Pereira SL, Baker AJ. 2006. A mitogenomic timescale for birds detects variable phylogenetic rates of molecular evolution and refutes the standard molecular clock. Mol Biol Evol. 23: 1731–1740.

Philippe H, Lopez P, Brinkmann H, Budin K, Germot A, 1215 Laurent J, Moreira D, Muller M, Le Guyader H. 2000. Early-branching or fast-evolving eukaryotes? An answer based on slowly evolving positions. Proc Roy Soc Lond B Biol Sci. 267:1213–1221.

Phillips MJ, Delsuc F, Penny D. 2004. Genome-scale phylogeny: 1220 sampling and systematic errors are both important. Mol Biol Evol. 21:1455–1458.

Phillips MJ, McLenachan PA, Down C, Gibb GC, Penny D. 2006. Combined mitochondrial and nuclear DNA sequences resolve the interrelations of the major Australasian marsupial 1225 radiations. Syst Biol. 55:122–137.

Phillips MJ, Penny D. 2003. The root of the mammalian tree inferred from whole mitochondrial genomes. Mol Phyl Evol. 28:171–185.

Poe S, Chubb AL. 2004. Birds in a bush: five genes indicate 1230 explosive evolution of avian orders. Evolution. 58:404–415.

Posada D, Crandall KA. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics. 14:817–818.
Proakis JG, Rader CM, Ling F, Nikias CL, Moonen M, Proudler IK. 2002. Algorithms for statistical signal processing. Upper Saddle River (NJ): Prentice Hall.
Rodriguez-Ezpeleta N, Brinkmann H, Roure B, Lartillot N, Lang BF, Phipippe H. 2007. Detecting and overcoming systematic errors in genome-scale phylogenies. Syst Biol. 56:389–399.
Shapiro LH, Dumbacher JP. 2001. Adenylate kinase intron 5: a new nuclear locus for avian systematics. Auk. 118:248–255.
Slack KE, Delsuc F, McLenachan PA, Arnason U, Penny D. 2007. Resolving the root of the avian mitogenomic tree by breaking up long branches. Mol Phyl Evol. 42:1–13.
Slack KE, Janke A, Penny D, Arnason U. 2003. Two new avian mitochondrial genomes (penguin and goose) and a summary of bird and reptile mitogenomic features. Gene. 302:43–52.
Slack KE, Jones CM, Ando T, Harrison GL, Fordyce E, Arnason U, Penny D. 2006. Early penguin fossils, plus mitochondrial genomes, calibrate avian evolution. Mol Biol Evol. 23:1144–1155.
Sorenson MD, Oneal E, García-Moreno J, Mindell DP. 2003. More taxa, more characters: the Hoatzin problem is still unresolved. Mol Biol Evol. 20:1484–1499.
Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 22:2688–2690.
Steel MA, Penny D. 1993. Distributions of tree comparison metrics—some new results. Syst Biol. 42:126–141.
Stidham TA. 1998. A lower jaw from a Cretaceous parrot. Nature. 396:29–30.
Susko E, Roger AJ. 2007. On reduced amino acid alphabets for phylogenetic inference. Mol Biol Evol. 24:2139–2150.
Swofford DL. 2001. PAUP* phylogenetic analysis using parsimony (* and other methods). Sunderland (MA): Sinauer Associates.
van Tuinen M, Hedges SB. 2001. Calibration of avian molecular clocks. Mol Biol Evol. 18:206–213.
van Tuinen M, Stidham TA, Hadly EA. 2006. Tempo and mode of modern bird evolution observed with large-scale taxonomic sampling. Hist Biol. 18:209–225.
Waterhouse DM. 2006. Parrots in a nutshell: the fossil record of Psittaciformes (Aves). Hist Biol. 18:223–234.
Whitfield JB, Lockhart PJ. 2007. Deciphering ancient rapid radiations. Trends Ecol Evol. 22:258–265.
Zwickl DJ. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion [PhD dissertation]. [Austin (TX)]: The University of Texas at Austin.

Manolo Gouy, Associate Editor

Accepted October 27, 2008

Available online at www.sciencedirect.com

Molecular Phylogenetics and Evolution 46 (2008) 594–605

www.elsevier.com/locate/ympev

Family-level relationships among the Australasian marsupial "herbivores" (Diprotodontia: Koala, wombats, kangaroos and possums)

Matthew J. Phillips [a,b,*], Renae C. Pratt [a]

[a] *Allan Wilson Centre for Molecular Ecology and Evolution, Massey University, Palmerston North, New Zealand*
[b] *Centre for Macroevolution and Macroecology, School of Botany and Zoology, Australian National University, Canberra, ACT 0200, Australia*

Received 24 May 2007; revised 13 August 2007; accepted 5 September 2007
Available online 18 September 2007

Abstract

The marsupial order Diprotodontia includes 10 extant families, which occupy all terrestrial habitats across Australia and New Guinea and have evolved remarkable dietary and locomotory diversity. Despite considerable attention, the interrelations of these families have for the most part remained elusive. In this study, we separately model mitochondrial RNA and protein-coding sequences in addition to nuclear protein-coding sequences to provide near-complete resolution of diprotodontian family-level phylogeny. We show that alternative topologies inferred in some previous studies are likely to be artifactual, resulting from branch-length and compositional biases. Subordinal groupings resolved herein include Vombatiformes (wombats and koala) and Phalangerida, which in turn comprises Petauroidea (petaurid gliders and striped, feathertail, ringtail and honey possums) and a clade whose plesiomorphic members possess blade-like premolars (phalangerid possums, kangaroos and their allies and most likely, pygmy possums). The topology resolved reveals ecological niche structuring among diprotodontians that has likely been maintained for more than 40 million years.
© 2007 Elsevier Inc. All rights reserved.

Keywords: Marsupial; Diprotodontia; Phalangerida; Molecular phylogeny; Long-term ecological niche differentiation

1. Introduction

It is well accepted that marsupials, which are emblematic of the Australian mammal fauna, evolved in the northern continents and dispersed to Australasia via South America and Antarctica during the latest Cretaceous or Early Tertiary (see Woodburne and Case, 1996; Nilsson et al., 2004). Combined analysis of mitochondrial and nuclear data (Phillips et al., 2006) resolved the basal split dividing the Australasian marsupials into two distinct groupings. On one side are the dentally plesiomorphic polyprotodontian orders, including the bandicoots, numbat and marsupial "wolves", "cats", "mice" and "moles". On the other is the order Diprotodontia, whose members are characterized by possessing two procumbent (forward-pointing) lower incisors. Diprotodontians make up approximately 40% of all extant marsupial species richness, with more than 125 species currently recognized.

Diprotodontians are often thought of as herbivorous browsers, grazers and arboreal folivores. While this is true of the most iconic species (e.g. kangaroos, wombats, and koala), other members have diversified into numerous additional ecological niches. These include semi-fossorial root/fungi feeders (bettongs, potoroos), terrestrial frugivores (musky rat-kangaroos), scansorial to arboreal omnivores (pygmy and feathertail possums, petaurid gliders), nectivores (honey possum), insectivores (striped possums) and frugivore/folivores (phalangerid possums). Much additional diversity existed among recently extinct forms (until ≈ 45–55 ka, Roberts et al., 2001), including large carnivorous marsupial lions (Thylacoleonidae).

The present consensus on the interrelations of the ten extant diprotodontian families leaves a seven-lineage basal

* Corresponding author. Fax: +61 2 61255573.
E-mail address: matt.phillips@anu.edu.au (M.J. Phillips).

1055-7903/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2007.09.008

M.J. Phillips, R.C. Pratt / Molecular Phylogenetics and Evolution 46 (2008) 594–605 595

polytomy unresolved, as is shown in Fig. 1. General agreement exists on the monophyly of Vombatiformes (wombats and koala), Macropodoidea (kangaroos and rat-kangaroos) and a petauroid clade that contains at least the petaurids (gliders and striped possums) and the pseudocheirids (ringtail possums). Both morphological and molecular disciplines have provided conflicting evidence on any further suprafamilial resolution. Accordingly, the matrix representation with parsimony (MRP) supertree of Cardillo et al. (2004) that incorporated source trees from all marsupial phylogenetic studies from 1995–2003 also recovered the seven-lineage basal diprotodontian polytomy, as shown in Fig. 1.

Resolving family-level diprotodontian phylogeny can essentially be reduced to testing four taxonomic hypotheses (A–D below) that are consistent with the major classification of Aplin and Archer (1987) and the favoured MRP weighted (by data quality) supertree of Cardillo et al. (2004), and then determining the affinities of the burramyids.

