The genetics and evolution of the critically endangered Trinidad Piping Guan *Pipile pipile*, synonym *Aburria pipile*.

Thesis submitted in accordance with the requirements of the University of Chester for the degree of Doctor of Philosophy

by Amelia Anne Grass

February, 2018

Declaration

"The material being presented for examination is my own work and has not been submitted for an award of this or another HEI except in minor particulars which are explicitly noted in the body of the thesis. Where research pertaining to the thesis was undertaken collaboratively, the nature and extent of my individual contribution has been made explicit."

Signed : 

Date : 1st February 2018

Acknowledgements

I would like to acknowledge the Pawi Study Group and Professor John Cooper for supplying the Trinidad Piping Guan samples used throughout this study, without which this research would not have been possible.

I would also like to acknowledge the contribution of Piping Guan samples from the British Natural History Museum ornithological collection and senior curator of the Bird Group, Mark Adams, for his assistance in the provision of reference specimens for this study. We would also like to acknowledge the contribution of Cracid samples from the Museu de Zoologica da Universidade de Säo Paolo in Brazil, and the Chester Zoological Gardens, the Cotswald Wildlife Park and the Lotherton Bird Gardens in the UK.

I wish to acknowledge the Department of Biological Sciences at the University of Chester for funding this study.

I would also like to extend my thanks and gratitude towards my supervisors, Ian McDowall and Lottie Hosie for their support and advice throughout this study, and their continued faith in me.

Finally, I would like to thank my family and friends for their enduring support, without whom I would never have gotten this far.

Abstract

The genetics and evolution of the critically endangered Trinidad Piping Guan *Pipile pipile*, synonym *Aburria pipile.*

Amelia A. Grass

The Trinidad Piping Guan, *Pipile pipile* synonym *Aburria pipile* (Jaquin, 1784) is the only endemic Cracid on the island of Trinidad. The species is currently listed as Critically Endangered and is considered to be in 'on-going decline' by the IUCN, BirdLife International and Cracid Specialist Group. This study aims to examine aspects of genetic variation and the evolution of the mitochondrial genome in the Trinidad Piping Guan utilising, for the first time, samples collected from individuals in the wild and reference specimens of the genus *Pipile* sourced from museum collections.

In this study the complete mitochondrial genome of the Trinidad Piping Guan was sequenced for the first time. Analysis of intra-specific variation of wild Trinidad Piping Guan individuals using single nucleotide polymorphisms demonstrates extremely limited variation within the genes of the mitochondrial genome and nuclear gene intron sequences. Limited variation within this population is consistent with both historical and contemporary contractions of populations within a restricted island system, which may have serious implications for the future of this species in terms of both genetic diversity and conservation management.

Phylogenetic analysis of the complete mitochondrial genome of the Trinidad Piping Guan enabled placement of the genus *Pipile* within the Galliforme evolutionary tree for the first time, and subsequently places the genus within the broader context of the Aves class. Mito-genomic analysis confirms that the Cracids are one of the basal Galliforme clades, and sister taxa to the *Megapodidae*. Phylogenetic placement of the *Pipile* genus is basal to that of the *Crax* species within the Cracidae family, indicative of an earlier evolutionary origin of the Piping Guans. The inclusion of the Trinidad Piping Guan, in the avian evolutionary tree using the whole mitochondrial genomes expands the current genetic phylogeny of the Cracid family, yielding a better understanding of evolutionary relationships among the Galliforme order and the diversification of modern avian lineages.

This study has established novel molecular techniques for the analysis of mitochondrial DNA in historical specimens of the genus *Pipile* from museum reference collections. The analysis of inter-specific relationships within the genus *Pipile* has clarified the evolutionary and biogeographic relationships between the Piping Guan species. Additionally, the Trinidad Piping Guan is genetically defined for the first time as an evolutionarily significant unit, which represents a unique evolutionary pathway within this important genus in a closed island system on the island of Trinidad.

Table of Contents

Declaration **2**

Acknowledgements **3**

Abstract **4**

Table of Contents **5**

List of Figures **9**

List of Tables **13**

List of Appendices **15**

List of Supplementary Documents **17**

Abbreviations **18**

Chapter 1. General Introduction **21**

1.1 Piping Guans of the genus *Pipile* **21**

1.2 The Trinidad Piping Guan **29**

1.3 Evolution, Population Genetics & the Trinidad Piping Guan **34**

1.4 Genetic Analysis of the Trinidad Piping Guan **38**

Chapter 2. General Experimental Methods **52**

2.1 Introduction **52**

2.2 The Sample Collections **52**

2.2.1 Biological Sample Types 52

2.2.2 Study samples obtained from the wild 53

2.2.3 Samples obtained from Captive Piping Guans 56

2.2.4 Samples from Museum Collections 56

2.2.5 Further taxa 59

2.2.6 CITES permits & DEFRA import licences 60

2.2.7 Ethical Approval 60

2.3 DNA Extraction Procedures **61**

2.3.1 DNA Isolation from Feathers 61

2.3.2 DNA Isolation from Egg /Embryo Tissue 63

2.3.3 DNA Isolation from Museum Samples 64

2.4 DNA Quantification **66**

2.4.1 Quantification of dsDNA using Qubit 3.0 Fluorometry 66
2.4.2 Quantification of dsDNA using gel visualisation 66
2.5 Polymerase Chain Reaction 67
2.5.1 Primers 67
2.5.2 PCR Reaction Components 71
2.5.3 PCR Cycle design 71
2.5.4 Advanced PCR Techniques 74
2.6 Gel Electrophoresis 76
2.6.1 Agarose Gel Electrophoresis 76
2.6.2 Visual Analysis 77
2.7 Quality Control 77
2.7.1 Authentication of DNA product 77
2.7.2 NUMTs 78
2.7.3 Controls 79
2.7.4 Ancient DNA 80
2.8 Purification of Target DNA 80
2.9 Sequencing 82
2.10 Health & Safety Considerations 82
2.11 Statistical Analysis 83
2.11.1 Methods for constructing Phylogenetic trees 83
2.11.2 Species & Gene Identification 85
2.11.3 Statistical Analysis Software 85
2.12 Standard Solutions 87
Chapter 3. Complete Mitochondrial Genome of the Trinidad Piping Guan 88
3.1 Introduction 88
3.2 Methodology 96
3.2.1 Trinidad Piping Guan Samples 96
3.2.2 PCR & Primers 97
3.2.2 PCR Amplification 97
3.2.3 Sequence Analysis 98
3.3 Results 102
3.3.1 Compositional Analysis of the Whole Mitochondrial Genome 102
3.3.2 Gene Identification & Features 103

3.3.3 Protein-coding Genes 104
3.3.4 tRNA genes 106
3.3.5 Non-coding Regions 108
3.3.6 Genomic Variation 109
3.4 Discussion 110
3.5 Conclusions 111
Chapter 4 - Intra-specific Genetic Variation in the Trinidad Piping Guan 112
4.1 Introduction 112
4.2 Experimental Methodology 137
4.2.1 Genetic Samples 137
4.2.2 Mitochondrial and Nuclear DNA Sequencing PCR 137
4.2.3 CHD Gender determination 140
4.2.4 Sequence Analysis 142
4.2.5 Sequence Alignment & Construction 144
4.2.6 Phylogenetic Analysis 144
4.2.7 Genetic Distance or Pairwise Distance Analysis 146
4.3 Results 149
4.3.1 Mitochondrial DNA Sequence Analysis 149
4.3.2 Nuclear DNA Sequence Analysis 156
4.3.3 Summary of Genetic Diversity 158
4.3.5 Phylogenetic Relationships 162
4.3.6 CHD Sexing Analysis 166
4.4 Discussion 170
4.5 Conclusions 182
Chapter 5. Evolution of the Trinidad Piping Guan 183
5.1 Introduction 183
5.2 Experimental Methodology 203
5.2.1 Whole Mitochondrial Genome Sequence Data 203
5.2.2 Phylogenetic Analysis 203
5.3 Results 213
5.3.1 Compositional Analysis of the Whole Mitochondrial Genome 213
5.3.2 Individual Genes 213
5.3.3 Phylogenetic Analysis 213

5.3.3.1 Galliforme Data Analysis 214
5.3.3.2 Avian Data Analysis 219
5.3.3.3 Avian with Outgroup Data 222
5.3.4 Metadata Table Analysis 227
5.3.5 Molecular Clock Analysis 229
5.4 Discussion 237
5.5 Conclusions 247
Chapter 6. Taxanomic Identification and Inter-Specific Genetic Variation of the Genus *Pipile* 248
6.1 Introduction 248
6.2 Experimental Methodology 270
6.2.1 The Samples 270
6.2.2 Ancient DNA Extractions 271
6.2.3 PCR & Primers 272
6.2.4 Sequence Preparation 275
6.2.5 Sequence Analysis 275
6.2.6 Statistical Analysis 277
6.2.7 Phylogenetic Analysis 278
6.3 Results 280
6.3.1. Data sets 280
6.3.2. Cytochrome *b* Gene Analysis 282
6.3.3. Cytochrome c oxidase subunit 1 Analysis 286
6.3.4. Control Region Analysis 289
6.3.5 Genetic Distances 292
6.4 Discussion 297
6.5 Conclusions 307
Chapter 7. General Discussion 308
Chapter 8. Further Research 325
Chapter 9. Conclusions 332
References 336
Appendix 361

List of Figures

Figure 1.1: MP tree from combined nuclear and mitochondrial DNA sequence data with bootstrapping, resolving relationships among the *Cracidae* (Pereira et al., 2002). **26**

Figure 1.2: Neighbour-Joining tree using Tamura-Nei's model of sequence evolution, for phylogenetic analysis of the genus *Pipile.* (Robinson, 2011). **28**

Figure 2.1: Samples from wild Trinidad Piping Guans; Moulted feathers, Egg & embryo and Quill sections. (All photographs provided by A. Grass except the Egg & embryo images provided and printed with permission of by Prof. John Cooper). **54**

Figure 2.2: Map of the island of Trinidad (A), highlighting the Northern Range region (B) and indicating the origins of the Trinidad Piping Guan samples. Samples Reference 1, Reference 2 and the Egg/embryo 1 (R1, R2 & E1) are of sourced from the wild in an unknown location of the 'Northern Range'. 55

Figure 2.3: Samples from Piping Guans held in the British Natural History Museum ornithological collection Tring, UK (BNHM). All photographs provided by A. Grass and printed with permission of the British Natural History Museum. **57**

Figure 2.4: Samples from Piping Guans held in the Museu de Zoologia da Universidade de São Paulo, Brazil (MZUSP). All photographs provided by A. Grass. **58**

Figure 2.5: Map of sample origins for the wild Trinidad Piping Guan samples (*Pipile Pipile**) and samples from the British Natural History Museum collection (*P. cumanensis, P. grayi & P. cujubi**) and specimens from the Museu de Zoologia da Universidade de São Paulo collection (*A. cumanensis, A. grayi* & *A. cujubi**). **59**

Figure 2.6: Trinidad Piping Guan Feathers; A: complete contour feather, B: close up of quill & C: illustration of the major features of the feather, diagram of a primary flight feather (photographs by A. Grass). **62**

Figure 2.7: DNA quantification by gel electrophoresis, using the Image Lab software (Biorad), from known DNA standards (Lane 1) in Norgen PCR Sizer 100bp DNA Ladder (Geneflow). **67**

Figure 2.8: Standard Polymerase chain reaction (PCR) cycling parameters. **72**

Figure 2.9: Touchdown Polymerase Chain Reaction cycle parameters, where T_m = 55 ºC. **73**

Figure 2.10: Primerless Polymerase Chain Reaction cycle parameters. **75**

Figure 3.1: Mitochondrial gene order from ND5 to 12S RNA as described for the typical vertebrate (non-avian) animal mitochondrial genome (a), the typical avian mitochondrial genome (b), and the alternate avian mitochondrial genome (c). Adapted from Mindell et al., 1998. **91**

Figure 3.2: Mitochondrial gene organisation in the Trinidad Piping Guan, *Pipile pipile*. **102**

Figure 3.3: The genomic organisation of the Trinidad Piping Guan mitochondrial genome, Nucleotide position 1 is located at the 5' end of the Control region. **103**

Figure 3.4: Secondary structures of tRNA genes of Pipile pipile, as constructed by tRNAscan-SE 1.21 software. **107**

Figure 3.5: Conserved features of the Control Region of the Trinidad Piping Guan. **108**

Figure 4.1: Karyotype diagram of loci used in this study, based on the Chicken genome (Gallus gallus), adapted from Kimball et al. (2009). **133**

Figure 4.2: CHD Sex determination using P2/P8 primers (Griffiths, Double, & Dawson, 1998). **142**

Figure 4.3: Hasegawa-Kishino-Yano substitution model matrix (Felsenstein, 2004; Nei & Kumar, 2000; Page & Holmes, 1998; Yang, 2014). **145**

Figure 4.4: Tamura-Nei substitution model matrix (Felsenstein, 2004; Nei & Kumar, 2000; Page & Holmes, 1998; Yang, 2014). **146**

Figure 4.5: Molecular Phylogenetic analysis of the mitochondrial DNA Control Region (A), the Cytochrome b gene (B), Cytochrome oxidase 1 (C) and NADH dehudrogenase 2 (D). Using the Maximum Likelihood method. **163**

Figure 4.6: Molecular Phylogenetic analysis of the combined mitochondrial sequence analysis of the CytB, COI, ND2 and Control region (4879bp) by Maximum Likelihood method. **164**

Figure 4.7: Molecular Phylogenetic analysis of the Beta-Fibrinogen intron 7 (A) and Serpin B14 intron c (B) by Maximum Likelihood method. **165**

Figure 4.8: Molecular Phylogenetic analysis of the combined nuclear DNA sequence analysis of the Beta-Fibrinogen intron 7 and Serpin B14 intron c (1482bp) by Maximum Likelihood method. **166**

Figure 4.9: Gel electrophoresis analysis of the CHD sex marker in *Pipile* specimens of known sex using the CHD1F/CHD1R primer set (Lee et al., 2010). **169**

Figure 5.1: Current Consensus on Galliforme Relationships based on DNA hybridisation, Cytochrome B and ND2 partitions adapted from Dimcheff et al., (2002) and Van Tuinen & Dyke (2004). **186**

Figure 5.2: Cladograms from various molecular phylogenetic analyses of the Galliforme genera. **189**

Figure 5.3: Cladograms from various molecular phylogenetic analyses of the Avian class. **195**

Figure 5.4: Representation of the scale of geological time, Age is represented in Millions of Years (Mya). **196**

Figure 5.5: General Time Reversible substitution model matrix (Felsenstein, 2004; Nei & Kumar, 2000; Page & Holmes, 1998; Yang, 2014) **208**

Figure 5.6: Phylogenetic relationships among Galliformes based on analysis of 10,858bp of Whole Mitochondrial Genome data (13 protein-coding genes) from 59 genera of the Galliforme order including the novel *Pipile pipile,* and 4 Anseriforme genera as a putative outgroup, Maximum Likelihood method. **217**

Figure 5.7: Phylogenetic relationships among Galliformes based on analysis of 10,858bp of Whole Mitochondrial Genome data (13 protein-coding genes) from 59 genera of the Galliforme order including the novel *Pipile pipile,* and 4 Anseriforme genera as a putative outgroup, Bayesian Inference method. **218**

Figure 5.8: Phylogenetic relationships among Avians based on analysis of 11,433bp of Whole Mitochondrial Genome data (13 protein-coding genes) from 68 genera across the Aves class; 23 of the Galliforme order including the novel *Pipile pipile,* Maximum Likelihood method. **220**

Figure 5.9: Phylogenetic relationships among Galliformes based on analysis of 11,433bp of Whole Mitochondrial Genome data (13 protein-coding genes) from 68 genera across the Aves class; 23 of the Galliforme order including the novel *Pipile pipile,* Bayesian Inference method. **221**

Figure 5.10: Phylogenetic relationships among Avians based on analysis of 11,010bp of Whole Mitochondrial Genome data (12 protein-coding genes) from 68 genera across the Aves class including the novel *Pipile pipile,* and 3 reptile outgroups, Maximum Likelihood method. **225**

Figure 5.11: Phylogenetic relationships among Galliformes based on analysis of 11,010bp of Whole Mitochondrial Genome data (12 protein-coding genes) from 68 genera across the Aves class including the novel *Pipile pipile,* and 3 reptile outgroups, Bayesian Inference method. **226**

Figure 5.12: Evolutionary relationships among Galliformes based on analysis of 10,858bp of Whole Mitochondrial Genome data (12 protein-coding genes) from 59 genera of the Galliforme order including the novel *Pipile pipile,* and 4 Anseriforme genera as a putative outgroup, Maximum Likelihood method. **232**

Figure 5.13: Evolutionary relationships among Galliformes based on analysis of 10,858bp of Whole Mitochondrial Genome data (12 protein-coding genes) from 59 Galliformes including the novel *Pipile pipile,* and 4 Anseriformes as an outgroup, Bayesian Inference method. **233**

Figure 5.14: Evolutionary relationships among Galliformes based on analysis of 11,433bp of Whole Mitochondrial Genome data (12 protein-coding genes) from 68 genera of the Avian class including the novel *Pipile pipile,* Maximum Likelihood method. **235**

Figure 5.15: Evolutionary relationships among Galliformes based on analysis of 11,433bp of Whole Mitochondrial Genome data (12 protein-coding genes) from 68 Avian species including the novel *Pipile pipile,* Bayesian Inference method. **236**

Figure 6.1: Photographs of facial features of the Piping Guan genus Aburria (photographs reproduced with permission of *Trini Eco Warriors and the British Natural History Museum). **241**

Figure 6.2: Estimated distribution of the Piping Guan species of the genus Pipile (synonym Aburria), adapted from Grau et al. (2005). **252**

Figure 6.3: Inferred phylogenetic relationships of the Cracidae family. **257**

Figure 6.4: Alignment of *Pipile* species (*Aburria*) sequences of the Cytochrome b gene region, homologous nucleotides indicated in grey and single nucleotide polymorphisms indicated by colours, in Geneious R8. **280**

Figure 6.5: Alignment of *Pipile* species (*Aburria*) sequences of the Cytochrome oxidase subunit 1 gene region, homologous nucleotides indicated in grey and single nucleotide polymorphisms indicated by colours, in Geneious R8. **281**

Figure 6.6: Alignment of Pipile species (Aburria) sequences of the Control Region, homologous nucleotides indicated in grey and single nucleotide polymorphisms indicated by colours, in Geneious R8. **282**

Figure 6.7: Phylogenetic analysis of *Pipile* species (*Aburria*) sequences of the Cytochrome b gene region, including specimens from the BNHM and MZUSP collections and wild Trinidad Piping Guans. **283**

Figure 6.8: Phylogenetic analysis of *Cracidae* order, using 448 base sequences of the Cytochrome b gene, including specimens from the BNHM and MZUSP collections and wild Trinidad Piping Guans. Bayesian inference method. **285**

Figure 6.9: Phylogenetic analysis of *Pipile* species (*Aburria*) sequences of the Cytochrome c oxidase subunit 1 gene region, including specimens from the BNHM and MZUSP collections and wild Trinidad Piping Guans. Bayesian inference method. **287**

Figure 6.10: Phylogenetic analysis of *Cracidae* order using 338 base sequences of the Cytochrome c oxidase subunit 1 gene, including specimens from the BNHM and MZUSP collections and wild Trinidad Piping Guans. Bayesian inference method. **288**

Figure 6.11: Phylogenetic analysis of *Pipile* species (*Aburria*) sequences of the Control Region, including specimens from the BNHM and MZUSP collections and wild Trinidad Piping Guans. Bayesian inference method. **290**

Figure 6.12: Phylogenetic analysis of *Cracidae* order using 633 base sequences of the Control Region, including specimens from the BNHM and MZUSP collections and wild Trinidad Piping Guans. Bayesian inference method. **291**

Figure 8.1: Gel electrophoresis image of the amplification of the Hughes & Larson (2000) Wattled Curassow microsatellite primer sets, using the Trinidad Piping Guan Pawi 5 sample. **328**

Figure 8.2: Polyacrylamide gel electrophoresis of ISSR $GACA_4$ amplifications in the Trinidad Piping Guan samples. **330**

List of Tables

Table 2.1: Standard Nucleotide Coding System. **70**

Table 3.1: Samples used in whole mitochondrial genome analysis. **96**

Table 3.2: Gene information for the complete mitochondrial genome of the Trinidad piping Guan (accession number KU221051). **105**

Table 3.3: Single Nucleotide Polymorphisms in the complete mitochondrial genomes of three individuals of the species *Pipile pipile*. **109**

Table 4.1: Primers used in amplification of the whole Control Region and whole Cytochrome B, Cytochrome Oxidase subunit I and NADH dehydrogenase 2 genes of the mitochondrial genome. **138-9**

Table 4.2: Primers used in amplification of the Beta-fibrinogen intron 7 and Serpin B14 intron c, regions of the nuclear genome. **140**

Table 4.3: CHD Primer sets and associated PCR Thermal Cycles. **141**

Table 4.4: Analysis of single nucleotide variation in the Control Region in the wild Trinidad Piping Guan and UK captive Piping Guan populations. **150**

Table 4.5: Analysis of single nucleotide variation in the Cytochrome *b* gene in the wild Trinidad Piping Guan and UK Captive Piping Guan populations. **152**

Table 4.6: Analysis of single nucleotide variation in the Cytochrome c oxidase subunit 1 gene in the UK captive Piping Guan population. **153**

Table 4.7: Analysis of single nucleotide variation in the NADH dehydrogenase subunit 2 gene in the wild Trinidad Piping Guan population. **154**

Table 4.8: Analysis of single nucleotide variation in the nuclear DNA Beta-fibrinogen intron 7 in the wild Trinidad Piping Guan population and captive Piping Guan population. **156**

Table 4.9: Analysis of single nucleotide variation in the nuclear DNA SERPIN B14 intron c in the wild Trinidad Piping Guan population **157**

Table 4.10: Measures of genetic diversity within the Wild Trinidad Piping Guan population. **159**

Table 4.11: Measures of genetic diversity within the captive Piping Guan population. **160**

Table 4.12: Gel electrophoresis analysis results for the sex determination of wild Trinidad Piping Guans and captive Piping Guans using the P2/P8 and CHD1F/CHD1R CHD tests (Griffiths et al., 1998; Lee et al., 2010). **167**

Table 5.1: Fossils & Geological events used in the calibration of avian Molecular Clocks. **201**

Table 5.2: Data sets used for phylogenetic and evolutionary analysis of *Pipile pipile.* **204**

Table 5.3: Priors included in analysis in BEAST 2.2 (BEAUti) Fossil and Molecular Calibration points. **211**

Table 5.4: Meta-analysis of bootstrap support for well accepted clades within avian phylogenetic trees, from this study and other published literature. **228**

Table 6.1: Holotype specimens of the genus Aburria (Pipile), as reported by Lopez (2009). **264**

Table 6.2: Samples used in Taxonomic analysis. **270**

Table 6.3: Primers used in amplification of the domain I of the Control Region, Cytochrome B, and Cytochrome Oxidase subunit I genes of the mitochondrial genome. **273-4**

Table 6.4: Specimen information for the final data sets used in taxonomic analysis of the Pipile genus and Cracidae **277**

Table 6.5: Average genetic distances (p-distance) between species of the *Pipile* genus, for the 338 base region of the Cytochrome oxidase 1 gene, 448 base region of the Cytochrome *b* gene and 633 base section of the Control Region. **292**

Table 6.6: Average genetic distances (p-distance) within species of the Cracidae order, for the 338 base region of the Cytochrome oxidase 1 gene, 448 base region of the Cytochrome *b* gene and 633 base section of the Control Region. **295**

Table 6.7: Average genetic distances (p-distance) within genera of the Cracidae order, for the 338 base region of the mitochondrial Cytochrome oxidase 1 gene, 448 base region of the Cytochrome b gene and 633 base section of the Control Region. **296**

List of Appendices

Appendix 1: Sample Information

Table 1: Complete list of all biological samples used in this study.

