

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.

**Evolutionary Genetics and the Major
Histocompatibility Complex of New Zealand Robins
(Petroicidae)**

Hilary C. Miller

A thesis submitted for the degree of Doctor of Philosophy in
Molecular BioSciences at Massey University, New Zealand
June 2003



Photo: Don Merton

The founding black robin pair,
Old Blue (above), and Old
Yellow (right)

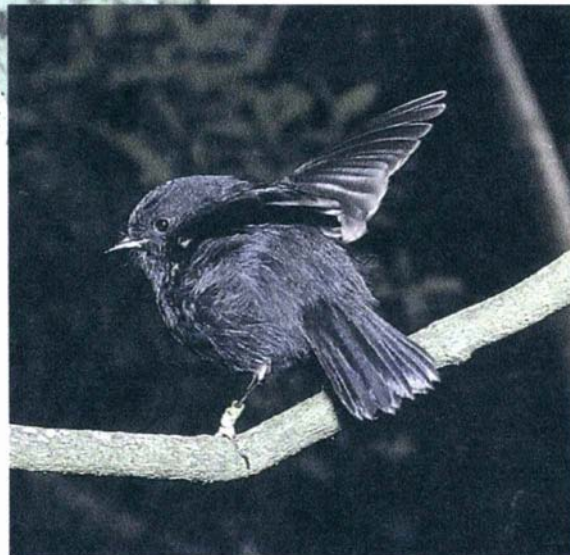


Photo: Rod Morris



The South Island robin
Photo: J. Kendrick (DOC)

Abstract

The genes of the major histocompatibility complex (MHC) are highly polymorphic and play a direct role in disease resistance. Loss of variation at MHC loci may increase extinction risk in endangered species, due to an inability to combat a range of pathogens. In this thesis, the evolution of class II B MHC genes is investigated, and levels of variation at these loci are measured in two species of New Zealand robin, the endangered Chatham Island black robin (*Petroica traversi*), and the non-endangered South Island robin (*Petroica australis australis*). Transcribed class II B MHC loci from both black robin and South Island robin were characterised prior to analysis of MHC variation. To this end, a non-lethal protocol for isolation of transcribed sequences from blood using 3'RACE and RT-PCR was developed. Four class II B cDNA sequences were isolated from black robin, and eight sequences were isolated from the South Island robin, indicating there are at least four class II B loci. RFLP analysis indicated that all class II MHC loci were contained in a single linkage group. Analysis of 3'untranslated region sequences enabled orthologous loci to be identified in the two species, and indicated that multiple rounds of gene duplication have occurred. A partial genomic DNA sequence of a putative pseudogene was also isolated from the black robin. Evolution of MHC genes in New Zealand robins appears to be influenced by gene conversion and balancing selection, resulting in loss of orthologous relationships in the coding region, and a highly diverse peptide-binding region. In order to assess the effect of population bottlenecks on MHC variation, levels of variation in the extant black robin population, which is descended from a single breeding pair, were compared with artificially bottlenecked populations of South Island robin and their respective source populations. Both RFLP and sequence analysis indicated that the black robin is monomorphic at class II B loci, while both source and bottlenecked populations of South Island robin have retained moderate to high levels of variation. Comparison of MHC variation with minisatellite DNA variation in each population indicated that genetic drift was the predominant force determining MHC diversity in bottlenecked populations in the short-term. Despite its lack of MHC variation, the black robin population appears to be viable under existing conditions. The evolutionary history of New Zealand's *Petroica* species, investigated by phylogenetic analysis of mitochondrial DNA sequences, is also discussed.

Acknowledgements

Firstly, I would like to thank my supervisor, Professor David Lambert, for the opportunity to work on such an interesting topic, and for his guidance, support and insights throughout the course of this degree.

This project was made possible by a Marsden grant to David Lambert (99-MAU-039). I wish to acknowledge the support of a Massey University Doctoral Scholarship, a scholarship from the New Zealand Federation of University Women, and an Institute of Molecular BioSciences postgraduate bursary.

The following people assisted with collection of blood samples for this project: Paul Barrett, Doug Armstrong, Lara Shepherd, and Hilary Aikman. Special thanks to Paul Barrett for teaching me how to blood sample small passerines. The black robin samples could not have been collected without support from the New Zealand Department of Conservation, in particular Christine Reed, Hilary Aikman, Mike Thorsen, and all the staff at the Chatham Islands Area Office. Hilary Aikman also kindly made available the black robin genealogy database. In addition I would like to thank Brent Beavan, Raewyn Empson, and David Pattemore for providing feathers for the phylogenetics study, Alan Tennyson of Te Papa for the museum specimens, and Doug Armstrong for making the North Island robin blood samples available.

I would like to thank Craig Millar for the many helpful discussions regarding both technical and theoretical aspects of this project, and Don Merton for help and support in early stages of this project, and for giving me a copy of the black robin “bible”. Thanks also to Rob Slade for constructing the BR6 probe and Judith Robins for CRDR1 primer sequence. I would also like to thank a number of people within the Institute of Molecular BioSciences who helped with various aspects of the project: Kathryn Stowell, for allowing me to use her spectrophotometer; Michelle McGill for various RNA-related protocols; and Lorraine Berry for her tireless work on the DNA sequencer.

I would like to thank my friends and colleagues who have passed through the Molecular Ecology laboratory, the Ecology department, and the Allan Wilson Centre during my four

years in Palmerston North, including Lara Shepherd, Peter Ritchie, Jenny Hay, Olly Berry, Jennifer Anderson, Leon Huynen, Gwilym Haynes, Gillian Gibb, Niccy Aitken, Tania King, Richelle Marshall, Wayne Linklater, and Ed Minot (and others I've probably missed out). I've really appreciated all the helpful discussions, advice, humour and coffee breaks that have made day-to-day life in the lab enjoyable. Special thanks to Jen for the black robin pathogen data, Pete for assistance with phylogenetic analysis, and for answering my endless inane questions about thesis writing and everything else, and to Jenny, Olly, and Lara for proof-reading various parts of this thesis. Thanks also to Steve Trewick for the helpful discussion about robin phylogenetics.

Finally, I would like to thank my parents for their support and interest in whatever obscure topic I might be working on, and especially for providing sustenance and a roof over my head during the final month of writing.

Thesis Structure and Format

This thesis is written as a series of papers, with a general introduction and final summary drawing together the main themes of the thesis. The general introduction (Chapter one) is a literature review providing background information on aspects of MHC genetics relevant to this thesis, and enabling the specific topics covered in later chapters to be put into the wider context. Some of this information is reiterated in reduced form in the later chapters, but the introduction provides additional detail that would not be appropriate in a scientific paper. An outline of the major aims of the study is also given.

Chapters two to six are data chapters and are written in the format of scientific papers. Each chapter is written to stand alone as an independent unit, which results in some repetition, particularly in the introduction and reference sections. This format, however, allows each aspect of the study to be considered as a whole, in preparation for publication. Chapters two and four are written in the format of short communications, with the results and discussion sections combined. A modified version of Chapter two (An evaluation of methods of blood preservation for RT-PCR from endangered species) has been accepted for publication as a technical note in *Conservation Genetics*, and a reprint is included in Appendix D. The remaining chapters are being prepared for publication. The second paper in Appendix D was also published during the course of this work. This paper also involved analysis of genetic variation in an endangered species, but is not directly relevant to main themes of this thesis.

Each of the data chapters includes a discussion, which covers important aspects of the empirical data presented and places the results in the context of existing work. The final chapter summarises the main findings of the study and outlines possible areas of future research. Appendices A and B consist of the raw sequence data from Chapters five and six, respectively, and a full list of samples used in this study is given in Appendix C.

Note on nomenclature: In chapter six, the term “New Zealand robin” refers specifically to *Petroica australis*, however in the remainder of the thesis “New Zealand robins” is used in more general terms for simplicity, and includes the Chatham Island black robin *Petroica traversi* as well as *Petroica australis*.

Table of Contents

Abstract	i
Acknowledgments	ii
Thesis Structure and Format	iv
Table of Contents	v
List of Tables	ix
List of Figures	x

Chapter One: The Major Histocompatibility Complex: Structure, Evolution and the Conservation of Endangered Species

1.1 Overview	1
1.2 The Major Histocompatibility Complex	2
1.2.1 Genomic organization of the MHC	2
1.2.2 Structure and function of classical MHC molecules	5
1.2.2.1 MHC diversity and the specificity of peptide binding	7
1.3 Polymorphism and evolution of MHC genes	8
1.3.1 Gene conversion	9
1.3.2 Concerted evolution and the birth-and-death model	11
1.3.3 Balancing selection	13
1.3.3.1 Sources of balancing selection	14
1.4 Pathogen resistance and MHC variation	16
1.4.1 Evidence for pathogen-driven selection	16
1.4.2 Models of pathogen-driven selection	18
1.4.3 How strong is balancing selection?	19
1.4.4 Pathogen resistance, MHC diversity and conservation	22
1.5 Conservation genetics and the black robin	23
1.6 Thesis aims and outline	25
1.7 References	26

Chapter Two: An Evaluation of Methods of Blood Preservation for RT-PCR from Endangered Species

2.1	Introduction	37
2.2	Methods	38
2.2.1	Blood preservation and RNA extraction	38
2.2.2	RT-PCR	39
2.3	Results and Discussion	41
2.3.1	Blood preservation and RNA extraction	41
2.3.2	Isolation of MHC cDNA sequences	43
2.4	References	43

Chapter Three: Class II B MHC Genes in New Zealand Robins: Evolution by Gene Duplication, Gene Conversion and Balancing Selection

3.1	Introduction	45
3.2	Methods	48
3.2.1	Birds and sample collection	48
3.2.2	Isolation of genomic DNA	48
3.2.3	Probes	49
3.2.4	Southern blot hybridisation	49
3.2.5	RNA isolation	50
3.2.6	Isolation of MHC cDNAs	50
3.2.7	Data analysis	52
3.3	Results	53
3.3.1	Restriction fragment length polymorphism	53
3.3.2	Inheritance/linkage of MHC fragments	53
3.3.3	Isolation of MHC cDNA sequences	54
3.3.4	Orthology, balancing selection and gene conversion	55
3.4	Discussion	63
3.4.1	Evolution of class II B MHC genes in NZ robins	65
3.5	References	68

Chapter Four: A Divergent Class II B MHC Sequence in the Black Robin

4.1 Introduction	72
4.2 Methods	72
4.3 Results and Discussion	74
4.4 References	81

Chapter Five: Population Bottlenecks and MHC Variation in New Zealand Robins

5.1 Introduction	83
5.2 Materials and Methods	87
5.2.1 Birds and sample collection	87
5.2.2 DNA techniques	87
5.2.3 Data analysis	89
5.3 Results	90
5.3.1 RFLP analysis	90
5.3.2 Isolation of intron sequences and design of PBR primers	90
5.3.3 Peptide binding region sequences	91
5.4 Discussion	97
5.4.1 MHC diversity in bottlenecked populations	98
5.4.2 Intron sequences and the evolution of class II B MHC genes	100
5.4.3 MHC monomorphism in the black robin: implications for conservation	102
5.5 References	104

Chapter Six: A Molecular Phylogeny of New Zealand's *Petroica* Species Based on Mitochondrial DNA Sequences

6.1 Introduction	110
6.2 Materials and Methods	112
6.2.1 Samples	112
6.2.2 Isolation of genomic DNA	115
6.2.3 PCR and sequencing	115

6.2.4	Data analysis	116
6.3	Results	117
6.3.1	Mitochondrial gene order	117
6.3.2	Mitochondrial DNA sequences	117
6.3.3	Phylogeny	121
6.4	Discussion	125
6.4.1	Phylogenetic placement of the black robin	125
6.4.2	Origins of New Zealand's <i>Petroica</i> species	128
6.4.3	Divergence within robin and tomtit radiations	129
6.5	References	131

Chapter Seven: Summary and Concluding Remarks

7.1	Introduction	133
7.2	Summary of major findings	133
7.3	Future directions	135
7.3.1	Does MHC monomorphism increase disease susceptibility in the black robin?	135
7.3.2	Locus specificity in studies of the passerine MHC	136
7.4	References	138

Appendix A: Class II B MHC peptide-binding region sequences 139

Appendix B: *Petroica* mitochondrial DNA sequences 143

Appendix C: Details of *Petroica* samples used in this study 149

Appendix D: Manuscripts 152

- I. *An evaluation of methods of blood preservation for RT-PCR from endangered species*
- II. *Minisatellite DNA profiling detects lineages and parentage in the endangered kakapo (*Strigops habroptilus*) despite low microsatellite DNA variation.*

List of Tables

1.1	Examples of MHC-disease associations	17
1.2	Species with low or no polymorphism at MHC loci	20
2.1	Storage conditions of sparrow samples	40
3.1	Sequences and annealing temperatures of primers for isolation of class II B MHC cDNA sequences	51
3.2	Nonsynonymous and synonymous substitution values for class II B MHC sequences	62
3.3	Gene conversion events between class II B MHC sequences from NZ robins	62
4.1	PCR primers used in this chapter	73
4.2	Pairwise sequence diversity between the Petr05 sequence and previously isolated cDNA sequences	75
5.1	Average Percent Difference values for each enzyme with the BR6 probe	90
5.2	Distribution and frequency of class II B MHC alleles in NZ robins	95
5.3	Nonsynonymous and synonymous substitution values for class II B sequences	96
5.4	Percentage of genetic diversity lost in bottlenecked populations compared with their source	97
6.1	<i>Petroica</i> taxa sampled in this study	113
6.2	Pairwise sequence divergence within and between ingroup taxa	119

List of Figures

1.1	Genomic organization of the MHC in birds and humans	5
1.2	Structure of class I and class II molecules	6
2.1	(A) Agarose gel electrophoresis of rRNA (B) RT-PCR products amplified using β -Actin primers (C) 3'RACE product amplified from black robin cDNA	42
3.1	Position of PCR primers used in isolation of class II B MHC cDNA sequences	51
3.2	RFLP profiles of class II B MHC genes in New Zealand robins	54
3.3	Nucleotide sequence alignment of class II B MHC cDNA sequences	58
3.4	Amino acid sequence alignment of MHC class II B exon 2 sequences	60
3.5	Neighbour-joining trees of class II B MHC 3'UTR and exon 2 sequences	61
4.1	Sequence of the putative pseudogene from black robin	77
4.2	Neighbour joining trees showing the relationship between Petr05 and other passerine sequences	79
5.1	(A) Map of New Zealand and the Chatham Islands (B) Schematic representation of populations	86
5.2	Position of PCR primers used in this study	88
5.3	Alignment of intron sequences flanking exon 2	94
5.4	Amino acid sequences of class II B MHC alleles	96
5.5	Plot of MHC diversity vs minisatellite DNA diversity	97
6.1	Map of New Zealand and Pacific islands	114
6.2	Mitochondrial gene order in birds	118
6.3	The observed number of transitions and transversions versus sequence divergence	120
6.4	Relationships of NZ <i>Petroica</i> species based on cyt <i>b</i> sequences	123
6.5	Relationships among <i>Petroica australis</i> and <i>Petroica macrocephala</i> subspecies based on CR sequences	124
A.1	Alignment of peptide-binding region sequences	139
B.1	Sequence of 1371 bp of the mitochondrial genome from <i>P. traversi</i> and <i>P. a. australis</i>	143
B.2	Alignment of cytochrome <i>b</i> sequences	144
B.3	Alignment of control region sequences	146

CHAPTER ONE

The Major Histocompatibility Complex: Structure, Evolution and the Conservation of Endangered Species

1.1 Overview

Small and declining populations often show reduced levels of genetic variation (e.g. Groombridge et al. 2000, Wisely et al. 2002, Miller et al. 2003 (Appendix DII)), however the consequences of this for the survival of endangered species has been the subject of much debate (Caughley 1994, Amos and Balmford 2001). The two main genetic outcomes of population decline are loss of variation through genetic drift and inbreeding, both of which cause a decrease in heterozygosity. It is argued that these processes may increase the risk of extinction in endangered species because of loss of adaptive flexibility, and decreased reproductive fitness from exposure of deleterious recessive alleles (Frankham 1995, Amos and Balmford 2001, Keller and Waller 2002). Genetic diversity may also be important for pathogen resistance in wild populations, as a number of studies have shown that inbred animals have increased susceptibility to pathogens (e.g. Coltman et al. 1999, Acevedo-Whitehouse et al. 2003). The genes of the major histocompatibility complex (MHC) are particularly important in this regard, as they play a critical role in the vertebrate immune system. Classical MHC genes (class I and class II) code for cell-surface proteins that present short peptides, usually fragments of bacterial or viral proteins, to T cells to initiate an immune response. MHC genes are highly polymorphic, and this diversity is thought to be linked to the diversity of pathogens faced by vertebrate species. Diversity at MHC loci appears to be an important component of differential disease resistance between individuals (Zekarias et al. 2002) and may also influence mate choice and kin recognition (Penn 2002). Populations with low levels of MHC variation are thought to be at risk from increased susceptibility to pathogens (O'Brien and Evermann 1988).

Analysis of genetic variation in declining populations has become an important component of conservation genetics. Analysis of MHC variation in particular is becoming increasingly relevant, as the threats posed to endangered species from disease are increasingly being recognised (Deem et al. 2001, Lafferty and Gerber 2002). In this thesis, MHC variation is analysed in an endangered passerine bird, the Chatham Islands black robin (*Petroica traversi*). Analysis of MHC variation in a non-model organism such as this requires an understanding of how variation is generated and maintained, as the evolutionary processes that determine the makeup of MHC genes appear to vary considerably among species (Edwards et al. 2000b). In this chapter I will outline key aspects of the genetics and evolution of MHC genes in birds and mammals and discuss their role in disease resistance. Specifically I will discuss the differences in MHC organisation among birds and mammals, the structure and function of classical MHC molecules, the generation and maintenance of MHC diversity, and the relationship between pathogen resistance and MHC diversity. The population history of the black robin and its relevance for conservation and MHC genetics is also outlined, along with the specific aims of this thesis.

1.2 The Major Histocompatibility Complex

The MHC is a large multigene family that plays a central role in the mounting of an immune response. The discovery of the MHC grew out of work on blood group antigens, tumour transplantation and skin graft experiments (reviewed by Klein 1986, 2001). Its physiological function was elucidated by Zinkernagel and Doherty, who found that in order for T lymphocytes to recognise an infected cell, they must simultaneously recognise the foreign antigen and self MHC molecule, a phenomenon termed MHC restriction (Zinkernagel and Doherty 1974a, 1974b). MHC molecules thus provide the context for the recognition of foreign antigens by T lymphocytes. MHC genes appear to be present in all vertebrate species with the exception of jawless fish. Invertebrates do not appear to have an MHC although alternative simple systems of self/nonself recognition are present (Kaufman et al. 1990).

1.2.1 Genomic organisation of the MHC

Genomic mapping of the MHC region in a variety of vertebrate species shows there are considerable differences in organisation and complexity between mammalian and non-mammalian species (Kulski et al. 2002). The mammalian MHC region appears to be large

and complex. The human MHC, often referred to as the HLA (human leukocyte antigen system), spans 4 mb and contains 224 gene loci. However, only 128 genes are predicted to be expressed, and more than 60% of these have functions outside of the immune system (MHC Sequencing Consortium 1999). Only a small number of loci represent classical antigen-presenting MHC genes, which are defined as those that determine skin-graft rejection and mixed lymphocyte response (Klein 1986). A number of non-classical MHC genes are also present. These are similar in structure to classical genes, but have only a weak influence on skin-graft rejection, are often non-polymorphic and may have specialised, (non-antigen presenting) functions. The human MHC is broadly divided into three regions: class I, class II, and class III (Figure 1.1). The class I region is characterised by a large number of pseudogenes and genes with non-immune function, amongst which the classical (HLA-A, B, and C) and nonclassical (HLA E-G) antigen presenting genes are interspersed. The class II region also contains a number of pseudogenes, however, almost all expressed genes have an immune function. In addition to the classical class II genes (DR, DQ, and DP), this region includes RING (nuclear kinase) genes, and genes involved in processing and transport of peptides such as the LMP (low molecular mass proteasome), TAP (transporter associated with processing) and DM genes (DMA and DMB), (Kelly et al. 1991, Trowsdale 1995, MHC Sequencing Consortium 1999). The class III region lies in between the class I and class II regions, and includes complement components such as C2, C4, and TNF genes, which are an important component of T-cell mediated immunity but are not directly involved in antigen presentation (Trowsdale 1995, The MHC Sequencing Consortium 1999).

Knowledge about the organization of the MHC in birds has largely been limited to galliformes, as to date only the chicken and quail MHC regions have been fully sequenced. In contrast to the mammalian MHC, the chicken MHC (or B locus) spans just 92kb and contains only 19 genes, and has been described as “minimal essential” (Kaufman et al. 1995b, Kaufman and Salomonsen 1997). The genes are more densely packed than in mammals, with shorter introns, and no pseudogenes and few repetitive elements (Kaufman et al. 1999a, 1999b). The class I and class II regions are not well defined and are not separated by the class III region (represented by only the C4 gene in chickens) (Guillemot et al. 1988, Kaufman et al. 1999b). There are two classical class I and two class II MHC genes, however only one gene of each class appears to be expressed at high levels (Kaufman et al. 1995b, Pharr et al. 1998). Homologues of a number of immune related genes found in the mammalian MHC are present, including the TAP genes, RING3 and DM genes. However a number of genes are

absent including LMP genes and most of the class III complement genes. There are also genes that are not present in the mammalian MHC including a lectin-like Natural Killer receptor (NKLR), a NK *Ig*-like receptor (NKIR) and the BG gene, which has some similarity to the *butryophilin* genes in the mammalian MHC (Kaufman et al. 1999b). Non classical class I and class II genes have also been found outside the B locus, in a genetically unlinked locus designated *Rfp-Y* by Briles et al. (1993).

Preliminary data on organisation of MHC in other birds suggests that minimal nature of the chicken MHC is an exception, as only closely related species such as the pheasant (*Phasianus colchicus*) (Witzell et al. 1999) and turkey (*Meleagris gallopavo*) (Zhu and Nestor 1995) appear to share the simple gene organisation. The quail MHC has a similar basic organisation to the chicken and contains homologs of many genes found in the B complex (Figure 1.1). However, it spans a larger region (191 kb), and contains multiple copies of many loci. There are seven class I and seven class II genes as well as multiple duplications of BG (8), NKLR (6), and NKIR (4) (Kulski et al. 2002).

MHC sequences have also been isolated from a number of passerine species, although there are little data on the organisation of MHC genes at the genomic level in these species. Large-scale genomic sequencing has only been carried out on red-winged blackbird (*Agelaius phoeniceus*) and house finch (*Carpodacus mexicanus*). In red-winged blackbird, 3 class II MHC genes (including one pseudogene), and 2 gene fragments have been isolated from a total of 84 kb of sequence from two cosmid clusters (Edwards et al. 1998, 2000a, Gasper et al. 2001), while a single MHC pseudogene has been isolated from 32kb of sequence from the house finch (Hess et al. 2000). From these results, it appears that the passerine MHC is much less gene dense than that of chicken or quail, although it has not yet been determined whether these regions are homologous to the B complex in chickens. MHC sequences isolated from other passerines also indicate that the MHC region in passerines may be larger and more complex than that of chicken. Seven class II sequences and four class I sequences have been isolated from a great reed warbler (*Acrocephalus arundinaceus*) cDNA library (Westerdahl et al. 1999, 2000), and a PCR survey of Darwins finches MHC sequences suggests there may be 6 class II MHC loci (Sato et al. 2000). The introns of class II MHC genes from red-winged blackbird and Darwins finches are also longer than those of the chicken class II genes (Sato et al. 2000, Gasper et al. 2001). Other orders of birds are not well represented in MHC studies to date. A recent analysis of great snipe (*Gallinago media*) MHC sequences found that the

number and size of genes appeared to be intermediate between that of passerine and chicken (Ekblom et al. 2003). Overall, the data available on the avian MHC suggests there may be considerable variation in MHC organisation and size among species.

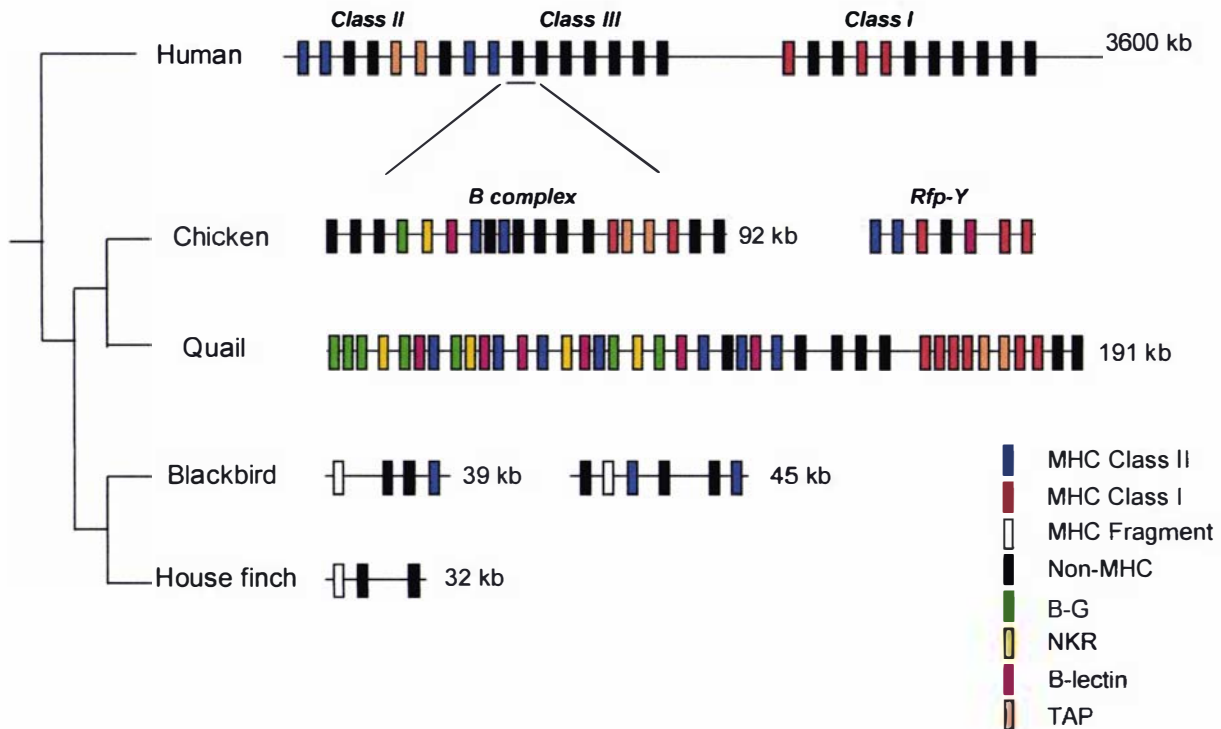


Figure 1.1 Genomic organisation of the MHC region in birds and humans. Only avian MHC regions for which the genomic structure has been elucidated are shown. A simplified version of the human MHC is shown, with each square representing multiple genes, except for the TAP genes. Non-MHC genes are represented by black boxes, except for B-G (green), NKR (yellow), B-lectin (pink) and TAP genes (orange). (Adapted from Hess and Edwards 2002).

1.2.2 Structure and function of classical MHC molecules.

The classical class I and class II MHC genes differ in their structure, tissue distribution, and in the types of T-cell they activate. The three-dimensional structures of the class I and class II molecules have been determined by X-ray crystallography (Bjorkman et al. 1987a, Brown et al. 1993). The class I molecule has a single transmembrane polypeptide chain with three extracellular domains (α_1 , α_2 , and α_3), which is non-covalently associated with a β_2 -microglobulin molecule (Figure 1.2). The α_1 and α_2 domains combine to form the peptide binding groove which interacts with the T-cell receptor (Bjorkman et al. 1987b Chelvanayagam et al. 1997). Each domain is coded for by a separate exon, while a gene

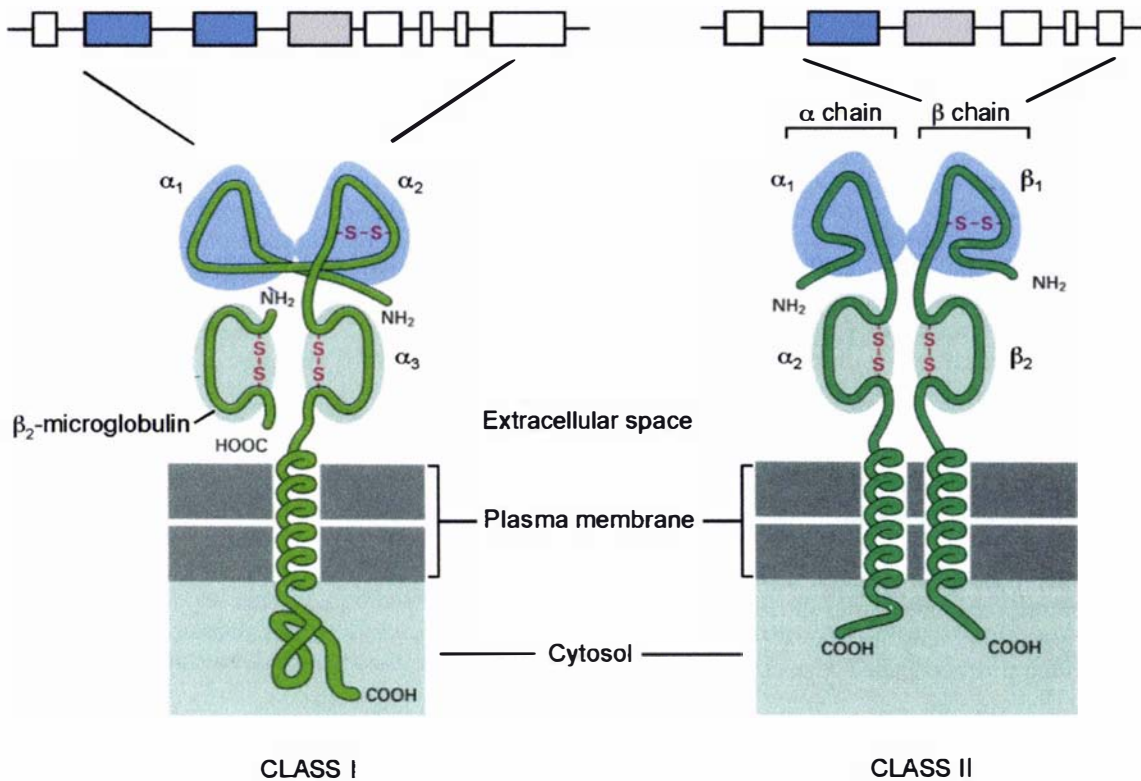


Figure 1.2 Structure of class I and class II molecules, showing the domains which interact to form the peptide binding site (blue). The corresponding gene structure in chickens is shown (boxes indicate exons) (Adapted from Alberts et al. 1994).

located outside the MHC codes for the β_2 -microglobulin molecule. Class II molecules are comprised of 2 polypeptide chains, α and β , coded for by separate genes (A and B). In mammals, these genes are found in pairs, whereas in chicken there appears to be only a single, monomorphic, class II A gene which is located outside the MHC (5 cM away) (Kaufman et al. 1995a). Each chain has 2 extracellular domains encoded for by separate exons: α_1 and β_1 which interact to form the peptide binding site; and the immunoglobulin-like α_2 and β_2 (Brown et al. 1993). The domains that form the peptide binding groove consist of an α -helical region, which forms the side of the groove, and β -sheet strands, which form the base of the groove (Chelvanayagam et al. 1997).

Class I molecules are found on all nucleated cells and present mainly endogenous peptides (derived from proteins produced inside infected cells) about 8-11 amino acids long on the surface of the cell, where they are recognised by CD8⁺ cytotoxic T lymphocytes (CTLs). The CTLs then proliferate and begin to kill infected cells. Class II molecules have a more

restricted expression pattern than class I molecules, being expressed only on the surface of specialised antigen presenting cells, such as dendritic cells, B cells, macrophages and other haematopoietic cells. They bind and present mainly exogenous peptides (approximately 15-24 amino acids long), derived from foreign antigens (e.g. viruses, bacteria or parasites) that are taken up into the cell from the extracellular fluid. Class II MHC molecules are recognised by CD4⁺ T helper cells, which then proliferate and activate antibody-producing B cells, and also enhance CTL activity. The mechanisms of peptide processing and presentation to T cells have been reviewed in Van Kaer (2002) (class I), and Robinson and Delvig (2002) (class II). Peptides derived from self-proteins are also presented by both classes of MHC molecule, but generally do not activate T cells. This is because during development negative selection of the T cell repertoire occurs, in which T cells that strongly recognise self peptide/self MHC complexes are eliminated to prevent self-reactivity (reviewed in Sebzda et al. 1999 and Starr et al. 2003). Thus, for every MHC molecule expressed in an organism, a proportion of T cells are eliminated.

1.2.2.1 MHC diversity and the specificity of peptide binding

Knowledge of how the structure of MHC molecules relates to their function is important for studies of MHC diversity. The polymorphic residues in MHC molecules are concentrated in peptide binding groove and have important effects on peptide binding specificity and T cell response. A higher rate of nonsynonymous than synonymous nucleotide substitution at these residues suggest that these sites are subject to positive selection for diversity in peptide binding (Hughes et al. 1990). Falk et al. (1991) confirmed that different MHC alleles bind a different spectrum of peptides and that the specificity of binding is determined by two or three anchor residues along the peptide that interact with amino acids lining the peptide binding groove. The rest of the amino acids along the peptide are free to vary, creating a broad, rather than tight, specificity where each MHC molecule can bind a range of peptides with high affinity as long as the anchor residues are conserved. As more crystal structures of MHC molecules have been determined, the sequences of thousands of peptide motifs that bind to specific alleles have become available (Brusic et al. 1998), and many theoretical approaches to predicting peptide motifs have been developed (e.g. Buus 1999, Donnes and Elofsson 2002). The ability to predict which epitopes from particular pathogenic proteins will bind to specific MHC molecules is important for vaccine development (Kast et al. 1994). Characterisation of peptide motifs from common HLA alleles has found that different alleles

may bind an overlapping spectrum of peptides, and can be arranged into supertypes on the basis of their peptide-binding specificity (Sidney et al. 1996, Ou et al. 1998). This means that allelic variation may not always result in functional polymorphism, as individuals who are heterozygous for two different MHC alleles may be homozygous with respect to peptide binding if the two alleles are from the same supertype (e.g. Flores-Villanueva et al. 2001).

A recent report has challenged the view that the MHC polymorphism enhances immune defence simply because it enables a broad array of peptides to be presented. Messaoudi et al. (2002) found that differences in pathogen resistance between MHC alleles was due to differences in T cell responses rather than the ability of the differing alleles to present different peptides. Both the H-2K^b and H-2K^{bm8} haplotypes bind the immunodominant peptide from herpes simplex virus type 1 (HVH-1) with equal affinity, however H-2K^{bm8} individuals have 4-5 times greater resistance to HVH-1 than H-2K^b. Despite the fact that both molecules are identical at T cell contact residues, it was found that the T cells produced in the H-2K^{bm8} response were more diverse and bound the MHC-peptide complex more strongly. This suggests that in some examples MHC diversity may be important for producing a diverse T-cell repertoire, rather than for binding a variety of peptides.

1.3 Polymorphism and evolution of MHC genes

Classical MHC genes are the most polymorphic genes known in vertebrates. There are a large number of alleles at each class I and class II locus in many species. In humans for instance, 330 DRB1 alleles, and 511 HLA-B alleles have been recognised (as of January 2003, Robinson et al. 2003). In individual populations the number of alleles can vary greatly, but allele frequencies are often evenly distributed (Meyer and Thomson 2001). In many cases there is a high level of sequence divergence between alleles, and as discussed above, this diversity is largely restricted to the exons which contain the peptide-binding region (PBR) – exon 2 of class II loci, and exons 2 and 3 of class I loci. Originally it was thought that the high polymorphism within the MHC reflected a high mutation rate followed by selection for diversity, i.e. rapid post speciation diversification. However, the mutation rate for MHC genes appears to be similar to that of other genes in the genome (Satta et al. 1993, Klein et al. 1993b), and is not sufficient to explain the high levels of diversity. There is now substantial evidence that MHC polymorphism arises from a combination of point mutation,

gene conversion, and balancing selection, however the relative contribution of each remains controversial, and appears to differ between loci and species.

1.3.1 *Gene conversion*

Gene conversion is a form of recombination, involving the non-reciprocal exchange of short segments of sequence, either between alleles at the same locus (intralocus), or between alleles at different loci (interlocus). Evidence for gene conversion comes from the fact that when MHC allele sequences are compared, they often differ from one another by a short segment of sequence that can itself be found within another MHC allele, rather than by point mutations. Gene conversion was first implicated in the generation of novel ('mutant') alleles of the murine class I H-2K gene (Weiss et al. 1983, Miyada et al. 1985), and there have since been numerous studies of MHC sequence patterns documenting gene conversion in both class I and class II genes in a variety of species, including mammals (She et al. 1991, Andersson et al. 1991 Andersson and Mikko 1995), birds (Wittzell et al. 1999) and fish (Langefors et al. 2001b). Direct evidence for gene conversion has been documented using PCR based assays to directly observe gene conversion events in sperm (Hogstrand and Bohme 1994, Zangenberg et al. 1995, Hogstrand and Bohme 1999).

Much of the debate on the role of gene conversion in MHC evolution centres on not whether gene conversion occurs, but to what degree it contributes to the generation of MHC variation. It has been suggested that gene conversion, rather than point mutation, is the principal mutational mode in the MHC (Martinson et al. 1999). Analysis of HLA-B alleles in native American populations found the majority of novel alleles were produced by gene conversion rather than point mutation (Parham and Ohta 1996). In addition, computer simulation studies have found that levels of MHC polymorphism can be elevated by gene conversion in combination with weak balancing selection (Ohta 1997, 1999). A contrasting view is that point mutation is the main generator of polymorphism and that patchwork patterns can be accounted for by convergent evolution due to selective constraints (O'hUigin 1995). However, the accumulation of mutations required for convergent evolution within a species is unlikely to occur within limited time periods, and the increased rate of synonymous substitutions often measured within the PBR is more suggestive of gene conversion, as synonymous substitutions are carried in concert with non-synonymous substitutions when blocks of sequence are exchanged (Bergstrom et al. 1998). Convergent evolution may

account for shared motifs between distantly related species such as humans and mice (Yeager and Hughes 1999). Erlich and Gyllensten (1991) suggested that polymorphism in human class II alleles can be explained by a mixture of gene conversion and convergent evolution, depending on the location of the shared motif.

It is evident that the relative rates of gene conversion within the MHC vary between loci and species. In general, intralocus conversion occurs at higher frequency than interlocus conversion, as gene conversion occurs more frequently between regions of high similarity. Zangenberg et al. (1995) estimate the rate of intralocus gene conversion at the HLA-DPB1 locus in human sperm was 0.81×10^{-4} (almost 1 event per 10,000 gametes per generation), while Hogstrand and Bohme (1999) reported rates of interlocus gene conversion in class II MHC genes in mouse sperm varying from 1.2×10^{-6} to 9.6×10^{-5} . In this study, it was found that gene conversion only occurs in the germ cells and not in somatic cells, and appears to be associated with CpG islands (regions containing high frequencies of C-G dinucleotides). Thus gene conversion is not a feature of all MHC genes, and those which have never been reported to undergo gene conversion do not contain these islands. In humans and primates, the HLA-B and DPB-1 loci appear to have a much higher rate of gene conversion than HLA-A, HLA-C, DRB-1, or DQB-1 (Hughes et al. 1993, Parham and Ohta 1996, Takahata and Satta 1998).

Where interlocus gene conversion occurs frequently, it would be expected that phylogenetic analysis of alleles would show no clear distinctions between loci. Low rates of interlocus gene conversion, however, will result in clustering of alleles by loci, and will allow for the retention of ancient allelic lineages, and the identification of orthologous relationships between loci in distantly related species. In mammals, orthologous class II loci can be identified among distantly related species, and it appears that these loci evolved prior to the major mammalian speciation events (Takahashi et al. 2000), and have been maintained separately ever since, with little interlocus gene conversion. However, the α helix region of class II loci does appear to undergo high rates of interlocus gene conversion, as sharing of sequence motifs within this region is evident between loci in different species (Andersson and Mikko 1995, Bergstrom et al. 1998). Interlocus gene conversion appears to be much less common in humans than in other, non-primate mammals (Ohta 1999), as there is a clear separation of alleles at HLA-A, -B and -C loci and maintenance of ancient allelic lineages, whereas alleles at mouse class I genes H2-K, -D and -L intermingle among loci (Lawlor et al.

1990). The rate of interlocus gene conversion in birds appears to be much higher than in mammals, as class II sequences cluster within species, and orthologous loci can only be identified in closely related species (Edwards et al. 1995b, 1999, Wittzell et al. 1999).

It is important to note that gene conversion can only generate diversity when it occurs over short tracts between already polymorphic sequences. This suggests that point mutation and balancing selection remain important factors in generating and maintaining polymorphisms (She et al. 1991). It has been argued that gene conversion alone is not sufficient to account for MHC diversity, and that balancing selection is required to explain the high levels of diversity, and the excess of nonsynonymous substitutions at peptide binding sites (Satta 1997, Takahata and Satta 1998, Martinsohn et al. 1999).

1.3.2 *Concerted evolution and the birth-and-death model*

Where interlocus gene conversion occurs frequently over longer fragments, it can homogenise sequences within a species, a process known as concerted evolution. This leads to paralogous genes within a species maintaining sequence similarity and evolving together, to the exclusion of orthologous genes in closely related species. Concerted evolution is thought to be prevalent in evolution of avian MHC genes (Edwards et al. 1995b, 1999, Wittzell et al. 1999), however it is difficult to distinguish between this and recent, post-speciation gene duplication, as both models result in clustering of loci within species (Hess and Edwards 2002). Concerted evolution does not appear to play an important role in evolution of mammalian MHC genes, because as described above, rates of interlocus gene conversion appear to be lower than for birds, and orthologous genes in closely related species are more similar to each other than paralogous genes within a species. It has been proposed that mammalian MHC genes evolve under a birth-and-death model, rather than a concerted evolution model. Under this model, new genes arise from gene duplication, some are maintained for long periods by balancing selection, and others are deleted or become pseudogenes. Mutation, gene conversion, and selection contribute to divergence of genes after duplication (Nei et al. 1997, Gu and Nei 1999). However, it has been suggested that the concerted evolution and birth-and-death models are not necessarily mutually exclusive, but are the result of the same processes occurring over different timescales (Edwards et al. 1999, Wittzell et al. 1999). Concerted evolution occurs primarily between closely related loci (i.e. recently duplicated genes) while birth-and-death processes may be occurring in the long term.

The presence of pseudogenes in passerines suggests that the avian MHC is also subject to gene duplication and decay (Hess et al. 2000, Edwards et al. 2000a). However, gene duplications may have occurred more recently than in mammals, and higher rates of interlocus gene conversion, and hence concerted evolution, may result.

In mammals, the pattern of duplication and deletion appears to be the same for class II as class I loci, however the length of time it takes for gene turnover in class II genes is much longer. This is reflected in the ability to identify orthologous class II loci in distantly related species such as humans and mice, but not orthologous class I loci, and the apparent plasticity of the class I region where the number of loci may vary with the haplotype (haplotype diversity) (Trowsdale 1995). Takahashi et al. (2000) estimated that most mammalian class II loci originated at least 170-200 MYA, while recent work by Piontkivska and Nei (2003) suggests that most primate class I loci diverged about 35-66 MYA. The class II DRB genes in mammals appear to be an exception, however, as they also have a level of plasticity similar to class I loci, with multiple independent duplications in different mammalian species (e.g. Sliereendregt et al. 1994). It has been suggested that this can occur because the DRA chain gene is monomorphic, and thus duplication of both DRA and DRB genes in tandem may not be required to form functional heterodimers, unlike the situation for the other class II loci (Yeager and Hughes 1999).

The presence of different numbers of genes in different haplotypes, even within the same or closely related species suggests that the processes of gene duplication and decay are always occurring. It has been suggested that the expansion and contraction of genes coincides with adaptive radiations of new species (Klein et al. 1998b), a process that is best seen in teleost fishes. Primitive teleost fish species such as carp, zebrafish and salmon have 1-3 highly differentiated class I and class II genes (e.g. Ono et al. 1992), whereas extensive gene duplication is present in the more advanced neoteleost species, such as cichlids, sticklebacks and Atlantic cod, which have undergone extensive species radiations in the last 1-20 million yrs. For instance, at least 17 class II loci have been identified in cichlid fishes (Malaga-Trillo et al. 1998), and at least 42 duplicated class I genes were identified in a single Atlantic cod individual (Miller et al. 2002). In theory, the larger the number of MHC loci, the greater the variety of peptides able to be presented, so gene duplication should provide a selective advantage. However, the corresponding decrease in T-cell repertoire may counteract the advantage of additional MHC genes and limit the number of loci (Nowak 1992). In species

where large numbers of duplicated loci have been identified there appear to be mechanisms which restrict the number of expressed MHC molecules, such as haplotype diversity, degeneration to non-classical genes, and loss of gene expression in some loci (Malaga-Trillo et al. 1998, Miller et al. 2002, Wegner et al. 2003).

1.3.3 *Balancing selection*

It is evident that neutral scenarios cannot explain the levels of polymorphism observed at MHC loci, even where diversity-generating processes such as gene conversion are common (Satta 1997, Takahata and Satta 1998). Several lines of evidence support the hypothesis that polymorphism is primarily maintained by balancing selection (reviewed in Meyer and Thomson 2001).

(1) High numbers of alleles exist for many loci and allele frequencies are uniform. The number of alleles observed at HLA loci is far higher than that expected under neutral conditions, unless unrealistically high mutation rates and population sizes are assumed (Potts and Wakeland 1993). Allele frequencies are also more even than expected under neutrality, with no 'wild type' allele prevailing in many populations. The Ewens-Watterson test measures the evenness of allele frequencies by comparing observed vs expected homozygosity. Using this test, homozygosity at MHC loci has been found to be lower than neutral expectations (i.e. allele frequencies are more even) in several human populations (Markow et al. 1993, Salamon et al. 1999, Valdes et al. 1999), and other mammalian populations including red wolves (Hedrick et al. 2002) and social tuco-tucos (Hambuch and Lacey 2002). Similarly, an excess of observed heterozygotes over Hardy-Weinberg proportions has been observed for MHC loci in some populations (e.g. Markow et al. 1993, Hambuch and Lacey 2002, Hedrick et al. 2002).

(2) Alleles are highly divergent, and lineages are maintained for long periods. Nucleotide diversity at MHC loci is about 10 times higher than for other nuclear or mitochondrial genes in humans and mice (Nei and Hughes 1991), and some alleles differ by as many as 60 substitutions in exon 2 (Meyer and Thomson 2001). Under neutrality, nucleotide diversity is a function of mutation rate and population size. These are expected to be similar for MHC genes as for other loci, so the most likely scenario for elevated diversity at MHC loci is that balancing selection maintains alleles in a population for longer than expected, and they have

time to accumulate more differences than neutral alleles. In mammals, there are many examples of allelic lineages that have persisted through successive speciation events, a phenomenon known as trans-species evolution (reviewed in Klein 1987 and Klein et al. 1998a). Gene conversion may act as a form of hypermutation and contribute to increased divergence in MHC alleles, but cannot by itself explain the presence of alleles that have been maintained for a long period.

(3) Sites involved in peptide binding have elevated rates of nonsynonymous substitutions and higher levels of heterozygosity. In the absence of balancing selection, the number of synonymous substitutions should be higher than the number of nonsynonymous substitutions, as amino acid changes are more likely to be selected against (purifying selection). Hughes and Nei (1988) found that, while this was true outside the peptide binding region, for sites forming the peptide binding region in both human and mouse class I genes there was a significantly higher number of nonsynonymous substitutions than synonymous substitutions. Since then this pattern has been found in both class I and class II genes across a wide range of species (e.g. Hughes and Nei 1989, Edwards et al. 1995a, Klein et al. 1993a, Hedrick et al. 2002). In addition, average heterozygosity at individual amino acid sites are in some instances an order of magnitude higher for peptide binding sites than sites in the rest of the molecule (Hedrick and Kim 1999).

(4) Increased levels of variation in non-coding regions close to classical MHC loci. The distribution of variation across the MHC region is not random, rather the level of polymorphism decreases as the distance from classical MHC loci increases (Satta et al. 1998). This is consistent with hitchhiking of non-coding regions with closely linked sites that are under selection (Thomson 1977).

1.3.3.1 Sources of balancing selection

The nature of the driving force behind the selective maintenance of MHC polymorphism has been the subject of much debate. Sources of balancing selection proposed include pathogen resistance, non-random mating and maternal-foetal interactions (Apanius et al. 1997, Hedrick and Kim 1999). Of these, pathogen resistance appears to be the most universal across all vertebrates, as it is directly linked to the main function of the MHC molecules in immune defence. Different MHC alleles are thought to confer differential resistance to particular

infectious diseases (e.g. Hill et al. 1991, Langefors et al. 2001a, Lohm et al. 2002), and thus increased MHC polymorphism may confer resistance to a greater range of pathogens. The relationship between pathogen resistance and MHC diversity will be discussed in detail in the next section.

Non-random mating, where individuals preferentially choose mates with differing MHC genotypes, may be a source of MHC diversity in some species (reviewed in Apanius et al. 1997, Penn and Potts 1999, and Penn 2002). This phenomenon has been reported in house mice (Egid and Brown 1989, Potts et al. 1991), humans (Wedekind et al. 1995, Ober et al. 1997), and fish (Landry et al. 2001, Reusch et al. 2001). However the evidence is often not conclusive, particularly in humans (e.g. Hedrick and Black 1997), and in some species there is no evidence for MHC-dependent mating preferences (Paterson and Pemberton 1997). It is thought that odour differences linked to MHC genotype are the basis for non-random mating (reviewed in Eggert et al. 1999) however the precise molecular mechanisms responsible for this are unknown. Mating preferences may also be indirectly linked to MHC diversity, as disease resistance can affect visual secondary sexual characteristics and subsequent mate choice (the Hamilton-Zuk hypothesis) (Hamilton and Zuk 1982). Support for this hypothesis has been found in pheasants (von Schantz et al. 1997).

Maternal foetal interactions have been identified as another potential reproductive-based source of balancing selection in mammals. In humans, couples with a history of spontaneous abortion often share alleles at MHC loci (Alberts and Ober 1993, Ober et al. 1998), and one explanation is that correct foetal implantation and growth requires an immune response that occurs when the mother and foetus differ at MHC loci (Hedrick and Kim 1999). This mechanism is unlikely to apply to oviparous animals such as birds, fish and reptiles, however. Non-random mating and spontaneous abortion based on MHC genotype may have developed in order to preferentially produce MHC heterozygotes (or individuals with disparate MHC alleles) who have increased pathogen resistance, or as a kin recognition mechanism to avoid inbreeding. These two factors are not mutually exclusive and evidence for both exists (reviewed in Penn and Potts 1999, Penn 2002). Reproduction-based sources of balancing selection may play an important role in maintaining MHC diversity in species in which they occur. However, these mechanisms are not as universal as pathogen-driven mechanisms, and it is likely that no single force is responsible for the maintenance of MHC diversity (Apanius et al. 1997).

1.4 Pathogen resistance and MHC variation

1.4.1 Evidence for pathogen-driven selection

Under pathogen-driven balancing selection, the diversity of MHC alleles present in a population will be largely determined by the diversity of past and present pathogen exposure. An important assumption under this model is that different MHC alleles or allelic supertypes provide different degrees of resistance or susceptibility to specific pathogens, and these types of associations have been found in several studies (Table 1.1). Also consistent with this model is the finding of pathogen evasion of MHC molecules (reviewed in Brodsky et al. 1999). For instance, the Epstein-Barr virus epitope bound by HLA-A11 contains a mutation in populations where this allele is common, allowing the virus to escape detection (de Campos-Lima et al. 1993). Similar escape mutants associated with particular MHC alleles have been found in the HIV-1 virus (Moore et al. 2002).

MHC-pathogen associations are often weak or inconsistent, however, and may vary considerably between populations. This suggests that non-MHC genes may also be important for disease resistance, and/or may reflect variation in disease strains with geographical location. For example, in humans the class I MHC allele B*5301 is protective against malaria in the Gambia, but not in Kenya. Instead, DRB1*0101 is protective in Kenya (reviewed in Hill 1998). Associations between HIV and MHC alleles are particularly inconsistent, possibly reflecting the extreme polymorphism of the virus, even within the host. For some diseases however, the same associations are consistently found in widely disparate populations. For example, the association between HLA-DR2 with tuberculosis has been found in India, Russia and Indonesia, and DRB1*1302 is associated with clearance of the hepatitis B virus in the Gambia and in European populations (reviewed in Hill 1998). Many non-MHC genes have also been associated with disease resistance. For instance, a variant in the CC chemokine receptor gene-5 (CCR-5) is associated with resistance to HIV-1 and slower rates of progression to AIDS, and polymorphisms in the tumour necrosis factor gene promoter, vitamin D receptor, and macrophage gene NRAMP-1 have been associated with differential susceptibility to a variety of infectious diseases (reviewed in Hill 1998).

Table 1.1 Examples of MHC-disease associations. DR2, DR7, B*35, Cw04, and Bw4 are allelic supertypes, all others are individual alleles. For chicken, correlations are with MHC haplotypes (a combination of linked class I and class II alleles) in particular inbred chicken lines. HTLV-1, human T-lymphotropic virus-1; INHV, infectious hematopoietic necrosis virus.