(A) *Phalangerida* (Aplin and Archer, 1987) includes all diprotodontians except Vombatiformes, such that the diprotodontian root is defined by the split between Phalangerida and Vombatiformes. Phalangerida is supported by a number of basicranial characters (Aplin and Archer, 1987; Springer and Woodburne, 1989), simplified yolk-sac placentation relative to Vombatiformes and many polyprotodont marsupials (Hughes, 1974), serological studies (e.g. Kirsch, 1977) and DNA–DNA hybridization (Kirsch et al., 1997). Recent evidence on the veracity of Phalangerida has been far from unequivocal. Morphological cladistic analyses (e.g. Horovitz and Sánchez-Villagra, 2003; Luo et al., 2003) and a mitochondrial (mt) genome study (Nilsson et al., 2004; but see Munemasa et al., 2006) have rendered Phalangerida paraphyletic, with Vombatifomes grouping with kangaroos and their allies. In contrast, a five-loci nuclear concatenation (Amrine-Madsen et al., 2003) strongly supported a phalangeridan grouping, albeit

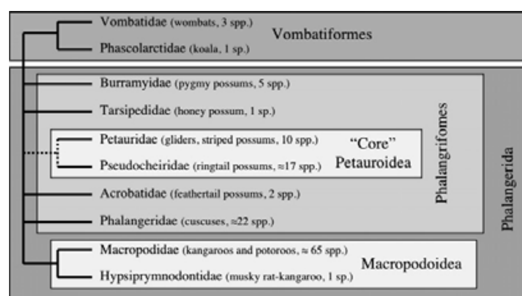


Fig. 1. The present consensus on the interrelations of the ten extant diprotodontian families. General agreement exists within both morphological and molecular studies on Vombatiformes, Macropodoidea and a Petauroidea clade that contains at least Petauridae and Pseudocheiridae. Both disciplines have provided conflicting evidence regarding the monophyly of Phalangerida and Phalangeriformes.

without sampling three key families (Acrobatidae, Tarsipedidae, Burramyidae) whose placements in previous partial mt genome studies (e.g. Osborne et al., 2002; Kavanagh et al., 2004) have rendered Phalangerida either polyphyletic or paraphyletic.

(B) *Phalangeriformes* (*sensu* Kirsch et al., 1997) is comprised of all the possum families (Phalangeridae, Petauridae, Pseudocheiridae, Tarsipedidae, Acrobatidae and Burramyidae) to the exclusion of Macropodoidea and Vombatiformes. A single basicranial character, the fusion of the ectotympanic to a pneumatised squamosal has been attributed great phylogentic weight by numerous authors who have favoured this possum monophyly (e.g. Springer and Woodburne, 1989; Marshall et al., 1990; but see Szalay, 1994). Flannery (1987) gave this character precedence over five dental characters (also see Pearson, 1950) and the presence of tail scales that support an alternative macropodoid plus phalangerid grouping. Molecular studies have had little impact on these arguments, only ever providing limited or contradictory support for Phalangeriformes or either of the macropodoid plus phalangerid, or macropodoid plus petauroid alternatives (for review, see Kirsch et al., 1997; Kavanagh et al., 2004).

(C) *Petauroidea* (Gill, 1872), previously conceived as including Petauridae and Pseudocheiridae, has been expanded on the basis of albumin microcompliment fixation (MC'F, Baverstock, 1984) and DNA–DNA hybridization (Edwards and Westerman, 1995) to incorporate Tarsipedidae and Acrobatidae. However, morphological evidence for this grouping has been difficult to interpret (see Aplin and Archer, 1987; Szalay, 1994) due to the extreme autapomorphy of the monotypic, pollen and nectar-feeding *Tarsipes rostratus* and to a lesser extent, the more plesiomorphic acrobatids. Moreover, the mt genome sequence analysis of Nilsson et al. (2004) found Petauroidea to be paraphyletic. Mostly though, DNA sequence analyses have provided limited to moderate statistical support for Petauroidea (e.g. Baker et al., 2004 using nuclear RAG-1 and Kavanagh et al., 2004 using mt RNA) or have otherwise been equivocal (e.g. Osborne et al., 2002, using NADH2).

(D) *Tarsipedoidea* (Aplin and Archer, 1987) groups together Tarsipedidae and Acrobatidae, the two smallest petauroid families in terms of both physical size and species richness. Aplin and Archer (1987) cited a number of cranial and male reproductive traits in support of this grouping. Additional support has come from the MC'F study of Baverstock et al. (1990) and the RAG-1 analysis of Baker et al. (2004). Results that have underpinned arguments against Tarsipedoidea have usually also contradicted Petauroidea by placing *Tarsipes* basal among possums or diprotodontians altogether. Examples include the serological work of Kirsch (1977) and evidence from studies of sperm morphology (see Harding et al., 1984 and Luckett, 1994). However, the mt 12S/16S and t^{val} RNA analysis of Kavanagh et al. (2004), which did support Petauroidea, placed acrobatids deepest among this superfamily, leaving

Tarsipes as the moderately well supported sister to Petauridae/Pseudocheiridae.

Burramyidae has been the most itinerant diprotodontian family in its affinities. Gunson et al. (1968) and Kirsch (1968) grouped these pygmy possums with the acrobatids based on shared $2N = 14$ karyotype, which is now considered plesiomorphic for Australasian marsupials. Morphological studies have subsequently placed burramyids basal among possums (e.g. Archer et al., 1987), with petauroids (Marshall et al., 1990) and in combination with molecular data, basal among all diprotodontians (Asher et al., 2004). Molecular analyses have provided little resolution for burramyid affinities, although affinities with phalangerids have been the most frequent (e.g. Springer and Kirsch, 1991; Baker et al., 2004). However, short mtDNA sequences have (weakly) favoured placements for burramyids that are separate from other possums, such as with Vombatiformes (Kavanagh et al., 2004) or basal among Diprotodontia (Osborne et al., 2002).

Munemasa et al. (2006) recently provided complete mt genome sequences for six diprotodontians and found moderate support for Phalangerida. However, their analyses did not include the mt genome sequences published in Nilsson et al. (2004) and Phillips et al. (2006), which between them provided 11 new diprotodontian and outgroup (Australasian polyprotodontian) sequences. In the present study all of these complete mt genome sequences are analysed together, along with the five-loci nuclear dataset of Amrine-Madsen et al. (2003) as separately modelled process partitions (see Phillips et al., 2006). In doing so, we provide the first inference of diprotodontian phylogeny from long sequences (20,306 nucleotides) with broad taxonomic coverage. The resulting phylogeny is near-fully resolved (including Vombatiformes, Phalangerida, Petauroidea and Macropodoidea–Phalangeridae–Burramyidae) and provides a novel context within which to discuss ecological niche differentiation among diprotodontians.

2. Methods

2.1. Data matrices

The primary dataset includes complete mt genome protein, rRNA and tRNA coding sequences, totaling 14,645 nucleotides after alignment in Se-Al 2.0a9 (Rambaut, 1996) and exclusion of apporoximately 690 nucleotide sites with ambiguous homology (mostly from RNA loop regions). Excepting for the burramyids and the monotypic *Hypsiprymnodon*, this alignment includes representatives from all diprotodontian families: *Phascolarctos cinereus* (Koala, NC_008133), *Vombatus ursinus* (common wombat, NC_003322), *Marcropus robustus* (wallaroo, NC_001794), *Potorous longipes* (long-nosed potoroo, NC_006524), *Lagorchestes hirsutus* (rufous hare-wallaby, NC_008136), *Trichosurus vulpecula* (common brushtail possum, NC_003039), *Phalanger interpositus* (Stein's cuscus, NC_008137), *Distoechurus pennatus* (feather-tailed possum, NC_008145), *Tarsipes rostratus* (honey possum, NC_006518), *Pseudocheirus peregrinus* (common ringtail possum, NC_006519), *Dactylopsila trivirgata* (striped possum, NC_008134) and *Petaurus sp.* (sugar glider, NC_008135).

The nine available complete mt genomes from non-diprotodontian Australasian marsupials provided data for the outgroup: *Isoodon macrourus* (northern brown bandicoot, NC_002746), *Perameles gunnii* (eastern barred bandicoot, NC_006521), *Echymipera rufescens* (rufous spiny bandicoot, NC_007632), *Macrotis lagotis* (bilby, NC_006520), *Notoryctes typhlops* (marsupial mole, NC_006522), *Phascogale tapoatafa* (brush-tailed phascogale, NC_006523), *Dasyurus hallucatus* (northern quoll, NC_007630), *Sminthopsis douglasi* (julia creek dunnart, NC_006517) and *Sminthopsis crassicaudata* (fat-tailed dunnart, NC_007631).

The 21-taxon complete mt genome matrix (Mt21) is analysed alone and together (MtNuc21) with 5,661 nucleotides from five nuclear protein-coding genes (IRBP, vWF, APOB, RAG-1, and BRCA1) from Amrine-Madsen et al. (2003). Although the nuclear data provides further information for only 12 (six outgroup and six diprotodontian) of the 21 taxonomic units among the MtNuc21 dataset, these data lend substantial additional power for distinguishing between hypotheses for rooting the diprotodontian phylogeny.