Appendix 2: Primer Sequence Information

Table 1: Complete list of oligonucleotide primers used in the sequencing of the complete mitochondrial genome and analysis of genetic variation.

Appendix 3: Publication of the Whole Mitochondrial Genome

Grass, A., Hosie, C., & McDowall, I. (2016). The complete mitochondrial genome and phylogenetic position of the critically endangered Trinidad Piping Guan, Pipile pipile synonym Aburria pipile (Aves: Galliformes). *Mitochondrial DNA*.

Appendix 4: Genetic variation in the Trinidad Piping Guans

Figure 1: Alignment of the mitochondrial Control Region sequences of eight Trinidad Piping Guan individuals, using Clustal W in Geneious R8.0.

Figure 2: Alignment of the mitochondrial Cytochrome b gene sequences of eight Trinidad Piping Guan individuals, using Clustal W in Geneious R8.0.

Figure 3: Alignment of the mitochondrial Cytochrome c oxidase subunit 1 gene sequences of eight Trinidad Piping Guan individuals, using Clustal W in Geneious R8.0.

Figure 4: Alignment of the mitochondrial NADH dehydrogenase subunit 2 gene sequences of eight Trinidad Piping Guan individuals, using Clustal W in Geneious R8.0.

Figure 5: Alignment of the nuclear DNA Beta-fibrinogen intron 7 sequences of four Trinidad Piping Guan individuals and nine UK captive Piping Guans, using Clustal W in Geneious R8.0.

Figure 6: Alignment of the nuclear DNA SERPIN B14 intron c sequences of three Trinidad Piping Guan individuals and nine UK captive Piping Guans, using Clustal W in Geneious R8.0.

Appendix 5: Reference Sequence Information

Table 1: Whole mitochondrial genome sequences used in the phylogenetic analysis of the genetic relationships between Galliformes.

Table 2: Whole mitochondrial genome sequences used in the phylogenetic analysis of the genetic relationships between Avian orders, in addition to those of the Galliforme order and including reptile outgroup sequences.

Appendix 6: Phylogenetic Reconstructions

Figure 5.16: Phylogenetic relationships among Galliformes based on analysis of Cytochrome b gene data (1143bp) from 59 genera of the Galliforme order including the novel *Pipile pipile,* and 4 Anseriforme genera as a putative outgroup.

Figure 5.17: Phylogenetic relationships among Galliformes based on analysis of Cytochrome c oxidase subunit 1 (COI) gene data (1551bp) from 59 genera of the Galliforme order including the novel *Pipile pipile,* and 4 Anseriforme genera as a putative outgroup.

Figure 5.18: Phylogenetic relationships among Galliformes based on analysis of NADH dehydrogenase subunit 2 (ND2) gene data (1038bp) from 59 genera of the Galliforme order including the novel *Pipile pipile,* and 4 Anseriforme genera as a putative outgroup.

Figure 5.19: Phylogenetic relationships among Galliformes based on analysis of Control Region gene data (1429bp) from 59 genera of the Galliforme order including the novel *Pipile pipile,* and 4 Anseriforme genera as a putative outgroup.

Figure 5.20: Phylogenetic relationships among Galliformes based on analysis of 13,365bp of Whole Mitochondrial Genome data (13 protein coding genes and 2 rRNA genes) from 59 genera of the Galliforme order including the novel *Pipile pipile,* and 4 Anseriforme genera as a putative outgroup.

Figure 5.21: Phylogenetic relationships among Galliformes based on analysis of 13,365bp of Whole Mitochondrial Genome data (13 protein-coding genes and 2 rRNA genes) from 59 genera of the Galliforme order including the novel *Pipile pipile,* and 4 Anseriforme genera as a putative outgroup.

Figure 5.22: Phylogenetic relationships among Avian class based on analysis of Cytochrome b gene data (1143bp) from 71 genera; 23 of the Galliforme order including the novel *Pipile pipile,* and 3 reptile outgroups.

Figure 5.23: Phylogenetic relationships among Avian class based on analysis of Cytochrome c oxidase subunit 1 gene data (1548bp) from 71 genera; 23 of the Galliforme order including the novel *Pipile pipile,* and 3 reptile outgroups.

Figure 5.24: Phylogenetic relationships among Avian class based on analysis of NADH dehydrogenase subunit 2 gene data (1038bp) from 71 genera; 23 of the Galliforme order including the novel *Pipile pipile,* and 3 reptile outgroups.

Figure 5.25: Phylogenetic relationships among Avians based on analysis of 11,433bp of Whole Mitochondrial Genome data (2 rRNA genes and 13 protein-coding genes) from 68 genera across the Aves class; 23 of the Galliforme order including the novel *Pipile pipile*.

Figure 5.26: Phylogenetic relationships among Galliformes based on analysis of 11,433bp of Whole Mitochondrial Genome data (2 rRNA genes and 13 protein-coding genes) from 68 genera across the Aves class; 23 of the Galliforme order including the novel *Pipile pipile*.

Appendix 7: Reference Sequence Information

Table 1: Species Identification and NCBI DNA database accession numbers for sequences used in phylogenetic analysis.

List of Supplementary Documents

Table S1: Genetic Distance Analysis Matrix (p-distance) of the Control Region in the wild and captive Piping Guan populations.

Table S2: Genetic Distance Analysis Matrix (p-distance) of the Cytochrome *b* gene in the wild and captive Piping Guan populations.

Table S3: Genetic Distance Analysis Matrix (p-distance) of the Cytochrome c oxidase subunit 1 gene in the wild and captive Piping Guan populations.

Table S4: Genetic Distance Analysis Matrix (p-distance) of the NADH dehydrogenase 2 gene in the wild and captive Piping Guan populations.

Table S5: Genetic Distance Analysis Matrix (p-distance) of the combined sequence data from the Control Region, COI, CytB, and ND2 mitochondrial genes in the wild Trinidad Piping Guans and captive Piping Guan populations.

Table S6: Genetic Distance Analysis Matrix (p-distance) of the nuclear DNA Beta-fibrinogen intron 7 in the wild and captive Piping Guan populations.

Table S7: Genetic Distance Analysis Matrix (p-distance) of the nuclear DNA SERPIN B14 intron in the wild and captive Piping Guan populations.

Table S8: Genetic Distance Analysis Matrix (p-distance) of the combined sequence data from the Beta-fibrinogen and SerpinB14 nuclear DNA gene introns in the wild Trinidad Piping Guans and captive Piping Guan populations.

Figure S1: Alignment of *Pipile* species (*Aburria*) sequences and Cracidae reference sequences of the Cytochrome *b* gene region.

Figure S2: Alignment of *Pipile* species (*Aburria*) sequences and Cracidae reference sequences of the Cytochrome oxidase subunit 1 gene region.

Figure S3: Alignment of *Pipile* species (*Aburria*) sequences and Cracidae reference sequences of the Control Region.

Abbreviations

Abbreviations which have been used throughout the text with regard to units, nomenclature and chemical abbreviations.

AIC	Akaike Information Criterion
AMNH	American Museum of Natural History
AU	Absorbance Units (eg. mAU/ml)
BIC	Bayesian Information Criterion
bp	Base pairs
BMNH	British Museum of Natural History
CBOL	Consortium for the Barcode ofLife
CITES	Convention on the International Trade in Endangered Species
COI	Cytochrome c oxidase subunit 1 (also subunits COII and COIII)
CR	Control Region
Cyt B	Cytochrome *b*
CSB	Conserved Sequence Block
DEFRA	Department for Environment, Food and Rural Affairs
Dloop	Displacement loop
DMSO	dimethyl sulphoxide
DTT	1,4-dithiothreitol
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide tri-phosphate
dsDNA	Double stranded DNA
EDTA	Ethylenediaminetetraacetic acid
ESU	Evolutionary Significant Unit
Fig.	Figure
g	Gram

H_2O	Water
IUCN	International Union for Conservation of Nature and Natural Resources
Kb	Kilo-base pairs
L	Litre
M	Molar concentration
MCMC	Markov Chain Monte Carlo [statistical model]
mg	Milligram (milli 10^{-3}x)
MgCl	Magnesium Chloride
mins	Minutes
ml	Millilitre (milli 10^{-3}x)
ML	Maximum Likelihood
mtDNA	Mitochondrial DNA
Mya	Million Years Ago
Myr	Million Years
MZUSP	Museu de Zoologica da Universidade de Sào Paolo
NADH	Nicotinamide adenienine dinucleotide
NCBI	National Centre for Biotechnology Information
ND2	NADH dehydrogenase subunit 2 (also subunits ND1, ND3, ND4L, ND4, ND5 and ND6)
nDNA	Nuclear DNA
ng	Nanogram (nano 10^{-9}x)
nm	Nanometer
NUMT	Nuclear Mitochondrial Pseudogene, or Nuclear Mitochondrial DNA segment
Pawi	Trinidad Piping Guan (local island name)
PCR	Polymerase Chain Reaction
pPCR	Primerless PCR

pH	$-\log_{10}$ Hydrogen ion concentration (H+) [numeric scale of acidity/basicity]
pmol	Picomoles (pico 10^{-12}x)
rpm	Revolutions per minute
RNA	Ribonucleic acid
rRNA	ribosomal RNA
RFLP	Restriction Fragment Length Polymorphism
sec	Seconds
SNP	Single Nucleotide Polymorphism
Tris	2-amino-2-(hydroxymethyl)-1,3-propanediol
tRNA	transfer RNA
TAE	Tris-acetate-EDTA
TPG	Trinidad Piping Guan
UV	Ultra Violet (light)
V	Volts
°C	Degrees Celsius
∞	Infinity
µl	Microliters (micro 10^{-6}x)

Chapter 1. General Introduction

The Trinidad Piping Guan, *Pipile pipile* (Jacquin 1784) is the only cracid species endemic to the Island of Trinidad. The Trinidad Piping Guan, known locally as the Pawi, is listed by the IUCN as Critically Endangered, with the most recent population estimates suggesting there are less than 200 individual birds left in the wild (Hayes, Sanasie, & Samad, 2009). The Trinidad Piping Guan population has been in drastic decline over recent decades due to unsustainable levels of hunting, destruction of its primary habitat and the slow reproductive rate of the species (Hayes et al., 2009). Significant decline in the population size over a prolonged period has undoubtedly led to restriction of the gene pool and reduced genetic variation within the species, which may have serious implications for the future of the species.

This study aims to develop molecular methods of analysis of both intra-specific variation within the species *Pipile pipile* and inter-specific variation within the genus *Pipile*, using both mitochondrial DNA and nuclear DNA analysis. The Trinidad Piping Guan will be genetically defined for the first time using the mitochondrial genome, through examination of the genetic relationships within the *Pipile* genus, between species of the Cracidae family and its broader placement within the Aves class. It is hoped that the development of this reference collection of molecular information and DNA samples from Trinidad Piping Guans will aid further research into the genetics of this important endemic species and assist in the development of conservation management strategies to ensure the survival of the species in the future.

1.1 Piping Guans of the genus *Pipile*

Avian Biodiversity & the Neotropics

Biodiversity levels in the Neotropical regions of the planet are some of the highest recorded in any terrestrial environment, the Neotropics maintains a high degree of species endemism and therefore has been recorded as key areas for conservation (Vegas-Vilarrúbia, Rull, Montoya, & Safont, 2011). The biodiversity in Neotropical

regions is theorised to have resulted through mechanisms including major climatic and geological changes such as river formation, mountain uplifts and glacial cycling, and the subsequent response of species to environmental shifts resulting in the specialisation and diversification of species (Lutz, Weckstein, Patané, Bates, & Aleixo, 2013). Taxonomic composition of the Neotropics altered in response to recent Quaternary climatic and geographic fluctuations and the reactions of species to environmental shifts, enhancing species diversifications in these regions (Vegas-Vilarrúbia et al., 2011).

The Neotropical regions of South America are one of the most biodiverse regions on the planet. It has been estimated that over one eighth of neotropical avifauna are endangered or vulnerable, many of which are large forest frugivores of the *Cracidae, Psittacidae, Tinamidae, Ramphastidae and Psophiidae* families (Cracids, Parrots, Tinamous, Toucans and Trumpeters respectively) (Strahl & Grajal, 1991). A lack of knowledge surrounding the neotropical avifauna has traditionally left them vulnerable due to a lack of management; widespread, unsustainable hunting and habitat destruction have significantly impacted on these species despite their ecological and socio-economic importance.

Recent evidence suggests that globally there has been a significant decline in the levels of biodiversity (Butchart et al., 2010; Duraes, Carrasco, Smith, & Karubian, 2013; Mooney et al., 2009; Moura et al., 2013; Rompre, Robinson, Desrochers, & Angehr, 2009). Anthropomorphic pressures and critically habitat loss are thought to have induced extinction levels comparable to the previous major extinction level events in geological history (Vegas-Vilarrúbia et al., 2011). Climate change has exacerbated the existing anthropomorphic pressures on environmental resources and species, cascading species extinctions in response to climatic shifts and accelerating global level changes in habitat and ecosystem interactions (Bellard, Bertelsmeier, Leadley, Thuiller, & Courchamp, 2012; Duraes et al., 2013; Mooney et al., 2009; Parmesan, 2006; Sgrò, Lowe, & Hoffmann, 2011). It has been predicted that 35% of avian species are susceptible to the influences of climate change, with endemic species being more vulnerable to species loss (Bellard et al., 2012).

Neotropical Aves

Historical distributions and biogeographic patterns of avifauna indicate distinct regions of endemism within South America which originate through physiogeographic barriers, geographic gradients and climatic refugia (Carcraft, 1985; Tilston Smith et al., 2014). Neotropical avian taxa are subject to high degrees of biodiversity (Duraes et al., 2013), largely as a consequence of high levels of regional endemism within clades. Avian endemic regions are convergent with that of other taxonomic classes, including invertebrates and plants (Carcraft, 1985; Haffer, 1985; Tilston Smith et al., 2014), species distribution is attributed to both biotic and abiotic factors (Rompre et al., 2009). Regional endemism is evidenced in numerous species including avians such as the Piping Guans (Grau, Pereira, Silveira, Höfling, & Wajntal, 2005), Toucans (Lutz et al., 2013; Patel, Weckstein, Patané, Bates, & Aleixo, 2011), as well as mammal and invertebrate taxa (Di Fiore et al., 2014; Matos-Maraví, Peña, Willmott, Freitas, & Wahlberg, 2013).

One of the dominant trophic groups of avifauna is the arboreal frugivores and insectivores (Blake & Loiselle, 2000), which includes most of the Cracidae species. The Cracids are predominantly Neotropical today, however fossil evidence suggests a wider distribution including North America (Haffer, 1985; Tordoff & MacDonald, 1957), which is indicative of major biogeographic and climatic shifts, Cracids being tropical specialists (Hosner, Braun, & Kimball, 2016). Fifteen of the Cracidae species are endemic to regions of South America (Carcraft, 1985), as a consequence of species specialisms, climatic and vegetative fragmentations.

The Cracids & the Genus Pipile

The Galliforme order comprises more than 250 species of around 70 genera, found globally, including some of the most common species of avians such as the chicken *Gallus gallus*, now a highly domesticated species found on almost every continent. Other common taxa within the order include the *Phasianidae* and *Numididae* (pheasants, grouse, partridges and guinea-fowl), and less well known birds such as the

Megapodidae (scrubfowl) and *Cracidae* (guans, chachalacas and curassows) (del Hoyo, Ellliot, & Sargatal, 1994; Dyke, Gulas, & Crowe, 2003).

The family *Cracidae* comprises eleven genera divided into 56 currently recognised species of guans, chachalacas, curassows and horned guans, endemic to the Neotropical regions of South America (Grau et al., 2005; Hosner et al., 2016; Pereira et al., 2002). The family is characterised as medium- to large-bodied forest dwelling avians of the Gallinaceous form, most of which are non-migratory, arboreal frugivores (Delacour & Amadon, 1973; del Hoyo et al., 1994; Hughes & Larson, 2000).

The *Cracidae* are one of the most endangered families of avians in the Neotropics, many of these species are now threatened in the wild due to loss of primary habitat and unsustainable hunting (Brooks, 1999; Grau et al., 2005; Hughes & Larson, 2000; Pereira et al., 2002). Of the 56 species of Cracids, 20 are listed as Vulnerable, Endangered or Critically endangered under the IUCN Red List of threatened species (Brooks & Strahl, 2000; IUCN, 2017) and 14 are protected under CITES legislation (CITES, 2017).

Cracids are key indicator species within the Neotropical ecosystems, birds of this family have been found to be highly sensitive to habitat quality and hunting pressures, and are therefore important bio-indicators of the health of their habitats (Galetti, Martuscelli, Olmos, & Aleixo, 1997; Grau, Pereira, Silveira, & Wajntal, 2003; Pereira & Wajntal, 2001; Strahl & Grajal, 1991). Cracids play a critical role in the ecosystem of Neotropical forests; as largely reliant on a fruit based diet, these birds are vital for the regeneration of forests via seed dispersal and control of plant diversity through seed predation (Pereira & Wajntal, 2001). Strahl & Grajal (1991) estimated that large avian frugivores, including Cracids, parrots, toucans and trumpeters, contribute 60-75% of the total biomass of avian seed dispersers in Neotropical forest ecosystems.

Cracids also have considerable socio-economic importance in South America, particularly for rural and Amerindian communities. Large avian species are one of the most important sources of protein for native populations and rural communities (Brooks, 1999; Strahl & Grajal, 1991); guans and curassows represent the highest biomass in subsistence hunting in some Neotropical regions (Brooks, 1999; Strahl & Grajal, 1991). Some species are hunted for local crafts and cultural or ceremonial purposes, however sport hunting of Cracids as game birds also occurs in the Neotropics, and licencing fees and hunting site fees provide an economic incentive to local communities (Strahl & Grajal, 1991). *Pipile* species are the second most hunted genus amongst the Cracids, and seventh most frequently hunted group of all game species in Latin America, however most are hunted unsustainably (Brooks, 1999). Critically endangered Piping Guans (Trinidad Piping Guans and Jacutinga) are considered more susceptible to the effects of hunting due to their fragmented and island habitat effects (Brooks, 1999; Galetti et al., 1997).

Taxonomic & Phylogenetic Classifications

The family Cracidae are generally considered to be one of the most primitive forms in the Galliforme family, with fossil records indicative of the ancient origins of the genus (Delacour & Amadon, 1973; del Hoyo et al., 1994; Hosner et al., 2016). The classification of Galliforme species has been much debated since initial classifications under the Linnean system (1758). The *Numididae*, *Megapodidae*, *Odontophoridae*, *Phasianidae* and *Cracidae* have conventionally been grouped together, however composition of these groups has varied significantly (Dyke et al., 2003).

Dyke et al. (2003) found phylogenetic relationships among the Galliformes inferred by morphological characteristics to be broadly congruent with traditional classifications and molecular evidence from DNA-DNA hybridisation analysis by Sibley & Ahlquist (1990). The monophyletic sister relationship of *Phasianidae* and the *Cracidae* has been confirmed by both molecular and morphological data (Dyke et al., 2003; Pereira et al., 2002).

Phylogenetic relationships at the genus level have been proposed using morphological characteristics and osteology, and molecular techniques (Pereira et al., 2002), however little consensus has been achieved. Dyke et al. (2003) found resolution of the phylogenetic relationships using morphological techniques was insufficient to determine relationships beyond the generic level. Delacour and Amadon (1973) and Del Hoyo et al. (1994) judged that the Cracids were most naturally divided into three groups, the Chachalacas, Guans and Curassows, similarities between the Chachalacas and Guans indicate that the two groups could be attached to a single sub-family *Penelopinae*, with Curassows being sufficiently dissimilar to be separated into the sub-family *Cracinae.* Pereira et al. (2002) utilised combined nuclear and mitochondrial gene sequences to discriminate relationships among the eleven genera of Cracidae, see Figure 1.1.

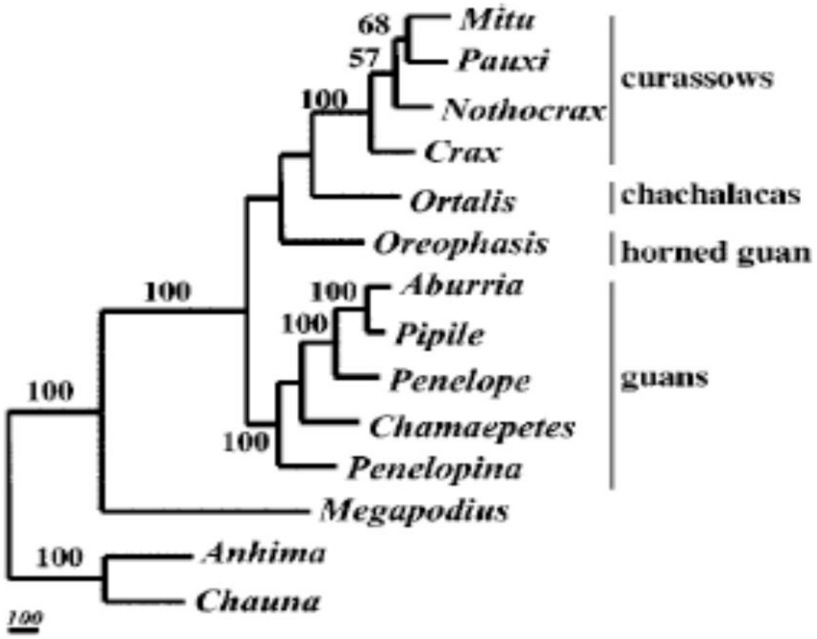


Figure 1.1: MP tree from combined nuclear and mitochondrial DNA sequence data with bootstrapping, resolving relationships among the *Cracidae* (Pereira et al., 2002).

Recent comprehensive analysis of Cracidae species by Hosner et al. (2016) supports the molecular relationships of Pereira et al. (2002), classifying the Cracids in four strongly supported groups; the curassows (*Crax*, *Mitu* and *Pauxi*), the chachalacas (*Ortalis*), the horned guans (*Oreophasis*) and the guans (*Penelope*, *Penelopina*, *Pipile*, *Aburria* and *Chamaepetes*).

