

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.

DNA Barcoding the Birds of New Zealand

A thesis presented in fulfillment of the requirements for the degree of
Doctor of Philosophy in Molecular BioSciences
at Massey University, Auckland, New Zealand

William John Waugh

2011

I dedicate this thesis to my parents Molly and Charles Waugh

both of whom died before it was complete.

I know that my achieving this degree would have

given them great pleasure.

No matter how much I imagined I would miss their

love, guidance and friendship,

nothing prepared me for the reality of the loss that I feel.

I also dedicate this thesis to my wife Fiona

who gave me the opportunity to undertake it and

the love and support needed to complete it.

Abstract

A comprehensive inventory of the life forms on earth is at the heart of any scientific study of evolution and biodiversity. The international "Barcode of Life" project is an attempt to identify the earth's biodiversity, at the species level, using short signature DNA sequences. The hypothesis underlying DNA barcoding is being comprehensively tested in different taxa. A database was constructed of DNA sequences from part of the mitochondrial gene cytochrome c oxidase subunit 1 for the avian fauna of New Zealand. To date, 833 sequences from 215 species have been added to this database, of which 628 sequences from 126 species are from native or endemic birds. This represents an average of 5 samples per species (minimum 1, maximum 18) for the latter group, which is the central focus of this thesis. Samples of species, from different geographical locations throughout New Zealand, have been collected to highlight any intraspecific nucleotide variation that may occur. Some samples analysed here were from historical specimens housed in museum collections and required specialised DNA extraction and amplification. These techniques were developed as part of the project and provide a means of collecting DNA barcodes where no modern material is available. In general, DNA barcoding proved effective at identifying avian species in New Zealand. However, some species were highlighted that contained distinct DNA barcode clusters, indicative of possible subspecies or cryptic species while in other cases two or more species that appear to be different share very similar DNA barcodes. Remains from aircraft birdstrikes were identified using this technique in order to inform wildlife management at airports around New Zealand. A review of and outlook for the uses of this technique are given.

Acknowledgements

Firstly, I thank my supervisors Professor David Lambert and Dr Craig Millar for providing me with a very interesting PhD project. This research has not only challenged me intellectually but also forms part of what is becoming regarded as an effort to speed the identification of species and, thereby increase global biological literacy. As someone who has taught biosciences for many years, this is a project tailored to augment my life's work. I thank them for providing guidance and resources, discussions and assistance with manuscripts. I also thank them for obtaining funding for this project and relieving me of the burdens relating to financial issues. I thank the funding bodies, the Allan Wilson Centre and Massey University for providing a stipend and tuition fees for this PhD.

I am indebted to a number people and institutions for providing samples for this project. Sylvia Durant provided many fresh samples from both live and dead specimens, as did Rosemary Tulley, Della Bennet, Bill Pohatu, Jon Eyely, Sondra Taylor, Nick Petkov and Phil Gotlieb. In addition, I thank my supervisors for providing many fresh bird samples from their collections as well as Brian Gill of Auckland War Memorial Museum, Ian Jamieson from Otago Museum and Janeatte Norman of Victoria Museum. Particular thanks to Paul Scofield for providing preserved samples and for assisting me to collect them from the Canterbury Museum's collection.

I am most grateful to Andrew Dodd, Jennie Hay, Jennifer Anderson and Leon Huynen for introducing me to the mysteries of molecular genetics and for their patience and forbearance with me during the early part of my research. Their wisdom and help has served me well throughout this project. I am further indebted to Andrew Dodd and Tamara Sirey for their help in all matters pertaining to data analysis and computing technology. My own levels of understanding in these matters have been elevated to a completely different plane thanks to their unstinting help and guidance. Their tolerance of nothing but the very highest standards has provided me with the best possible example.

I thank David Lambert, Craig Millar, Selina Patel and Allan Baker for all their work helping to build the New Zealand Birds DNA barcode database and for sharing their insights and wisdom with me on everything from the minutia of molecular techniques to the broad strategies for developing this project.