The Mt21 and MtNuc21 matrices emphasize sequence length and adhere to the outgroup sampling strategy of Phillips and Penny (2003), in that taxa are excluded if they provide long, isolated branches that are unstable in the tree, but tend to fall adjacent to the ingroup on the unrooted tree. One such taxon, the monotypic South American Monito del Monte (*Dromiciops gliroides*: Microbiotheria) is included in an additional matrix (Mt29) that emphasizes taxon sampling. Mt29 also includes additional phalangerids and pseudocheirids as well as *Hypsiprymnodon moschatus* and both burramyid genera *Burramys* and *Cercartetus*, such that all diprotodontian families are represented. The cost of this additional taxon-sampling is reduced gene-sampling (12S rRNA, 16S rRNA and NADH2, totaling 3174 nucleotides). GenBank accessions for the additional mt sequences in Mt29 and for the nuclear sequences in MtNuc21 are provided in Appendix A (supplementary information online). All alignments are available from the corresponding author.

Numerous studies have revealed compositional heterogeneity to be a particular concern for mammalian mitogenomic phylogenetics (e.g. Schmitz et al., 2002; Gibson et al., 2005). Recent examinations of compositional heterogeneity among marsupials (Phillips et al., 2006) and birds (Harrison et al., 2004) provide several relevant findings: (a) the influence of compositional heterogeneity on phylogenetic reconstruction was exacerbated among data partitions for which saturation has greatly eroded phylogenetic signal, such as 3rd codon positions, (b) Compositional χ^2 tests are poor indicators of potential for phylogenetic bias and they are not directly comparable between datasets, because their statistical power depends

on factors such as the number of variable sites, (c) RY-coding 3rd codon position data excluded sources of compositional heterogeneity more efficiently than did using the protein (amino acid) sequence, in terms of retaining phylogentic signal and (d) much of the compositional heterogeneity was attributable to the proportion of Thymine relative to Cytosine being higher in peramelemorphians and didelphimorphians than among other marsupial orders. Skewed nucleotide (and amino acid) composition in these outgroup taxa alone will not directly bias inference of the diprotodontian tree. However, two metrics described below, stemminess and relative compositional variability are used here respectively to examine phylogenetic signal erosion and the magnitude of compositional non-stationarity among the diprotodontian mt genome data.

2.1.1. Stemminess

Fiala and Sokal (1985) defined stemminess as the proportion of overall tree-length that internal branches contribute. As in Phillips et al. (2001), this definition is employed on minimum evolution (ME) trees derived from uncorrected distances. Given the same topology, greater phylogenetic signal erosion results in shorter internal branches relative to their attendant external branches (hence, lower stemminess).

2.1.2. Relative compositional variability (RCV, Phillips and Penny, 2003)

RCV is the average variability in composition between taxa; for nucleotides this is:

$$RCV = \sum_{i=1}^{n}(|A_i - A^*| + |T_i - T^*| + |C_i - C^*| + |G_i - G^*|)/n.t$$

A_i, T_i, C_i and G_i are the frequencies of each nucleotide for the ith taxon. A^*, T^*, C^* and G^* are averages across the n taxa, and t is the number of sites. Uninformative sites effectively dilute apparent non-stationarity so were excluded (along with gapped sites) from RCV calculations.

Inspection of the stemminess and RCV results (Table 1) suggests that RY-coding is warranted for the mt protein 3rd codon positions. These sites have the highest relative compositional variability among all of the partitions. Furthermore, stemminess among the mt 3rd codon positions is on average only 62% of that among the other partitions. Low stemminess, effectively compounds the influence of the strong compositional bias, because less of the "historical" signal remains to compete against the bias signal. However, RY-coding the 3rd codon positions results in RCV comparable to and stemminess that is higher than the other partitions (Table 1). Hence, in the primary phylogenetic analyses, the 3rd codon positions are RY-coded in each of the Mt21, MtNuc21 and Mt29 datasets. This excludes the source of compositional bias in the data that has the greatest potential to mislead phylogenetic reconstruction. Nevertheless, the possible influence of remnant compositional heterogeneity on phylogenetic reconstruction is also examined, using base frequency distance trees (see below).

Table 1
Relative compositional variability and stemminess among the diprotodontian complete mt genomes

Coding sequence	Partition	RCV	Stemminess
RNA coding[a]	Stems	0.0748	0.0951
	Loops	0.0726	0.0934
Protein coding	Codon 1	0.0704	0.0840
	Codon 2	0.0802	0.0839
	Codon 3	0.0884	0.0549
	Codon 3 (RY-coded)	0.0756	0.1337

[a] 12S/16S rRNA and tRNA coding sequences are concatenated and divided into stem and loop sites.

2.2. Phylogenetic analysis

Diprotodontian phylogeny was inferred from the Mt21, MtNuc21 and Mt29 datasets as single concatenations and partitioned into codon positions for the protein-coding data (separately for mt and nuclear) and structure (stem and loop sites) for the RNA-coding data. With only one protein-coding gene included in the Mt29 dataset there are 65 free parameters requiring estimation for the 2nd codon position, more than half the number of variable sites (115). By combining the 1st and 2nd position data (Mt29 only) the same number of free parameters are estimated from 331 variable sites. Substitution model categories for each partition were assigned according to whichever of the ModelTest 3.06 (Posada and Crandall, 1998) AIC or hLRT recommendations was more general. In each case these were GTR+I+Γ_4 for standard nucleotide partitions and F81+I+Γ_4 (equivalent to CF87+I+Γ_4, Cavender and Felsenstein, 1987) for the RY-coded mitochondrial 3rd codon partitions.

Bayesian inference (MrBayes 3.1.2; Huelsenbeck and Ronquist, 2001) analyses were run with unlinked substitution models and branch-length rate multipliers among the partitions. Four mcmc chains for each of two independent runs proceeded for 5,000,000 (Mt21, MtNuc21) or 3,000,000 (Mt29) generations with trees being sampled every 1000 generations. The burn-in for each MrBayes run (which varied between 400,000 and 1,200,000 generations) ensured that $-\ln L$ had plateaued, clade frequencies had converged between runs and estimated sample sizes for substitution parameter estimates were above 200 (using Tracer v1.0, Rambaut and Drummond, 2003).

Maximum parsimony (MP), Minimum evolution (ME) and maximum likelihood (ML) analyses were performed within PAUP*4.0b10 (Swofford, 2002). ML bootstrapping (500 replicates) applied heuristic searches to random starting trees for the datasets as single concatenations. Following Phillips and Penny (2003) the TN93 (Tamura and Nei, 1993) substitution model was applied to these concatenations, such that the transversions in the standard nucleotide

and RY-coded data are effectively weighted equally. As an alternative to optimizing substitution parameters on an NJ distance tree (as in ModelTest), these were optimized on the MP tree, then employed in a ML heuristic search and re-optimized on the resulting tree for use in the bootstrap analysis. In order to ensure computational feasibility for these bootstrap analyses, clades that are uncontroversial in all recent molecular and morphological classifications and also received posterior probabilities of 1.00 in the Bayesian analysis were constrained. These include Sminthopsinae, *Dasyurus* + *Phascogale*, Dasyuridae, Peramelinae, Peramelidae and Peramelemorphia among the outgroup and Diprotodontia, Vombatiformes, Macropodinae, Phalangeridae and Petauridae among the ingroup.

Support among alternative topologies was further examined with KH (Kishino and Hasegawa, 1989) and approximately unbiased (AU; Shimodaira, 2002) tests, using the RELL method (100,000 replications) within CONSEL (Shimodaira and Hasegawa, 2001). The AU test is related to the SH test (Shimodaira and Hasegawa, 1999) and has been developed in order to overcome tree selection biases that affect the latter test when multiple topologies are being simultaneously compared.

The ML significance tests were applied with the sequences treated as separately modelled process partitions among codon positions (1, 2, 3), RNA structures (stems, loops) and genomes-of-origin (mt, nuclear), as for the Bayesian inference analyses (see Table 2). Partitioning the data allows for more accurate models of sequence evolution that address differential influences on substitution patterns across the sequence (e.g., Yang, 1996; Caterino et al., 2001). Substitution model categories again followed the ModelTest recommendations, as employed in the Bayesian inference analyses described above. All substitution parameters and branch-lengths were ML optimized for each partition, for each tree hypothesis.