Disease	MHC association	Reference
Humans		
Malaria	HLA-B*5301 protective DRB1*1302 protective	Hill et al. 1991
HIV	HLA-B*35/Cw*04 susceptible HLA-B*27 protective HLA-B*57 protective HLA-Bw4 protective Heterozygote advantage (class I)	Carrington et al. 1999 Kaslow et al. 1996 Migueles et al. 2000 Flores-Villanueva et al. 2001 Carrington et al. 1999
Hepatitis B	DRB1*1302 protective HLA-DR7 susceptible Heterozygote advantage	Thursz et al. 1995 Almarri and Batchelor 1994 Thursz et al. 1997
Hepatitis C	DQB1*0301 protective DRB1*1101 protective	Alric et al. 1997, Minton et al. 1998
Epstein-Barr virus	HLA-A11 protective	de Campos-Lima et al. 1993
Leprosy	HLA-DR2 susceptible	Visentainer et al. 1997
Pulmonary tuberculosis	HLA-DR2 susceptible	Brahmajothi et al. 1991
HTLV-1	DRB1*0101 susceptible HLA-A*02 protective	Jeffery et al. 1999
Chicken		
Mareks Disease	B ²¹ protective B ¹⁹ susceptible	Briles et al. 1983
Rous Sarcoma virus	B ¹² protective B ⁴ susceptible	rev. in Kaufman and Venugopal 1998
<i>Emeria tenella</i> (Coccidiosis)	B ³ , B ²¹ protective B ² susceptible	Caron et al. 1997
Fowl cholera	B ¹ protective	Lamont et al. 1987
Cellulitis	B ¹³ protective B ²¹ susceptible	Macklin et al. 2002
Atlantic salmon		
<i>Aeromonas salmonicida</i>	Class II e protective Class II j susceptible	Langefors et al. 2001a, Lohm et al. 2002
Chinook salmon		
INHV	Heterozygote advantage (class II)	Arkush et al. 2002
Soay Sheep		
Intestinal nematodes	OLA-DRB263 protective OLA-DRB257 susceptible	Paterson et al. 1998

The strongest correlations between MHC alleles and disease have been found in chicken, where only one class I and one class II gene are expressed at high levels. Class I and II genes are tightly linked so can evolve together as an allelic group (Kaufman 1999, 2000). This gives rise to stable haplotypes comprising particular combinations of class I and II alleles which are strongly associated with differential disease resistance. Non-MHC genes also

appear to be important for disease resistance in the chicken however, as in many instances the specific MHC haplotype associated with disease resistance varies in different genetic backgrounds (ie different chicken lines) (Plachy et al. 1992, Lamont 1998). Mareks disease is perhaps the strongest example of a specific MHC haplotype associated with disease resistance, as chickens with the B²¹ haplotype have 80-95% survival when infected, while other haplotypes have high mortality (Briles et al. 1983, Plachy et al. 1992). This is thought to be related to the level of cell-surface expression of class I molecules and the subsequent level of Natural Killer cell activity, as the most resistant haplotype has the fewest molecules on the cell surface, while the most susceptible haplotype B¹⁹ has the most (reviewed in Kaufman 1998).

An association has also been found between differing environmental stressors and the pattern of nucleotide substitution in the estuarine fish *Fundulus heteroclitus* (Cohen 2002). In this study, heavily parasitised fish from a heavily PCB contaminated site showed elevated rates of amino acid substitution in the α -helix region in exon 2 of a class II B MHC gene, and a higher d_N/d_S ratio at peptide-binding sites, compared with fish from a clean site. Amino acid substitutions in fish from the clean site were mainly concentrated in the β -sheet region, producing functional differences in the suite of peptides that fish in each geographical location can bind.

1.4.2 Models of pathogen-driven selection

Three major models have been described for how pathogen-driven balancing selection contributes to MHC diversity. (1) Frequency dependent selection, in which rare alleles have a selective advantage over common alleles as pathogens may evolve resistance to the most common alleles. Under this model, as rare alleles are selected for they become common and previously common genotypes may become rare, resulting in a constant cycling of alleles (Bodmer 1972). (2) Heterozygote advantage (overdominance), where all heterozygotes have an advantage over all homozygotes, as they have 2 molecules for each locus so are able to present peptides from a larger range of pathogens than homozygotes (Doherty and Zinkernagel 1975). (3) Fluctuating selection, where the selection intensity varies over time and space due to the presence or absence of particular pathogens. Under this model, different MHC molecules are selected for at different times as MHC variation tracks pathogen variation (Hedrick and Kim 1999).

Mathematical modelling shows that all three models can theoretically account for MHC polymorphism (e.g. Maruyama and Nei 1981, Takahata and Nei 1990, Takahata et al. 1992, Hedrick 2002). In practice, however, it is not clear which model is most important. It has been suggested that the three models are not mutually exclusive and that all contribute to the maintenance of MHC variation (Hedrick and Kim 1999, Meyer and Thomson 2001). Most correlations between pathogens and MHC variation involve specific MHC alleles, which appears to fit better with frequency-dependent or fluctuating selection models. Heterozygote advantage has only been observed in a few studies of specific diseases (Table 1.1), but may be more relevant in the context of resistance to multiple pathogens. This was recently demonstrated by Penn et al. (2002), who found that MHC heterozygous mice were more resistant to infection with multiple strains of *Salmonella* than homozygous mice. However, resistance was dominant rather than overdominant, as although MHC heterozygotes were more resistant than homozygotes on average, they were not more resistant than the most resistant parental homozygote. The effect of multiple pathogens on MHC diversity has also been demonstrated in the three-spined stickleback (*Gasterosteus aculeatus*) by Wegner et al. (2003), who found a positive correlation between the diversity of parasites present in a population and the diversity of MHC alleles.

1.4.3 How strong is balancing selection?

Although a number of studies have shown a link between particular MHC alleles and resistance or susceptibility to pathogens, the number of strong, consistent associations is low, particularly in mammals. In addition, there are a number of examples of viable populations with little or no MHC polymorphism (Table 1.2) indicating that MHC polymorphism is not a prerequisite for at least short-term survival. This has led to the suggestion that the strength of balancing selection is generally low (Klein et al. 1993b). In many cases, low MHC polymorphism is linked to small population size or past population bottlenecks (e.g. Hedrick et al. 2000a), suggesting that genetic drift can override balancing selection (Seddon and Baverstock 1999), while in other cases it is possible that selection pressure is reduced due to behavioural factors and decreased exposure to pathogens (Slade 1992, Ellegren et al. 1996, Sommer et al. 2002).

Table 1.2. Examples of species with low or no polymorphism at MHC loci. Levels of genetic variation at neutral (minisatellite or microsatellite) loci, and the hypothesized cause of low MHC diversity is given. RFLP = Restriction fragment length polymorphism (Southern blotting using heterologous probes), SSCP = Single stranded conformation polymorphism, Seq = nucleotide sequencing. *Hoelzel et al. (1999) reported high levels of diversity in the Southern Elephant seal.

Species	MHC variation	Assay method	Variation at neutral loci	Hypothesized cause	Reference
Fin Whale, Sei Whale	Low (Class II)	RFLP	Moderate	Fewer pathogens	Trowsdale et al. 1989
Southern Elephant seal	Low (Class II)	RFLP	Moderate	Fewer pathogens	Slade 1992*
Northern Elephant seal	Low (Class II)	Seq	Low	Bottleneck	Hoelzel et al. 1999
Balkan mole rat	Low (Class I & II)	RFLP	?	Bottleneck	Nizetic et al. 1988
Malagasy jumping rat	Low (Class II)	Seq	?	Mating system, Bottleneck	Sommer et al. 2002
Beaver	Monomorphic (Class I & II)	RFLP	low	Bottleneck	Ellegren et al. 1993
Cheetah	Low (Class I)	RFLP	low	Bottleneck	Yuhki and O'Brien 1990
Asiatic Lion (Gir forest)	Monomorphic (Class I)	RFLP	low	Bottleneck	Yuhki and O'Brien 1990
Lion (Ngorongoro crater)	Low (Class I)	RFLP	low	Bottleneck	Yuhki and O'Brien 1990
Sweedish moose	Low (Class II)	RFLP, SSCP	Moderate	Solitary lifestyle Ancient bottleneck?	Ellegren et al. 1996
Musk Ox, Fallow Deer	Monomorphic (Class II)	Seq	Low	?	Mikko et al. 1999
Prezwalski's Horse	Low (Class II)	Seq	?	Bottleneck	Hedrick et al. 1999
Arabian Oryx	Low (Class II)	Seq	?	Bottleneck	Hedrick et al. 2000a
Bontebok	Low (Class II)	Seq	?	Bottleneck	van der Walt et al. 2001
Hungarian meadow viper	Low (Class I)	RFLP	?	Bottleneck	Ujvari et al. 2002

Selection intensity is measured using the selection coefficient (s), and can be defined as the reduction in fitness of a genotype compared to the fittest genotype (Klein et al. 1993b), or the selective advantage of heterozygotes over homozygotes (Edwards and Hedrick 1998). The value of s is between 0 and 1, with $s = 0$ indicating no selection for polymorphism. There are a number of ways of estimating s , either from population data or from the ratio of non-synonymous to synonymous substitutions (reviewed in Edwards and Hedrick 1998, and

Hedrick and Kim 1999). Most estimates of s for MHC genes are low (less than 0.05). For example, Satta et al. (1994) found that for human class I and class II loci, the selection coefficients ranged from 0.0007 for *DPB1* to 0.042 for *HLA-A*. Similarly, the levels of selection for protective alleles against malaria are estimated to be 0.029 and 0.038 for *HLA-B*5301* and *DRB1*1302*, respectively (Hedrick and Kim 1999). However, s may vary greatly depending on the experimental context, for example estimates based on deviation from Hardy-Weinberg equilibrium can give values as high as 0.3 in the generation examined (Hedrick and Kim 1999). It is also important to note that the selective advantage of heterozygotes over homozygotes may not be constant for all genotypes. For instance, Richman et al. (2001) found that heterozygotes with divergent alleles have a selective advantage relative to heterozygotes with similar alleles. This is consistent with the idea of allelic supertypes, as discussed previously.

Where the selection coefficient is low, large sample sizes are required to demonstrate selection experimentally. This may partly account for inconsistencies in studies of correlations between MHC alleles and pathogens, as many studies may employ a sample size that is too small to reliably detect an association (Hill 1998). Selection coefficients are likely to be lower in organisms where multiple MHC loci are expressed (i.e. mammals), as the chance of all peptides from a pathogen being able to evade all MHC molecules is small. By contrast, in the chicken only one MHC molecule from each class is expressed, and this lack of redundancy means that associations between MHC haplotypes and pathogens are likely to be stronger and easier to detect (Kaufman et al. 1995b). For other birds, however, the situation may be more like that for mammals.

The apparent redundancy of the mammalian MHC appears to be at odds with the high levels of polymorphism observed in many mammalian populations. However, mathematical models show that even weak selection can maintain a significant level of diversity, particularly when gene conversion also occurs (Ohta 1997). It has been suggested that even selection pressure of less than 0.05 may be enough to retain polymorphisms for long periods of time, provided the effective population size remains large (Satta et al. 1994). Also, selection intensity may vary considerably over time and space, and elevated rates of nonsynonymous to synonymous substitutions may reflect selection pressure from past epidemics rather than current selective forces. It has been suggested that selection is only strong when the spectrum of parasites an

organism is exposed to changes dramatically (for instance when a species moves into a totally different niche) and alleles may be nearly neutral the rest of the time (Klein 1987).

1.4.4 Pathogen resistance, MHC diversity and conservation

Infectious diseases are increasingly being recognised as a major threat to endangered species (Deem et al. 2001, Lafferty and Gerber 2002). It is thought that low levels of MHC diversity may increase this threat by rendering populations more susceptible to pathogens (O'Brien and Evermann 1988). Thus, analysis of MHC diversity is becoming common in studies of endangered species (e.g. Hedrick and Parker 1998, Hedrick et al. 1999, 2000a, 2000b), and such functional markers are becoming increasingly relevant in defining conservation units (Hedrick 2001). It has even been suggested that captive breeding programs should focus primarily on preserving MHC diversity (Hughes 1991, but see Vrijenhoek and Leberg 1991, Gilpin and Wills 1991, and Miller and Hedrick 1991). However, the link between population decline from disease and low MHC diversity is difficult to demonstrate. Several of the populations in Table 1.2 have not only survived for a long period with low MHC variation, but have increased dramatically in size, with no apparent increased susceptibility to disease (e.g. Scandinavian beavers, Ellegren et al. 1993, Swedish moose and musk ox, Mikko et al. 1999).

The cheetah is the most commonly cited example of a species with low MHC diversity and increased susceptibility to disease (O'Brien et al. 1983, 1985). Multiple outbreaks of FIPV (feline infectious peritonitis virus) have decimated several captive cheetah populations (O'Brien et al. 1985), however it was not possible to establish an explicit connection between particular MHC alleles or MHC homozygosity and disease susceptibility. In the wild, the cheetah appears to be more at risk from predation on juveniles than low genetic diversity (Caro and Laurenson 1994). In fact, there are few examples of correlations between MHC diversity and particular diseases in wild animal populations, as the majority of such studies have occurred humans, or animals of economic importance (see Table 1.1). One of the few extensive studies has been done on Soay sheep, where Paterson et al. (1998) showed that both juvenile survival and resistance to intestinal nematodes were significantly associated with particular MHC alleles. A correlation between inbreeding and susceptibility to parasites has also been measured in Soay sheep, as sheep with high homozygosity at microsatellite loci are more susceptible to parasitism by nematodes and are less likely to survive harsh winters

(Coltman et al. 1999). The link between the MHC-associated disease susceptibility and inbreeding-associated susceptibility is not clear, however. Other studies have also reported a correlation between the level of outbreeding and disease resistance, without invoking a specific MHC related effect (e.g. Cassinello et al. 2001, Hedrick et al. 2001, Acevedo-Whitehouse et al. 2003).

Disease susceptibility is not always related to low MHC variation. For example, bighorn sheep have declined 40-fold from the time of European settlement, primarily because of infectious disease transmission from livestock, yet have extensive levels of MHC variation (Gutierrez-Espeleta et al. 2001). Thus, it is clear that the link between MHC variation and pathogen resistance is not universal. Non-genetic factors such as host density, proximity to disease reservoirs (i.e. domestic livestock), climate, and habitat modification may significantly contribute to disease susceptibility in some populations (Daszak et al. 2000, Lafferty and Gerber 2002). In addition, disease resistance is usually polygenic, and many non-MHC loci may also be associated with disease resistance (Hill 1998). For this reason, maintaining genome-wide diversity may be a more appropriate goal of captive breeding programs.

1.5 Conservation genetics and the black robin

The Chatham Islands black robin *Petroica traversi* (Passeriformes: Petroicidae) is a small insectivorous passerine found only on South East (Rangatira) and Mangere Islands in the Chatham Islands group, 850 km east of the South Island of New Zealand. The black robin was widespread throughout the Chatham Islands prior to European settlement in the 19th century, however habitat loss and introduction of predators saw the species restricted to an isolated rock stack known as Little Mangere Island by the end of the 19th century (Butler and Merton 1992). Little Mangere Island contains less than 9ha of forest and scrub and is likely to have supported only 20-30 birds at one time (Merton 1990). By 1979 the entire species consisted of only 5 individuals, including only one viable breeding pair, and was on the brink of extinction. However, an intensive management program, in which birds were translocated to nearby Mangere and South East Islands, and black robin egg production was increased by fostering out eggs to the Chatham Island tomtit, was successful, and by 1990 numbers had increased to over 100 birds (Butler and Merton 1992). The black robin population currently numbers around 250 birds (H. Aikman, pers comm), all of whom are derived from a single

breeding pair: the female “Old Blue” and the male “Old Yellow” (Merton, 1990, Butler and Merton, 1992).

As a result of its long history of isolation in a single small population, the black robin is highly inbred, with frequent matings between siblings, and between parents and offspring (Butler and Merton 1992), so is ideal for studying the genetic consequences of a severe population bottleneck. Their close relatives, the New Zealand bush robins (*Petroica australis*) are not endangered, but are patchily distributed throughout the North and South Islands of New Zealand. In 1973 new populations of South Island robin (*P. a. australis*) were established on offshore islands as a “practice run” for the black robin translocations (Flack 1974), and now provide a useful comparison to the black robin. The black robin has amongst the lowest levels of neutral minisatellite DNA variation for any wild bird, with profiles being almost identical from one bird to the next (Arden and Lambert 1997), whereas even translocated populations of South Island robin have retained moderate levels of variation (Arden et al. 1997). There is little evidence for inbreeding depression in the black robin population however (Butler and Merton 1992, Holmes 1994, Arden and Lambert 1997), and numbers are continuing to increase without intensive management.

Analysis of functional, fitness-related markers (such as MHC loci) may complement existing genetic and demographic information in assessing the long-term viability of the black robin population. The population does not appear to be particularly susceptible to pathogens, although parasitic mite infestations and episodes of avian pox have been reported (Tisdall and Merton 1988, Butler and Merton 1992). A more detailed analysis of pathogen loads in the black robin is currently being undertaken (J. Anderson, pers comm), and may allow the link between genetic diversity and disease susceptibility to be explored. The black robin and South Island robin populations are ideal for testing ideas about the effect of population bottlenecks on MHC variation, and isolation of MHC sequences from these two closely related species will add to the data on the organisation and complexity of MHC genes in passerines.

1.6 Thesis aims and outline

The major aims of this thesis are two-fold:

- (1) to characterise expressed class II B MHC loci from New Zealand robins and to analyse the evolutionary processes influencing MHC genes in birds.
- (2) to analyse MHC variation in the black robin and compare it with the source and bottlenecked populations of South Island robin in order to assess the effect of population bottlenecks on MHC variation.

Analysis of expressed MHC genes in an endangered species is complicated by the requirement for high quality RNA. Blood samples are often the only available tissue, and in **Chapter Two** the most appropriate method for preserving blood for RNA extraction is investigated. In addition, a protocol is given for isolating nearly full-length class II B MHC cDNA sequences using an RT-PCR based method, as an alternative to constructing a cDNA library. **Chapter Three** details the isolation and characterisation of class II B MHC genes in New Zealand robins. RFLP analysis is used to gain an overview of the complexity of the class II B MHC region, and the RT-PCR methods developed in chapter two are used to isolate transcribed class II B genes from both black robin and South Island robin. Key features of the evolution of class II MHC genes in New Zealand robins are investigated. Preliminary data on a divergent, and possibly non-functional class II B MHC sequence isolated from black robin, is given in **Chapter Four**. In **Chapter Five**, MHC variation is analysed using RFLP, and PCR/sequencing. Intron sequences of the transcribed MHC loci are isolated and provide additional information on the evolution of class II B genes. The effect of population bottlenecks on MHC variation is measured by comparing black robin individuals with South Island robins from artificially bottlenecked populations and their respective source populations. The implications of low MHC diversity for conservation are also discussed. The comparison between the black robin and South Island robin forms an important component of this thesis. Thus, in **Chapter Six**, the level of relatedness between these two species is assessed using mitochondrial DNA sequences, and the evolutionary history of the New Zealand *Petroica* species is discussed. **Chapter Seven** summarises the major findings of this study, and discusses possible future research.

1.7 References

- Acevedo-Whitehouse K, Gulland F, Greig D, Amos W (2003) Disease susceptibility in California sealions. *Nature* 422: 35
- Alberts B, Bray D, Lewis J, Raff M, Roberts K, Watson JD (1994) *Molecular Biology of the Cell*. Garland Publishing, New York
- Alberts SC and Ober C (1993) Genetic variability in the major histocompatibility complex: a review of non-pathogen mediated selective mechanisms. *Yrbk. Phy. Anthropol* 36: 71-89
- Almarri A and Batchelor JR (1994) HLA and hepatitis-B infection. *Lancet* 344: 1194-1195
- Alric L, Fort M, Izopet J, Vinel JP, Charlet JP, Selves J, Puel J, Pascal JP, Duffaut M, Abbal M (1997) Genes of the major histocompatibility complex class II influence the outcome of hepatitis C virus infection. *Gastroenterology* 113: 1675-1681
- Amos W and Balmford A (2001) When does conservation genetics matter? *Heredity* 87: 257-265
- Andersson L, Gustafsson K, Jonsson A, Rask L (1991) Concerted evolution in a segment of the first domain exon of polymorphic MHC class II B loci. *Immunogenetics* 33: 235-242
- Andersson L and Mikko S (1995) Generation of MHC class II diversity by intra- and intergenic recombination. *Immunol. Rev.* 143: 5-12
- Apanius V, Penn D, Slev PR, Ruff LR, Potts WK (1997) The nature of selection on the major histocompatibility complex. *Crit. Rev. Immunol.* 17: 179-224
- Arden SL and Lambert DM (1997) Is the black robin in genetic peril? *Mol. Ecol.* 6: 21-28
- Arden SL, Lambert DM, Rodrigo AG, McLean IG (1997) The effects of population bottlenecks on multilocus DNA variation in robins. *J. Heredity* 88: 179-186
- Arkush KD, Giese AR, Mendonca HL, McBride AM, Marty GD, Hedrick PW (2002) Resistance to three pathogens in the endangered winter-run chinook salmon (*Oncorhynchus tshawytscha*): effects of inbreeding and major histocompatibility complex genotypes. *Can. J. Fish. Aquat. Sci.* 59: 966-975
- Bergstrom TF, Josefsson A, Erlich HA, Gyllensten U (1998) Recent origin of *HLA-DRB1* alleles and implications for human evolution. *Nature Genetics* 18: 237-242
- Bjorkman PJ, Saper MA, Samraoui B, Bennett WS, Strominger JL, Wiley DC (1987a) Structure of the human class I histocompatibility antigen, HLA-A2. *Nature* 329: 506-512
- Bjorkman PJ, Saper MA, Samraoui B, Bennett WS, Strominger JL, Wiley DC (1987b) The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. *Nature* 329: 512-518
- Bodmer W (1972) Evolutionary significance of the HLA system. *Nature* 237: 139-145
- Brahmajothi V, Pitchappan RM, Kakkanaiah VN, Sashidhar M, Rajaram K, Ramu S, Palanimurugan K, Paramasivan CN, Prabhakar R (1991) Association of Pulmonary Tuberculosis and Hla in South-India. *Tubercle* 72: 123-132
- Briles WE, Briles RW, Taffs RE, (1983) Resistance to a malignant lymphoma in chickens is mapped to subregion of major histocompatibility (B) complex. *Science* 219: 977-978
- Briles WE, Goto RM, Auffray C, Miller MM (1993) A polymorphic system related to but genetically independent of the chicken major histocompatibility complex. *Immunogenetics* 37: 408-414

- Brodsky FM, Lem L, Solache A, Bennett EM (1999) Human pathogen subversion of antigen presentation. *Immunol. Rev.* 168: 199-215
- Brown JH, Jardetzky TS, Gorga JC, Stern LJ, Urban RG, Strominger JL, Wiley DC (1993) Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature* 364: 33-39
- Brusic V, Rudy G, Harrison LC (1998) MHCPEP, a database of MHC-binding peptides: update 1997. *Nucl. Acids Res.* 26: 368-371
- Butler D and Merton D (1992) *The Black Robin: Saving the World's Most Endangered Bird.* Oxford University Press
- Buus S (1999) Description and prediction of peptide-MHC binding: the 'human MHC project'. *Curr. Op. Immunol.* 11: 209-213
- Caro TM and Laurenson MK (1994) Ecological and genetic factors in conservation: a cautionary tale. *Science* 263: 485-486
- Caron LA, Abplanalp H, Taylor RL (1997) Resistance, susceptibility, and immunity to *Eimeria tenella* in major histocompatibility (B) complex congenic lines. *Poultry Science* 76: 677-682
- Carrington M, Nelson GW, Martin MP, Kissner T, Vlahov D, Goedert JJ, Kaslow R, Buchbinder S, Hoots K, O'Brien SJ (1999) HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage. *Science* 283: 1748-1752
- Cassinello J, Gomendio M, Roldan ERS (2001) Relationship between coefficient of inbreeding and parasite burden in endangered gazelles. *Cons. Biol.* 15: 1171-1174
- Caughley G (1994) Directions in conservation biology. *J. Anim. Ecol.* 63: 215-244
- Chelvanayagam G, Apostolopoulos V, McKenzie IFC (1997) Milestones in the molecular structure of the major histocompatibility complex. *Protein Engineering* 10: 471-474
- Cohen S (2002) Strong positive selection and habitat-specific amino acid substitution patterns in *Mhc* from an estuarine fish under intense pollution stress. *Mol. Biol. Evol.* 19: 1870-1880
- Coltman DW, Pilkington JG, Smith JA, Pemberton JM (1999) Parasite-mediated selection against inbred Soay sheep in a free-living, island population. *Evolution* 53: 1259-1267
- Daszak P, Cunningham AA, Hyatt AD (2000) Emerging infectious diseases of wildlife - threats to biodiversity and human health. *Science* 287: 443-449
- de Campos-Lima P-O, Gavioli R, Zhang Q-J, Wallace LE, Dolcetti R, Rowe M, Rickinson AB, Masucci MG (1993) HLA-A11 epitope loss isolates of Epstein-Barr virus from a highly A11+ population. *Science* 260: 98-100
- Deem SL, Karesh WB, Weisman W (2001) Putting theory into practice: wildlife health in conservation. *Cons. Biol.* 15: 1224-1233
- Doherty PC and Zinkernagel RM (1975) Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. *Nature* 256: 50-52
- Donnes P and Elofsson A (2002) Prediction of MHC class I binding peptides, using SVMHC. *Bmc Bioinformatics* 3: art. no.-25
- Edwards SV, Grahn M, Potts WK (1995a) Dynamics of *Mhc* evolution in birds and crocodylians: amplification of class II genes with degenerate primers. *Mol. Ecol.* 4: 719-729

- Edwards SV, Wakeland EK, Potts WK (1995b) Contrasting histories of avian and mammalian *Mhc* genes revealed by class II B sequences from songbirds. *Proc. Natl. Acad. Sci. USA* 92: 12200-12204
- Edwards SV, Gasper J, March M (1998) Genomics and polymorphism of *Agph-DAB1*, an MHC Class II B gene in red-winged blackbirds (*Agelaius phoeniceus*). *Mol. Biol. Evol.* 15: 236-250
- Edwards SV and Hedrick PW (1998) Evolution and ecology of MHC molecules: from genomics to sexual selection. *Trends Ecol. Evol* 13: 305-311
- Edwards SV, Hess CM, Gasper J, Garrigan D (1999) Toward an evolutionary genomics of the avian Mhc. *Immunol. Rev.* 167: 119-132
- Edwards SV, Gasper J, Garrigan D, Martindale D, Koop BF (2000a) A 39kb sequence around a blackbird *Mhc* Class II gene: Ghost of selection past and songbird genome architecture. *Mol. Biol. Evol.* 17: 1384-1395
- Edwards SV, Nusser J, Gasper JS (2000b) Characterization and evolution of major histocompatibility complex (MHC) genes in non-model organisms, with examples from birds. *In* AJ Baker (ed.): *Molecular Methods in Ecology*. Blackwell Science, pp. 168-207
- Eggert F, Muller-Ruchholtz W, Ferstl R (1999) Olfactory cues associated with the major histocompatibility complex. *Genetica* 104: 191-197
- Egid K and Brown JL (1989) The major histocompatibility complex and female mating preferences in mice. *Animal Behaviour* 38: 548-549
- Ekblom R, Grahn M, Hoglund J (2003) Patterns of polymorphism in the MHC class II of a non-passerine bird: the great snipe (*Gallinago media*). *Immunogenetics* 54: 734-741
- Ellegren H, Hartman G, Johansson M, Andersson L (1993) Major histocompatibility complex monomorphism and low levels of DNA fingerprinting variability in a reintroduced and rapidly expanding population of beavers. *Proc. Natl. Acad. Sci. USA* 90: 8150-8153
- Ellegren H, Mikko S, Wallin K, Andersson L (1996) Limited polymorphism at major histocompatibility complex (MHC) loci in the Swedish moose A-alces. *Mol. Ecol.* 5: 3-9
- Erlich HA and Gyllenstein UB (1991) Shared epitopes among HLA class II alleles: gene conversion, common ancestry, and balancing selection. *Immunol. Today* 12: 411-414
- Falk K, Rotzschke O, Stevanovic S, Jung G, Rammensee H-G (1991) Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature* 351: 290-296
- Flack JAD (1974) Chatham Island black robin. *Wildlife - a Review, New Zealand Wildlife Service* 5: 25-34
- Flores-Villanueva PO, Yunis EJ, Delgado JC, Vittinghoff E, Buchbinder S, Leung JY, Ugliarolo AM, Clavijo OP, Rosenberg ES, Kalams SA, Braun JD, Boswell SL, Walker B, Goldfield AE (2001) Control of HIV-1 viremia and protection from AIDS are associated with HLA-Bw4 homozygosity. *Proc. Natl. Acad. Sci. USA* 98: 5140--5145
- Frankham R (1995) Conservation Genetics. *Ann. Rev. Genetics* 29: 305-327
- Gasper JS, Shiina T, Inoko H, Edwards SV (2001) Songbird genomics: Analysis of 45 kb upstream of a polymorphic Mhc class II gene in red-winged blackbirds (*Agelaius phoeniceus*). *Genomics* 75: 26-34
- Gilpin M and Wills C (1991) MHC and captive breeding: a rebuttal. *Cons. Biol.* 5: 554-555
- Groombridge JJ, Jones CG, Bruford MW, Nichols RA (2000) 'Ghost' alleles of the Mauritius kestrel. *Nature* 403: 616

- Gu X and Nei M (1999) Locus specificity of polymorphic alleles and evolution by a birth-and-death process in mammalian MHC genes. *Mol. Biol. Evol.* 16: 147-156
- Guillemot F, Billault A, Pourquie O, Behar G, Chausse A-M, Zoorob R, Kreiblich G, Auffray C (1988) A molecular map of the chicken major histocompatibility complex: the class II B genes are closely-linked to the class I genes and the nucleolar organiser. *EMBO J.* 7: 2775
- Gutierrez-Espeleta GA, Hedrick PW, Kalinowski ST, Garrigan D, Boyce WM (2001) Is the decline of bighorn sheep from infectious disease the result of low MHC variation? *Heredity* 86: 439-500
- Hambuch TM and Lacey EA (2002) Enhanced selection for MHC diversity in social tuco-tucos. *Evolution* 56: 841-845
- Hamilton WD and Zuk M (1982) Heritable true fitness and bright birds: A role for parasites? *Science* 218: 384-387
- Hedrick PN and Black FL (1997) HLA and mate selection: no evidence in South Amerindians. *Am. J. Hum. Genet.* 61: 505-511
- Hedrick PW and Parker KM (1998) MHC variation in the endangered Gila topminnow. *Evolution* 52: 194-199
- Hedrick PW and Kim TJ (1999) Genetics of complex polymorphisms: parasites and the maintenance of the major histocompatibility complex variation. In RS Singh and CB Krimbas (eds.): *Evolutionary Genetics: From Molecules to Morphology*. Cambridge University Press
- Hedrick PW, Parker KM, Miller EL, Miller PS (1999) Major histocompatibility complex variation in the endangered Przewalski's horse. *Genetics* 152: 1701-1710
- Hedrick PW, Karen MP, Gutierrez-Espeleta GA, Rattink A, Lievers K (2000a) Major histocompatibility complex variation in the Arabian oryx. *Evolution* 54: 2145-2151
- Hedrick PW, Lee RN, Parker KM (2000b) Major histocompatibility complex (MHC) variation in the endangered Mexican wolf and related canids. *Heredity* 85: 617-624
- Hedrick PW (2001) Conservation genetics: where are we now? *Trends Ecol. Evol.* 16: 629-636
- Hedrick PW, Kim TJ, Parker KM (2001) Parasite resistance and genetic variation in the endangered Gila topminnow. *Animal Conservation* 4: 103-109
- Hedrick PW (2002) Pathogen resistance and genetic variation at MHC loci. *Evolution* 56: 1902-1908
- Hedrick PW, Lee RN, Garrigan D (2002) Major histocompatibility complex variation in red wolves: evidence for common ancestry with coyotes and balancing selection. *Mol. Ecol.* 11: 1905-1913
- Hess CM, Gasper J, Hoekstra HE, Hill CE, Edwards SV (2000) MHC Class II pseudogene and genomic signature of a 32 kb cosmid in the house finch (*Carpodacus mexicanus*). *Genome Res.* 10: 613-623
- Hess CM and Edwards SV (2002) The evolution of the major histocompatibility complex in birds. *BioScience* 52: 423-431
- Hill AVS, Allsopp CEM, Kwiatkowski D, Anstey NM, Twumasi P, Rowe PA, Bennett S, Brewster D, McMichael AJ, Greenwood BM (1991) Common West African HLA antigens are associated with protection from severe malaria. *Nature* 352: 595-600
- Hill AVS (1998) The immunogenetics of human infectious diseases. *Ann. Rev. Immunol.* 16: 593-617
- Hoelzel AR, Stephens JC, O'Brien SJ (1999) Molecular genetic diversity and evolution at the MHC DQB locus in four species of pinnipeds. *Mol. Biol. Evol.* 16: 611-618

- Hogstrand K and Bohme J (1994) A determination of the frequency of gene conversion in unmanipulated mouse sperm. *Proc. Natl. Acad. Sci. USA* 91: 9921-9925
- Hogstrand K and Bohme J (1999) Gene conversion can create new MHC alleles. *Immunol. Rev.* 167: 305-317
- Holmes SL (1994) Parentage and Molecular Genetics of New Zealand Robins. MSc thesis, University of Auckland, New Zealand
- Hughes AL and Nei M (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* 335: 167-170
- Hughes AL and Nei M (1989) Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. *Proc. Natl. Acad. Sci. USA* 86: 958-962
- Hughes A, Ota T, Nei M (1990) Positive Darwinian selection promotes charge profile diversity in the antigen-binding cleft of class I major-histocompatibility-complex molecules. *Mol Biol Evol* 7: 515-524
- Hughes AL (1991) MHC polymorphism and the design of captive breeding programs. *Cons. Biol.* 5: 249-251
- Hughes AL, Hughes MK, Watkins DL (1993) Contrasting roles of interallelic recombination at the *HLA-A* and *HLA-B* loci. *Genetics* 133: 669-680
- Jeffery KJM, Usuku K, Hall SE, Matsumoto W, Taylor GP, Procter J, Bunce M, Ogg GS, Welsh KI, Weber JN, Lloyd AL, Nowak MA, Nagai M, Kodama D, Izumo S, Osame M, Bangham CRM (1999) HLA alleles determine human T-lymphotropic virus-I (HTLV-I) proviral load and the risk of HTLV-I-associated myelopathy. *Proc. Natl. Acad. Sci. USA* 96: 3848-3853
- Kaslow RA, Carrington M, Apple R, Park L, Munoz A, Saah AJ, Goedert JJ, Winkler C, O'Brien SJ, Rinaldo C, Detels R, Blattner W, Phair J, Erlich H, Mann DL (1996) Influence of combinations of human major histocompatibility complex genes on the course of HIV-1 infection. *Nature Medicine* 2: 405-411
- Kast WM, Brandt RMP, Sidney J, Drijfhout JW, Kubo RT, Grey HM, Melief CJM, Sette A (1994) Role of HLA-A motifs in identification of potential CTL epitopes in human papillomavirus type-16 E6 and E7 proteins. *J. Immunol.* 152: 3904-3912
- Kaufman J, Skjoeedt K, Salomonsen J (1990) The MHC molecules of nonmammalian vertebrates. *Immunol. Rev.* 113: 83-117
- Kaufman J, Bumstead N, Miller M, Riegert P, Salomonsen J (1995a) The chicken class II a gene is located outside of the B complex. *In* TF Davison, N Bumstead, and P Kaiser (eds.): *Advances in avian immunology research*. Carfax, Abingdon, pp. 119-127
- Kaufman J, Volk H, Wallny H-J (1995b) A "Minimal Essential MHC" and an "Unrecognised MHC": Two extremes in selection for polymorphism. *Immunol. Rev.* 143: 63-88
- Kaufman J and Salomonsen (1997) The "minimal essential MHC" revisited: Both peptide-binding and cell surface expression level of MHC molecules are polymorphisms selected by pathogens in chickens. *Hereditas* 127: 67-73
- Kaufman J and Venugopal K, (1998) The importance of MHC for Rous sarcoma virus and Marek's disease virus - some Payne-ful considerations. *Avian Pathology* 27: S82-S87
- Kaufman J (1999) Co-evolving genes in MHC haplotypes: the "rule" for nonmammalian vertebrates? *Immunogenetics* 50: 228-236
- Kaufman J, Jacob J, Iain S, Walker B, Milne S, Beck S, Salomonsen J (1999a) Gene organisation determines evolution of function in the chicken MHC. *Immunol. Rev.* 167: 101-117

- Kaufman J, Milne S, Gobel TWF, Walker BA, Jacob JP, Auffray C, Zoorob R, Beck S (1999b) The chicken B locus is a minimal essential major histocompatibility complex. *Nature* 401: 923-925
- Kaufman J (2000) The simple chicken major histocompatibility complex: life and death in the face of pathogens and vaccines. *Phil. Trans. R. Soc. Lond. B* 355: 1077-1084
- Keller LF and Waller DM (2002) Inbreeding effects in wild populations. *Trends Ecol. Evol* 17: 230-241
- Kelly AP, Monaco JJ, Cho S, Trowsdale J (1991) A new human HLA class II-related locus, DM. *Nature* 353: 571-573
- Klein D, Ono H, O'hUigin C, Vincek V, Goldschmidt T, Klein J (1993a) Extensive MHC variability in cichlid fishes of Lake Malawi. *Nature* 364: 330-334
- Klein J (1986) *Natural History of the Major Histocompatibility Complex*. John Wiley and Sons, New York,
- Klein J (1987) Origin of major histocompatibility complex polymorphism: the trans-species hypothesis. *Human Immunol.* 19: 155-162
- Klein J, Satta Y, O'hUigin C, Takahata N (1993b) The molecular descent of the major histocompatibility complex. *Ann. Rev. Immunol* 11: 269-295
- Klein J, Sato A, Nagl S, O'hUigin C (1998a) Molecular trans-species polymorphism. *Ann. Rev. Ecol. Syst.* 29: 1-21
- Klein J, Sato A, O'hUigin C (1998b) Evolution by gene duplication in the major histocompatibility complex. *Cytogenet Cell Genet* 80: 123-127
- Klein J (2001) George Snell's first foray into the unexplored territory of the Major Histocompatibility Complex. *Genetics* 159: 435-439
- Kulski JK, Shiina T, Anzai T, Kohara S, Inoko H (2002) Comparative genomic analysis of the MHC: the evolution of class I duplication blocks, diversity and complexity from shark to man. *Immunol. Rev.* 190: 95-122
- Lafferty KD and Gerber LR (2002) Good medicine for conservation biology: the intersection of epidemiology and conservation theory. *Cons. Biol.* 16: 593-604
- Lamont SJ, Bolin C, Chevillat N, (1987) Genetic resistance to fowl cholera is linked to the major histocompatibility complex. *Immunogenetics* 25: 284-289
- Lamont SJ (1998) Impact of genetics on disease resistance. *Poultry Science* 77: 1111-1118
- Landry C, Garant D, Duchesne P, Bernatchez L (2001) 'Good genes as heterozygosity': the major histocompatibility complex and mate choice in Atlantic salmon (*Salmo salar*). *Proc. R. Soc. Lond. B.* 268: 1279-1285
- Lanfords A, Lohm J, Grahn M, Anderson O, von Schantz T (2001a) Association between major histocompatibility complex class IIB alleles and resistance to *Aeromonas salmonicida* in Atlantic salmon. *Proc. R. Soc. Lond. B.* 268: 479-485
- Lanfords K, Lohm J, von Schantz T (2001 b) Allelic polymorphism in MHC class IIB in four populations of Atlantic salmon (*Salmo salar*). *Immunogenetics* 53: 329-336
- Lawlor DA, Zemmour J, Ennis PD, Parham P (1990) Evolution of class-I MHC genes and proteins: from natural selection to thymic selection. *Ann. Rev. Immunol* 8: 23-63

- Lohm J, Grahn M, Langefors A, Andersen O, Storset A, von Schantz T (2002) Experimental evidence for major histocompatibility complex-allele-specific resistance to a bacterial infection. *Proc. R. Soc. Lond. B.* 269: 2029-2033
- Macklin KS, Ewald SJ, Norton RA (2002) Major histocompatibility complex effect on cellulitis among different chicken lines. *Avian Pathology* 31: 371-376
- Malaga-Trillo E, Zaleska-Rutczynska Z, McAndrew B, Vincek V, Figueroa F, Sultmann H, Klein J (1998) Linkage Relationships and Haplotype Polymorphism Among Cichlid Mhc Class II B Loci. *Genetics* 149: 1527-1537
- Markow T, Hedrick PW, Zuerlein K, Danilovs J, Martin J, Vyvial T, Armstrong C (1993) HLA polymorphism in the Havasupai: evidence for balancing selection. *Am. J. Hum. Genet.* 53: 943-952
- Martinsohn JT, Sousa AB, Guethlein LA, Howard JC (1999) The gene conversion hypothesis of MHC evolution: a review. *Immunogenetics* 50: 168-200
- Maruyama T and Nei M (1981) Genetic variability maintained by mutation and overdominant selection in finite populations. *Genetics* 98: 441-459
- Merton D (1990) The Chatham Island black robin: How the world's most endangered bird was saved from extinction. *Forest and Bird* 21: 14-19
- Messaoudi I, Guevara Patino JA, Dyal R, LeMaout J, Nikolich-Zugich J (2002) Direct link between mhc polymorphism, T cell avidity, and diversity in immune defense. *Science* 298: 1797-1800
- Meyer D and Thomson G (2001) How selection shapes variation of the human major histocompatibility complex: a review. *Annals of Human Genetics* 65: 1-26
- MHC Sequencing Consortium (1999) Complete sequence and gene map of a human major histocompatibility complex. *Nature* 401: 921-923
- Migueles SA, Sabbaghian MS, Shupert WL, Bettinotti MP, Marincola FM, Martino L, Hallahan CW, Selig SM, Schwartz D, Sullivan J, Connors M (2000) HLA B*5701 is highly associated with restriction of virus replication in a subgroup of HIV-infected long term nonprogressors. *Proc. Natl. Acad. Sci. USA* 97: 2709-2714
- Mikko S, Roed K, Schmutz S, Andersson L (1999) Monomorphism and polymorphism at *Mhc DRB* loci in domestic and wild ruminants. *Immunol. Rev.* 167: 169-178
- Miller HC, Lambert DM, Millar CD, Robertson BC, Minot EO (2003) Minisatellite DNA profiling detects lineages and parentage in the endangered kakapo (*Strigops habroptilus*) despite low microsatellite DNA variation. *Conservation Genetics* 4: 265-274
- Miller KM, Kaukinen KH, Schulze AD (2002) Expansion and contraction of major histocompatibility complex genes: a teleostean example. *Immunogenetics* 53: 941-963
- Miller P and Hedrick PW (1991) MHC polymorphism and the design of captive breeding programs: simple solutions are not the answer. *Cons. Biol.* 5: 556-557
- Minton EJ, Smillie D, Neal KR, Irving WL, Underwood JCE, James V (1998) Association between MHC class II alleles and clearance of circulating hepatitis C virus. *J. Infect. Dis.* 178: 39-44
- Miyada CG, Klofolt C, Reyes AA, McLaughlin-Taylor E, Wallace RB (1985) Evidence that polymorphism in the murine major histocompatibility complex may be generated by the assortment of subgene sequences. *Proc. Natl. Acad. Sci. USA* 82: 2890-2894

- Moore CB, John M, James IR, Christiansen FT, Witt CS, Mallal SA (2002) Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level. *Science* 296: 1439-1443
- Nei M and Hughes AL (1991) Polymorphism and evolution of the major histocompatibility complex loci in mammals. *In* RK Selander, AG Clark, and TS Whittam (eds.): *Evolution at the Molecular Level*. Sinauer Associates Inc
- Nei M, Gu X, Sitnikova T (1997) Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proc. Natl. Acad. Sci. USA* 94: 7799-7806
- Nizetic D, Stevanovic M, Soldatovic B, Savic I, Crkvenjakov R (1988) Limited polymorphism of both classes of MHC genes in four different species of the Balkan mole rat. *Immunogenetics* 28: 91-98
- Nowak M (1992) The optimal number of major histocompatibility complex molecules in an individual. *Proc. Natl. Acad. Sci. USA* 89: 10896-10899
- Ober C, Weitkamp LR, Cox H, Dytch H, Kostyu D, Elias S (1997) HLA and mate choice in humans. *Am. J. Hum. Genet.* 61: 497-504
- Ober C, Hyslop T, Elias S, Weitkamp LR, Hauck WW (1998) HLA matching and fetal loss: results of a 10 year prospective study. *Human Reproduction* 13: 33-38
- O'Brien SJ, Wildt DE, Goldman D, Merrill C, Bush M (1983) The cheetah is depauperate in genetic variation. *Science* 221: 459-462
- O'Brien SJ, Roelke ME, Marker L, Newman A, Winkler CA, Meltzer D, Colly L, Evermann JF, Bush M, Wildt DE (1985) Genetic basis for species vulnerability in the cheetah. *Science* 227: 1428-1434
- O'Brien SJ and Evermann JF (1988) Interactive influence of infectious disease and genetic diversity in natural populations. *Trends Ecol. Evol.* 3: 254-259
- Ohta T (1997) Role of gene conversion in generating polymorphisms at major histocompatibility complex loci. *Hereditas* 127: 97-103
- Ohta T (1999) Effect of gene conversion on polymorphic patterns at major histocompatibility complex loci. *Immunol. Rev.* 167: 319-325
- O'hUigin C (1995) Quantifying the degree of convergence in primate Mhc-DRB genes. *Immunol. Rev.* 143: 123-140
- Ono H, Klein D, Vincek V, Figueroa F, O'hUigin C, Tichy H, Klein J (1992) Major histocompatibility complex class II genes of zebrafish. *Proc. Natl. Acad. Sci. USA* 89: 11886-11890
- Ou DW, Mitchell LA, Tingle AJ (1998) A new categorization of HLA DR alleles on a functional basis. *Human Immunology* 59: 665-676
- Parham P and Ohta T (1996) Population biology of antigen presentation by MHC class I molecules. *Science* 272: 67-74
- Paterson S and Pemberton JM (1997) No evidence for major histocompatibility complex-dependent mating patterns in a free-living ruminant population. *Proc. R. Soc. Lond. B.* 264: 1813-1819
- Paterson S, Wilson K, Pemberton JM (1998) Major histocompatibility complex variation associated with juvenile survival and parasite resistance in a large unmanaged ungulate population (*Ovis aries L.*). *Proc. Natl. Acad. Sci. USA.* 95: 3714-3719
- Penn D and Potts WK (1999) The evolution of mating preferences and major histocompatibility complex genes. *Am. Nat.* 153: 145-164

- Penn DJ (2002) The scent of genetic compatibility: Sexual selection and the major histocompatibility complex. *Ethology* 108: 1-21
- Penn DJ, Damjanovich K, Potts WK (2002) MHC heterozygosity confers a selective advantage against multiple-strain infections. *Proc. Natl. Acad. Sci. USA* 99: 11260-11264
- Pharr GT, Dodgson JB, Hunt HD, Bacon LD (1998) Class II MHC cDNAs in 15I₅ B-congenic chickens. *Immunogenetics* 47: 350-354
- Piontkivska H and Nei M (2003) Birth-and-death evolution in primate MHC class I genes: divergence time estimates. *Mol Biol Evol* 20: 601-609
- Plachy J, Pink JRL, Hala K (1992) Biology of the Chicken Mhc (B-Complex). *Crit. Rev. Immunol.* 12: 47-79
- Potts WK, Manning CJ, Wakeland EK, (1991) Mating patterns in seminatural populations of mice influenced by MHC genotype. *Nature* 352: 619-621
- Potts WK and Wakeland EK (1993) Evolution of MHC genetic diversity: a tale of incest, pestilence and sexual preference. *Trends Genet.* 9: 408-412
- Reusch TBH, Haberli MA, Aeschlimann PB, Milinski M (2001) Female sticklebacks count alleles in a strategy of sexual selection explaining MHC polymorphism. *Nature* 414: 300-302
- Richman AD, Herrera LG, Nash D (2001) MHC class II beta sequence diversity in the deer mouse (*Peromyscus maniculatus*): implications for models of balancing selection. *Mol. Ecol.* 10: 2765-2773
- Robinson J, Waller MJ, Parham P, Groot Nd, Bontrop R, Kennedy LJ, Stoehr P, Marsh SGE (2003) IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex. *Nucl. Acids. Res.* 31: 311-314
- Robinson JH and Delvig AA (2002) Diversity in MHC class II antigen presentation. *Immunology* 105: 252-262
- Salamon H, Klitz W, Easteal S, Gao XJ, Erlich HA, Fernandez-Vina M, Trachtenberg EA, (1999) Evolution of HLA class II molecules: allelic and amino acid site variability across populations. *Genetics* 152: 393-400
- Sato A, Figueroa F, Mayer WE, Grant PR, Grant BR, Klein J (2000) Mhc class II genes of Darwin's Finches: divergence by point mutations and reciprocal recombination. *In* M Kasahara (ed.): Major Histocompatibility Complex: Evolution, Structure and Function. Springer Verlag
- Satta Y, O'hUigin C, Takahata N, Klein J (1993) The synonymous substitution rate of the major histocompatibility complex loci in primates. *Proc. Natl. Acad. Sci. USA* 90: 7480-7484
- Satta Y, O'hUigin C, Takahata N, Klein J (1994) Intensity of natural selection at the major histocompatibility complex loci. *Proc. Natl. Acad. Sci. USA.* 91: 7184-7188
- Satta Y (1997) Effects of intra-locus recombination on *HLA* polymorphism. *Hereditas* 127: 105-112
- Satta Y, Li Y-J, Takahata N (1998) The neutral theory and natural selection in the HLA region. *Frontiers in BioScience* 3: 459-467
- Sebzda E, Mariathasan S, Ohteki T, Jones R, Bachmann MF, Ohashi PS (1999) Selection of the T cell repertoire. *Annu. Rev. Immunol.* 17: 829-874
- Seddon JM and Baverstock PR (1999) Variation on islands: major histocompatibility complex (*Mhc*) polymorphism in populations of the Australian bush rat. *Mol. Ecol.* 8: 2071-2079
- She JX, Boehme SA, Wang TW, Bonhomme F, Wakeland EK (1991) Amplification of major histocompatibility complex class II gene diversity by intraexonic recombination. *Proc. Natl. Acad. Sci. USA* 88: 453-457

- Sidney J, Grey HM, Kubo RT, Sette A (1996) Practical, biochemical and evolutionary implications of the discovery of HLA class II supermotifs. *Immunology Today* 17: 261-266
- Slade RW (1992) Limited MHC polymorphism in the southern elephant seal: implications for MHC evolution and marine mammal population biology. *Proc. R. Soc. Lond. B.* 249: 163-171
- Sliereendregt BL, Otting N, Vanbesouw N, Jonker M, Bontrop RE (1994) Expansion and Contraction of Rhesus Macaque Drb Regions by Duplication and Deletion. *J. Immunol.* 152: 2298-2307
- Sommer S, Schwab D, Ganzhorn JU (2002) MHC diversity of endemic Malagasy rodents in relation to geographic range and social system. *Behav. Ecol. Sociobiol.* 51: 214-221
- Starr TK, Jameson SC, Hogquist KA (2003) Positive and negative selection of T cells. *Annu. Rev. Immunol.* 21: 139-176
- Takahashi K, Rooney AP, Nei M (2000) Origins and divergence times of mammalian class II MHC gene clusters. *J. Heredity* 91: 198-204
- Takahata N and Nei M (1990) Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. *Genetics*: 967-978
- Takahata N, Satta Y, Klein J (1992) Polymorphism and balancing selection at major histocompatibility complex loci. *Genetics* 130: 925-938
- Takahata N and Satta Y (1998) Selection convergence, and intragenic recombination in HLA diversity. *Genetica* 103: 157-169
- Thomson G (1977) The effect of a selected locus on linked neutral loci. *Genetics* 85: 753-788
- Thursz MR, Kwiatkowski D, Allsopp CEM, Greenwood BM, Thomas HC, Hill AVS (1995) Association between an MHC class-II allele and clearance of hepatitis-B virus in the Gambia. *N. Engl. J. Med.* 332: 1065-1069
- Thursz MR, Thomas HC, Greenwood BM, Hill AVS (1997) Heterozygote advantage for HLA class II type in hepatitis B virus infection. *Nature Genetics* 17: 11-12
- Tisdall DJ and Merton DV (1988) Disease surveillance in the Chatham Islands black robin. *Surveillance* 15
- Trowsdale J, Groves V, Arnason A (1989) Limited MHC polymorphism in whales. *Immunogenetics*: 19-24
- Trowsdale J (1995) "Both man and bird and beast": comparative organisation of MHC genes. *Immunogenetics* 41: 1-17
- Ujvari B, Madsen T, Kotenko T, Olsson M, Shine R, Wittzell H (2002) Low genetic diversity threatens imminent extinction for the Hungarian meadow viper (*Vipera ursinii rakosiensis*). *Biol. Conserv.* 105: 127-130
- Valdes AM, McWeeney SK, Meyer D, Nelson MP, Thomson G (1999) Locus and population specific evolution in HLA class II genes. *Annals of Human Genetics* 63: 27-43
- van der Walt JM, Nel LH, Hoelzel AR (2001) Characterisation of major histocompatibility complex DRB diversity in the endemic South African antelope *Damaliscus pygargus*: a comparison in two subspecies with different demographic histories. *Mol. Ecol.* 10: 1679-1688
- Van Kacer L (2002) Major histocompatibility complex class I-restricted antigen processing and presentation. *Tissue Antigens* 60: 1-9
- Visentainer JEL, Tsuneto LT, Serra MF, Peixoto PRF, PetzIerler ML (1997) Association of leprosy with HLA-DR2 in a Southern Brazilian population. *Brazilian J. Med. Biol. Res.* 30: 51-59

- von Schantz T, Wittzell H, Goransson G, Grahn M, (1997) Mate choice, male condition-dependent ornamentation and MHC in the pheasant. *Hereditas* 127: 133-140
- Vrijenhoek RC and Leberg PL (1991) Let's not throw the baby out with the bathwater: A comment on management for MHC diversity in captive populations. *Cons. Biol.* 5: 252-254
- Wedekind C, Seebeck T, Bettens F, Paepke A (1995) MHC-dependent mate preferences in humans. *Proc. R. Soc. Lond. B.* 260: 245-249
- Wegner KM, Reusch TBH, Kalbe M (2003) Multiple parasites are driving major histocompatibility complex polymorphism in the wild. *J. Evol. Biol.* 16: 224-232
- Weiss EH, Mellor A, Golden L, Fahrner K, Simpson E, Hurst J, Flavell RA (1983) The structure of a mutant *H-2* gene suggests that the generation of polymorphism in *H-2* genes may occur by gene conversion-like events. *Nature* 301: 671-674
- Westerdahl H, Wittzell H, von Schantz T (1999) Polymorphism and transcription of Mhc class I genes in a passerine bird, the great reed warbler. *Immunogenetics* 49: 158-170
- Westerdahl H, Wittzell H, von Schantz T (2000) Mhc diversity in two passerine birds: no evidence for a minimal essential Mhc. *Immunogenetics* 52: 92-100
- Wisely SM, Buskirk SW, Fleming MA, McDonald DB, Ostrander EA (2002) Genetic diversity and fitness in black-footed ferrets before and during a bottleneck. *J. Heredity* 93: 231-237
- Wittzell H, Bernot A, Auffray C, Zoorob R (1999) Concerted evolution of two MHC Class II loci in pheasants and domestic chickens. *Mol. Biol. Evol.* 16: 479-490
- Yeager M and Hughes AL (1999) Evolution of the mammalian MHC: natural selection, recombination, and convergent evolution. *Immunol. Rev.* 167: 45-58
- Yuhki N and O'Brien SJ (1990) DNA variation of the mammalian major histocompatibility complex reflects genomic diversity and population history. *Proc. Natl. Acad. Sci. USA* 87: 836-840
- Zangenberg G, Huang MM, Arnheim N, Erlich HA (1995) New HLA-DPB1 alleles generated by interallelic gene conversion detected by analysis of sperm. *Nature Genetics* 10: 407-414
- Zekarias B, Ter Huurne A, Landman WJM, Rebel JMJ, Pol JMA, Gruys E (2002) Immunological basis of differences in disease resistance in the chicken. *Veterinary Research* 33: 109-125
- Zhu J and Nestor KE (1995) Survey of major histocompatibility complex class II haplotypes in four turkey lines using restriction fragment length polymorphism analysis with nonradioactive DNA detection. *Poultry Science* 74: 1067-1073
- Zinkernagel RM and Doherty PC (1974a) Restriction of *in vitro* T cell-mediated cytotoxicity in lymphocytic choriomeningitis within a syngeneic or semiallogeneic system. *Nature* 248: 701-702
- Zinkernagel RM and Doherty PC (1974b) Immunological surveillance against altered self components by sensitised T lymphocytes in lymphocytic choriomeningitis. *Nature* 251: 547-548

CHAPTER TWO

An Evaluation of Methods of Blood Preservation for RT-PCR from Endangered Species

2.1 Introduction

The study of conservation genetics traditionally uses an array of DNA based methods (Lambert and Millar 1995, Sunnucks 2000). However there are a number of instances where the isolation of RNA may be required, particularly in studies of genetic variation and disease resistance, where analysis of expression of major histocompatibility complex (MHC) genes is important (Edwards & Potts 1996). Usually such studies use spleen or liver snap-frozen in liquid nitrogen for the isolation of high-quality RNA and the subsequent construction of cDNA libraries (e.g. Vincek et al. 1995). However this approach may pose particular difficulties in studies of endangered species, as a non-lethal method of sampling will inevitably be required, such as taking blood. In addition, often only small samples can be taken, making the construction of cDNA libraries difficult, as traditional methods of library construction may require as much as 5 µg of mRNA. RNA extraction from blood may also be required for the detection and typing of RNA viruses in wildlife (e.g. Frisk et al. 1999).