I am grateful to Paul Hebert, Mark Stoeckle, Sujeevan Ratnasingham, Gregory Downs, Riadul Mannan Riad, Janet Topan, Megan Milton, Brianne St. Jacques, Claudia Kleint-Steinke, Natalia Ivanova, Andras Papp, Alex Borisenko, Taika von Königslöw and others in Canada at the DNA barcoding facilities at the University of Guelph, Ontario for helping to develop the Birds of New Zealand project and in assisting in building the database that now resides on the Barcode of Life Data Systems (BOLD) website.

When examining specimens from aircraft birdstrikes, I received invaluable help and data from Max Evans of the New Zealand Civil Aviation Authority, advice and the opportunity to attend a Flight Safety Forum, which he chaired, from John Martin of Air Nelson, birdstrike samples from Della Bennet, Wildlife Control Officer at Christchurch International Airport and, Jon Eyely, Airside Operations Manager as well as Bill Pohatu, Sondra Taylor, Nick Petkov and Phil Gotlieb, Wildlife Control Officers at Wellington International Airport. I also received much helpful advice from Carla Dove of the Smithsonian Institute in Washington, USA.

I thank my colleagues, past and present, at the Allan Wilson Centre/Institute of Natural Sciences in Albany: Hayley Lawrence, Andrew Dodd, Jennie Hay, Leon Huynen, Elmira Mohandesan, Martina Dautel, Katie Hartnup, Gabrielle Beans-Pecon, Jyothsna Viswesweriah, Andrew Cridge, Chris Rodley, Tim Heupink, Bill Peacock, Barbara Binney, Subashchandran Sankarasubaramanian, Muharam Khoussainova, Bupanelo Mbuba, Sujung Lee, Jarod Young, Saumya Agrawal, Diane Brunton, Gabriele Schmidt-Adam, Justin O'Sullivan, Evelyn Sattlegger, Wayne Patrick and others too numerous to mention for their companionship, conversation and company as well as for their help, guidance and support throughout this

research.

Finally I thank my wife Fiona and my two daughters Sarah and Rebecca for providing me with much love, tolerance and support throughout the years of this research, without which it would never have been completed

Thesis Structure, Financial Support and Regulatory Compliance

This thesis begins with a general introduction (Chapter 1), which is a review of DNA barcoding and provides the background and intellectual framework that underpins the thesis. It is an update of a review paper published in BioEssays. Chapter 2 is a general methods section outlining the collection and processing protocols associated with the project. Chapters 3 and 6 have been written as stand-alone scientific papers. Therefore some information provided in the introduction will be briefly outlined again in these chapter introductions. Chapter 4 is an analysis of the data. Chapter 5 outlines and discusses interesting anomalies within the data and Chapter 7 discusses the variety of uses that DNA barcoding has and may be put to and the conclusions and applications of the research findings and potential future research.

Financial Support

Funding for this project was provided by: the Allan Wilson Centre for Molecular Ecology and Evolution, Massey University and the University of Auckland. Personal financial support was provided through a Massey University Doctoral scholarship and an Allan Wilson Centre scholarship.

Ethics Approval

This research had ethics approval from the New Zealand Department of Conservation (DOC) Animal Ethics Committee for sampling.

Table of Contents

Abstract		i
Acknowledgements		iii
Thesis Structure, Financial Support and Regulatory Compliance		vii
Table of Contents		ix
List of Figures		xiii
List of Tables		xvii
Chapter 1	General Introduction to DNA barcoding	1
1.1	A brief history of modern taxonomy	1
1.2	The size of the taxonomic task	3
1.3	Potential solutions to the taxonomic impediment	4
1.4	DNA barcoding	5
1.5	A review of the efficacy of DNA barcoding in published studies	8
	1.5.1 <i>Efficacy at identifying species</i>	16
	1.5.2 <i>Problematic taxa</i>	18
1.6	Discussion	19
1.7	Conclusion and outlook	24
1.8	The scope of this thesis	24
1.9	References	26
Chapter 2	Methods	37
2.1	Overview of sample collection	37
2.2	Samples obtained for this study	38
2.3	Tissue collection and recording	39
2.4	DNA extraction, amplification and sequencing of fresh samples	42
2.5	DNA extraction, amplification and sequencing of historical samples	43
2.6	Analysis of sequence data	45
2.7	References	48