The status of the *Pipile* and *Aburria* genera has been greatly disputed, the separation of the two genera has been questioned both in terms of morphology and molecular research. The merge of the two genera has been suggested based on morphological characteristics (Delacour & Amadon, 1973; Grau et al., 2005) and close molecular divergences (Pereira et al., 2002; Robinson, 2011). In the analysis of Hosner et al. (2016) some mitochondrial DNA evidence is supportive of the merge of *Aburria* and *Pipile*, however the addition of nuclear DNA markers to the analysis excludes the *Aburria* from the *Pipile* clade. Currently the merge of the *Pipile* and *Aburria* genera has been rejected by the South American Classification Committee due to the lack of species genetic data, and *Pipile* and *Aburria* remain separate genera (Remsen et al., 2012).

There are currently six accepted forms of the genus *Pipile* (Grau et al., 2005; del Hoyo et al., 1994), taxonomy and the phylogenetics of which are still disputed. Four species of *Pipile* are well recognised in research; *Pipile pipile* (Trinidad Piping-guan), *Pipile cumanensis* (Blue-throated or Common Piping-guan), *Pipile cujubi* (Red-throated Piping-guan) and *Pipile jacutinga* (Black-fronted Piping-guan). Further to this the species *P.cumanensis* and *P. cujubi* are sub-divided into the sub-species, *P. cumanensis cumanensis, P. cumanensis grayi, P. cujubi cujubi* and *P. cujubi nattereri* (Grau et al., 2005).

In 1967 Vaurie published systematic notes of the genus *Pipile,* in which the six geographical forms were recognised; the Trinidad form confined to the island and the five widely distributed mainland South American forms. Similarities in both appearance and size suggest the six forms are closely related; the most important differentiating morphological features determining relationships are the bare blue face and throat common to *P. cumanensis cumanensis, P. cumanensis grayi,* and *P.pipile*, the bare blue face and bare blue and red throat common to *P. cujubi cujubi* and *P. cujubi nattereri*, and the black and feathered face of *Pipile jacutinga.*

The Trinidad form of Piping Guan *Pipile pipile*, is currently accepted as monospecific, it being a single species diverged from mainland Piping guans. Grau et al. (2005) assessed morphological and molecular evidence from Piping Guans, including a proposed Trinidad Piping Guan which supported the monospecific status, and suggested its closest relationship with the mainland species *Pipile cumanensis* (Blue-throated Piping Guan). However the origin of the individual reported to be a Trinidad Piping Guan in the research by Grau et al. (2005) was disputed by Robinson (2011), due to the sample being obtained from a captive bird in Granja 'La Siberia' in Mexico, of unconfirmed ancestry, which is neither a voucher nor reliable reference sample. Phylogenetic analysis by Robinson (2011) using molecular information from a true Trinidad Piping Guan reference sample (an individual originally from the wild in Trinidad, with species confirmed by both zoologists and veterinarians) re-evaluated relationships among the genus *Pipile*, see Figure 1.2. The Trinidad Piping-Guan was found to be more basal in the phylogenetic tree than in previous research by Grau et al. (2005), highly supported by bootstrap analysis (>90), suggesting the Trinidad Piping Guan was previously misrepresented. The analysis indicates that the Trinidad form is more closely related to *Pipile cujubi* than *P. cumanensis* or *P. cumanensis grayi*.

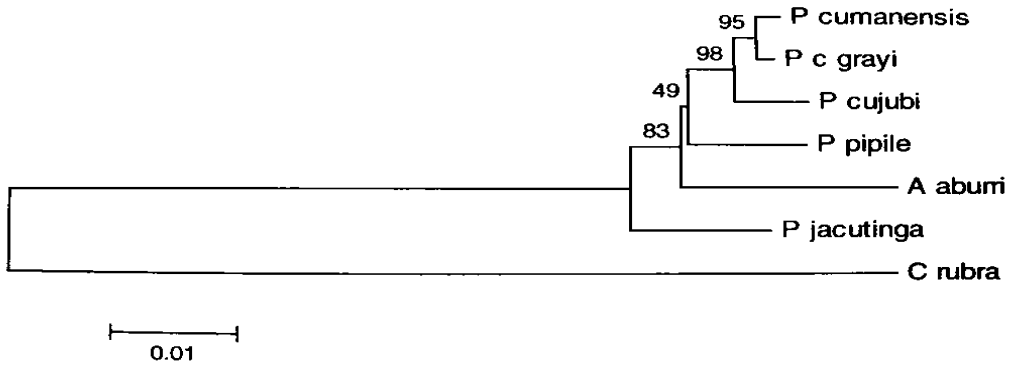


Figure 1.2: Neighbour Joining tree using Tamura-Nei's model of sequence evolution, for phylogenetic analysis of the genus *Pipile.* (Robinson, 2011)

The lack of taxonomic consensus does not alter the conservation need of these species, they remain one of the most threatened groups of avians in South America, and the lack of knowledge of these species highlights the need for continued research.

1.2 The Trinidad Piping Guan

The monospecific Trinidad Piping Guan *Pipile pipile* (Jacquin 1784), known locally as the Pawi, is the only endemic Cracid species on the island of Trinidad. The Trinidad Piping Guan historically was abundant throughout Trinidad with the exception of the drier western coast, is now confined to north-eastern regions of the Northern Range where forest habitat is most prevalent and human population is low (Hayes, Shameerudeen, et al., 2009). There has been a drastic decline in the population in recent decades and current population size of the Trinidad Piping Guan is estimated to be unlikely to exceed 200 individuals, in an area of around 150km^2 of suitable primary forests (Hayes, Sanasie, et al., 2009). Sightings of the Trinidad Piping Guans outside the Northern Range are rare, with no confirmed reports of sightings in the central range since the 1980's and the last unconfirmed reports in the south in the Trinity Hills in 2000 (Hayes, Sanasie, et al., 2009). In 2005 Project Pawi conducted an extensive pilot study in the Northern Range to establish a current population size estimate, however despite local information on sightings, no Trinidad Piping Guans were found during distance sampling. This provides evidence of the cryptic nature of this species and the difficulties involved in the detection of Trinidad Piping Guans. The population is generally considered to be in ongoing decline (BirdLife International, 2015; IUCN, 2017), however the general opinion of members of the local communities in the villages around the Trinidad Piping Guans suspected range is divided, with 38.4% of interviewees indicating that they thought populations were increasing based on their own sightings (Alemu et al., 2005).

Morphology & Ecology of the Trinidad Piping Guan

The Trinidad Piping Guan is morphologically similar to other members of the genus, approximately 60cm in length, with a wingspan of over 75cm (Vaurie, 1967). The Trinidad Piping Guan demonstrates typical Gallinaceous form; large rotund body shape, long necked with a small head and long, strong legs. The species has strongly rounded, concave wings with well-developed pectoral musculature despite its moderate capacity for flight. Distinctive to the guans are the modified quills on the outermost three or four primary feathers of the wing, which permits their characteristic

wing-drumming; this feature is evident in the Trinidad Piping Guan (Delacour & Amadon, 1973; del Hoyo et al., 1994; del Hoyo, Motis, Delacour, & Amadon, 2004).

Pipile pipile is characterised by the much darker feathers of the crest of the head and nape than observed in other members of the genus, the feathers of which have the white colouration restricted to the edges of the feather. The upper wing covert feathers of the species have extensive white colouration, with colouration of the upper sections of the feathers a moderately glossy purplish-brown. The skin of the face and throat is completely bare and a bright cobalt blue, with a broad throat wattle forming a dewlap attached to the throat along its entire length. (Delacour & Amadon, 1973; del Hoyo et al., 1994; Vaurie, 1967).

Little is known of the ecology, behaviour and biology of the Trinidad Piping Guan, recent research has found the species to be elusive and difficult to study (Alemu et al., 2005). The area around Grande Riviere had a relatively stable population of Trinidad Piping Guans between 1967 and the research in 2005 (Alemu et al., 2005); in 2001 there were six individuals regularly observed in an area of around 19 hectares allowing some research to be carried out on the ecology and behaviour of the species (Hayes, Shameerudeen, et al., 2009; Hayes, Sanasie, et al., 2009).

Generally considered a gregarious species, Trinidad Piping Guans were commonly seen in flocks, more recent research has observed Trinidad Piping Guans in small groups, pairs or more commonly individually, possibly as a result of the reduced population size (del Hoyo et al., 1994; Hayes, Shameerudeen, et al., 2009). Recent recorded sightings of free living Trinidad Piping Guans between 2015 and 2016 on the eBird avian species recording system, support small group sizes of one to three individuals, regularly sighted in the Morne Bleu, Brasso Secco and Grand Riviere regions of the Northern Range, Trinidad (eBird, 2012). Grande Riviere records from eBird also demonstrate sightings of small aggregations of six to twenty individuals, which may be indicative of the higher degree of protection in this region associated with Eco-Tourism.

The species is predominantly diurnal, demonstrating increased activity levels during dawn and dusk periods, however the difficulties in observing the species in canopy environments has meant there is little information available on behaviour of the species (Alemu et al., 2005; Hayes, Shameerudeen, et al., 2009). Trinidad Piping Guans are essentially frugivores, a large portion of their diet is primarily fruits and seeds (over 40 species reported), indicating their reliance on healthy primary forest habitats. Cultivated Nutmeg *Myristica sp.* has recently been identified as a significant food source for Trinidad Piping Guans outside primary forest habitats in areas of agriculture (Alemu et al., 2005).

The two greatest threats to the current population of Trinidad Piping Guans come from deforestation leading to the loss of primary habitat, and unsustainable hunting pressure. Deforestation continues to be a major issue in Trinidad's rural communities, aside from conventional timber harvesting and agroforestry, deforestation for cultivation of small scale agricultural crops and construction of settlements is an ongoing concern. Forest cover in Trinidad & Tobago in 2000 was estimated to have reduced to 44%, a decline of 2% since 1990 (Government of the Republic of Trinidad and Tobago, 2010b); actual primary forests are expected to be less than 44% as data includes agroforestry and timber plantations. Quarrying is also increasing in Trinidad in line with the countries construction and economic improvements, with major quarry sites in the Northern Range in the Arima Valley and Valencia regions, increasing deforestation and causing water and noise pollution concerns in this sensitive ecosystem (Government of the Republic of Trinidad and Tobago, 2010b). Trinidad Piping Guans are primarily frugivores and canopy dwellers, therefore deforestation, pollution and habitat destruction has a significant effect on this species, and others in the Northern Range forest ecosystems.

In the past Trinidad Piping Guans had an important role as a source of protein in rural communities, hunting for subsistence is no longer considered to be a current issue however it has become a common past-time and has cultural importance (Hayes, Sanasie, et al., 2009). Between 1999 and 2008 wildlife hunting incidences increased by almost 100%, and has been recognised as a method of supplementing income by

rural demographic groups (Government of the Republic of Trinidad and Tobago, 2010b). Wild meat is known to fetch premium prices during hunting seasons and prices increase out of season, resulting in black market trade in wild meats. Research by Waylen (2009) suggests community support for conservation has increased in line with socio-economic benefits of ecotourism in areas connected to endangered species such as the Trinidad Piping Guan and Turtles. Hunting is declining in Trinidad with increased education and awareness of conservation issues, however it still poses a major threat to the species, and changing culturally entrenched attitudes towards species such as the Trinidad Piping Guan may be difficult.

Hayes et al. (2009) reported evidence of Trinidad Piping Guans becoming habituated to the presence of humans in areas of small scale agricultural plantations, where humans are active on a regular basis within Piping Guan territory, provided canopy trees were left intact in nearby forest. This shows potential for the Trinidad Piping Guan to live in conjunction with human activity, however in the current situation it provides worrying opportunity for hunting of this rare species. Trinidad Piping Guans are thought to have undergone behavioural changes in line with their declining population numbers, once reported to have been seen foraging in large groups on the forest floor (Alemu et al., 2005; Delacour & Amadon, 1973; del Hoyo et al., 1994), current sightings suggest the species has evolved arboreal behaviours rarely being seen outside of the upper forest canopy (Hayes, Shameerudeen, et al., 2009), possibly as a consequence of excessive hunting pressure and habitat contraction.

The Trinidad Piping Guan has been the focus of several major conservation incentives in recent years and is internationally recognised as a species of great concern. The IUCN Red List of Threatened species lists the Trinidad Piping Guan as Critically Endangered, in the category CR C2a(i,ii); the species is currently considered to be in ongoing decline (IUCN, 2017). The IUCN and the Species Survival Committee (SSC) Cracid Specialist Group consider the Trinidad Piping Guan to be an 'Immediate Conservation Priority', ranking third highest in the conservation priority list of Cracid species (Brooks & Strahl, 2000). Since 1975 the Trinidad Piping Guan has been listed in Appendix I by the Convention on the International Trade in Endangered Species

(CITES). The CITES provides the international framework for legislation governing the trade in animals (whole or in part) supported by 175 member states (parties) globally, affording the species the highest level of international protection against exploitation (CITES, 2017).

At a national level in Trinidad, *Pipile pipile* has been designated as an Environmentally Sensitive Species. In addition a large area of the Northern Range known as Matura, which includes primary Piping Guan habitat, has been designated as an Environmentally Sensitive Area affording protection of both the Trinidad Piping Guan and other endemic species (Alemu et al., 2005). However, despite the highest levels of legislative protection for both the Trinidad Piping Guan and the habitat under Trinidad law, environmental and wildlife legislation are some of the most difficult to laws to enforce. The ability to assess accurately the population size through the genetic identification of individuals may be critical to the conservation of this species, however distance sampling techniques have been generally unsuccessful (Alemu et al., 2005). Therefore, genetic analysis may prove critical to population size estimation and provide unambiguous evidence of Trinidad Piping Guans in areas outside the current known locations.

Captive breeding programs have previously been considered for the Trinidad Piping Guan, but no significant efforts have been made to establish a program either on Trinidad or globally. Currently, Species360 (formerly International Species Information Service) lists 35 Trinidad Piping Guans in captivity in Europe, and a further 31 birds in the Americas, however some of the currently listed individuals in the captive population in the UK were found to be misclassified; genetic testing of the mitochondrial DNA suggests evidence of *Pipile cumanensis* maternal ancestry in the captive population, supported by morphological evidence (Robinson, 2011). Were a significant captive breeding program to be established for the Trinidad Piping Guan genetic monitoring would be critically important to avoid loss of species integrity and genetic diversity, and to avoid founder effects such as genetic drift and relaxed selection, which could be detrimental to the future health and survival of the species.

1.3 Evolution, Population Genetics & the Trinidad Piping Guan

Biodiversity is a dynamic balance between speciation and the immigration of species, and the extinction or emigration of species within an ecological niche (Terborgh, 1974). In light of current trends towards diminishing natural diversity it is necessary to understand key concepts of evolutionary and extinction processes within an environment, and the ecology and genetics of the species concerned. Genetic research can inform many questions of evolution, ecology and diversity in endangered species, and the study of population genetics and genetic variation may be critical to the future of many species. Frankham (1997) suggested that the evolution of species' is directly linked to genetic variation; populations evolve and adapt as a direct response to environmental and climatic change, therefore genetic variation within a species is critical for future success of a species.

Diversification within the Cracids is thought largely to be due to ecological and biogeographic causes such as habitat changes, climate changes and marine transgression inducing speciation by fragmentation or isolation of populations (Hosner et al., 2016; Pereira et al., 2002). The analysis of Hosner et al. (2016) suggests the extant Cracidae family originate from a Mesoamerican ancestral Cracid ca. 48-68 Mya, stem Cracids present in North America being influenced by global cooling during the Oligocene and diversifying in South America following the uplift of the Panamanian Isthmus ca. 3.5-10 Mya. Subsequently Cracids diverged in the South Americas in response to biogeographical and climatic events such as the Andean uplift and Amazonian basin river formations; the typical Guan divergence estimated at ca. 3.0-4.8Mya (Hosner et al., 2016).

Grau et al. (2005) suggest a biogeographic hypothesis for the inter-generic diversification of the *Pipile* species, according to correlation of divergence time estimates and major geographical events of the Pliocene-Pleistocene epochs. Geographical events in the history of South America, including the uplift of the Cordillera of the Andes, formation of the Brazilian and Guyanan Shields, glaciation

events, and the formation of major river drainage systems, forced physical separation of the species allowing species to diverge and adapt in accordance with natural surroundings. As a consequence significant areas of endemism arose in the Neotropical regions of South America, with significant influence on the Cracidae species and *Pipile* genus, many of which are endemic to specific biogeographical regions within South America (Carcraft, 1985; Haffer, 1985; Tilston Smith et al., 2014).

Glaciation and the interglacial periods are thought to be significant in the allopatric speciation of the *Pipile* genus, through the fragmentation of Neotropical forest regions during glacial periods and formation of isolated pockets of forest refuge, and by means of sea transgressions during interglacial periods, the resulting rise in sea levels fragmenting forest landscapes and isolating species in the elevated areas. Unique genetic populations may result in speciation events resulting from genetic isolation and adaptation to environmental pressures, or conversely may become genetically impoverished through their isolation and therefore more susceptible to extinction events (Bush et al., 2010).

Trinidad is a continental island off the north-east coast of South America, it is the most southerly of the Caribbean islands and offers a unique ecosystem more akin to the mainland ecosystem than other island ecosystems (Waylen, 2005). Trinidad is a predominantly lowland island (with peaks below 300m) connected to the mainland by a continental platform less than 200 metres below sea level. Speciation of the Trinidad form of Piping Piping Guan, is hypothesised to have occurred as a result of rising sea levels during interglacial periods, isolating a population of Piping Guans on the island. Island environments provide an important system of evolution within species due to the defined geographic boundaries, small geographical size, diverse habitats and ecosystems, and critically the restriction of gene flow within and between populations in relation to oceanic barriers (Emerson, 2002). The diversification of island species may result from processes such as colonisation of neighbouring islands or continental land masses, adaptive radiation (the diversification of foundation populations into species which have adapted to environmental niches), adaptation to vicariant events such as volcanic activity or extreme weather, and speciation through founder events or

population bottlenecks (Emerson, 2002). It has been suggested that the Trinidad Piping Guan form is still found on the Venezuela coastal cordillera (highland peninsula rising up from the continental shelf connecting the mainland and Trinidad), which suggests evidence of the species' origin (Grau et al., 2005). The bio-geographic theory is supported by the limited dispersal ability of Piping Guan species, which are essentially incapable of sustained flight. The species, similar to many other Cracid and Galliforme species, is capable of short flights and gliding as a means of moving within the forest environment, but is unable to attain prolonged flight sufficient to migrate large distances to islands such as Trinidad. Other theories have been suggested for island colonisations in many species including displacement by hurricanes. The molecular analysis of Grau et al. (2002) suggests an estimated diversification of *Pipile pipile* from the mainland form *Pipile cumanensis* between 0.4 and 1.6 Mya.

Island organisms have been shown to be particularly subject to adaptative pressures which may manifest in high levels of morphological diversity. Morphological variation and inter-species molecular variation reflects the dispersal histories of the species and environmental selective pressures, which can therefore be demonstrated by phylogenetic analysis. The use of molecular phylogenetics may provide information about adaptive radiation of species and help to determine processes which have led to species richness in island ecosystems (Emerson, 2002).

Genetic diversity within a species may be directly affected by anthropogenic processes, particularly in relation to habitat fragmentation or destruction and unsustainable hunting pressures. Anthropogenic or geographical fragmentation of habitats and populations leads to decreased geneflow due to reduced dispersal of individuals and reduced population size. Habitat fragmentation and range contraction are particularly destructive to sedentary species such as Galliforme birds which are less resilient to the effects of fragmentation due to limited dispersal strategies. Bush et al (2010) suggests that species with strong dispersal abilities demonstrate stronger genetic structure and differentiation. Population genetic variation in Cracid species has been seen to decrease in relation to habitat destruction in recent decades, as observed

in the *Crax fasicolata,* Bare-faced curassow in South America (Pereira & Wajntal, 2001).

The study of population genetics and genetic variation may be critical to the future of endangered species such as the Trinidad Piping Guan where increasing habitat alteration and fragmentation and limited dispersal strategies are exacerbating population declines. Understanding the genetic effects of fragmentation of habitat and population decline on the Trinidad Piping Guan may be critical to the conservation of species and their ecosystem. Population management of the Trinidad Piping Guan may become necessary to ensure the future of the species, both in situ and potentially in a captive situation. Management of the habitat may be key to the species' survival and maintaining a genetically and physically healthy population.

Genetic factors directly influence the survival and adaptive potential of species, particularly threatened, small or declining populations, where inbreeding and loss of genetic diversity are unavoidable (Frankham, 2005). Declining populations result in a loss of genetic diversity, inbreeding and fixation of deleterious alleles, which may result in reduced adaptive potential and an increased probability of extinction (Bush et al., 2010). The development of genetic bottlenecks in declining populations have long been a concern in conservation genetics, where population numbers decline to such an extent that the genepool becomes critically limited. Population decline or maintenance of bottleneck population levels may be due to genetically determined reduced survival or reproduction rates, environmental or population catastrophes, which limit the growth and recovery of population numbers (Beissinger, Wunderle, Meyers, & Engen, 2008).

Endemic island species are significantly prone to higher levels of extinction, where they are constrained by their environmental niche and more susceptible to environmental change, catastrophic events and anthropomorphic pressures, as is the case with the Trinidad Piping Guan. The type of species is also a factor in extinction risk; large bodied avians with slow reproductive rates are prone to extinction

(Terborgh, 1974). Larger species are more susceptible to anthropogenic influences such as habitat loss and hunting, and are critically limited by higher metabolic demand for resources especially when confined to islands or fragmented habitats.

1.4 Genetic Analysis of the Trinidad Piping Guan

Genetic Analysis & Avian Biological Samples

Non-human biological samples are utilised routinely in a number of research fields including conservation genetics, population research and forensics. Provision of large sample sizes can be a significant limitation to population studies; samples from a mixed range of sources including e*x-situ* conservation samples (zoos etc.) and museum samples can be used to supplement field collections. Several types of biological samples are routinely used in molecular analysis of avian species; feathers, blood, tissue, faecal, eggshell, eggshell membranes or eggshell swabs, each of which varies in the invasiveness of samples collection and the success in the application of genetic analysis (Bush, Vinsky, Aldridge, & Paszkowski, 2005). Genetic samples collected in the field are commonly less than optimal for identification purposes, such as environmentally degraded samples, damaged samples or morphologically ambiguous samples.

Invasive samples include blood or tissue sampling, buccal sampling, plucked feathers and egg collection, these methods have been shown to result in high DNA yields and success rates in genetic analysis, however these techniques require capture and handling of the birds, invasive procedures which may result in direct injury, stress or induce behavioural changes (Bush et al., 2005; Egloff, Labrosse, Hebert, & Crump, 2009; Martín-Gálvez et al., 2011). The most widely used sampling method for avian species is blood, however this method is highly invasive and restricted by the ability of the researchers to collect the birds (Martín-Gálvez et al., 2011; Trimbos, Broekman, Kentie, Musters, & Snoo, 2009).