In this chapter, an RT-PCR based method was developed to isolate expressed class II B MHC sequences from the Chatham Island black robin (*Petroica traversi*). This species is highly endangered (Butler and Merton 1992), and only small amounts of blood (less than 200 µL per bird) can be sampled. In order to determine the most appropriate method of preserving blood for RNA extraction, a pilot study was first carried out on the common sparrow (*Passer domesticus*) to investigate the quality of RNA extracted from blood preserved in buffers commonly used when sampling blood. Several methods for the preservation of blood for DNA extraction exist (reviewed in Carter 2000). However, these methods have not been

tested for their suitability for RNA extraction. Commercially available preparations for tissue storage such as RNAlater (Ambion) are not suitable for whole blood as they cause serum proteins to coagulate (M. Kracklauer, Ambion, pers comm). In addition, RNA is more sensitive to degradation than DNA, so this can potentially impose additional logistical difficulties when blood sampling in field situations.

2.2 Methods

2.2.1 Blood preservation and RNA extraction

Blood (ca. 40 -100 μ L per bird) was sampled from 20 common sparrows as described in Ardem et al. 1994, and collected in Nunc Cryotube™ vials containing one of the following: (1) lysis buffer (Seutin et al. 1991), 1 mL; (2) absolute ethanol, 1 mL; (3) K₂EDTA, 100 μ L; (4) No buffer. We also collected samples directly into 500 μ L of the RNA extraction reagent Trizol-LS™ (Gibco-BRL), with and without the prior addition of 100 μ L of K₂EDTA. Trizol-LS is a rapid RNA extraction reagent that has been shown to be reliable at extracting RNA from peripheral blood for use in RT-PCR (Chadderton et al. 1997). Because RNA is highly sensitive to degradation, we tested whether samples could withstand a short period at room temperature before freezing. For a single sample collected in K₂EDTA + Trizol-LS this room temperature period was extended to 5 days in order to test the usefulness of Trizol-LS as a medium-term storage option during fieldwork. Details of the storage conditions are given in Table 2.1.

For samples collected in lysis buffer, 250 μ L of the sample was added to 750 μ L of Trizol-LS. Samples in absolute ethanol were spun down and washed once with phosphate-buffered saline. Ten microlitres of washed cells were added to 200 μ L of DEPC H₂O and 750 μ L Trizol-LS. Unbuffered (20 μ L) and K₂EDTA-buffered blood (50 μ L) was diluted up to 250 μ L with DEPC-H₂O, and then added to 750 μ L of Trizol-LS. Samples collected with immediate addition of Trizol-LS were divided into 2 tubes and diluted to 1 mL total volume with extra Trizol-LS. The samples were incubated at room temperature for 15-30 mins, and then any visible clumps of cellular matter were broken up using a needle and syringe. RNA was extracted following the manufacturers instructions, and then resuspended in 25 μ L DEPC-H₂O before being treated with RQ1 RNase-free DNase (Promega). Following

DNase treatment, samples were re-extracted with phenol/chloroform, precipitated with 3 M sodium acetate and 3 volumes of absolute ethanol, then resuspended in 20 μ L DEPC-H₂O with 5U RNAsin (Promega) and 5 mM DTT. The yield of RNA was measured by UV spectrophotometry at 260 and 280 nm, and the product-moment correlation coefficient (r) for the relationship between RNA yield and buffer type or time at room temperature, was measured in Microsoft Excel.

2.2.2 RT-PCR

In order to test whether RNA extracted from sparrows was suitable for RT-PCR, 495 bp of β -actin cDNA was amplified using the primers R- β Actin-ex4 (5'-CTTGCTGATCCACATCTGCTGGAAGG-3') and F- β Actin-ex2 (5'-GAGAGGCTACAGCTTCACCACCAC-3') (P. Ritchie, unpublished data). β -actin was chosen as it is constitutively expressed in most cell types and highly conserved across diverse taxa. cDNA was produced by reverse transcription of 0.5-1.0 μ g of RNA at 42°C for 50 min in a 20 μ L reaction containing 50 mM Tris-HCl pH 8.3, 75 mM KCl, 3 mM MgCl₂, 0.5 μ M Oligo-dT, 10 U RNAsin (Promega), 10 mM DTT, 500 μ M each dNTP, and 200 U Superscript™ II (Gibco-BRL). Two microlitres of cDNA was then amplified in a 25 μ L reaction of 10 mM Tris-HCl pH 8.3, 50 mM KCl, 1.5 mM MgCl₂, 200 μ M each dNTP, 0.4 μ M each primer, and 0.5 U of *Taq* polymerase (Roche). Thermal cycling was performed in a Hybaid OmniGene for 30 cycles at 94°C, 30 sec, 55°C, 30 sec, and 72°C, 45 sec, with a final extension of 72°C for 2 min.

MHC cDNA sequences were amplified using 3'RACE (Rapid Amplification of cDNA Ends, (Frohman et al. 1988)). RNA (0.5-1.0 μ g) was reverse transcribed at 45°C for 50 min in a 50 μ L reaction containing 50 mM Tris-HCl pH 8.3, 75 mM KCl, 3 mM MgCl₂, 0.4 μ M Oligo-dT-AP, 15 U RNAsin (Promega), 10 mM DTT, 0.5 mM each dNTP, and 200 U Superscript™ II (Gibco-BRL). Oligo-dT-AP (5'-GGCCACGCGTCGACTAGTAC(T₁₇)-3') adds an adaptor primer sequence to each cDNA, which is then used as a template for PCR. Class II B MHC cDNA sequences were amplified from 2 μ L of cDNA using the adaptor primer RACE-AP (5'- GGCCACGCGTCGACTAGTAC-3') and an MHC-specific primer MHC05 (5'-CGTRCTGGTGGCACTGGTGGYGCT-3'), designed from aligned passerine class II B MHC sequences to conserved regions of exon 1. PCR was performed in a Hybaid

OmniGene thermal cycler in a 25 μ L reaction using the Expand *HiFi* PCR system (Roche) with 2 mM MgCl₂, 200 μ M each dNTP, and 0.4 μ M each primer. Thermal cycling was initially performed for 5 cycles at 95°C, 30 sec, 63°C, 30 sec, and 72°C, 2 min using only the BRMHC05 primer. The RACE-AP primer was then added for a further 30 cycles at 95°C, 30 sec, 63°C, 10 sec, and 72°C, 1 min 20 sec, with a final extension of 72°C for 3 min.

PCR products of the expected size were excised from an agarose gel and purified using a HighPure PCR product purification kit (Roche), then cloned into a pGEM-T Easy vector (Promega). Positive clones were sequenced using the PRISM BigDye Terminator Cycle Sequencing kit (Applied Biosystems) and analyzed on an ABI 377A automated sequencer.

Table 2.1. Storage conditions of sparrow samples tested in this study. RNA quality was assessed according to whether ribosomal bands were clearly visible on an agarose gel. Samples frozen immediately were placed in a thermos flask containing aluminium “bullets” snap-frozen in liquid nitrogen. For long-term storage prior to RNA extraction, all samples except those stored in lysis buffer or ethanol were stored at -80°C. *Samples were transferred to liquid nitrogen within 3 hours of collection. **The ratio of blood to K₂EDTA ranged from 0.5:1 to 1.4:1.

Buffer	Storage	No. samples	RNA Quality	β -actin amplification
Lysis buffer	Room temp, 4-5 hrs then 4°C, 3-4 weeks	2	X	X
Ethanol	Room temp, 4-5 hrs then 4°C, 3-4 weeks	2	X	✓
No buffer	Frozen immediately*	4	✓	✓
	Room temp, 1 hr*	2	X	✓
K ₂ EDTA**	Frozen immediately*	4	✓	✓
	Room temp, 1 hr*	2	✓	✓
Trizol-LS only	Frozen immediately*	1	X	X
Trizol-LS + K ₂ EDTA	Frozen immediately*	1	✓	✓
	Room temp, 1 hr*	1	✓	✓
	Room temp, 5 days	1	✓	✓

2.3 Results and Discussion

2.3.1 Blood preservation and RNA extraction

RNA quality was visualized by agarose gel electrophoresis (Figure 2.1A) and its suitability for use in RT-PCR was tested by amplification of 495 bp of the β -actin gene from 0.5-1.0 μ g of each sample (Figure 2.1B). The results are summarised in Table 2.1. RNA extracted from blood stored in lysis buffer (lane 1) showed extensive degradation. RNA from samples stored in ethanol was not visible on an agarose gel, suggesting an extremely low yield, however amplification of the β -actin fragment was still possible. High quality RNA could be extracted from blood that was unbuffered, frozen immediately, or blood collected into K₂EDTA. After one hour at room temperature, RNA from unbuffered blood showed some degradation, however no significant degradation was seen in blood stored in K₂EDTA. Blood collected directly into Trizol-LS in the field coagulated instantly, and appeared to prevent cell lysis, thus rendering RNA extraction virtually impossible. However, this could be overcome by collecting the blood into 100 μ L of K₂EDTA and then immediately adding the Trizol-LS. Samples collected in this manner showed no appreciable loss of RNA quality after five days at room temperature. Although the entire blood sample is used for RNA extraction when collected straight into Trizol-LS, good quality DNA could subsequently be extracted from the phenol and interphase of the initial homogenate (data not shown).

Yields of RNA ranged from 54 to 348 ng per μ L of whole blood used in the extraction. Samples collected in lysis buffer or ethanol were not quantified due to the extremely low yield and extensive degradation seen with these methods. Among samples that were quantified, there was only a weak correlation between method of preservation ($r = 0.257$) or time at room temperature ($r = -0.241$), and yield of RNA, suggesting that the efficiency of RNA extraction varies considerably. Absorbance ratios at 260 nm and 280 nm ranged from 1.41 to 1.98. Pure RNA has a 260/280 nm absorbance ratio of 1.8-2.0. Although we had a number of samples with ratios lower than this, the suitability of the RNA for RT-PCR did not appear to be affected as all showed good amplification of the β -actin gene. Samples collected in K₂EDTA with immediate addition of Trizol-LS had higher absorbance ratios (1.7-1.98) than samples collected by other methods.

The results of the pilot study show that two common methods of blood preservation (lysis buffer and ethanol) are not suitable for the extraction of RNA. RNA of acceptable quality can be obtained from whole blood, frozen immediately after collection, or blood collected into K₂EDTA with or without immediate addition of Trizol-LS. K₂EDTA may protect RNA from degradation by nucleases as it chelates free magnesium ions that are required for nucleases to be active (Carter 2000). Thus, it is suitable for storing blood at room temperature for short periods, where immediate freezing is not possible. Immediate addition of the Trizol-LS reagent in the presence of K₂EDTA may allow even greater freedom in field situations, as these samples can be stored for several days at room temperature without affecting RNA quality.

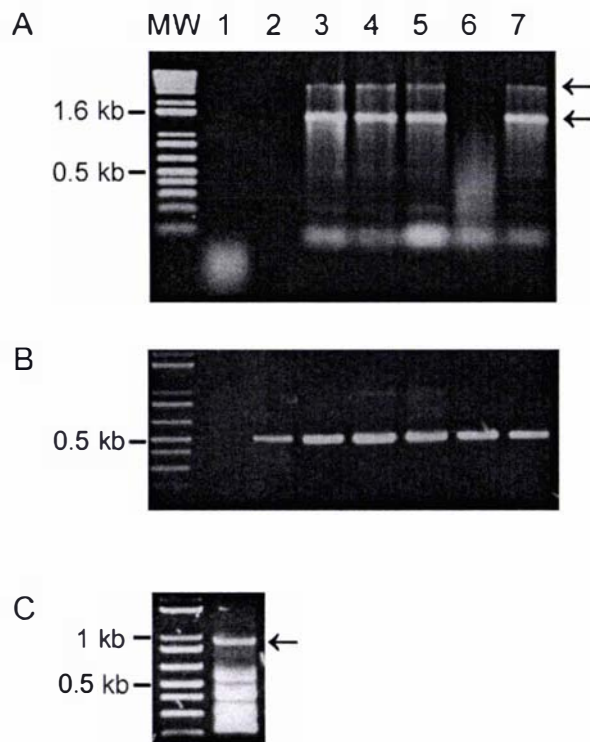


Figure 2.1. (A) Agarose gel electrophoresis of RNA, showing ribosomal bands at approximately 1.5 and 4 kb (←), visualized over UV light after ethidium bromide staining. RNA was extracted from blood stored under the following conditions: lane 1, lysis buffer; lane 2, absolute ethanol; lane 3, no buffer, blood frozen immediately; lane 4, K₂EDTA, frozen immediately; lane 5, K₂EDTA + Trizol, room temp 5 days; lane 6, no buffer, room temp 1 hr; lane 7, K₂EDTA, room temp 1 hr. MW = 1kb+ DNA ladder (GibcoBRL). (B) RT-PCR products (samples as in lanes 1-7, above) amplified using β -Actin oligonucleotide primers. The amplified products correspond to the predicted β -actin cDNA size of 495 bp. (C) 3'RACE product amplified from black robin cDNA using a class II B MHC-specific primer. The 920-bp fragment (←) was gel-purified and found to contain class II B MHC sequences.

2.3.2 Isolation of MHC cDNA sequences

MHC cDNA sequences were isolated from the Chatham Island black robin using RNA extracted from blood (ca 140 μ L) collected into K₂EDTA, with immediate addition of Trizol-LS, as this proved to be the most effective method in the pilot study. A 920-bp class II B MHC fragment was isolated using 3'RACE (Figure 2.1C). RACE is a modification of RT-PCR that is ideal for isolating rare transcripts or in instances where small amounts of starting material make cDNA library construction difficult (Frohman et al. 1988). When 5' and 3' RACE are used together, full-length transcripts can be obtained. Despite considerable non-specific amplification, a single clear band at the correct size was present and could be purified without the need for a second, nested PCR. Non-specific amplification is common when only a single specific primer and a non-specific primer are used (Ohara et al. 1989). Three different class II B MHC sequences were isolated after sequencing 14 positive clones. These sequences spanned part of exon 1, all of exons 2-5, and the entire 3' untranslated region and have been deposited in Genbank (Accession numbers AY258333-AY258335).

Functional markers, such as MHC genes, are becoming increasingly relevant in conservation genetics (Crandall et al. 2000, Hedrick 2001), yet studies of gene expression in endangered species are often avoided due to a reluctance to sacrifice an individual animal. Here I have shown that for studies of MHC gene expression whole blood, collected and stored appropriately, is a suitable source of mRNA. Nearly full-length class II B MHC sequences were successfully isolated from blood samples from the Chatham Island black robin by using 3'RACE as an alternative to constructing a cDNA library. This is a useful non-lethal protocol for isolation of MHC genes from endangered species when only small amounts of sample are available.

2.4 References

- Ardern SL, McLean IG, Anderson S, Maloney R, Lambert DM (1994) The effects of blood sampling on the behaviour and survival of the endangered Chatham Island Black Robin (*Petroica traversi*). *Cons. Biol.* 8: 857-862
- Butler D, and Merton, D (1992) *The Black Robin: Saving the World's Most Endangered Bird*. Oxford University Press
- Carter RE (2000) General molecular biology. In: *Molecular Methods in Ecology* (ed. Baker, AJ). Blackwell Science

- Chadderton T, Wilson C, Bewick M, and Gluck S (1997) Evaluation of three rapid RNA extraction reagents: Relevance for use in RT-PCR's and measurement of low level gene expression in clinical samples. *Cell. Mol. Biol.* 43: 1227-1234
- Crandall K, Bininda-Emonds ORP, Mace GM, Wayne RK (2000) Considering evolutionary processes in conservation biology. *Trends Ecol. Evol.* 15: 290-295
- Edwards SV and Potts WK (1996) Polymorphism of genes in the major histocompatibility complex (MHC): Implications for conservation genetics of vertebrates. In: *Molecular Genetic Approaches in Conservation* (eds Smith TB, and Wayne RK), pp. 214-237. Oxford University Press
- Frisk AL, Konig M, Moritz A, Baumgartner W (1999) Detection of canine distemper virus nucleoprotein RNA by reverse transcription-PCR using serum, whole blood, and cerebrospinal fluid from dogs with distemper. *J. Clin. Microbiol.* 37: 3634-3643
- Frohman MA, Dush MK, and Martin GR (1988) Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci. USA* 85: 8998-9002
- Hedrick, PW (2001) Conservation genetics: where are we now? *Trends Ecol. Evol* 16: 629-636
- Lambert DM, and Millar CD (1995) DNA Science and Conservation. *Pacific Conservation Biology* 2: 21-38
- Ohara O, Dorit RL, and Gilbert W (1989) One-sided polymerase chain reaction: the amplification of cDNA. *Proc. Natl. Acad. Sci. USA* 86: 5673-5677
- Seutin G, White B, and Boag PT (1991) Preservation of avian blood and tissue samples for DNA analyses. *Can J. Zool.* 69: 82-90
- Sunnucks P (2000) Efficient genetic markers for population biology. *Trends Ecol. Evol.* 15: 199-203
- Vincek V, Klein D, Graser RT, Figueroa F, O'hUigin C, Klein J (1995) Molecular cloning of major histocompatibility complex class II B gene cDNA from the Bengalese finch *Lonchura striata*. *Immunogenetics* 42: 262-267

CHAPTER THREE

Class II B MHC Genes in New Zealand Robins: Evolution by Gene Duplication, Gene Conversion and Balancing Selection.

3.1 Introduction

The major histocompatibility complex (MHC) is a multigene family that plays an important role in pathogen resistance. Classical class I and class II MHC genes are highly polymorphic, and encode cell-surface proteins that present short peptides, usually of bacterial or viral origin, to T cells, initiating a specific immune response (Klein 1986). The MHC region has been completely sequenced in human (MHC Sequencing Consortium 1999) and chicken (Kaufman et al. 1999b). The chicken MHC (or B locus) is smaller and simpler than the mammalian MHC. It has been described as “minimal essential” (Kaufman et al. 1995; Kaufman et al. 1999a), as it contains fewer genes, few repetitive elements, and no pseudogenes, and the intron and intergenic sizes are much smaller than for mammalian MHC genes. Class I and class II genes have also been found outside the MHC region in chicken, in a separate locus designated *Rfp-Y* by Briles et al. (1993). However, these genes appear to be non-classical, as they show little polymorphism and low levels of expression in haematopoietic tissues (Zoorob et al. 1993; Kaufman et al. 1999a), and may have specialised functions (Afanassieff et al. 2001).

Recent studies of other avian taxa suggest that the chicken MHC may not be a good model for birds in general. For example, the MHC of another galliform, the quail (*Coturnix japonica*), does not appear to be “minimal essential”. Although it has the same basic organisation as the chicken MHC, it contains multiple copies of many loci (Kulski et al. 2002). The passerine MHC also appears to be larger and more complex than that of the chicken. In most passerine species, Southern blot analysis using a variety of cloned class II B

probes spanning exons 1-4 has revealed a more complex pattern of restriction fragments than in chicken (Wittzell et al. 1999a; Edwards et al. 2000b; Freeman-Gallant et al. 2002). Analysis of exon 2 sequences from Darwin's finches indicates there may be six class II B MHC loci in these species (Sato et al. 2000), and seven cDNA sequences from class II B genes have been isolated from the great reed warbler (*Acrocephalus arundinaceus*) (Westerdahl et al. 2000). There also appears to be four transcribed class I loci in the great reed warbler (Westerdahl et al. 1999). In addition, passerine MHC genes appear to be larger than those of chicken, as all three full-length class II B MHC sequences isolated from a red-winged blackbird (*Agelaius phoeniceus*) cosmid library have much larger introns than chicken genes (Edwards et al. 1998; Edwards et al. 2000a; Gasper et al. 2001). Pseudogenes have also been identified in red-winged blackbird (Edwards et al. 2000a), and house finch (*Carpodacus mexicanus*) (Hess et al. 2000).

Key features of evolution of MHC genes include gene duplication and decay, gene conversion, and balancing selection, however the relative importance of each of these mechanisms has been much debated. MHC genes are the most polymorphic genes found in vertebrates, and there is evidence that MHC diversity is maintained by balancing selection (reviewed in Meyer and Thomson 2001). Balancing selection is thought to be largely pathogen-driven, where increased MHC diversity confers resistance to a greater range of pathogens. However in at least some species MHC diversity may also be driven by reproductive mechanisms such as disassortative mating based on MHC genotype (reviewed in Apanius et al. 1997).

Gene conversion, a form of non-reciprocal recombination, has been identified as an important mechanism by which variation in the MHC can be generated (Martinson et al. 1999). Evidence for gene conversion has been documented in both class I and class II genes across a range of species, from mammals (Miyada et al. 1985, Andersson et al. 1991, She et al. 1991), to birds (Wittzell et al. 1999b) and fish (Langefors et al. 2001). However, gene conversion can only create new alleles when it operates over short stretches of sequence, and when there are already a number of polymorphic alleles (Satta 1997). When gene conversion occurs over long tracts it can homogenise sequences within species, leading to concerted evolution in which duplicated genes evolve together. Concerted evolution is thought to be important in avian MHC evolution, as exon 2 and 3 sequences tend to cluster within species under phylogenetic analysis (Edwards et al. 1995b; Edwards et al. 1999). Evidence for concerted

evolution has been detected in the pheasant MHC (Wittzell et al. 1999b). Two loci orthologous to the chicken *BLB-I* and *BLB-II* loci have been isolated from pheasant, however these orthologous relationships are not retained in exons 2 and 3. Instead, these sequences group together within the species, suggesting that gene conversion has occurred between the coding regions of the two pheasant genes.

However in mammals, there appears to be little evidence for concerted evolution, rather loci are maintained independently of each other (Nei et al. 1997; Gu and Nei 1999). Nei et al. (1997), proposed the birth-and-death model as the primary mode of evolution of mammalian MHC genes. Under this model new genes arise from gene duplication, some are maintained for long periods, and others are deleted or become pseudogenes. However, it has been pointed out that the birth-and-death models and concerted evolution models are not mutually exclusive, but may occur over different timescales (Edwards et al. 1999; Wittzell et al. 1999b). The finding of pseudogenes in passerines suggests that gene duplication and decay may be a universal feature of MHC evolution (Hess et al. 2000; Edwards et al. 2000a). However evolution of mammalian genes appears to be characterised by ancient duplications predating speciation, followed by divergent evolution, whereas in birds there is a higher rate of concerted evolution and/or more recent duplications. This means that in mammals, distantly related species such as humans and mice share orthologous loci (Trowsdale 1995), whereas the only report of orthologous loci in birds is in the closely related species chicken and pheasant (Wittzell et al. 1999b).

In this chapter, the evolution of class II B MHC genes in two species of New Zealand robin – the Chatham Islands black robin (*Petroica traversi*) and the South Island robin (*Petroica australis australis*) - is investigated by isolation of transcribed class II B sequences and Restriction Fragment Length Polymorphism (RFLP) analysis. The orthology of sequences isolated from these two closely related species is investigated, allowing the evolutionary processes influencing avian MHC genes over short timescales (i.e. within genera) to be analysed.

New Zealand robins are Australo-Papuan songbirds (family Petroicidae) and are not closely related to the North American or European robin (family Turdidae), or to the Eurasian robins (Old-World flycatchers, Muscicapidae). The Chatham Islands black robin is found on South East (Rangatira) and Mangere Islands, in the Chatham Islands group, 850 km east of New

Zealand. In 1980 the entire black robin population consisted of only 5 birds, after at least a century of population decline (Butler and Merton 1992). The entire species is now descended from a single breeding pair, and is one of the most inbred species ever recorded. Levels of neutral minisatellite DNA variation are among the lowest measured for any wild bird (Ardern and Lambert 1997). However, intensive management since 1980 has seen the species increase in number to around 250 birds. The South Island robin is a close relative of the black robin, and is found throughout the South Island of New Zealand. New populations of South Island robins were established on two off shore islands in 1973 from small numbers of founders (Flack 1974), providing an ideal comparison to the Chatham Island black robin. A central aim of this thesis is to measure levels of MHC variation in the black robin population and to compare it with South Island robin populations. Because of the heterogeneity of the passerine MHC, isolation of transcribed class II MHC genes from these species is an essential precursor to analysing variation, and will be used as a basis for designing a genotyping method for class II MHC alleles in New Zealand robins (see chapter five).

3.2 Materials and Methods

3.2.1 *Birds and sample collection*

Black robins were sampled from South East (Rangatira) Island, as described in Ardern et al. (1994). Black robin blood samples used for RFLP analysis were collected in March-April 1992, and subsequent sampling to collect fresh blood for RNA extraction was undertaken in January 2001. All blood samples from South Island robins were collected during 1992 and 1993, from four populations: Motuara Is, Allports Is, Nukuwaiata Is and Kaikoura (see chapter five and Ardern et al. 1997a). Motuara Is and Allports Is are “bottlenecked” populations founded by a small number of individuals translocated from the Nukuwaiata Is and Kaikoura populations, respectively. A random sample of eight individuals from each population were analysed for RFLP variation, and samples from five South Island robin families from Motuara Island were analysed for inheritance of MHC fragments.

3.2.2 *Isolation of genomic DNA*

Genomic DNA was extracted from 5-10 μ L of whole blood. Blood cells were lysed as

described in Ardern and Lambert (1997). DNA was then extracted and purified using standard phenol/chloroform methods and precipitation and resuspension of DNA was performed in accordance with Sambrook et al. (1989)

3.2.3 Probes

Two class II B MHC probes were isolated for RFLP analysis: one spanning 209 bp of exon 2 (BR6) and the other spanning 260 bp of exon 3 (BRex3C). BR6 was isolated by PCR amplification of black robin genomic DNA using degenerate primers 305 and 306 from Edwards et al. (1995a)¹. BRex3C was isolated by PCR from black robin cDNA using primers MHC06 and MHC07 (see Figure 3.1 and Table 3.1 for primer details). Probes were amplified in 25 μ L volumes which included 10 mM Tris-HCl, 50 mM KCl pH 8.3, 1.5 mM MgCl₂, 200 μ M each dNTP, 0.4 μ M each primer, 0.5 units of *Taq* polymerase (Roche) and 1 μ L of DNA. Thermal cycling was performed on a Hybaid OmniGene for 30 cycles consisting of denaturing at 94°C, annealing at 50°C (BR6) or 64°C (BRex3C), and extension at 72°C, each for 30 sec. PCR products were purified using High Pure PCR Purification Kit (Roche), cloned into a pks+ vector (Stratagene) (BR6), or a pGEM[®]-T Easy vector (Promega) (BRex3C), then sequenced using the BigDye Terminator Cycle Sequencing kit (Applied Biosystems) and analysed on an ABI 377A automated sequencer. Homology of probe sequences to avian MHC sequences was confirmed using BLAST (NCBI), and for each probe one positive clone was randomly selected for use in RFLP analysis.

3.2.4 Southern blot hybridisation

DNA samples of 10-15 μ g were digested overnight with 10 units each of *Pvu*II or *Taq*I enzymes in the presence of spermidine trihydrochloride and 2 mg/mL bovine serum albumin (BSA) with the manufacturer's recommended buffer. The following morning a further 10 units of enzyme were added and incubation continued for a minimum of 1 hour. The concentration of the digested DNA was determined with a Hoefer TKO-100 DNA fluorometer. For *Taq*I digests the DNA was diluted following fluorometry with the enzyme buffer mix and incubated overnight with a further 10 units of enzyme to ensure complete digestion.

¹ The BR6 probe was constructed by Dr Rob Slade, Australian Genome Research Facility, Univ. of Queensland.

The digested DNA was then electrophoresed through 0.8% agarose in TBE buffer (134 mM Tris, 74.9 mM boric acid, 2.55 mM EDTA pH 8.8) for 27 hrs at 55 V. After electrophoresis, the DNA was transferred to a nylon membrane (Boehringer Mannheim) by Southern blotting overnight as described in Ardern et al. (1997a). Membranes were hybridised with probes labelled with [$\alpha^{32}\text{P}$]dCTP (GibcoBRL RTS RadPrime DNA Labelling System OR Amersham Megaprime kit) as described in Miller et al. (2003) (see Appendix EII). Hybridisation temperatures used were 60°C (BR6) and 65°C (BRex3C). Membranes were washed twice with 5 x SSC, 0.1% SDS for 30 min at the hybridisation temperature, then exposed to Fuji Medical X-ray film (RX) at -80°C for up to 3 weeks. All bands on the autoradiographs were scored using a bin width of 1.5 mm, and only samples run on the same gel were compared.

3.2.5 RNA isolation

RNA was isolated from blood in this study, as it was not possible to obtain a fresh spleen or liver from any New Zealand robin species due to their protected status. Black robin blood from a single individual (140 μL) was collected into K_2EDTA and Trizol-LS for the purpose of RNA isolation, as this was previously determined to be the best method of blood preservation for RNA extraction (see chapter two). South Island robin blood collected in this manner was not available, so RNA was isolated from 50 μL of whole blood, from a single individual from Kaikoura, which had been stored at -80°C for several years. RNA was extracted using Trizol-LS reagent according to manufacturers instructions, then DNase treated and resuspended as described in chapter two.

3.2.6 Isolation of MHC cDNAs

RNA was reverse-transcribed and class II B MHC cDNA sequences were amplified using a combination of 3'RACE (Rapid Amplification of cDNA Ends, Frohman et al. 1998) and standard RT-PCR. Details of the cDNA synthesis and 3'RACE methods are given in chapter two. The positions of the PCR primers used are shown in Figure 3.1 and their sequences are detailed in Table 3.1. The primers MHC05 and MHC06 were designed from conserved regions in other passerine MHC sequences obtained from Genbank, and the primers MHC07, MHCvar3', and MHCpetr3' were designed from black robin and South Island robin

sequences obtained using the MHC06 + RACE-AP primer pair. All PCR amplifications were carried out in 25 μ L volumes containing 2 μ L of cDNA and using the Expand *HiFi* PCR system (Roche) with 2 mM MgCl₂, 200 μ M each dNTP, and 0.4 μ M each primer. Thermal cycling was performed using a Hybaid Omni-gene Thermal Cycler for 3'RACE reactions or a BioRad iCycler for all other PCRs, for 30 cycles consisting of 95°C for 30 sec, annealing temperature (see Table 3.1) for 10 sec, and 72 °C for 45-80 sec.

All PCR products were gel-purified using a HighPure PCR product purification kit (Roche), and cloned into pGEM[®]-T Easy vector (Promega). For each PCR product 12-20 clones were sequenced in both directions using the BigDye Terminator Cycle Sequencing kit (Applied Biosystems) and analysed on an ABI 377A automated sequencer. Details of the number of PCR amplifications and clones sequenced using each pair of primers is given in Table 3.1.

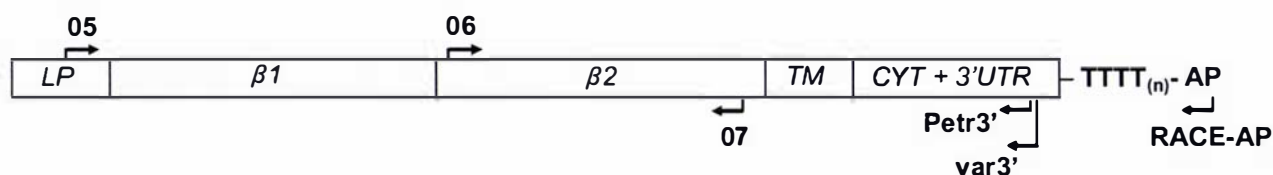


Figure 3.1 Position of PCR primers used in isolation of class II B MHC cDNA sequences from black robin and South Island robin (see Table 3.1). The poly-T primer containing an adaptor sequence (AP) used to prime cDNA synthesis is shown. Each domain is coded for by a separate exon, and named according to Edwards et al. (1995b): LP, leader peptide (exon 1); $\beta 1$, peptide-binding region (exon 2); $\beta 2$, immunoglobulin-like domain (exon 3); TM, transmembrane domain (exon 4); CYT, cytoplasmic tail (exon 5); 3'UTR, 3' untranslated region.

Table 3.1 Sequences and annealing temperatures of primers used for isolation of class II B MHC cDNA sequences from black robin (BR) and South Island robin (SR). The number of PCRs performed using each primer pair and the total number of clones sequenced (and found to contain MHC sequences) is also given.

Primer pair	Primer sequence	Annealing Temp	# PCRs		Total # clones	
			BR	SR	BR	SR
MHC05 + MHC07	5'-CGTRCTGGTGGCACTGGTGGYGCT-3' 5'-TGCTCCAGGCTCACGTGCTCCAC-3'	63°C	3	3	36	33
MHC06 + RACE-AP	5'-AGTGCCTCCCAGCGTGTCCATCTC-3' 5'-GGCCACGCGTCTGACTAGTAC-3'	64°C	2	3	27	27
MHC05 + MHCvar3'	5'-CGTRCTGGTGGCACTGGTGGYGCT-3' 5'-GGAACAACATGGGGACAAAGCGA-3'	58°C	2	2	21	23
MHC05 + MHC Petr3'	5'-CGTRCTGGTGGCACTGGTGGYGCT-3' 5'-GGGAACAACAGGGAACAACCAGGGT-3'	63°C	1	2	12	25
MHC05 + RACE-AP	5'-CGTRCTGGTGGCACTGGTGGYGCT-3' 5'-GGCCACGCGTCTGACTAGTAC-3'	63°C	2	-	21	-

In order to identify and eliminate clones containing PCR errors, each PCR and subsequent cloning and sequencing was repeated up to three times, and only sequences that could be verified in independent PCRs (either using the same, or different but overlapping, primer pairs) were included in the analysis.

3.2.7 Data analysis

Contigs were built from overlapping fragments using Sequencher™ 4.1.2 (Gene Codes Corporation), and checked by eye. Full-length sequences from these contigs were aligned with ClustalW (Thompson et al. 1994). Phylogenetic analysis and calculation of sequence identity was performed in PAUP*4.0b10 (Swofford 2002). Calculation of the number of non-synonymous (d_N) and synonymous (d_S) substitutions per site was performed using *MEGA* version 2.1 (Kumar et al. 2001) using the Nei and Gojobori (1986) method with Jukes-Cantor (1969) correction. A Z-test was performed in *MEGA* to test for positive selection, i.e., whether d_N is significantly greater than d_S .

The occurrence of gene conversion was assessed using the program GENECONV version 1.81 (Sawyer 1999). GENECONV analyses the distribution of nucleotide substitutions to detect gene conversion events by looking for stretches of nucleotides in a pair of sequences that are more similar to each other than would be expected by chance (Drouin et al. 1999). GENECONV contains an option that allows only synonymous sites of coding sequences to be included in the analysis, thus avoiding the possible effects of selection. However, it was not possible to include the 3'UTR in the analysis using this option, as it is non-coding sequence. To overcome this problem, a separate data file containing only non-amino acid changing sites across the full-length cDNA transcript (i.e. all sites in the 3'UTR plus synonymous sites in the coding region) was created. Putative gene conversion events were considered significant when the simulated global p -value, based on 10,000 permutations of the original data, was less than 0.05. The analysis was performed on alignments consisting of cDNA sequences from both species. The within-group option was also used to detect gene conversions specifically within sequences from the same species. Gscale values of 0, 1, and 2 were used, allowing for varying levels of mismatches (i.e. subsequent mutation) within the gene conversion event to be taken into account. A gscale value of 0 means no mismatches are allowed within the converted region, whereas values of 2 and 1 allow for progressively more mismatches.

3.3 Results

3.3.1 Restriction fragment length polymorphisms

MHC RFLP patterns were analysed in eight black robins and 32 South Island robins (eight from each population) in order to gain an overview of the complexity and polymorphism of the class II MHC region. An example of an RFLP profile for black robin and South Island robin is given in Figure 3.2. The BR6 (exon 2) probe gave the clearest profiles with both *PvuII* and *TaqI* restriction enzymes, with a total of six hybridising fragments in black robin, and 6-15 hybridising fragments of varying intensity in South Island robin. A higher number of fragments were seen with the exon 3 (BRex3C) probe, due to a higher number of weakly hybridising bands. Using *PvuII*, six hybridising bands were detected in black robin, and 8-12 bands in South Island robin, and with *TaqI* there were eight bands in black robin profiles, and 9-17 bands in SI robin. Identical RFLP patterns were observed across eight black robin samples for both *PvuII* and *TaqI* enzymes for both probes. Higher levels of variation were observed in South Island robin populations, as every individual from the Kaikoura and Nukuwaiata Island populations had a unique RFLP genotype for both enzymes. Seven unique genotypes were found in eight individuals from Motuara Is, and among eight birds from Allports Island four (*TaqI*) and five (*PvuII*) unique genotypes were recorded.

3.3.2 Inheritance/linkage of MHC fragments

In order to assess whether any MHC fragments segregate independently, the pattern of inheritance of bands was analysed in profiles of five South Island robin families hybridised with the BR6 probe. Each family contained two chicks, and paternity had previously been confirmed using minisatellite DNA analysis (Ardern et al. 1997b). For each family there were between 3 and 8 variable bands (found in only one of the parents) and no evidence of unlinked bands segregating independently was found. That is, profiles for each parent contained two haplotypes containing linked bands within each haplotype, and each chick consistently inherited one RFLP haplotype from its mother and one from its father. This suggests that class II MHC genes are contained in a single linkage group in robins.

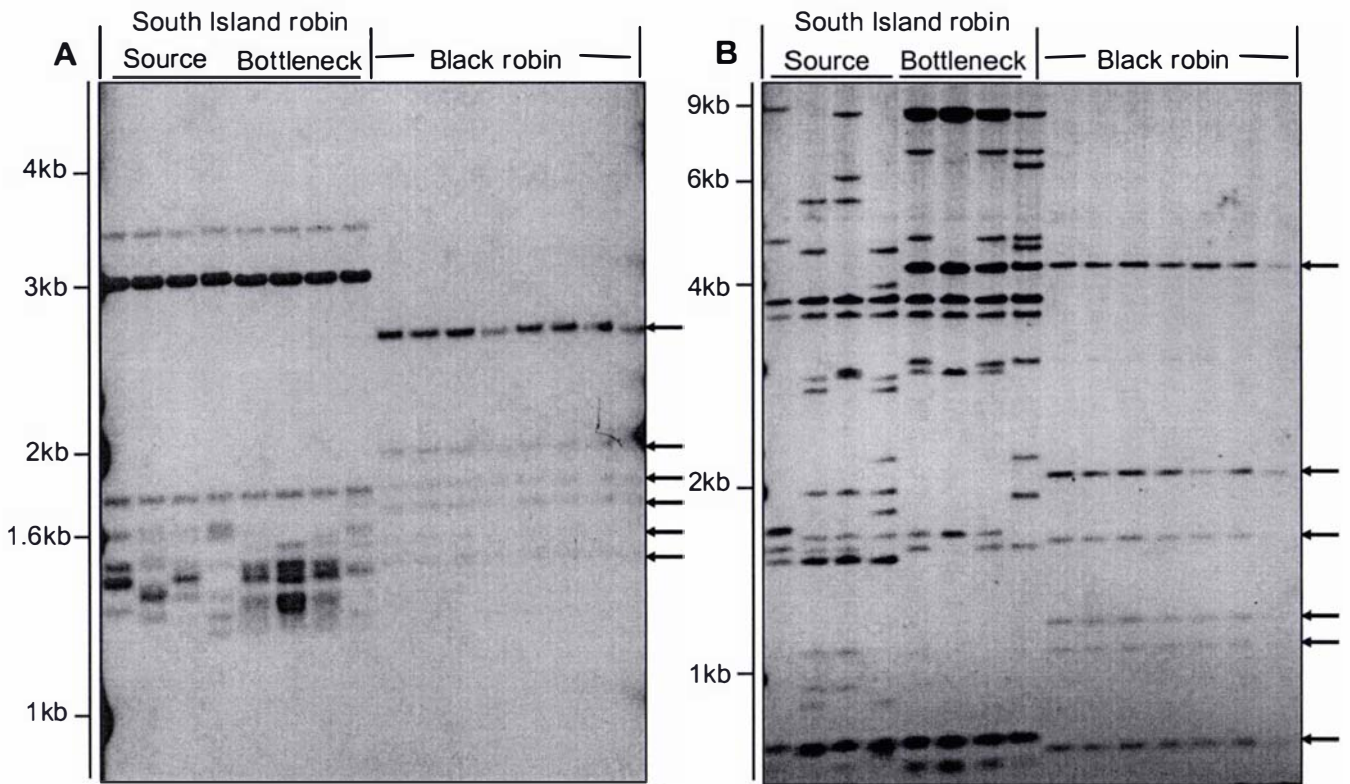


Figure 3.2 RFLP profiles of class II B MHC genes in New Zealand robins using the probe BR6 and the restriction enzymes (A) *PvuII*, and (B) *TaqI*. The source and bottlenecked South Island robin individuals are from Kaikoura and Allports Is respectively. Arrows indicate monomorphic bands in black robin.

3.3.3 Isolation of MHC cDNA sequences

We used an RT-PCR based approach to isolate class II B MHC cDNA sequences for this study, as we could not obtain a large enough tissue sample to construct a representative cDNA library due to the requirement for non-lethal sampling. Initially, the 3' end from exon 3 through to the 3'UTR was amplified with the primer pair MHC06 + RACE-AP. Sequences obtained from this primer pair were used to design the MHC07, MHPetr3', and MHCvar3' primers, which in combination with MHC05 primer, produced overlapping fragments covering almost the full length of the gene, excluding the 5'UTR and the 5' end of exon 1 (Figure 3.1). The MHC05 + RACE-AP primer pair was also used to amplify almost full-length sequences from black robin (see chapter two), however sequences from these primers could not be successfully amplified from South Island robin due to degradation of RNA. Where sequences amplified with the MHC05 + MHPetr3' and MHC05 + MHCvar3' primer pairs were identical in the region of overlap to those obtained using MHC05 + MHC07 or MHC06 + RACE-AP primers, they were assumed to represent transcripts from the same

allele. On this basis, the three overlapping fragments were assembled into contigs. These contigs appear to represent four different class II B MHC cDNA sequences in black robin and eight different sequences in South Island robin (Figure 3.3).

The class II B cDNA sequences we have isolated from assembled contigs have been named Petr01-04 (black robin) and Peau01-08 (South Island robin), as it is not possible to unambiguously distinguish loci. These sequences span part of exon 1, all of exons 2, 3, 4, and 5, and the 3'UTR, with an open reading frame of 222 amino acids from the start of exon 2 to the 3'UTR. The Peau06 sequence is missing the 3' end as it was only amplified using the MHC05 + MHC07 primer pair, and overlapping fragments using additional primers could not be obtained. Two different forms of the Petr04 sequence were isolated. The majority of clones containing this sequence had a 53 bp deletion at the end of exon 2 (Petr04D, Fig 3.3), however a few near-full length Petr04 clones without the deletion were also found in this individual. The truncated sequence appears to be the result of an mRNA splicing error and is unlikely to produce a functional molecule. For all subsequent analyses the Petr04 sequence without the deletion was used. All of the near-full length sequences isolated contain residues expected in functional class II genes according to Kaufman et al. (1994), such as those involved in glycosylation, disulphide bond formation, salt bridges and peptide binding (Figure 3.4).

The sequence of the BR6 probe used for RFLP analysis differed from the cDNA sequences we isolated from the black robin. In fact, this sequence appears to be divergent from the cDNA sequences as it shares 76.1-78.3% identity with the black robin cDNA sequences, while sequence identity among these cDNA sequences across this region is 85.2-90.9%. The BR6 sequence was isolated from a different individual from that which the cDNA sequences were isolated, however that is unlikely to be the cause of the divergence as sequence identity between the black robin and South Island robin cDNA sequences is also higher (83 – 91.7%) than identity between BR6 and the black robin cDNA sequences. The BR6 sequence is further investigated in chapter four. The BRex3C probe was identical to Petr01, 02 and 03.

3.3.4 Orthology, balancing selection and gene conversion

The sequence differences between the transcripts isolated in this study are mostly confined to exon 2 and the 3'UTR (see Figure 3.3). Other than a few fixed differences between black

robin and South Island robin sequences, exons 3, 4 and 5 are highly conserved. The 3'UTR sequences can be grouped according to length and sequence, and form two distinct clusters on a neighbour joining tree (Figure 3.5A). One cluster consists of the black robin sequence Petr04, and the South Island robin sequences Peau01 and Peau05, indicating that these sequences represent orthologous loci. Peau01 and Peau05 may represent two alleles of the same gene. The remaining 3'UTR sequences cluster together with strong bootstrap support, although Petr03 appears somewhat divergent. The 3'UTRs of Peau02 and Peau08 are 100 and 85 bp longer respectively than the other 3'UTR sequences in this group, but share significant sequence similarity in the region of overlap. No such extended 3'UTR regions were isolated from black robin, but this may be due to the preferential amplification of shorter fragments in 3'RACE, and the presence of such sequences in black robin cannot be ruled out.

The orthologous relationships identified at the 3'end of the gene are lost when coding sequences are compared. Exons 3, 4 and 5 appear to be homogenised within each species as five fixed nucleotide changes across this region distinguish the black robin from the SI robin sequences (see Figure 3.3). Phylogenetic analysis of exon 2 produces a "star-like" tree where few branches are well supported by bootstrap analysis (Figure 3.5B). The sequences that share 3'UTR similarity do not cluster together on the exon 2 tree. In the absence of recombination or interlocus gene conversion, phylogenetic relationships should be retained across the length of the gene. Thus, the loss of phylogenetic signal in exon 2, and the breakdown of orthologous relationships, is indicative of gene conversion across the length of the sequence, and may also indicate balancing selection within exon 2 (Edwards et al. 1998).

The rate of nonsynonymous (d_N) to synonymous (d_S) substitutions is commonly used as one indicator of balancing selection (Hughes and Nei 1989). This analysis is usually performed on alleles at the same locus, however in this instance we have measured the divergence of sequences within putative groups of orthologous loci (based on 3'UTR sequences), which are expected to be closely related. Group 1 consists of sequences Petr04, Peau01 and Peau05, and group 2 includes Petr01, Petr02, Peau03 Peau04 and Peau07. In exon 2, d_N was significantly higher than d_S for both groups (Table 3.2). In exons 3, 4, and 5, the ratio of d_N/d_S was slightly greater than one, but this was not significant. Within exon 2, d_N and d_S were also compared for the 24 sites that are known to contact the peptide in the human DRB molecule (Brown et al. 1993, see Figure 3.4). Of these 24 sites, 79% contain nonsynonymous changes, compared

with 37% of non peptide-binding (PB) codons. The ratio of d_N/d_S is higher for PB codons than for non-PB codons for both groups of sequences, however for group 2 sequences the probability of $d_N > d_S$ for PB codons is only on the margin of the conventional significance value.

GENECONV was used to specifically look for gene conversion events, after removal of sites that may be influenced by balancing selection (i.e., non-synonymous sites). A total of 13 significant gene conversion events ranging from 55 to 788 bp in length were detected using GENECONV (Table 3.3). Two of these were only significant when the analysis was confined to within-species groups. Initially, the *gscale* parameter was set to one. This allows for mismatches within putative gene conversion events, and thus allows for the detection of older events. When the *gscale* parameter was set to zero (no mismatches) or two (some mismatches), an additional gene conversion event was identified, and the length of the converted region changed in some cases. Peau06 was not included in the analysis as this sequence lacks exons 4, 5, and the 3'UTR. Most of the putative gene conversion events identified involve Peau01, Peau05 or Petr04, and the position of the events identified using GENECONV can account for the loss of orthologous relationships between 3'UTR and the coding region.