Chapter 3	Primer Design	49
3.1	An overview of primer design	49
3.2	Primer design for fresh samples from New Zealand Birds	50
	3.2.1 <i>Introduction</i>	50
	3.2.2 <i>Method</i>	50
3.3	Primers for historical samples from New Zealand birds	53
	3.3.1 <i>Introduction</i>	53
	3.3.2 <i>Method</i>	55
3.4	Results	56
3.5	Discussion	59
3.6	References	62
Chapter 4	Data Analysis	65
4.1	Overview of data	65
4.2	Data analysis methods	68
	4.2.1 <i>Kimura 2-Parameter distance correction</i>	68
	4.2.2 <i>Bayesian inference of phylogeny</i>	71
4.3	Results	77
	4.3.1 <i>Anseriformes</i>	82
	4.3.2 <i>Apterygiformes</i>	85
	4.3.3 <i>Chradriiformes</i>	88
	4.3.4 <i>Ciconiformes, Falconiformes and Gruiformes</i>	91
	4.3.5 <i>Passeriformes</i>	95
	4.3.6 <i>Pelecaniformes</i>	101
	4.3.7 <i>Procellariiformes</i>	103
	4.3.8 <i>Psittaciformes</i>	108
	4.3.9 <i>Sphenisiformes</i>	111
4.4	References	114
Chapter 5	Ambiguous Results	117
5.1	Overview of ambiguous results	117
5.2	Species that contain DNA barcode clusters	119
	5.2.1 <i>Acanthasitta chloris</i>	119
	5.2.2 <i>Petroica australis</i>	120
	5.2.3 <i>Eudyptula minor</i>	122
	5.2.4 <i>Pelagodroma marina</i>	126
	5.2.5 <i>Pterodroma axillaris</i>	127
	5.2.6 <i>Eudyptes chrysocome</i>	128

5.3	Species that are indistinguishable using DNA barcodes	130
	5.3.1 <i>Apteryx</i> species	130
	5.3.2 <i>Haematopus</i> species	134
	5.3.3 <i>Thalassarche melanophris</i> and <i>T. chrysostoma</i>	139
	5.3.4 <i>Larus</i> species	140
	5.3.5 <i>Puffinus assimilis</i> , <i>P. gavia</i> and <i>P. griseus</i>	143
	5.3.6 <i>Cyanorhamphus unicolor</i> and <i>C. navaezelandiae</i>	145
5.4	Discussion	147
5.5	References	151
Chapter 6	Aircraft Birdstrike	155
6.1	Introduction to birdstrike	155
6.2	Method	159
6.3	Results	161
	6.3.1 <i>Birdstrike statistics</i>	161
	6.3.2 <i>DNA barcoding</i>	163
6.4	Discussion	164
6.5	References	172
Chapter 7	General Conclusion	175
7.1	Introduction	175
7.2	Present and future applications of DNA barcoding	178
	7.2.1 <i>Accurate species identification</i>	178
	7.2.2 <i>verification of claims made about biological material</i>	179
	7.2.3 Biological data collection	180
7.3	Progress, potential and pitfalls	181
7.4	References	189

Appendix A	Papers published from this study	199
	1. Waugh, J., <i>DNA barcoding in animal species: progress, potential and pitfalls</i> . Bioessays, 2007. 29 (2): p. 188-197.	201
	2. Waugh, J., Huynen, L, Millar, C, Lambert, D, <i>DNA barcoding of animal species - response to DeSalle</i> . Bioessays, 2008. 30 (1): p. 92-93.	211
	3. Patel, S., Waugh, J., Millar, C.D., & Lambert, D.M., <i>Conserved primers for DNA barcoding historical and modern avian samples</i> . Molecular Ecology Resources, 2009. 10, 431–438	215
	4. Waugh, J., Evans, MW, Millar, CD, Lambert, DM, <i>Birdstrikes and barcoding: can DNA methods help make the airways safer?</i> Molecular Ecology Resources, 2010. doi: 10.1111/j.755-0998.2010.02884.x.	225
Appendix B	BOLD submission sheets	233
Appendix C	BOLD published primer list	239
Appendix D	GenBank accession numbers	247
Appendix E	Barcoding New Zealand Birds project list	251