Non-invasive samples are those which present no adverse effects to the bird in terms of capture stress or injury, and include moulted feathers, egg swabbing or hatched eggshells. Non-invasive techniques to sample avian DNA are important in genetic research and conservation strategies; to gain vital genetic information for individualisation and population research whist minimising disturbance of the individual animals and without interfering with natural behaviours. Traditional sampling techniques require trapping and handling of birds or removing eggs from nest sites, which may result in injuries, stress and behavioural changes or the destruction of viable eggs which may be detrimental to the population (Egloff et al., 2009). Invoking stress may have negative consequences on behaviour and reproductive performance such as nest abandonment, reduced reproductive success and reduced survival rates (Martín-Gálvez et al., 2011). Non-invasive methods of sampling are becoming more common in avian genetics, but often the DNA yield is lower or of poorer quality than in samples collected by invasive techniques. Additionally the origin of the sample may be unknown, as often non-invasive samples are collected from the wild where the material is deposited without directly being observed. Feathers have been used as a non-invasive source of DNA over recent years in many research studies, particularly where the species is endangered, rare or difficult to catch (Hogan, Cooke, Burridge, & Norman, 2008; Johannson, McMahon, Hoglund, & Segelbacher, 2012; Segelbacher, 2002).

Feathers have obvious disadvantages of low DNA yield compared to blood & tissue, however the ease of collection especially with endangered and elusive species is a significant advantage (Johannson et al., 2012). In research by Bush et al (2005) whole blood, tissue, hatched eggshell membranes and plucked feathers result in high DNA yields and 95-100% successful recovery, as compared to 60% recovery rate in moulted feathers and 55% from predated eggshell membranes (Bush et al., 2005). Feather condition is influenced by factors including age, time since shedding and environmental conditions; often the history of a sample is unknown, therefore the degree of degradation is unknown prior to analysis. Feather condition can be characterised by visual examination for signs of degradation to the calamus and barbs of the vane (Hogan et al., 2008). Degradation of the genetic material can be significantly detrimental to the recovery of nuclear DNA and to the recovery of larger

molecular weight fragments of both the mitochondrial and nuclear genomes. Hogan et al (2008) demonstrated that the DNA yield and quality has been shown to vary greatly with feather type, condition and species.

The collection of samples in sufficient numbers for genetic research is problematic in critically endangered species where populations are small or difficult to access. Canopy dwelling avians such as the Trinidad Piping Guan are notoriously difficult to locate (Alemu et al., 2005), therefore access to genetic material is hindered both by life history traits and geography of their habitat landscape. The use of invasive sampling techniques is both labour intensive and distressing to the animal, and impractical in elusive species such as the Trinidad Piping Guan. Shed feathers are an important source of genetic material in such cases, and are a readily available non-invasive genetic source of conservation data. Both the quality and quantity of DNA recovered from shed feathers has been shown to vary greatly according to the feather type, condition and species (Hogan et al., 2008).

The use of mitochondrial DNA as a source of genetic information, allows for the use of almost any biological material as a potential source including those which traditionally have low nuclear DNA recovery, such as feathers (Hogan et al., 2008; Johannson et al., 2012; Segelbacher, 2002; Sorenson & Quinn, 1998). Mitochondrial DNA is particularly useful in the analysis of degraded samples (Ogden, Dawnay, & McEwing, 2009).

Mitochondrial DNA & Analysis of the Mitochondrial Genome

The mitochondrial genome is a compact, closed circular molecule of approximately 16Kb in length, tightly packed with genes, the genome contains little extra-genic or repetitive DNA sequence. The structure of the mitochondrial DNA (mtDNA) molecule is highly conserved among vertebrate and invertebrate species, and has been extensively studied and sequenced in its entirety in many species. The mtDNA genome in vertebrate species contains 13 protein coding genes, two rRNA genes, 22 tRNA genes and a non-coding control region. Gene structure and function are highly

conserved among vertebrate species however the order of the genes within the genome are known to differ in some species (Brown, George, & Wilson, 1979; Mindell et al., 1998). The avian mitochondrial genome differs significantly from other vertebrate orders in the organisation of genes within the genome, notably the rearrangement of the Cytochrome *b* and NADH dehydrogenase subunit 6 gene order within the genome (Mindell et al., 1999, 1998).

Mitochondrial DNA is suited to population genetics and phylogenetic analysis of species due to its high mutation rate, high copy number per cell, lack of recombination and maternal inheritance. The mutation rate of mtDNA is thought to be 6-17 times higher than that of nuclear DNA due to the lack of efficient repair mechanisms as seen in nuclear DNA, this high mutation rate results in sequence variability between individuals and species (Avise et al., 1987; Ballard & Dean, 2001; Brown et al., 1979). Extensive intra-specific and inter-specific variation due to the rapid rate of evolution is of great utility in systematics and population genetics research at the micro and macro evolutionary level. mtDNA is ideal indicator of parental populations and familial relationships, due to the mode of inheritance through the maternal line, and is therefore indicative of evolutionary changes. However maternal lineage bias may be of significant importance to the assessment of genetic variation at the species or population level, both in terms of genetic inheritance and behavioural bias of female dispersal. The high evolutionary rate gives rise to both intra-species and inter-species variations within the mtDNA sequence which provides the basis for species identification. However analysis of mtDNA within a population is limited by its maternal mode of inheritance, therefore providing an incomplete picture of a population. Mitochondrial DNA demonstrates great potential for analysis of genetic variation in populations, however has a lower discriminatory power than bi-parental analysis of nuclear DNA (Rhymer & Simberloff, 1996; Szibor, Michael, Plate, & Krause, 2000).

In the case of non-human biological samples it is common that animal material is morphologically indistinct, highly degraded or are incomplete, such as partial feather quills, which may mean analysis using DNA is difficult (Tobe & Linacre, 2008). The

quantity of mitochondria present in a cell vary with the type of tissue, however mitochondria often number in the thousands per cell, as compared to the single copy of nuclear DNA (Tobe & Linacre, 2009). Mitochondrial DNA composes a significant percentage of the total volume of cellular DNA means that it is useful in such circumstances. The circular structure and smaller size of mtDNA also means that it suffers less from the effects of degradation than nuclear DNA, meaning mtDNA is often successful when nuclear DNA analysis is incomplete.

Various locations on the mitochondrial DNA genome have been used in research involving the identification of non-human species and phylogenetics including the Control Region, Cytochrome *b*, and Cytochrome Oxidase sub-unit I. These genes all contain regions which are highly conserved, useful indicators of species identification and phylogenetic relationships, and regions which are more variable, which aid in intra species identification (Linacre, 2008). More recently analysis has focussed on complete mitochondrial genome analysis to infer both species identity and relationships at a genomic rather than gene level, enabling a more accurate representation of the genetics of a species.

Protein Coding Mitochondrial Genes

Protein-coding genes contain regions which are highly conserved which facilitate cellular function, these regions are selectively constrained resulting in high levels of conservation which can provide useful markers for species identification and phylogenetic relationships. Variation within species in protein-coding structures is largely confined to third codon position variations, which don't necessarily alter the amino acid code for the gene, therefore preserving cellular function. However, as distance between species increases more variation is seen in the first and second codon positions in addition to increased third position substitutions, which alters the amino acid sequence and may be informative of species (Sorenson, 2003).

Cytochrome *b* (Cyt *b*) is a widely used mtDNA gene, validated for species identification in forensic investigations (Branicki, Kupiec, & Pawlowski, 2003), and is also extensively used in taxonomic research and as a standard reference site in

international DNA reference databases (Tobe, Kitchener, & Linacre, 2009). Research by Grau et al (2005) used both Cytochrome *b*, to infer phylogenetic relationships among the *Pipile* genus, this sequence data has been published in the GenBank DNA database. Pereira et al. (2002) also used Cytochrome *b* combined with other protein-coding mtDNA genes and nuclear gene sequences to disseminate relationships within the *Cracidae* family.

Cytochrome Oxidase subunit I (COI) is used commonly as a species identifying tool, known to show significant variation between even closely related species, its use is increasing in reference collections including Genbank and Bar-coding for Life (CBOL). COI is particularly useful to species identification in that only small fragments (~100bp) are necessary to identify sufficient differences in the sequence to identify some species (Dawnay, Ogden, McEwing, Carvalho, & Thorpe, 2007).

NADH dehydrogenase subunit 2 (ND2) is the third most variable gene in avians, after ATPase8 and NADH dehydrogenase subunit 6 (Sorenson, 2006). ND2 benefits from ease of amplification in most birds, compared to ND6 which is more difficult, and is sufficiently long to be valuable in species comparisons and intra-specific variation compared to ATPase8 which is comparatively short and therefore less useful (Sorenson, 2003). Recent research by Robinson (2011) enabled the identification of Trinidad Piping Guan samples using species-specific nucleotides in both the Cytochrome *b* and ND2 genes.

Non-coding Mitochondrial DNA

The non-protein coding region of the mitochondrial genome centres around the variable length Control Region which responsible for regulation of transcription and replication of the mitochondrial DNA (Pereira, Grau, & Wajntal, 2004). The Control Region is suited to a variety of identification techniques both inter- and intra- specific, its hypervariable regions have a high rate of evolutionary change, therefore may be indicative of phylogenetic relationships both within and between species. Control

Region analysis has been considered challenging due to the high mutation rate, large variation in length between species, the presence of hypervariable nucleotide positions and unequal nucleotide frequencies which confound analysis (Ruokonen & Kvist, 2002).

Pereira et al. (2004) analysed the control region of several species of Cracid and found the Control Region to be consistent in structure to other Galliformes and Anseriformes, with length ranging from 1145bp in *Penelope orchrogaster* to 1197bp in *Pauxi pauxi.* Variation between cracid species was substantial with only 77% of nucleotides homologous, with the highest levels of variation observed in domains I and III flanking a central conserved domain.

Species identification

Molecular identification of the species is often best achieved by the use of mtDNA, which is more stable and resilient than nuclear DNA and so less prone to degradation, mtDNA is also present in higher copy numbers per cell making it more readily available for use in molecular identification (Bellis, Ashton, Freney, Blair, & Griffiths, 2003; Tobe & Linacre, 2009). The current standard method of species identification is to align and compare the mtDNA sequence of the unknown animal with the reference sequence for the suspected species(s) in a database, such as GenBank or CBOL (Bar-coding Life) (Ogden et al., 2009). The level of confidence of the identification is based on the sequence regions compared (conserved or variable), length of the sequence compared and the percentage of the sequence that matches the reference sequence (Linacre, 2008; Ogden et al., 2009). There is no set guideline as to what percentage variation constitutes an inclusion or exclusion in species identifications using mtDNA sequencing (Tobe et al., 2009). It should be noted that it is possible in cases where the species is rare or where little research has been carried out on the species, that the reference sequence in the DNA database may be unreliable.

In research into the genetic identification of the Trinidad Piping Guans by Robinson (2011), species identification was proposed using gene sequencing, species-specific

primers and RFLP analysis of the Cytochrome *b* and ND2 mtDNA genes. The research however demonstrated that sequences in Genbank attributed to *Pipile pipile* (accession numbers JQ175861 and JQ175862) were misidentified, and were more probably attributed to *P. cumanensis grayi* (Robinson, 2011).

Nuclear DNA & Analysis of the Nuclear Genome

Much research has been conducted on the avian genome and the genomics of both natural and domestic bird populations to attempt to understand fundamental evolutionary processes and population dynamics. Avian genomes are largely homogenous in size with less variation than seen in mammals, and strong conservation of the karyotype has been observed between bird species (Backström, Fagerberg, & Ellegren, 2008).

Nuclear DNA sequencing to estimate the relationships among taxa is increasingly being used; nuclear genes provide independent estimates of evolutionary history, whereas mtDNA genes are linked and higher levels of homoplasmy and substitutional bias occur due to the high rate of evolution in the mitochondrial genome (Pereira, Baker & Wajntal, 2002). Nuclear introns have been demonstrated to be useful in avian phylogenetics due to the higher rate of genetic variation, in comparison to mitochondrial genes (Wang, Kimball, Braun, Liang, & Zhang, 2013).

In the analysis of avian inter-specific and intra-specific variation many genes and gene regions have been proposed with varied utility across taxonomic divisions (Kimball et al., 2009; Matthee et al., 2007). Commonly used gene regions include the exon sites of the RAG1, CLTC and ZENK genes (Chojnowski, Kimball, & Braun, 2008; Chubb, 2004; Miller et al., 2008) and intron sites of the Beta-fibrinogen and Ovomucoid genes (Crowe et al., 2006; Morgan-Richards et al., 2008; Prychitko & Moore, 2000). Nuclear introns evolve at a more rapid rate than other gene regions therefore they demonstrate utility in phylogenetic analysis of inter-specific relationships (Armsrtong, Braun, & Kimball, 2001; Prychitko & Moore, 2000).

Research by Pereira, Baker & Wajntal (2002) found combined nuclear and mitochondrial DNA sequences provided the best estimate of phylogenetic relationships among the 11 genera of the family Cracidae; sequences complement each other and support nodes at various depths within phylogenetic trees. The longer sequence length reduces stochastic error in substitutions, evidenced by the increased bootstrap support at key nodes and congruence between the independent trees and combined trees increases the confidence in the phylogenetic and therefore evolutionary relationships (Pereira et al., 2002). Comparative analysis by Hosner et al. (2016) of the evolution of the Cracidae family used a multi-gene approach incorporating both mitochondrial DNA and nuclear DNA sequence information. A combined gene approach produces a robust analysis which conforms to morphological and biogeographical knowledge, however variation was observed between the mitochondrial DNA and nuclear DNA analyses individually which represents differing evolutionary mechanisms of the genomes (Hosner et al., 2016).

The analysis of multiple gene region sequences has been greatly debated. Separate sequence analysis can determine the congruence between genes and reflect on the similarities between species trees, providing confidence in the assignment of clades. Combined sequence analysis gives a more robust estimate of relationships due to the larger amount of data. Combined data can be useful in resolving relationships and providing improved tree resolution, however incongruence between the data sets may be obscured or remain undetected (Pereira et al., 2002).

Microsatellite markers are one of the most commonly used genetic markers in population and conservation biology research. Microsatellites can provide essential information for conservation and species management, including genetic diversity, population structure, parentage and dispersal patterns (Ping-ping, Ping, & Sheng-Guo, 2006). Multi-locus genotyping is more reliably achieved in high quality DNA samples, Hogan et al (2008) recommended a screening process for samples prior to genotyping; including visual analysis of condition and test amplification of a single nuclear DNA loci of large size (~400bp) such as CHD-1. Genotyping of non-invasive feather samples with low volumes of DNA has been shown to be difficult in previous research.

Segelbacher (2002) identified issues associated with the use of low DNA quality feather samples in microsatellite analysis of the Capercaille (*Tetrao urogallus*). Johannson et al (2012) compared the microsatellite amplification success of shed feathers and plucked feathers from the Red Grouse (*Lagopus lagopus hibernicus*), and found success of multilocus genotyping was reduced in shed feathers (23% of shed feathers successful, compared to 61% overall successful amplification). Johannson et al (2012) found microsatellite amplification success to be highly dependent upon size of the feather; large feathers from the wing or tail yield better quality DNA than coverts or body contour feathers, and larger birds also have higher amplification success rates. Feathers have the obvious disadvantages of low DNA yield compared to other biological sample types, however the ease of collection especially with endangered and cryptic species is a significant advantage.

Cross-species microsatellite use may be advantageous in developing molecular markers for the Trinidad Piping Guan, through the use of 'universal' primers which can amplify homologous PCR products in multiple species. Some primers amplify microsatellites in closely related species, however success in cross-species microsatellite amplification relates to the evolutionary distance between species (Galbusera, van Dongen, & Matthysen, 2000; Primmer, Moller, & Ellegren, 1996; Primmer, Painter, Koskinen, Palo, & Merila, 2005). Conservation of microsatellite loci between related bird species has been demonstrated to be significant (Reed, Mendoza, & Beattie, 2000). Cross-species amplification has been demonstrated extensively in Galliforme species, most in relation to the domestic chicken or turkey, including the Red-legged partridge (Ferrero, González-Jara, Blanco-Aguiar, Sánchez-Barbudo, & Dávila, 2007), and the Common Pheasant (Baratti, Alberti, Groenen, Veenendaal, & Fulgheri, 2001). Several microsatellite markers have been developed in for use in the Wattled Curassow (*Crax globulosa*), as part of genetic monitoring in a breeding programme (Hughes & Larson, 2000). The utility of such cross-species microsatellite analysis in Trinidad Piping Guans will depend largely on the degree of variation observed between the species and the homology of loci within the species' genomes.

Avian Gender Determination

Accurate gender determination is important for veterinary care, breeding programs, ecological research, enforcing legislation and resolving paternity and maternity (Griffiths et al., 1998; Lee et al., 2010; Ong & Vellayan, 2008). Gender determination in many species is simple, in that most species are sexually dimorphic. In avian species, including many of the Galliformes, it is common for the sexes to display colour dimorphism, with females and males of the same species presenting characteristically different plumage.

Conventional techniques for sexing animals, where there are no external morphological indicators or where only a small biological sample is available, often use the analysis of genes present on the sex chromosomes. Molecular gender identification in avian species uses genes on the W and Z chromosomes, commonly the Chromo-helicase DNA binding gene (CHD) which is present on both chromosomes. Female birds have both W and Z chromosomes (W,Z), whereas male birds have two Z chromosomes (Z,Z), therefore CHD-W is specifically female allowing accurate gender identification. Variation in the fragment size for the CHD-W and CHD-Z genes observed between species allows for the determination of both sex and species (even between closely related species) (Dawson et al., 2001; Griffiths et al., 1998; Lee et al., 2010; Ong & Vellayan, 2008).

Gender determination is important to many facets of conservation genetics including establishing the health and genetic status of wild populations (sex ratios, seasonality and breeding structures) and establishing successful captive breeding programmes through the use of genetic monitoring. Establishing a successful genetic test for gender determination in Trinidad Piping Guans may be beneficial to both in-situ and ex-situ conservation strategies in the future.

Phylogenetic Inference

Phylogenetics is an essential tool in modern conservation management and taxonomic practice. Modern phylogenetic techniques enable evolutionary hypotheses including

common ancestors, adaptation, diversification and biogeography to be tested in the context of genetic relationships at the gene and genome level. The evolutionary relationships within the Avian class have been highly debated, modern insights through genetic analysis has provided many insights into species relationships however deep relationships and early divergences in evolutionary history remain unresolved (Hackett et al., 2008). Conflicting phylogenetic analyses compound the uncertainties within the avian evolutionary tree; discrepancies between morphology, molecular markers, mitochondrial and nuclear genomic data highlight these conflicts. Clear discrepancies and conflicts have been reported in the analysis of modern avian (Neoaves) relationships between mitochondrial and nuclear genomics, and the analysis of gene trees rather than species trees (Braun & Kimball, 2002). Biases in sampling of taxonomic groups has also resulted in significant conflicts, where taxonomic representation directly influences the phylogenetic reconstruction.

Current practices in genetic taxonomic systematics have highlighted the deficit in taxon sampling within collections and databases (Eo, Bininda-Emonds, & Carroll, 2009). Often less well studied species, such as the Piping Guans, are poorly characterised and lack sufficient sampling in taxonomic collections and DNA databases. Such biases in taxon sampling and insufficiencies in genetic data collections infer error within phylogenetic hypotheses.

Mitochondrial DNA has been used in phylogenetic analysis extensively, its relatively low rates of evolution mean it is a good indicator of evolutionary relationships, patterns of radiation and diversification and temporal indication of evolutionary diversification (Braun & Kimball, 2002). Mitochondrial gene function constrains mutation within the genome, however synonymous site mutation (3^{rd} codon) enables the genome to evolve quickly whilst maintain function. The mitochondrial genome is widely considered to conform to a predominantly selectively neutral model, with a relatively constant substitution rate through time and through the retention of functionality of genes within the genome reducing the influence on fitness. However, response to environmental stress or genetic adaptation may infer positive or purifying selection without loss of functionality or fitness effects (Bazin, Glémin, & Galtier, 2006).

Phylogenetic hypotheses may be heavily influenced by the method of analysis, the model used and the amount and type of genetic data selected for analysis. Complete mitochondrial genomes have been widely used in avian genomic and evolutionary analysis and have demonstrated significant utility in resolving complex relationships and histories (Cooper et al., 2001; Guan et al., 2009; Slack et al., 2006). New techniques in molecular biology have enabled the use of whole genomes in phylogenetic analysis of avian relationships (Jarvis et al., 2014; Prum et al., 2015). The use of whole genes or complete genomes is widely accepted as a robust and reliable technique to demonstrate phylogenetic support and analyse genetic relationships.

Aims & objectives of the research

The main focus of the research is to develop methods of analysing genetic variation within the Trinidad Piping Guan species, and examine the evolutionary relationships within the genus *Pipile.* Specifically the study will focus on the relationship of the critically endangered Trinidad Piping Guan to mainland conspecific species, and the relationships of the genus *Pipile* within the Galliforme order. The study will analyse the complete mitochondrial genome of the Trinidad Piping Guan for the first time, using specimens verified as having origination from the wild in Trinidad. In order to achieve this, the study will examine techniques for the extraction and amplification of DNA from specimens of varing tissue type, and develop protocols for molecular analysis using feathers, museum skins and embryonic tissue. The production of a reference mitochondrial genome for the Trinidad Piping Guan will enable the examination of the species for the first time at a genomic level, and provide insights into both intra-specific and inter-specific variation within the genus *Pipile*.

The study will demonstrate the use of advanced PCR techniques to amplify DNA fragments from temporally and environmentally degraded avian samples, and outline molecular analysis techniques to characterise intra-specific and inter-specific variations in the *Pipile* genus. Samples from museum specimens will supplement contemporary samples from both the wild in Trinidad and captive Piping Guans in the UK, providing a historical perspective of inter-specific relationships. Voucher samples sourced from museum collections, where the species has been identified and classified

by experts, will therefore enable the development of taxonomically informative markers for species of the genus *Pipile* for the first time. In addition, reference samples collected from contemporary wild Trinidad Piping Guans which have been identified by expert zoologists and veterinarians, will be utilised to ensure the accuracy of species assignment throughout this analysis. The research aims to develop a database of samples from wild and captive Trinidad Piping Guans, congeneric species and museum voucher specimens. This collection enables comprehensive comparisons to be made between the samples and analyse the current status of the Trinidad Piping Guan population.

Throughout this study the analysis will attempt to answer several important questions in relation to the genetics of the Trinidad Piping Guan, which may be beneficial to future conservation management of the species.

Is there sufficient variation in the 'Trinidad island type' to distinguish it from mainland species of Piping Guans, of the genus Pipile? Establishing the Trinidad Piping Guan as an evolutionarily distinct endemic species, will reinforce the conservation status of the species and highlight the need for protection of this unique island form of the genus *Pipile*.

Can the Trinidad Piping Guan be unambiguously identified using Single Nucleotide Polymorphisms in the DNA sequence? Developing species indicative genetic markers will aid in the unambiguous identification of these species, particularly with reference to captive population management and the introgression of hybrid common Piping Guans in the UK cative population.