Exon 2

1 21 41 61 81

Petr01 -----TG----TTC----G-----AC-

Petr02 -----TT----T---TT- A-----G-----

Petr03 -----A-----

Petr04D -----TT----TT-----G-----AC-

Peau01 -----C---T-----C---CT

Peau05 -----TG---G-----

Peau02 -----T--ATCC---T-----

Peau03 -----AG-----T-----

Peau04 -----TG---G-----C-----

Peau06 -----A---G-----

Peau07 -----TG-----C-----

Peau08 -----C-----

consensus GGAGCCCCCGCTGCGGG CGTGGGGCTCTCGGGGTCT TCCAGGAGATGGGAAAGGCC GAGTGTCACTTCATTAACGG CACGGAGAAGGTGAGTTGG

101 121 141 161 181

Petr01 ---T---A-----A---T-----ACA-----T-

Petr02 ---C---C---C-----C---T-----T-----

Petr03 ---C---A-C---A---T-----AAG- T-----ACA---A---

Petr04D ---C---AC---C-----A-C---T-----AAG- T-----AAA---T---

Peau01 ---C---C-----A-CC---T-----A-----GG-----T---A-T-

Peau05 ---C---A-----A---T-----C-----T-----T-----

Peau02 ---T---A-----A-----G-----C-----C---T---

Peau03 ---C---C-----C-----C-----T-----

Peau04 ---T---C-----A-C-----A---T-----GG---ACA-----T-

Peau06 ---A---T-----A-----C-----C-----T-----T---

Peau07 ---T---A-----A-C-----GG-----ACA-----T-

Peau08 ---T-----A-C-----A-----ACA-----T-

consensus TGGAGAGGTACTTCTACAAC CGGGAGGAGTTCGTGAGGTA CGACAGCGATGTGGGGCGCT ATGTGGGGCTCACCCCTAT GGGGAGAAGGWGGCCGGAA

201 221 241 261 281

Petr01 -----A-A-----C T-A-----AG-----G-----A---TC---

Petr02 G-----C T-A-----AG-----G-----A-----CGA-----

Petr03 -----G A-----A-----G-TA-----A---T---

Petr04D -----G A-----A-----A-C -C-----***** *****

Peau01 -----A-----T-----A-----G-TA-----TGA-G-----

Peau05 ---T-----G-----TA-CTT--AAT -C-----T-----G-----TC-----CT---

Peau02 ---T-----A-----G A-A-----A-----TA-----GG-----C-TGC---CT---T---

Peau03 -----A-----G-----TA-CTT--AAT -C-----T-----A-TG-----T---

Peau04 -----G-----G-----A-----G-A-----T---

Peau06 -----G A-A-----A-----G-----T-----T---

Peau07 -----T---A-----T-----T-----G-----A-----G-A-----CT---T---

Peau08 -----G-----A-A-----G-----A-----TC-----CT---T---

consensus CTGGAACAGCGACCCGGCAA RACTGGAGCGCAGACGGGCA GAGGTGGACACGKTCTGCCG GCACAACACTACGAGGTGKTCA CCCCCTTACCGTGGAGCGC

Exon 3

301 321 341 361 381

Petr01 -----

Petr02 -----

Petr03 -----

Petr04D ***-----A-----

Peau01 -----A-----G-----

Peau05 -----G-----

Peau02 -----

Peau03 -----G-----

Peau04 -----

Peau06 -----G-----

Peau07 -----G-----

Peau08 -----G-----

consensus CGAGTGCCTCCAGCGTGTG CATCTCGCTGGTGCCTCGA GCTCCAGCCCGGTCCCGC CGCTGTCTGCTCCGTGCT GGATTTCTACCCCTCCGCCA

401 421 441 461 481

Petr01 -----G-----

Petr02 -----G-----

Petr03 -----G-----

Petr04D -----G-----G-----

Peau01 -----

Peau05 -----

Peau02 -----

Peau03 -----

Peau04 -----

Peau06 -----

Peau07 -----

Peau08 -----

consensus TCCAGTGAGGTGGTTCCAG GGCCAGCAGGACTCTCGGC CCACGTGGTGGCCACCGACA TTGTCCCCAACGGGACTGG ACCTACCAGCTGCTGGTGT

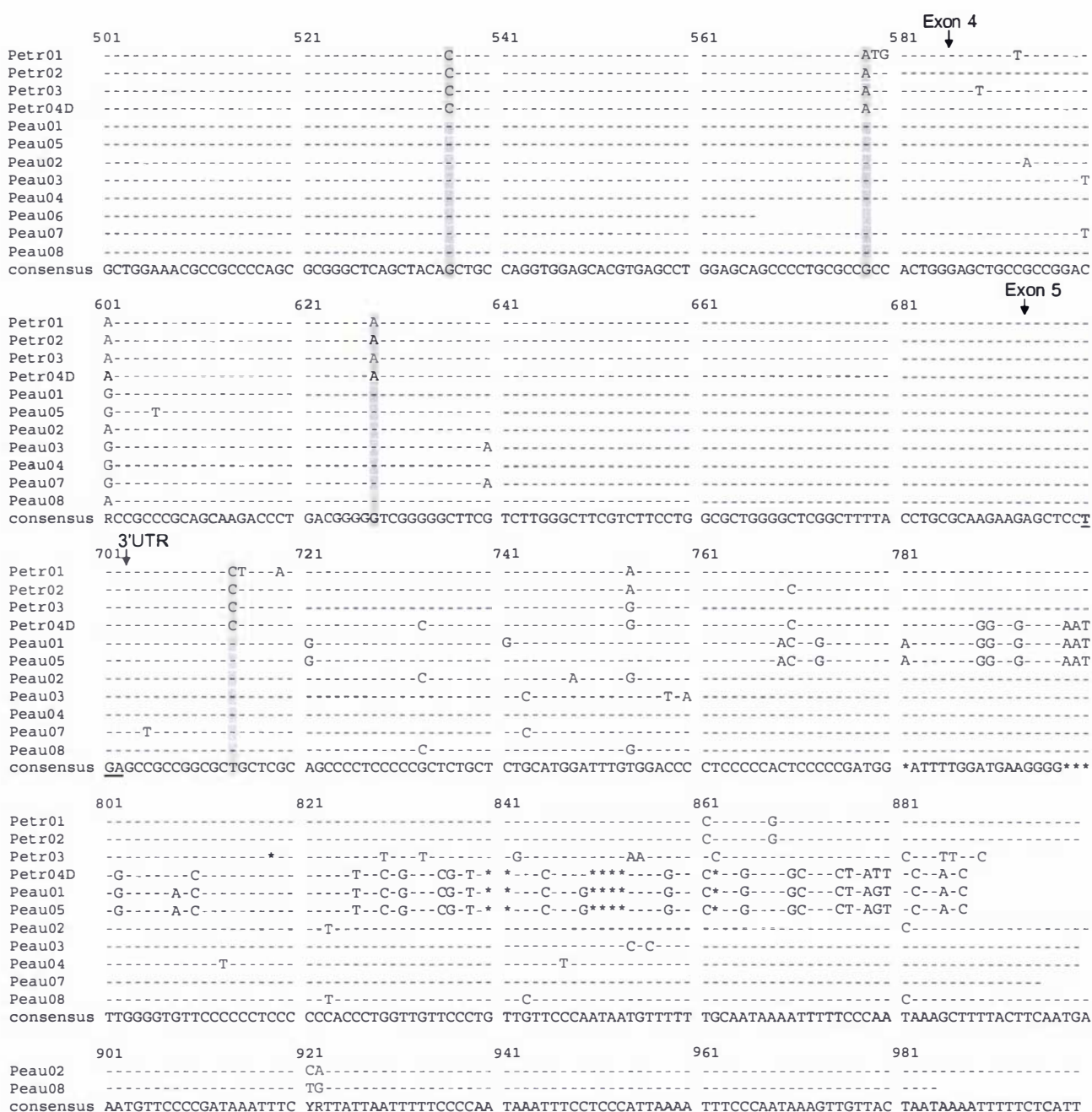


Figure 3.3 Nucleotide sequence alignment of class II B MHC cDNA sequences isolated from black robin (Petr) and South Island robin (Peau). Identity with the consensus sequence is shown with dashes, and gaps are indicated with asterisks. Fixed nucleotides that distinguish black robin from South Island robin sequences are shaded grey. Exon boundaries are marked according to the Agph-DAB1 gene (Edwards et al. 1998).

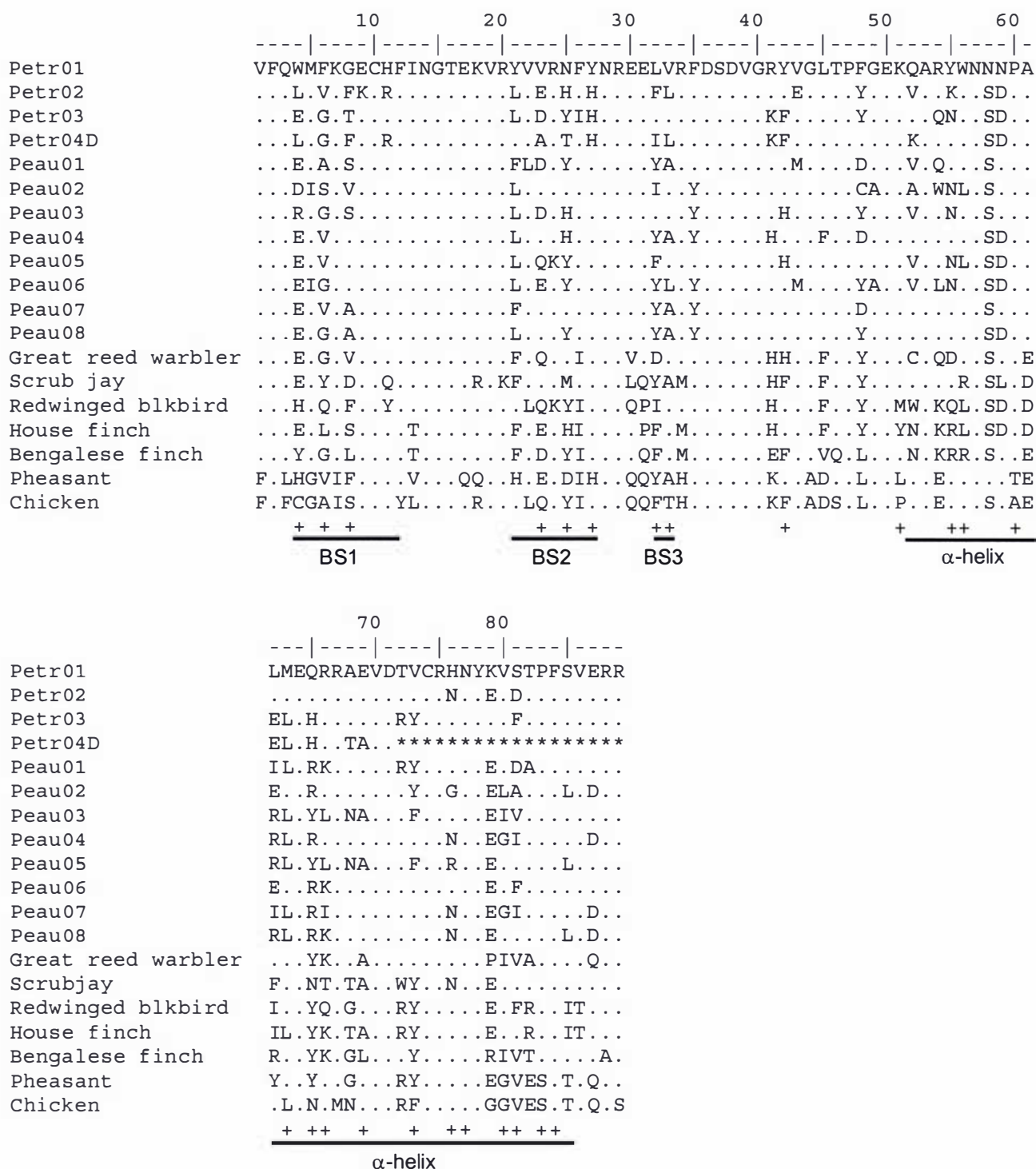


Figure 3.4 Amino acid sequence alignment of MHC class II B exon 2 sequences. Sequences are translated from cDNA sequences of black robin (Petr) and South Island robin (Peau), and compared with other avian cDNA sequences. Dots indicate identity with the Petr01 sequence and gaps are indicated with asterisks. Polymorphic subdomains recognised by She et al. (1991) are underlined (BS = β sheet), and amino acids that contact the peptide in the human DR β molecule (Brown et al. 1993) are indicated with a plus sign. References and accession numbers for other sequences are: chicken (BLBII from B¹² haplotype), AL023516, Kaufman et al. (1999b); pheasant (Phco-DAB1), AJ224344, Wittzell et al. (1999b); great reed warbler, AF404371 (cO2), Westerdahl et al. (2000); Bengalese finch, L42335 (Lost 5), Vincek et al. (1995); red-winged black bird, U23970 (Agph1.1), house finch, U23976 (Came 2.1), Scrubjay, U23975 (Apco 2.1), all Edwards et al. (1995b).

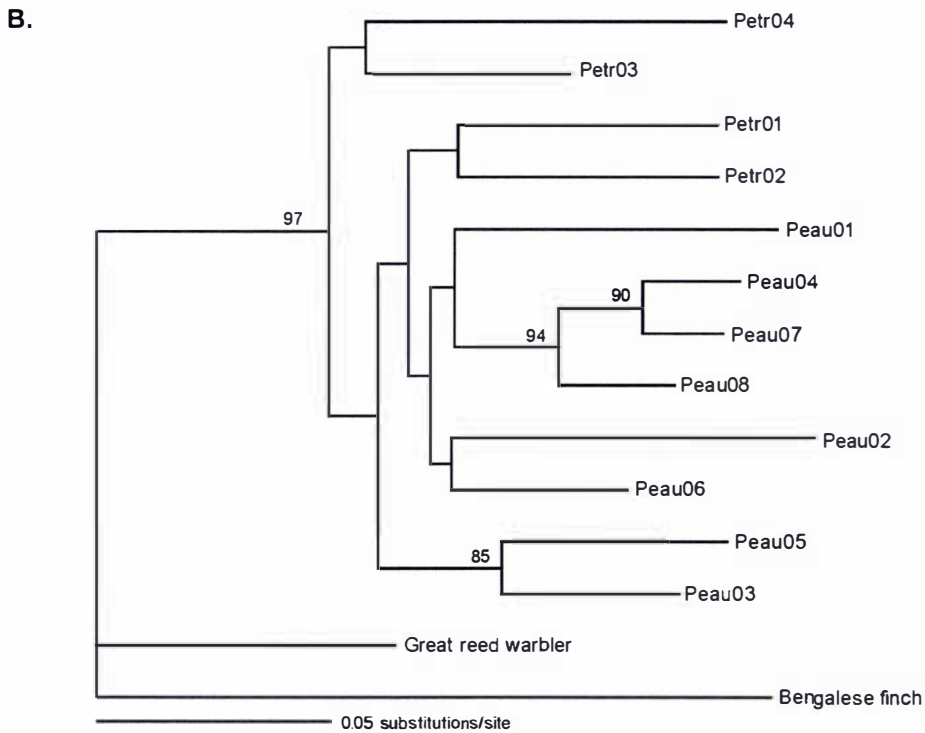
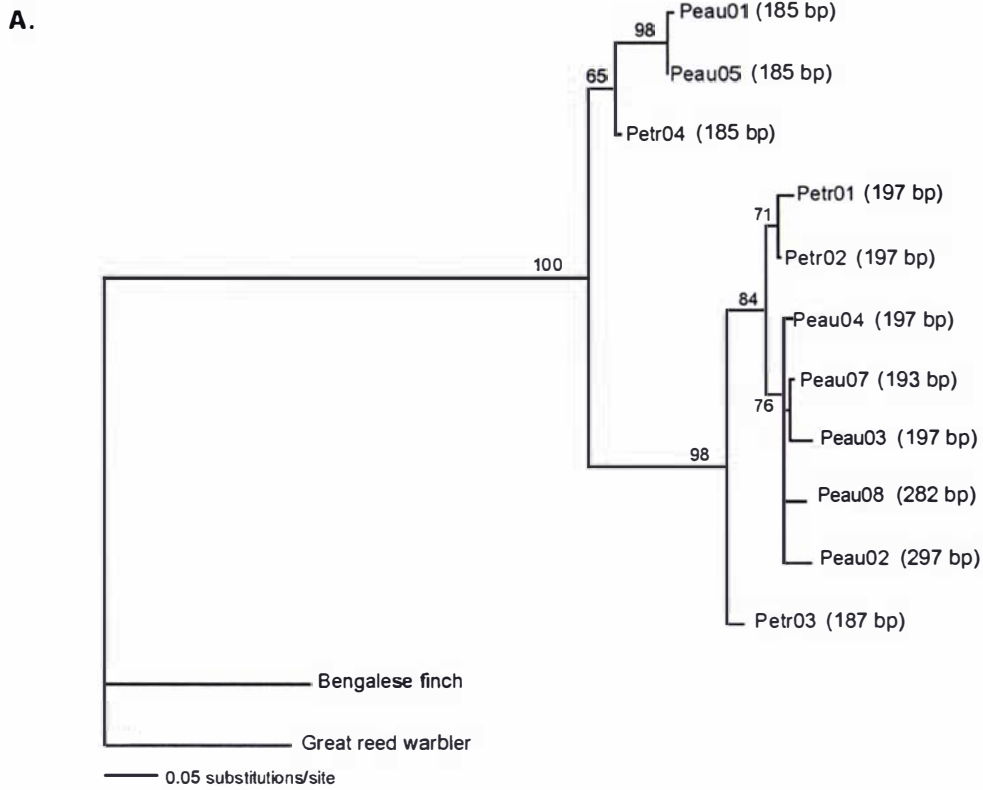


Figure 3.5 Neighbour joining trees using Jukes-Cantor correction of 3'UTR sequences (A), and class II B MHC exon 2 sequences (B) from black robin (Petr) and South Island robin (Peau). Bootstrap values from 500 replicates are given above the branches. The length of the 3'UTR is given in brackets after the sequence name. The Bengalese finch and great reed warbler sequences which form the outgroup are cDNA sequences from Vincek et al. (1995) and Westerdahl et al. (2000).

Table 3.2 Nonsynonymous (d_N) and synonymous (d_S) substitution values (\pm SE) for class II B MHC sequences from black robin (Petr) and South Island robin (Peau). The analysis was performed on groups of putative orthologous loci. Grp 1 = Petr04, Peau01, Peau05; Grp 2 = Petr01, Petr02, Peau03, Peau04, Peau07. PB = peptide binding (amino acids that contact the peptide). P indicates the probability that $d_N > d_S$.

Sequences		d_N	d_S	d_N/d_S	p
exons 3,4,5	Grp1	0.019 \pm 0.006	0.013 \pm 0.009	1.46	0.297
	Grp2	0.018 \pm 0.007	0.010 \pm 0.006	1.14	0.128
exon 2	Grp1	0.201 \pm 0.033	0.097 \pm 0.033	2.08	< 0.005
	Grp2	0.140 \pm 0.026	0.075 \pm 0.024	1.87	< 0.05
exon 2 PB	Grp1	0.527 \pm 0.122	0.208 \pm 0.100	2.54	< 0.05
	Grp2	0.393 \pm 0.095	0.194 \pm 0.100	2.03	0.055
exon 2 non-PB	Grp1	0.115 \pm 0.030	0.061 \pm 0.032	1.9	0.077
	Grp2	0.068 \pm 0.018	0.042 \pm 0.021	1.64	0.164

Table 3.3 Gene conversion events between class II B MHC sequences from New Zealand robins, identified using GENECONV. Sim p = simulated p values based on 10,000 permutations. *Identified using the within-species group option. [†]Pairwise p -values with Bonferroni-correction for multiple comparisons. All other p -values are global values, which have a built in correction for multiple comparisons. Begin = first nucleotide of the converted region, End = last nucleotide of the converted region, based on nucleotide numbering in Figure 3.3. Length = length of the converted region. Gscale indicates the mismatch penalty (see text).

Seq 1	Seq 2	Sim p	Begin	End	Length	gscale
Peau01	Peau02	0.0365*	184	720	537	1
Peau01	Peau04	0.0123	190	768	579	1
Peau01	Peau08	0.0431*	337	720	384	1
Peau02	Peau05	0.0060	1	720	720	1
		0.0282*	1	605	605	0
Peau02	Peau07	0.0385 [†]	755	880	126	1
Peau05	Peau04	0.0008	190	768	579	1
Peau05	Peau03	0.0420	1	757	757	1
Petr01	Peau07	0.0495 [†]	755	887	133	1
		0.0442	755	860	106	0
Petr01	Petr04	0.0245	1	714	714	0
		0.0482*	1	714	714	1
Petr03	Petr04	0.0009	1	769	769	2
		0.0000	1	788	788	1
Petr04	Peau02	0.0194	1	788	788	1
Petr04	Peau04	0.0428	184	788	605	1
Petr04	Peau08	0.0032	95	788	694	1
		0.0245	715	769	55	0
		0.0110	236	769	534	2

3.4 Discussion

Work to date on the passerine MHC suggests that it is somewhat more complex than the chicken MHC, being characterised by a higher number of loci, pseudogenes and gene fragments, recent duplications and/or concerted evolution (reviewed in Hess and Edwards 2002). In this chapter, RFLP analysis was used to investigate the complexity of the MHC in New Zealand robins, and gain an approximate estimate of the number of loci. All black robins had an identical RFLP genotype, however moderate to high levels of variation were recorded in the South Island robin profiles, suggesting the monomorphism evident in black robin profiles is a reflection of the black robin population history, rather than an inherent feature of the MHC region in New Zealand robins.

Using a probe to exon 2, we found six bands per individual in black robin and 6-15 bands in the South Island robin. Additional bands were detected when hybridisation was carried out using a probe to exon 3. This trend was also observed in red-winged blackbird, house finch, and scrub jay (Edwards et al. 2000b). Exon 3 is a conserved, immunoglobulin-like domain and thus these additional faint bands may represent non-specific hybridisation. The number of hybridising bands obtained in other passerine RFLP profiles using an exon 2 probe varies considerably, from 2-4 bands per individual in the house finch (Edwards et al. 2000b), to up to 30 bands per individual in the willow warbler (Westerdahl et al. 2000), and savannah sparrow (Freeman-Gallant et al. 2002). The number of bands in New Zealand robin profiles is similar to the number of bands identified in red-winged blackbird, and scrub jay profiles (8-11 bands, Edwards et al. 1999), and slightly lower than number in great reed warbler (13-17 bands, Westerdahl et al. 2000). Our data add to the suggestion that the number of class II MHC genes in passerines is higher than that of chicken, but that the exact number may vary considerably among species.

An inheritance analysis of RFLP bands in families of robins suggests that class II MHC genes in robins are contained in a single linkage group. In chicken, a second linkage group containing non-classical MHC genes, the *Rfp-Y* locus, has been identified (Briles et al. 1993). The *Rfp-Y* locus is located on the same microchromosome as the B locus but is separated from it by the nucleolar organiser region, a region of high recombination, and is therefore unlinked to the B locus (Miller 1996). A similar locus has been found in pheasants (although so far is only known to contain class II B sequences (Wittzell et al. 1995)), but it is not

known whether it is present in other avian families. Work on the great reed warbler suggests a single MHC linkage group (Westerdahl et al. 2000), however two linkage groups containing class II genes were identified in the savannah sparrow (Freeman-Gallant et al. 2002). Our study was limited by small family size (only two chicks per family), and the small number of variable bands in these families. Although all the variable bands in our study appeared to be linked, we cannot rule out the possibility of a second linkage group containing non-polymorphic bands.

Based on the number of hybridising bands in the RFLP analysis, there may be six class II B MHC loci in New Zealand robins. However, it was not possible to determine an accurate number of loci from this method, or determine whether the hybridising bands represent functional loci. Although the sequence of the BR6 probe used in RFLP analysis was divergent from the cDNA sequences, the level of divergence (76-78%) would be unlikely to prevent hybridisation with the cDNA sequences. Thus a number of the bands seen in the RFLP profiles may represent the cDNA sequences.

Using an RT-PCR based approach, we found four transcribed class II B MHC sequences in the black robin, and eight transcribed sequences in the South Island robin, indicating that there are at least four transcribed class II B MHC loci in New Zealand robins. However, we cannot make any inference about the level of expression of any of these loci from our RT-PCR based approach. This is a conservative estimate of the number of class II B genes in New Zealand robins. Additional sequences were occasionally found with the 05 +07 and 06 + RACE-AP primers, but only in a single clone (for example, see chapter four). These rare sequences were not included in our results as they could not be verified in independent PCRs, however they generally differed from other transcripts by more than 5 bp so did not seem to have arisen as a result of PCR error or *in vitro* recombination. It is more likely that these transcripts were rare species in the PCR product, and therefore many more clones would need to be sequenced to obtain these transcripts more than once. Also, the nature of the sequences obtained is restricted by the primers we used. We have used primers placed in conserved regions (and degenerate primers if necessary) in an attempt to amplify all class II B MHC sequences present, but cannot rule out the possibility of preferential amplification of some sequences.

The results presented here are consistent with the suggestion that the black robin, which is represented by a small, bottlenecked population is homozygous, and the South Island robin is heterozygous for these four loci. However, distinguishing alleles from loci is a problem in avian MHC studies and we also cannot rule out the possibility that there may be a difference in gene copy number between the black robin and South Island robin. Heterogeneity in gene number within species or between closely related species is common, for instance among DRB genes in primates (Brandle et al. 1992, Slierendregt et al. 1994), and within the class II MHC region cichlid fishes (Malaga-Trillo et al. 1998). However, the prevalence of this in closely related bird species has not been established.

One of the black robin sequences (Petr04) appears to contain an mRNA splicing error and may not be functional. However all other cDNA sequences contain the hallmarks of functional class II loci, as they have a high level of divergence in exon 2 with an excess of non-synonymous nucleotide substitutions at peptide-binding codons, and contain the conserved amino acids typical of classical class II molecules. The sequence of the BR6 probe also contains all conserved residues expected in a functional class II MHC gene and does not contain any deletions or premature stop codons. However, it is highly divergent from the transcribed sequences, raising the possibility that it represents a pseudogene. This possibility is further investigated in chapter four by isolation of a longer portion of this locus.

3.4.1 Evolution of class II B MHC genes in NZ robins.

It has been suggested that the 3'UTR is more conserved within than between loci and may be a more accurate indicator of orthology than the coding sequences (Westerdahl et al. 1999, Wittzell et al. 1999b). Among the sequences isolated here there were considerable differences among the 3'UTRs, and on this basis, one orthologous locus (represented by sequences Petr04 in black robin, and Peau01/05 in South Island robin) could be putatively identified between the two species. The remaining sequences share significant sequence similarities and it is difficult to establish whether there are orthologous pairs of loci among them. However, sequences Petr01, Petr02, and Peau03, Peau04 and Peau07 are the same length and therefore may represent a second orthologous group, perhaps containing two loci.

These data are consistent with the idea of multiple rounds of gene duplication occurring in the New Zealand robin MHC. The putative orthologs Petr04, Peau01, and Peau05 may be

the result of an older duplication, while more recent duplication events may have produced multiple loci with the other 3'UTR. The number of non-synonymous and synonymous substitutions was consistently higher across the whole gene for the first group of sequences (Grp 1, Table 3.2), suggesting that these sequences have been diverging from each other for a longer period of time. Orthologous genes have been identified in chicken and pheasant (Wittzell et al. 1999b), however this is the first demonstration in passerines of orthologous genes identified on the basis of non-coding sequences. Geological and mitochondrial DNA evidence suggests that the black robin and South Island robin diverged within the last 4 million years (see chapter six), so at least the first duplication event must predate this split. However, it is possible that some of the more recent duplications within the second group of sequences occurred after the divergence of the two species.

Among passerines, full-length cDNA sequences from class II B MHC genes have been isolated from great reed warbler (Westerdahl et al. 2000), and the Bengalese finch (Vincek et al. 1995), both of which are members of the Passerida (*sensu* Sibley and Ahlquist 1990) group of songbirds. However, a BLAST search revealed no similarity between the 3'UTR sequences from these species and New Zealand robins, and there was no evidence of orthology either between the New Zealand robin and Passerida sequences, or between the warbler and finch sequences on a phylogenetic tree (data not shown). The Passerida are thought to have diverged from the basal Australo-Papuan lineages about 15-25 million years ago (Ericson et al. 2002). By contrast, orthologous loci are evident in chicken and pheasant, which are estimated to have diverged about 20 million years ago (Helm-Bychowski and Wilson 1986). This indicates that gene turnover may occur more quickly in passerines than in galliforms. It should be noted however, that it is difficult to distinguish between recent gene duplications and concerted evolution. Thus, the duplications evident in NZ robins may have occurred within the lineage separating Petroicidae from the Passerida (i.e. between about 4-25 million years ago) or alternatively, concerted evolution across long tracts of sequence, including the entire the 3'UTR, may have occurred within this lineage and obscured more distant relationships. Our results suggest that if such concerted evolution does occur, the rate is not high enough to obscure orthologous loci within the *Petroica* genus, but does obscure orthologous relationships in more distantly related species. Further taxonomic sampling, particularly of other Australo-Papuan songbirds, may aid in narrowing down the timing of these events.

The rate of gene conversion across short tracts of coding sequence, however, does appear to be high enough to obscure orthologous relationships within the coding regions of the sequences isolated here. Sequences within the coding region grouped within species, and using GENECONV we identified numerous putative gene conversion events from 55 to 788 bp in length among sequences from both species. These gene conversion events can account for the loss of orthologous relationships between the 3'UTR and the coding region, and homogenisation (concerted evolution) of exons 3, 4 and 5 within species. Sequences within exon 2, however, are not homogenised. The elevated diversity in this region, the excess of non-synonymous changes in peptide-binding codons and the elevated d_N/d_S ratios in exon 2 suggest these sequences have diverged from each other under a model that includes point mutation and balancing selection. However, it is difficult to determine the magnitude of balancing selection as the ratio of d_N/d_S may be underestimated in comparisons between divergent sequences (i.e. interlocus comparisons), due to saturation of nonsynonymous changes (Takahata et al. 1992, Edwards et al. 1995a). Further evidence of balancing selection, such as high numbers of alleles, uniform allele frequencies, and heterozygote excess, will require population level data on allelic diversity at single loci.

It is also possible that gene conversion within exon 2 adds to the diversity of exon 2 sequences, but was not detected in our analysis. Shared motifs across short tracts of sequence within exon 2 are present, and the number of synonymous changes was higher in exon 2 than for the remainder of the gene, which may also be indicative of gene conversion in this region (Bergstrom et al. 1998). Our analysis of gene conversion events using GENECONV used silent sites only in order to avoid any confounding effects due to selection. This meant that a large number of nucleotides from exon 2 were removed from the analysis, and gene conversion events occurring over short tracts in this region may have been missed. Motif shuffling by gene conversion with exon 2 has been reported in several studies (e.g. She et al. 1991, Edwards et al. 1998, Langefors et al. 2001), and in combination with point mutation and balancing selection, can create a highly divergent peptide-binding region. In fact, it has been suggested that gene conversion is the principal mutational mode in the MHC (Martinson et al. 1999).

Analysis of class II B MHC sequences isolated from the black robin and South Island robin supports the hypothesis that gene conversion and concerted evolution are important components of MHC evolution in passerines. Features of the birth-and-death model of

evolution (i.e. gene duplication and decay) are also evident, adding to the suggestion that birth-and-death and concerted evolution models are not mutually exclusive (Witzzell et al. 1999b). The data presented here also suggest that care should be taken when inferring loci/allelic relationships from exon 2 sequences alone. Isolation of further full-length MHC cDNA sequences from related songbird species is likely to establish additional orthologous relationships between avian MHC genes and provide further information on the timescales over which the evolutionary processes influencing MHC genes occur.

3.5 References

- Afanassieff M, Goto RM, Ha J, Sherman MA, Zhong LW, Auffray C, Coudert F, Zoorob R, Miller MM (2001) At least one class I gene in restriction fragment pattern-Y (Rfp-Y), the second MHC gene cluster in the chicken, is transcribed, polymorphic, and shows divergent specialization in antigen binding region. *J. Immunol.* 166: 3324-3333
- Andersson L, Gustafsson K, Jonsson A, Rask L (1991) Concerted evolution in a segment of the first domain exon of polymorphic MHC class II B loci. *Immunogenetics* 33: 235-242
- Apanius V, Penn D, Slev PR, Ruff LR, Potts WK (1997) The nature of selection on the major histocompatibility complex. *Crit. Rev. Immunol.* 17: 179-224
- Ardern SL, McLean IG, Anderson S, Maloney R, Lambert DM (1994) The effects of blood sampling on the behaviour and survival of the endangered Chatham Island Black Robin (*Petroica traversi*). *Cons. Biol.* 8: 857-862
- Ardern SL and Lambert DM (1997) Is the black robin in genetic peril? *Mol. Ecol.* 6: 21-28
- Ardern SL, Lambert DM, Rodrigo AG, McLean IG (1997a) The effects of population bottlenecks on multilocus DNA variation in robins. *J. Heredity* 88: 179-186
- Ardern SL, Ma W, G. EJ, Armstrong DP, Lambert DM (1997b) Social and Sexual Monogamy in Translocated New Zealand Robin Populations Detected Using Minisatellite DNA. *The Auk* 114: 120-126
- Bergstrom TF, Josefsson A, Erlich HA, Gyllensten U (1998) Recent origin of *HLA-DRB1* alleles and implications for human evolution. *Nature Genetics* 18: 237-242
- Brandle U, Ono H, Vincek V, Klein D, Golubic M, Grahovac B, Klein J (1992) Transspecies Evolution of Mhc-Drb Haplotype Polymorphism in Primates - Organization of Drb Genes in the Chimpanzee. *Immunogenetics* 36: 39-48
- Briles WE, Goto RM, Auffray C, Miller MM (1993) A polymorphic system related to but genetically independent of the chicken major histocompatibility complex. *Immunogenetics* 37: 408-414
- Brown JH, Jardetzky TS, Gorga JC, Stern LJ, Urban RG, Strominger JL, Wiley DC (1993) Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature* 364: 33-39
- Butler D and Merton D (1992) *The Black Robin: Saving the World's Most Endangered Bird*. Oxford University Press
- Drouin G, Prat F, Ell M, Clarke P (1999) Detecting and characterising gene conversion events between multigene family members. *Mol. Biol. Evol.* 16: 1369-1390

- Edwards SV, Grahn M, Potts WK (1995a) Dynamics of *Mhc* evolution in birds and crocodilians: amplification of class II genes with degenerate primers. *Mol. Ecol.* 4: 719-729
- Edwards SV, Wakeland EK, Potts WK (1995b) Contrasting histories of avian and mammalian *Mhc* genes revealed by class II B sequences from songbirds. *Proc. Natl. Acad. Sci. USA* 92: 12200-12204
- Edwards SV, Gasper J, March M (1998) Genomics and polymorphism of *Agph-DABI*, an MHC Class II B gene in red-winged blackbirds (*Agelaius phoeniceus*). *Mol. Biol. Evol.* 15: 236-250
- Edwards SV, Hess CM, Gasper J, Garrigan D (1999) Toward an evolutionary genomics of the avian Mhc. *Immunol. Rev.* 167: 119-132
- Edwards SV, Gasper J, Garrigan D, Martindale D, Koop BF (2000a) A 39kb sequence around a blackbird *Mhc* Class II gene: Ghost of selection past and songbird genome architecture. *Mol. Biol. Evol.* 17: 1384-1395
- Edwards SV, Nusser J, Gasper JS (2000b) Characterization and evolution of major histocompatibility complex (MHC) genes in non-model organisms, with examples from birds. *In* AJ Baker (ed.): *Molecular Methods in Ecology*. Blackwell Science, pp. 168-207
- Ericson PGP, Christidis L, Cooper A, Irestedt M, Jackson J, Johansson US, Norman JA (2002) A Gondwanan origin of passerine birds supported by DNA sequences of the endemic New Zealand wrens. *Proc. R. Soc. Lond. B.* 269: 235-241
- Flack JAD (1974) Chatham Island black robin. *Wildlife - a Review, New Zealand Wildlife Service* 5: 25-34
- Freeman-Gallant CR, Johnson EM, Saponara F, Stanger M (2002) Variation at the major histocompatibility complex in Savannah sparrows. *Mol. Ecol.* 11: 1125-1130
- Frohman MA, Dush MK, Martin GR (1988) Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci. USA* 85: 8998-9002
- Gasper JS, Shiina T, Inoko H, Edwards SV (2001) Songbird genomics: Analysis of 45 kb upstream of a polymorphic Mhc class II gene in red-winged blackbirds (*Agelaius phoeniceus*). *Genomics* 75: 26-34
- Gu X and Nei M (1999) Locus specificity of polymorphic alleles and evolution by a birth-and-death process in mammalian MHC genes. *Mol. Biol. Evol.* 16: 147-156
- Helm-Bychowski KM and Wilson AC (1986) Rates of nuclear DNA evolution in pheasant-like birds: evidence from restriction maps. *Proc. Natl. Acad. Sci. USA* 83: 688-692.
- Hess CM, Gasper J, Hoekstra HE, Hill CE, Edwards SV (2000) MHC Class II pseudogene and genomic signature of a 32 kb cosmid in the house finch (*Carpodacus mexicanus*). *Genome Research* 10: 613-623
- Hess CM and Edwards SV (2002) The evolution of the major histocompatibility complex in birds. *BioScience* 52: 423-431
- Hughes AL and Nei M (1989) Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. *Proc. Natl. Acad. Sci. USA* 86: 958-962
- Jukes TH and Cantor CR (1969) Evolution of protein molecules. *In* HN Monroe (ed.): *Mammalian Protein Metabolism*. Academic Press, New York, pp. 21-132
- Kaufman J, Salomonsen J, Flajnik M (1994) Evolutionary conservation of MHC class I and class II molecules - different yet the same. *Seminars in Immunology* 6: 411-424

- Kaufman J, Volk H, Wallny H-J (1995) A "Minimal Essential MHC" and an "Unrecognised MHC": Two extremes in selection for polymorphism. *Immunol. Rev.* 143: 63-88
- Kaufman J, Jacob J, Iain S, Walker B, Milne S, Beck S, Salomonsen J (1999a) Gene organisation determines evolution of function in the chicken MHC. *Immunol. Rev.* 167: 101-117
- Kaufman J, Milne S, Gobel TWF, Walker BA, Jacob JP, Auffray C, Zoorob R, Beck S (1999b) The chicken B locus is a minimal essential major histocompatibility complex. *Nature* 401: 923-925
- Klein J (1986) *Natural History of the Major Histocompatibility Complex*. John Wiley and Sons, New York.
- Kulski JK, Shiina T, Anzai T, Kohara S, Inoko H (2002) Comparative genomic analysis of the MHC: the evolution of class I duplication blocks, diversity and complexity from shark to man. *Immunol. Rev.* 190: 95-122
- Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 17: 1244-1245
- Langefors K, Lohm J, von Schantz T (2001) Allelic polymorphism in MHC class IIB in four populations of Atlantic salmon (*Salmo salar*). *Immunogenetics* 53: 329-336
- Malaga-Trillo E, Zaleska-Rutczynska Z, McAndrew B, Vincek V, Figueroa F, Sultmann H, Klein J (1998) Linkage Relationships and Haplotype Polymorphism Among Cichlid Mhc Class II B Loci. *Genetics* 149: 1527-1537
- Martinsohn JT, Sousa AB, Guethlein LA, Howard JC (1999) The gene conversion hypothesis of MHC evolution: a review. *Immunogenetics* 50: 168-200
- Meyer D and Thomson G (2001) How selection shapes variation of the human major histocompatibility complex: a review. *Annals of Human Genetics* 65: 1-26
- MHC Sequencing Consortium (1999) Complete sequence and gene map of a human major histocompatibility complex. *Nature* 401: 921-923
- Miller HC, Lambert DM, Millar CD, Robertson BC, Minot EO (2003) Minisatellite DNA profiling detects lineages and parentage in the endangered kakapo (*Strigops habroptilus*) despite low microsatellite DNA variation. *Conservation Genetics* 4: 265-274
- Miller MM, Goto, R., Bernot, A., Zoorob, R., Auffray, C., Briles, R. W., Briles, W. E., Bloom, S. E., (1996) Assignment of Rfp-Y to the chicken MHC/NOR microchromosome and evidence for high frequency recombination associated with the nucleolar organiser region. *Proc. Natl. Acad. Sci. USA* 93: 3958-3962
- Miyada CG, Klofolt C, Reyes AA, McLaughlin-Taylor E, Wallace RB (1985) Evidence that polymorphism in the murine major histocompatibility complex may be generated by the assortment of subgene sequences. *Proc. Natl. Acad. Sci USA* 82: 2890-2894
- Nei M and Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* 3: 418-426
- Nei M, Gu X, Sitnikova T (1997) Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proc. Natl. Acad. Sci. USA* 94: 7799-7806
- Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular cloning: a laboratory manual*, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

- Sato A, Figueroa F, Mayer WE, Grant PR, Grant BR, Klein J (2000) Mhc class II genes of Darwin's Finches: divergence by point mutations and reciprocal recombination. *In* M Kasahara (ed.): Major Histocompatibility Complex: Evolution, Structure and Function. Springer Verlag
- Satta Y (1997) Effects of intra-locus recombination on *HLA* polymorphism. *Hereditas* 127: 105-112
- Sawyer SA (1999) GENECONV: A computer package for the statistical detection of gene conversion. Distributed by the author, Department of Mathematics, Washington University, St Louis (available at <http://www.math.wustl.edu/~sawyer>)
- She JX, Boehme SA, Wang TW, Bonhomme F, Wakeland EK (1991) Amplification of major histocompatibility complex class II gene diversity by intraexonic recombination. *Proc. Natl. Acad. Sci. USA* 88: 453-457
- Sibley CG and Ahlquist JE (1990) Phylogeny and classification of birds: a study in molecular evolution. Yale University Press, New Haven & London
- Slierendregt BL, Otting N, Vanbesouw N, Jonker M, Bontrop RE (1994) Expansion and Contraction of Rhesus Macaque Drb Regions by Duplication and Deletion. *J. Immunol* 152: 2298-2307
- Swofford DL (2002) PAUP* Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, MA
- Takahata N, Satta Y, Klein J (1992) Polymorphism and balancing selection at major histocompatibility complex loci. *Genetics* 130: 925-938
- Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* 22: 4673-4680
- Trowsdale J (1995) "Both man and bird and beast": comparative organisation of MHC genes. *Immunogenetics* 41: 1-17
- Vincek V, Klein D, Graser RT, Figueroa F, O'hUigin C, Klein J (1995) Molecular cloning of major histocompatibility complex class II B gene cDNA from the Bengalese finch *Lonchura striata*. *Immunogenetics* 42: 262-267
- Westerdahl H, Wittzell H, von Schantz T (1999) Polymorphism and transcription of Mhc class I genes in a passerine bird, the great reed warbler. *Immunogenetics* 49: 158-170
- Westerdahl H, Wittzell H, von Schantz T (2000) Mhc diversity in two passerine birds: no evidence for a minimal essential Mhc. *Immunogenetics* 52: 92-100
- Wittzell H, von Schantz T, Zoorob R, Auffray C (1995) *Rfp-Y*-like sequences assort independently of pheasant *Mhc* genes. *Immunogenetics* 42: 68-71
- Wittzell H, Madsen T, Westerdahl H, Shine R, von Schantz T (1999a) MHC variation in birds and reptiles. *Genetica* 104: 301-309
- Wittzell H, Bernot A, Auffray C, Zoorob R (1999b) Concerted evolution of two MHC Class II loci in pheasants and domestic chickens. *Mol. Biol. Evol.* 16: 479-490
- Zoorob R, Bernot A, Renoir DM, Choukri F, Auffray C (1993) Chicken major histocompatibility complex class II B genes: analysis of interallelic and interlocus sequence variance. *Eur J. Immunol.* 23: 1139-1145

CHAPTER FOUR

A Divergent Class II B MHC Sequence in the Black Robin

4.1 Introduction

MHC genes appear to evolve according to a birth-and-death model of evolution, in which new genes arise by duplication, while others are deleted or become pseudogenes (Nei et al. 1997). In accordance with this, pseudogenes appear to be ubiquitous in both the class I and class II MHC regions in mammals (Brandle et al. 1992, Hughes 1995, Grimsley et al. 1998, MHC Sequencing Consortium 1999, Renard et al. 2001) and have also been described in fish (Shum et al. 2002). Little is known, however, about how widespread the occurrence of pseudogenes is in birds. The chicken MHC appears to be an exception among vertebrate MHC, as no pseudogenes are present (Kaufman et al. 1999b), however pseudogenes have been identified in two species of songbird, the House finch *Carpodacus mexicanus* (Hess et al. 2000), and red-winged blackbird *Agelaius phoeniceus* (Edwards et al. 2000).

In this chapter, the isolation of a partial genomic DNA sequence of a possible pseudogene in the black robin is reported. Four transcribed class II B MHC sequences have been isolated from black robin (see chapter three), however the sequence of the BR6 probe used for RFLP analysis, is divergent from these sequences. BR6 was isolated from black robin genomic DNA using the PCR primers 305 and 306 from Edwards et al. (1995a), and spans part of exon 2 of a class II B MHC gene. The possibility that the BR6 sequence represents a pseudogene is investigated in this chapter by using primers specific to BR6 to isolate a longer portion of this locus, and to analyse whether this sequence is transcribed.

4.2 Methods

PCR primers specific for the BR6 sequence (named MHC01: 5'-TCAATGGCACTGAGCG

AGTGAGG-3' and MHC02: 5'-AGTTGTGACGGCAGTACGTGTCC-3') were designed slightly internal to the 305 and 306 primers from Edwards et al. (1995a). In order to test whether the BR6 sequence is transcribed, RT-PCR was performed using MHC01 paired with MHC07, and MHC02 paired with MHC05 (see Table 4.1 for details of PCR primers used in this chapter, and chapter three for the sequences of MHC05 and MHC07). These primer pairs both span an intron, enabling cDNA sequences to be distinguished from genomic sequences. Reverse transcription was performed as described in chapters two and three. PCR amplifications were carried out in 25 μ L volumes using the Expand *HiFi* PCR system with 1.5 mM MgCl₂, 200 μ M each dNTP, 0.4 μ M each primer and 2 μ L of cDNA. Thermal cycling was performed in a BioRad iCycler for 35 cycles consisting of 95°C for 30 sec, annealing temperature (see Table 4.1) for 10 sec, and 72°C for 1 min.

A longer fragment of the BR6 sequence was isolated from genomic DNA using the primer combinations MHC01 + MHC07, MHC05 + MHC07, and MHCpseudo3' + MHC05 or MHC06. MHCpseudo3' (5'-CCGGGCAGGGGCAGAGTGACAG-3') was designed from the 3' end of the cDNA clone 06AP/4 (see results). All PCR amplifications were performed as described above, except that 1 mM MgCl₂ was used in amplifications using the MHC05 + MHC07 primer combination, and the extension at 72°C was extended to 2 min for products expected to be longer than 1.5 kb. All PCR products were gel-purified using a HighPure PCR product purification kit (Roche), and cloned into a pGEM[®]-T Easy vector (Promega). Positive clones were sequenced in both directions (using internal primers where necessary) using the BigDye Terminator Cycle sequencing kit (Applied Biosystems) and analysed on an ABI 377A automated sequencer. The resulting sequences were edited using Sequencher[™] 4.1.2 (Gene Codes Corporation) and aligned using ClustalW (Thompson et al. 1994). Uncorrected *p*-distances were calculated in MEGA version 2.1 (Kumar et al. 2001), and neighbour-joining trees were constructed in PAUP*4.0b10 (Swofford 2002).

Table 4.1. PCR primers used in this chapter. The sequences of the MHC05, MHC06 and MHC07 primers are given in chapter three.

Primer pair	Annealing temp	Region amplified
MHC01 + MHC07	55°C	Exon 2 – exon 3
MHC05 + MHC02	55°C	Exon 1 – exon 2
MHC05 + MHC07	63°C	Exon 1 – exon 3
MHC05 + MHCpseudo3'	65°C	Exon 1 – 3'UTR
MHC06 + MHCpseudo3'	65°C	Exon 3 – 3'UTR

4.3 Results and Discussion

No PCR products were seen when RT-PCR was performed using MHC01 + MHC07, however a faint product at the correct size was seen with the MHC05 + MHC02 primer pair. Amplification from genomic DNA using MHC01 + MHC07 was successful, however. Three positive clones obtained using these primers were found to contain MHC sequences spanning approximately 1 kb, including most of exons 2 and 3, and all of intron 2. The exon 2 sequence of these clones matched exactly with the BR6 sequence. The sequence of exon 3 was also divergent from the previously isolated cDNA sequences, however it did match a single cDNA clone which was previously found using 3'RACE. This clone (named 06AP/4) was amplified using the primers MHC06 + RACE-AP, but could not be verified in any subsequent PCRs so was discarded from the analysis in chapter three. Alignment of the 06AP/4 sequence with the other cDNA sequences revealed a large deletion spanning 56 bp of exon 3 and all of exon 4. This deletion includes the binding site for the MHC07 primer, which may explain why no amplification from cDNA was possible with this primer. These results suggest that the BR6 sequence may be transcribed (although possibly at only low levels), but it does not make a functional protein.

In order to isolate a longer fragment and to confirm that the 06AP/4 cDNA sequence and the BR6 sequence do represent the same gene, the primer MHCpseudo3' was designed to the 3'UTR of 06AP/4. PCR amplification using this primer in combination with MHC05 produced a faint band at approximately 3 kb, however this product could not be successfully cloned and sequenced. Products obtained using the primer pairs MHC06 + MHCpseudo3' and MHC05 + MHC07 were successfully cloned and sequenced, however. Eight positive clones from the MHC05 + MHC07 amplification were sequenced. Two were found to match the BR6 sequence in exon 2, while the others matched the cDNA sequences Petr02 and Petr04. Two positive clones were sequenced from the MHC06 + MHCpseudo3' amplification. These sequences were identical to the MHC05 + MHC07 sequences in the region of overlap (exon 3), and could be easily aligned to the 06AP/4 sequence (Figure 4.1). These overlapping fragments thus appear to represent the same gene, which from here on is referred to as Petr05.

The Petr05 sequence spans a total of 2.7 kb and lacks only the first 27 bp of exon 1. Although the sequence does not appear to contain any insertions, deletions or stop codons, and contains

all the conserved residues in exon 2 thought to be characteristic of functional MHC genes, it is significantly divergent from the cDNA sequences (Petr01-04) isolated in chapter three. The intron sequences of Petr05 are substantially different in both length and sequence from the introns of Petr01-04 (see chapter five), and there is a significant increase in pairwise distance values for comparisons between Petr05 and Petr01-04, when compared with distance values among Petr01-04 sequences alone ($p < 0.0001$ for all exons, Table 4.2). This divergence is also reflected in phylogenetic analysis of all passerine MHC sequences. Neighbour-joining trees were constructed for exon 2 and exon 3 using almost all passerine sequences available on Genbank (Figure 4.2). These include sequences isolated using Edwards et al. (1995a) primers 305 and 306, sequences isolated from cDNA, and sequences from full-length genomic clones. A striking feature of both the exon 2 and exon 3 trees is the clustering together of all New Zealand robin cDNA sequences to the exclusion of the Petr05 sequence. Exon 3 sequences cluster according to species with high bootstrap support, with the exception of Petr05 and the House finch pseudogene CameDAB1 (Hess et al. 2000), which contains frameshift mutations in the second and third exons. However these two sequences do not have a high level of similarity to each other. The pseudogene identified in red-winged blackbird AgphDAB2, also does not appear to be closely related to the Petr05 sequence. The Petr05 exon 2 sequence falls outside almost all other passerine MHC exon 2 sequences. However a similar exon 2 sequence has been amplified from genomic DNA in two other species of songbird, the red-winged blackbird (Agph2.3) and the Savannah sparrow (Pasa 39/43). These three sequences cluster together with high bootstrap support, to the exclusion of all other sequences, indicating that loci orthologous to Petr05 may exist in other species.

The large genetic distance between Petr05 and the expressed Petr sequences may be a reflection of the loss of functional constraint on a gene that is no longer expressed and in the process of decaying. The long branch lengths separating exon 3 sequences of both Petr05 and the pseudogene CameDAB1 appear to support loss of constraint on the amount of change possible in this exon. Exon 3 codes for an immunoglobulin-like domain and is generally

Table 4.2. Average pairwise sequence diversity between the Petr05 sequence and previously isolated cDNA sequences, compared to diversity within the cDNA sequences alone.

	Exon 2	Exon 3	Exon 4
Within cDNA	0.1210 ± 0.0149	0.0071 ± 0.0036	0.0093 ± 0.0061
Between Petr05-cDNA	0.2065 ± 0.0205	0.1356 ± 0.0186	0.1782 ± 0.0360

highly conserved, as indicated by the short branch lengths separating functional exon 3 sequences. Although there is no indication of loss of function from the genomic DNA sequence, the deletion of part of exon 3 and all of exon 4 in the cDNA indicates that Petr05 does not make a functional protein. However the cDNA sequence has not been verified in a separate PCR, so the possibility that the deletion is the result of a PCR error cannot yet be ruled out. The inability to amplify Petr05 from cDNA using the MHC07 primer lends weight to the suggestion that the deletion is real, however. RT-PCR amplification with Petr05-specific primers could be used to confirm the cDNA sequence, however as only a single cDNA clone has been isolated to date, it is possible that the sequence is only transcribed at low levels.

Until the cDNA deletion is verified, we also cannot rule out the possibility that the genetic distance between Petr05 and Petr01-04 reflects the fact that Petr05 is a non-classical MHC gene, analogous to those found in the Rfp-Y locus in chicken (Briles et al. 1993). Petr05 does not appear to be orthologous to the non-classical class II MHC genes located in the Rfp-Y locus, as both classical and non classical class II MHC sequences from chicken group separately from all passerine sequences (data not shown). We have not found any evidence for a second, independently segregating cluster containing MHC genes in New Zealand robins (see chapter three), however genomic mapping of Petr05 and the expressed MHC genes is required to confirm this. In addition, analysis of population-level variation of the Petr05 gene will aid in determining whether any of the hallmarks of classical class II genes (such as high polymorphism, and elevated rates of nonsynonymous substitutions) are present.

In summary, although the genomic DNA sequence of Petr05 appears to contain all the features of a functional class II MHC gene, it is divergent from all the transcribed black robin MHC sequences, and appears to have a large deletion in the mRNA. These results indicate it is a pseudogene, providing further support for the birth-and-death model of evolution in the New Zealand robin MHC. However further work, including verification of the mRNA sequence, is required to confirm the status of this locus.