List of Figures

Chapter 1

1.1	The mitochondrial genome of a Eukaryote	6
1.2	A diagrammatic representation of cytochrome c oxidase subunit I (COI)	7

Chapter 2

2.1	An example of a data form from the database of birds sampled.	40
2.2	Part of the Barcoding New Zealand Birds project on the BOLD website	41
2.3	Amplification product of the DNA barcode region of the COI gene	43
2.4	A data.xls sheet for uploading ab1 or .scf and .phd files to the BOLD website	46
2.5	Summary page for the Barcoding NZ Birds project on the BOLD website	47

Chapter 3

3.1	Part of the mass alignment used to locate suitable primer binding sites	51
3.2	A diagrammatic representation of the COI barcoding region	55

Chapter 4

4.1	An asymptotic divergence with time in the evolution of DNA	69
4.2	The Jukes-Cantor correction to distance model	70
4.3	Flow chart of the series of steps of the MCMC algorithm	73
4.4	A sequence record for a <i>Puffinus griseus</i> specimen on the BOLD website	75
4.5	A specimen data record for a <i>Puffinus griseus</i> specimen on the BOLD website	76
4.6	New Zealand Anseriformes K2P neighbour joining tree	83
4.7	New Zealand Anseriformes Bayesian tree	84
4.8	New Zealand Apterygiformes K2P neighbour joining tree	86
4.9	New Zealand Apterygiformes Bayesian tree	87
4.10	New Zealand Charadriiformes K2P neighbour joining tree	89
4.11	New Zealand Charadriiformes Bayesian tree	90
4.12	New Zealand Ciconiiformes K2P neighbour joining tree	91
4.13	New Zealand Ciconiiformes Bayesian tree	91
4.14	New Zealand Falconiformes K2P neighbour joining tree	92
4.15	New Zealand Falconiformes Bayesian tree	92
4.16	New Zealand Gruiformes K2P neighbour joining tree	93
4.17	New Zealand Gruiformes Bayesian tree	94
4.18	New Zealand Passeriiformes K2P neighbour joining tree	97

4.18	New Zealand Passeriiformes K2P neighbour joining cont'd	98
4.19	New Zealand Passeriiformes Bayesian tree	99
4.19	New Zealand Passeriiformes Bayesian tree cont'd	100
4.20	New Zealand Pelecaniformes K2P neighbour joining tree	101
4.21	New Zealand Pelecaniformes Bayesian tree	102
4.22	New Zealand Procellariiformes K2P neighbour joining tree	104
4.22	New Zealand Procellariiformes K2P neighbour joining tree cont'd	105
4.23	New Zealand Procellariiformes Bayesian tree	106
4.23	New Zealand Procellariiformes Bayesian tree cont'd	107
4.24	New Zealand Psittaciformes K2P neighbour joining tree	109
4.25	New Zealand Psittaciformes Bayesian tree	110
4.26	New Zealand Sphenisciformes K2P neighbour joining tree	112
4.27	New Zealand Sphenisciformes Bayesian tree	113