What is the current status of genetic variation in the wild Trinidad Piping Guan population? Has the decline in population size of the Trinidad Piping Guan led to a contemporary genetic bottleneck which could be detrimental to the future of the species.

Chapter 2. General Experimental Methods

2.1 Introduction

The study focuses on the evaluation of genetic information from the Trinidad Piping Guan and its conspecifics. Predominantly the analysis utilises sequence data from the mitochondrial DNA genome and short regions of nuclear DNA genome. Amplification of both mtDNA and nDNA requires the extraction of high quality DNA from a variety of biological sample types, which in the course of this study have included feathers, fresh tissue and historical tissue samples.

Polymerase Chain Reaction (PCR) provides the basis for the amplification and subsequent analysis of DNA throughout this study; as a relatively quick, robust highly specific tool for the exponential amplification of DNA for genetic analyses. PCR is an enzymatic reaction under specialised thermal conditions which, through the highly specific binding of oligonucleotides to a template DNA, copies and replicates the template exponentially. In all cases standard PCR techniques were used throughout, with modifications made as necessary to optimise the recovery of target regions of mitochondrial and nuclear DNA. Modification to the standard PCR protocol was used for the analysis of old DNA and ancient DNA, to enable the amplification of highly degraded DNA from museum specimens of significant age and unknown preservation history.

2.2 The Sample Collections

2.2.1 Biological Sample Types

The use of non-invasively collected genetic material is one of the most appropriate methods of gaining genetic information for critically endangered species. Biological materials shed by avian species such as feathers, faeces or eggshells are readily available and can be collected without distress to the individual. However non-invasive sampling has a significant disadvantage in that there may be issues associated with

environmental degradation and contamination. Degradation over time and as a result of environmental conditions has a direct impact on both the DNA quality and the quantity of DNA recovered; the recovery of nuclear DNA can be particularly problematic with degraded samples (Hogan et al., 2008).

In order to counter the problems associated with low quality DNA recovery from samples such as shed feathers Hogan et. al. (2008) and Segelbacher (2002) recommended, pre-screening of samples, replication of analysis and use of larger sample sets to compensate for lost data. Such visual analysis of the shed sample can be indicative of the extent of degradation that can be expected and as a consequence the likelihood of amplification success (Hogan et al., 2008).

2.2.2 Study samples obtained from the wild

DNA was extracted from moulted feathers, egg contents including embryo liver tissue and footpad skin samples, ethically collected from Piping Guans of the genus *Pipile (Aburria)*, and sourced both internationally and from the UK. Reference material in the form of moulted feathers was sourced from captive Trinidad Piping Guans in the collection at Port of Spain Zoological Gardens, Trinidad. Samples were obtained from two individual Trinidad Piping Guans at the zoo; the first samples were from a bird which died in 2007 (Reference individual 1) and the second samples are from the Piping Guan housed in the zoo in 2010 (Reference individual 2). Both of the individuals from the Port of Spain Zoo collection originated in the wild in Trinidad (unknown locations) and were rescued and rehabilitated by the zoo. These two samples can therefore be considered reference examples of the species as they originated in the wild on the island of Trinidad, and have been independently classified by both veterinarians and ornithological experts as conforming to Trinidad Piping Guan morphological and taxonomic species indicators (personal communication Cooper, J).

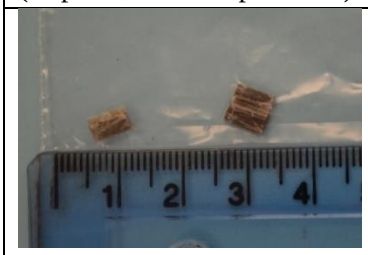
		
TPG Reference 1. Northern Range, Trinidad. (Captive, Port of Spain Zoo)	**TPG Reference 2.** Northern Range, Trinidad. (Captive, Port of Spain Zoo)	**TPG Egg/embryo 1.** Northern Range, Trinidad
		
TPG Pawi 1. (Quill only) Morne bleu, Trinidad	**TPG Pawi 2.** Plato Plantation, Grande Riviere, Trinidad	**TPG Pawi 3.** Plato Plantation, Grande Riviere, Trinidad
		
TPG Pawi 4. Matelot, Grande Riviere, Trinidad	**TPG Pawi 5.** Matelot, Grande Riviere, Trinidad	

Figure 2.1: Samples from wild Trinidad Piping Guans; Moulted feathers, Egg & embryo and Quill sections. (All photographs provided by A. Grass except the Egg & embryo images provided and printed with permission of by Prof. John Cooper).

In addition to the reference individuals, further samples were collected by field researchers working in the Northern Range in Trinidad. Moulted feathers were collected opportunistically in the field by researchers from University of the West Indies and the Pawi Study Group in the Northern Range, Trinidad.

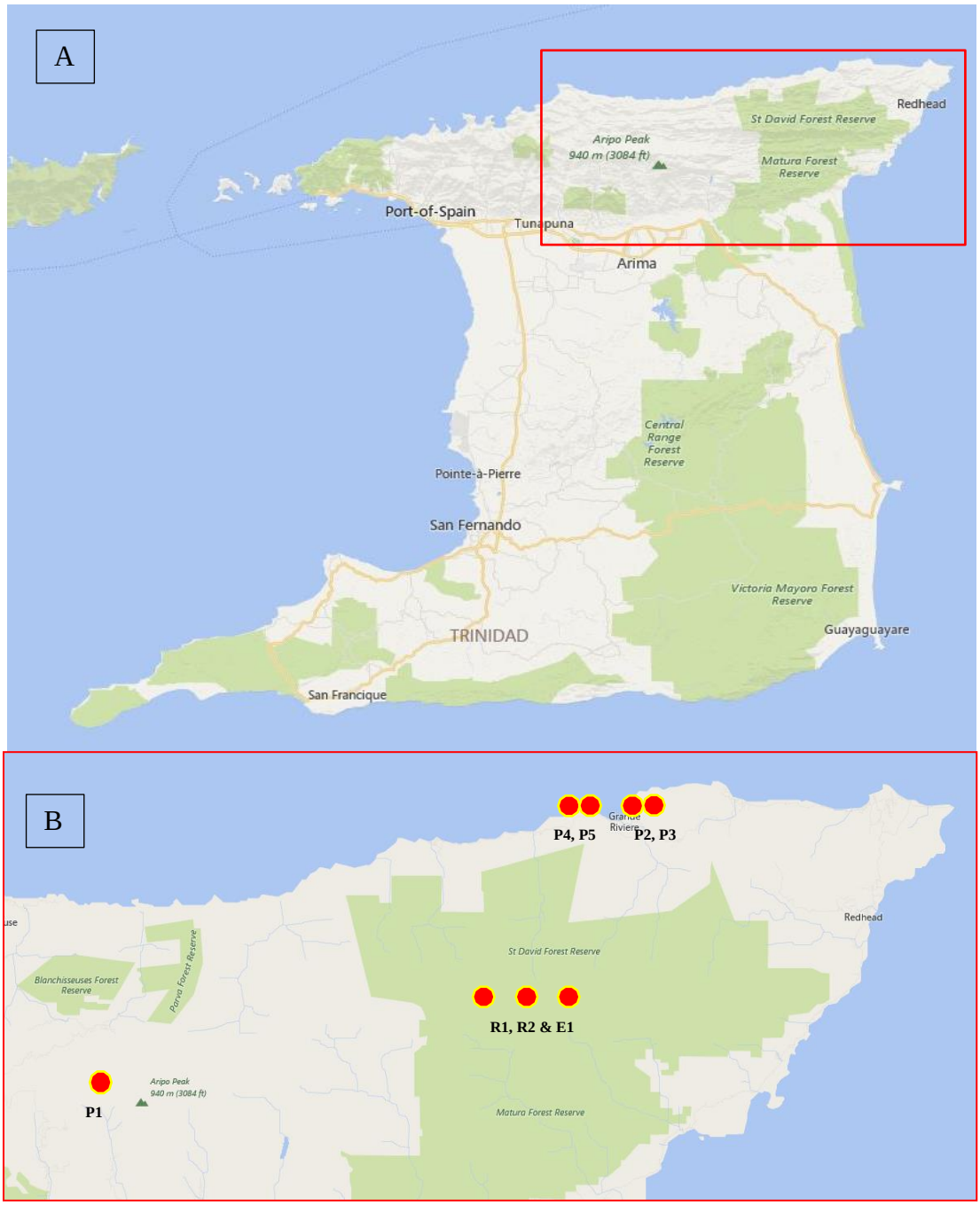


Figure 2.2: Map of the island of Trinidad (A), highlighting the Northern Range region (B) and indicating the origins of the Trinidad Piping Guan samples. Samples Reference 1, Reference 2 and the Egg/embryo 1 (R1, R2 & E1) are of sourced from the wild in an unknown location of the 'Northern Range'.

Samples were also obtained from an egg recovered from the ground below a nest site in the Northern Range, the embryo being non-viable. The egg and embryo were morphologically identified as Trinidad Piping Guan by a veterinarian and ornithological expert and samples of liver tissue, developing feathers and

chorioallantoic membranes were subsequently harvested for use in DNA analysis. In total five Trinidad Piping Guan feathers (including one Quill only sample) and one Egg/embryo sample were sourced from the wild in the Northern Range region of Trinidad; and thus a total of eight individual Trinidad Piping Guans of wild origin were analysed throughout this research. Trinidad Piping Guan sample details and images are shown in Figure 2.1 and geographical origins of the samples are illustrated in Figure 2.2. Further information of all samples are provided in Appendix 1.

2.2.3 Samples obtained from Captive Piping Guans

In addition to the samples collected from wild Trinidad Piping Guans, nine samples were also sourced from UK zoological collections and museum collections. In this case moulted feathers were collected from captive Piping Guans from zoological collections in the UK (Details provided in Appendix 1). These individuals are listed under the Species360 (formerly International Species Information System, ISIS) taxon records as Trinidad Piping Guans, however morphological and molecular evidence suggests these birds may have been miss-classified and are more likely to be Blue Throated Piping Guans, *Pipile cumanensis* or hybrids (Robinson, 2011).

2.2.4 Samples from Museum Collections

Footpad skin samples, comprising of nine individuals from the genus *Pipile* (*Aburria*), were obtained from Piping Guan specimens held in the British Natural History Museum, Tring Ornithological collection, under guidance of the curator. These birds were collected from a wide range of locations across South America, and range in age from 64 to 161 years old (collected between 1856 and 1953). Classification of these specimens by the British Natural History Museum and subsequent research (Lopes, 2009), allows the use of these specimens as vouchers in the analysis of inter-specific genetic variation.

Details of the museum specimen Piping Guan samples and images of the specimens are shown in Figures 2.3 and 2.4. The origins of the specimens collection is illustrated in Figure 2.5 (in accordance with specimen labels).

		
BNHM-AC1 1902.3.13.1852 Pipile cumanensis, male Peru, The Andes, S. Am. P.O. Simons, 12th March 1900	**BNHM-AC2** 1953. 68. 44 Pipile cumanensis, male Cerro Galera, Orienie, Equador Leopoldo Gomez, 12th July 1940	**BNHM-AC3** 1940. 12. 5. 43 Pipile cumanensis, male Rio Juno, E. Equador L. Gomez, 8th Dec 1938
		
BNHM-AC4 1856.11.5.10 Pipile cumanensis Columbia Verreaux, 1856(?)	**BNHM-AC5** 1922. 3. 5. 185 Pipile cumanensis, (male 2002) Pomeroon River, Brit Guiana McConnell collection	**BNHM-AC6** 1892.1.16.138 Pipile cumanensis, Male Takutu River, Brit. Guiana H. Whitely, 1888
		
BNHM-ACG1 1910.7.9.107 Pipile cumanensis (grayi), Male Concireucia, Alto, Paraguay G. W. Tudor, 9th Sept 1909	**BNHM-ACJ1** 1889.6.1.254 Amazonian Guan, Pipile cujubi Lower Amazon Zoological Society of London (Ex-captivity) 25th May 1873 (died 9th May 1875)	**BNHM-ACJ2** 1850.11.30.21 Penelope atrofurpurea? Pipile cujubi S. America Zoological Society of London Burton 19. V. 97

Figure 2.3: Samples from Piping Guans held in the British Natural History Museum ornithological collection Tring, UK (BNHM). All photographs provided by A. Grass and printed with permission of the British Natural History Museum.

In addition to the Tring samples, a further six feather samples were made available by the Museu de Zoologia da Universidade de Sãо Paulo, Brazil (MZUSP) from Cracids in their ornithological collection. This sample set includes individuals of the *Aburria* genus (synonym *Pipile*). Information regarding the geographical origins of these specimens was sourced from Lopes (2009).

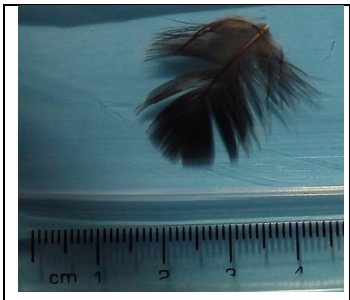
		
MZSUP-AC1 MZUSP 76361 Aburria cumanensis Rio Macauà, Brazil	**MZSUP-ACJ1** MZUSP 20832 Aburria cujubi (male) Caxiricatuba, Rio Tapajós, Brazil	**MZSUP-ACG1** MZUSP 'no number' Aburria grayi Mato Grosso (?)
		
MZSUP-CF1 MZUSP 79216 Crax fasciolata	**MZSUP-PT1** MZUSP 28037 Pauxi tuberosa	**MZUSP-ACJ2** MZUSP 21947 Aburria cujubi (male) Piquiatuba, Rio Tapajós, Brazil

Figure 2.4: Samples from Piping Guans held in the Museu de Zoologia da Universidade de Sà̀o Paulo, Brazil (MZUSP). All photographs provided by A. Grass.

The geographical distributions of the specimens are estimated in Figure 2.5 based on information supplied on specimen labels. Although the accuracy of location is limited by the information available, the geographical distributions are consistent with current information regarding species ranges in South America (Grau et al., 2005).

Further information of all museum collection specimens is provided in Appendix 1.

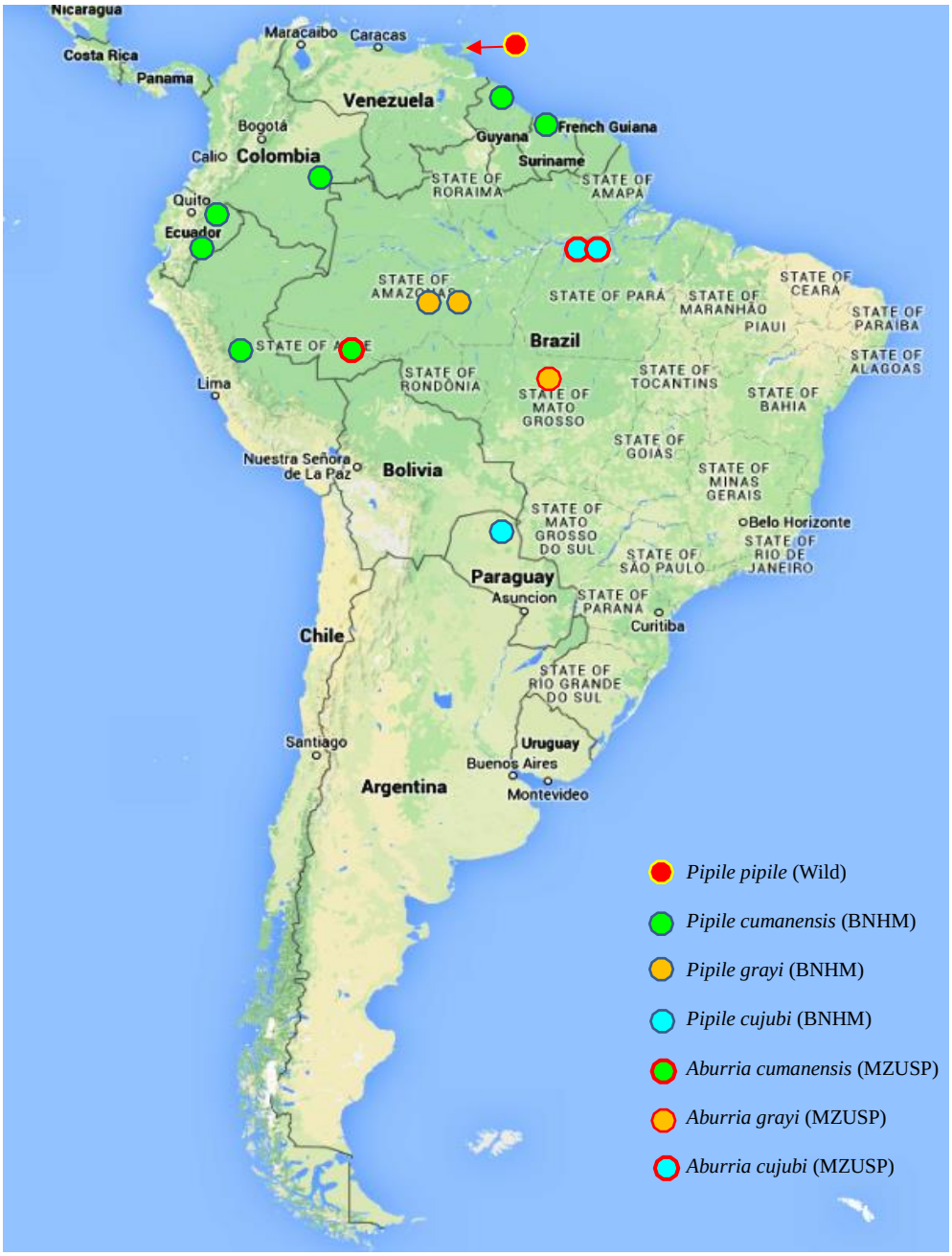


Figure 2.5: Map of sample origins for the wild Trinidad Piping Guan samples (*Pipile Pipile**) and samples from the British Natural History Museum collection (*P. cumanensis, P. grayi & P. cujubi**) and specimens from the Museu de Zoologia da Universidade de São Paulo collection (*A. cumanensis, A. grayi* & *A. cujubi**).
* genus classification *Aburria* synonym *Pipile*

2.2.5 Further taxa

In addition to samples from the genus *Pipile* other Galliforme species were used as control samples and test specimens for optimisation of techniques. Samples of *Crax blumencachii* (Red-billed Curassow), family Cracidae, were used as a species control for amplifications and as an outgroup for species variation analysis. In this respect,

moulted feather samples of *Crax Blumenbachii* were donated by Chester Zoological Gardens, Chester, UK. In addition domestic Chicken (*Gallus gallus*) and Ring-necked Pheasant (*Phasianus colchicus*) samples were also selected as model specimens due to their extensive use in published research (critically the use of the domestic chicken in the design of universal avian primers), and utilised as positive reference controls.

2.2.6 CITES permits & DEFRA import licences

The Trinidad Piping Guan (*Pipile pipile* synonym *Aburria pipile*) is listed under the Convention on the International Trade in Endangered Species (CITES), Appendix I species list. Therefore the importation of Trinidad Piping Guan biological samples including feathers and tissues, requires a CITES permit to cross international borders; all samples from Trinidad Piping Guans were imported with CITES permits as listed in the full sample inventory (Appendix 1), approval granted by the Trinidad Forestry Division.

In addition to the CITES permit the Department for Environment, Food and Rural Affairs (DEFRA) require an import licence for vertebrate biological samples from countries outside the Europe. DEFRA conditions of import required precautionary heat treatment of samples at 56°C for a period of 30 minutes to prevent the importation and spread of infectious agents into the UK. DEFRA import licence numbers are listed in the full samples inventory information in Appendix 1.

2.2.7 Ethical Approval

Ethical approval to conduct the research was granted by the Faculty of Applied Sciences Research Ethics Committee in December 2012 (FREC reference number 734/12/AG/BS). No ethical concerns were raised regarding the sampling of critically endangered avians, as all samples were sourced through non-invasive techniques, primarily through opportunistic collection of moulted feathers.

The British Natural History Museum granted permission to use destructively invasive techniques on specimens from their collection under guidance of a curator, therefore no ethical concerns were raised in this respect.

2.3 DNA Extraction Procedures

DNA was extracted from all samples using a silica based technique. The Qiagen DNeasy Blood and Tissue Kit™ (Qiagen) enables extraction of high yields of both nuclear and mitochondrial DNA and can be modified to support DNA extraction from varying tissue types and quality.

Silica based extraction techniques, such as the DNeasy Spin Columns (Qiagen), have been shown to be of great utility in the extraction of DNA from feathers and other biological materials (Johannson et al., 2012; Trimbos et al., 2009) and have been widely used in the extraction of DNA from animal tissue, including those of significant age (ancient DNA). Due to the high binding capacity of silica and its minimal effects on PCR inhibition (Höss & Pääbo, 1993; Yang, Eng, Waye, Dudar, & Saunders, 1998), silica based spin columns are an ideal method for the extraction of DNA from biological samples. Silica membranes are efficient in both the adsorption and desorption of DNA molecules, in the presence of suitable buffers, therefore silica based techniques are highly efficient at maximising the DNA yield (Tian, Hühmer, & Landers, 2000). In addition DNA molecules are easily released from silica based surfaces in the presence of low ionic strength buffers and elution of the DNA in this fashion is advantageous for further downstream applications such as PCR which is salt sensitive (Tian et al., 2000).

2.3.1 DNA Isolation from Feathers

Qiagen DNeasy Blood and Tissue Kit, User Developed Protocol for Feathers

Cell lysis was carried out by incubation of the sample in a Proteinase K solution. Approximately 10mm of quill section was cut from the feather sample and placed in a

1.5ml micro-centrifuge tube, 10 µl of DTT (1,4-dithiothreitol), 360 µl of ATL Buffer and 40 µl of Proteinase K (>600 mAU/ml) were added, and mixed by vortexing. The samples were incubated at 56 °C overnight to complete lysis. Where possible quill sections were taken from the region of the inferior umbilicus of the feather to ensure the highest possible quality of both nuclear and mitochondrial DNA was extracted (see Figure 2.6).