				┌ Intron 1 →
BR0507.G7	CGTGCTGGTGGCACTGGTGG	AGCTGGGAGCCCCCGGCT	GGGACGCGGAACTCTCGGG	TGAGCGGGGGGAAGGGTGG
BR0507.G7	GAACGGTGGGGAGTGGAGGG	AAAACGGGGAGGTGGTGATG	CCAGAATGAGATTTCTTTT	CTTTTCTATACCTTTTACAC
BR0507.G7	AGTTCTGTGTATCCACAAGT	ATTAGCAGTATGCATACTTG	TAACCCTTCACCTAGTCTAG	TGCTTAGTGTTTTGTAAAGT
BR0507.G7	TTCTGCAGGCCCTGAGACTG	AACACCCAGAACTGCAGAG	CCAGGGAGACCAAACATTC	TACTGTCTTAGGAAACCAAG
BR0507.G7	GTCTCCTAGTAGGACAGATT	CTGTAAGCTAGAGGTGAGCC	CAGCCAGCCCCATTATG	GCCTCCTGTGTAGATACGCT
BR0507.G7	GATGAGTAAAGTCTATAACT	TGTGTACGTGGTTGATGGAG	TGTGTGAATTGCCACGCTGT	CACCTGCGTGAAGTGAATGC
BR0507.G7	ATCAAACAGGATCCTAATTA	AAATACCCAGSTAACGACCC	CCAACCCTCTAACTGTGTCT	AGCTTAATTTTTGTAACGCC
BR0507.G7	AGGAAAAGGCAGCAGTGGG	ACCCTCAAACGGATGTGGGG	GTGTTGCTGGCGATGGGGGA	GCTATGGGAAAGGGAGAAGG
BR0507.G7	GTGAAAAGAGGAGGTTATG	GGACCACCACAAGGGGTGGC	ACGTGGGGCACAGGGAGTGC	TGAGTGGGGTCCGGGGTGG
BR0507.G7	CCCTCTGAGCTCTTGATTT	CTGTGGGGTGTGGGGCTGC	TGGAGGGGCACCCCTGACC	TGTGTCTGCACACTCAGGG
Petr01-04				..
BR0507.G7	GTCTTCCAGGTGATGGTAAA	GTGCGAGTGTCACTTCAATA	ACGGCACCCGACAGAGTGAGG	TTCGTGAAGAGGTTTCACTA
Petr01-04T.....K...	..T.....R.....T..G..G..AG.....	.WS...G.....AA.T..C..
BR0507.G7	CAACCGGAAGCAGTTCCTGC	ACTTCGACAGCGATGTGGGG	CTGTTTGAGGGGGACACCCC	CTATGGGGAGAAGGTTGCAA
Petr01-04G..G....MS..A	GG.....	MRS.W..T....CT.....	..W.....ACAG..CC
BR0507.G7	GGTACTGGAACAGAGACTTG	GAATGGATGGAGTACAGACG	GGCTGCGGTGGACAGGCACT	GCCGGCACAACACTACAAGTGG
Petr01-04	..A.....C...CC.	.CASWAM.....C.S.....	..A.A.....S.KW..R..G..
			┌ Intron 2 →	
BR0507.G7	TCCACGCCGTCCCCGTGGG	TCGCAGAGGTGAGCGTGGGG	CAGAGCGTCCCTCGGGCCCT	GCCCTAGGAATGACTCTGGA
Petr01-04	G....C.....AG.....A	G...C..		
BR0507.G7	GCCCCAAAACCTCCCTGGG	AATTACCCAGACCTCTCAG	TTCTCCCTGTGCCACGCCC	GTGGCCCCTTGCCCACCCCG
BR0507.G7	TAGTGTCCCGAAGCTCCCC	TGTCTCTCCAGTCTCTCC	AGTGCCCGCCCGCTTTTCC	ATCCCGCCGTGCGTGGAGC
BR0507.G7	TGGAGTCTTTAGCCAATCA	GAGTGCATCTCTGATGAC	TCATCAGTTGCCAGGCAGCT	CCGACGCGGAGTGCCCCG
BR0507.G7	CGCTGGACTCGGCTCCAG	CGCCGCAGCTTCATTCCCA	GTCTTTCCAGTTCATTCCC	ATCCCATTCCTAATCCTTCT
BR0507.G7	CAGTCCACTCCAGACCTCC	ATCCTCTCTCCAGTGCCT	TCCCAGTCCCTCCCCGTCC	CTCTCAATGTCACCCAGCC
BR0507.G7	CTCCCCGTGTCTCCATTT	TACCCACATCTCATCCAGT	GTCCCCACTCTCTCCAGTG	CCCCCAGCGTGTCCATCTC
Petr01-04			T.....
BR0507.G7	GCTGATGCCCTCGAGCTCCC	AGCCGATCCCGGCCACCTG	CTCTACTCCGTGATGGATT	CTACCCTGCCAAGATCCAGG
BR06Psd.1G.....
06AP/4G.....
Petr01-04	...G.....G.....G.....	...G.....C.....CT..GCC...G..

```

BR0507.G7 TGAGGTGGTTCCAGGGTCAG CAGGAGCTCTCAGAGCATGT GGTGACCAGCAACATTATCC CCAATGGGGACTGGACCTAC
BR06Psd.1 .....
06AP/4 .....
Petr01-04 .....C.....G.CC..C... ..G...C.G....G... ..C.....

BR0507.G7 CAGCTCCTGGTGCTGCTGGA AATCCCCTCAAGGCACGGAG TCAGGTACACCTGCCAGGTG GAGCACGTGAGCCTGGAGCA
BR06Psd.1 .....
06AP/4 .....*****
Petr01-04 .....G.....CG..GC.CCA..G...GC .....C.....

          ┌ Intron 3 ──┐
BR06Psd.1 CCCCTGAGCCGGCACTGGG GTACGCCCCAGTGGGACTGG GGGGACTAGGAGGGGAATTG GGATGTGCTAGGAGTAATTG
06AP/4 *****
Petr01-04 G.....C...AC.....

BR06Psd.1 GGAGGGAGTTTGAGGGAGTT AGGTGGTAACTGAGAGGAGA GTTTGGACAGCCTGGGAGGG CAGAGAAGATATTTAGAGGA
06AP/4 *****

BR06Psd.1 CCCTGGGAGGACTGGGATGC CATGAGGGGTTTCTGGTGG CTACTGGGAGGGAATTGGG GTGCTGGATGGGTTCTGTTG
06AP/4 *****

BR06Psd.1 GAGTTGGAGGCAGGTTGGGG ATCTAAAAGCAGGGGTGGGA GGAAGTCGAATGTGCTTCA GTCTATTGGCAGTAGTGTTG
06AP/4 *****

BR06Psd.1 GTGGGTCTAGTGGGACTGG GAAGGAAGTGGGAAAGGATT TGGAGGATCATCGGGGATA GGTATCAACTGAGAGGGCGT
06AP/4 *****

BR06Psd.1 TTCGGGTCTGGAGGGACGG GGAGGGGATTTGGGGGGTA CTAGGGAGGTTGGATGGGGT TAAGAAAGAGTCTCATGGAG
06AP/4 *****

BR06Psd.1 CCTGGGTGGGCTGGAGGCA ACTGGGAGGGTCCCTGGAGG GGCTTGGGGTGTCTGTGAGA GATTTTGGGGGTTCTGAC
06AP/4 *****

          ┌ Exon 4 ──┐
BR06Psd.1 CCCTGCGGACCCCCAGAGA TGCCACCAGATACTGCCCGG AGCAAGATGCTGACAGGGAT TGGGGGCTTTGTGGTGGGTT
06AP/4 *****
Petr01-04 ...C ...G..G..C..C....C .....CC....G.... C.....C..CT...C.

          ┌ Intron 4 ──┐
BR06Psd.1 TTGTCTTCTGCACTGGGG CTTGGCTTCTACCTGTGCAA GAAGGTGAGAGGGGTCCCA GGGATCGTGTCCCCCGGGG
06AP/4 *****
Petr01-04 .C.....G..... ..C....T.....C....

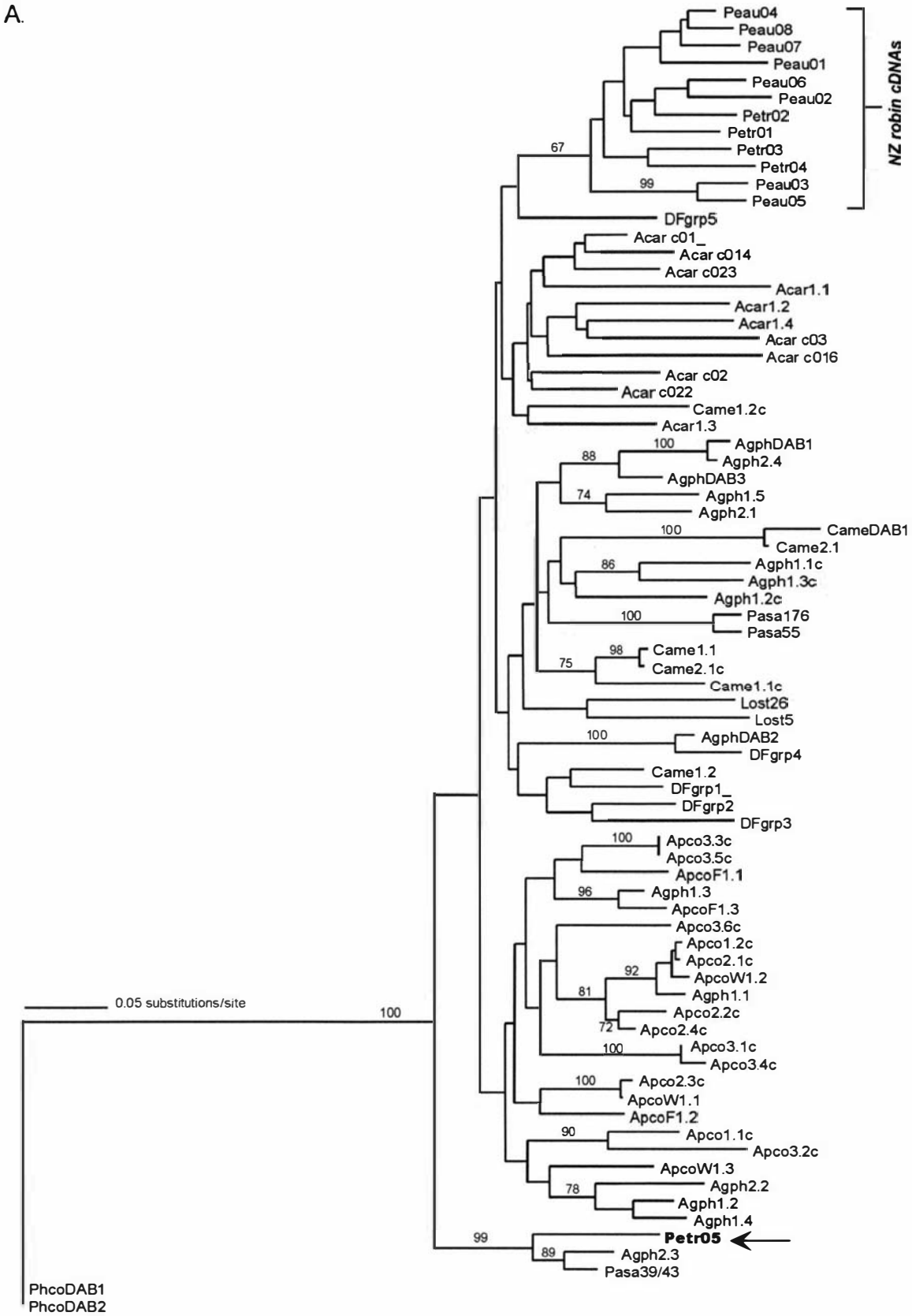
          ┌ Exon 5 ┌ 3'UTR ──┐
BR06Psd.1 CCCAGCCCGGTGTCACCT CTTTTCTCTGCCCGCAGAGC TCCTGATCCAGTGGCAGCTG CAGCCCTCTCTGTGACCCC
06AP/4 *****
Petr01-04 .....

BR06Psd.1 AGGCTGGCTGGCTACCCCT GCTCCATCCCCACGCTGATT TTGGGGATCCTGTGTCCC CCCCAGCCCTGCTGCTACT
06AP/4 .....

BR06Psd.1 CTGCCCTGCCCGG
06AP/4 .....CTGCTC CCAGTGTCCCAATAAAGCT TCCAGTTTTCCC

```

Figure 4.1 Sequence of the putative black robin pseudogene Petr05. The Petr05 sequence is a consensus of genomic clones BR0507.G7 and BR06Psd.1. The sequence of the corresponding cDNA clone 06AP/4 is also shown. The exon sequences are compared with the consensus sequence of the previously isolated class II B MHC cDNAs from black robin (Petr01-04). Dots indicate identity with the genomic DNA sequence, and asterisks indicate gaps. Intron and exon boundaries are marked according to the Agph-DAB1 gene (Edwards et al. 1998a), and the stop codon and transcription termination site are underlined.



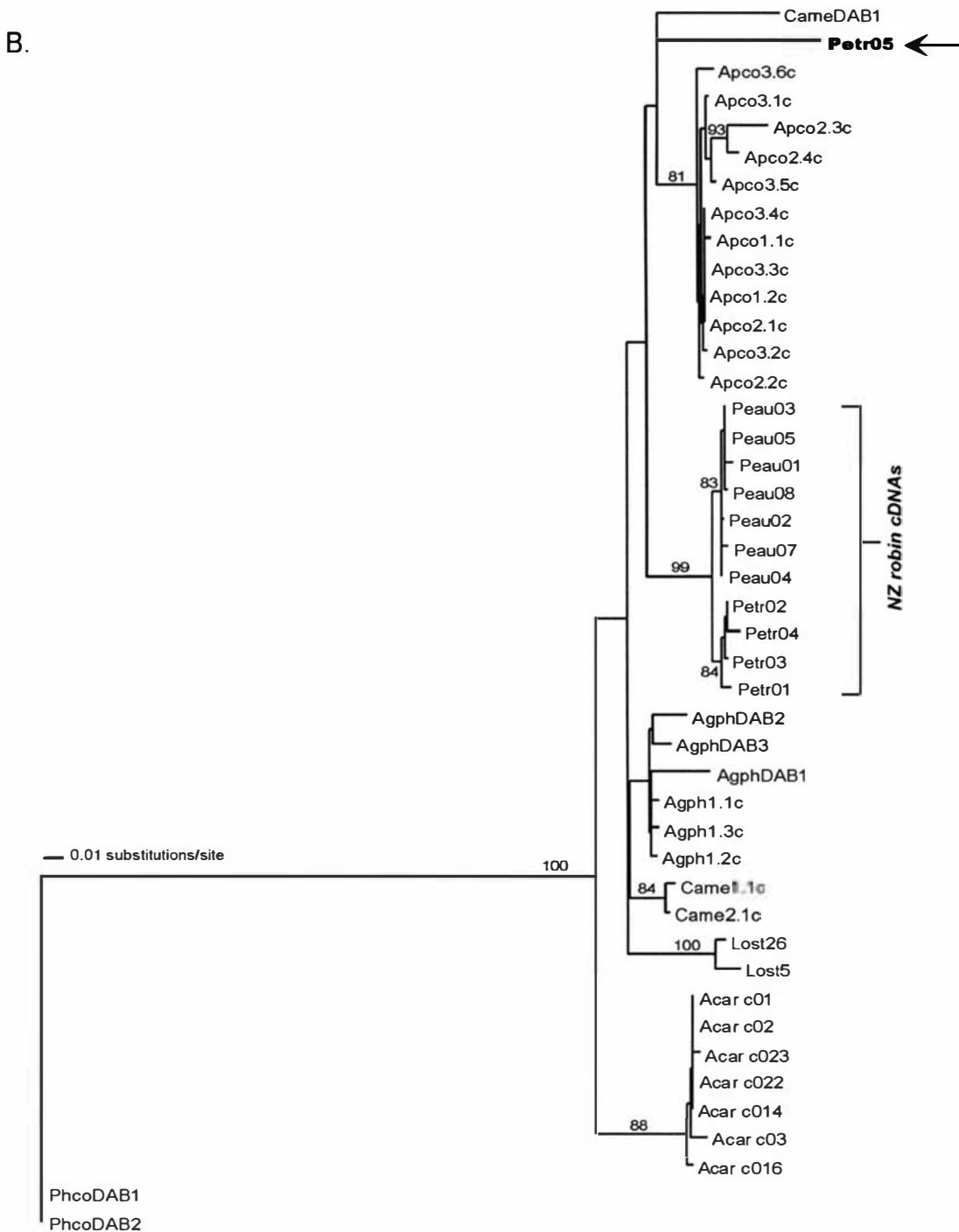


Figure 4.2. Neighbour-joining trees using Jukes-Cantor correction of exon 2 (A) and exon 3 (B) sequences, showing the relationship between the Petr05 sequence and other passerine sequences. Petr (black robin) and Peau (South Island robin) are cDNA sequences from chapter three. Nomenclature and references for other sequences are as follows: Acar (great reed warbler) Edwards et al. 1995a, Westerdahl et al. 2000; Agph (redwinged blackbird) Edwards et al. 2000, Edwards et al. 1998a, Edwards et al. 1995a, Edwards et al. 1995b, Gasper et al. 2001; Apco (Scrub Jay) Edwards et al. 1995a, Edwards et al. 1995b; Came (House finch) Edwards et al. 1995a, Edwards et al. 1995b, Hess et al. 2000; DF (Darwins finches) Sato et al. 2000, Sato et al. 2001; Lost (Bengalese finch) Vincek et al. 1995; Pasa (Savannah sparrow) Freeman-Gallant et al. 2002. Bootstrap values from 500 replicates are indicated where the value is greater than 70%. The trees are rooted with Pheasant (Phco) sequences (from Wittzell et al. 1999b). A is based on 157 bp of exon 2; B: 285 bp of exon 3.

4.4 References

- Brandle U, Ono H, Vincek V, Klein D, Golubic M, Grahovac B, Klein J (1992) Transspecies Evolution of Mhc-Drb Haplotype Polymorphism in Primates - Organization of Drb Genes in the Chimpanzee. *Immunogenetics* 36: 39-48
- Briles WE, Goto RM, Auffray C, Miller MM (1993) A polymorphic system related to but genetically independent of the chicken major histocompatibility complex. *Immunogenetics* 37: 408-414
- Edwards SV, Grahn M, Potts WK (1995a) Dynamics of *Mhc* evolution in birds and crocodylians: amplification of class II genes with degenerate primers. *Mol. Ecol.* 4: 719-729
- Edwards SV, Wakeland EK, Potts WK (1995b) Contrasting histories of avian and mammalian *Mhc* genes revealed by class II B sequences from songbirds. *Proc. Natl. Acad. Sci. USA* 92: 12200-12204
- Edwards SV, Gasper J, March M (1998a) Genomics and polymorphism of *Agph-DABI*, an MHC Class II B gene in red-winged blackbirds (*Agelaius phoeniceus*). *Mol. Biol. Evol.* 15: 236-250
- Edwards SV, Gasper J, Garrigan D, Martindale D, Koop BF (2000) A 39kb sequence around a blackbird *Mhc* Class II gene: Ghost of selection past and songbird genome architecture. *Mol. Biol. Evol.* 17: 1384-1395
- Freeman-Gallant CR, Johnson EM, Saponara F, Stanger M (2002) Variation at the major histocompatibility complex in Savannah sparrows. *Mol. Ecol.* 11: 1125-1130
- Gasper JS, Shiina T, Inoko H, Edwards SV (2001) Songbird genomics: Analysis of 45 kb upstream of a polymorphic Mhc class II gene in red-winged blackbirds (*Agelaius phoeniceus*). *Genomics* 75: 26-34
- Grimsley C, Mather KA, Ober C (1998) HLA-H: A pseudogene with increased variation due to balancing selection at neighboring loci. *Mol. Biol. Evol.* 15: 1581-1588
- Hess CM, Gasper J, Hoekstra HE, Hill CE, Edwards SV (2000) MHC Class II pseudogene and genomic signature of a 32 kb cosmid in the house finch (*Carpodacus mexicanus*). *Genome Research* 10: 613-623
- Hughes AL (1995) Origin and Evolution of Hla Class-I Pseudogenes. *Mol. Biol. Evol.* 12: 247-258
- Kaufman J, Milne S, Gobel TWF, Walker BA, Jacob JP, Auffray C, Zoorob R, Beck S (1999b) The chicken B locus is a minimal essential major histocompatibility complex. *Nature* 401: 923-925
- Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 17: 1244-1245
- MHC Sequencing Consortium (1999) Complete sequence and gene map of a human major histocompatibility complex. *Nature* 401: 921-923
- Nei M, Gu X, Sitnikova T (1997) Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proc. Natl. Acad. Sci. USA* 94: 7799-7806
- Renard C, Vaiman M, Chiannikulchai N, Cattolico L, Robert C, Chardon P (2001) Sequence of the pig major histocompatibility region containing the classical class I genes. *Immunogenetics* 53: 490-500
- Sato A, Figueroa F, Mayer WE, Grant PR, Grant BR, Klein J (2000) Mhc class II genes of Darwin's Finches: divergence by point mutations and reciprocal recombination. *In* M Kasahara (ed.): Major Histocompatibility Complex: Evolution, Structure and Function. Springer Verlag
- Sato A, Mayer WE, Tichy H, Grant PR, Grant BR, Klein J (2001) Evolution of MHC class II B genes in Darwin's finches and their closest relatives: birth of a new gene. *Immunogenetics* 53: 792-801

- Shum BP, Mason PM, Magor KE, Flodin LR, Stet RJM, Parham P (2002) Structures of two major histocompatibility complex class I genes of the rainbow trout (*Oncorhynchus mykiss*). *Immunogenetics* 54: 193-199
- Swofford DL (2002) PAUP* Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, MA
- Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acid Res.* 22: 4673-4680
- Vincek V, Klein D, Graser RT, Figueroa F, O'hUigin C, Klein J (1995) Molecular cloning of major histocompatibility complex class II B gene cDNA from the Bengalese finch *Lonchura striata*. *Immunogenetics* 42: 262-267
- Westerdahl H, Wittzell H, von Schantz T (2000) Mhc diversity in two passerine birds: no evidence for a minimal essential Mhc. *Immunogenetics* 52: 92-100
- Wittzell H, Bernot A, Auffray C, Zoorob R (1999b) Concerted evolution of two MHC Class II loci in pheasants and domestic chickens. *Mol. Biol. Evol.* 16: 479-490

CHAPTER FIVE

Population Bottlenecks and MHC Variation in New Zealand Robins

5.1 Introduction

In 1980, the entire Chatham Island black robin (*Petroica traversi*) species consisted of only five individuals, including only one successful breeding pair, making it one of the world's rarest birds (Butler and Merton 1992). Following intensive management, numbers have increased to approximately 250 individuals (H. Aikman, pers comm.) and populations have been established on South East (Rangatira) and Mangere Islands, part of the Chatham Islands group (Figure 5.1A). Theoretical models (Wright 1931, Nei et al. 1975) predict that small, bottlenecked populations will show reduced levels of genetic variation at neutral loci (e.g. Groombridge 2000, Wisely et al. 2002). In accordance with these predictions, extremely low levels of neutral minisatellite DNA variation have been reported in the black robin (Ardern and Lambert 1997). However, analysis of whether there has been a parallel loss of variation at genes with known fitness consequences is important in assessing the long-term viability of small populations. Such loci may be under selection, and be more likely to retain variation in small populations (Maruyama and Nei 1981, Nevo et al. 1997). In this chapter, levels of variation at Major Histocompatibility Complex (MHC) loci, which are thought to be influenced by balancing selection, are investigated in the black robin and its non-endangered relative, the South Island robin (*Petroica australis australis*).

MHC molecules are central to the vertebrate immune response, as they present short peptides, usually of bacterial or viral origin, to T cells. A hallmark of classical class I and class II MHC genes is their high polymorphism, particularly at residues involved in peptide binding and presentation (the peptide-binding region, PBR), and there is evidence that balancing selection plays a role in the maintenance of this diversity (Meyer and Thomson 2001). Balancing selection is thought to be largely driven by the diversity of pathogens encountered

by vertebrates, however there is also evidence that reproductive mechanisms such as disassortative mating may contribute to MHC diversity in some species (Apanius et al. 1997). Because of their central role in disease resistance, levels of variation at MHC genes may be particularly relevant for the conservation of endangered species, as it is thought that loss of variation may render a population susceptible to novel pathogens (O'Brien and Evermann 1988).

Although most natural populations of vertebrates have high levels of diversity at MHC genes (Klein 1986, Hedrick and Kim 1999), there are a number of examples of populations or species that exhibit little variation, including the cheetah *Aconyx jubatus* (Yuhki and O'Brien 1990), Scandinavian beavers *Castor fiber* (Ellegren et al. 1993), Swedish moose *Alces alces* (Ellegren et al. 1996), Musk Ox *Ovibos moschatus*, (Mikko et al. 1999), Arabian oryx *Oryx leucoryx* (Hedrick et al. 2000) the Malagasy giant jumping rat *Hypogeomys antimena* (Sommer et al. 2002) and a number of marine mammals (e.g. Trowsdale et al. 1989, Slade 1992, Murray and White 1998) (see chapter one, Table 1.2, for a full list). This indicates that balancing selection may not always be the predominant force shaping MHC variation and that behavioural and demographic factors may play an important role. In some cases, it appears that low MHC variation can be explained by reduced selection pressure due to solitary lifestyles (Ellegren et al. 1996, Hambuch and Lacey 2002) and decreased exposure to pathogens (Trowsdale et al. 1989, Slade 1992). It has also been suggested that monogamous mating systems may contribute to low MHC variability (Sommer et al. 2002). In many instances low MHC variation is associated with small population size or past population bottlenecks (e.g. Yuhki and O'Brien 1990, Hedrick et al. 2000). Although balancing selection may slow the loss of variation in small populations, the effect of genetic drift at times appears to override that of selection (Boyce et al. 1997, Seddon and Baverstock 1999, Hedrick et al. 2001a).

In this chapter, the effect of population bottlenecks on MHC diversity is measured by comparing levels of MHC variation in the black robin with artificially bottlenecked populations of South Island robin, and their respective source populations. To investigate whether MHC variation correlates with neutral genetic variation through population bottlenecks, MHC data is compared with minisatellite DNA data, which is available for all populations (Arden and Lambert 1997, Arden et al. 1997). The South Island robin is closely related to the black robin, and thus shares similar behavioural characteristics and

occupies a similar type of habitat. In 1973 two new populations of South Island robin were deliberately established on offshore islands. Five individuals from Kaikoura and Nukuwaiata Island were transferred to Allports Island and Motuara Island, respectively (Figure 5.1B). The Motuara Island population appears to have originated from only a single pair (Flack 1974), mirroring the 1980 bottleneck of the black robin. There is an important distinction, however, as although the black robin was widespread throughout the Chatham Islands prior to European settlement in the 19th century, it had been reduced to a single, small population of approximately 20 – 30 individuals on Little Mangere Island for approximately 100 years prior to the extreme bottleneck in 1980 (Butler and Merton 1992). The South Island robin bottlenecks, however, represent a sudden contraction in population size, followed by a rapid increase in population size after the founder event (Flack 1978). This enables the effect of two types of bottleneck (long vs short) to be compared.

Here, I have used restriction fragment length polymorphism (RFLP) and PBR sequences to analyse variation in class II B MHC genes in these populations. Extensive RFLP variation has been reported in many natural populations of birds (Wittzell et al. 1999, Freeman-Gallant et al. 2002), however data on sequence variation in wild populations is lacking compared to that available for mammalian populations. This is likely to be due to the inherent difficulties in studying the avian MHC, as unlike the situation for mammals, the avian MHC (excluding that of chicken) appears to be characterised by multiple closely related loci (Hess and Edwards 2002, and chapter three, this study). Recent duplication events and/or homogenisation of loci outside the PBR by concerted evolution can obscure orthologous relationships and make the design of single locus primers difficult. Analysis of cDNA sequences from black robin and South Island robins suggests there are at least four expressed class II B loci, which have evolved by a mixture of gene duplication, gene conversion and balancing selection (chapter three), plus a potential pseudogene (chapter four). Prior to analysis of PBR variation, the sequences of introns flanking the PBR of these loci were characterised, in order to assess the feasibility of designing single locus primers for genotyping class II B MHC alleles in New Zealand robins.

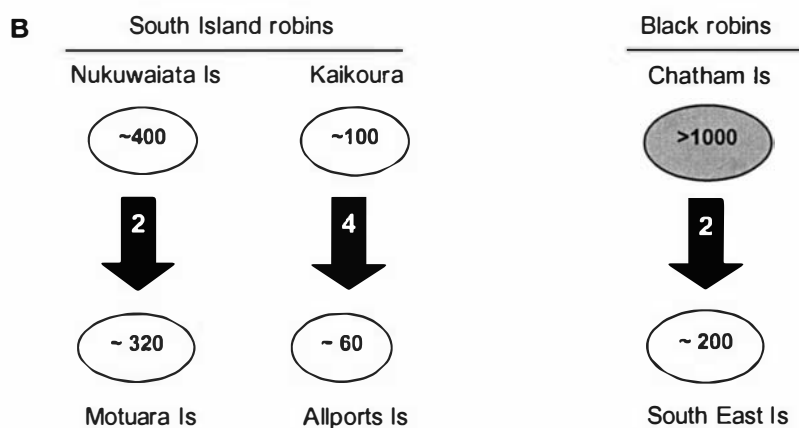
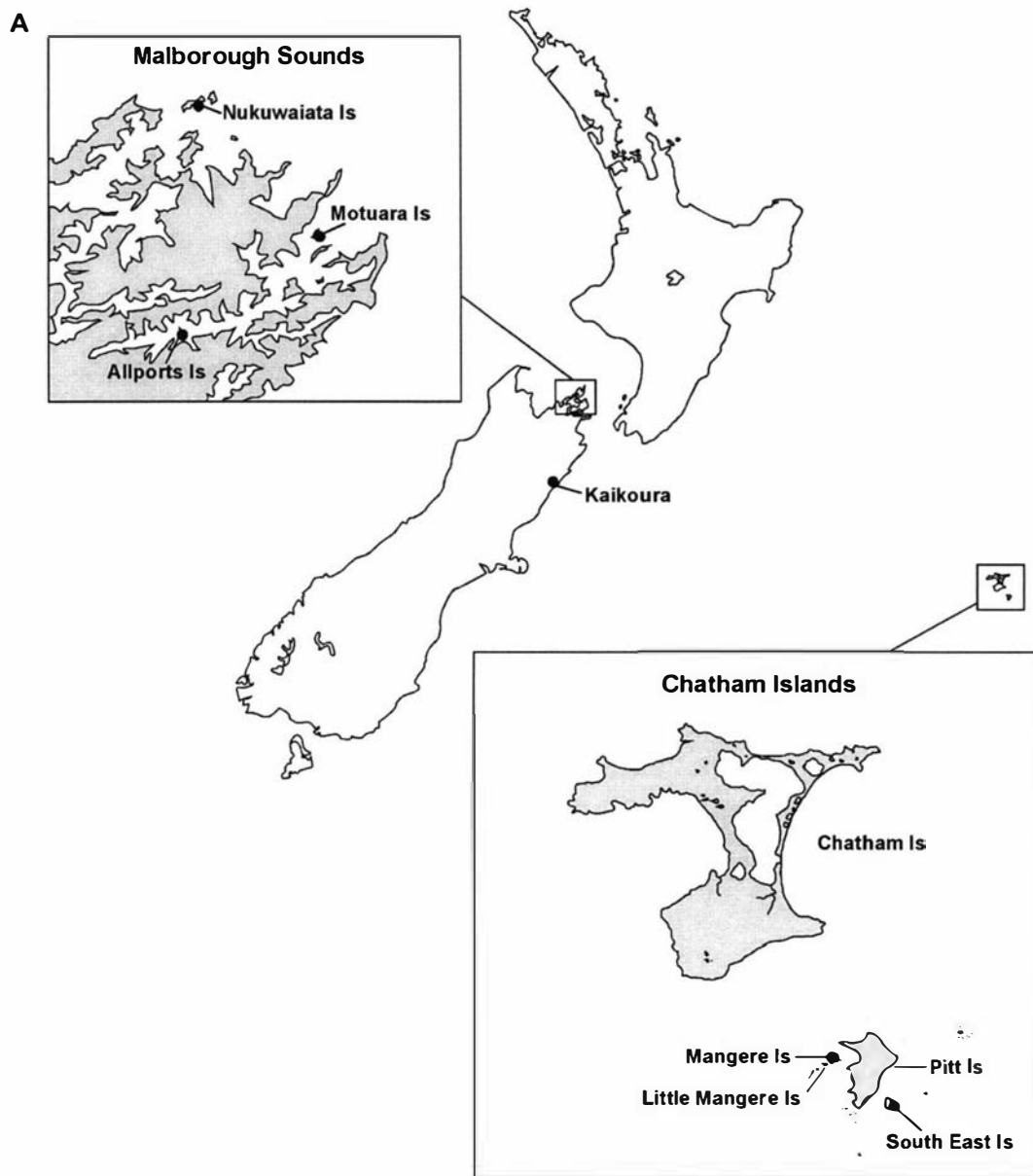


Figure 5.1 A. Map of New Zealand and the Chatham Islands, showing the location of robin populations sampled in this study. B. Schematic representation of populations, showing approximate population sizes at the time samples were collected (circles) and the number of individuals thought to have given rise to bottlenecked populations (arrows). Samples from the ancestral black robin population (grey) were not available.

5.2 Materials and Methods

5.2.1 Birds and sample collection

Black robin blood samples were collected from South East Island and Mangere Island, in the Chatham Islands group, as described in Ardem et al. (1994). Samples were collected from Mangere Island in February 1993, and South East Island in March-April 1992, January 2001 and February 2002. The parents and grandparents of all individuals used were checked on the black robin genealogy database (NZ Dept of Conservation, unpublished data) to ensure that the sample set contained no first or second order relatives. All blood samples from South Island robins were collected during 1992 and 1993, from the source populations on Nukuwaiata Island and Kaikoura, and their respective bottlenecked populations on Motuara Island and Allports Island, (Figure 5.1 and Ardem et al. 1997).

5.2.2 DNA techniques

Analysis of MHC variation by RFLP analysis, using the restriction enzymes *PvuII* and *TaqI* and the BR6 probe, was conducted on eight black robins (all South East Island), and eight South Island robins from each source and bottlenecked population. Methods for DNA extraction, probe construction, Southern blotting and hybridisation were described in chapter three. Peptide-binding region (PBR) sequence analysis was conducted on ten black robins ($n = 8$ from South East Island and $n = 2$ from Mangere Is), and eight South Island robins each from the Nukuwaiata Is and Motuara Is populations.

In order to isolate sequences spanning the entire peptide-binding region (exon 2) it was necessary to design PCR primers to the flanking introns. To isolate these intron sequences primers were designed to cDNA sequences from chapter three to amplify across introns 1 and 2. The cDNA sequences were aligned and conserved regions within exons 1, 2, and 3 were chosen as primer sites (Figure 5.2). Amplification of introns was performed in a BioRad iCycler for 30 cycles, using the Expand *HiFi* PCR system containing 200 μM each dNTP, 0.4 μM each primer, and 1 μL of genomic DNA in a 25 μL reaction volume. For amplification of intron 1, 1 mM MgCl_2 was added and an annealing temperature of 60°C was used. For intron 2, 1.5 mM MgCl_2 was added and an annealing temperature of 58°C was used. Both PCR amplifications produced multiple products between 500 bp and 1 kb for intron 1, and

between 900 bp and 2 kb for intron 2: all were excised from an agarose gel and purified using a HighPure PCR product purification kit (Roche), then cloned into a pGEM-T Easy vector. Positive clones (8-15 per PCR) were sequenced using the BigDye Terminator Cycle Sequencing kit (Applied Biosystems) and analysed on an ABI 377A automated sequencer. The sequences obtained using these primers contained the entire intron and part of the flanking exons, which allowed the intron sequences to be matched with the previously isolated cDNA sequences.

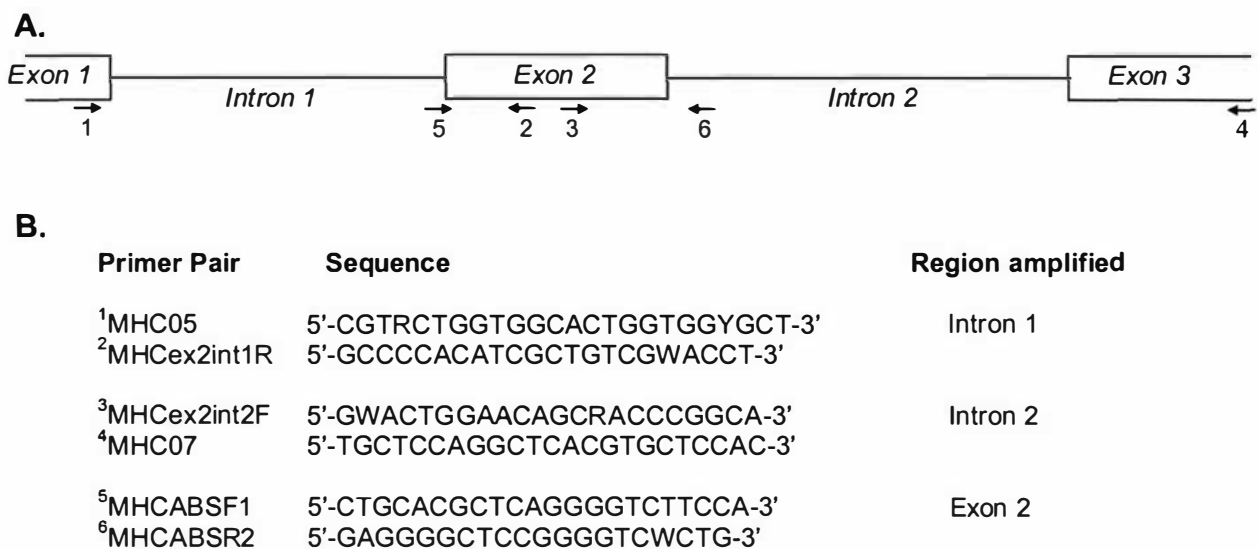


Figure 5.2. Position of PCR primers used in this study (A), with corresponding primer names and sequences given in B.

Intron sequences were aligned in Clustal W (Thompson et al. 1994), and conserved regions immediately flanking exon 2 were chosen as primer sites for amplification of PBR sequences. Because these flanking regions appear to be highly conserved across all expressed loci in both the black robin and South Island robin (see results), it was not possible to design single locus primers. Hence the PBR primers MHCABSF1 and MHCABSR2 were designed to amplify all expressed sequences in a single PCR, providing a “multilocus genotype” similar to that obtained by RFLP analysis. PCR was carried out in a 15 μ L reaction using the Expand *HiFi* PCR system (Roche) with 1.5 mM MgCl₂, 1 M Betaine (Sigma), 200 μ M each dNTP, 0.4 μ M each primer, and 1 μ L of genomic DNA. Thermal cycling was performed in a BioRad iCycler for 25 cycles consisting of 95°C for 30 sec, 61°C for 20 sec, and 72°C for 30 sec. Two PCRs were performed per individual and the products were pooled before

purification using a HighPure PCR product purification kit (Roche). Purified PCR products were then cloned into pGEM-T Easy vector (Promega).

Eighteen clones per individual were analysed for MHC variation using single-strand conformation polymorphism (SSCP) analysis (Orita et al. 1989) prior to sequencing. Use of SSCP made it possible to screen a large number of clones per individual, and avoided the sequencing of multiple clones from one PCR with the same MHC sequence. Products for SSCP analysis were generated by PCR amplification of the insert directly from the colonies using M13 primers. M13 PCR products were then diluted 1:5, and 0.5 μ L was used in a nested PCR using primers BRABSF1 and BRABSR2. For SSCP analysis, 2 μ L of the nested PCR product was added to 3 μ L of loading dye (96% formamide, 0.05% bromophenol blue, 20 mM EDTA) denatured for 7 min at 100°C, then chilled on ice before loading on a 9% polyacrylamide (37.5:1 acrylamide:bisacrylamide) gel with 2.5% glycerol.

Electrophoresis was carried out in 0.5 X TBE at 12W for 6 hours at 4°C, using a Whatman Biometra[®] V16-2 Vertical Gel Electrophoresis apparatus. Gels were then stained with SYBR[®] Gold (Molecular Probes) for 15-30 min and visualised over UV light. From each gel one representative of each SSCP banding pattern was chosen for sequencing. Plasmids were prepared from the corresponding colony using a HighPure plasmid purification kit (Roche), and sequencing performed as described above. In order to distinguish PCR artifacts (i.e. PCR recombinants and *Taq* errors) from real alleles, this entire process (PCR, cloning, SSCP and sequencing) was repeated for each individual and only sequences that were found in independent PCRs were included in the analysis.

5.2.3 Data analysis

Sequences were edited, aligned and translated using Sequencher[™] 4.1.2 (Gene Codes Corporation). Average Percent Difference (APD) was calculated for each population as an index of within-population genetic variation. This was calculated for both RFLP profiles, where it is defined as the average percentage of restriction fragments that differ between individuals, and the sequence data, where it is the average percentage of sequences that differ between individuals. APD values were calculated in Excel according to Yuhki and O'Brien (1990). For RFLP profiles, the mean APD is the average of APD values for both enzymes. The difference in APD values between source and bottlenecked populations of South Island robins was evaluated with the Mann-Whitney *U* test. The correlation between minisatellite

and MHC values was calculated in Excel using the Pearson product-moment correlation coefficient. Calculation of pairwise amino acid differences, and non-synonymous (d_N) to synonymous (d_S) nucleotide changes were performed using *MEGA* version 2.1 (Kumar et al. 2001). The Nei and Gojobori (1986) method with Jukes-Cantor correction was used to calculate d_N and d_S , and a Z-test of selection was performed to test whether $d_N > d_S$.

5.3 Results

5.3.1 RFLP analysis

Identical RFLP patterns (APD = 0) were observed in eight black robin individuals for both enzymes (see Fig 3.2). All four South Island robin populations had substantially higher levels of diversity with mean APD values ranging from 19.91% to 43.20% (Table 5.1). Bottlenecked populations of South Island robin (Allports and Motuara) had significantly lower APD values for each enzyme than their respective source populations (Kaikoura and Nukuwaiata), however these bottlenecked populations still retained moderate levels of diversity.

Table 5.1 Average Percent Difference (APD) values for each enzyme with the BR6 probe. n = number of individuals analysed. * $p < 0.05$, ** $p < 0.001$, Mann Whitney U test.

Population	n	APD <i>PvuII</i>	APD <i>TaqI</i>	Mean APD
Black robin	8	0.00	0.00	0.00
South Island robin (Kaikoura)	8	32.93	43.72	38.32
South Island robin (Allports)	8	20.96*	18.86**	19.91
South Island robin (Nukuwaiata)	8	37.38	49.03	43.20
South Island robin (Motuara)	8	30.02	31.42**	30.72

5.3.2 Isolation of intron sequences and design of PBR primers

The class II B MHC cDNA sequences previously isolated from black robin and South Island robin (see chapter three) were used to isolate intron sequences, in order to design primers for amplifying the full length of exon 2. Introns were amplified from both black robin and South

Island robin using the primer pairs MHC05 + MHCex2int1R (intron 1) and MHCex2int2F + MHC07 (intron 2), and the resulting sequences were matched with their corresponding cDNA sequences by aligning the overlapping part of exon 2. The introns of the cDNA sequences appeared to be characterised by repeated motifs, and differ considerably in length (from 411 to 591 bp for intron 1, and 666 to 1460 bp for intron 2), largely because of the varying number of imperfect repeats. Intron sequences could not be amplified for every previously isolated cDNA, however across all the sequences that were isolated the regions flanking exon 2 are highly conserved (Figure 5.3). The exception to this is the intron sequences of the putative pseudogene Petr05, which diverge soon after the intron/exon boundary, and differ considerably in length (739 bp for intron 1, and 515 bp for intron 2).

It was necessary to place PCR primers for amplifying exon 2 sequences in the regions immediately flanking exon 2, as this would produce a small enough product to enable indirect methods of mutation detection (SSCP) to be used, while still allowing the entire PBR to be amplified. It has been suggested that intron sequences may be better indicators of loci in MHC genes than coding sequences (Sato et al. 2000). However in this study, orthologous and paralogous loci could not be distinguished from the intron sequences, preventing the design of single locus primers. For instance, the sequences Petr04 and Peau01/Peau05 were identified as putative orthologs in chapter three on the basis of their 3' untranslated region (UTR) sequence, but their intron sequences do not show clear differences from those of the other, paralogous sequences, especially in the region immediately flanking exon 2. Similarly, the lengths of the introns do not reflect the orthologous relationships identified in chapter three (see Figure 5.3). For these reasons, the PBR primers MHCABSF1 and MHCABSR2 were designed to amplify all transcribed MHC sequences. Amplification of the Petr05 sequence was excluded however, by a mismatch at the 3' end of MHCABSR2.

5.3.3 Peptide binding region sequences

PBR sequences were amplified with MHCABSF1 and MHCABSR2, and the resulting products were cloned, screened using SSCP, and sequenced. A total of 36 clones per individual from two independent PCR amplifications (each one consisting of products from two pooled PCR reactions) were screened by SSCP, and a total of 5-13 clones per individual for black robin, and 20-28 clones per individual for South Island robins were sequenced. Four different sequences were identified in the 10 black robin individuals, and a total of 41

sequences were found in the Nukuwaiata Is and bottlenecked Motuara Is populations of South Island robin (Table 5.2). The black robin sequences are identical to the previously isolated cDNA sequences so have been given the same number, while the South Island robin sequences are numbered in order of isolation. We did not attempt to assign alleles to loci on the basis of exon 2 sequences alone due to the likelihood of gene conversion and recent gene duplications (see chapter three). For simplicity, different sequences are referred to as alleles even though they may be from different loci. The alignment of PBR nucleotide sequences is given in Appendix A.

All sequences included in this analysis were verified by amplification from more than one independent PCR. In addition, a number of single clones which appeared to contain PCR artifacts were identified. *In vitro* recombination and heteroduplex formation are known to occur at detectable frequencies when PCR amplification of a mixture of sequences is performed (Ennis et al. 1990, Zylstra et al. 1998). Out of a total of 936 clones screened by SSCP (452 sequenced) in this study, 10 clones resulting from heteroduplex formation were identified, and 34 chimeric clones resulting from PCR recombination were found. Chimeras were only found in South Island robin clones (representing 5.9% of the total number of South Island robin clones screened) and in all cases, the two “donor” sequences could be identified among previously verified sequences. In addition, approximately 9% of clones screened appeared to contain *Taq* errors (*Taq* error rate of 0.36 per 1000 bp).

The lack of variation seen in the black robin RFLP profiles is also apparent in the sequence data. The alleles Petr01, Petr02, and Petr03 appear to be fixed in the population, as they were found in all black robin individuals sampled, while the allele Petr04 was found in half the individuals sampled, resulting in an overall APD value of 7.65%. From this it appears that there is some MHC variation in the black robin population, however it is likely that Petr04 was simply not amplified as successfully as the other alleles, so was not sampled in all individuals. The cDNA sequence of Petr04 contains considerable differences in the 3'UTR, indicating that it is a separate locus from the other sequences, which lends weight to this suggestion.

In contrast to the black robin, the South Island robin populations had high levels of MHC variation. However there was significantly less diversity in the bottlenecked Motuara Island population. Thirty-five alleles were found in the Nukuwaiata Island population, while only

28 alleles were present in Motuara Island individuals, and this is reflected in a significant decrease in APD value (65.5% vs 49.02%). However, there was no difference in the number of alleles per individual. APD values for both MHC sequence and RFLP data were highly positively correlated with those obtained from minisatellite DNA from the same populations ($r = 0.946$, $p < 0.05$ for RFLP data, $r = 0.991$, $p < 0.05$ for sequence data, see Figure 5.5). The proportion of genetic variation lost in the bottlenecked populations compared with the source populations was similar for both minisatellite DNA and MHC, except in the Allports Island population, which appeared to lose more variation at MHC loci than in minisatellite DNA (Table 5.4).

The exon 2 DNA sequences were translated into amino acids and the alignment is shown in figure 5.4. Once the PCR primers are removed, the sequences span 86 out of 89 amino acids of exon 2. No stop codons or indels were found, and all alleles contain the conserved amino acids involved in glycosylation, disulfide bond formation and salt bridges (according to Kaufman et al. 1994). Some alleles have changes in the conserved residues involved in peptide binding, however in all cases the replacement amino acid has the same properties as the conserved residue. Most alleles were highly divergent from one another, with the number of amino acid differences between alleles ranging from 19-25 for black robin alleles (average 21.5), and 0-28 for South Island robin alleles (average 17). Alleles Peau-5 and Peau-6 have the same amino acid sequence, and differ by two silent nucleotide substitutions. There are also a number of South Island robin alleles that differ by only one or two amino acids e.g. Peau-14, Peau-15 and Peau-16. The amino acid variation is largely confined to the polymorphic subdomains (BS1, BS2, BS3 and the α -helix) identified in She et al. (1991). Almost all the South Island robin alleles contain sequence motifs that are present in other alleles (e.g. boxed sequences Figure 5.4), and many alleles appear to have arisen by shuffling of these motifs. Of the 24 sites known to contact the peptide in the human DR β molecule (Brown et al. 1993), 20 (83%) contain non-synonymous changes, compared with 43.5% of non peptide-binding (PB) codons. This is reflected in the higher values of d_N for PB codons than non-PB codons for all alleles (Table 5.3). Among the South Island robin alleles, the ratio of d_N/d_S was higher among PB codons than for the remainder of the exon. However, d_N/d_S ratios greater than one were also found for non-PB codons (although this is marginally non-significant for SIR), and for black robin alleles the ratio for non-PB codons was actually higher than for PB codons.

A

```

Petr01 (531) G*****AGAGGATGTGA AAGGGGAAATGGCGA***** *****GAA GGGCGGGAAGGAGGAAGCGG CAGCGCGGGCAGGAGGGT*
Petr02 (590) -----
Petr03 (591) -----
Petr04 (465) -----
Peau01 (493) ---AGGGG-----
Peau02 (411) -----
Peau03 (539) -GGAGGGG-----
Peau05 (521) ---AGGGG-----
Peau07 (514) -G-AGGGG-----
Peau09 (412) -----
Petr05 (739) -GACCCTCA-AC-----G GG-T-TTGC-----TGGGG GAGCTATGGGAAAGGGA--- --T--A--A----G-TTA TG-GA-CAC--CAAG----G
    
```

```

Petr01 *CCC*CGGGGCTCAGCGGGG CCCGCGCGGGATCCGGTGG GCGCC*CGCAGCTGTGGG** *****GGTGTCTGGGG** *****GGCACCCCTGA
Petr02 -----G-G-----G-----G-----C-----
Petr03 -----G-G-----G-----C-----
Petr04 -----G-G-----G-----
Peau01 -----G-G-----G-----T-----A-----
Peau02 -----G-GA-----G-----
Peau03 -----G-G-----G-----A-----
Peau05 -----G-G-----G-----
Peau07 -----G-G-----G-----T-----
Peau09 -----G-G-----G-----
Petr05 G-A--T---A--G-A-T G-T-A-T-G--G---G-T ---T-TG---T--AT TTCTGTGG---CT GCTGGAGG---
    
```

Intron 1 → Exon 2

```

Petr01 CCTGGGTCCTGCACGCTCAG GGGTCTTCCAGTGGATGTTTC 21 AAGGGCGAGTGTCACTTCAT 41 TAACGGCACGGAGAAGGTGA 61 GGTACGTGGTGAGGAAGTTC
Petr02 -----T---G-A---TT-A---G-----TG---A---C-----
Petr03 -----GA---GGA---AC-----TG---AC---T---A---
Petr04 -----T---GGA---TT-----G-----C---C---C-----
Peau01 -----GA---GCA---TC-----C---T---T---AC---T---
Peau02 -----A-----GAT---A---C---T-----TG-----
Peau03 -----A-----A---GGA---TC-----TG---AC---C-----
Peau05 -----T-----GA---G---G-----TG---CA---A---T---
Peau07 -----A-----GA---G---G---C-----T-----
Peau09 -----A-----GA---AGGA---T-----TG---A---C-----
Petr05 ---T-----A-----GT---G---A---T-----A-----C---C---GA---T---AA---TT---A---
    
```

B

```

Petr02 (990) 190 TGGAGCAGAGACGGGCAGAG 210 GTGGACACGGTCTGCCGGA 230 CAACTACGAGGTCGACACC 250 CGTTCAGCGTGAGCGCCGA Intron 2
Petr03 (1800) -----C-----G-TA-----C-----A---GT-----
Petr04 (1033) -----C-----A-C-C-----G-TA-----C-----A-T-C-----CT---G-----
Peau01 (881) -----GC-A-----G-TA-----C-----T---G-----
Peau02 (808) -----GC-----TA-----CG-----C-T-C-----CT---T-----
Peau03 (889) -----T-CCTT---AAT-C-----T-----C-----A-T-T-----
Peau06 (887) -----GC-A-----GTT-----C-----
Peau08 (666) -----GC-A-A-----GTC-----CT---T-----
Peau09 (744) -----GC-A-----C-----GTT-----CT-----
Petr05 (515) -----T-C-----T-C-----G-CA-----C-----A---T-GTC---G---CC---GT---A---T-----*
    
```

```

Petr02 AGGCCCTCGGGCCCTGCC CGGCAGAGACCCCGGAGCCC CTCGGAGCCTCGCTGGGAAT CGCCCCAGAGCCCTCAGCCC TCCTGGACCCGTCCCGGG
Petr03 -----G-----T-----
Petr04 -----G-----
Peau01 -----T-----C-----A---
Peau02 -----T---T---G-----
Peau03 -----T-----C---T---
Peau06 -----T---T---T-----G-----
Peau08 -----T---T-----T---T---AT-C-*GT-C---A-TC
Peau09 -----T---T---G-----
Petr05 **-T-----TA-G-AT---T-T-----*AA-A---C---TA-----C-T---TT-----TG---ACG---C-T-
    
```

```

Petr02 CCTTTC*TGTCATCCAGT CCCTCCAGTCCCT****** *****
Petr03 -----C-----T-----CCCTCCAGTCCCTCCAG-----CCAGT CCCTCCAGTTCATCCAGT
Petr04 -----A-----T-----CCATCCAGTCCCTCCAG-----*-----
Peau01 -----G-----G-T-----CCCTCCAGTCCATCCCGG-----T-A-----
Peau02 -----*** *****C-----
Peau03 G-C--TC-----*** *****G---C-----A-----
Peau06 -----A-----G-----A---A-----
Peau08 -A-CC-A---C---*** *****-MC---A---A---G---T-----
Peau09 -----*** *****G-----
Petr05 G-CCCT--C---C---G** *****T---*G---GA-----AG -T-C---T---T-----
    
```

Figure 5.3. Alignment of intron sequences flanking exon 2, showing (A) the last 220 bp of intron 1 and first 80 bp of exon 2, and (B) the last 80 bp of exon 2 with the first 220 bp of intron 2. The length of the full intron is given in brackets at the start of each section. Dashes indicate identity with the reference sequence, asterisks indicate gaps. The numbering of exon 2 nucleotides is shown and sequences are named according to the cDNA sequence from chapter three. An additional sequence not included in chapter three was also identified (Peau09). Petr05 is the putative pseudogene identified in black robin (chapter four). The shaded blocks show the position of PCR primers MHCABSF1 and MHCABSR2. A repeated motif identified in intron 2 is shown in the box.