Chapter 5

5.1	<i>Acanthisitta chloris</i> K2P neighbour joining tree	119
5.2	K2P distance summary for <i>Acanthisitta chloris</i>	120
5.3	<i>Petroica australis</i> K2P neighbour joining tree	121
5.4	Distance summary for <i>Petroica australis</i>	122
5.5	<i>Eudyptula minor</i> K2P neighbour joining tree	124
5.6	Distance summary for <i>Eudyptula minor</i>	125
5.7	An alignment of <i>Eudyptula minor</i> specimens	126
5.8	<i>Pelagodroma marina</i> K2P neighbour joining tree	127
5.9	Distance summary for <i>Pelagodroma marina</i>	127
5.10	<i>Pterodroma axillaris</i> K2P neighbour joining tree	127
5.11	Distance summary for <i>Pterodroma axillaris</i>	128
5.12	<i>Eudyptes chrysocome</i> K2P neighbour joining tree	129
5.13	Distance summary for <i>Eudyptes chrysocome</i>	130
5.14	New Zealand Apterygiformes K2P neighbour joining tree	131
5.15	Distance summaries for 47 <i>Apteryx</i> species	132
5.16	Nearest neighbour analyses for <i>Apteryx</i> species	133
5.17	<i>Haematopus finschi</i> (a), <i>H. unicolor</i> (b) and <i>H. chathamensis</i> (c)	135
5.18	New Zealand <i>Haematopus</i> species K2P neighbour joining tree	136
5.19	Nearest neighbour analysis for New Zealand <i>Haematopus</i> species.	137
5.20	Distance summary for New Zealand <i>Haematopus</i> species	137
5.21	The results of a nearest neighbour search for <i>Haematopus</i> species	138
5.22	<i>Thalassarche chrysostoma</i> and <i>T. melanophrys</i> K2P neighbour joining tree	139
5.23	Distance summary for <i>Thalassarche chrysostoma</i> and <i>T. melanophrys</i>	139

5.24	Nearest neighbour analysis for <i>Thalassarche chrysostoma</i> and <i>T.melanophrys</i>	140
5.25	<i>Larus novaehollandiae</i> (a) and <i>L. bulleri</i> (b)	140
5.26	New Zealand <i>Larus</i> species K2P neighbour joining tree	141
5.27	Distance summary for <i>Larus novaehollandiae</i> and <i>L. bulleri</i>	142
5.28	Nearest neighbour analysis for <i>Larus novaehollandiae</i> and <i>L. bulleri</i>	142
5.29	<i>Puffinus assimilis</i> , <i>P. gavia</i> and <i>P. griseus</i> K2P neighbour joining tree	144
5.30	Distance summary for <i>Puffinus assimilus</i> , <i>P. gavia</i> and <i>P. griseus</i>	144
5.31	Nearest neighbour analysis for <i>Puffinus assimilus</i> , <i>P. gavia</i> and <i>P. griseus</i>	145
5.32	<i>Cyanorhamphus</i> species K2P neighbour joining tree	146
5.33	Distance summary for <i>Cyanorhamphus novaezelandiae</i> and <i>C. unicolor</i>	146
5.34	Nearest neighbour analysis for <i>Cyanorhamphus novaezelandiae</i> and <i>C. unicolor</i>	147

Chapter 6

6.1	Boeing 737 passenger jet striking a flock of more than 200 starlings	156
6.2	Sample collection kit	159
6.3	Total annually reported birdstrikes from airports in New Zealand 2003-2009	161
6.4	The species most highly represented in reported birdstrikes in New Zealand	162
6.5	Unidentified birds involved in birdstrike in New Zealand 2003 - 2009	163
6.6	US Airways flight 1549 after making a forced landing in the Hudson River	166

Chapter 7

7.1	Stages of development of a major biological project such as the HGP	183
-----	---	-----

List of Tables

Chapter 1

- 1.1 Design details and major results from selected DNA barcode studies 9

Chapter 2

- 2.1 Sample data by category entered in a database for each sample sequenced 40

Chapter 3

- 3.1 Table used during testing of the primers for the Barcoding NZ Birds project 52
3.2 PCR primers and 5' tags used for the amplification of the 648 bp COI region 57
3.3 Results of PCR success of modern and historical samples 58
3.4 PCR primer success on 17 different avian orders 59

Chapter 4

- 4.1 An example of Nexus file format 74
4.2 Native and endemic bird species with number of samples obtained for each 77
4.3 Intraspecific divergence within Columbiformes, Coraciiformes and Strigiformes 82

Chapter 5

- 5.1 The six subspecies that have been recognised in *Eudiptula minor* 123
5.2 K2P distance measures for 57 specimens of *Eudiptula minor* 126

Chapter 6

- 6.1 Birdstrike samples identified using DNA barcoding 165

Chapter 7

- 7.1 Published examples of current usage of DNA barcoding 178
7.2 A summary of campaigns overseen by the Consortium for the Barcode of Life 185