2.3. Bias exploration

Minimum evolution base-frequency (BF) distance trees were constructed from matrices of pairwise base-frequency distances to assess the potential for compositional bias to affect phylogenetic inference. The basic idea is to compare BF distance tree-length differences between alternative topologies. In this way, Phillips and Penny (2003) showed that composition bias could explain earlier mt genome studies grouping marsupials with monotremes, a grouping that contradicts the well accepted sister relationship between marsupial and placental mammals. Base-frequency distances are half the sum of absolute frequency differences between taxon pairs for each nucleotide category. So the pairwise base-frequency (BF) distance between taxa *i* and *j* is:

$$BF\ distance = (|A_i - A_j| + |T_i - T_j| + |C_i - C_j| + |G_i - G_j|)/2$$

A_i, T_i, C_i, and G_i, and A_j, T_j, C_j and G_j are the frequencies of each nucleotide for the *i*th and *j*th taxa, respectively. Dividing by two is necessary for ME differences between BF distance trees to be comparable with ME differences on standard (absolute) distance trees. Otherwise, a substitution at a site in taxon *i* that previously had the same base as for taxon *j* will result in one unit of standard distance, but two units of base-frequency distance. Parsimony-uninformative characters were excluded from base-frequency distance calculations, as these cannot explain ME differences between standard distance trees. For all of the ME trees, any negative branch-lengths were treated as absolute values for computing tree-length.

The possible influence of branch-length related artifacts were examined in a simulation study akin to that of Goremykin et al. (2005), which showed basal relationships among flowering plants inferred from chloroplast sequences to be mislead by long-branch attraction. Pseudoreplicates of the Mt21 matrix were simulated in SeqGen 1.3 (Rambaut and Grassly, 1997) with substitution parameters and branch-lengths as optimized for the original data, using the topology (shown in Fig. 3) that was favoured by both Bayesian inference and ML hypothesis testing. The optimal trees for each of 200 simulated data matrices were identified with heuristic searches under the heterogeneous rates-across-sites model they were simulated under. Rate-homogenous ML and MP were also employed on the same simulated datasets, so as to provide "model-misspecified" parametric bootstrap support for clade hypotheses.

Table 2
Partitions included for the three different datasets[a]

Dataset	Number of partitions	Partitions
Mt21	5	mt protein codons (1,2,3), mt RNA (stems, loops)
MtNuc21	8	mt protein codons (1,2,3), mt RNA (stems, loops), nuclear protein codons (1,2,3)
Mt29	4	NADH2 protein codons (12,3), rRNA (stems, loops)

[a] Each partition was modeled separately in the Bayesian inference and ML hypothesis testing analyses.

3. Results

3.1. Phylogenetic inference

The analyses focus primarily on the Mt21 and MtNuc21 data matrixes, which lend considerably more resolving power than does the shorter Mt29 data matrix, which instead emphasizes taxon sampling. Nevertheless, Bayesian inference and ML analyses of Mt29 (Fig. 2) recover a tree topology that is fully congruent with those for Mt21 and MtNuc21 (Fig. 3) and additionally, allows the placement of the burramyids (pygmy possums) to be examined. In all analyses, the macropodoids (macropodids plus hypsiprymnodontids), phalangerids, and burramyids (when included) group together as sister to Petauroidea. These

M.J. Phillips, R.C. Pratt / Molecular Phylogenetics and Evolution 46 (2008) 594–605 599

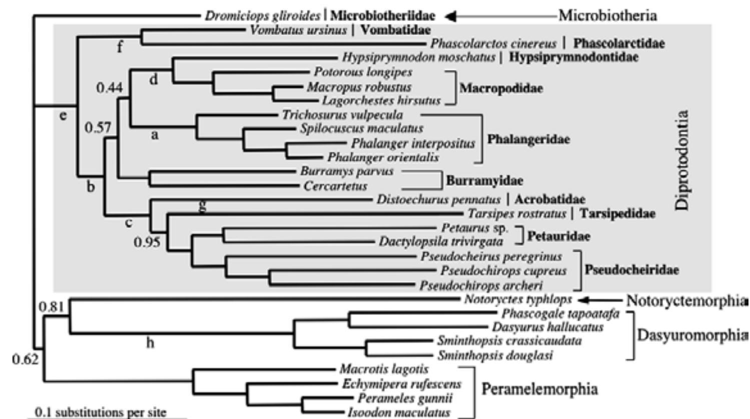


Fig. 2. Bayesian inference phylogram for the Mt29 dataset. Bayesian posterior probabilities (BPP) are shown to the left of the clades they refer to. Where not shown, BPP = 1.00. Partitioned maximum likelihood analysis favoured the same tree, including Burramyidae as sister to a clade containing Phalangeridae and Macropodoidea (Macropodidae–Hypsiprymnodontidae). AU P-values for rejecting the six nearest (as well as previously published) alternative burramyid placements are a: 0.500, b: 0.446, c: 0.269, d: 0.080, e: 0.102, f: 0.042, g: 0.002, h: <0.001, where the branches referred to are those above a–h.

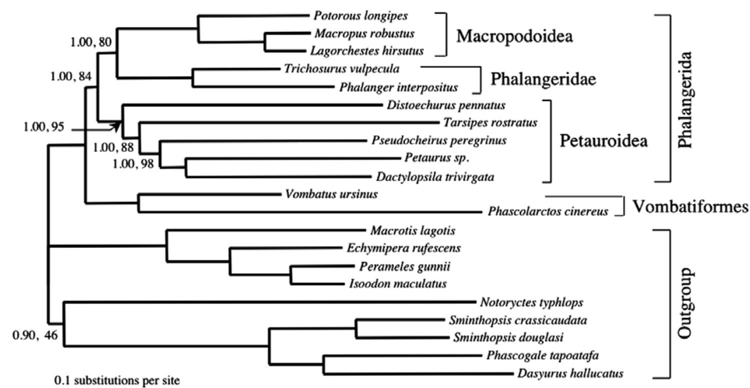


Fig. 3. Bayesian inference phylogram for the Mt21 dataset. Bayesian posterior probabilities (BPP) followed by ML bootstrap support (BP) are shown to the left of the clades they refer to. Where not shown, BPP = 1.00 and BP = 100%.

possum and kangaroo groups comprise the suborder Phalangerida, which diverges basally within Diprotodontia as sister to the other suborder, Vombatiformes (wombats and koala).

Bayesian inference and ML both favour burramyids grouping with macropodoids and phalangerids (Fig. 2). While the Bayesian posterior probability (BPP) for this grouping is only 0.57, conflict derives mainly from alternative placements for the macropodoids. If the grouping of Macropodoidea and Phalangeridae relative to Petauroidea is fixed, as is resolved with the longer sequences, then the BPP for Macropodoidea–Phalangeridae–Burramyidae rises to 0.97. The more conservative AU testing rejects all placements of Burramyidae outside of Phalangerida at $P \leqslant 0.102$. Placements as sister to all other phalangeridans or to petauroids specifically cannot however be rejected even at $P \leqslant 0.25$. The most likely burramyid affinities are as sister to Phalangeridae or as sister to Macropodoidea–Phalangeridae, with the latter placement favoured by only 1.8 $-\ln L$ units.

The full mt genome dataset (Mt21) provides increased resolution (over Mt29) for relationships among the non-burramyid diprotodontians. Indeed, as shown in Fig. 3, the BPP values for all diprotodontian clades are 1.00. ML bootstrapping gives an identical topology, with groupings attaining $\geqslant 95\%$ bootstrap support including Diprotodontia, Vombatiformes and Petauroidea (among others). Three groupings, Phalangerida, Tarsipedidae–Petauridae/Pseudocheiridae, and Macropodoidea–Phalangeridae receive slightly lower bootstrap support (80–90%).

Robust support for relationships among diprotodontians from the Bayesian inference and ML bootstrapping analyses of Mt21 is encouraging. However, PAUP* does not allow for separate modeling of the protein codon and RNA structural partitions, for which substitution processes differ markedly (e.g. Pupko et al., 2002; Phillips et al., 2006). Furthermore, BPP values from the Bayesian inference analysis (which was partitioned) may be positively misleading. This is because the assumption that stochastic error is accurately inferred from the posterior probability ratio of proposed versus current trees along the mcmc chain depends on the dubious assumption that the substitution models closely reflect the evolutionary processes under which the sequences have evolved (Felsenstein, 2004; Kelcher and Thomas, 2006).

Phylogenetic hypothesis testing associated with separately modeled process partition ML circumvents both the concatenation problem associated with the ML bootstrapping in PAUP* and the restriction to parametric error estimation in MrBayes. Shimodaira and Hasegawa (1999) showed that such AU and KH hypothesis tests can more closely reflect sampling error than do bootstrap values, which in turn are far more faithful than typically overconfident BPP values (Suzuki et al., 2002; Gontcharov et al., 2004).

The 95% support consensus level for the ML bootstrap analysis provides full resolution of Diprotodontia, except for a 4-way basal polytomy and a 3-way polytomy within Petauroidea. Of the 45 trees that these polytomies represent, all but six trees can be rejected at $P \leqslant 0.036$ for the AU test (0.032 for the KH test) on the Mt21 data alone. ML score differences and associated AU and KH P-values are shown in Table 3 for the favoured clade hypotheses and for each of these, the two alternatives that are least easy to reject (on AU P-values).