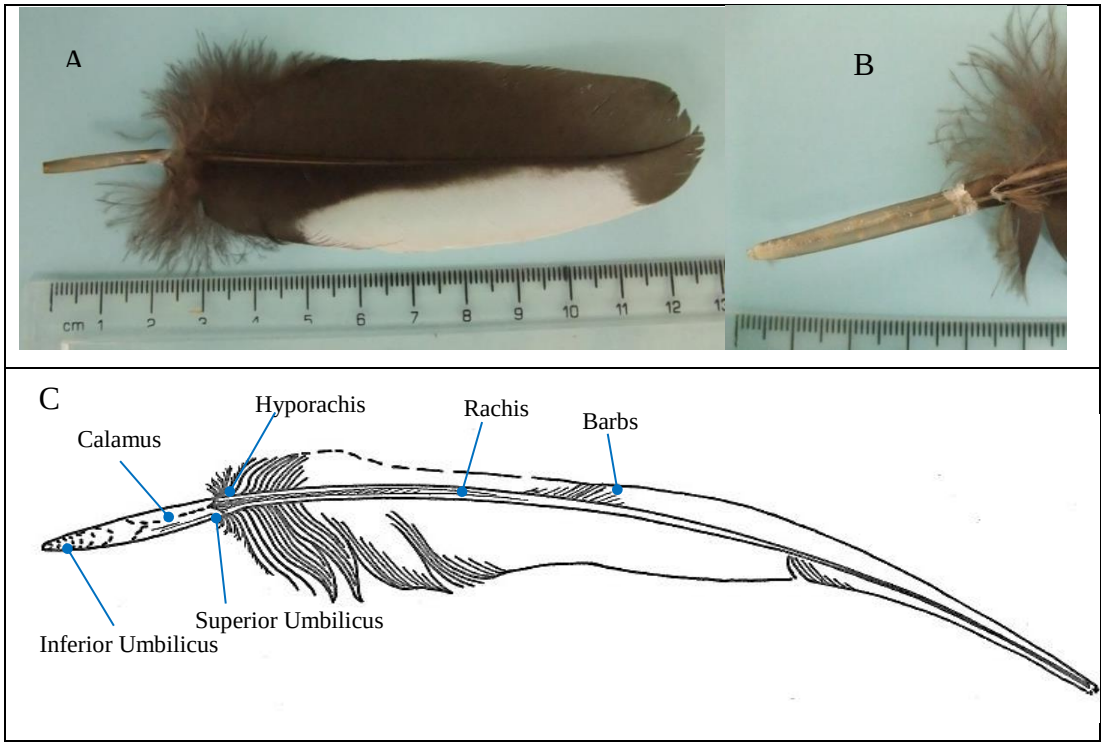


Figure 2.6: Trinidad Piping Guan Feathers; A: complete contour feather, B: close up of quill & C: illustration of the major features of the feather, diagram of a primary flight feather (photographs by A. Grass)

Following tissue lysis, DNA was isolated using silica based membrane spin-columns; centrifugation of the sample in the spin-column allows DNA to selectively bind to the DNeasy membrane and contaminants or waste to be removed by the washing steps under optimised conditions.

200 µl of AL Buffer and 200 µl of Ethanol were added to the lysed sample and mixed by vortexing. The resulting solution was removed and added to a spin-column, which was then centrifuged at ≥6,000 x *g* for 1min and the flow through discarded. 500 µl of

Buffer AW1 was added to the spin column, the spin column was centrifuged at ≥6,000 x *g* for 1min and the flow through discarded. 500 µl of Buffer AW2 was then added to the spin column, the spin column was centrifuged at 20,000 x *g* for 3min and the flow through discarded. Finally the spin column was placed in a new 1.5ml micro-centrifuge tube and 200 µl of Buffer AE added for the elution of the DNA product from the spin-column membrane. After incubation for 1min at room temperature the spin column was centrifuged for 1min at ≥6,000 x *g*. A second elution sample was collected by repeating the addition of Buffer AE followed by centrifugation, to maximise the yield of DNA from each sample. The resulting DNA extractions were stored on ice for immediate use, or frozen at -20 °C for use in the future.

To eliminate contaminants and ensure sample integrity all equipment and disposable plastics used in the extraction process were sterilised by autoclaving at 121 ˚C and subsequent Ultra-Violet Light exposure (254 nm shortwave UV for 1hour). In addition to prevent cross-contamination between samples, each sample was extracted using separate sterile equipment and plastics. Negative extraction controls (no sample) were performed during all sample extractions to ensure that no contamination resulted from the procedures employed.

2.3.2 DNA Isolation from Egg /Embryo Tissue

Qiagen DNeasy Blood and Tissue Kit, Modified Tissue Protocol

Lysis was carried out by incubation of the sample in a Proteinase K solution. Approximately 5mm^3 of tissue/ egg contents were removed from the sample and placed in a 1.5 ml micro-centrifuge tube, 10 µl of DTT, 180 µl of ATL Buffer and 20 µl of Proteinase K was added, and the solution mixed by vortexing. The samples were incubated at 56 °C overnight or until lysis was complete (determined visually by the absence of solid particles in the solution).

Following incubation DNA was isolated from the samples using the silica based DNeasy membrane spin-columns (methodology as previously described for feather samples, in section 2.3.1). The resulting DNA extractions were stored on ice for immediate use, or frozen at -20 °C for use in the future.

2.3.3 DNA Isolation from Museum Samples

Museum specimens are valuable sources of genetic material, however sampling is both invasive and destructive. As such, efforts were taken to ensure minimal damage to the specimen and sampling site on the specimen was selected to ensure minimal interference in taxonomic characterisation.

Tissue samples obtained from the British Natural History Museum specimens were collected from the proximal phalanx of the footpad, or close to this region depending on preservation limitations of foot position. Skin from the footpad is considered to be minimally invasive and of little impact to taxonomic considerations, whilst providing a good source of old or ancient DNA for mitochondrial or nuclear analysis (Mundy, Unitt, & Woodruff, 1997). Feet of avian museum specimens are typically not given any additional treatment during preservation. In comparison, traditional preservation treatments on the main body of the skin may include the use of arsenic compounds, for example. As such there is reduced DNA degradation in the footpad compared to other tissue sources (Mundy et al., 1997).

The Museu de Zoologia da Universidade de São Paulo provided flank feathers cut close to the skin to ensure minimum disturbance to the taxonomic features of the specimens. Historical feather samples have been shown to be useful sources of genetic materials, however demonstrate more significant levels of degradation and present issues related to the use of preservative compounds and treatments (Leeton & Christidis, 1993; Mundy et al., 1997).

Ancient DNA and Old DNA

For the purposes of the extraction of DNA from old or ancient samples (>50 years) additional measures were implemented to avoid contamination with modern DNA including; the use of separate facilities and equipment, sterilisation of all equipment and consumables by autocalving and UV light (254 nm for three hours), and stringent cleaning procedures to remove all traces of modern DNA from the lab and working areas. Fresh reagents were used in each case to prevent cross-contamination with previously extracted Piping Guan tissues or DNA samples. No extractions were

performed where Piping Guan DNA was stored or had been previously in use. Negative control extractions (no sample/no DNA) were also performed to identify any potential contamination of the reagents with foreign DNA.

Qiagen DNeasy Blood and Tissue Kit, Modified Tissue Protocol for Footpad Skin (>50 yrs old)

As previously described in sections 2.3.1 and 2.3.2 lysis was carried out by incubation of the tissue sample in a Proteinase K solution. Approximately 2 mm^3 of footpad skin section was cut from the proximal phalanx of the digit (footpad) of the preserved specimen, placed in a 1.5 ml micro-centrifuge tube and dry stored until use. 10 µl of DTT, 360 µl of ATL Buffer and 40 µl of Proteinase K, was added to the tissue sample and the solution mixed by vortexing. The samples were then incubated at 56 °C for between 24 and 72 hours, until lysis was complete (determined visually by the absence of solid tissue material within the solution).

DNA was isolated from the samples using silica based DNeasy membrane spin-columns, methodology as previously described for feathers in section 2.3.1. The resulting DNA extractions were stored on ice for immediate use, or frozen at -20 °C for long term storage.

Qiagen DNeasy Blood and Tissue Kit, Modified Tissue Protocol for Feathers (>50yrs old)

As previously described in section 2.3.1 cell lysis was carried out by incubation of the sample in a Proteinase K solution. Approximately 100mm of feather section was cut from the quill of the specimen (including any barbs), and 10 µl of DTT, 180 µl of ATL Buffer and 20 µl of Proteinase K, was added to the sample. The solution mixed by vortexing. The samples were then incubated at 56 °C for between 24 and 72 hours, until lysis was complete (determined visually by the absence of feather/barb particles within the solution).

DNA was isolated from the lysed samples using silica based DNeasy membrane spin-columns, methodology as previously described for feather samples. The resulting

DNA extractions were stored on ice for immediate use, or frozen at -20 °C for long term storage.

2.4 DNA Quantification

2.4.1 Quantification of dsDNA using Qubit 3.0 Fluorometry

DNA extractions were quantified using the Qubit 3.0 Fluorometer (Invitrogen, Life Technologies) with the Qubit® dsDNA HS Assay Kit (Invitrogen). DNA quantifications were achieved using the manufacturers protocols as follows; Qubit® dsDNA HS reagent was diluted 1:200 in Qubit® dsDNA HS buffer to a final volume of 200µl including 1-20 µl of DNA extract, the assay was performed at room temperature (22-24 °C). DNA extractions were prepared with 10 µl of DNA extract in addition to 190 µl Qubit® dsDNA HS buffer. Standards were prepared with 10 µl of Stock in addition to 190 µl Qubit® dsDNA HS buffer. Samples and standards were mixed by vortexing for 2-3 seconds and incubating for 2 minutes at room temperature. The Qubit 3.0 Fluorometer was calibrated with the manufacturers standards and samples were analysed within 1hour of preparation (Assay reagents being stable for a maximum of 3 hours at room temperature). Calibration was performed prior to each assay. All samples were analysed in triplicate, the sample was removed from the Qubit 3.0 Fluorometer for 30 seconds between readings to avoid heating of the sample, which may reduce accuracy. DNA concentration was calculated by the Qubit 3.0 Flurometer (ng/ml).

2.4.2 Quantification of dsDNA using gel visualisation

DNA extractions and amplified DNA products were quantified through visual analysis of the dsDNA via gel electrophoresis and image analysis in ImageLab software (Biorad). Known dsDNA concentrations in the DNA ladder (PCR Sizer 100 bp Ladder by Geneflow) were used to calculate a standard curve (see Figure 2.7). Samples were quantified from the standard curve using linear regression to determine absolute quantity.

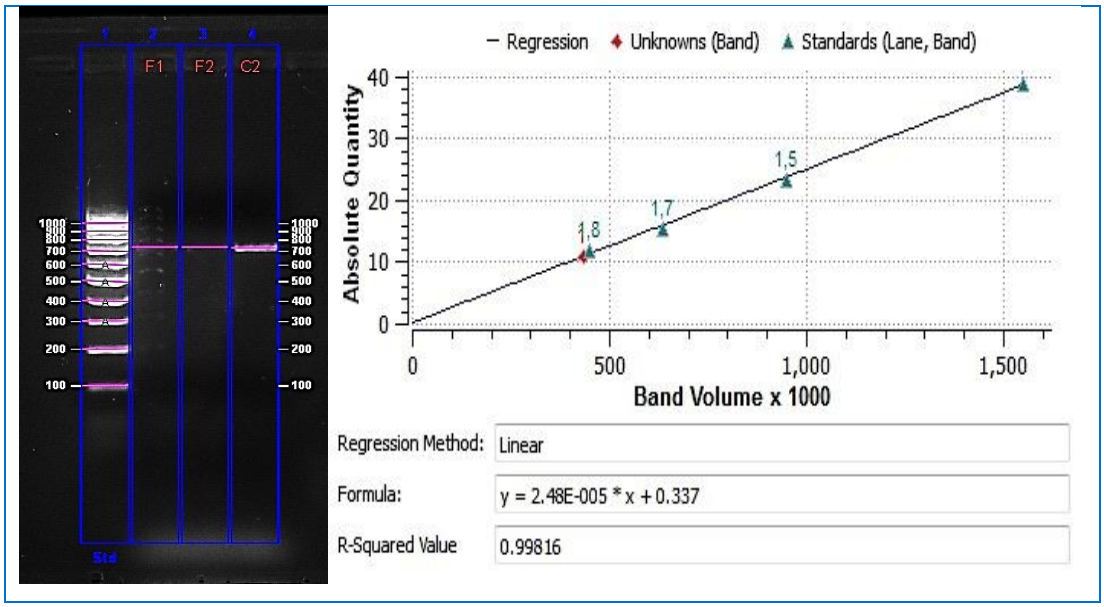


Figure 2.7: DNA quantification by gel electrophoresis, using the Image Lab software (Biorad), from known DNA standards (Lane 1) in Norgen PCR Sizer 100bp DNA Ladder (Geneflow).

2.5 Polymerase Chain Reaction

Automated thermal cycling was performed using either the Techne Flexigene Thermal Cycler or the Techne 5100 Gradient Thermal Cycler, with cycles designed specifically to suit the primer sets used and the melting temperatures of each primer. Cycling parameters were optimised using the Reference Trinidad Piping Guan sample (R1) and *Gallus gallus* DNA to establish the most productive amplification parameters for each primer set.

2.5.1 Primers

Primer sets were designed using sequence data from the Reference Trinidad Piping Guan sample (R1) and modification of primers used in previously published research. Robinson (2011) developed primers for the amplification of regions in the Cytochrome B, NADH dehydrogenase 2 and Control Regions of the mitochondrial genome for the Trinidad Piping Guan, which were adopted or modified for this study. In addition, universal avian primers were used as a basis for the development of additional primer

sets for use in this study (Sorenson, Ast, Dimcheff, Yuri, & Mindell, 1999 and Sorenson 2003).

Primer Design

Primers are traditionally designed by taking advantage of stable regions of the genome, regions which are conserved among species or are active sites within a gene (stable due to their role in the protein coding function of the gene). Conventional primer design platforms, including Primer3 and Geneious R8 Primer Design, were used to develop new primers based on information from the mitochondrial DNA sequence of the Trinidad Piping Guan and other Galliform species, including the reference *Gallus gallus* (sequences accessed through the NCBI database https://www.ncbi.nlm.nih.gov/).

The following general procedures for the design of oligonucleotide primers were used to ensure optimum results. Primers were designed to be 17-28 bp length, with a melting temperature (T_m) range of between 55 °C and 80 °C, according to the nucleotide content. Nucleotide composition was assessed to optimise stable annealing; GC-content of 40-60 % promotes stronger annealing, and use of a G, C or GC/CG nucleotide(s) at the 3'end increases priming efficiency due to the stronger triple-hydrogen bond structure in the Guanidine-Cytosine pairing (Dieffenbach, Lowe, & Dveksler, 1993).

During design, primers were analysed for self-priming and hairpin structures; avoidance of dinucleotide repeats, multiple tandem nucleotides of the same base and in particular avoidance of three or more G or C nucleotides at the 3'end of the primer ensure mis-priming is not promoted (Dieffenbach et al., 1993).

Cross-homology with other areas of the genome and other species (human) was, where possible, avoided to maintain oligo efficiency and therefore specificity; primers are not 100% efficient therefore annealing can be successful at lower levels of homology with template or contaminant DNA. Short primers (<20 n) are subject to higher degrees

of cross-homology due to length effects therefore the use of short primers was avoided. Where possible the use of degenerate bases within primers was avoided to ensure accurate species specific priming with the target template DNA. Homology of the primers with the target DNA is also important in the avoidance of spurious amplifications and Nuclear Mitochondrial Pseudogenes (NUMT) amplifications (Sorenson & Quinn, 1998).

Performance of newly designed primers was assessed by amplification of DNA products of appropriate size in the Trinidad Piping Guan reference samples, along with positive and negative DNA controls to identify potentially spurious amplifications and/or contamination.

Universal Avian Primers

Universal primers were used to amplify regions of both the mitochondrial and nuclear genomes, present in the DNA extracts. Universal primers are usually designed across class groups and use conserved regions of the genome as priming sites to maximise their utility in many taxa (Kocher et al., 1989; Sorenson et al., 1999; Sorenson, 2003). The use of universal primers is also beneficial where there is a lower than expected primer-template homology (Sorenson et al., 1999); primers have been shown to be effective with up to a 20 % internal miss-match, providing there is complete homology at the 3'end.

Differences between species in sequences were accommodated in the design of new primers through the use of degenerate nucleotides at specific nucleotide positions, allowing the design of primers to anneal at priming sites with inter-specific variation. Standard nucleotide coding systems for mixed base definitions were used in primer design to account for variations at a given nucleotide position (see Table 2.1 for details).

Table 2.1: Standard Nucleotide Coding System

Symbol	Nucleotide base equivalents
R	A,G
Y	C,T
M	A,C
K	G,T
S	C,G
W	A,T
H	A,C,T
B	C,G,T
V	A,C,G
D	A,G,T
N	A,C,G,T

The redesign of primers from their original 'Universal' sequence to a Species-specific sequence was carried out by using overlapping regions of confirmed Trinidad Piping Guan sequence data. The modified primers subsequently improve template-primer annealing by increasing the homology of the primer to the target sequence.

The use of 'Universal' primers has been reported to increase the occurrence of NUMT amplification (Sorenson et al., 1999; Sorenson & Quinn, 1998), due to the use of degenerate sites in the oligonucleotide sequence. However the use of such degenerate sites is necessary for such primers to be applicable in many cases where there is not the necessary conserved sequence between all of the species within the group (in this case avians), to design a primer universally specific for a target region. As such the identification of potential NUMTs through visualisation of the PCR products is necessary to eliminate the use of erroneous data. The high ratio of mitochondrial copies to nuclear copies in PCR products is useful in the visual identification of such NUMTs (Sorenson & Quinn, 1998). (Further information on NUMTs is given in section 2.6.2 Quality Control).

Primers were supplied by Eurofins MWG Operon, either to the manufacturers 'Standard Oligonucleotide' specifications or 'Sequencing Oligonucleotide'

specifications. A complete list of primers used in this research can be found in Appendix 2.

2.5.2 PCR Reaction Components

The PCR reaction outcome or success is influenced strongly by cycling parameters and reaction components (Mg++, H+, dNTP, Primer and Template concentrations), each reaction having its own optimum conditions for specificity of product and yield (Hecker & Roux, 1996). This study used a standard PCR technique for the amplification of DNA products from the extracts, however where amplifications failed or were of insufficient quality a more advanced PCR approach was adopted, including the use of high fidelity Taq in PCR reactions, primerless PCR and Touchdown PCR techniques.

Standard PCR Reactions

Standard PCR was performed using Illustra™ puRe Taq Ready-To-Go PCR Beads (GE Healthcare), preparation of the PCR reaction mixture was carried out in a UV sterilised safety cabinet (254 nm shortwave UV for 120 mins). A solution of 19 µl of sterile distilled nucleotide free water and 1µl of the Primer mix (containing both forward and reverse primers at a concentration of 25 pmol/µl) was added to the PCR tube containing the puRe Taq bead. Amplifications using the puRe Taq reaction bead contain BSA, 2.5 u of puRe Taq DNA polymerase, 200 µM of each dNTP, 10 mM Tris-HCl (pH 9.0), 50 mM KCl and 1.5 mM $MgCl_2$.The PCR tubes were moved to the separate DNA cabinet where 5µl of the extracted DNA was added to the reaction mixture (final reaction volume 25 µl).

2.5.3 PCR Cycle design

Standard PCR Cycles

PCR follows a standard three phase cycle of denaturation, annealing and extension, see Figure 2.8. Initial denaturation of the DNA template is performed at 94°C to

facilitate the deconstruction of the double-stranded molecule into a single stranded template. The annealing phase is temperature sensitive according to the oligonucleotide sequence melting temperature (T_m), to achieve maximum binding to the target template DNA (see below). The final phase is the extension of the DNA template at an optimum temperature of 72 °C. On completion the cycle is held at 4 °C, samples can then be used immediately or stored at -20 °C for future use. See Figure 2.8 for the standard PCR cycle.

Fidelity of the PCR relies on an accurate melting temperature (T_m) to facilitate appropriate hybridisation and amplification of the target product. The T_m is calculated from the nucleotide constituents of the primers using the following formula;

$$T_m = 4\,(G + C) + 2\,(A + T) \qquad \text{(Dieffenbach et al., 1993)}$$

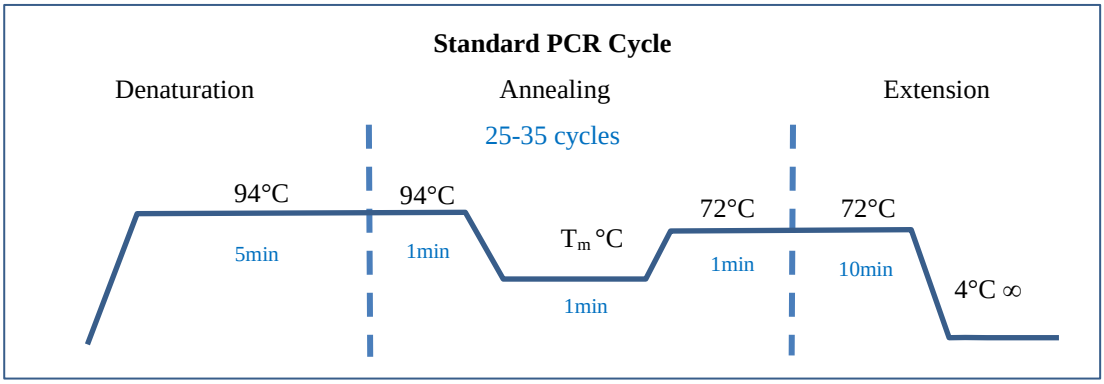


Figure 2.8: Standard Polymerase chain reaction (PCR) cycling parameters.

Touchdown PCR

Touchdown PCR is primarily used to improve the fidelity of priming of one or both of the oligonucleotides to the template DNA to eliminate or reduce the amplification of spurious products during the PCR process (Don, Cox, Wainwright, Baker, & Mattick, 1991; Korbie & Mattick, 2008). Amplification of spurious products is a common problem where oligonucleotides bind to multiple regions of the genome producing alternative products which may subsequently dominate the reaction. The co-amplification of products which are smaller in size than the desired product may result

in stochastic advantage due to the nature of replication in the PCR cycle, where smaller amplification products have the advantage over larger ones, through a more rapid reconstruction during extension phase of the replication process (Don et al., 1991).

The Touchdown PCR process takes advantage of the exponential growth which occurs during PCR by starting the annealing temperature cycling at a higher temperature than required in a normal PCR cycle. The annealing temperature is set to approximately 10ºC higher than the T_m of the oligonucleotides and is reduced by 1 ºC per cycle for the first ten cycles, at which point the optimum T_a is achieved and normal cycling resumes (see Figure 2.9). The Touchdown technique favours the highest fidelity primer-template interaction and therefore enhances the formation the desired product, even when reaction conditions are sub-optimal (Hecker & Roux, 1996).