Allele	Black robin	South Island robin	
		Nukuwaiata Is	Motuara Is
Petr01	1.000	-	-
Petr02	1.000	-	-
Petr03	1.000	-	-
Petr04	0.500	-	-
Peau-1	-	0.375	0.500
Peau-2	-	0.250	-
Peau-3	-	0.500	0.375
Peau-4	-	0.500	0.625
Peau-5	-	0.875	0.875
Peau-6	-	0.375	0.125
Peau-7	-	-	0.750
Peau-8	-	0.625	0.875
Peau-9	-	0.500	-
Peau-10	-	-	0.250
Peau-11	-	0.250	-
Peau-12	-	0.250	-
Peau-13	-	0.125	-
Peau-14	-	0.375	0.125
Peau-15	-	0.125	0.375
Peau-16	-	-	0.750
Peau-17	-	0.250	0.375
Peau-18	-	0.750	0.750
Peau-19	-	0.250	0.250
Peau-20	-	0.250	0.125
Peau-21	-	0.375	-
Peau-22	-	0.375	0.375
Peau-23	-	0.375	-
Peau-24	-	0.500	0.125
Peau-25	-	0.125	-
Peau-26	-	0.250	0.250
Peau-27	-	0.500	-
Peau-28	-	0.125	0.125
Peau-29	-	0.375	0.500
Peau-30	-	0.250	-
Peau-31	-	-	0.500
Peau-32	-	0.125	0.250
Peau-33	-	-	0.750
Peau-34	-	0.250	-
Peau-35	-	0.125	-
Peau-36	-	-	0.750
Peau-37	-	0.125	-
Peau-38	-	0.125	0.375
Peau-39	-	0.125	0.125
Peau-40	-	0.375	0.125
Peau-41	-	0.250	0.125
APD	7.65	65.50	49.02*
Alleles /individual	3.5 (3-4)	11.4 (8-14)	11.5 (9-14)

Table 5.2. Distribution and frequency of class IIB MHC alleles in NZ robins. The proportion of individuals in each population who have each allele is given, along with the APD value for each population. The average number of alleles per individual is shown, along with the range (in brackets). * $p < 0.005$, Mann Whitney U test.

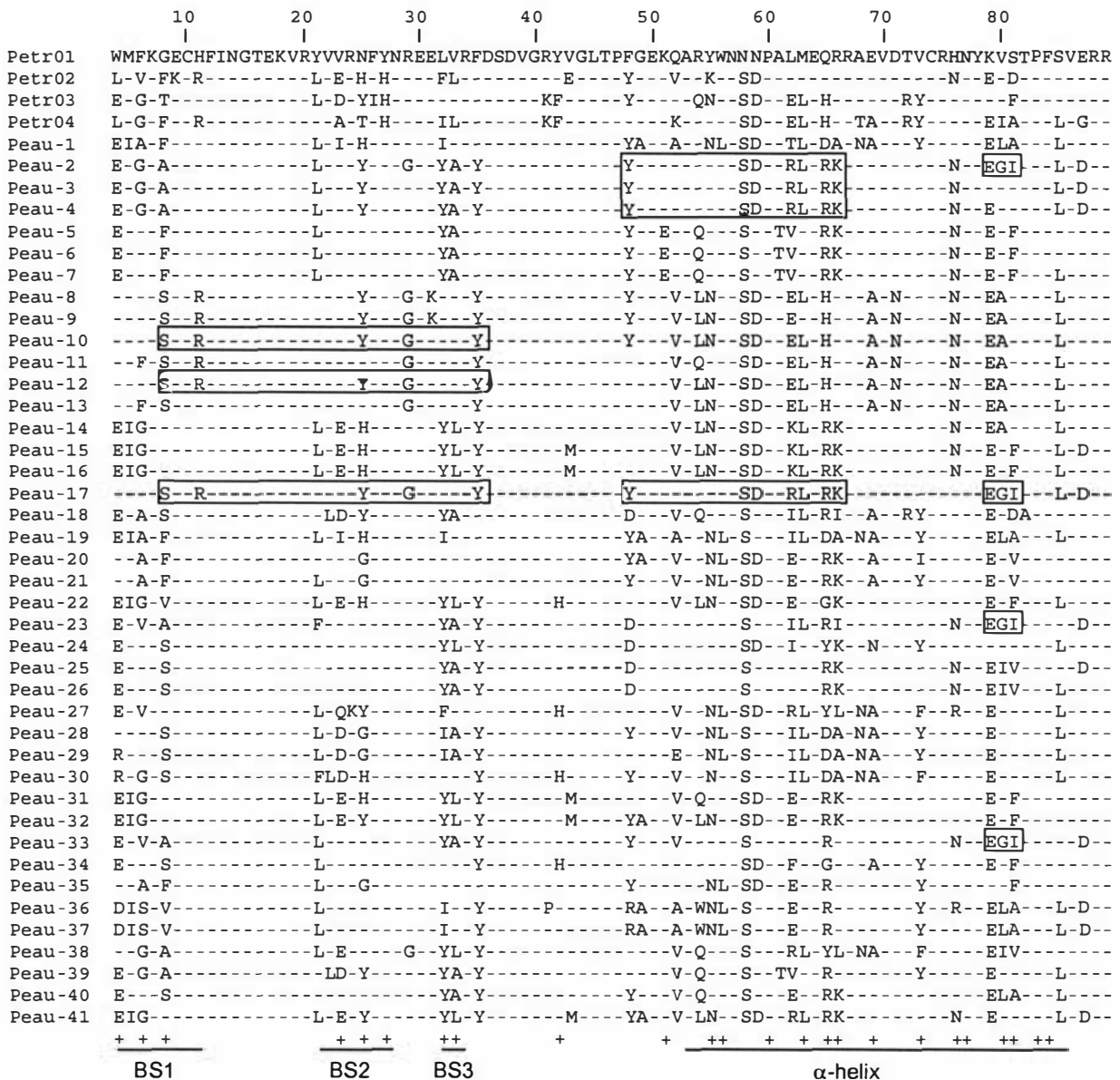


Figure 5.4. Amino acid sequences of class II B MHC alleles from black robin (Petr), and South Island robin (Peau). Identity with the Petr01 sequence is shown with dashes. Amino acids that contact the peptide in the human DR β molecule (Brown et al. 1993) are indicated with a plus sign, and polymorphic subdomains recognised by She et al. (1991) are underlined (BS = Beta sheet). Examples of common sequence motifs are shown in boxes.

Table 5.3. Nonsynonymous (d_N) and synonymous (d_S) substitution values (\pm SE) for class II B sequences from black robin (BR) and South Island robin (SIR). PB = amino acids thought to be involved in peptide binding. P indicates the probability that $d_N > d_S$.

Sequences		d_N	d_S	d_N/d_S	p
Exon 2 PB	BR	0.373 \pm 0.086	0.135 \pm 0.078	2.76	< 0.05
	SIR	0.339 \pm 0.078	0.094 \pm 0.059	3.6	< 0.005
Exon 2 non-PB	BR	0.099 \pm 0.024	0.020 \pm 0.014	5.05	< 0.005
	SIR	0.076 \pm 0.019	0.039 \pm 0.013	1.95	0.055

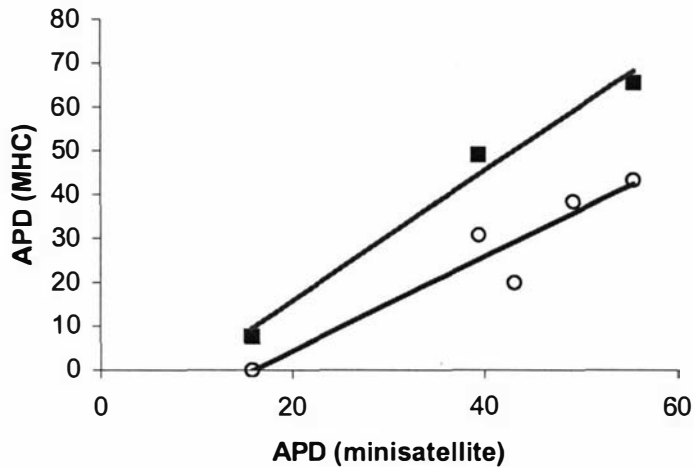


Figure 5.5. Plot of MHC diversity vs minisatellite DNA diversity in the populations measured in this study. Filled squares, MHC sequence diversity; open circles, MHC RFLP diversity. Minisatellite DNA APD values are from Ardern and Lambert (1997), and Ardern et al. (1997).

Table 5.4. Percentage of genetic variation lost in bottlenecked populations compared with their source populations. Minisatellite DNA data is from Ardern et al. (1997).

	Minisat	MHC (RFLP)	MHC (Sequence)
Allports Is	13%	50%	-
Motuara Is	28.9%	28.9%	25%

5.4 Discussion

In this chapter, the loss of MHC diversity through population bottlenecks was investigated. MHC variation was measured using both RFLP and sequence analysis of exon 2, in the highly endangered black robin, and in source and bottlenecked populations of South Island robin. Both measures suggest that the black robin is monomorphic at class II B loci, whereas both source and bottlenecked South Island robin populations have substantial levels of variation. South Island robin APD values are comparable to those reported in outbred mammalian and reptile populations (e.g. Plante et al. 1991, Madsen et al. 2000). High levels of MHC RFLP variation have also been documented in other natural populations of songbirds, including Savannah sparrows (Freeman-Gallant et al. 2002) and Great reed warblers (Wittzell et al. 1999).

APD values for the PBR sequence data are consistently higher than the RFLP values, indicating that RFLP analysis underestimates the actual levels of MHC diversity. Hybridising

fragments on RFLP profiles may include non-polymorphic pseudogenes and gene fragments, which will reduce the APD values. In addition, differences between alleles of only a few base pairs, as was found between some South Island robin alleles in this study, may not be detected using RFLP. The sequence data are likely to be a more accurate reflection of the true levels of MHC diversity, however there may still be inaccuracies introduced by the stochastic nature of the PCR reaction. Pooling of multiple PCR reactions, the use of Betaine (which can equalise the amplification efficiencies of diverse target sequences), and a low cycle number were employed in this study in an attempt to minimize PCR artifacts (Wagner et al. 1994, Weissensteiner and Lanchbury 1996, Zylstra et al. 1998). However, PCR recombination and heteroduplex formation was still evident, and it is impossible to rule out preferential amplification of some alleles and/or non-amplification of others. In particular, the low level of sequence variation measured in the black robin was likely to be due to non-amplification of Petr04 in some samples, and may not reflect the true levels of variation. Nevertheless, the validity of every sequence that was included in the analysis was confirmed by amplification in an independent PCR.

In addition, all PBR sequences appear to represent functional loci (with the possible exception of Petr04, see chapter three). The PBR PCR primers were designed specifically to amplify expressed sequences and to exclude the sequence of the putative pseudogene, and translation of sequences into amino acids showed no evidence of non-functional sequences. The number of PBR sequences found in South Island robin individuals (up to 14 per individual) indicates that the previous estimate of 4 transcribed loci (see chapter three) may be an underestimate. This adds to the suggestion in chapter three that there may be a difference in gene copy number between the black robin and South Island robin as a result of recent, within-species gene duplications.

5.4.1 MHC diversity in bottlenecked populations

Levels of MHC variation are highly correlated with levels of neutral minisatellite DNA variation (from Ardem et al. 1997, and Ardem and Lambert 1997). For both minisatellite and MHC data there is a loss of overall allelic diversity in the bottlenecked populations of South Island robin (measured by the total number of hybridising fragments for minisatellite and RFLP data, or total number of alleles in the population for sequence data), and a significant decrease in APD values, but no significant decrease in the number of alleles per individual.

Under balancing selection, it would be expected that relatively more MHC variation would be retained through the bottleneck compared to minisatellite DNA variation (Maruyama and Nei 1981). However, according to the neutral theory of evolution (Kimura 1983), alleles are effectively neutral when $s < 1/2N_e$ (where s = selection coefficient and N_e is the effective population size). Thus, the smaller N_e is, the stronger selection must be to overcome the effect of genetic drift. Where a population has been founded by only two individuals (as in the black robin and Motuara South Is robin populations), s must be greater than 0.25 before any effect due to selection is seen. Although selection intensities may vary greatly over time and space, and can depend on the species and experimental context (Edwards and Hedrick 1998b), most estimates of s for MHC genes are less than 0.05 (Satta et al. 1994, Slatkin and Muirhead 2000, Garrigan and Hedrick 2001, Langefors et al. 2001b). This suggests that MHC genes will be effectively neutral through a bottleneck of a single pair, even if the bottleneck only lasts for a single generation, as selection coefficients considerably larger than normally measured would be required otherwise. No relative increase in MHC variation compared with minisatellite DNA variation was measured in the bottlenecked populations in this study, in accordance with these theoretical predictions. In fact, the Allports Is population appears to have lost proportionally more MHC variation whereas magnitude of decrease was the same for Motuara Is.

The effect of genetic drift on MHC variation can be expected to be dominant through both short and long bottlenecks. The black robin population was confined to Little Mangere Island, a small rock stack containing less than 9 ha of bush that could have supported only 20-30 birds, for approximately 100 years (about 50 robin generations) (Butler and Merton 1992). In the bottlenecked South Island robin populations, however, the founding pairs produced offspring in the first breeding season following their release, and population numbers increased rapidly (Flack 1974, 1978). Only a small fraction of genetic variation ($1/2N_e$) is predicted to be lost per generation through genetic drift (Wright 1931), so populations must be at low levels for a long period for substantial losses to occur. This explains the difference in MHC variation between the black robin and bottlenecked South Island robin populations. The number of alleles/hybridising fragments per individual did not significantly decrease in the bottlenecked South Island robin populations, suggesting there has been little decrease in heterozygosity through the short bottleneck. As for the minisatellite DNA data (Arden and Lambert 1997), the monomorphism evident in the black

robin MHC reflects its long history of low population size, rather than the 1980 bottleneck of five individuals.

Although genetic drift appears to be the predominant force determining the current makeup of MHC alleles in each population, the excess of nonsynonymous changes at peptide binding residues suggests that balancing selection has influenced the long-term evolution of these sequences. However, it is difficult to accurately measure the intensity of selection in this instance. Tests of nonsynonymous (d_N) to synonymous (d_S) changes are usually performed on alleles at a single locus, however the comparison in this study includes alleles of several closely related loci. Although the test is still valid for interlocus comparisons, it may lead to an underestimate of d_N values because nonsynonymous changes may become saturated as sequences become more divergent (Takahata et al. 1992, Edwards et al. 1995a). This means our measures of d_N/d_S for PB codons (where d_S values are high) may be underestimates. This effect may be particularly problematic for the black robin sequences, as the four sequences isolated are highly divergent and appear to represent a single allele at each of four loci. The high ratio of d_N/d_S recorded at non-PB codons in the black robin sequences also appears to be anomalous, as most estimates for non-PB codons in other studies are close to one (e.g. Langefors et al. 2001b, but see Hedrick et al. 2001c). The high ratio is due to a lower than expected d_S value, rather than an elevated d_N value, and may be an artifact of codon usage bias (Yang and Bielawski 2000), possibly exacerbated by the small number of highly divergent sequences in the analysis. It is also possible the exact locations of the peptide binding codons in birds may differ from those of humans, as the crystalline structure of the class II MHC molecule has not been determined for birds. Nonetheless, for all sequences much higher levels of nonsynonymous changes were recorded in PB codons than among non-PB codons, suggesting that balancing selection has been a feature of the evolution of these sequences. This may represent selection occurring over thousands of generations and does not necessarily reflect current selective forces. Similar patterns of MHC variation, where a “footprint” of balancing selection is evident, but short-term neutral forces dominate current patterns of diversity, have been reported in other studies (e.g. Hedrick et al. 2001c)

5.4.2 Intron sequences and the evolution of class II B MHC genes

In this chapter, the prospect of delineating loci on the basis of intron sequences was investigated, with a view to designing single locus primers for MHC genotyping. In

mammals, introns appear to be good indicators of loci, and amplification of single loci using primers targeted to intron sequences has been successfully used in a number of mammalian species (e.g. Figueroa et al. 1994, Edwards et al. 1997). However, passerine MHC genes appear to undergo higher rates of gene conversion and/or more recent gene duplications producing multiple, closely related loci with high sequence similarity between loci, particularly outside the PBR (reviewed in Hess and Edwards 2002, and Zelano and Edwards 2002, and see chapter three, this study). Because of this, the use of introns for analysis of variation at single loci in birds has not been as successful and has been largely limited to amplification of pseudogenes and non-polymorphic genes (e.g. Hess et al. 2000, Gasper et al. 2001). In this study, intron sequences from the putative pseudogene were sufficiently divergent to allow for the design of separate PCR primers. However we could not design single locus primers for genotyping each of the multiple expressed class II B MHC loci individually because the intron sequences immediately flanking exon 2 are highly conserved. In addition, intron sequences of the expressed loci do not group according to the orthologous relationships identified from the 3'UTR in chapter three. This lack of orthology appears to provide support for role of gene conversion at MHC loci in robins. Analysis of cDNA sequences in chapter three suggested that gene conversion is not confined to the PBR, but occurs across the length of the coding region, so it can be expected that the action of gene conversion will extend into the intron. Both introns 1 and 2 appear to be characterised by repeated motifs, and the introns vary considerably in length rather than sequence, largely due to varying numbers of these repeats. Unequal gene conversion between repeated units can create length differences between alleles at the same locus (Armour 1996), meaning intron length may not be a reliable indicator of loci.

Gene conversion may also contribute to homogenisation of sequences between distantly related species (Edwards et al. 1998a). Other passerine species for which intron sequences have been characterised (red-winged blackbird and Darwins finches) show similar patterns of length variation, repeated motifs, and conserved flanking regions (Edwards et al. 1998a, Garrigan and Edwards 1999, Sato et al. 2000). Despite the fact that Darwins finches and red-winged blackbirds are only distantly related to New Zealand robins, the repeat unit present in intron 2 in robins can also be identified in the Darwins finches and red-winged blackbird, and the 30-40 bp of both introns immediately flanking exon 2 appears to be conserved in all species.

The action of gene conversion can also be seen among the PBR sequences, as a patchwork pattern of sequence motifs is evident among many alleles, and much of the diversity among sequences appears to have arisen by shuffling of these motifs. This is not as obvious among black robin sequences, as fewer alleles are available for comparison, and divergence times between sequences may be greater. The higher rates of synonymous substitutions among PB codons are also indicative of gene conversion in these regions (Bergstrom et al. 1998). Previous studies have found evidence for gene conversion within the α -helix, and the domains of the β -sheet, where most of the PB codons are located (Erlich and Gyllensten 1991, Edwards et al. 1995b, 1998a).

5.4.3 MHC monomorphism in the black robin: implications for conservation.

Genetic variation at MHC loci is thought to be important for the long-term survival of populations, as low levels of variation may increase susceptibility to pathogens (O'Brien and Evermann 1988). There are several reports of increased pathogen resistance among MHC heterozygotes (Carrington et al. 1999, Thurz et al. 1997, Arkush et al. 2002, Penn et al. 2002), or associations between specific MHC alleles and resistance or susceptibility to specific pathogens (Paterson et al. 1998, Flores-Villanueva et al. 2001, Langefors et al. 2001a, Lohm et al. 2002). Several studies have reported a general effect of increased pathogen susceptibility in inbred individuals, without invoking a specific MHC related effect (Coltman et al. 1999, Cassinello et al. 2001, Hedrick et al. 2001b, Acevedo-Whitehouse et al. 2003). However, few studies of natural populations have established a direct link between pathogen-mediated population decline and low MHC variation. In addition, disease susceptibility is not always associated with MHC variation (e.g. Gutierrez-Espeleta et al. 2001), suggesting the link between MHC variation and pathogen resistance is not universal.

The black robin population does not presently appear to be adversely affected by a lack of MHC variation, as numbers have continued to increase, despite the fact that the population is no longer intensively managed (H. Aikman, pers comm). There is little evidence that the population suffers from increased susceptibility to pathogens. Several episodes of avian pox have been reported, however, this appears to have only had a small effect on overall population numbers (Butler and Merton 1992, Tisdall and Merton 1988). A preliminary study on pathogen loads in the black robin found no evidence of viral particles, protozoa or

bacterial infection and only low levels of ectoparasites (J. Anderson, unpublished data). However, these results should be interpreted with caution, as it is not possible to determine whether they reflect resistance to disease, or simply lack of exposure. Many populations that are apparently viable despite low MHC variation are thought to be under decreased selection pressure from pathogens due to solitary lifestyles and/or decreased exposure to pathogens (Slade 1992, Ellegren et al. 1996, Hambuch and Lacey 2002). The possibility of low levels of pathogen exposure in the black robin cannot be ruled out. Pathogens may not be able to maintain themselves where host density is low (as occurs during a population bottleneck), unless other reservoirs are available. South East and Mangere Islands have high densities of avian fauna, including other native and introduced passerines, which may act as reservoirs for pathogens able to infect the black robin population. Data on pathogen prevalence in these species may aid in determining the likely level of pathogen exposure to black robins. It has also been suggested that the impact of pathogens is less in cold climates, as pathogens cannot maintain high densities as in tropical regions (Moller 1998). The Chatham Islands are geographically isolated, windswept islands located at 44°S, where the air temperature ranges from 5-20°C, so the effect of pathogens on Chatham Islands fauna may be less than that observed in more tropical regions.

It is possible that the four class II B MHC sequences found in the black robin are sufficient to confer resistance to pathogens commonly encountered by the population. The sequences are highly divergent, and if they all represent functional class II molecules, they may be able to present a wide variety of peptides between them. It could be argued these sequences have been selected for as the population decreased in size, because they confer the greatest resistance to commonly encountered pathogens. However, for such selection to override genetic drift, a slow decrease in population numbers would be required, which appears to be an unlikely scenario. The black robin was once widespread throughout the Chatham Islands, and appears to have suffered mostly from introduced predators and severe habitat loss at the time of European settlement (Butler and Merton 1992). It is likely that large numbers were wiped out quickly, and the population on Little Mangere Island only survived because the island is an isolated, sheer sided rockstack impermeable to predators.

It is difficult to assess the significance of the low level of MHC variation in the extant black robin population without knowing what the “ancestral” levels of variation were. Levels of

diversity in the source populations of South Island robin are one indicator of pre-bottleneck levels of variation in the black robin. However, it is possible that the black robin population has always had substantially lower levels of genetic variation than the mainland robins, as colonisation of the Chatham Islands by *Petroica* may have only involved a few individuals. The bottleneck over the last 100 years may therefore not have resulted in a substantial additional decrease in MHC diversity. Levels of diversity in Chatham Island passerines that have not gone through a severe recent bottleneck (such as the tomtit (*Petroica macrocephala chathamensis*) or grey warbler (*Gerygone albofrontata*) are likely to be a more accurate indicator of the pre-bottleneck levels of variation in the black robin. Interestingly, levels of minisatellite DNA variation in the Chatham Island tomtit are similar those reported in the black robin (Ma and Lambert 1997).

In conclusion, the generation of diversity at class II B MHC genes in New Zealand robins appears to be influenced by gene conversion and balancing selection, however the current forces determining the makeup of MHC alleles in bottlenecked populations are neutral. The black robin appears to be monomorphic at class II B loci but is viable under existing conditions, however predictions about the long-term viability of the population will require extensive data on pathogen exposure in both the black robin and its close neighbours. It is also important to note that although the current makeup of MHC alleles in the black robin may be sufficient to counter pathogens currently present, the population may still be highly vulnerable to novel pathogens.

5.5 References

- Acevedo-Whitehouse K, Gulland F, Greig D, Amos W (2003) Disease susceptibility in California sealions. *Nature* 422: 35
- Apanius V, Penn D, Slev PR, Ruff LR, Potts WK (1997) The nature of selection on the major histocompatibility complex. *Crit. Rev. Immunol.* 17: 179-224
- Arden SL, McLean IG, Anderson S, Maloney R, Lambert DM (1994) The effects of blood sampling on the behaviour and survival of the endangered Chatham Island Black Robin (*Petroica traversi*). *Conserv. Biol.* 8: 857-862
- Arden SL and Lambert DM (1997) Is the black robin in genetic peril? *Mol. Ecol.* 6: 21-28
- Arden SL, Lambert DM, Rodrigo AG, McLean IG (1997) The effects of population bottlenecks on multilocus DNA variation in robins. *J. Heredity* 88: 179-186

- Arkush KD, Giese AR, Mendonca HL, McBride AM, Marty GD, Hedrick PW (2002) Resistance to three pathogens in the endangered winter-run chinook salmon (*Oncorhynchus tshawytscha*): effects of inbreeding and major histocompatibility complex genotypes. *Can. J. Fish. Aquat. Sci.* 59: 966-975
- Armour JAL (1996) Tandemly repeated minisatellites: generating human genetic diversity via recombinational mechanisms. *In* M Jackson, T Strachan, and G Dover (eds.): *Human Genome Evolution*. BIOS Scientific Publishers, Oxford
- Bergstrom TF, Josefsson A, Erlich HA, Gyllensten U (1998) Recent origin of *HLA-DRB1* alleles and implications for human evolution. *Nature Genetics* 18: 237-242
- Boyce WM, Hedrick PW, Muggli-Cockett NE, Kalinowski S, Penedo MCT, Ramey RR (1997) Genetic variation of major histocompatibility complex and microsatellite loci: a comparison in Bighorn sheep. *Genetics* 145: 421-433
- Brown JH, Jardetzky TS, Gorga JC, Stern LJ, Urban RG, Strominger JL, Wiley DC (1993) Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature* 364: 33-39
- Butler D and Merton D (1992) *The Black Robin: Saving the World's Most Endangered Bird*. Oxford University Press.
- Carrington M, Nelson GW, Martin MP, Kissner T, Vlahov D, Goedert JJ, Kaslow R, Buchbinder S, Hoots K, O'Brien SJ (1999) HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage. *Science* 283: 1748-1752
- Cassinello J, Gomendio M, Roldan ERS (2001) Relationship between coefficient of inbreeding and parasite burden in endangered gazelles. *Conserv. Biol.* 15: 1171-1174
- Coltman DW, Pilkington JG, Smith JA, Pemberton JM (1999) Parasite-mediated selection against inbred Soay sheep in a free-living, island population. *Evolution* 53: 1259-1267
- Edwards SV, Grahn M, Potts WK (1995a) Dynamics of *Mhc* evolution in birds and crocodilians: amplification of class II genes with degenerate primers. *Mol. Ecol.* 4: 719-729
- Edwards SV, Wakeland EK, Potts WK (1995b) Contrasting histories of avian and mammalian *Mhc* genes revealed by class II B sequences from songbirds. *Proc. Natl. Acad. Sci. USA* 92: 12200-12204
- Edwards SV, Chesnut K, Satta Y, Wakeland EK (1997) Ancestral polymorphism of *Mhc* Class II genes in mice: Implications for balancing selection and the mammalian molecular clock. *Genetics* 146: 655-668
- Edwards SV, Gasper J, March M (1998a) Genomics and polymorphism of *Agph-DAB1*, an MHC Class II B gene in red-winged blackbirds (*Agelaius phoeniceus*). *Mol. Biol. Evol.* 15: 236-250
- Edwards SV and Hedrick PW (1998b) Evolution and ecology of MHC molecules: from genomics to sexual selection. *Trends Ecol. Evol* 13: 305-311
- Ellegren H, Hartman G, Johansson M, Andersson L (1993) Major histocompatibility complex monomorphism and low levels of DNA fingerprinting variability in a reintroduced and rapidly expanding population of beavers. *Proc. Natl. Acad. Sci. USA* 90: 8150-8153
- Ellegren H, Mikko S, Wallin K, Andersson L (1996) Limited polymorphism at major histocompatibility complex (MHC) loci in the Swedish moose A-alces. *Mol. Ecol.* 5: 3-9
- Ennis PD, Zemmour J, Salter RD, Parham P (1990) Rapid cloning of HLA-A, B cDNA by using the polymerase chain reaction: frequency and nature of errors produced in amplification. *Proc. Natl. Acad. Sci. USA* 87: 2833-2837

- Erlich HA and Gyllensten UB (1991) Shared epitopes among HLA class II alleles: gene conversion, common ancestry, and balancing selection. *Immunology Today* 12: 411-414
- Figueroa F, O'hUigin C, Tichy H, Klein J (1994) The origin of the primate *Mhc-DRB* genes in allelic lineages deduced from the study of prosimians. *J. Immunol* 152: 4455-4465
- Flack JAD (1974) Chatham Island black robin. *Wildlife - a Review, New Zealand Wildlife Service* 5: 25-34
- Flack JAD (1978) Interisland transfers of New Zealand black robins. *In* SA Temple (ed.): *Endangered Birds: Management Techniques for the Preservation of Threatened Species*. University of Wisconsin Press, Madison
- Flores-Villanueva PO, Yunis EJ, Delgado JC, Vittinghoff E, Buchbinder S, Leung JY, Ugialoro AM, Clavijo OP, Rosenberg ES, Kalams SA, Braun JD, Boswell SL, Walker B, Goldfield AE (2001) Control of HIV-1 viremia and protection from AIDS are associated with HLA-Bw4 homozygosity. *Proc. Natl. Acad. Sci. USA* 98: 5140--5145
- Freeman-Gallant CR, Johnson EM, Saponara F, Stanger M (2002) Variation at the major histocompatibility complex in Savannah sparrows. *Mol. Ecol.* 11: 1125-1130
- Garrigan D and Edwards SV (1999) Polymorphism across an exon-intron boundary in an avian *Mhc* class II B gene. *Mol. Biol. Evol.* 16: 1599-1606
- Garrigan D and Hedrick PN (2001) Class I MHC polymorphism and evolution in endangered California Chinook and other Pacific salmon. *Immunogenetics* 53: 483-489
- Gasper JS, Shiina T, Inoko H, Edwards SV (2001) Songbird genomics: Analysis of 45 kb upstream of a polymorphic *Mhc* class II gene in red-winged blackbirds (*Agelaius phoeniceus*). *Genomics* 75: 26-34
- Groombridge JJ, Jones CG, Bruford MW, Nichols RA, (2000) 'Ghost' alleles of the Mauritius kestrel. *Nature* 403: 616
- Gutierrez-Espeleta GA, Hedrick PW, Kalinowski ST, Garrigan D, Boyce WM (2001) Is the decline of bighorn sheep from infectious disease the result of low MHC variation? *Heredity* 86: 439-500
- Hambuch TM and Lacey EA (2002) Enhanced selection for MHC diversity in social tuco-tucos. *Evolution* 56: 841-845
- Hedrick PW and Kim TJ (1999) Genetics of complex polymorphisms: parasites and the maintenance of the major histocompatibility complex variation. *In* RS Singh and CB Krimbas (eds.): *Evolutionary Genetics: From Molecules to Morphology*. Cambridge University Press
- Hedrick PW, Karen MP, Gutierrez-Espeleta GA, Rattink A, Lievers K (2000) Major histocompatibility complex variation in the Arabian oryx. *Evolution* 54: 2145-2151
- Hedrick PW, Gutierrez-Espeleta GA, Lee RN (2001a) Founder effect in an island population of bighorn sheep. *Mol. Ecol.* 10: 851-857
- Hedrick PW, Kim TJ, Parker KM (2001b) Parasite resistance and genetic variation in the endangered Gila topminnow. *Anim. Cons.* 4: 103-109
- Hedrick PW, Parker KM, Lee RN (2001c) Using microsatellite and MHC variation to identify species, ESUs, and MUs in the endangered Sonoran topminnow. *Mol. Ecol.* 10: 1399-1412
- Hess CM, Gasper J, Hoekstra HE, Hill CE, Edwards SV (2000) MHC Class II pseudogene and genomic signature of a 32 kb cosmid in the house finch (*Carpodacus mexicanus*). *Genome Res.* 10: 613-623

- Hess CM and Edwards SV (2002) The evolution of the major histocompatibility complex in birds. *BioScience* 52: 423-431
- Kaufman J, Salomonsen J, Flajnik M (1994) Evolutionary conservation of MHC class I and class II molecules - different yet the same. *Seminars in Immunology* 6: 411-424
- Kimura M (1983) *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge, UK.
- Klein J (1986) *Natural History of the Major Histocompatibility Complex*. John Wiley and Sons, New York.
- Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 17: 1244-1245
- Langefors A, Lohm J, Grahn M, Anderson O, von Schantz T (2001a) Association between major histocompatibility complex class IIB alleles and resistance to *Aeromonas salmonicida* in Atlantic salmon. *Proc. R. Soc. Lond. B* 268: 479-485
- Langefors K, Lohm J, von Schantz T (2001b) Allelic polymorphism in MHC class IIB in four populations of Atlantic salmon (*Salmo salar*). *Immunogenetics* 53: 329-336
- Lohm J, Grahn M, Langefors A, Andersen O, Storset A, von Schantz T (2002) Experimental evidence for major histocompatibility complex-allele-specific resistance to a bacterial infection. *Proc. R. Soc. Lond. B* 269: 2029-2033
- Ma W and Lambert DM (1997) Minisatellite DNA markers reveal hybridisation between the endangered black robin and tomtit. *Electrophoresis* 18: 1682-1687
- Madsen T, Olsson M, Wittzell H, Stille B, Gullberg A, Shine R, Andersson S, Tegelstrom H (2000) Population size and genetic diversity in sand lizards (*Lacerta agilis*) and adders (*Vipera berus*). *Biol. Conserv.* 94: 257-262
- Maruyama T and Nei M (1981) Genetic variability maintained by mutation and overdominant selection in finite populations. *Genetics* 98: 441-459
- Meyer D and Thomson G (2001) How selection shapes variation of the human major histocompatibility complex: a review. *Annals of Human Genetics* 65: 1-26
- Mikko S, Roed K, Schmutz S, Andersson L (1999) Monomorphism and polymorphism at *Mhc DRB* loci in domestic and wild ruminants. *Immunol. Rev.* 167: 169-178
- Moller AP (1998) Evidence of larger impact of parasites on hosts in the tropics: investment in immune function within and outside the tropics. *Oikos* 82: 265-270
- Murray BW and White BN (1998) Sequence variation at the major histocompatibility complex DRB loci in beluga (*Delphinapterus leucas*) and narwhal (*Monodon monoceros*). *Immunogenetics* 48:242-252.
- Nei M, Maruyama T, Chakraborty R (1975) The bottleneck effect and genetic variability in populations. *Evolution* 29: 1-10
- Nei M and Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* 3: 418-426
- Nevo E, Kirzhner V, Beiles A, Korol A (1997) Selection versus random drift: long-term polymorphism persistence in small populations (evidence and modelling). *Phil. Trans. R. Soc. Lond. B* 352
- O'Brien SJ and Evermann JF (1988) Interactive influence of infectious disease and genetic diversity in natural populations. *Trends Ecol. Evol.* 3: 254-259

- Orita M, Iwahana H, Kanazawa H, Hayashi K, Sekiya T (1989) Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proc. Natl. Acad. Sci. USA* 86: 2766-2770
- Paterson S, Wilson K, Pemberton JM (1998) Major histocompatibility complex variation associated with juvenile survival and parasite resistance in a large unmanaged ungulate population (*Ovis aries L.*). *Proc. Natl. Acad. Sci. USA*. 95: 3714-3719
- Penn DJ, Damjanovich K, Potts WK (2002) MHC heterozygosity confers a selective advantage against multiple-strain infections. *Proc. Natl. Acad. Sci. USA* 99: 11260-11264
- Plante Y, Boag PT, White BN, Boonstra R (1991) Highly polymorphic genetic markers in meadow voles (*Microtus pennsylvanicus*) revealed by a murine major histocompatibility complex (MHC) probe. *Can J. Zool.* 69: 213-220
- Sato A, Figueroa F, Mayer WE, Grant PR, Grant BR, Klein J (2000) Mhc class II genes of Darwin's Finches: divergence by point mutations and reciprocal recombination. *In* M Kasahara (ed.): Major Histocompatibility Complex: Evolution, Structure and Function. Springer Verlag
- Satta Y, O'hUigin C, Takahata N, Klein J (1994) Intensity of natural selection at the major histocompatibility complex loci. *Proc. Natl. Acad. Sci. USA*. 91: 7184-7188
- Seddon JM and Baverstock PR (1999) Variation on islands: major histocompatibility complex (*Mhc*) polymorphism in populations of the Australian bush rat. *Mol. Ecol.* 8: 2071-2079
- She JX, Boehme SA, Wang TW, Bonhomme F, Wakeland EK (1991) Amplification of major histocompatibility complex class II gene diversity by intraexonic recombination. *Proc. Natl. Acad. Sci. USA* 88: 453-457
- Slade RW (1992) Limited MHC polymorphism in the southern elephant seal: implications for MHC evolution and marine mammal population biology. *Proc. R. Soc. Lond. B.* 249: 163-171
- Slatkin M and Muirhead CA (2000) A method for estimating the intensity of overdominant selection from the distribution of allele frequencies. *Genetics* 156: 2119-2126
- Sommer S, Schwab D, Ganzhorn JU (2002) MHC diversity of endemic Malagasy rodents in relation to geographic range and social system. *Behav. Ecol. Sociobiol.* 51: 214-221
- Takahata N, Satta Y, Klein J (1992) Polymorphism and balancing selection at major histocompatibility complex loci. *Genetics* 130: 925-938
- Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* 22: 4673-4680
- Thursz MR, Thomas HC, Greenwood BM, Hill AVS (1997) Heterozygote advantage for HLA class II type in hepatitis B virus infection. *Nature Genetics* 17: 11-12
- Tisdall DJ and Merton DV (1988) Disease surveillance in the Chatham Islands black robin. *Surveillance* 15
- Trowsdale J, Groves V, Amason A (1989) Limited MHC polymorphism in whales. *Immunogenetics*: 19-24
- Wagner A, Blackstone N, Cartwright P, Dick M, Misof B, Snow P, Wagner GP, Bartels J, Murtha M, Pendleton J (1994) Surveys of gene families using polymerase chain reaction: PCR selection and PCR drift. *Syst. Biol.* 43: 250-261
- Weissensteiner T and Lanchbury JS (1996) Strategy for controlling preferential amplification and avoiding false negatives in PCR typing. *Biotechniques* 21: 1102-1108

- Wisely SM, Buskirk SW, Fleming MA, McDonald DB, Ostrander EA (2002) Genetic diversity and fitness in black-footed ferrets before and during a bottleneck. *J. Heredity* 93: 231-237
- Wittzell H, Madsen T, Westerdahl H, Shine R, von Schantz T (1999) MHC variation in birds and reptiles. *Genetica* 104: 301-309
- Wright S (1931) Evolution in mendelian populations. *Genetics* 16: 97-159
- Yang Z and Bielawski JP (2000) Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol.* 15: 496-503
- Yuhki N and O'Brien SJ (1990) DNA variation of the mammalian major histocompatibility complex reflects genomic diversity and population history. *Proc. Natl. Acad. Sci. USA* 87: 836-840
- Zelano B and Edwards SV (2002) An Mhc component to kin recognition and mate choice in birds: Predictions, progress, and prospects. *Am. Nat.* 160: S225-S237
- Zylstra P, Rothenfluh HS, Weiller GF, Blanden RV, Steele EJ (1998) PCR amplification of murine immunoglobulin germline V genes: Strategies for minimization of recombination artefacts. *Immunol. Cell Biol.* 76: 395-405

CHAPTER SIX

A Molecular Phylogeny of New Zealand's *Petroica* Species Based on Mitochondrial DNA Sequences.

6.1 Introduction

The Petroicidae family of Australo-Papuan robins comprises approximately 44 species distributed around Australia, New Zealand, New Guinea, and the western Pacific (Sibley and Monroe 1990). Australo-Papuan robins are small flycatchers, and are not closely related to the North American or European robin (family Turdidae), or to the Eurasian robins (Old-World flycatchers, Muscicapidae). Instead, Petroicidae are part of the Australo-Papuan radiation of oscines (parvorder Corvida *sensu* Sibley and Ahlquist 1990), now thought to be basal to the Afro-Eurasian oscines (parvorder Passerida) (Barker et al. 2002, Ericson et al. 2002a). However the exact placement of Petroicidae within the oscine phylogeny remains unresolved (Christidis and Schodde 1991, Ericson et al. 2002b).

Three species of Petroicidae endemic to New Zealand have been named: the New Zealand robin (*Petroica australis*), the Chatham Islands black robin (*Petroica traversi*), and the tomtit (*Petroica macrocephala*) (Fleming 1950). The black robin is highly endangered and is confined to South East (Rangatira), and Mangere Islands in the Chatham Islands group, 850 km east of the South Island of New Zealand (Butler and Merton 1992). Both the New Zealand robin and tomtit were widely distributed throughout mainland New Zealand and its nearby offshore islands at the time of European settlement (Heather and Robertson 2000). However their numbers have declined since then, largely due to clearance of habitat and introduction of predators, particularly stoats and rats (Worthy and Holdaway 2002). The New Zealand robin is now patchily distributed: in the North Island apart from artificially established populations on island sanctuaries, they are restricted to Little Barrier and Kapiti

Islands, and a band across the central North Island, including strongholds in Pureora and Mamaku forests (see Figure 6.1). In the South Island they are locally common in the north and west, but patchily distributed further south, and are also moderately common on Stewart Island. The New Zealand tomtit is more widespread than the robin and is particularly common in forests of the central North Island and South Island (Heather and Robertson 2000).

The different island forms of robins and tomtits have been described as subspecies on the basis of morphological differences (e.g., plumage, wing/tail length) (Fleming, 1950). However, in the nearly 200 years since the New Zealand members of the *Petroicidae* family were first described, the division of species, subspecies and even genera has undergone many changes, with each of these subspecies at times being described as separate species. In 1950, Fleming undertook a comprehensive analysis of museum skins and significantly revised the taxonomy of the genus. He placed the New Zealand robins in the subgenus *Miro* and named three subspecies: the North Island robin *P. a. longipes*, the South Island robin *P. a. australis*, and the Stewart Island robin *P. a. rakiura*. The Chatham Island black robin was described as a melanistic form of *Miro*, but appeared to have a longer history of separation and greater morphological differences than the mainland subspecies so was named as a separate species, *P. traversi*. Five subspecies of New Zealand tomtit, each restricted to an island, were described: the North Island tomtit *P. m. toitoi*; the South Island tomtit *P. m. macrocephala*; the Chatham Island tomtit *P. m. chathamensis*; the Auckland Island tomtit *P. m. marrineri*; and the Snares Island tomtit *P. m. dannefaerdi*. The Snares Island tit, like the black robin, is a melanistic form and was originally recorded as being the same species as the black robin by Buller, (Buller 1890), and later as a subspecies of black robin (Mathews and Iredale 1913). In naming the Snares Island tit as a subspecies of *P. macrocephala*, Fleming (1950) recognised its distinctiveness from the black robin for the first time. Fleming (1950) also noted that the *P. macrocephala* subspecies appear to be more closely related to the Australian *Petroica* species, and suggested there had been two separate invasions of ancestral *Petroica* stock, an early (Pliocene) invasion giving rise to robins and a more recent colonisation (early Pleistocene) of the ancestral *P. macrocephala*. He suggested that *P. macrocephala* was a derivative of the scarlet robin *P. multicolor*, which is found throughout the south-east and south-west of Australia, and in the western Pacific, including Norfolk Island, Fiji and Samoa (Higgins and Peter 2002).

Fleming's nomenclature has been the most widely accepted to date, however it has recently been proposed that most subspecies of *P. australis* and *P. macrocephala* should be treated as separate species (Holdaway et al. 2001, Worthy and Holdaway 2002). On the basis of the range of morphological variation present, Holdaway and Worthy named a total of eight species within the New Zealand *Petroica* genus, a similar division to that proposed in the first half of the 19th century (reviewed in Fleming 1950).

The various classifications of New Zealand's *Petroica* species have so far been based entirely on morphological and behavioural characteristics and have never been investigated using genetic methods. In addition, comparisons between the black robin and South Island robin form an important component of this thesis, and understanding the evolutionary history of these species is important for placing comparisons of MHC sequences in context (e.g. timing MHC gene duplication events, chapter three). Thus, in this chapter, the evolutionary history of New Zealand's *Petroica* species is investigated by phylogenetic analysis of mitochondrial DNA markers. Cytochrome *b* (*cyt b*) and mitochondrial control region sequences are analyzed from representatives of all New Zealand *Petroica* taxa, as well as two subspecies of *P. multicolor* - *P. m. multicolor* from Norfolk Island, and *P. m. kleinschmidti* from Fiji. Specifically, I investigate: (1) how closely related is the black robin to the New Zealand robin? (2) Are the DNA markers consistent with current taxonomic classifications based on morphology? (3) Is there support for the hypothesis of two separate invasions of *Petroica*, and is *P. macrocephala* a derivative of *P. multicolor*?

6.2 Materials and Methods

6.2.1 Samples

We analyzed multiple samples from all *Petroica* taxa present in New Zealand (details of taxa, location and type of sample are given in Table 6.1 and geographic location of samples is given in Figure 6.1). Species and subspecies are named as in Fleming, 1950. Where possible, we used samples from throughout the geographic range of the taxon. Modern samples (i.e. blood and feather) were used where possible. Museum specimens were used where fresh sources of DNA were not available for a particular taxon, or to provide additional geographic or temporal data for taxa where blood or feather samples were also available. The *cyt b*

sequence from the outgroup *Eopsaltria australis* (Australian yellow robin) was obtained from Genbank (Accession No. AF096455).

Table 6.1 *Petroica* taxa sampled in this study, including location and sample type. For museum specimens the voucher number is also given. All museum specimens are from the Museum of New Zealand (Te Papa). The year the specimen was collected is given in brackets. Nomenclature of Fleming (1950) is followed.

Taxon	Location (No. sampled)	Sample type	Voucher No.
<i>P. traversi</i> (black robin)	South East Is (8)	Blood	
	Little Mangere Is (3)	Toe-pad	DM 1687 (1871); DM 1688 (1897); DM 1693 (1900)
<i>P. australis longipes</i> (North Is robin)	Mamaku (6)	Blood	
	Pureora (5)	Feather	
	Kapiti Is (2)	Feather	
	Kapiti Is (1)	Toe-pad	NM 22198 (1976)
	Little Barrier Is (1)	Toe-pad	NM 24160 (1990)
<i>P. australis australis</i> (South Is robin)	Nukuwaiata Is (6)	Blood	
	Kaikoura (8)	Blood	
	D'Urville Is (1)	Toe-pad	DM 5257 (1905)
	Eglinton Valley (1)	Toe-pad	NM 22247 (1980)
<i>P. australis rakiura</i> (Stewart Is robin)	Stewart Is (4)	Feather	
	Ulva Is (2)	Feather	
	Pouawaitai Is (1)	Toepad	DM 8701 (1956)
<i>P. macrocephala toitoi</i> (North Is tomtit)	Wgtn region (1)	Toe-pad	DM 5207 (1906)
	Wanganui (1)	Toe-pad	NM 9695 (1933)
	Wellington (2)	Toe-pad	NM 21785 (1966); NM 21786 (1973)
	Gisborne (1)	Toe-pad	NM 24841 (1974)
	Whirinaki (1)	Toe-pad	NM 26790 (2000)
<i>P. macrocephala macrocephala</i> (South Is tomtit)	Stewart Is (1)	Toe-pad	DM 8703 (1956)
	Fiordland (1)	Toe-pad	NM 22047 (1958)
	Westland (2)	Toe-pad	NM 22691 (1981); NM 22720 (1981)
	Hunter Mtns (1)	Toe-pad	NM 22870 (1983)
	Takaka (1)	Toe-pad	NM 24297 (1989)
<i>P. macrocephala chathamensis</i> (Chatham Is tomtit)	Mangere Is (2)	Blood	
	South East Is (6)	Feather	
	South East Is (1)	Toe-pad	DM 1833 (1926)
<i>P. macrocephala dannefaerdi</i> (Snares Is tomtit)	Snares Is (4)	Toe-pad	DM 5242 (1947); DM 18614 (1975); DM 23668 (1987); DM 26238 (1999)
<i>P. macrocephala marrineri</i> (Auckland Is tomtit)	Auckland Is (2)	Toe-pad	DM 7943 (1952); DM 13226 (1943)
<i>P. multicolor multicolor</i>	Norfolk Is (1)	Toe-pad	NM 19900 (1901)
<i>P. multicolor kleinschmidti</i> (scarlet robin)	Viti Levu, Fiji (1)	Toe-pad	DM 20177 (1973)

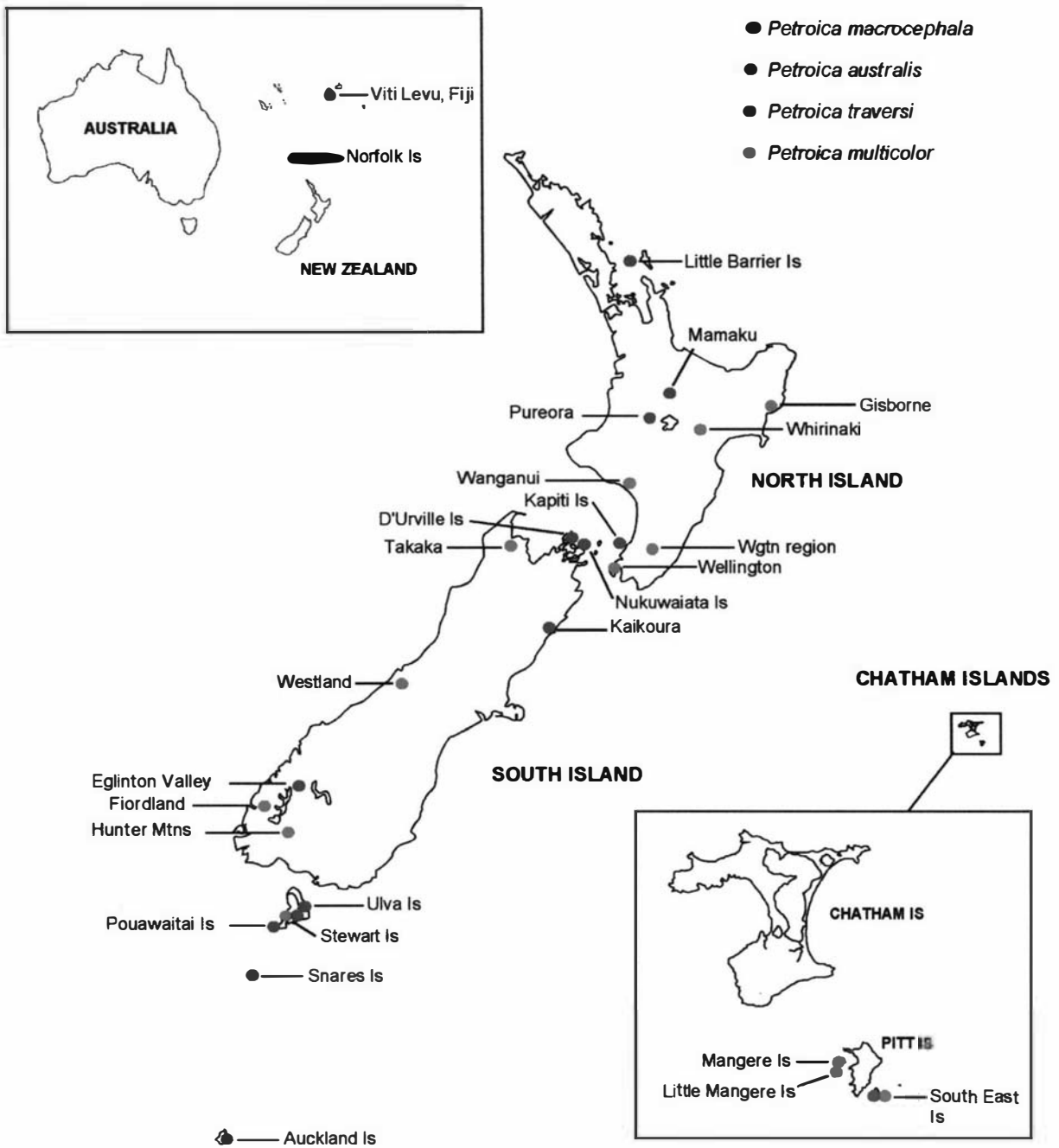


Figure 6.1 Map of New Zealand and Pacific islands, showing the location of samples used in this study.

6.2.2 Isolation of Genomic DNA

Genomic DNA was extracted from whole blood, feathers, or museum skins, and purified using standard phenol/chloroform methods. Blood cells (5-10 μL) were lysed and digested as described in Ardem and Lambert (1997). For extraction from feathers, 2-3 mm of the tip of the feather was digested for 2-3 hrs at 50°C in 400 μl of digestion buffer (10 mM Tris-HCl pH 8.0, 50 mM NaCl, 10 mM EDTA) with 0.25% SDS and 500 $\mu\text{g}/\text{mL}$ proteinase-K. Museum specimens were sampled by taking a small scraping (approximately 1 mm x 1 mm x 3 mm) from the toe-pad using a sterile scalpel. This was cleaned and hydrated by soaking in two changes of milliQ-H₂O for 30 min each. The tissue was then cut into smaller pieces using a sterile scapel blade and incubated overnight at 50°C in 400 μl of digestion buffer (as above) with 0.25% SDS, 500 $\mu\text{g}/\text{mL}$ proteinase-K, and 10 mg/mL DTT. All museum specimen extractions were performed in a dedicated ancient DNA laboratory. Following purification, DNA was ethanol precipitated and resuspended in accordance with Sambrook et al. (1989).

6.2.3 PCR and sequencing

A 485 bp fragment of the *cyt b* gene was isolated by PCR amplification using the primers H15649 and L15212 (Baker et al. 1995). This fragment could not be amplified in its entirety from some museum specimens, so internal primers PetCytB1 and PetCytB2 were used in combination with H15649 and L15212, respectively, to amplify two overlapping fragments (236 bp and 311 bp) covering the same portion of the gene. The approximate position of primers used in this study and their sequences are given in Figure 6.2. To obtain control region sequences, we initially amplified a segment of the mitochondrial genome from *cyt b* through to the centre of the control region (see results), using the L15212 primer with CRDR1, which was designed to the central conserved domain of the control region. From this sequence, we designed primers BRtRNAThr and BRCR2 to amplify 506 bp of the 5' end of the control region. For museum specimens, this fragment was amplified in two overlapping segments, using primer pairs BRCR2 and BRCRInt2, and BRtRNAThr with either BRCRInt1 (black robin and tomtit taxa), or SRCRInt1 (New Zealand robin taxa).