Alternatives to the favoured topology (that of Fig. 3) include paraphyly of Phalangerida at $P = 0.244$ (0.134), Macropodoidea–Petauroidea at $P = 0.134$ (0.058) and Tarsipedoidea (Acrobatidae-*Tarsipes*) at $P = 0.237$ (0.149). Inclusion of the nuclear sequences (MtNuc21) further resolves the rooting of Diprotodontia in favour of a basal split between Phalangerida and Vombatiformes, with all alternatives rejected at $P \leqslant 0.001$ (0.001). The nuclear data has less influence on support among hypotheses for the relative placements of Macropodoidea, Phalangeridae and Petauroidea. A possible explanation here is sparse taxon-sampling; the five-loci nuclear dataset of Amrine-Madsen et al. (2003) includes only one macropodoid, one phalangerid and two petauroids. The nuclear sequences were also unavailable for Tarsipedidae and Acrobatidae, preventing additional examination of their relative placements.

Despite the shortfalls in taxonomic coverage among the nuclear sequences within MtNuc21, the expanded gene coverage relative to Mt21 lends confidence to the resolution at higher taxonomic levels. Alternatives to Diprotodontia, Phalangerida and Petauroidea can all be rejected at ($AU_{MtNuc21}$) $P < 0.01$, including the placement of *Tarsipes* as sister to all other diprotodontians (as was recovered in the mt genome analysis of Nilsson et al., 2004).

3.2. Bias exploration

The phylogenetic analyses of the combined mt and nuclear data strongly reject all but two topological alterna-

Table 3
ML hypothesis testing for (a) Mt21 and (b) MtNuc21[a]

Alternative Groupings	−ln*L*	*P*-values	
		AU	KH
(a) Mt21			
Phalangerida (Vomb, (Petauro, (Macro, Phal)))[b]	<92562.9>	—	—
1. (Petauro, (Vomb, (Macro, Phal)))	+10.7	0.244	0.134
2. (Petauro, (Macro, (Vomb, Phal)))	+24.8	0.129	0.060
Macropodoidea+Phalangeridae (Petauro, (Macro, Phal))	<92562.9>	—	—
4. (Phal, (Macro, Petauro))	+19.5	0.134	0.058
5. (Marco, (Phal, Petauro)) = possum monophyly	+28.8	0.002	0.004
(Acrobatidae, (*Tarsipes*, (Petauridae, Pseudocheir)))[c]	<92562.9>	—	—
6. (*Tarsipes*, Acrobatidae) = Tarsipedoidea	+9.7	0.237	0.149
7. (*Tarsipes*, (Acrobatidae, (Petauridae, Pseudocheir)))	+16.2	0.012	0.034
(b) MtNuc21			
Phalangerida (Vomb, (Petauro, (Macro, Phal)))	<116001.3>	—	—
1. (Petauro, (Vomb, (Macro, Phal)))	+62.6	<0.001	0.001
2. (Petauro, (Macro, (Vomb, Phal)))	+83.7	<0.001	<0.001
Macropodoidea+Phalangeridae (Petauro, (Macro, Phal))	<116001.3>	—	—
4. (Phal, (Macro, Petauro))	+18.4	0.090	0.099
5. (Marco, (Phal, Petauro)) = possum monophyly	+33.9	<0.001	0.003

[a] Each favoured hypothesis has its −ln*L* score is provided. The two alternative hypotheses that are least easy to reject (on AU *P*-values) are numbered below each favoured hypothesis and correspond between the (a) Mt21 and (b) MtNuc21 results.

[b] *Abbreviations:* Vomb (Vombatiformes), Petrauro (Petauroidea), Marco (Macropodoidea), Phal (Phalangeridae), Pseudocheir (Pseudocheiridae).

[c] Results for the placement of *Tarsipes* and Acrobatidae among Petauroidea are only shown for Mt21. The MtNuc21 dataset provides no additional information because the former two taxa are not represented among the 5-loci nuclear dataset of Amrine-Madsen et al. (2003).

tives to the favoured tree, which is shown in Fig. 3. One of these alternatives concerns the rooting of Petauroidea and the other concerns the rooting of Phalangerida. In both cases, AU-testing strongly rejects one "local" rearrangement of the favoured rooting, but not the other. So with respect to the favoured *Tarsipes*–Petauridae/Pseudocheiridae basal petauroid grouping, Acrobatidae–Petauridae/Pseudocheiridae is rejected at $P = 0.012$, while the third alternative, *Tarsipes*–Acrobatidae (i.e. Tarsipedoidea) can only be rejected at $P = 0.237$. Similarly with respect to the favoured Macropodoidea–Phalangeridae relationship at the base of Phalangerida, the Phalangeridae–Petauroidea local rearrangement is rejected at $P > 0.001$, while Macropodoidea–Petauroidea is more difficult to reject at $P = 0.090$.

As noted in Phillips et al. (2004), such asymmetry in the rejection of local rearrangements can indicate that the power to reject a hypothesis is limited by the presence of a non-phylogenetic signal resulting from model misspecification, rather than by stochastic error. As a first examination for such biases, base frequency (BF) distance trees are compared among the local rearrangements for the rooting of both Phalangerida and Petauroidea. Table 4 reveals that for the rooting of both Phalangerida and Petauroidea, compositional heterogeneity does not favour the ML tree (groupings 1 and 4, consistent with Fig. 3), but instead contributes "non-phylogenetic" signal to the local rearrangements of this tree that were most difficult for the AU testing to reject (respectively groupings 2. Macropodoidea–Petauroidea and 5. Tarsipedoidea in Table 4). In fact, composition bias handicaps the ML tree relative to Macropodoidea-Petauroidea by 36.47 distance units and Tarsipedoidea by 149.42 distance units. *Tarsipes* and Acrobatidae are not represented among the nuclear sequences. However, composition bias favouring Macropodoidea–Petauroidea is also present among the nuclear data (ME on BF distances, not shown), though the difference from Macropodoidea–Phalangeridae is only 5.27 distance units.

As a second examination for bias, simulating Mt21 sequence evolution along the favoured tree (Fig. 3) under the original ML-optimized substitution model (including $I+\Gamma_4$) allows the potential for artifacts related to branch-length estimation to be examined. Table 4 shows that under rate-homogenous ML and MP, the expected phalangeridan root (as simulated) was recovered, but that the simulated *Tarsipes*–Petauridae/Pseudocheiridae grouping was not recovered. Instead, *Tarsipes*–Acrobatidae (=Tarsipedoidea) was strongly supported. Hence, inference of the relative placements of *Tarsipes* and Acrobatidae are highly sensitive to the modeling of heterogeneous rates-across-sites (RAS). Insufficient allowance for RAS (and hence, underestimation of branch-lengths, (see Waddell, 1995; Sullivan and Swofford, 1997) provides a phylogenetic bias that favours Tarsipedoidea.

In summary, the Bayesian inference and ML analyses firmly resolve for Diprotodontia, Vombatiformes, Phalangerida, Petauroidea and Petauridae/Pseudocheiridae and provide moderately strong support for both a *Tarsipes*–Petauridae/Pseudocheiridae grouping and an association between Macropodoidea and Phalangeridae (±Burramyidae). The examinations for compositional and branch-length artifacts add further weight to the latter two hypotheses in so much as they reveal that support for their closest alternative topologies is being bolstered by "non-phylogenetic" signals. The major remaining uncertainty is the placement of Burramyidae, which although falling within a Macropodoidea–Phalangeridae–Burramyidae grouping in all analyses, can only be confidently placed within the higher-level taxon, Phalangerida.

4. Discussion

Previous mt genome studies (Nilsson et al., 2004; Munemasa et al., 2006; Phillips et al., 2006) have provided low or conflicting support for interrelations among diprotodontian families. The present study combines the sequences from the aforementioned mt genome studies along with five nuclear loci (from Amrine-Madsen et al., 2003) and employs independent models for genome-of-origin × protein codon (or RNA structure) partitions. As a result, the

Table 4
Comparison of alternative groupings among the "basal" members of Phalangerida and Petauroidea for (a) Mt21 BF distance minimum evolution scores and (b) parametric bootstrap support under heterogeneous rates-across-sites (RAS) ML, rate-homogenous ML and MP for the simulated Mt21 datasets

Alternative groupings	(a) ME scores on BF distances	(b) Parametric bootstrap support (%)		
		ML (GTR+I+Γ)	ML (GTR)	MP
Within Phalangerida				
1. (Macropodoidea, Phalangeridae)[a]	+36.47	100	100	99
2. (Macropodoidea, Petauroidea)	<2434.30>	0	0	0
3. (Petauroidea, Phalangeridae)	+13.74	0	0	1
Within Petauroidea				
4.(*Tarsipes*, (Petauridae, Pseudocheir)[a,b]	+149.42	100	9	4
5. (*Tarsipes*, Acrobatidae)	<2321.35>	0	91	95
6. (Acrobatidae, (Petauridae, Pseudocheir)	+174.37	0	0	0

[a] The tree recovered by the primary phylogentic analyses (e.g. Fig. 3) is consistent with grouping 1 within Phalangerida and grouping 4 within Petauroidea.
[b] *Abbreviation:* Pseudocheir (Pseudocheiridae).

diprotodontian basal polytomy in Fig. 1 (representing 10,395 possible rooted trees) is resolved to all but a handful of possible trees.