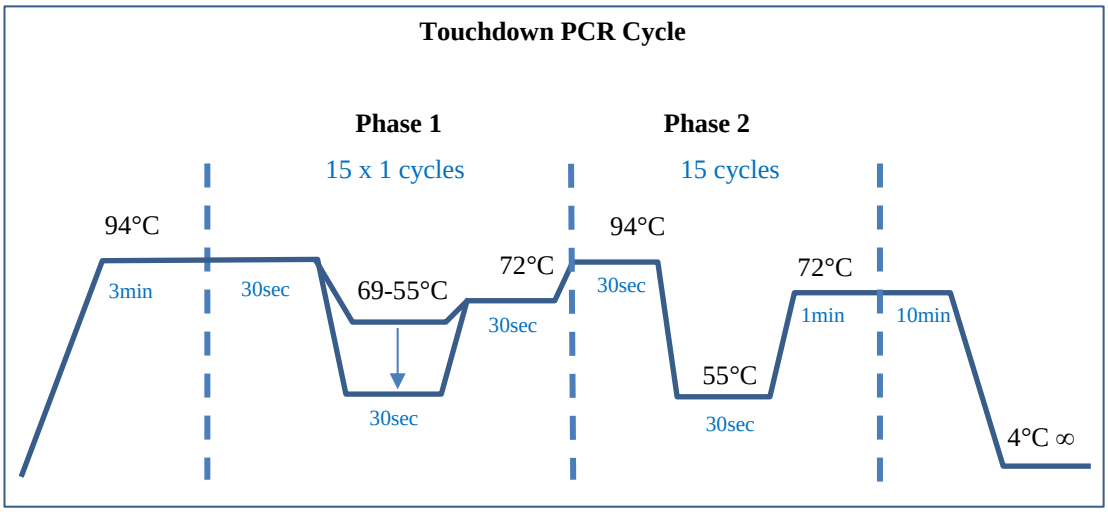


Figure 2.9: Touchdown Polymerase Chain Reaction cycle parameters, where T_m = 55 ºC

Don et al. (1991) recorded an advantage of between two-fold and four-fold in specific priming, per ºC difference in the annealing temperature of the touchdown temperature and the 'correct' annealing temperature, in the use of Touchdown PCR. This selective amplification advantage of Touchdown PCR allows the correct or desired amplification product to dominate in the reaction, eliminating or reducing secondary spurious amplifications.

In addition, the use of Touchdown PCR reduces the need for lengthy optimisation experiments to reduce spurious priming during amplification; including the alteration of Mg++ and adjustment of the annealing temperature to find the optimum reaction conditions. Where there are limited resources available in terms of sample availability, such as the use of samples from critically endangered species, the use of techniques such as touchdown PCR is advantageous in that it does not require large volumes of DNA to optimise each individual reaction. This therefore enables the amplification of multiple PCR products with a minimal outlay of DNA sample.

2.5.4 Advanced PCR Techniques

Primerless PCR

Primerless PCR is a technique in which template DNA is denatured and replicated under standard PCR conditions with the exclusion of any targeting oligonucleotides. The technique of Primerless PCR (pPCR) initially outlined in research by Stemmer (1994), uses to Polymerase Chain Reaction to increase the volume of DNA within a sample by accurately replicating the template sequence. The pPCR process denatures the template DNA and replicates low volumes of DNA present through extension of the product without specific primers being present. Stemmer (1994) demonstrated a high degree of template accuracy using this technique in the re-assembly of DNA fragments with the pPCR process (however the technique is not widely used).

Primerless PCR is ideally suited to the amplification of ancient and historical DNA samples due to the nature of degradation, where by DNA strands are broken in a random fashion throughout the genome. In such cases pPCR is able to partially restore and rebuild products when the DNA strand is partially complete or complete in the form of a single strand and will therefore improve the yield of general template DNA within a sample prior to the targeted PCR amplification using specific primers. Weber (2004) and Weber, Stewart, Garza, & Lehman (2000) utilised the primerless PCR technique in the amplification of aDNA from museum specimens of Guadalupe Fur Seals and Northern Elephant Seals, respectively, with success in the recovery of DNA

products where standard, direct PCR techniques had failed to amplify DNA from historical seal specimens.

Primerless PCR reactions utilised standard PCR reagents, as previously described for standard PCR (section 2.4.2) with the exclusion of primers, Illustra™ puRe Taq Ready-To-Go PCR Beads (GE Healthcare) in a total reaction volume of 25 µl, including 10 µl of template DNA and Nuclease free water.

Primerless PCR reactions are amplified in the PCR cycle as outlined in Figure 2.10, the products of which are then used in standard PCR techniques to amplify specific Target DNA products.

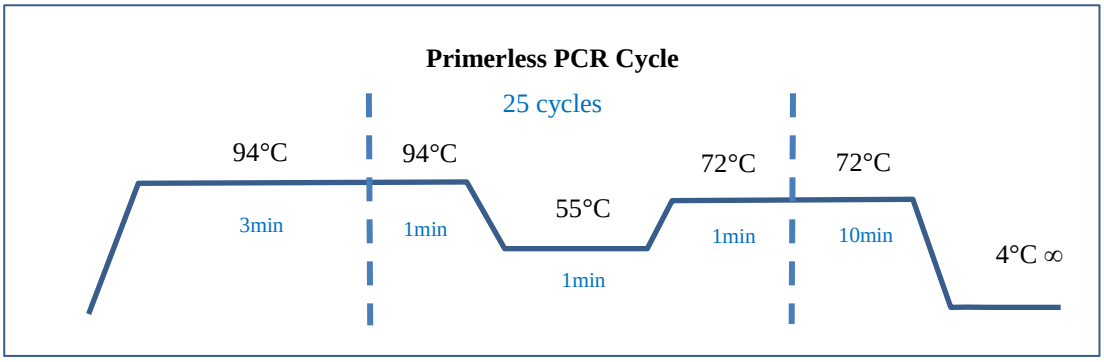


Figure 2.10: Primerless Polymerase Chain Reaction cycle parameters.

Independent analyses of multiple Primerless PCR amplification products were used to prevent the accidental use of erroneous amplicons, as a direct control measure (Weber et al., 2000).

High Fidelity PCR

High fidelity DNA polymerase reaction mixtures were also adopted in an attempt to improve fidelity and therefore improve the amplification yield of aDNA PCR. Standard *Thermus aquaticus* (Taq) DNA polymerase demonstrates a high degree of fidelity, with base substitution errors in the reported as one in 9,000 and single base frameshift errors in the region of one in 41,000 (Eckert & Kunkel, 1990). However,

fidelity and action of the Taq polymerase is influenced by a number of factors including the Mg^{++} ion concentration, pH and dNTP concentration. Eckert & Kunkel (1990) reported Taq fidelity was at its highest when $MgCl_2$ concentration is equimolar to total concentration of dNTPs, and a low reaction pH of around pH5.8. Specific 'High Fidelity' DNA polymerases have been demonstrated to improve the accuracy of nucleotide homology in the annealing and extension phases of the PCR cycle. Q5® High Fidelity 2X Master Mix (New England Biolabs Inc.) supports a PCR reaction with error rates >100-fold lower than that of Taq DNA Polymerase (New England Biolabs Inc., 2016).

A standard Q5® High Fidelity 2X Master Mix protocol was followed; 12.5 µl of Q5® High Fidelity 2X Master Mix, 10 µM Forward primer, 10µM Reverse primer, 5µl of template DNA and Nuclease Free Water in a 25 µl reaction total volume. At a 1X final concentration Q5® High Fidelity Master Mix reactions contain 2.0 mM $MgCl_2$, 0.5 µM forward and reverse primers, and 200 µM of each deoxynucleotide.

2.6 Gel Electrophoresis

The amplification success of all targeted DNA products were determined through visual analysis of the target DNA via gel electrophoresis and quantification of the PCR product.

2.6.1 Agarose Gel Electrophoresis

Amplification of the target DNA from the extracted samples was confirmed by Agarose Gel Electrophoresis, using 2 % w/v Agarose gels with GelRed™ Nucleic acid Stain (Cambridge Bioscience) in a 1x Tris-Acetate-EDTA (TAE) Running Buffer pH 8.0. Agarose (2 g) was dissolved by heating in 98.0 ml of de-ionised water. The resulting solution was cooled to approximately 50 °C and 2.0 ml of 50x TAE Buffer (see standard solutions section 2.11) and 10 µl of GelRed™ was added and gently mixed. The Agarose solution was poured into the gel tank mould and the sample comb inserted. The solution was allowed to set for approximately 1hour and electrophoresis carried out in a 1x TAE Running Buffer. PCR amplifications along with positive and

negative controls were loaded into the gel wells and a 100bp ladder (Norgen PCRSizer 100 bp, Geneflow) was used as a marker standard. In each case, the gel was run for approximately 1 hour 15 mins at 100V (constant voltage).

2.6.2 Visual Analysis

The samples were visualised by UV-trans-illumination on a Bio-Rad Gel Doc™ EZ Imager system and analysed using Image Lab software (Bio Rad). Analysis of gel images using Image Lab software allowed determination of molecular weight through linear regression of the standard curve produced from the molecular weight DNA ladder (Norgen PCR Sizer 100bp DNA ladder by Geneflow). Image Lab software was also used for the absolute DNA quantification of amplified fragments, using standard curve analysis of the concentration of markers within the DNA ladder.

2.7 Quality Control

All DNA extractions and PCR based amplification reactions were conducted under strict protocols to ensure the validity and authenticity of the DNA sequence produced.

2.7.1 Authentication of DNA product

The amplification of the appropriate target DNA product was critical in all stages of the research, care was taken to ensure the authenticity of all DNA products both mitochondrial and nuclear.

In the analysis of large regions of the genomes including complete genes and the whole mitochondrial genome, it was important that all DNA sequence should overlap by a minimum of 50-100 bp, to ensure a contiguous sequence link between short fragment amplifications. In all instances DNA sequences were obtained for both the Heavy strand and Light strand, which allowed the direct comparison of both strands to ensure the fidelity of the nucleotide asignment in the amplified sequence. Where possible, replications of the PCR amplifications were made to ensure the validity of the DNA sequence products.

All DNA sequences were subsequently analysed in the Basic Local Alignment Search Tool (BLAST) software by the National Centre for Biotechnology Information (https://blast.ncbi.nlm.nih.gov/Blast.cgi). The BLAST software searches for the best sequence homology of the query DNA within the combined DNA sequence databases of GenBank, the EMBL and the DDBJ[1]. Analysis in BLAST finds the closest matching DNA sequences, therefore giving a reliable indicator of species or genus and the gene region of the query sequence.

In addition to the species identification and gene region identification of the BLAST tool, more detailed analysis of the DNA sequences and the sequence features provides further verification of the authenticity of the DNA sequence. Conserved gene features such as stop codons, domains, conserved sequence blocks, and wider genomic features such as gene organisation are indicative of the appropriate amplification of the target DNA products.

2.7.2 NUMTs

Nuclear mitochondrial pseudogenes (NUMTs) are a concern in the amplification of mitochondrial DNA sequences. NUMTs have been identified in a number of avian species including several species of Anseriformes (Sorenson & Quinn, 1998), and represent transpositions of several genes including the Control Region, Cytochrome b and Cytochrome oxidase 1 genes. NUMTs are subject to nuclear evolutionary rates which ensure measurable differences between mitochondrial and nuclear sequences (Bensasson, Zhang, Hartl, & Hewitt, 2001), with varying degrees of homology. Although nuclear copies may be co-amplified with the targeted mitochondrial region, the high copy number of mitochondrial genomes facilitates the identification of mitochondrial products (Bensasson et al., 2001; Sorenson & Quinn, 1998).

NUMTs may be identified in a number of ways;

---

[1] GenBank - National Centre for Biotechnology Information (NCBI), USA www.ncbi.nlm.nih.gov

EMBL - European Molecular Biology Laboratory, Cambridge UK www.ebi.ac.uk

DDBJ - DNA Data Bank of Japan www.ddbj.nig.ac.jp

- The presence of multiple bands of differing sizes in the visualisation of PCR products.
- Sequence ambiguities as a result of heterozygosities caused by sequencing of background nuclear copies.
- Unexpected indels, frameshifts and stop codons the result of mutation in the nuclear copy without functional constraint (mitochondrial copies being functionally constrained in their evolution).
- Nucleotide sequences radically different from the expected sequences, when compared to mitochondrial references.
- Unexpected phylogenetic topology as a result of the use of nuclear copy sequence data.

In addition a number of techniques favour the avoidance of the amplification of NUMTs, particularly the use of mitochondrial DNA rich tissues such as feathers (a poor nuclear DNA source), the amplification of long fragments in PCR and purification of mitochondrial DNA (Sorenson & Quinn, 1998). The use of sequence specific primers, without degenerate nucleotide positions additionally may reduce the likelihood of amplifying NUMTs.

2.7.3 Controls

At all stages, extraction, PCR and purifications, controls were employed to identify the presence of any contaminant problems. Negative control samples (no DNA) were used to highlight the presence of contamination in the reagents or the infiltration of human DNA. Positive controls (known avian species DNA) were used to identify the appropriate amplification of PCR products.

In addition following species analysis of amplicons in BLAST (http://www.ncbi.nlm.nih.gov/), all sequences were compared to Human DNA reference sequences to identify contamination of the samples with human DNA. At no point in the research was human DNA identified in the amplification of DNA products from any of the samples.

2.7.4 Ancient DNA

The use of ancient and old DNA samples (aDNA) requires the addition of more stringent laboratory controls and procedures, (as previously outlined in Section 2.3.3.) due to the increased susceptibility of these samples to the infiltration of contaminant modern DNA (Graham, Ferrier, Huettman, Moritz, & Peterson, 2004). Positive and negative controls were used throughout all procedures. All reagents and consumables were appropriately cleaned through autoclaving, UV light sterilisation (254nm shortwave UV for 3hours) and bleaching to ensure no contaminant DNA was present. Strict cleaning procedures were adhered to at all times. Sample preparations were made in cabinets where no previous DNA samples had been used, that could present contamination issues.

The infiltration of contaminant human DNA was particularly of concern, as all samples have previously been handled by humans the presence of Touch DNA was a possibility (Williamson, 2012). To avoid amplification of human DNA, procedures such as species/genus specific primers or avian specific primers were used to prevent the accidental amplification of undesirable human DNA contaminants.

2.8 Purification of Target DNA

Purification of the PCR product removes excess components such as primers, nucleotides, DNA polymerase and salts from the sample for downstream services such as DNA sequencing.

Purification using GenElute™ PCR Clean Up Kit

Purification of the PCR products was carried out using GenElute™ PCR Clean Up Kit; using silica binding columns and centrifugation. The GenElute system is optimised for the purification of double or single stranded amplifications between 100 bp and 10 kb (Sigma-Aldrich, 2016).

A GenElute Miniprep Binding Column was inserted into a collection tube. 0.5 ml of the Column Preparation Solution was added to the binding column, which was then centrifuged at 12,000 x *g* for 30 seconds and the flow through discarded. 125 µl of Binding Solution was added to 25 µl of the amplified PCR product in a 1.5 ml microcentrifuge tube, and mixed by vortexing. The binding solution PCR product mixture was transferred to the Binding Column and centrifuged at maximum speed (~16,000 x *g*) for one minute. The flow through was again discarded. 0.5 ml of diluted Wash Solution was added to the Binding Column and centrifuged at maximum speed for one minute. The flow through was discarded. Excess ethanol was removed by further centrifugation for two minutes at maximum speed. The binding column was transferred to a 1.5 ml microcentrifuge tube. Elution Solution (50 µl) was added to the binding column and incubated for one minute at room temperature. The purified DNA was eluted by centrifugation at maximum speed for one minute and the resulting purified DNA stored on ice for immediate use, or frozen at -20 °C for long term storage.

Purification using Qiagen MinElute Gel Extraction Kit (standard protocol)
Purification of the PCR products was carried out using MinElute Gel Extraction Kit; isolation of the amplicon via agarose gel electrophoresis and purification using silica binding columns and centrifugation. The Qiagen MinElute system is designed for the purification of amplifications between 70 bp and 4 kb (Qiagen, 2016).

DNA fragments were separated on a 1 % agarose gel via electrophoresis (see previous method section 2.5.1), and visualised using a UV light source. The desired fragment was excised from the agarose gel with a sterile scalpel, the gel slice was weighed in a sterile 1.5ml microcentrifuge tube (weight in mg = 1 gel unit). Three gel volumes of Buffer QG were added and the solution was incubated at 50 ºC for 10 mins, or until the gel completely dissolved. One gel volume of Propan-2-ol (Isopropanol) was added and the solution was mixed by vortexing. A MinElute Binding Column was inserted into a collection tube and 0.8 ml of the solution was added to the binding column, which was then centrifuged at 20,000 x *g* for 1 minute and the flow through discarded. The remaining solution was added to the column, centrifuged and the flow through discarded. 500µl of Buffer QG was added to the column and centrifuged at 20,000 x *g*

for one minute, the flow through was again discarded. Buffer PE (750 µl) was added to the column, and the column allowed to stand for 2-5 minutes, it was then centrifuged at 13,000 x rpm for one minute, the flow through was discarded. Excess buffer was removed by a further centrifugation step for one minute at 20,000 x *g*. The binding column was transferred to a sterile 1.5 ml microcentrifuge tube and 11µl of the Elution buffer (Buffer EB) added to the binding column before incubation for one minute at room temperature (~24 °C). The purified DNA was eluted by centrifugation at 20,000 x *g* for one minute and stored on ice for immediate use, or frozen at -20°C.

2.9 Sequencing

Sequencing of the amplified mtDNA samples was carried out using the customised Sequencing Services of MWG Eurofins (MWG Biotech). Samples were prepared as per MWG Eurofins requirements (www.eurofinsdna.com) for sequencing services. PCR products were diluted with nucleotide-free sterile water or concentrated by ethanol precipitation to the appropriate concentration, according to the product size (bp) and sequencing service specifications.

2.10 Health & Safety Considerations

All experimental work was performed in a sterile, controlled environment, using sterilised equipment and personal protective equipment (double gloves, lab coats, disposable aprons and masks). PCR reactions were prepared in a laminar flow cabinet with UV sterilisation to avoid contamination of the samples. DNA samples were added to PCR reaction mixes in a separate PCR workstation under sterile conditions. All procedural COSHH Risk Assessments used in the research were reviewed and updated as necessary prior to commencement of the study. Disposal of all chemical reagents was performed according to individual chemical Material Safety Data Sheets. All disposable plastics, DNA and DNA products were disposed of by autoclaving and appropriate waste disposal according to local regulatory requirements.

2.11 Statistical Analysis

2.11.1 Methods for constructing Phylogenetic trees

Phylogenetic inference through tree-searching methods, such as Maximum Likelihood modelling and Bayesian Inference modelling, utilise a method which estimates many trees then computes the 'best' tree from the selection. Tree-searching methods in their most basic form typically employ a branch building strategy adding each branch/taxa individually and evaluating the structure for minimum evolution. Heuristic algorithm approaches build an initial tree, evaluate it then make re-arrangements to improve the tree structure and score, through probabilistic analysis in relation to a random tree structure.

Character based methods of phylogenetic analysis (Maximum Likelihood and Bayesian Inference) use the alignment information directly, by comparing individual character (nucleotide) changes in each column of the alignment, and subsequently constructing trees based on the minimum number of events (Nei & Kumar, 2000). Most character methods utilise a probabilistic approach to the likelihood of observing the data under a given substitution model. In theory, the accuracy of the tree would be based on the probability of recovering the 'true tree', however in reality this is an impossibility, as we cannot know the 'true tree' it is impossible to measure how well the analysis has performed. All methods will infer some level of bias and error associated with the substitution model and the algorithm employed to analyse the data. Therefore, the use of multiple methods to confirm tree topology is employed to allow errors to be observed more easily and where possible to confirm topology to the best of ones ability with the available data.

Maximum Likelihood models construct the 'best' tree from the topologies which maximises the likelihood or probability of observing the data under a specific model of substitution, the tree with the highest likelihood is selected from the topologies (Vandamme, 2011). Maximum Likelihood is a well accepted, robust method for assessing genetic relationships and constructing phylogenetic trees, used widely in evolutionary and conservation modelling.

Maximum Likelihood methodologies are typically assessed using Bootstrap analysis methods to estimate reliability and test the topology. The Bootstrap method randomly re-draws the tree multiple times and presents the reliability of each node within the tree as a percentage according to the number of times the node occurs in a given position. Bootstrap support is considered an estimate of statistical support for the node, with values over 95 % (or 99 % depending on confidence required) being indicative of confidence in the topology (Nei & Kumar, 2000).

Bayesian Inference models constructs the 'best' tree based on a consensus of many trees using the greatest likelihood of the given the data, where the frequency of the branch or clade in the set of trees is linked to the probability of that branch or clade existing (Drummond & Bouckaert, 2015; Vandamme, 2011). Bayesian Inference models utilise a generational system (heuristic approach), where each generation involves the selection of a tree, calculation of the trees likelihood, modification of the tree, re-calculation of the likelihood and finally the acceptance of the tree with the greatest likelihood. Generations are iterated repeatedly until the likelihood values do not significantly change, thereby accepting the most likely tree under the given parameters for the data. At each sampling a Burn-in fraction of samples was discarded, removing the first proportion of the samples so that initial values for posterior inferences which are less accurate do not skew or affect the overall product.

Bayesian Inference methods use a probability based estimate of reliability and topology, presenting the probability of each node occurring at a given position within the topology, given as a mean of the distribution of probabilities at that node (posterior mean) (Drummond & Bouckaert, 2015; Ronquist, van der Mark, & Huelsenbeck, 2011).

The consistency between genome and gene nomenclature is of great importance to the analysis of DNA sequence data, discrepancies within the sequence annotation would result in significant error in the alignment of gene trees. The use of the *Gallus gallus* reference sequence as a model for genome structure is instrumental in the allocation of gene features, in addition software was used to identify the key features including tRNA secondary structures, start and stop codon positions and protein coding genes.

In the construction of whole genome sequences the identification of both overlapping gene regions and short non-coding intergenic spacers is critical to the correct alignment of complete mito-genome data. The correct alignment of sequences is vital to the construction of phylogenetic trees, the more diverse the group of taxa within a tree the more divergent the DNA sequences and the more difficult the sequence alignment, and therefore the phylogenetic tree construction.

2.11.2 Species & Gene Identification

The Basic Local Alignment Search Tool, nucleotide algorithm (nBLAST) of the National Centre for Biotechnology Information (Altschul, Gish, Miller, Myers, & Lipman, 1990), may be used to identify species and gene matches for DNA sequences. The nucleotide BLAST algorithm searches the combined DNA databases of GenBank, EMBL and DDJB for the closest nucleotide sequence homolog match to the query DNA sequence. All matches are scored according to percentage coverage, Expectation value and percentage identity.

2.11.3 Statistical Analysis Software

Several statistical packages are used for the alignment and phylogenetic construction of datasets throughout this research, depending on the nature of analysis performed.

Clustal X2.0 Software to perform the alignment of multiple nucleic acid sequences (Larkin et al., 2007)**.** Multiple sequence alignments provide the structure for phylogenetic reconstructions, through the pairing of primary homologous sites between sequences and length conversion to sequences of equal length through the inference of indels (Ogden & Rosenberg, 2006).

MEGA 7.0 Molecular Evolutionary Genetics Analysis (Kumar, Stecher, & Tamura, 2016), software for the alignment and statistical analysis of DNA sequences. The MEGA software has multiple functions in the analysis of DNA and was primarily used for alignment of DNA sequences, analysis of pairwise distances, model testing and phylogenetic tree construction under the Maximum Likelihood method. Initial statistical analysis was performed in the previous version of the software MEGA6.0

(Tamura, Stecher, Peterson, Filipski, & Kumar, 2013). Confidence interval for Maximum Likelihood phylogenetic inferences was estimated using the Bootstrap re-sampling method (Felsenstein, 1985). Character sites are re-sampled to infer variability of topology, nucleotide sites having four possible outcomes for each character (A, T, C & G) (Felsenstein, 1985). Support at the 95% confidence level is inferred by 95 % or more of the bootstrap estimates. Nodes with lower than 70% confidence level were excluded from all tree topologies.