PCR amplifications were carried out in 25 μ L reaction volumes containing 10 mM Tris-HCl, 50 mM KCl pH 8.3, 1.5 mM MgCl₂, 200 μ M each dNTP, 0.4 μ M each primer, and 0.5 units of *Taq* polymerase (Roche). Thermal cycling was performed in a Hybaid OmniGene or BioRad iCycler for 35 cycles consisting of 94°C 30 sec, 54°C 10 sec, and 72°C 35 sec, with a final extension of 72°C for 2 min. For museum specimens, BSA (2 mg/ml) was added, and thermal cycling was performed for 40 cycles. PCR products were purified using a High Pure PCR Product Purification Kit (Roche), sequenced using the BigDye Terminator Cycle Sequencing kit ver. 3.0 and 3.1 (Applied Biosystems) and analysed on an ABI 377A automated sequencer.

6.2.4 Data analysis

Sequences were edited in Sequencher™ 4.1.2 (Gene Codes Corporation), and aligned with Clustal W (Thompson et al. 1994). See Appendix B for sequence alignments relating to this chapter. The number of variable sites, uncorrected pairwise sequence divergence (*p*-distance), base frequencies and mean pairwise transition/transversion ratio (*R*) were calculated in MEGA2 (Kumar et al. 2001). Phylogenetic analyses were performed using distance, parsimony and likelihood approaches in PAUP*4.0b10 (Swofford 2002). Maximum parsimony (MP) and maximum likelihood (ML) analyses were performed using a heuristic search with tree-bisection-reconnection (TBR) branch swapping and random addition of taxa (10 replicates). For maximum likelihood analyses the program ModelTest ver. 3.06 (Posada and Crandall 1998) was used to determine the optimal evolutionary model. For the *cyt b* dataset, the hierarchical likelihood ratio test (hLRT) selected the general time-reversible model with the following parameters: empirical base frequencies A = 0.2915, C = 0.3567, G = 0.1375, T = 0.2143; substitution rate matrix A↔C = 1.4230, A↔G = 4.8689, A↔T = 0.0000, C↔G = 0.0000, C↔T = 16.6965, and G↔T = 1.0; gamma distribution shape parameter = 0.1489. The Akaike Information Criterion (AIC) selected a model with slightly different parameters, however using these parameters did not result in any difference in tree topology. For the control region dataset, *P. australis* and *P. macrocephala* sequences were analyzed separately. For *P. australis*, the Hasegawa-Kashino-Yano (HKY) model was selected (hLRT) with the following parameters: empirical base frequencies A = 0.3128, C = 0.3299, G = 0.1314, T = 0.2260, ti/tv ratio = 12.8484, gamma shape parameter = 0.1622. The Tamura-Nei (TrN) model was selected under AIC, however using these parameters did not

result in any difference in tree topology. The HKY and TrN models were also selected for the *P. macrocephala* data, with the following parameters: (HKY) empirical base frequencies A = 0.2999, C = 0.3169, G = 0.1384, T = 0.2448, ti/tv ratio = 6.0233, proportion invariable sites (I) = 0.8343, gamma shape parameter = 0.8594; (TrN) empirical base frequencies A = 0.2960, C = 0.3207, G = 0.1344, T = 0.2489, substitution rate matrix A↔C = 1.0, A↔G = 16.2295, A↔T = 1.0, C↔G = 1.0, C↔T = 7.3439, and G↔T = 1.0, I = 0.88741, equal rates for all sites. In this case the different models produced some difference in tree topology (see results).

Maximum likelihood models described above were also used for distance analysis, and trees were drawn using neighbour-joining (NJ) and minimum evolution (ME) methods. Tree topologies were evaluated using bootstrap analysis with 500 replicates for maximum parsimony and neighbour-joining methods, and 100 replicates for maximum likelihood.

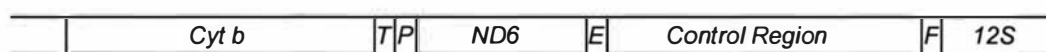
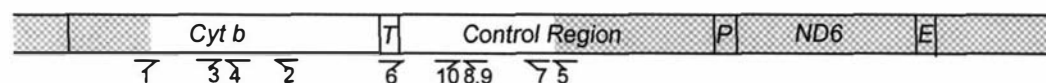
6.3 Results

6.3.1 Mitochondrial gene order

Amplification of a fragment of the mitochondrial genome from *P. traversi* using primers L15212 and CRDR1 produced a 1457 bp product. This is smaller than expected based on the most common mitochondrial gene order found in birds (e.g. Desjardins and Morais 1990), and a BLAST search revealed that this fragment spans 846 bp of *cyt b*, 69 bp of tRNA-Thr, and 539 bp of the 5' end of the control region (Figure 6.2A & B). Amplification from *P. a. australis* using the same primers revealed the same gene order arrangement (see Appendix B for full sequence. This gene order is identical to that found in the *Phylloscopus* warblers (Bensch and Harlid 2000), suboscine passerines, and Falconiformes, Cuculidae, and Picidae (Mindell et al. 1998)

6.3.2 Mitochondrial DNA sequences

Cytochrome *b* and control region sequences were obtained from representatives of all New Zealand *Petroica* taxa, plus two subspecies of *P. multicolor* from Norfolk Island and Fiji.

A Original mt gene order in birds**B** *Petroica* mt gene order**C** Primer sequences

Reference

¹ L15212	5'-GGACGAGGCTTTTACTACGGCTC-3'	Baker et al. 1995
² H15649	5'-TTGCTGGGGTGAAGTTTTCTGGGTC-3'	"
³ PetCytB1	5'-GGTTAGGGTGGGATTGTCTACTGAGA-3'	This study
⁴ PetCytB2	5'-CGGCCAAACACTAGTAGAATGAGCCT-3'	"
⁵ CRDR1	5'-CCAGTGGCGCAAAAGAGCAAGTT-3'	"
⁶ BRtRNAThr	5'-TGGTCTTGTAACCAAAGATTGAAG-3'	"
⁷ BRCR2	5'-GACGTTGAGTAGCTCGGTTCTCGT-3'	"
⁸ BRCRInt1	5'-ATAGCTCGAACATTATCTCATTGGA-3'	"
⁹ SRCRInt1	5'-ATGTCTTGACACAACATCTCATTGGA-3'	"
¹⁰ BRCRInt2	5'-CAAACCAAACATYTTCTCCAATGA-3'	"

Figure 6.2 (A) Original mitochondrial gene order reported in birds (Desjardins and Morais 1990). (B) Gene order found in Petroicidae, which appears identical to the rearrangement reported by Mindell et al. (1998) and Bensch and Harlid (2000). The shaded region was not isolated in this study, but is inferred from the gene order reported in the references above. The approximate position of primers used in this study is shown, with corresponding primer names and sequences given in C.

For *cyt b*, 19 different haplotypes were obtained from 78 individuals (excluding the outgroup). In 425 bp of sequence, there were 85 (20%) variable sites, of which 70 (16.5%) were parsimony informative. The variable sites were largely confined to the third codon position (82%), with fewer at the first (15%) and second (2%) positions. The base composition is similar to that observed for other passerine birds (e.g. Helm-Bychowski and Cracroft 1993), with a bias towards C (0.332), a deficiency in G (0.134), with approximately equal frequencies of T (0.25) and A (0.284). A strong bias against G and T was observed at third codon positions.

The control region dataset contains 77 sequences, as it was not possible to obtain clean sequence from the *P. traversi* museum specimen DM1687. There were 37 haplotypes among the 77 sequences. A homopolymeric tract of cytosines was present at the 5' end of the control region sequences. This produced a region of ambiguous double sequence immediately adjacent to it when sequenced directly, suggesting length heteroplasmy within

the polycytidine stretch. To overcome this, all samples were directly sequenced using the reverse primer BR CR2, and the sequence from the homopolymeric tract to the tRNA-Thr was excluded. This left 381 bp of control region sequence (including 5 indels) in the dataset. There were 157 (41.2%) variable sites, and 142 (37.3%) parsimony informative sites. As in the *cyt b* sequences, the control region sequences showed asymmetric base frequencies (A = 0.298, T = 0.239, C = 0.329, G = 0.134).

Considering all the ingroup taxa, the mean uncorrected sequence divergence was 6.8% (± 0.7) for *cyt b* sequences, and 14.4% (± 1.0) for control region sequences. Between the outgroup *E. australis* and each ingroup taxa the sequence divergence for *cyt b* was 13.4% (± 1.6), 14.3% (± 1.7), 14.6% (± 1.7), and 13.4% (± 1.6) for *P. traversi*, *P. macrocephala*, *P. australis*, and *P. multicolor*, respectively. No control region sequences were available for the outgroup taxa. Uncorrected *p*-distances for sequence divergence within and between each species are given in Table 6.2. Mean pairwise sequence divergence for control sequences was consistently higher than for *cyt b* sequences, indicating a higher rate of nucleotide substitution in the control region. The mean pairwise sequence divergence between North and South Island subspecies of robin (*P. australis*) is 5.9% ± 1.04 for *cyt b* and 10.34% ± 1.35 for the control region, but divergence within the North Island subspecies is only 0.24% ± 0.22 and 2.12% ± 0.04 , and within the South Island subspecies is only 0.39% ± 0.23 and 2.1% ± 0.43 for *cyt b* and control region sequences respectively. The divergence between South Island and Stewart Island individuals is similar to the level of divergence within South Island robins (*cyt b*: 0.35% ± 0.18 , cntrl reg: 2.2% ± 0.49). Among tomtit (*P. macrocephala*) subspecies the level of sequence divergence is similar to that within subspecies of *P. australis*

Table 6.2 Uncorrected pairwise sequence divergence (*p*, given as %) within and between ingroup taxa.

Comparison	Cyt <i>b</i>	Control region
Within <i>P. traversi</i>	0	0.3 \pm 0.3
Within <i>P. macrocephala</i>	0.6 \pm 0.2	2.0 \pm 0.4
Within <i>P. australis</i>	2.7 \pm 0.4	6.4 \pm 0.8
Within <i>P. multicolor</i>	4.0 \pm 0.9	2.7 \pm 0.8
<i>P. traversi</i> – <i>P. macrocephala</i>	4.1 \pm 0.8	7.3 \pm 1.2
<i>P. traversi</i> – <i>P. australis</i>	8.7 \pm 1.2	22.6 \pm 1.9
<i>P. macrocephala</i> – <i>P. australis</i>	9.2 \pm 1.2	21.7 \pm 1.8
<i>P. multicolor</i> – <i>P. australis</i>	10.6 \pm 1.3	23.0 \pm 1.9
<i>P. multicolor</i> – <i>P. macrocephala</i>	11.4 \pm 1.4	18.1 \pm 1.8
<i>P. multicolor</i> – <i>P. traversi</i>	10.8 \pm 1.4	18.2 \pm 1.8

(cyt *b*: 0.64% \pm 0.23, cntrl reg 2.7% \pm 0.25). Only the Chatham tomtit *P. m. chathamensis* appears to be slightly more divergent on the basis of cyt *b* sequences, with an average of 1.1% (\pm 0.5) divergence between it and other *P. macrocephala* subspecies. A single cyt *b* haplotype was found among 11 black robin (*P. traversi*) sequences, whereas control region sequences from the two museum specimens of *P. traversi* differ at one position (G \leftrightarrow A) from all extant individuals sampled.

For both control region and cyt *b* datasets, the relative rates for each substitution type were biased towards transitions. The degree of saturation within cyt *b* and control region sequences was investigated by plotting the number of transitions and transversions versus *p*-distances (Figure 6.3). Among cyt *b* sequences there was no evidence of saturation of either transitions or transversions. Similarly, there was no evidence of saturation among closely related control region sequences (*p*-distance up to 0.13). However there was evidence for saturation of transitions among distant comparisons, as Figure 6.3B shows a downward shift in the observed number of transitions when *p* > 0.15, relative to that observed for smaller *p*-distances. The mean pairwise transition/transversion ratio (R) was higher for cyt *b* sequences (7.5) than for control region sequences (2.1). However, control region R-values for within species comparisons are higher than for all comparisons (*P. macrocephala*, R = 4.0, *P. australis*, R = 6.7), providing additional support for saturation of transitions in distant comparisons.

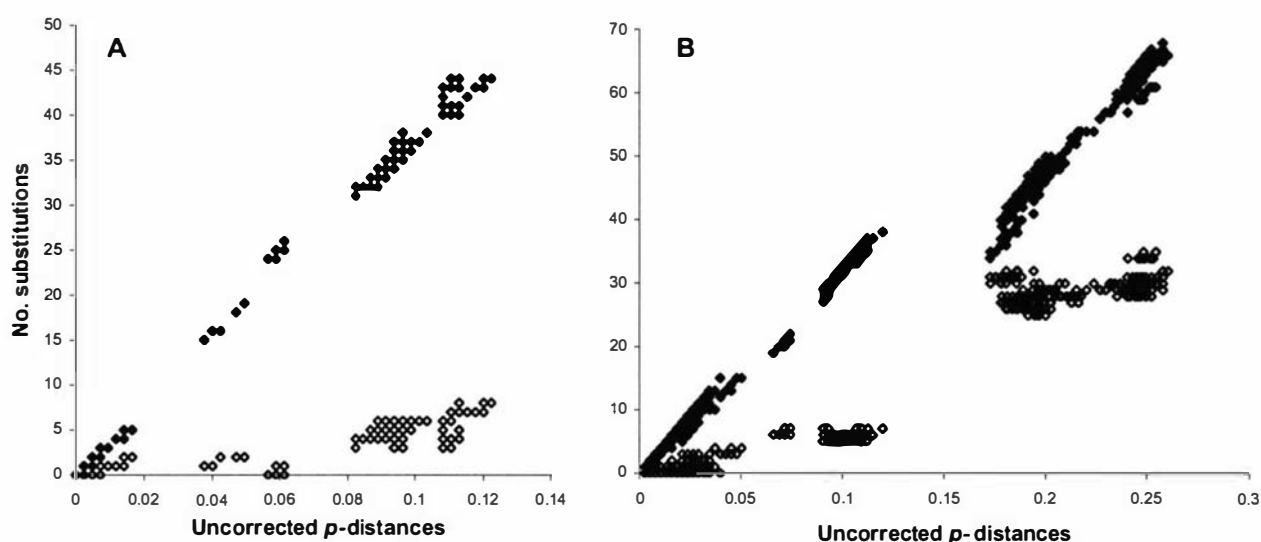


Figure 6.3 The observed number of transitions (filled diamonds) and transversions (open diamonds) versus pairwise sequence divergence for (A) cytochrome *b* and (B) control region sequences.

6.3.3 Phylogeny

The *cyt b* dataset was used to determine the overall relationships of the New Zealand Petroicidae species. An Australian member of the Petroicidae family *Eopsaltria australis* (yellow robin) was chosen as an outgroup. A neighbour-joining tree rooted with 2 non-Petroicidae species *Pomatostomus temporalis* and *Orthonyx temminckii* (Australo-Papuan babbler and logrunner, respectively) was initially constructed, which showed that the New Zealand and Pacific Petroicidae species form a monophyletic group to the exclusion of *E. australis* (data not shown), thus confirming the suitability of *E. australis* as an outgroup.

The maximum parsimony and maximum likelihood trees for the *cyt b* dataset are shown in Figure 6.4. The main difference between these trees is in the position of the scarlet robin *P. multicolor*, as in the ML tree it clusters with *P. australis*, while in the MP tree all the New Zealand *Petroica* taxa form a monophyletic group to the exclusion of *P. multicolor*. On all trees the black robin *P. traversi* clusters with *P. macrocephala* rather than with *P. australis*. The topology of the NJ tree constructed using the ML model is congruent with the ML tree, however the ME tree (strict consensus of 15 trees) is congruent with the MP topology (trees not shown). On all trees there is clear separation between North Island (*P. a. longipes*) and South Island robins (*P. a. australis*), and between Chatham Island tomtit and other *P. macrocephala* subspecies, but little resolution within or between other subspecies.

Control region sequences were used to investigate relationships within *P. australis* and *P. macrocephala* subspecies. Unrooted maximum likelihood trees for each species are presented in Figure 6.5. Neighbour-joining trees gave similar topologies (not shown), however maximum parsimony trees provided little phylogenetic information due to the large number of equally parsimonious trees (> 200). As with the *cyt b* sequences, the North Island and South Island subspecies of *P. australis* form distinct monophyletic clades with strong bootstrap support, but there is little evidence of geographical structuring within each subspecies. None of the internal nodes (except that defining the Kapiti Island individuals) have strong bootstrap support. The Stewart Island robin subspecies (*P. a. rakiura*) does not form a separate clade, instead sequences from this subspecies group within the South Island sequences. Among the *P. macrocephala* subspecies, distinct mitochondrial lineages with moderate bootstrap support separate North Island, South Island, Chatham Island and Snares Island subspecies. The branch lengths separating these *P. macrocephala* lineages are much

shorter than those separating North and South Island *P. australis* lineages, however. There is some indication of geographical structuring within *P. m. toitoi* as individuals from the Wellington region, and the eastern north island (Whirinaki and Gisborne) form separate clusters. The single control region sequence from the Auckland Island tit (*P. m. marrineri*) clusters with the North Island sequences. The placing of the Snares Island tit (*P. m. dannefaerdi*) in both maximum likelihood and neighbour-joining trees differs depending on whether the Tamura-Nei, or HKY models are used. Under the Tamura-Nei model, the South Island tit is paraphyletic with respect to the Snares Is (tree not shown), whereas under the HKY model the Snares Is tit falls outside the other subspecies.

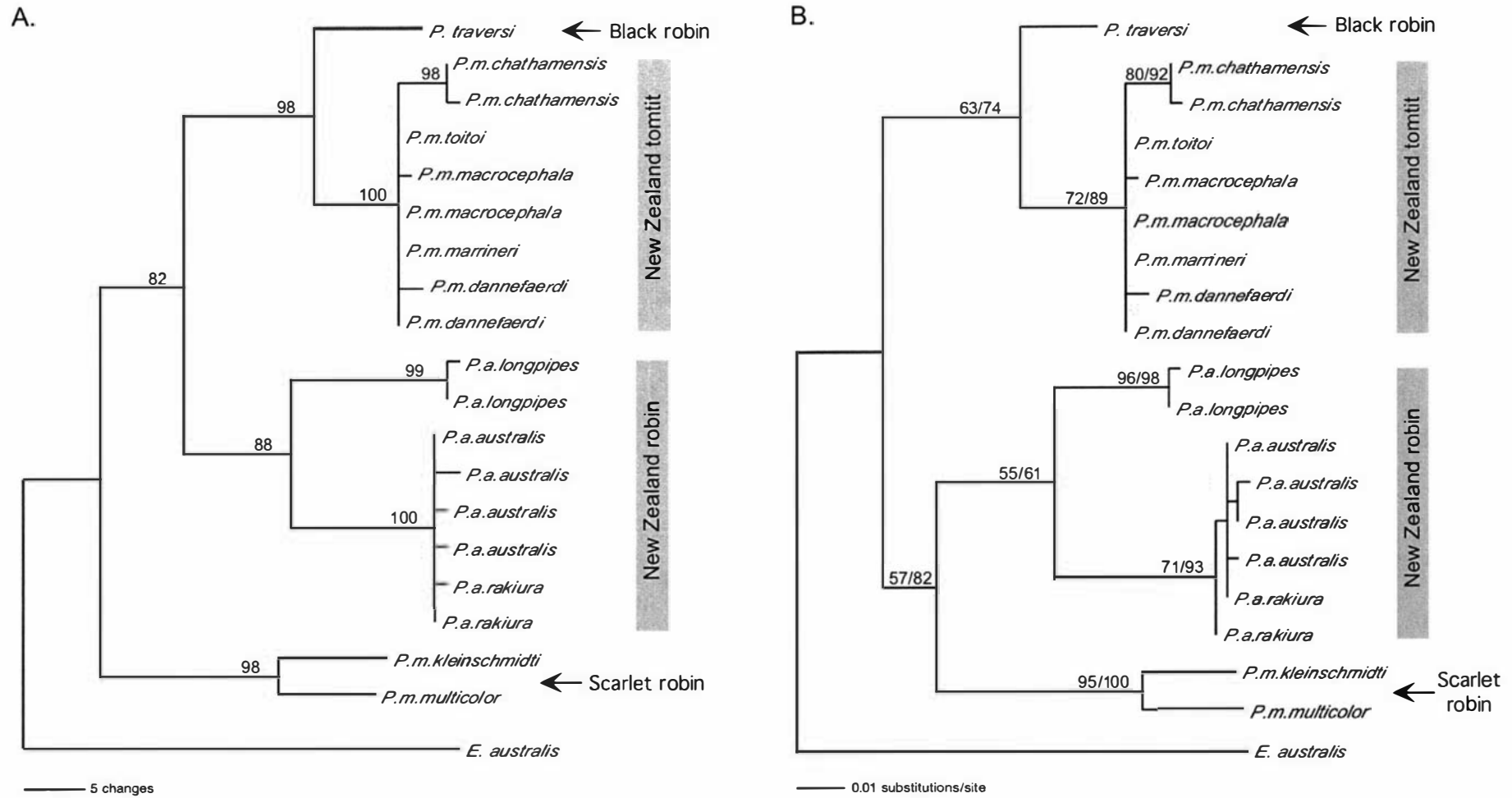


Figure 6.4 Relationships of New Zealand *Petroica* species, based on cytochrome *b* sequences (485 bp). Trees shown are (A) maximum parsimony and (B) maximum likelihood. Bootstrap values > 50% are shown above each branch. Both NJ (left) and ML (right) bootstrap values are shown on the ML tree.

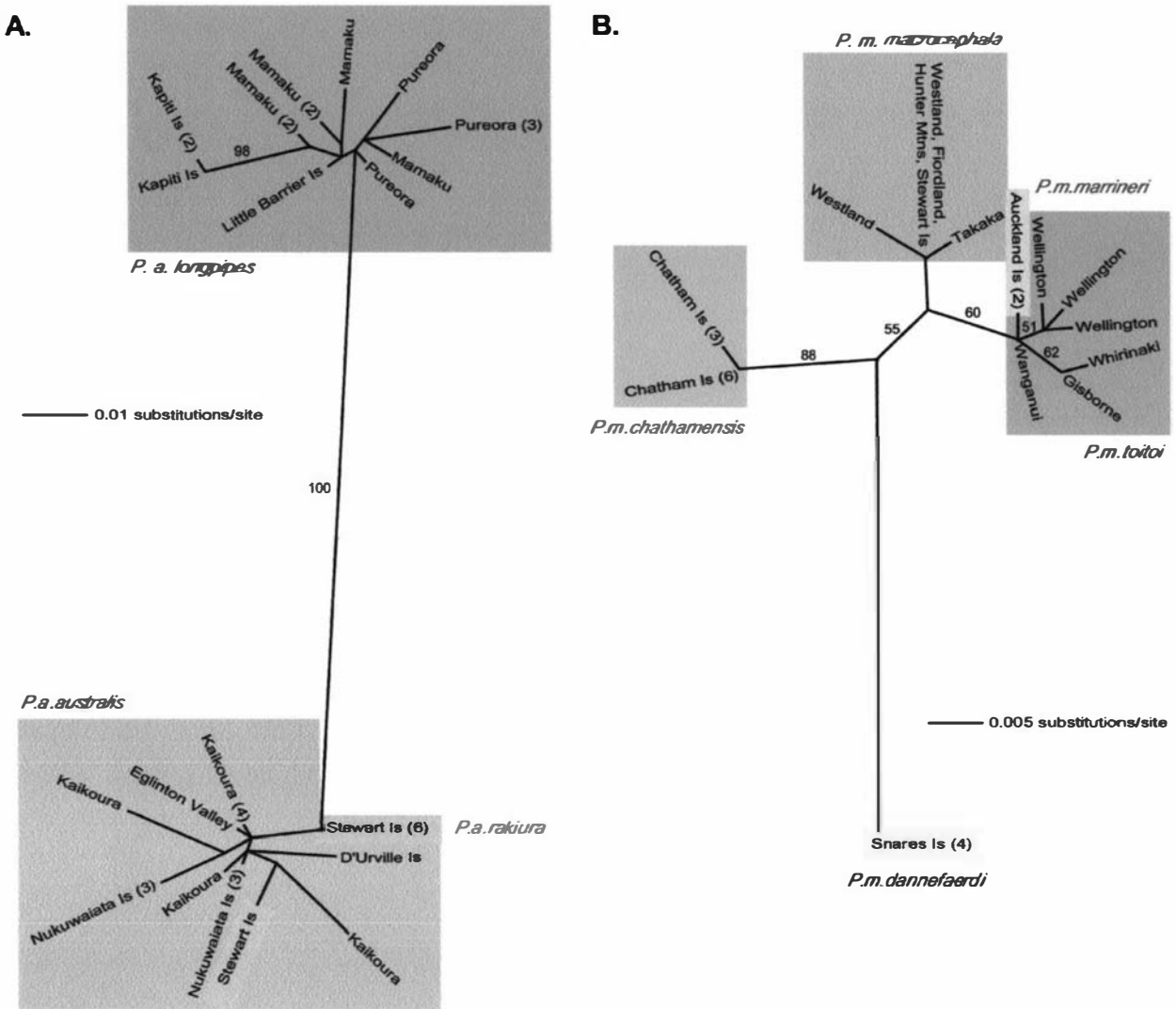


Figure 6.5 Relationships among (A) *Petroica australis* subspecies, and (B) *Petroica macrocephala* subspecies based on control region sequences. Trees shown are unrooted maximum likelihood trees with bootstrap values > 50% shown. The *P. macrocephala* tree shown was obtained using parameters estimated under the HKY model. Currently recognized subspecies (Fleming, 1950) are distinguished by coloured boxes, and sequences are named according to location (see Fig 6.1). Where the same sequence was found in more than one individual, the number of individuals is given in brackets.

6.4 Discussion

In this chapter, I have used mitochondrial DNA sequence data to investigate the phylogenetic relationships among the species of Petroicidae endemic to New Zealand. Cytochrome *b* sequences were used to determine the overall relationships among species, but provided little resolution at the subspecies level. Control region sequences provided higher resolution as a higher rate of nucleotide substitution was evident, and genetic distances within and among species were consistently higher (average 2.1 times higher for all pairwise comparisons) than for *cyt b* sequences. Because of the large genetic distances between species, and the likelihood of saturation between distant comparisons (Fig 6.2), analysis using control region sequences was restricted to within-species groups. Isolation of a fragment of the mitochondrial genome from *P. traversi* and *P. australis* in order to design control region primers revealed a rearrangement in gene order identical to that observed in several other avian taxa (Mindell et al. 1998, Bensch and Harlid 2000). This mitochondrial rearrangement is thought to have been independently derived from the ancestral arrangement on multiple occasions, and there is only one other example of this arrangement in an oscine passerine (Bensch and Harlid 2000).

6.4.1 Phylogenetic placement of the black robin

In general, the phylogenetic relationships estimated in this study support the current taxonomy (*sensu* Fleming 1950) of New Zealand's *Petroica* species. Distinct mitochondrial lineages separate the New Zealand tomtit *P. macrocephala*, the New Zealand robin *P. australis*, and the black robin *P. traversi*. The position of the black robin is anomalous however, as it clusters strongly with the tomtit, rather than the New Zealand robin on the *cyt b* tree. The same placement was also evident in the control region dataset as the genetic distance between black robin and tomtit lineages is far less than between black robin and New Zealand robin lineages (7.3% versus 22.6%, uncorrected *p*-distances). This placement is at odds with morphological and behavioural data, which strongly suggest that the affinities of the black robin lie with the New Zealand robin. New Zealand robins and tomtits are separated by distinct differences in overall size, wing length, tarsus length, song and foraging ecology. The black robin is superficially similar in colour and size to the Snares Is tit, and in measurements of tarsus/wing length ratio appears intermediate between tomtits and robins. However the length of the tarsus and first primary wing clearly group the black robin with

New Zealand robins (Fleming 1950). Similarly the song and foraging style of the black robin are clearly “robin-like” rather than “tit-like” (Fleming 1950, McLean et al. 1994).

Throughout mainland New Zealand each robin is sympatric with a tomtit, and subspecies of each appear to have evolved in allopatry. The hypothesis that the black robin is a Chatham Island derivative of the New Zealand robin, although separated by a greater gap than the mainland subspecies are to each other, is consistent with this pattern.

If the mitochondrial gene tree for the New Zealand *Petroica* taxa does reflect the true species tree, considerable morphological and behavioural convergence between the black robin and New Zealand robin must have occurred. The mitochondrial gene phylogeny appears to suggest that the ancestral *P. macrocephala* colonised the Chatham Islands twice with an early invasion giving rise to the black robin, and a later invasion of the Chatham tomtit. However, convergence between the black robin and New Zealand robin would have to have occurred on many disparate aspects of their morphology and behaviour, which appears unlikely. It is more likely that the gene tree does not reflect the species tree and that lineage sorting or hybridisation accounts for the anomalous position of the black robin. Ancient lineage sorting, i.e. the random segregation of polymorphic mitochondrial lineages present in the ancestor of the New Zealand *Petroica* species, could erroneously suggest that the black robin and New Zealand tomtit are sister taxa. However this would require all the New Zealand *Petroica* taxa to have originated from the same ancestral population, and it is not possible to confirm whether this is the case from our analyses (see below). In fact, it is possible that *P. australis* and *P. macrocephala* originated from separate invasions into New Zealand from separate ancestral populations.

An ancient hybridisation event and subsequent lineage sorting, resulting in introgression of the tomtit mitochondrial genome into the black robin, may be the best explanation for the anomalous position of the black robin. There is increasing evidence for the occurrence of hybridization between sympatric species, particularly in birds (Grant and Grant 1997, Allendorf et al. 2001), and several examples of anomalous mitochondrial gene trees in birds resulting from introgressive hybridization have been reported (e.g. Andersson 1999, Weckstein et al. 2001). The black robin and Chatham tomtit do have the ability to hybridize, as black robin nestlings cross-fostered to tomtit pairs in the 1980s have subsequently paired with tomtits and produced fertile, viable offspring (Ma and Lambert 1997). However, hybridisation does not appear to occur under natural conditions among the extant populations.

Morphological data are consistent with an early split of the black robin from the New Zealand robins, as the differences between black robin and New Zealand robins are much greater than the differences between North and South Island robin subspecies. Conversely, colonisation of the Chatham Islands by the tomtit appears to have occurred as part of a recent range expansion by *P. macrocephala*, as evidenced by the shallow branches on the *P. macrocephala* phylogeny (Figures 6.4 and 6.5B). Although the black robin lineage clearly groups with the *P. macrocephala* lineage, it falls outside this recent range expansion, and importantly, the black robin and Chatham tomtit are not sister taxa. This indicates that hybridization must have occurred with the ancestral *P. macrocephala*, possibly prior to the colonization of the Chatham Islands by the black robin. The original robin mitochondrial lineages may have been lost in the black robin by genetic drift during the process of colonization. Current differences between the black robin and tomtit lineages may be due to sequence divergence of the two lineages following hybridisation, or loss of the “black robin like” tomtit lineage in the tomtit. Alternatively, if hybridization did occur on the Chatham Islands, an early colonization of the Chathams by *P. macrocephala* must be postulated in addition to the recent colonization, with subsequent loss of this early *P. macrocephala* lineage in the extant Chatham tomtit population.

Analysis of nuclear markers would provide an independent estimate of the species phylogeny (Moore 1995) and may aid in resolving the relationship between the black robin and the other *Petroica* species. Based on the data presented here it is difficult to determine how closely related the black robin and New Zealand robins are, and to accurately date their divergence because of the apparent lack of congruence between the mitochondrial and species tree. However, if hybridisation between the ancestor of the black robin and the ancestral *P. macrocephala* did occur on the New Zealand mainland, this may give some idea of the date of the colonisation of the Chatham Islands, as the colonisation event cannot be older than the time to coalescence of the black robin and tomtit lineages. A poor fossil record for *Petroica* in New Zealand (Fleming 1950) means it is not possible to accurately calibrate divergence rates. However a rough approximation of divergence times can be estimated from the *cyt b* data, as it has been suggested that *cyt b* evolves at a similar rate (2% per million years (Myr)) in diverse avian taxa (Klicka and Zink 1997). This may be a more conservative estimate than rates from control region sequences, because of the possibility of saturation of control region sequences, and possible variation in control region divergence rates between taxa (Ruokonen

and Kvist 2002). Using the *cyt b* estimate, the 4.1% sequence divergence between black robin and tomtit lineages indicates a coalescence of 2 Myr ago, suggesting that colonisation of the Chatham Islands occurred within the last two million years. This approximate date is concordant with geological data, which suggests that the Chatham Islands were completely submerged four million years ago (Campbell 1998).

6.4.2 Origins of New Zealand's *Petroica* species

The *Petroica* genus is thought to have developed on the Australian continent and invaded New Zealand on two occasions. The first invasion, of the ancestral *P. australis*, may have occurred during the Pliocene, and a second invasion of the ancestral *P. macrocephala*, is thought to have occurred in the early Pleistocene (Fleming 1950). The morphological characteristics of *P. macrocephala* are considered to be closer to the Australian *Petroica* species, whereas *P. australis* appears to have a longer history of derivation, consistent with an earlier arrival. The fact that *P. australis* and *P. macrocephala* form distinct monophyletic clades in our phylogenetic analyses is consistent with the hypothesis of separate invasions. There is a deep divergence between the North Island and South Island lineages of *P. australis*, but not between the different island lineages of *P. macrocephala*, which suggests a longer history of separation of *P. australis* lineages, concordant with an earlier invasion. The North Island and South Island lineages of *P. australis* may have diverged prior to the Pleistocene, and represent the only *P. australis* lineages that survived the Pleistocene glacial cycles. An approximate estimate of divergence times (from *cyt b* data) of the North and South Island lineages of 3 Myr ago, is consistent with this hypothesis. The short branch lengths within the North and South Island *P. australis* clades are consistent with a post-Pleistocene range expansion into previously glaciated areas. In *P. macrocephala*, expansion of a single lineage appears to have occurred. This suggests that either *P. macrocephala* is a recent arrival and had not spread significantly throughout New Zealand prior to the onset of glaciation, or that all but one of the *P. macrocephala* lineages were lost during the Pleistocene glacial cycles. If the hypothesis of an ancient hybridisation event between the *P. traversi* and *P. macrocephala* is correct, the colonisation of New Zealand by *P. macrocephala* may have occurred at least 2 Myr ago.

Fleming (1950) suggested that *P. macrocephala* was derived from colonisation of the scarlet robin *P. multicolor*, which is found throughout the western Pacific. There was no evidence

for this in our data, however. ML and NJ trees suggest that the scarlet robin is in fact more closely related to *P. australis*, although this was not supported by the maximum parsimony or minimum evolution trees, which suggest that all New Zealand *Petroica* species form a monophyletic group to the exclusion of *P. multicolor*. Monophyly of the New Zealand *Petroica* species is not consistent with the hypothesis of 2 invasions (Emerson 2002). However it is not possible to determine whether the group actually is monophyletic, or to resolve the relationship between *P. multicolor* and the New Zealand taxa, without including all other related taxa from Australia in the analysis. It is apparent, however that *P. macrocephala* is not a recent derivative of *P. multicolor*, as the sequence divergence between these two species (11.4% for *cyt b*) is as large as the separation between *P. macrocephala* and *P. australis*.

6.4.3 Divergence within robin and tomtit radiations

Both *P. australis* and *P. macrocephala* have radiated to some degree within New Zealand, as different island forms can be distinguished on the basis of morphology. There are consistent differences in plumage and size separating the North Island and South Island subspecies of *P. australis* (described in detail in Fleming 1950), and Holdaway et al. (2001) suggested that these differences warranted elevation of the two subspecies to full species status. There is less to distinguish the Stewart Island subspecies, in fact in size and some plumage characteristics it appears more similar to the geographically more distant North Island robin, than the nearby South Island robin. The deep separation between the North Island and South Island *P. australis* clades in the mitochondrial DNA phylogeny does not contradict the suggestion they should be regarded as separate species, however there is little to distinguish the Stewart Island and South Island forms. Although no control region haplotypes were shared between the Stewart Island and South Island individuals, the Stewart Island haplotypes cluster within the South Island group. This suggests that the Stewart Island population is part of the recent radiation of South Island robins. The extent of gene flow (if any) between the South Island and Stewart Island populations cannot be determined from our data, as only a single individual from the southern South Island was haplotyped.

The morphological divergence between the five *P. macrocephala* subspecies is somewhat less marked than between the *P. australis* forms, with a more continuous series in many traits, although each island form can still be distinguished by size and plumage (Fleming

1950). Holdaway et al. (2001) named the North, South, Snares and Auckland Island forms as separate species on the basis of morphological differences, but kept the Chatham Island form as a subspecies of *P. macrocephala*, as it is morphologically similar to the South Island form. However, this study shows that each island subspecies is indistinguishable on the basis of *cyt b* data, with the exception of the Chatham Island lineage. Distinct lineages are present within the control region sequences, with the exception of the Auckland Island haplotype, which clusters within the geographically distant North Island lineage. The Auckland Island and Snares Island forms are virtually indistinguishable by morphology except for completely black plumage of the Snares Island form, which led Fleming to suggest that the Auckland Islands were colonised from the Snares Island population. Although there is nothing in our data to suggest this, incomplete lineage sorting during the recent radiation may have obscured the true relationships, erroneously suggesting that the Auckland Island and North Island forms are sister taxa. The short branch lengths among the control region *P. macrocephala* lineages suggest they are the result of a recent radiation, probably since the Pleistocene. The *P. macrocephala* forms are assumed to have evolved in allopatry, and there is evidence of genetic structuring and lack of gene flow between them, as would be expected between geographically isolated populations. This alone, however, does not warrant their elevation to full species status, as differences in morphology do not always correlate with genetic distance (as indicated by the Chatham Island tomtit *cyt b* data), and measures of genetic distance alone cannot be used to determine whether populations belong to separate species (Ferguson 2002).

In summary, the results presented in this chapter represent a preliminary phylogenetic analysis of the New Zealand Petroicidae species. In general, the mitochondrial DNA phylogeny supports Fleming's (1950) taxonomy and hypothesis of two separate invasions of the ancestral *Petroica*. However, additional data from Australian *Petroica* taxa are required to elucidate whether *P. australis* and *P. macrocephala* share a most recent common ancestor, or whether they are more distantly related, and originate from different ancestral populations. The placement of the black robin in the mitochondrial DNA phylogeny appears at odds with the species phylogeny, making it difficult to accurately measure the divergence of the black robin from the New Zealand robins. This highlights potential pitfalls of relying solely on mitochondrial DNA sequence data to infer evolutionary relationships, particularly among isolated island populations.

6.5 References

- Allendorf FW, Leary RF, Spruell P, Wenburg JK (2001) The problems with hybrids: setting conservation guidelines. *Trends Ecol. Evol* 16: 613-622
- Andersson M (1999) Hybridisation and skua phylogeny. *Proc. R. Soc. Lond. B.* 266: 1579-1585
- Baker AJ, Daugherty CH, Colbourne R, McLennan JL (1995) Flightless brown kiwis of New Zealand possess extremely subdivided population structure and cryptic species like small mammals. *Proc. Natl. Acad. Sci. USA* 92: 8254-8258
- Barker FK, Barrowclough GF, Groth JG (2002) A phylogenetic hypothesis for passerine birds: taxonomic and biogeographic implications of an analysis of nuclear DNA sequence data. *Proc. R. Soc. Lond. B.* 269: 295-308
- Bensch S and Harlid A (2000) Mitochondrial genomic rearrangements in songbirds. *Mol. Biol. Evol.* 17: 107-113
- Buller WL (1890) An exhibition of new and interesting forms of New Zealand birds, with remarks thereon. *Trans. NZ. Inst* 23: 36-43
- Butler D and Merton D (1992) *The Black Robin: Saving the World's Most Endangered Bird*. Oxford University Press
- Campbell HJ (1998) Fauna and flora of the Chatham Islands: Less than 4 my old? *Geogenes* March 1998: 15-16
- Christidis L and Schodde R (1991) Relationships of Australo-Papuan songbirds - protein evidence. *Ibis* 133: 277-285
- Desjardins P and Morais R (1990) Sequence and gene organization of the chicken mitochondrial genome - a novel gene order in higher vertebrates. *J. Mol. Biol.* 212: 599-634
- Emerson BC (2002) Evolution on oceanic islands: molecular phylogenetic approaches to understanding pattern and process. *Molecular Ecology* 11: 951-966
- Ericson PGP, Christidis L, Cooper A, Irestedt M, Jackson J, Johansson US, Norman JA (2002a) A Gondwanan origin of passerine birds supported by DNA sequences of the endemic New Zealand wrens. *Proc. R. Soc. Lond. B.* 269: 235-241
- Ericson PGP, Christidis L, Irestedt M, Norman JA (2002b) Systematic affinities of the lyrebirds (Passeriformes: *Menura*), with a novel classification of the major groups of passerine birds. *Mol. Phylogenet. Evol.* 25: 53-62
- Ferguson JWH (2002) On the use of genetic divergence for identifying species. *Biol. J.Lin. Soc.* 75: 509-516
- Fleming CA (1950) New Zealand flycatchers of the genus *Petroica* Swainson (Aves), Part I and II. *Trans. Roy. Soc. NZ* 78: 14-47, 126-160
- Grant PR and Grant BR (1997) Genetics and the origin of bird species. *Proc. Natl. Acad. Sci. USA* 94: 7768-7775
- Heather BD and Robertson HA (2000) *The Field Guide to the Birds of New Zealand* (revised edn). Viking, Auckland,
- Helm-Bychowski K and Cracraft J (1993) Recovering phylogenetic signal from DNA sequences: relationships within the Corvine assemblage (Class Aves) as inferred from complete sequences of the mitochondrial DNA cytochrome *b* gene. *Mol. Biol. Evol.* 10: 1196-1214

- Higgins PJ and Peter JM (2002) Handbook of Australian, New Zealand, and Antarctic birds. Oxford University Press,
- Holdaway RN, Worthy TH, Tennyson AJD (2001) A working list of breeding bird species of the New Zealand region at first human contact. *New Zealand J. Zool.* 28: 119-187
- Klicka J and Zink RM (1997) The importance of recent Ice Ages in speciation: A failed paradigm. *Science* 277: 1666-1669
- Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 17: 1244-1245
- Ma W and Lambert DM (1997) Minisatellite DNA markers reveal hybridisation between the endangered black robin and tomtit. *Electrophoresis* 18: 1682-1687
- Mathews GM and Iredale T (1913) A reference list of the birds of New Zealand. *Ibis* April 1913: 201-263 and July, 402-452
- McLean IG, Holzer C, Sagar PM (1994) Niche overlap and foraging ecology of island *Petroica* species. *Notornis* S41: 39-48
- Mindell DP, Sorenson MD, Dimcheff DE (1998) Multiple independent origins of mitochondrial gene order in birds. *Proc. Natl. Acad. Sci. USA.* 95: 10693-10697
- Moore WS (1995) Inferring phylogenies from mtDNA variation: mitochondrial gene trees versus nuclear gene trees. *Evolution* 49: 718-726
- Oliver WRB (1930) New Zealand birds. Fine Arts (NZ), Wellington
- Posada D and Crandall KA (1998) ModelTest: testing the model of DNA substitution. *Bioinformatics* 14: 817-818
- Ruokonen M and Kvist L (2002) Structure and evolution of the avian mitochondrial control region. *Mol. Phylogenet. Evol.* 23: 422-432
- Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- Sibley CG and Ahlquist JE (1990) *Phylogeny and classification of birds: a study in molecular evolution*. Yale University Press, New Haven & London
- Sibley CG and Monroe BL (1990) *Distribution and Taxonomy of the Birds of the World*. Yale University Press, New Haven, Connecticut
- Swofford DL (2002) *PAUP* Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Version 4. Sinauer Associates, Sunderland, MA
- Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* 22: 4673-4680
- Weckstein JD, Zink RM, Blackwell-Rago RC, Nelson DA (2001) Anomalous variation in mitochondrial genomes of white-crowned (*Zonotrichia leucophrys*) and golden-crowned (*Z. atricapilla*) sparrows: pseudogenes, hybridisation, or incomplete lineage sorting? *The Auk* 118: 231-236
- Worthy TH and Holdaway RN (2002) *The lost world of the moa: prehistoric life of New Zealand*. Indiana University Press, Bloomington, Indiana

CHAPTER SEVEN

Summary and Concluding Remarks

7.1 Introduction

In this thesis, I have analysed the genetics of the Major Histocompatibility Complex in a highly endangered species, the Chatham Islands black robin, and its non-endangered relative, the South Island robin. This study used source and bottlenecked populations of South Island robin as a comparison to the extremely bottlenecked black robin population, enabling the effect of population bottlenecks on MHC variation to be measured, and the significance of the low levels of variation in the black robin to be assessed. Characterisation of transcribed class II B MHC genes in these two species, and analysis of the key processes driving the evolution of these loci, was an important precursor to analysis of MHC variation, and a non-lethal protocol for isolation of transcribed sequences from blood was developed to achieve this. Use of two congeneric species in this study (whose relationship was investigated using mitochondrial DNA sequences) has facilitated a deeper understanding of the evolution and structure of the passerine MHC.

7.2 Summary of major findings

The major findings from the results of the preceding chapters can be summarised as follows:

1. Whole blood, when collected and stored appropriately, can be used as a source of RNA for isolation of transcribed MHC genes. Nearly full-length class II B MHC cDNA sequences could then be isolated using 3'RACE. This provides a non-lethal protocol for isolation of expressed MHC genes from endangered avian species, where use of conventional methods (such as cDNA library construction from spleen) may not be possible.

2. Using 3'RACE and RT-PCR, four transcribed class II B MHC sequences from black robin, and eight sequences from the South Island robin were isolated, suggesting there are at least four loci. RFLP analysis indicates that the class II B MHC genes are contained in a single linkage group. By analysis of the 3' untranslated regions of the cDNA sequences it was possible to identify an orthologous locus between the two species, as well as a second orthologous group of sequences containing multiple loci, suggesting multiple rounds of gene duplication have occurred within the MHC of New Zealand robins. However, these orthologous relationships are not retained within the coding region of the gene, and a number of putative gene conversion events were identified that may account for this. Exon 2 sequences are highly diverse and appear to be influenced by balancing selection, and gene conversion involving short stretches of sequences within exon 2 may add to this diversity. This is the first report of orthologous MHC loci in passerines, and provides further evidence that the evolution of the passerine MHC is characterised by high rates of gene duplication, concerted evolution, and balancing selection.
3. A partial genomic DNA sequence of an additional class II B MHC locus in black robin was isolated. This sequence was highly divergent all the class II B cDNA sequences isolated from black robin and South Island robin. Although the genomic DNA sequence appears to represent a functional locus, a large deletion apparent in the mRNA suggests it is a pseudogene.
4. Analysis of MHC variation using RFLP and sequence data indicates that the black robin is monomorphic at class II B loci, while both source and bottlenecked populations of South Island robin have retained moderate levels of variation. Comparison of MHC variation with minisatellite DNA variation indicates that the current forces determining MHC diversity in bottlenecked populations are neutral, however balancing selection and gene conversion appear to influence MHC diversity over evolutionary timescales. The orthologous loci identified in chapter three could not be distinguished on the basis of intron sequences, and homogenisation of regions immediately flanking exon 2 prevented the design of single locus primers for PBR sequence analysis.

5. Phylogenetic analyses using mitochondrial DNA sequences indicate that the black robin and New Zealand robin are not sister taxa, instead the black robin groups with the New Zealand tomtit. This is at odds with morphological and behavioural data, which indicates that the black robin is an early derivative of the New Zealand robin, but is consistent with ancient lineage sorting, morphological convergence, or introgressive hybridisation.

7.3 Future directions

7.3.1 *Does MHC monomorphism increase disease susceptibility in the black robin?*

Although low levels of MHC variation are thought to increase susceptibility to pathogens, it has been difficult to provide evidence for this in natural populations (e.g. Caro and Laurenson 1994, Gutierrez-Espeleta et al. 2001). Thus, the problem of how much MHC diversity is required to ensure long-term population viability remains a fundamental question in conservation genetics. The black robin population is ideal for investigating this question, as the results of this study indicate that it is monomorphic at class II MHC loci, yet the population appears to be viable under current conditions. The lack of MHC variation should be confirmed by analysis of more individuals, however, and diversity at class I loci should also be investigated. The degree of MHC diversity may vary between loci (e.g. the cotton-top tamarin has low class I diversity but extensive polymorphism at DRB loci (Gyllensten et al. 1994)), however this is unlikely to be the case in black robin given their population history. Analysis of expression of class II B sequences at the protein level is also required to confirm whether the four sequences all represent functional molecules expressed at high levels on the cell surface, and to investigate the range of peptides able to be presented by each individual. The pre-bottleneck levels of variation should also be investigated to assess the relevance of the current low levels of variation. Data from the South Island robin suggests that low genetic variation is not a general feature of this group, however the black robin may have had low levels of diversity since the initial colonisation of the Chatham Islands and not suffered a rapid decrease in variation through the recent bottleneck. Analysis of museum specimens may aid in determining the pre-bottleneck levels of variation, although most of the available specimens originate from the small remnant population on Little Mangere Island (A. Tennyson, pers comm.) so are likely to also show reduced levels of diversity. Analysis of

MHC diversity in the more abundant Chatham Island tomtit may provide a more accurate reflection of pre-bottleneck levels of diversity in the black robin.

A comprehensive analysis of pathogen exposure in the black robin population will be crucial to understanding the link between MHC diversity and population viability. Preliminary work has found little evidence for high pathogen loads in the black robin population (J. Anderson, pers comm). However this study needs to be extended to incorporate large sample sizes, and should also include analysis of pathogens in possible reservoir species, such as the sea birds, and other native and introduced passerines present in high densities on South East and Mangere Islands. An important question is whether the range of pathogens that the black robin population is exposed to is static, or whether the population is frequently exposed to new pathogens. The MHC alleles present in the black robin population may provide resistance to pathogens commonly encountered, but the ability to counter novel pathogens may be the key to long-term population viability. A long-term study of pathogen loads in the population will be required to assess whether the population is exposed to, and able to counter novel pathogens.

7.3.2 *Locus specificity in studies of the passerine MHC*

A major problem with MHC studies in passerines is distinguishing orthology from paralogy (e.g. alleles from loci), and this may hinder population level studies where allelic data from single loci are required. Because the passerine MHC appears to undergo high rates of gene conversion and/or gene duplication, loci within species share increased sequence similarity compared with loci from other species, particularly within coding regions (e.g. Edwards et al. 1995). Delineation of individual MHC loci and the design of single-locus primers for genotyping is something of a “holy grail” in passerine MHC studies, but has so far proved difficult because of this mode of evolution. PCR surveys (e.g. Vincek et al. 1997, Sato et al. 2000) are likely to be of little use in this regard. Vincek et al. and Sato et al. were able to distinguish four groups of class II B exon 2 sequences, and concluded that these groups represented different MHC loci. However they did not take the possibility of gene conversion into account, which may erroneously group sequences from different loci together. It has been suggested that introns are better indicators of loci, as they may be less subject to gene conversion (Hess and Edwards 2002). PCR primers amplifying single loci have been successfully designed by making use of sequence differences between loci in the introns

flanking exon 2 (e.g. Edwards et al. 2000, Gasper et al. 2001). However, it should be noted that this approach has so far been most successful for amplification of pseudogenes or non-polymorphic genes that appear to be considerably divergent from other loci. In this study, I found that introns were of limited use for distinguishing the multiple, expressed loci, as introns differed from each by length, rather than sequence, but intron length did not appear to be a reliable indicator of loci.

Nevertheless, genomic approaches, such as the construction of large-insert genomic libraries, are likely to provide the most success in unambiguous delineation of separate loci, and characterisation of differences between loci for PCR-typing. Isolation of microsatellites in tight linkage with MHC genes may facilitate studies of MHC variation (e.g. Paterson et al. 1998), as microsatellites may be easier to genotype than the MHC genes themselves. Large-scale MHC sequencing from genomic libraries has been performed in chicken, quail, red-winged blackbird and house finch. The chicken and quail MHCs have been sequenced in their entirety (Kaufman et al. 1999, Kulski et al. 2002), and partial genomic sequences are available from red-winged blackbird (Edwards et al. 1998, 2000, Gasper et al. 2001), and house finch (Hess et al. 2000). Studies of this nature from other songbird taxa (particularly the basal passerine lineages from Gondwana which are not well represented in studies to date) will allow a detailed comparative analysis of MHC organisation in passerines, and thus aid in understanding the key processes driving MHC evolution in birds. Use of closely related taxa are especially important in clarifying the timescales over which processes such as gene duplication and gene conversion occur. Bacterial Artificial Chromosome (BAC) libraries may be particularly useful, as they allow inserts in excess of 100 kb, which means large tracts of the MHC region may be found on a single vector (Graser et al. 1998). Further analysis of the patterns of expression of MHC genes is also required to distinguish functional from non-functional loci, and determine whether features of the minimal essential MHC in chickens also apply to other avian taxa. In conclusion, a greater understanding of the structure and evolution of the passerine MHC will facilitate the use of MHC genes in avian conservation genetics and behavioural ecology studies.