Two of the prior hypotheses, Phalangerida (non-vombatiform diprotodontians) and Petauroidea (gliders plus striped, ringtail, feathertail and honey possums) have been favoured by some molecular studies (e.g. Baker et al., 2004; Munemasa et al., 2006), though with low to moderate statistical support, and also by morphological character analyses (e.g. Aplin and Archer, 1987; Marshall et al., 1990). On the other hand, algorithmic cladistic analyses of morphology (e.g. Horovitz and Sánchez-Villagra, 2003; Luo et al., 2003) have assigned petauroids paraphyletically at the base of Diprotodontia and grouped Vombatiformes variously within a paraphyletic Phalangerida. All of the present analyses unequivocally confirm the monophyly of both Petauroidea and Phalangerida.

Setting aside the affinities of the burramyids (pygmy possums) for the moment, the basal division of Phalangerida becomes a question of the interrelations among the three major extant diprotodontian radiations, the macropodoids (kangaroos and allies), the phalangerid possums and the petauroid possums. ML hypothesis-testing clearly favours macropodoids associating with phalangerids, rather than with petauroids ($AU_{MtNuc21}$, $P = 0.090$). Indeed, the present results understate the confidence that may be attributed to a macropodoid–phalangerid association. Firstly, BF distance analyses (Table 4) indicate that apparent signal for the alternative macropodoid–petauroid hypothesis is inflated, due to compositional biases. Secondly, recent results from the BRCA1 study of Raterman et al. (2006) suggest that increased taxon sampling for the nuclear sequences increases support for the macropodoid–phalangerid association.

The third phalangeridan rooting possibility, possum monophyly (Phalangeriformes), is very strongly rejected in favour of the macropodoid–phalangerid association (see Table 3, groupings a.5 and b.5). This result is particularly interesting, because potential morphological synapomorphies for grouping macropodoids and phalangerids outweigh those for Phalangeriformes. Despite this, morphologists have typically given special status to one basicranial character that favours Phalangeriformes, fusion of the ectotympanic to a pneumatised squamosal (e.g. Flannery, 1987; Springer and Woodburne, 1989; but see Szalay, 1994). The present result adds further weight to the warnings of numerous authors (e.g. Gatesy and O'Leary, 2001; Luo et al., 2001) that special phylogenetic status should not be given to any one character. Moreover, particular caution should be exercised when (as for the present case) lineages have diverged in quick succession, such that the developmental and ecological conditions under which a "rare" transformation occurred are more likely to be retained between successive divergences.

Inferring the basal divergence within Petauroidea presents similar problems as for rooting Phalangerida. Each of the Bayesian and ML analyses favour *Tarsipes* grouping with Petauridae/Pseudocheiridae, although Tarsipedoidea (*Tarsipes* plus Acrobatidae) is only rejected at AU_{Mt21}, $P = 0.237$. Again however, exploring for possible causes of asymmetry in rejection patterns for local rearrangements about the Petauroidea root is instructive and lends further confidence for grouping *Tarsipes* with Petauridae/Pseudocheiridae. ME score differences between BF distance trees (Table 4, groupings 4–6) show a substantial base frequency bias favouring Tarsipedoidea. Hence, the secondary signal in the ML analysis for Tarsipedoidea may be largely explained as an artifact of model misspecification related to compositional non-stationarity.

The Mt21 simulation study reveals an additional model misspecification that may play a role in artifactually inflating support for Tarsipedoidea. Table 4 shows a dramatic shift in support from *Tarsipes*-Petauridae/Pseudocheiridae (as simulated) to Tarsipedoidea under models (rate-homogenous ML and no common mechanism MP) that are unable to account for parallelism that is "hidden" by among-site rate heterogeneity. The implication for our primary phylogenetic analyses is that any additional accounting for parallelism that is required because of among-site rate heterogeneity that the partitioned GTR+I+Γ_4 model ignores would likely further increase support for the presently favoured grouping of *Tarsipes*–Petauridae/Pseudocheiridae.

Tarsipes rostratus is unique among non-volant mammals by feeding almost exclusively on pollen and nectar. The reflection of this specialization in its morphology led Aplin and Archer (1987) to refer to the honey possum as "the paragon of autapomorphic specialization within Diprotodontia". On balance however, these authors favoured a relationship with acrobatids (Tarsipedoidea), particularly citing similarities among bascicranial characters. If however, *Tarsipes* affinities lie with Petauridae/Pseudocheiridae, then those basicranial similarities may be convergent, which Szalay (1994) suggested may be shown by future developmental studies. An alternative homoplasy option should also be considered. The "basal" petauroids (*Tarsipes* and acrobatids) attain less than 10% of the adult body mass of most other petauroids. Hence, if petauroids were small ancestrally, then at least some symplesiomorphic retentions associated with allometric constraints might be expected to link *Tarsipes* and Acrobatids, so providing misleading support for Tarsipedoidea. Again, developmental studies may distinguish between the homoplasy options.

The only statistically well-founded advocacy for Tarsipedoidea has come from the partial RAG-1 nucleotide sequence analysis of Baker et al. (2004). However, 100% bootstrap support from a single highly-conserved nuclear protein-coding sequence looks suspicious, especially given the poor resolution in that same study, for clades that have otherwise been far easier to resolve (e.g. Diprotodontia and Petauridae/Pseudocheiridae). This concern is upheld by our own analyses of RAG-1 sequences (see supplementary information online), for which a basal petauroid tricotomy

is unresolved (Tarsipedoidea received only 25% (MP) and 39% (ML) bootstrap support). Hence, the present ML and Bayesian results provide the strongest resolution yet for basal petauroid affinities. Given the strength of the "non-phylogenetic" biases among the present mt genome data that lend artifactual support to Tarsipedoidea, we expect additional taxon-sampling and concatenated nuclear sequences to cement support for *Tarsipes* affinities lying with Petauridae/Pseudocheiridae to the exclusion of the acrobatids.

Although burramyids could only be included in the Mt29 dataset, the 3174 sites still provide for the largest sequence comparison involving pygmy possums and provide better resolution of their affinities than previous studies have. All placements outside of Phalangerida receive no Bayesian posterior support and are rejected in ML hypothesis testing at $P \leqslant 0.102$ (AU_{Mt29}). More specifically, the weakly favoured placement with Vombatiformes in Kavanagh et al. (2004) and the traditional placement with acrobatids (e.g. Kirsch, 1968) are rejected in the AU tests at $P = 0.042$ and 0.002, respectively. Instead, the Bayesian and ML analyses favour burramyids as sister to a Macropodoidea–Phalangeridae clade (with the isolated *Dromiciops* included, e.g. Fig. 2) or with Phalangeridae specifically (with *Dromiciops* excluded, not shown). The latter of these hypotheses may now be regarded as the best estimate for burramyid affinities, with support also coming from DNA hybridization (Springer and Kirsch, 1991; Kirsch et al., 1997) and from the only nuclear loci (RAG-1) to be published for burramyids (Baker et al., 2004).

It is a curious irony that Flannery's (1987) primary argument against considering blade-like third premolars of macropodoids and phalangerids as synapomorphic was the possession of a similar condition in *Burramys*. It now appears that such secodont premolars may in fact be synapomorphic, at least among modern diprotodontians, for the more inclusive grouping of macropodoids, phalangerids, and burramyids.

A general implication of this study is the importance of taxon sampling, contra Rosenberg and Kumar (2001). Comparing clade support in the present study and in those less densely sampled studies that the Mt21 dataset largely derives from illustrates this point. Here Phalangerida and Petauroidea provide benchmark examples, because MtNuc21 AU-testing essentially places them beyond doubt (all H_a $P < 0.01$). Returning to the previous mt genome studies, Munemasa et al. (2006) was unable to consistently resolve for monophyly of either grouping and indeed, Nilsson et al. (2004) favours non-monophyly of both groupings. Combining the mt sequences from these studies together in the Mt21 dataset provides robust support for both groupings even without the inclusion of the nuclear sequences (Fig. 3). The conclusions from the Mt21 results also stand without the modifications made to Mt21 (RNA inclusion and RY-coding protein 3rd codon positions) relative to the data treatments of Nilsson et al. (2004) and Munemasa et al. (2006). It is likely that the major reason for improved phylogenetic performance here with denser taxon sampling (and a caveat to the simulation study of Rosenberg and Kumar, 2001) is that breaking up long branches becomes increasingly valuable when available models of sequence evolution do not closely match reality, such as under non-stationarity.

Munemasa et al. (2006) were pessimistic about the potential for molecular sequences to resolve the family-level diprotodontian tree and implied that other data, such as SINEs (e.g. Shedlock and Okada, 2000) would be necessary. Efforts to gather retroposition and other "rare event" data from marsupial genomes should certainly be encouraged. However, the present study shows the value of combining mt and nuclear data, of considering potential phylogenetic biases and of breaking up long branches with additional taxon sampling; only one other placement for *Tarsipes* relative to acrobatids and at most 5 other placements for burramyids remain as plausible alternatives to the favoured diprotodontian tree (Fig. 2). Here it is relevant that Tarsipedidae, Acrobatidae and Burramyidae (along with Hypsiprymnodontidae) are the only families from which the nuclear portion of the data matrix was missing. Sequencing these nuclear loci as well as complete mt genomes for Burramyidae will likely provide a fully resolved diprotodontian family-level phylogeny.

A fascinating feature of the diprotodontian tree is that it reveals long-term maintenance of ecological niche differentiation. The Vombatiformes are large (adults > 7 kg), strict herbivores that can typically subsist on lower quality forage than their phalangeridan contemporaries (Cork and Sanson, 1990; Hume and Barboza, 1998). Wombats and koalas have low field metabolic rates (Hume, 1999) and low birth rates (summarized in Tyndale-Biscoe, 2005). These are K-selected features that might be considered typical of "survivors", to borrow a term that Grime (1974) used in much the same sense in his classification of plant ecological strategies, for species that out-endure, rather than directly out-compete or out-disperse for ecological leverage in resource competition.

Within Phalangerida, the petauroids are all small (adults < 1.5 kg) and forage arboreally. This is consistent with all three independent origins of gliding among marsupials occurring within this superfamily. It may be inferred that the other (secodont) phalangeridan clade, Macropodoidea–Phalangeridae–Burramyidae, retained more evolutionary flexibility among its early crown-group members than did Vombatiformes (for diet and size) and Petauroidea (for terrestriality and size). Modern members of the secodont clade include some of the smallest and the largest living marsupials. Many species are rather generalist in diet and are scansorial to varying degrees (e.g. brushtail and pygmy possums). Indeed, even the more specialized members tend to occupy ecospace that lies within bounds "set" by, or intermediate between vombatiform and petauroid marsupials. For example, tree kangaroos, which exploit spatially heterogeneous rainforest food resources (Newell, 1999) have higher metabolic rates and agility than

Vombatiformes, while being larger and better terrestrially adapted than petauroid folivores.

In concert with progress in resolving diprotodontian phylogeny, new evolutionary questions are being brought into view, such as the mechanisms involved in maintaining ecological niche differentiation among suprafamilial groups whose divergences trace back over 40 million years (Nilsson et al., 2004; Drummond et al., 2006). Older questions too, are reinvigorated, such as the morphological nature of the diprotodontian most recent common ancestor, and the placement of enigmatic fossil diprotodontians (e.g. marsupial lions). Having a robust tree itself is crucial for improving the accuracy of phylogenetic inference on these and other questions and should motivate further investigation into the evolution of this arguably most ecologically diverse of mammalian orders.

Acknowledgments

This work has been enriched by discussions with the vertebrate evolution group at the Allan Wilson Centre. Richard Carter, Andrew Clarke and two anonymous reviewers provided useful comments on the manuscript and the work has been supported by the Marsden Fund of New Zealand and an Australian Research Council grant to M.J.P.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2007.09.008.

References

Amrine-Madsen, H., Scally, M., Westerman, M., Stanhope, M.J., Krajewski, C., Springer, M.S., 2003. Nuclear gene sequences provide evidence for the monophyly of australidelphian marsupials. Mol. Phylogenet. Evol. 28, 186–196.

Aplin, K.P., Archer, M., 1987. Recent advances in marsupial systematics with a new syncretic classification. In: Archer, M. (Ed.), Possums and Opossums: Studies in Evolution. Surrey Beatty & Sons, Chipping Norton, Nsw, pp. Xv–Lxxii.

Archer, M., Tedford, R.H., Rich, T.H., 1987. The Pilkipildridae, a new family and four new species of ?petauroid possums (Marsupialia: Phalangerida) from the Australian Miocene. In: Archer, M. (Ed.), Possums and Opossums: Studies in Evolution. Surrey Beatty & Sons, Chipping Norton, Nsw, pp. 607–627.

Asher, R.J., Horovitz, I., Sánchez-Villagra, M.R., 2004. First combined cladistic analysis of marsupial mammal interrelationships. Mol. Phylogenet. Evol 33, 240–250.

Baker, M.L., Wares, J.P., Harrison, G.A., Miller, R.D., 2004. Relationships among the families and orders of marsupials and the major mammalian lineages based on recombination activating gene-1. J. Mamm. Evol. 11, 1–16.

Baverstock, P.R., 1984. The molecular relationships of Australasian possums and gliders. In: Smith, A.P., Hume, I.D. (Eds.), Possums and Gliders. Australian Mammal Society, Sydney, pp. 1–8.

Baverstock, P.R., Krieg, M., Birrell, J., 1990. Evolutionary relationships of Australian marsupials as assessed by Albumin immunology. Aust. J. Zool. 37, 273–287.

Cardillo, M., Bininda-Emonds, O.R.P., Boakes, E., Purvis, A., 2004. A species-level phylogenetic supertree of marsupials. J. Zool. (Lond.) 264, 11–31.

Caterino, M.S., Reed, R.D., Kuo, M.M., Sperling, F.A.H., 2001. A partitioned likelihood analysis of swallowtail butterfly phylogeny (Lepidoptera: Papilionidae). Syst. Biol. 50, 106–127.

Cavender, J.A., Felsenstein, J., 1987. Invariants of phylogenies in a simple case with discrete states. J. Classif. 4, 57–71.

Cork, S.J., Sanson, G.D., 1990. Digestion and nutrition in the koala: a review. In: Lee, A.k., Handasyde, K.A., Sanson, G.D. (Eds.), Biology of the Koala. Surrey Beatty, Sydney, pp. 129–144.

Drummond, A.J., Ho, S.Y.W., Phillips, M.J., Rambaut, A., 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4, 699–710.

Edwards, D., Westerman, M., 1995. The molecular relationships of possum and glider families as revealed by DNA–DNA hybridisations. Aust. J. Zool. 43, 231–240.

Felsenstein, J., 2004. Inferring Phylogenies. Sinauer, Sunderland.

Fiala, K.L., Sokal, R.R., 1985. Factors determining the accuracy of cladogram estimation: evaluation using computer simulation. Evolution 39, 609–622.

Flannery, T.F., 1987. The relationships of the macropodoids (Marsupialia) and the polarity of some morphological features within the Phalangeriformes. In: Archer, M. (Ed.), Possums and Opossums: Studies in Evolution. Surrey Beatty & sons, Sydney, pp. 741–747.

Gatesy, J., O'Leary, M.A., 2001. Deciphering whale origins with molecules and fossils. Trends Ecol. Evol. 16, 562–570.

Gibson, A., Gowri-Shankar, V., Higgs, P.G., Rattray, M., 2005. A comprehensive analysis of mammalian mitochondrial genome base composition and improved phylogenetic methods. Mol. Biol. Evol. 22, 251–264.

Gill, T., 1872. Arrangement of the families of mammals with analytical tables. Smithsonian Misc. Coll. 11, 1–98.

Gontcharov, A.A., Marin, B., Melkonian, M., 2004. Are combined analyses better than single gene phylogenies? A case study using SSU rDNA and rbcL sequence comparisons in the Zygnematophyceae (Streptophyta). Mol. Biol. Evol. 21, 612–624.

Goremykin, V.V., Holland, B.R., Hirsch-Ernst, K.I., Hellwig, F.H., 2005. Analysis of *Acorus calamus* chloroplast genome and its phylogenetic implications. Mol. Biol. Evol. 22, 1813–1822.

Grime, J.P., 1974. Vegetation classification by reference to strategies. Nature 250, 26–31.

Gunson, M.M., Sharman, G.B., Thomson, J.A., 1968. The affinities of *Burramys* (Marsupialia: Phalangeroidea) as revealed by a study of its chromosomes. Aust. J. Sci. 31, 40–41.