7.4 References

- Caro TM and Laurenson MK (1994) Ecological and genetic factors in conservation: a cautionary tale. *Science* 263: 485-486
- Edwards SV, Wakeland EK, Potts WK (1995) Contrasting histories of avian and mammalian *Mhc* genes revealed by class II B sequences from songbirds. *Proc. Natl. Acad. Sci. USA.* 92: 12200-12204
- Edwards SV, Gasper J, March M (1998) Genomics and polymorphism of *Agph-DAB1*, an MHC Class II B gene in red-winged blackbirds (*Agelaius phoeniceus*). *Mol. Biol. Evol.* 15: 236-250
- Edwards SV, Gasper J, Garrigan D, Martindale D, Koop BF (2000) A 39kb sequence around a blackbird *Mhc* Class II gene: Ghost of selection past and songbird genome architecture. *Mol. Biol. Evol.* 17: 1384-1395
- Gasper JS, Shiina T, Inoko H, Edwards SV (2001) Songbird genomics: Analysis of 45 kb upstream of a polymorphic *Mhc* class II gene in red-winged blackbirds (*Agelaius phoeniceus*). *Genomics* 75: 26-34
- Graser R, Vincek V, Takami K, Klein J (1998) Analysis of zebrafish *Mhc* using BAC clones. *Immunogenetics* 47: 318-325
- Gutierrez-Espeleta GA, Hedrick PW, Kalinowski ST, Garrigan D, Boyce WM (2001) Is the decline of bighorn sheep from infectious disease the result of low MHC variation? *Heredity* 86: 439-500
- Gyllensten U, Bergstrom TF, Josefsson A, Sundvall M, Savage A, Blumer ES, Giraldo LH, Soto LH, Watkins DI (1994) The cotton-top tamarin revisited: MHC class I polymorphism of wild tamarins, and polymorphism and allelic diversity of the class II DQA1, DQB1, and DRB1 loci. *Immunogenetics* 40: 167-176
- Hess CM, Gasper J, Hoekstra HE, Hill CE, Edwards SV (2000) MHC Class II pseudogene and genomic signature of a 32 kb cosmid in the house finch (*Carpodacus mexicanus*). *Genome Research* 10: 613-623
- Hess CM and Edwards SV (2002) The evolution of the major histocompatibility complex in birds. *BioScience* 52: 423-431
- Kaufman J, Milne S, Gobel TWF, Walker BA, Jacob JP, Auffray C, Zoorob R, Beck S (1999) The chicken B locus is a minimal essential major histocompatibility complex. *Nature* 401: 923-925
- Kulski JK, Shiina T, Anzai T, Kohara S, Inoko H (2002) Comparative genomic analysis of the MHC: the evolution of class I duplication blocks, diversity and complexity from shark to man. *Immunological Reviews* 190: 95-122
- Paterson S, Wilson K, Pemberton JM (1998) Major histocompatibility complex variation associated with juvenile survival and parasite resistance in a large unmanaged ungulate population (*Ovis aries L.*). *Proc. Natl. Acad. Sci. USA.* 95: 3714-3719
- Sato A, Figueroa F, Mayer WE, Grant PR, Grant BR, Klein J (2000) *Mhc* class II genes of Darwin's Finches: divergence by point mutations and reciprocal recombination. *In* M Kasahara (ed.): *Major Histocompatibility Complex: Evolution, Structure and Function*. Springer Verlag
- Vincek V, O'Huigin C, Satta Y, Takahata N, Boag PT, Grant PR, Grant BR, Klein J (1997) How large was the founding population of Darwin's finches? *Proc. R. Soc. Lond. B.* 264: 111-118

Appendix A

Class II B MHC peptide-binding region sequences

Figure A.1 Alignment of peptide-binding region (exon 2) sequences of class II B MHC alleles isolated from black robin (Petr) and South Island robin (Peau). The sequences were amplified with primers MHCBSF1 and MHCBSR2 and span 259 bp of exon 2 and 43 bp of intron 2.

```
Petr01 GTGGATGTTCAAGGGCGAGTGTCACTTCATTAACGGCACGGAGAAGGTGAGGTACGTGGTGAGGAACTTCTACAACCGGAGGAGTTAGT
Petr02 --T---G-A---TT-A-----G-----TG---A---C---C-----CC-
Petr03 -GA---GGA---AC-----TG---AC---T--A--C-----
Petr04 --T---GGA---TT-----G-----C---C---C-----A-CC-
Peau-1 -GA---AGC---TT-----TG---A-C---C-----A-T--
Peau-2 -GA---GGA---C-----C---TG---T-----G-----AC-C
Peau-3 -GA---GGA---C-----C---TG---T-----AC-C
Peau-4 -GA---GGA---C-----C---TG---T-----AC-C
Peau-5 -GA-----TT-----C---TG-----AC-C
Peau-6 -GA-----TT--A-----TG-----AC-C
Peau-7 -GA-----TT-----C---TG-----AC-C
Peau-8 -----TC-----G-----T-----T-----G---A--C-C--
Peau-9 -----TC-----G-----T-----T-----G---A--C-C--
Peau-10 -----TC-----G-----T-----T-----G-----C-C--
Peau-11 -----TC-----G-----T-----T-----G-----C-C--
Peau-12 -----TC-----G-----T-----T-----G-----C-C--
Peau-13 -----TC-----G-----T-----T-----G-----C-C--
Peau-14 -GA---AGGA-----TG---A---C-----ACT-
Peau-15 -GA---AGGA-----TG---A---C-----ACT-
Peau-16 -GA---AGGA-----TG---A---C-----ACT-
Peau-17 -----TC-----G-----T-----T-----G-----C-C--
Peau-18 -GA---GCA---TC-----C---T---AC---T-----AC-C
Peau-19 -GA---AGCT---TT-----TG---A-C---C-----A-T--
Peau-20 -----GCA---TT-----GG-----C-C--
Peau-21 -----GCA---TT-----C---TG-----GG-----C-C--
Peau-22 -GA---AGGA---T-----C---TG---A---C-----ACT-
Peau-23 -GA---G-G---C-----T-----ACT-
Peau-24 -GA---AGGA---TC-----TC-----ACT-
Peau-25 -GA-----TC-----TC-----AC-C
Peau-26 -GA-----TC-----TC-----AC-C
Peau-27 -GA---G-G---TG---CA--A-T-----C--
Peau-28 -----TC-----TG---AC---GG-----A-C-C
Peau-29 -A-----TC-----TG---AC---GG-----A-C-C
Peau-30 -A---GGA---TC-----C---T-T---AC---C-----C-C--
Peau-31 -GA---AGGA-----TG---A---C-----ACT-
Peau-32 -GA---AGGA-----TG---A---T-----ACT-
Peau-33 -GA---G-G---C-----TG-----AC-C
Peau-34 -GA-----TC-----C---TG-----C-C--
Peau-35 -----GCA---TT-----C---TG-----GG-----C-C--
Peau-36 -GAT--A-C---T-----TG-----A-C--
Peau-37 -GAT--A-C---T-----TG-----A-C--
Peau-38 -----GGA---C-----TG---A-----G-----ACT-
Peau-39 -GA---GGA---C-----C---T---AC---T-----AC-C
Peau-40 -GA-----TC-----TC-----AC-C
Peau-41 -GA---AGGA-----TG---A---T-----ACT-
```

```

Petr01 GAGGTTCCGACAGCGATGTGGGGCGCTATGTGGGGCTCACCCCCTTTGGGGAGAAAACAGGCCCGTACTGGAACAACAACCCGGCACTAAT
Petr02 -----A-----A-----GGT-----A-G-----G-G-----
Petr03 -----AAG-T-----A-----A-A-----G-G-----GA-C-
Petr04 -----AAG-T-----A-----A-----G-G-----GA-C-
Peau-1 -----C-----C-----T-----A-C-----GGC-----A-A--T-----G-G-----AC-C-
Peau-2 -----A-----A-----G-----G-----AG-C-
Peau-3 -----A-----A-----G-----G-----AG-C-
Peau-4 -----A-----A-----G-----G-----AG-C-
Peau-5 -----A-----G-----A-----T-G-----A-G-----
Peau-6 -----A-----G-----A-----T-G-----A-G-----
Peau-7 -----A-----G-----A-----T-G-----A-G-----
Peau-8 -----A-----A-----GGT-----T-A-----G-G-----GA-C-
Peau-9 -----A-----A-----GGT-----T-A-----G-G-----GA--
Peau-10 -----A-----A-----GGT-----T-A-----G-G-----GA-C-
Peau-11 -----A-----A-----GGT-----A-----G-G-----GA-C-
Peau-12 -----A-----A-----GGT-----T-A-----G-G-----GA-C-
Peau-13 -----A-----A-----GGT-----T-A-----G-G-----GA-C-
Peau-14 -----A-----A-----GGT-----T-A-----G-G-----AA-C-
Peau-15 -----A-----A-----GGT-----T-A-----G-G-----AA-C-
Peau-16 -----A-----A-----GGT-----T-A-----G-G-----AA-C-
Peau-17 -----A-----A-----G-----G-----AG-C-
Peau-18 -----A-----GGA-----GGT-----A-----G-----A-C-
Peau-19 -----C-----A-C-----GGC-----A--T-----G-----A-C-
Peau-20 -----A-C-----GGT-----A--T-----G-G-----GA--
Peau-21 -----A-----A-----GGT-----A--T-----G-G-----GA--
Peau-22 -----A-----C-----GGT-----T-A-----G-G-----GA--
Peau-23 -----A-----GGA-----T-G-----A-C-
Peau-24 -----A-----GGA-----G-G-----A--
Peau-25 -----A-----GGA-----G-----
Peau-26 -----A-----GGA-----G-----
Peau-27 -----C-----GGT-----A--T-----G-G-----AG-C-
Peau-28 -----A-----A-----GGT-----A--T-----G-----A-C-
Peau-29 -----A-----GG-----A--T-----G-----A-C-
Peau-30 -----A-----C-----A-----GGT-----A-----G-----A-C-
Peau-31 -----A-----A-----T-----GGT-----A-----G-G-----GA--
Peau-32 -----A-----A-----A-C-----GGT-----T-A-----G-G-----GA--
Peau-33 -----A-----A-----GGT-----G-----
Peau-34 -----A-----C-----G-G-----T-C-
Peau-35 -----A-----A-----T-----G-G-----GA--
Peau-36 -----A-----C-----CG-C-----GGC-----T-A--T-----G-----GA--
Peau-37 -----A-----CG-C-----GGC-----T-A--T-----G-----GA--
Peau-38 -----A-----GGT-----A-----T-G-----AG-C-
Peau-39 -----A-----GGT-----A-----T-G-----A-G-----
Peau-40 -----A-----A-----GGT-----A-----G-----GA--
Peau-41 -----A-----A-----A-C-----GGT-----T-A-----G-G-----AG-C-

```

└─ Intron 2

```

Petr01 GGAGCAGAGACGGGCAGAGGTGGACACGGTCTGCCGGCACAACCTGCAAGGTGTCCACCCCGTTACAGCGTGGAGCGCCGAGGTGAGCGCGG
Petr02 -----A-----G---CGA-----
Petr03 -----C-----G-TA-----T-----
Petr04 -----C-----A-C-C-----G-TA-----G-A-TG-----CT-----G-----
Peau-1 ---G-CGCT---AAT-C-----TA-----G-C-TG-----CT-----
Peau-2 ---GC-A---AAT-C-----A-----G---G-AT-----CT-----T-----
Peau-3 ---GC-A-----A-----A-----G-----CT-----T-----
Peau-4 ---GC-A-A-----A-----G-----CT-----T-----
Peau-5 ---GC-A-----A-----G-----T-----
Peau-6 ---GC-A-----A-----G-----T-----
Peau-7 ---GC-A-----A-----G-----T-----CT-----
Peau-8 ---C-----C---A-----A-----G---C-----CT-----
Peau-9 ---C-----C---A-----A-----G---C-----CT-----
Peau-10 ---C-----C---A-----A-----G---C-----CT-----
Peau-11 ---C-----C---A-----A-----G---C-----CT-----
Peau-12 ---C-----C---A-----A-----G---C-----CT-----
Peau-13 ---C-----C---A-----A-----G---C-----CT-----
Peau-14 ---GC-A---A-----A-----G---Y-----CT-----
Peau-15 ---GC-A-----A-----G-----T-----CT-----T-----
Peau-16 ---GC-A-----A-----G-----T-----CT-----
Peau-17 ---GC-A-----A-----G---G-AT-----CT-----T-----
Peau-18 ---GC-T---C-----G-TA-----G---TGA-G-----
Peau-19 ---G-CGCT---AAT-C-----TA-----G-C-TG-----CT-----
Peau-20 ---GC-A---C-----A-----G---TGT-----
Peau-21 ---GC-A---C-----TA-----G---TGT-----
Peau-22 ---GGC-A-----A-----G-----T-----CT-----
Peau-23 ---GC-T-----A-----G---G-AT-----T-----
Peau-24 ---T-C-A---A-T-----TA-----CT-----
Peau-25 ---GC-A-----A-----G-A-TGT-----T-----
Peau-26 ---GC-A-----A-----G-A-TGT-----CT-----
Peau-27 ---T-CCTT---AAT-C-----T-----G-----G-----CT-----
Peau-28 ---G-CGCT---AAT-C-----TA-----G-----CT-----
Peau-29 ---G-CGCT---AAT-C-----TA-----G-----CT-----
Peau-30 ---G-CGCT---AAT-C-----T-----G-----CT-----
Peau-31 ---GC-A-----G-----T-----
Peau-32 ---GC-A-----G-----T-----
Peau-33 ---GC-A-----A-----G---G-AT-----T-----
Peau-34 ---GGC-----C-----TA-----G-----T-----
Peau-35 ---GC-----TA-----T-----
Peau-36 ---GC-----TA-----G-----G-C-TG-----CT-----T-----
Peau-37 ---GC-----TA-----R-----G-C-TG-----CT-----T-----
Peau-38 ---T-CCTT---AAT-C-----T-----G-A-TGT-----
Peau-39 ---GC-----TA-----G-----CT-----
Peau-40 ---GC-A-----G-C-TG-----CT-----
Peau-41 ---GC-A-----A-----G-----CT-----T-----

```

```

Petr01 GGCAGAGCGAGGCCCTCGGGCCCTGTCCCGG
Petr02 -----
Petr03 -----G-----
Petr04 -----G-----
Peau-1 -----
Peau-2 -----
Peau-3 -----TC----
Peau-4 -----TC----
Peau-5 -----W-----
Peau-6 -----
Peau-7 -----A-----
Peau-8 -----
Peau-9 -----
Peau-10 -----
Peau-11 -----A-----
Peau-12 -----CT----
Peau-13 -----A-----
Peau-14 -----
Peau-15 -----A-----TC----
Peau-16 -----A-----TC----
Peau-17 -A-----TC----
Peau-18 -----C-----
Peau-19 -A-----
Peau-20 -----C-----
Peau-21 -----C-----
Peau-22 -----
Peau-23 -----
Peau-24 -----
Peau-25 -----T-----C----
Peau-26 -----C-----
Peau-27 -----T-----
Peau-28 -A-----C-----
Peau-29 -----C-----
Peau-30 -----T-----C----
Peau-31 -A-----C-----
Peau-32 -A-----C-----
Peau-33 -----
Peau-34 -----T-----T-----
Peau-35 -----C-----
Peau-36 -----
Peau-37 -----
Peau-38 -----T-----C----
Peau-39 -----C-----
Peau-40 -----
Peau-41 -----TC----

```

Appendix B

Petroica mitochondrial DNA sequences

Figure B.1 Sequence of 1371 bp of the mitochondrial DNA genome from *P. traversi* (BR) and *P. a. australis* (SIR). The sequence spans from cytochrome *b* to the control region, and was amplified using primers L15212 and CRDR1. The boundaries of tRNA^{Thr} and cytochrome *b* genes were determined by alignment with sequences from Mindell et al. (1998).

```
BR  AGTCATCCTTCTCCTAACCCATAGCAACTGCCTTTGTGGGCTACGTACTCCCATGAGGACAAATATCATTTTGAGGAGCTACAGTAAT
SIR  ...T.....C.....G.....T..C..A..T.....C..T.....G.....C.....C.....C.....

BR  CACCAACCTATCTCAGCCATCCCTTACATCGGCCAAACACTAGTAGAATGAGCCTGAGGAGGATTCTCAGTAGACAATCCCACTCTAAC
SIR  T..T.....C.....C.....C.....C.....C.....C.....C.....C.....

BR  TCGATTCTTCGCCCTGCACCTTCTCCTACCATTTCGTAATCGCAGGCCTCACACTAGTCCACCTCACCTTCTCCACGAAACAGTTCAAA
SIR  C.....T.....A.....T.....T.....T.....T.....T.....T.....T.....T.....

BR  CAATCCACTAGGCATTCCCTCAGACTGCGACAAAATCCCATCCACCATACTACTCAACAAAAGACATCCTAGGCTTTGTCTAATACT
SIR  ...C.....G.....T.....T.....T.....T.....T.....T.....T.....C.....

BR  CATTCCACTCGTCTCACTAGCTTTATCTCCCCAACTTACTAGGAGACCCAGAAAATTCACGCCGCCAACCCCTAGCCACACCCCC
SIR  ...C.....TA.....T.....C.....T.....C.....G.....T.....T.....T.....T.....

BR  CCATATCAAACCTGAATGATACTTCTATTTGCATACGCCATCCTACGATCCATCCCAAACAACTAGGAGGAGTACTAGCCCTAGCAGC
SIR  ...C.....C.....C.....C.....C.....C.....C.....C.....C.....C.....

BR  CTCGGTCTAGTCCTATTCTCATCCCCATACTCCACGTATCCAAACAACGCTCAATGACTTTCGCCCCACTCTCACAAATCCTATTCTG
SIR  .....T.....T.....T.....T.....T.....T.....T.....T.....C.....G.....C.....C.....

BR  AACCTAGTAACCAACCTCCTCATCCTAACCTGAGTCGGCAGCCAACCCAGTCGAACCCCATTCATCATCATTGGACAAGTAGCCTCATT
SIR  .....G.....T.....C.....A..T.....T.....T.....T.....T.....T.....T.....T.....

BR  TACCTATTTCCGCATCATCTCGTCCTATTTCCCTCGCGCCGCCTTAGAAAATAAACTACTAAAGCTCTAATTAACCTAATAGTTTA
SIR  C.....C...A..T.....T...A.....A..TA..T..TC.....C..G.....C.....C.....C.....

BR  TAAAAACATTGGTCTTGTAACCAAAGATTGAAGACTAAGCCCCTTCTTAGAGTTAACCT*TTCCCCCCCCCTCCCCCCCCATGCCCCAG
SIR  .....G.....C.....C.....C.....C.....C.....C.....GCTT.A.....C.....C.....C.....A...

BR  ATCCTGGCGCATTTCTAGGGTATGTATTACTTTGCATTACACTATTTACCCCATCAGACACTAATGTAATGTAGGATATCCAGACAACA
SIR  C*.T.....T.C..T.....AT.....G..T..C.....CCC....C.T....

BR  CCCAGCCTCTCTTCCCAGCCAAACCAACATCTTCTCCAATGAGATAATGTTTCGAGCTATACCCCTTCTAGGCACATTCCCACTTCAGGT
SIR  ...A.....T.A...C.T.....T.....GT...G.A..AC..TA...C.....AA.T.C.G...

BR  ACCATTAACCCAGGTAATCCTACCACAGGAC*ACGCAAGCGTAGCCAAGGTCGAAACCAGCCAATCGTACTAAAGAACCACCTCG*CAT
SIR  ..T.CA.....AC..C.....TTA...Y..A..G...CA.....A.....GA.....T.....ATG.A.C..CT.C..A..C

BR  ACGGATAATGTCACAGTACTCCTTTGAATGCCCAATGCATACAATTCGCCACCTCCTAGGTCGGATTCTTCTCCTACAGCCTTCAGGA
SIR  ...AC.....CT.CC.....TT.T...A.....G.CA.TT.....C.A.CA.A..C.C.....C.G.A.TC...A.C

BR  ACTCCAAGCCAGAGGACCTGGTTATCTATTAACCGTGTCTTCTCACGAGAACCGAGCTACTCAACGTCAGTTATACCCTTGTATTGGCT
SIR  .....A.....T...GGT...A.CC.....C.....C.....C.....C.A...

BR  TCAGGGACATACATTCCCCCG
SIR  .....A.....
```

Figure B.2 Alignment of cytochrome *b* sequences. Sequences are named according to their species and sample ID (BR = black robin, CIT = Chatham Island tomtit, NIT = North Island tomtit, SIT = South Island tomtit, AIT = Auckland Island Tomtit, SnIT = Snares Island tomtit, NIR = North Island robin, SIR = South Island robin, StIR = Stewart Island robin, ScR = Scarlet robin).

```

BR PT1      TTGGGGAGAATAAAGCTAGTGAGACGAGTGGAAATGAGTATTAGGACAAAGCCTAGGATGTCTTTTGTGAGTAGTATGGGTGGAATGGGA
CIT 1      -----C-----G-----G-----A-----
CIT 2      -----C-----G-----G-----A-----
NIT 26790   -----C-----G-----G-----A-----
SIT 22720   -----C-----G-----G-----A-----
SIT 24297   -----C-----G-----G-----A-----
AIT 13226   -----C-----G-----G-----A-----
SnIT 26238  -----C-----G-----G-----A-----
SnIT 23668  -----C-----G-----G-----A-----
NIR DP1     -----GG-----TA-----G-----C-----TG-----A-----A-----A-----
NIR 11Rot   -----GG-----TA-----G-----C-----TG-----A-----A-----A-----
SIR K3      -----A---G---A---TA---G-----G-----A-----A-----A-----
SIR C5      -----A---G---A---TA---G-----G-----A-----A-----A-----
SIR K5      -----A---G---A---TA---G-----G-----A-----A-----A-----
SIR K8      -----A---G---A---TA---G-----G-----A-----A-----A-----
StIR St3    -----A---G---A---TA---G-----G-----A-----A-----A-----
SrIR 8701   -----A---G---A---TA---G-----G-----A-----A-----A-----
ScR 19900   -----G-----T-AC-G-----AG-G-A---A-----AC-----
ScR 20177   -----GG-----T-AC-G-----AG-G-A---A-----AC-----

```

```

BR PT1      TTTTGTGCGAGTCTGAGGGAATGCCTAGTGGATTGTTTGAACCTGTTTCGTGGAGGAAGGTGAGGTGGACTAGTGTGAGGCCTGCGATTA
CIT 1      -----G-----G-----A-----
CIT 2      -----G-----G-----A-----
NIT 26790   -----G-----G-----A-----
SIT 22720   -----G-----G-----A-----
SIT 24297   -----G-----G-----A-----
AIT 13226   -----G-----G-----A-----
SnIT 26238  -----G-----G-----A-----
SnIT 23668  -----G-----G-----A-----
NIR DP1     -----A-----G-----G-----G-----A-----A-----A-----
NIR 11Rot   -----A-----G-----G-----G-----A-----A-----A-----
SIR K3      -----A-----C-----G-----A-----A-----
SIR C5      -----A-----G-----C-----G-----A-----A-----
SIR K5      -----A-----C-----G-----A-----A-----
SIR K8      -----A-----G-----C-----G-----A-----A-----
StIR St3    -----A-----C-----G-----A-----A-----
SrIR 8701   -----A-----C-----G-----A-----A-----
ScR 19900   -----G-----G-----G-----A-A-----A-C-----
ScR 20177   -----G-----G-----G-----A-A-A-----C-A-----A-----

```

```

BR PT1      CGAATGGTAGGAGGAAGTGCAGGGCGAAGAATCGAGTTAGAGTGGGATTGTCTACTGAGAATCCTCCTCAGGCTCAGGCTACTAGTGTTT
CIT 1      -----AA-----G-----A-----
CIT 2      -----AA-----G-----A-----
NIT 26790   -----A-----G-----C-----
SIT 22720   -----A-----G-----C-----
SIT 24297   -----A-----G-----C-----
AIT 13226   -----A-----G-----C-----
SnIT 26238  -----A-----T-G-----
SnIT 23668  -----A-----G-----
NIR DP1     -----A---T---A---G---G---G---G---C-----
NIR 11Rot   -----A---T---A---G---G---G---G---C-----
SIR K3      -----A---T---A---G---G---G---G-----
SIR C5      -----A---T---A---G---G---G---G-----
SIR K5      -----A---T---A---G---G---G---G-----
SIR K8      -----A---T---A---G---G---G---G-----
StIR St3    -----A---T---A---G---G---G---G-----
SrIR 8701   -----A---T---A---G---G---G---G-----
ScR 19900   -----A---A---T-----G---AG-A-G-----A-C-----A-----C-----
ScR 20177   -----A---A---T-----G---G-A-G-----A-----A-----G-----

```

```

BR PT1      GGCCGATGTAAGGGATGGCTGAGAATAGGTTGGTGATTACTGTAGCTCCTCAAATGATATTTGTCCTCATGGGAGTACGTAGCCACAA
CIT 1      -----G-----G-----A-----G-----
CIT 2      -----G-----G-----A-----G-----
NIT 26790  -----G-----G-----A-----G-----
SIT 22720  -----G-----G-----A-----G-----
SIT 24297  -----G-----G-----A-----G-----
AIT 13226  -----G-----G-----A-----G-----
SnIT 26238 -----G-C---G-----A-----G-----
SnIT 23668 -----G-----G-----A-----G-----
NIR DP1    -----G-----A---A---G-A---G-----A---A-T-G-
NIR 11Rot  -----G-----A---A---G-A---G-----A---A-T-G-
SIR K3     -----G-----A-A---G-G---G---C-----A-G---A-T-G-
SIR C5     -----G-----A-A---G-G---G---C-----A-G---A-T-G-
SIR K5     -----G-----A-A---G-G---G---C-----A-G---A-T-G-
SIR K8     -----G-----A-A---G-G---G---C-----A-G---A-T-G-
StIR St3   -----G-----A-A---G-G---G---C-----A-G---A-T-G-
SrIR 8701  -----G-----A-A---G-G---G---C-----A-G---A-T-G-
ScR 19900  -----A-G-----A-----A---G---G---G-----A-A-----T-G-
ScR 20177  -----A-G-----A-----A---G---G---G-----C-----A-----T-G-

```

```

BR PT1      AGGCAGTTGCTATGAGGGTTAGGAGAAGGATGACTCCGATGTTTCAGGTTTCTTTGTTAGGTAT
CIT 1      -----T-----T-----C-----A---A---
CIT 2      -----T-----T-----C-----C---A---A---
NIT 26790  -----T-----C-----A-----
SIT 22720  -----T-----C-----A-----
SIT 24297  -----T-----C-----A-----
AIT 13226  -----T-----C-----A-----
SnIT 26238 -----T-----C-----A-----
SnIT 23668 -----T-----C-----A-----
NIR DP1    -A-----A-G---A-----C-----
NIR 11Rot  -A-----A-G---A-----C-----
SIR K3     -A-----C---G---A---A-----
SIR C5     -A-----C---G---A---A-----
SIR K5     -A-----C---G---A---A-----
SIR K8     -A-----C---G---A---A-----
StIR St3   -A-----C---G---A---A-----
SrIR 8701  -A-----C---G---A---A-----
ScR 19900  -----A-----A-G---A---A-----
ScR 20177  -----A-G---A-----

```

Figure B.3 Alignment of control region sequences. Sequences are named according to their species and sample ID, as in Figure B.2

```

BR PT1      CATGCCCCAGATCCTGGCGCATTTCTAGGGTATGTATTACTTTGCATTACACTATTTACCCCATCAGACACTAATGTAATGTAGGATATT
BR 1693     -----
CIT PMC17  -----C-----T-----
CIT CIT4    -----C-----T-----
NIT 9695    -----C-----T-----
NIT 21785   -----C-----T-----
NIT 21786   -----C-----T-----
NIT 26790   -----C-----T-----
NIT 24841   -----C-----T-----
NIT 5207    -----C-----T-----
SIT 22870   -----C-----T-----
SIT 24297   -----C-----T-----
SIT 22720   -----C-----T-----
AIT 7943    -----C-----T-----
SnIT 23668  -----C-----T-----
NIR 9Tiri   --C---A--CCTTA---T-C--T-----C-----G--T--C-----CC--
NIR 24160   --C---A--CCTTA---T-C--T-----C-----G--T--C-----CC--
NIR 11Rot   --C---A--CCTTA---T-C--T-----C-----G--T--C-----CC--
NIR DP3     --C---A--CCTTA---T-C--T-----C-----G--T--C-----CC--
NIR 15Tiri  --C---A--C--TTA---T-C--T-----T-C-----G--T--C-----CC--
NIR Nstlg2  --C---A--CCTTA---T-C--T-----C-----G--T--C-----CC--
NIR 8Rot    --C---A--CCTTA---T-C--T-----CCT-----G--T--C-----CC--
NIR Nstlg1  --C---A--CCTTA---T-C--T-----C-----G--T--C-----CC--
NIR 22198   --C---A--CCTTA---T-C--T-----C-----G--TC-----CC--
NIR KWS2    --C---A--CCTTA---T-C--T-----C-----G--TC-----CC--
SIR K8      --C---A--CCTT:---T-C--T-----AC-----T-----G--T--C-----CC--
SIR C2      --C---A--CCTT:---T-C--T-----AT-----G--T--C-----CC--
SIR 22247   --C---A--CCTT:---T-C--T-----AT-----G--T--C-----CCC--
SIR K4      --C---A--CCTT:---T-C--T-----AT-----G--T--C-----CCC--
SIR C27a    --C---A--CCTT:---T-C--T-----AT-----G--T--C-----CCC--
SIR K10     --C---A--CCTT:---T-C--T-----AC-----G--T--C-----CCC--
SIR K5      --C---A--CCTT:-A--T-C--T-----C--GT-----G--T--C-----CCC--
SIR 5257    --C---A--CCTT:---T-C--T-----AT-----G--T--C-----CCC--
StIR St4    --C---A--CCTT:---T-C--T-----AC-----G--T--C-----CCC--
StIR 8701   --C---A--CCTT:---T-C--T-----C--AT-----G--T--C-----CCC--
ScR 19900   -GCTGATTGAC-T-A-CGCAC-A-----AC---C---GAC--
ScR 20177   -GCTGATTGAC-T-A-CGCAC-A-----AC---C---GAC--

```

```

BR PT1      CCAGACAACACCCAGCCTCTCTCCAGCCAAACCAACATCTTCTCCAATGAGATAATGTTTCGAGCTATACCCCTTCTAGGCACATTCC
BR 1693     -----A-----
CIT PMC17  --A-----AT---AC--T-----C-----TA--T-----
CIT CIT4    --A-----AT---AC--T-----C-----TA--T-----
NIT 9695    --AG-----AT---C--T-----C-----TA--T-----
NIT 21785   --AG-----AT---C--T-----C-----TA--T-----
NIT 21786   --AG-----AT---C--T-----C-----TA--T-----
NIT 26790   --AG-----AT---C--T-----C-----TA--T-----
NIT 24841   --AG-----AT---C--T-----C-----TA--T-----
NIT 5207    --AG-----AT---C--T-----C-----TA--T-----
SIT 22870   --AG-----AT---T--T-----C-----TA--T-----
SIT 24297   --AG-----AT---T--T-----C-----TA--T-----
SIT 22720   --AG-----AT---T--T-----C-----TA--T-----
AIT 7943    --AG-----AT---C--T-----C-----TA--T-----
SnIT 23668  --AG-----AT---TA-----C-----C--TA--T-----
NIR 9Tiri   --T-T-----A---TCA--T-C---G-----T-----GT--G-A--AC--TA---C-----A
NIR 24160   --T-T-----A---TCA--T-C---G-----T-----GT--G-A--AC--TA---C-C-----A
NIR 11Rot   --T-T-----A---TCA--T-C-----T-----GT--G-A--AC--TA---C-----A
NIR DP3     --T-T-----A---TCA--T-C-----T-----GT--G-A--AC--TA---C-----A
NIR 15Tiri  --T-T-----A---TCA--T-C---G-----T-----GT--G-A--AC--TA---C-----A
NIR Nstlg2  --T-T-----A---TCA--T-C-----T-----GT--G-A--AC--TA---C-----A
NIR 8Rot    --T-T-----A---TCA--T-C---G-----T-----GT--G-A--AC--TA---C-----A
NIR Nstlg1  --T-T-----A---TCA--T-C-----T-----GT--G-A--AC--TA---C-----A
NIR 22198   --T-T-----A---TCA--T-C-T-----T-----GT--G-A--AC--TA---C-----A
NIR KWS2    --T-T-----A---TCA--T-C-T-----T-----GT--G-A--AC--TA---C-----A
SIR K8      --C-T-----A---T-A---C-T-----T-----GT--G-A--AC--TA---C-----A
SIR C2      --C-T-----A---T-A---C-T-----T-----GT--G-A--AC--TA---C-----A
SIR 22247   --C-T-----A---T-A---C-T-----T-----GT--G-A--AC--TA---C-----A
SIR K4      --C-T-----A---T-A---C-T-----T-----GT--G-A--AC--TA---C-----A
SIR C27a    --C-T-----A---T-A---C-T-----T-----GT--G-A--AC--TA---C-----A
SIR K10     --C-T-----A---T-A---C-T-----T-----GT--G-A--AC--TA---C-----A
SIR K5      --C-T-----A---TCA---C-T-----T-----GT--G-A--AC--TA---C-----A
SIR 5257    --C-T-----A---T-A---C-T-----T-----GT--G-A--GC--TA---C-C-----A
StIR St4    --C-T-----A---T-A---C-T-----T-----GT--G-A--AC--TA---C-----A
StIR 8701   --C-T-----A---T-A---C-T-----T-----GT--G-A--AC--TA---C-----A
ScR 19900   --C-TC-----A---A---A---T-----T-----GT--G--A--AC--A--TC-----
ScR 20177   --C-TC-----A-T-----A--T-----T-----GT--G-----A-----

```

BR PT1 CACTTCAGGTACCATTAACCCAGGTAATCCTACCAC : AGGACACG : CAAGCGTAGCCCAAGGTCGAAACCAGCCAATCGTACTAAAGAAC
BR 1693 -----
CIT PMC17 -----A--A-----TA-----G-----C-----AG--
CIT CIT4 -----A--A-----TA-----A-----G-----C-----AG--
NIT 9695 -----A-----TA-----G-T-----C-----AG--
NIT 21785 -----T--A-----TA-----G-T-----C-----AG--
NIT 21786 -----T--A-----TA-----G-T-----C-----AG--
NIT 26790 -----A-----TA-----G-T-----C-----G--AG--
NIT 24841 -----A-----TA-----G-T-----G--AG--
NIT 5207 -----AT--A-----TA-----G-T-----C-----AG--
SIT 22870 -----A--A-----TA-----T-----C-----AG--
SIT 24297 -----A--A-----TA-----T-----C-----AG--
SIT 22720 -----A--A-----TA-----T-----C-----AGG--
AIT 7943 -----A-----TA-----G-T-----C-----AG--
SnIT 23668 -----A--G-----A--A-----TA--G--A--G-----C-----AGG--
NIR 9Tiri A-T-C-G-----A-----A-----AC--AA-----CA-----A-----GA-----C-----ACC--
NIR 24160 A-T-C-G-----A-----A-----AC--AA-----CA-----A-----GA-----C-----G-ACC--
NIR 11Rot A-T-C-G-----A-----A-----AC--AA-----TA-----AC-----GA-----C-----G-ACC--
NIR DP3 A-T-C-----A-----A-----AC--AA-----CA-----A-----GA-----C-----G-ACC--
NIR 15Tiri A-T-C-G-----A-----A-----AC--AA-----CA-----A-----GA-----C-----G-ACC--
NIR Nstlg2 A-T-C-----A-----A-----AC--AA-----CA-----A-----GA-----C-----G-ACC--
NIR 8Rot A-T-C-G-----A-----A-----AC--AA-----CA-----A-----A-----C-----G-ACC--
NIR Nstlg1 A-T-C-G-----A-----A-----AC--AA-----CA-----A-----GA-----C-----G-ACC--
NIR 22198 A-T-C-G--N-----A-G-----A-----AC--AA-----CA-----A-----GA-----C-T-----G-ACC--
NIR KWS2 A-T-C-G-----A-G-----A-----AC--AA-----CA-----A-----GA-----C-----G-ACC--
SIR K8 A-T-C-G-----T-CA-----AC--C-----TT--AC--A-----C--G--A-----GA-----T-----ATG-A-C
SIR C2 A-T-C-G-----T-CA-----AC--C-----TT--AC--A-----CA--G--A-----GA-----TC-----AT--A-C
SIR 22247 A-T-C-G-----T-CA-----AC--C-----TT--AC--A--G-----CA--G--A-----GA-----T-----ATG-A-C
SIR K4 A-T-C-G-----T-CA-----AC--C-----TT--AC--A--G-----CA--G--A-----GA-----T-----ATG-A-C
SIR C27a A-T-C-G-----T-CA-----AC--C-----TT--ACT--A--G-----CA-----A-----GA-----T-----ATG-A-C
SIR K10 A-T-C-G-----T-CA-----AC--C-----TT--ACT--A--G-----CA--G--A-----GA-----T-----ATG-A-C
SIR K5 A-T-C-G-----T-CA-----AC--C-----TT--ACT--A-----CA--G--A-----GA-----T-----ATG-A-C
SIR 5257 A-T-C-G-----T-CA-----AC--C-----TT--ACT--A--G-----CA--G--A-----GA-----T-----ATG-A-C
StIR St4 A-T-C-G-----T-CA-----AC--C-----TT--AC--A--G-----CA--G--A-----GA-----T-----ATG-A-C
StIR 8701 A-T-C-G-----T--A-----AC--C-----T--ACT--A-----CA--G--A-----GA-----T-----ATG-A-C
ScR 19900 -----A-----AGAACA-----CA--G--A-----GAT-----GT-----C-G-ACC--
ScR 20177 -----A-----GGAACA--A--G--C--G--A-----GAT-----T-----C-G-ACT--

BR PT1 CACCTCG : CACACGGATAATGTCACAGTACTCCTTTGAATGCCCAATGCATACAATTCGCCACCTCCTAGGTCGGATTCTTCTCCTAC
BR 1693 -----
CIT PMC17 -----CT-----AC-----TC-----
CIT CIT4 -----CT-----AC-----TC-----
NIT 9695 -----A-----AC-----A-----TC-----
NIT 21785 -----A-----AC-----G--A-----TC-----
NIT 21786 -----A-----AC-----A-----TC-----
NIT 26790 -----G-----AC-----A-----TC-----
NIT 24841 -----G-----AC-----A-----TC-----
NIT 5207 -----A-----AC-----A-----TC-----
SIT 22870 -----A-----AC-----A-----TC-----
SIT 24297 -----A-----AC-----C-----A-----TC-----
SIT 22720 -----A-----AC-----A-----C-----TC-----
AIT 7943 -----A-----AC-----A-----G-----TC-----
SnIT 23668 --:--C-TG-----AC-----A-----T-----TC-----
NIR 9Tiri -C::C--A--C--AC-----A-----CA-T-----A--A-AG-C--C--C-A
NIR 24160 -C::C--A--C--AC-----A-----CA-T-----A--A-AG-C-----CGA
NIR 11Rot -C::C--A--C--AC-----A-----CA-T-----A--A-AG-C-----CGA
NIR DP3 -C::C--A--C--AC-----A-----CA-T-----C-A-AAA-C-----CGA
NIR 15Tiri -C::C--A--C--AC-----A-----CA-T-----A--A-AG-C-----CGA
NIR Nstlg2 -C::C--A--C--AC-----A-----CA-T-----A--A-AG-C-----CGA
NIR 8Rot -C::C--A--C--AC-----A-----CA-T-----A--A-AG-C-----CGA
NIR Nstlg1 -C::CAAA--C--AC--C-----T--A-----CA-T-----A--A-AG-C--C--CGA
NIR 22198 -C::C--AT-C--AC-----A-----CA-T-----A--A-AG-C--C--C-A
NIR KWS2 -C::C--AT-C--AC-----A-----CA-T-----A--A-AG-C--C--C-A
SIR K8 -C::C--AA--C--AC-----CT-CC-----TT-T--A-----G-CA-TT-----C-A-CA--C--C--CGG
SIR C2 -T::C--A--C--AC-----CT-CC-----TT-T--A-----G-CA-TT-----C-A-CA--C--C--C-G
SIR 22247 -CT-C--A--C--AC-----CT-CC-----TT--A-----G-CA-TT-----C-A-CA--A--C--C--C-G
SIR K4 -CT-C--A--C--AC-----CT-CC-----TT-T--A-----G-CA-TT-----C-A-CA--A--C--C--C-G
SIR C27a -CT-C--A--C--AC-----CT-CC-----TT-T--A-----G-CA-TT-----C-A-CA--A--C--C--C-G
SIR K10 -CT-C--A--C--AC-----CT-CC-----TT-T--A-----G-CA-T-----C-A-CA--A--C--C--C-G
SIR K5 -CT-C--A--C--AC-----CT-CC-----TT--A-----G-CAATT-----C-A-CA--A--C--C--C-G
SIR 5257 -CT-C--AA-C--AC-----CT-CC-----TT--A-----G-CAATT-----C-A-CA--A--C--C--C-G
StIR St4 -CT-C--A--GC--AC-----CT-CC-----TT-T--A-----CA-TT-----C-A-CA--A--C--C--C-G
StIR 8701 -CT-C--A--GC--AC-----CT-CC-----TT-T--A-----G-CA-TT-----C-A-CA--A--C--C--C-G
ScR 19900 -----C-----AC-----A-----TTA--GTC-T-----
ScR 20177 -----C-----AC-----G-----A-----TTA--GTC-T-----

```

BR PT1      AGCCTTCAGGAACTCCCAAGC
BR 1693     -----
CIT PMC17   -----A-----
CIT CIT4    -----A-----
NIT 9695    -----A-----
NIT 21785   -----A-----
NIT 21786   -----A-----
NIT 26790   -----A-----
NIT 24841   -----A-----
NIT 5207    -----A-----
SIT 22870   -----A-----
SIT 24297   -----A-----
SIT 22720   -----A-----
AIT 7943    -----A-----
SnIT 23668  -----A-----
NIR 9Tiri   -A-T---A-T-----
NIR 24160   -A-T---A-T-----
NIR 11Rot   -A-T---A-T-----
NIR DP3     -A-T---A-T-----
NIR 15Tiri  -A-T---A-T-----
NIR Nstlg2  -A-T---A-T-----
NIR 8Rot    -A-T---A-T-----
NIR Nstlg1  -A-T---A-T-----
NIR 22198   -A-T---A-T-----
NIR KWS2    -A-T---A-T-----
SIR K8      -A-TC---A-C-----
SIR C2      -A-TC---A-C-----
SIR 22247   -A-TC---A-C-----
SIR K4      GA-TC---A-C-----
SIR C27a    -A-TC---A-C-----
SIR K10     -A-TC---A-C-----
SIR K5      -A-TC---A-C-----
SIR 5257    -A-TC---A-C-----
StIR St4    -A-TC---A-T-----
StIR 8701   -A-TC---A-C-----
ScR 19900   -----A-C-----
ScR 20177   -----A-C-----

```

Appendix C:

Details of *Petroica* samples used in this study

Sample ID	Species	Individual ID	Location	Date collected	Sample type	Chapter ¹
BR2	<i>P.traversi</i>	B67510 (Keith)	South East Is	March 1992	Blood	2, 4a, 5
BR7	<i>P.traversi</i>	B67507 (Fredda)	South East Is	March 1992	Blood	4a, 5
BR8	<i>P.traversi</i>	B41499 (Ted)	South East Is	March 1992	Blood	2, 4a
BR10	<i>P.traversi</i>	B66161 (Magic)	South East Is	March 1992	Blood	4a
BR14	<i>P.traversi</i>	B41455	South East Is	March 1992	Blood	2, 4a, 5
BR28	<i>P.traversi</i>	B41445	South East Is	March 1992	Blood	2
BR30	<i>P.traversi</i>	B66188 (Skimmer)	South East Is	March 1992	Blood	2, 4a
BR38	<i>P.traversi</i>	B44846 (Elsa)	South East Is	March 1992	Blood	2, 4a, 5
BR40	<i>P.traversi</i>	B43218 (Bill)	South East Is	April 1992	Blood	4a
PT1	<i>P.traversi</i>	B81125 (Papagato)	South East Is	Jan 2001	Blood	2, 5
PT3	<i>P.traversi</i>	B80654 (Casey)	South East Is	Jan 2001	Blood	2, 3, 4b, 5
PT4	<i>P.traversi</i>	B79743 (Allan)	South East Is	Jan 2001	Blood	5
PT8	<i>P.traversi</i>	B79714 (Mango)	South East Is	Feb 2002	Blood	5
BR75	<i>P.traversi</i>	B68046 (Hariroa)	Mangere Is	Feb 1993	Blood	5
BR87	<i>P.traversi</i>	B68043 (Old Heck)	Mangere Is	Feb 1993	Blood	5
DM1687	<i>P.traversi</i>	(Museum sample)	Little Mangere Is	1871	Toe pad	2
DM1688	<i>P.traversi</i>	(Museum sample)	Little Mangere Is	1897	Toe pad	2
DM1693	<i>P.traversi</i>	(Museum sample)	Little Mangere Is	1900	Toe pad	2
K2	<i>P.a.australis</i>		Kaikoura	Sept 93	Blood	2, 4a
K3	<i>P.a.australis</i>	40737	Kaikoura	Sept 93	Blood	2, 4a
K4	<i>P.a.australis</i>	40738	Kaikoura	Sept 93	Blood	2, 4a
K5	<i>P.a.australis</i>	40739	Kaikoura	Sept 93	Blood	2, 4a, 4b
K7	<i>P.a.australis</i>	40741	Kaikoura	Sept 93	Blood	2, 4a
K8	<i>P.a.australis</i>		Kaikoura	Oct 93	Blood	2, 4a
K10	<i>P.a.australis</i>		Kaikoura	Oct 93	Blood	2, 4a
K11	<i>P.a.australis</i>		Kaikoura	Oct 93	Blood	2, 4a
A2	<i>P.a.australis</i>		Allports Is	Feb 92	Blood	4a
A5	<i>P.a.australis</i>		Allports Is	March 92	Blood	4a
A7	<i>P.a.australis</i>		Allports Is	March 92	Blood	4a
A12	<i>P.a.australis</i>		Allports Is	March 92	Blood	4a
A19	<i>P.a.australis</i>		Allports Is	March 92	Blood	4a
A23	<i>P.a.australis</i>		Allports Is	March 92	Blood	4a
A26	<i>P.a.australis</i>		Allports Is	March 92	Blood	4a
A31	<i>P.a.australis</i>		Allports Is	March 92	Blood	4a
M3	<i>P.a.australis</i>		Motuara Is	Feb 92	Blood	4a, 5
M6	<i>P.a.australis</i>		Motuara Is	Feb 92	Blood	4a, 5
M10	<i>P.a.australis</i>		Motuara Is	Feb 92	Blood	4a, 5

¹ Chapter refers to the analysis the sample was used for: 2 - Phylogenetics; 3 - Test of blood preservation; 4a - RFLP; 4b - Characterisation of cDNA; 5 - Analysis of MHC variation

Sample ID	Species	Individual ID	Location	Date collected	Sample type	Chapter ¹
M12	<i>P.a.australis</i>		Motuara Is	Feb 92	Blood	4a, 5
M13	<i>P.a.australis</i>		Motuara Is	Feb 92	Blood	4a, 5
M16	<i>P.a.australis</i>		Motuara Is	Feb 92	Blood	4a, 5
M22	<i>P.a.australis</i>		Motuara Is	Feb 92	Blood	4a, 5
M26	<i>P.a.australis</i>		Motuara Is	Feb 92	Blood	4a, 5
C2	<i>P.a.australis</i>	B72682	Nukuwaiata Is	May 93	Blood	2, 4a, 5
C5	<i>P.a.australis</i>	B72686	Nukuwaiata Is	May 93	Blood	2, 4a, 5
C7	<i>P.a.australis</i>	B72688	Nukuwaiata Is	May 93	Blood	4a, 5
C8	<i>P.a.australis</i>	B72689	Nukuwaiata Is	May 93	Blood	2, 4a, 5
C16	<i>P.a.australis</i>	B72670	Nukuwaiata Is	May 93	Blood	2, 4a, 5
C23	<i>P.a.australis</i>	B72677	Nukuwaiata Is	May 93	Blood	2, 4a, 5
C26	<i>P.a.australis</i>	B72680	Nukuwaiata Is	May 93	Blood	4a, 5
C27a	<i>P.a.australis</i>	B72681	Nukuwaiata Is	May 93	Blood	2, 4a, 5
DM5257	<i>P.a.australis</i>	(Museum sample)	D'Urville Is	1905	Toe pad	2
NM22247	<i>P.a.australis</i>	(Museum sample)	Eglinton Valley	1980	Toe pad	2
8Rot	<i>P.a.longpipes</i>	B72692	Mamaku Forest	June 1993	Blood	2
11Rot	<i>P.a.longpipes</i>	B72695	Mamaku Forest	June 1993	Blood	2
9Tiri	<i>P.a.longpipes</i>	C48690	Mamaku Forest	Nov 1992	Blood	2
15Tiri	<i>P.a.longpipes</i>	C48665	Mamaku Forest	Nov 1992	Blood	2
18Tiri	<i>P.a.longpipes</i>	C13699	Mamaku Forest	Nov 1992	Blood	2
22Tiri	<i>P.a.longpipes</i>	C48699	Mamaku Forest	Nov 1992	Blood	2
DP1	<i>P.a.longpipes</i>	Clutch 1	Pureora Forest	2002	Feather	2
DP2	<i>P.a.longpipes</i>	Clutch 2	Pureora Forest	2002	Feather	2
DP3	<i>P.a.longpipes</i>	Dead adult	Pureora Forest	2002	Feather	2
Nestling 1	<i>P.a.longpipes</i>	B80137 (Gavin and Tessa)	Pureora Forest (Tahae)	Dec 1997	Feather	2
Nestling 2	<i>P.a.longpipes</i>	B80147 (Smarty and Nina)	Pureora Forest (Tahae)	Dec 1997	Feather	2
NIR-KWS2	<i>P.a.longpipes</i>	B83995	Kapiti Is	May 2002	Feather	2
NIR-KWS3	<i>P.a.longpipes</i>	B83986	Kapiti Is	May 2002	Feather	2
NM22198	<i>P.a.longpipes</i>	(Museum sample)	Kapiti Is	1976	Toe pad	2
NM24160	<i>P.a.longpipes</i>	(Museum sample)	Little Barrier Is	1990	Toe pad	2
St1	<i>P.a.rakiura</i>	B81985 BG-WM	Stewart Is	2002	Feather	2
St2	<i>P.a.rakiura</i>	B81976 WM-GG	Stewart Is	2002	Feather	2
St3	<i>P.a.rakiura</i>	B81977 WM-GR	Ulva Is	2002	Feather	2
St4	<i>P.a.rakiura</i>	B81982 GR-WM	Stewart Is	2002	Feather	2
St5	<i>P.a.rakiura</i>	B81987 WG-WM	Stewart Is	2002	Feather	2
St6	<i>P.a.rakiura</i>	B81991 WM-YB	Ulva Is	2002	Feather	2
DM8701	<i>P.a.rakiura</i>	(Museum sample)	Pouawaitai Is	1956	Toe pad	2
Tomtit 1	<i>P.m.chathamensis</i>		Mangere Is?	1996?	Feather	2
PMC 17	<i>P.m.chathamensis</i>		Mangere Is	1995-1996	Blood	2
PMC 23	<i>P.m.chathamensis</i>		Mangere Is	1995-1996	Blood	2
CIT1	<i>P.m.chathamensis</i>		South East Is	April 2002	Feather	2
CIT2	<i>P.m.chathamensis</i>		South East Is	April 2002	Feather	2
CIT3	<i>P.m.chathamensis</i>		South East Is	April 2002	Feather	2
CIT4	<i>P.m.chathamensis</i>		South East Is	April 2002	Feather	2
CIT5	<i>P.m.chathamensis</i>		South East Is	April 2002	Feather	2
DM1833	<i>P.m.chathamensis</i>	(Museum sample)	South East Is	1926	Toe pad	2
NM21786	<i>P.m.toitoti</i>	(Museum sample)	Wellington	1973	Toe pad	2
NM26790	<i>P.m.toitoti</i>	(Museum sample)	Whirinaki	2000	Toe pad	2
NM24841	<i>P.m.toitoti</i>	(Museum sample)	Gisbourne	1974	Toe pad	2
NM21785	<i>P.m.toitoti</i>	(Museum sample)	Wellington	1966	Toe pad	2
NM9695	<i>P.m.toitoti</i>	(Museum sample)	Wanganui	1933	Toe pad	2
DM5207	<i>P.m.toitoti</i>	(Museum sample)	Wgtn region	1906	Toe pad	2
NM22720	<i>P.m.macrocephala</i>	(Museum sample)	Franz Josef	1981	Toe pad	2

Sample ID	Species	Individual ID	Location	Date collected	Sample type	Chapter ¹
NM22870	<i>P.m.macrocephala</i>	(Museum sample)	Otago	1983	Toe pad	2
NM24297	<i>P.m.macrocephala</i>	(Museum sample)	Takaka	1989	Toe pad	2
NM22047	<i>P.m.macrocephala</i>	(Museum sample)	Fiordland	1958	Toe pad	2
NM22691	<i>P.m.macrocephala</i>	(Museum sample)	Westland	1981	Toe pad	2
DM8703	<i>P.m.macrocephala</i>	(Museum sample)	Stewart Is	1956	Toe pad	2
NM23668	<i>P.m.dannefaerdi</i>	(Museum sample)	Snares Is	1987	Toe pad	2
NM26238	<i>P.m.dannefaerdi</i>	(Museum sample)	Snares Is	1999	Toe pad	2
DM18614	<i>P.m.dannefaerdi</i>	(Museum sample)	Snares Is	1975	Toe pad	2
DM5242	<i>P.m.dannefaerdi</i>	(Museum sample)	Snares Is	1947	Toe pad	2
DM13226	<i>P.m.marrineri</i>	(Museum sample)	Auckland Is	1943	Toe pad	2
DM7943	<i>P.m.marrineri</i>	(Museum sample)	Auckland Is	1952	Toe pad	2
NM19900	<i>P.multicolor</i>	(Museum sample)	Norfolk Is	1901	Toe pad	2
DM20177	<i>P.multicolor</i>	(Museum sample)	Viti Levu, Fiji	1973	Toe pad	2

Appendix D: Manuscripts

- I. ***An evaluation of methods of blood preservation for RT-PCR from endangered species. Conservation Genetics, 4: 651-654, 2003***

- II. ***Minisatellite DNA profiling detects lineages and parentage in the endangered kakapo (*Strigops habroptilus*) despite low microsatellite DNA variation. Conservation Genetics 4: 265-274, 2003.***