Harding, H.R., Carrick, F.N., Shorey, C.D., 1984. Sperm ultrastructure and development in the honey possum, *Tarsipes rostratus*. In: Smith, A.P., Hume, I.D. (Eds.), Possums and Gliders. Australian Mammal Society, Sydney, pp. 451–461.

Horovitz, I., Sánchez-Villagra, M.R., 2003. A morphological analysis of marsupial mammal higher-level phylogenetic relationships. Cladistics 19, 181–212.

Huelsenbeck, J., Ronquist, F., 2001. MrBayes: Bayesian inference on phylogenetic trees. Bioinformatics 17, 754–755.

Hughes, R.L., 1974. Morphological studies on implantation in marsupials. J. Reprod. Fert. 39, 173–186.

Hume, I.D., Barboza, P.S., 1998. The gastrointestinal tract and digestive physiology of wombats. In: Wells, R.T., Pridmore, P.A. (Eds.), Wombats, Surrey Beatty. Chipping Norton, NSW, pp. 67–74.

Hume, I.D., 1999. Marsupial Nutrition. Cambridge University Press, Cambridge.

Kavanagh, J.R., Burk-Herrick, A., Westerman, M., Springer, M.S., 2004. Relationships among families of Diprotodontia (Marsupialia) and the phylogenetic position of the autapomorphic honey possum (*Tarsipes rostratus*). J. Mamm. Evol. 11, 207–222.

Kelcher, S.A., Thomas, M.A., 2006. Model use in phylogenetics: nine key questions. Trends Ecol. Evol. 22, 87–94.

Kirsch, J.A.W., 1968. Prodromus of the comparative serology of Marsupialia. Nature 217, 418–420.

Kirsch, J.A.W., 1977. The comparative serology of Marsupialia, and the classification of marsupials. Aust. J. Zool. (Suppl.) 38, 1–152.

Kirsch, J.A.W., Lapointe, F.-J., Springer, M.S., 1997. DNA-hybridisation studies of marsupials and their implications for metatherian classification. Aust. J. Zool 45, 211–280.

Kishino, H., Hasegawa, M., 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data and the branching order Hominidae. J. Mol. Evol. 29, 170–179.

Harrison, G.L., McLenachan, P.A., Phillips, M.J., Slack, K.L., Penny, D., 2004. Four new avian mitochondrial genomes help get to basic evolutionary questions in the Late Cretaceous. Mol. Biol. Evol. 21, 974–983.

Luckett, W.P., 1994. Suprafamilial relationships within Marsupialia: resolution and discordance from multidisciplinary data. J. Mamm. Evol. 2, 255–288.

Luo, Z.-X., Cifelli, R., Kielan-Jaworowska, Z., 2001. Dual origin of tribosphenic mammals. Nature 409, 53–57.

Luo, Z.-X., Ji, Q., Wible, J.R., Yuan, C.-X., 2003. An early Cretaceous tribosphenic mammal and metatherian evolution. Science 302, 1934–1940.

Marshall, L.G., Case, J.A., Woodburne, M.O., 1990. Phylogenetic relationships of the families of marsupials. In: Genoways, H.H. (Ed.), Current Mammalogy. Plenum Press, New York, pp. 433–505.

Munemasa, M., Nikaido, M., Donnellan, S., Austin, C.C., Okada, N., Hasegawa, M., 2006. Phylogenetic analysis of diprotodontian marsupials based on complete mitochondrial genomes. Genes Genet. Syst. 81, 181–191.

Newell, G.R., 1999. Home range and habitat use by Lumholtz's tree kangaroo (Dendrolagus Lumholtzi) within a rainforest fragment in North Queensland. Wildlife Res. 26, 129–145.

Nilsson, M., Arnason, U., Spencer, P.B.S., Janke, A., 2004. Marsupial relationships and a timeline for marsupial radiation in South Gondwana. Gene 340, 189–196.

Osborne, M.J., Christidis, L., Norman, J.A., 2002. Molecular Phylogenetics of the Diprotodontia (kangaroos, wombats, koala, possums, and allies). Mol. Phylogenet. Evol. 25, 219–228.

Pearson, J., 1950. The relationships of the Potoroidae to the Macropodidae (Marsupialia). Pap. Proc. Roy. Soc. Tas. 1949, 211–229.

Phillips, M.J., Lin, Y.-H., Harrison, G.L., Penny, D., 2001. Mitochondrial genomes of a bandicoot and a brushtail possum confirm the monophyly of australiadelphian marsupials. Proc. Roy. Soc. Lond. B 268, 1533–1538.

Phillips, M.J., Penny, D., 2003. The root of the mammalian tree inferred from whole mitochondrial genomes. Mol. Phylogenet. Evol. 28, 171–185.

Phillips, M.J., Delsuc, F., Penny, D., 2004. Genome-scale phylogeny and the detection of systematic biases. Mol. Biol. Evol. 21, 1455–1458.

Phillips, M.J., MacLenachan, P.A., Down, C., Gibb, G.C., Penny, D., 2006. Combined mitochondrial and nuclear protein-coding DNA sequences resolve the Interrelations of the major Australasian marsupial radiations. Syst. Biol. 55, 122–137.

Posada, D., Crandall, K.A., 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14, 817–818.

Pupko, T., Huchon, D., Cao, Y., Okada, N., Hasegawa, M., 2002. Combining multiple data sets in a likelihood analysis: which models are the best? Mol. Biol. Evol. 19, 2294–2307.

Rambaut, A., 1996. Sequence Alignment Editor, Oxford University (http://evolve.zps.ox.ac.uk/software/index.html).

Rambaut, A., Grassly, N.C., 1997. Seq-Gen: an application for the Monte-Carlo simulation of DNA sequence evolution along phylogenetic trees. Comp. App. Biosci. 13, 235–238.

Raterman, D., Meredith, R.W., Reudas, L.A., Springer, M.S., 2006. Phylogenetic relationships of the cuscuses and brushtail possums (Marsupialia: Phalangeridae) using the nuclear gene BRCA1. Aust. J. Zool. 54, 353–361.

Roberts, R.G., Flannery, T.F., Ayliffe, L.K., Yoshida, H., Olley, J.M., Prideaux, G.J., Laslett, G.M., Baynes, A., Smith, M.A., Jones, R., Smith, B.L., 2001. New ages for the last Australian megafauna: continent-wide extinction about 46,000 years ago. Science 292, 1888–1892.

Rosenberg, M.S., Kumar, S., 2001. Incomplete taxon sampling is not a problem for phylogenetic inference. Proc. Natl. Acad. Sci. USA 98, 10751–10756.

Schmitz, J., Ohme, M., Zischler, H., 2002. The complete mitochondrial sequence of *Tarsius bancanus*: evidence for extensive nucleotide compositional plasticity of primate mitochondrial DNA. Mol. Biol. Evol. 19, 544–553.

Shedlock, A.M., Okada, N., 2000. SINE insertions: powerful tools for molecular systematics. Bioessays 22, 148–160.

Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16, 1114–1116.

Shimodaira, H., Hasegawa, M., 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17, 1246–1247.

Shimodaira, H., 2002. An approximately unbiased test of phylogenetic tree selection. Syst. Biol. 51, 492–508.

Springer, M.S., Woodburne, M.O., 1989. The distribution of some basicranial characters within the Marsupialia and the phylogeny of the Phalangeriformes. J. Vert. Paleont. 9, 210–221.

Springer, M.S., Kirsch, J.A.W., 1991. DNA hybridization, the compression effect and the radiation of diprotodontian mammals. Syst. Zool. 40, 131–151.

Sullivan, J., Swofford, D.L., 1997. Are guinea pigs rodents? The importance of adequate models in molecular phylogenetics. J. Mammal. Evol. 4, 77–86.

Suzuki, Y., Glazko, G.V., Nei, M., 2002. Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. Proc. Natl. Acad. Sci. USA 99, 16138–16143.

Swofford, D.L., 2002. PAUP*. Phylogenetic analysis using parsimony (* and other methods). Sinaur Associates, Sunderland, MA.

Szalay, F.S., 1994. Evolutionary History of the Marsupials and an Analysis of Osteological Characters. Cambridge University Press, New York.

Tamura, K., Nei, M., 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. 10, 512–526.

Tyndale-Biscoe, H., 2005. Life of Marsupials. CSIRO Publishing, Collingwood, Vic.

Waddell, P.J., 1995. Statistical methods of phylogenetic analysis: including Hadamard conjugations, LogDet transforms and maximum likelihood. Ph.D. Thesis, Massey University, Palmerston North, New Zealand.

Woodburne, M.O., Case, J.A., 1996. Dispersal, vicariance, and the Late Cretaceous to Early Tertiary land mammal biogeography from South America to Australia. J. Mamm. Evol. 3, 121–161.

Yang, Z., 1996. Maximum-likelihood models for combined analyses of multiple sequence data. J. Mol. Evol. 42, 587–596.