Geneious R8.0 Version 8.1.7, (2015), Biomatters Ltd. (Kearse et al., 2012)**.** The Geneious software platform has multiple functions in the analysis of DNA data, primarily it has been used in this study for the alignment of multiple sequences, genome construction and genome structural analysis.

CIPRES Science Gateway This portal operates numerous phylogenetic software packages through the San Diego Supercomputer Centre. The Science Gateway is designed to enable the inference of large phylogenetic trees (Miller, Pfeiffer, & Schwartz, 2011), in this research analysis of whole genome datasets was performed through this portal.

- ***MrBayes 3.2.6*** Through the CIPRES Science gateway. MrBayes employs tree inference using Bayesian Analysis (Huelsenbeck & Ronquist, 2001)**.**
- ***BEAST 2.2*** Through the CIPRES Science Gateway. Bayesian Evolutionary Analysis by Sampling Trees (Drummond, Suchard, Xie, & Rambaut, 2012)**.**

MrBayes 3.2.6 (Huelsenbeck & Ronquist, 2001)**,** the MrBayes software package reconstructs phylogenetic concensus trees through the use of Bayesian Inference posterior probability estimates through Marcov Chain Monte Carlo sampling under a specified substitution model with a Dirichlet distribution prior. Support for Bayesian Inference phylogenetic trees in MrBayes was estimated using the mean posterior distribution of the internal nodes and 95% highest posterior density (HPD) at the given node (support at the 95% confidence level is equivalent to a 95% HPD). Nodes with lower than 70% HPD were excluded from all tree topologies.

BEAST 2.2 Bayesian Evolutionary Analysis by Sampling Trees (Drummond, Suchard, Xie, & Rambaut, 2012), reconstructs rooted time-trees through the use of a Bayesian Inference method and MCMC model. BEAST infers the amount and duration of evolution between sequences, using prior information to infer age estimations under a clock model, time-tree node heights therefore correspond to the age of a node relative to the prior calibration point used (Drummond & Bouckaert, 2015). Data was prepared in the BEAUti v2.4.4 software (part of the BEAST package), to write the parameters for analysis, partition data and set priors. This analysis file was then analysed in BEAST 2.2 on the CIPRES Science Gateway. Node ages are given as a distribution mean, corresponding to the average node height calculated in all generated trees (excluding burin-in), and are supported by the 95% highest posterior density mean (HPD) which is equivalent to a 95% confidence interval.

FigTree v1.4.2 is a software package for editing raw phylogenetic tree files (http://tree.bio.ed.ac.uk/software/figtree/).

DensiTree v2.2.5 is software to visualise all trees generated by an MCMC analysis in a single file (Bouckaert & Heled, 2014), as part of the BEAST software suite.

Models used differ with each data set, see individual research chapter methodologies for full details of the statistical tests used.

2.12 Standard Solutions

50x Tris-Acetate-EDTA Buffer, pH 8.0

242g Trizma base (Sigma-Aldrich), 100ml 0.5M EDTA pH8.0 (Ethylenediaminetetraacetic acid) and 57.1ml Glacial Acetic Acid in a final volume of 1Litre, pH adjusted with 5M HCl.

1x Tris-Acetate-EDTA Buffer, Running Buffer

20ml 50x Tris-Acetate-EDTA Buffer, pH 8.0 in a final volume of 1Litre (980ml De-ionised water).

Chapter 3. Complete Mitochondrial Genome of the Trinidad Piping Guan

3.1 Introduction

The Mitochondrial Genome

Mitochondria are highly functional sub-cellular organelles involved in the processes of oxidative phosphorylation and are essential in the production of ATP, as well as other biological functions. Mitochondria organelles contain a small circular genome of between 15 to 20 kb, formed in a covalently closed circular structure with compact and conserved organisation, comprising 37 functional genes and a non-coding control region. The complete mitochondrial genome is uniquely suited to phylogenetic studies in that its is a relatively small compact genome which evolves at a faster rate than nuclear DNA (Kocher et al., 1989; Prychitko & Moore, 2000; Sorenson, Ast, Dimcheff, Yuri, & Mindell, 1999). The mtDNA genome demonstrates compact gene packing, with short non-coding intergeneric spacers, and limited gene overlaps (Pereira, 2000). The mutation rate of the mitochondrial genome is much higher than that of the nuclear genome; the average rate of divergence for mitochondrial genomes in mammals is 2% per million years compared to 0.7% for synonymous sites in mammal nuclear DNA (Armsrtong et al., 2001; Hewitt, 2000; Kan et al., 2010). In addition its lack of recombiation and maternal mode of inheritance make the mtDNA genome ideal for molecular analyses in phylogenetics and evolutionary research.

Complete mitochondrial genomes were first sequenced in 1989, initially in humans (Anderson et al. 1981) and then subsequently in a number of domestic and model species. The first complete avian mitochondrial genome was published in 1990 by Desjardin and Morais, for the Chicken or Red Junglefowl *Gallus gallus*, standard model avian system used in molecular and other research. The application of a 'Universal' avian primer set by Sorenson et al. (1999, 2003), for the amplification of whole mitochondrial genomes by PCR, has made avian genomics more accessible for many research applications.

Vertebrate mitochondrial genomes are well suited to the reconstruction of evolutionary relationships, producing robust well supported topologies (Meiklejohn et al., 2014). Whole mitochondrial genomes have been shown to improve the resolution of phylogenetic relationships, provide more informative sites and larger amounts of mitogenomic evolutionary data (Sorenson, Ast, Dimcheff, Yuri, & Mindell, 1999). Complete mitochondrial genome sequencing is advantageous in that it avoids issues of short sequence resolution, gene bias and variable gene substitution or evolutionary rates. The use of complete mitochondrial genomes may alter the traditional classifications of avian clades developed using single gene regions (Kan et al., 2010; Meiklejohn et al., 2014).

Mitochondrial DNA can offer valuable insights into matrilineal population structure, potentially demonstrating large amounts of variation within and between conspecific populations, within the context of a conserved gene arrangement. The maternal mode of inheritance of the mitochondrial genome, without recombination also offers insights into the evolutionary history of populations and species relationships. Evolutionary processes have been directly influenced by physio-geographic changes throughout history; geographical structuring, barrier effects on geneflow, fragmentation and isolation, extinctions and re-colonisations, which may be evident in the evolution of the mitochondrial genome.

Nuclear markers have been utilised widely in phylogenetic analysis, however there are several limiting factors such as recombination, multiple gene copies and allelic variation which are not limitations in the use of mitochondrial genes or genomes. However, the mitochondrial genome does demonstrate a number of limiting characteristics, notably the uni-parental inheritance, variable evolutionary rates and therefore a non-constant molecular clock, and non-neutrality (Zhang & Hewitt, 1996). Matrilineal inheritance can have significant influences on the interpretation of information from mitochondrial DNA, one of the most important being in dispersal bias against females (Lucchini & Randi, 1998).

Avian Mitochondrial Genome Structure

The standard vertebrate mitochondrial genome consists of 37 genes; 13 protein coding genes, 22 tRNAs, 2 rRNAs and a non-coding control region. The order of genes within the mitochondrial genome has been shown to vary between Class and in some instances between Orders. The typical vertebrate mitochondrial genome, on which nomenclature is usually based, is derived from the human mitochondrial genome published in 1981 by Anderson et al. This gene order has since been found to be conserved among many species of mammals, amphibians and some fish (Macey, Larson, Ananjeva, Fang, & Papenfiss, 1997; Pereira, 2000), however in avians the gene order between the NADH dehydrogenase 5 and the Control region differs from the typical form. Gene rearrangement of the mitochondrial genome is theorised to result from a number of mechanisms including inversion, transposition, intra-molecular recombination, replication errors of the L strand or tandem duplication and deletion (Macey et al., 1997; Mindell, Sorenson, & Dimcheff, 1998; Pereira, 2000). However rearrangement of the genes is functionally constrained at both the gene and genome level, therefore the mitochondrial genome remains largely consistent between taxa within orders.

The use of complete mitochondrial genomes as opposed to single gene or multi-gene analysis may be able to resolve traditionally held views regarding the avian evolutionary tree. The implementation of new molecular techniques will make whole genome sequences more accessible to research, as such, knowledge of the evolution of species and the mitochondrial genome may be elucidated.

A number of whole mitochondrial genomes of avians have now been published which allow a better understanding of the phylogenetic relationships amongst avian orders. The avian mitochondrial genome is typically 16.5-18.5kb in length (the largest reported avian genome is that of the Great Cormorant *Phalacrocorax carbo* at 18,995nt), with much of the length variation reported in the control region the result of small repeat units and Cytosine strings (Slack, Janke, Penny, & Arnason, 2003). The typical ‘conserved’ gene order for avians, was described by Desjardain & Morais

(1990) for the *Gallus gallus*, avian reference mitochondrial genome, see Figure 3.1. Research into other avian families has identified several other avian mitochondrial gene organisations which vary from the typical gene order relative to the evolution of the families (Mindell et al., 1998; Pereira, 2000), see Figure 3.1.

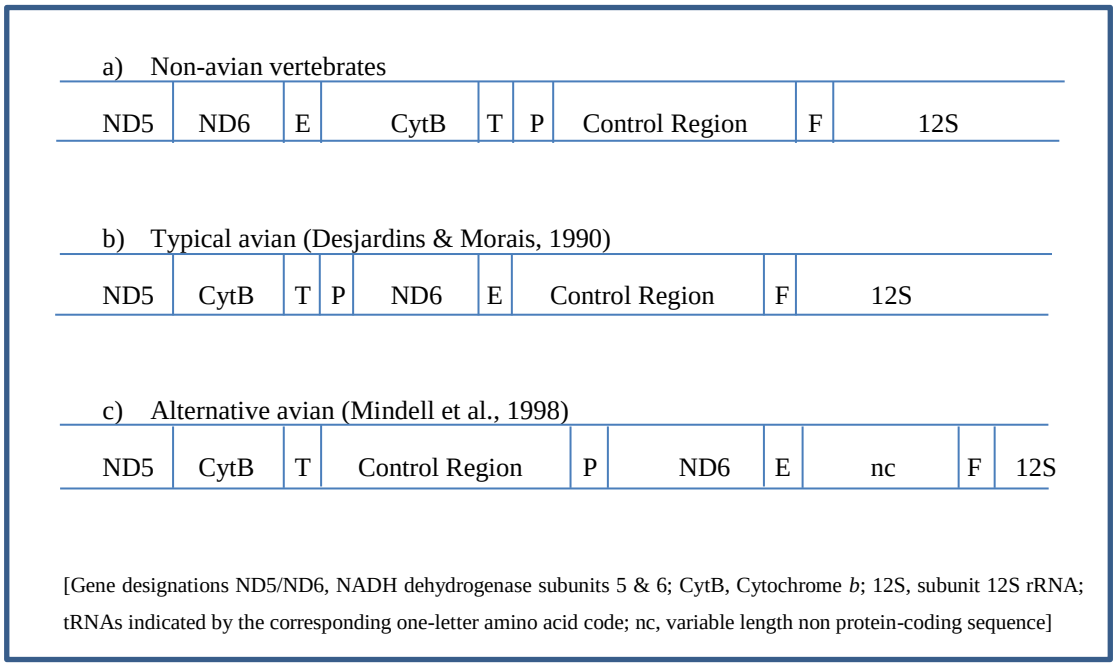


Figure 3.1: Mitochondrial gene order from ND5 to 12S RNA as described for the typical vertebrate (non-avian) animal mitochondrial genome (a), the typical avian mitochondrial genome (b), and the alternate avian mitochondrial genome (c). Adapted from Mindell et al., 1998.

Avian History and the Mitochondrial Genome

The avian class is one of the major lineages which diverged from the Archosaurs over 100 million years ago, being closely allied with the Crocodilians and reptiles (Janke & Arnason, 1997). The phylogenetic relationships among extant avian orders remains unresolved, general consensus implies that extant birds are divided into two groups, the Paleognaths (Ratites, Tinamous etc.) and the Neognaths (all other modern avian orders) (Mindell et al., 1999; Slack et al., 2003). Within the Neognaths several conflicting evolutionary trees have been proposed, and much research has been undertaken to resolve the avian relationships and estimate molecular timescales for the diversification of extant avians using both mitchondrial and nuclear genomic data.

Basal among the Neognaths are the Galliformes (landfowl), allied with the Anseriformes (waterfowl), often referred to as the Galloanserae (Dimcheff, Drovetski, & Mindell, 2002; Dyke, Gulas, & Crowe, 2003; van Tuinen & Hedges, 2001; van Tuinen, Sibley, & Hedges, 2000).

The Galliforme order is considered to be one of the most important groups of avians to human society, comprising 281 species in 81 genera, the order is of significant socio-economic and ecological importance featuring some of the most prominent domestic and commercial avian species (Chickens, Pheasants and Turkeys). Despite being one of the most studied avian orders, genetic structure and taxonomic relationships within the Galliformes remain unresolved. Within the Galliforme order there are seven well accepted groups, however their true classification remains unresolved (Dyke et al., 2003; Kan et al., 2010). The seven accepted classifications within the Galliforme order comprise; Megapodidae, Cracidae, Odontophoridae, Numudidae, Phasianidae, Meleagrididae and Tetraonidae.

Genetic relationships within the Galliforme order are reported to have been heavily influenced by temporal and geographical events across the world. The glacial oscillations of the last Ice age have had a significant influence of the structure of the Avian class including the Galliforme order, influencing the evolution and diversification of species. This diversification has been well documented both morphologically and genetically (Crowe et al., 2006; Dyke et al., 2003; Kimball, Braunj, Ligonl, Lucchini, & Randi, 2001; Van Tuinen, Sibley, & Hedges, 2000; Van Tuinen & Dyke, 2004), although much debate still remains and little research has incorporated the Cracidae family in analysis of Galliformes. Previous research into Galliforme lineages and phylogenetics has largely focused on single gene or multi-gene analysis, rather than whole genome, and have resulted in significant inconsistencies in topologies (Armsrtong et al., 2001; Dimcheff et al., 2002; Nishibori, Hayashi, Tsudzuki, Yamamoto, & Yasue, 2001).

Mitochondrial DNA is an ideal model to evaluate the genetic relationships of Galliformes in response to recent glacial fluctuations and the subsequent diversification of the order. As mitochondrial DNA evolves at a relatively slow rate, 2% per million years, there has been only a short period in which mutations have become fixed within the genomes of sister phylogroups since the last glacial cycle approximately 11,000 years ago (Zink, Klicka, & Barber, 2004). Therefore, taking into account effective population sizes, polymorphism levels and generation lengths, the origin of new mutations within the genome can be accredited to relatively recent isolation and demographic events.

The Cracidae are well established within the Galliformes as one of the oldest lineages, closely allied with the Megapodidae (Dimcheff et al., 2002; Dyke et al., 2003; Sibley & Ahlquist, 1990). At the time of writing only two other complete mitochondrial genomes have been published for Cracidae species, both of the Crax genus (*Crax blumenbachi* and *Crax rubra*), these having only recently been used in phylogenetic analysis of the Galliforme order in research by Meiklejohn et al. (2014). This research acknowledges the limitations in the analysis of Cracidae species within the Galliforme lineage, as a distinct gap in the basal topological analysis of the phylogenetic tree (Meiklejohn et al., 2014). The addition of a mitochondrial genome for the Trinidad Piping Guan has the potential therefore to improve the resolution of this clade and add support the Cracidae branches within the Galliforme tree.

As discussed in Chapter 1, the Piping Guans have been subject to much controversy regarding their inter-specific relationships and the genus is not well defined genetically, with little research to support the genetic history of the *Pipile* genus and the wider Cracidae clade. Piping Guans of the genus *Pipile* (or *Aburria*) are represented in a number of phylogenetic studies. However, all research conducted on these species uses single gene or multi-gene analysis, as opposed to whole mitochondrial genomes (Grau, Pereira, Silveira, Höfling, & Wajntal, 2005; Pereira, Grau, & Wajntal, 2004), and as such their genetic status remains undefined.

The Trinidad Piping Guan, is theorised to have been subject to direct isolation in the last glacial climate shift (Grau et al., 2005). This species, as well as other flora and fauna on the island have therefore diversified in isolation from their continental congeners, as a consequence of this geographical shift and adaptive responses to the environment. The novel sequencing of the complete mitochondrial genome for the Trinidad Piping Guan may inform evolutionary hypotheses regarding the evolution of this genus. Additionally, the application of genetic markers to the analysis of the population structure and evolutionary history of the Trinidad Piping Guan may provide important information for the future management of the declining population. The sequencing of the complete mitochondrial genome represents a significant advance towards determining suitable genetic markers for analysis in this species and is also of significant benefit to the direct analysis of its evolutionary history.

The illegal trade of rare and endangered birds is of international concern, and thus molecular methods of species identification are of increasing importance to conservation and wildlife forensics. The use of samples of traceable origins is critical, where possible living specimens or preserved specimens in collections which have been well documented. Species identification by molecular techniques is underpinned by the information available in the collaborative DNA databases, with quality of data and genetic coverage being of significant import. The underlying taxonomic provenance of samples within the databases is critical to the reliability of the information provided and the subsequent integrity of analysis (voucher and reference samples). Currently there is only a small amount of genetic information for *Pipile pipile* and its congeneric species published on DNA databases, including single mitochondrial genes and limited nuclear DNA markers. The inclusion of complete mitochondrial genomes for this species, has potential benefits in species identification and forensic applications, and therefore the conservation of this critically endangered species.

Sample Types & Whole Genome Amplification

The preferred source of genetic material from avians is blood, as a rich source of both mitochondrial and nuclear DNA, however the use of other biological materials is widely published, as discussed in Chapter 1(Bush et al., 2005; Coghlan et al., 2012; Egloff et al., 2009; Martín-Gálvez et al., 2011; Oskam et al., 2010; Trimbos et al., 2009; Wink, 2006). Feather samples typically yield lower levels of DNA, depending on the feather type, condition of the feather (age and degradation) and the species (Hogan et al., 2008). Feathers, both moulted and plucked are widely used in avian genomics and phylogenetic analyses (Gebhardt, Brightsmith, Powell, & Waits, 2009; Harvey, Bonter, Stenzler, & Lovette, 2006; Hogan et al., 2008; Johannson et al., 2012; Leeton & Christidis, 1993; Ong & Vellayan, 2008), however they are not typically used in the sequencing and analysis of whole mitochondrial genomes. Advancements in PCR and sequencing technologies, now make feathers a suitable genetic source for the amplification of complete mitochondrial genomes.

Study Objectives

This study focuses on the production of a high quality, reference mitochondrial genome for the Trinidad Piping Guan. The study will use PCR techniques to amplify the complete mitochondrial genome from the reference Trinidad Piping Guan samples, which are considered to be of voucher standard. The Trinidad Piping Guan mitochondrial genome may then be ustilised in subsequent research to inform questions regarding intra-specific variation and inter-specific relationships.

3.2 Methodology

3.2.1 Trinidad Piping Guan Samples

Three of the Trinidad Piping Guans samples imported from the wild in Trinidad were selected for analysis of the whole mitochondrial genome; the chosen samples were those of voucher standard independently verified by experts as being Trinidad Piping Guans, of wild origin.

Table 3.1: Samples used in whole mitochondrial genome analysis

Sample Identifier	Description	Comments
TPG-R1	Reference 1	Captive, Originated in the wild, Deceased 2007 CITES permit 1053 Species confirmed by veterinary examination & DNA
TPG-R2	Reference 2	Captive, Originated in the wild, 2010 CITES permit 1264 Species confirmed by veterinary examination & DNA
TPG-E1	Egg-embryo	Wild, 2009 CITES permit 1116 Species confirmed by veterinary examination & DNA

The embryo liver tissue sample (TPG-E1) was initially selected for development and optimisation of primers, due to the high quality DNA available from the liver tissue of this sample, facilitating larger fragment size amplifications. The reference samples (TPG-R1 and TPG-R2) from Trinidad Piping Guans caught from the wild in Trinidad and held in captivity, were used to further establish verifiable complete mitochondrial genomes for the purpose of publication on Genbank. In this respect, the use of three unrelated individuals allowed the sequence data accuracy to be assessed and confirmed, genome structure to be unambiguously tested and genetic relationships to be tested at a genomic level, as well as informing the analysis of intra-specific mito-genomic variation.

3.2.2 PCR & Primers

Optimisation of the method for sequencing was achieved by PCR amplification of the samples using the universal primer sets published by Sorenson (2003) and primers designed from Trinidad Piping Guan sequence data in this research. The Trinidad Piping Guan genome was amplified in fragments of up to 1.2kb and were sequenced in both directions (Heavy and Light strand sequenced). All fragments were designed to overlap by 100bp to allow accurate construction of the genome. A total of 90 oligonucleotides (45 primer sets) were used to amplify and sequence the whole genome (see Appendix 2), 24 of which were specifically designed from Trinidad Piping Guan sequence obtained in this study. Species-specific oligonucleotides were designed using nucleotide sequence from previously successful amplifications on the boundaries of the desired region, using Primer3 as previously described in Chapter 2.

3.2.2 PCR Amplification

Polymerase Chain Reaction were performed using Illustra™ puRe Taq Ready-To-Go PCR Beads (GE Healthcare). Preparation of the PCR reaction mixture was carried out as previously described in Chapter 2. At all stages sample integrity was maintained through accurate record keeping and labelling with the appropriate sample identification, samples were isolated to prevent cross-contamination with DNA from other Piping Guan individuals or species. Replications were performed using separate PCR reactions from the same original DNA extraction. Each amplification was performed independently (in triplicate), to ensure the accuracy and validity of sequence data produced.

Thermal cycling was performed using either cycles designed specifically for the oligonucleotides being used or a generic Touchdown PCR cycle designed to work with oligonucleotides with a T_m within the appropriate range. Further details of PCR cycles and Touchdown PCR cycles is given in Chapter 2.

All amplified PCR products were examined by agarose gel electrophoresis to determine the quality of the amplification and allow basic quantification of the

amplified product. Standard 2% w/v agarose gels were used for the analysis of fragments, as previously described in the Chapter 2. The gel images were analysed using the Image Lab software (Bio-Rad). Purification of the PCR products was carried out using GenElute™ PCR Clean Up Kit (Sigma-Aldrich) or MinElute Gel Extraction Kit (Qiagen), see Chapter 2 for further details.

Prior to sequencing DNA amplifications were subjected to quantitative analysis and concentration or dilution as appropriate for the sequencing of the DNA sample, as recommended by the MWG Eurofins sample preparation guidelines. Image Lab software (Bio-Rad) was used to determine a DNA concentration for all PCR products using Norgen PCR Sizer 100bp (Geneflow) molecular ladder for the standard curve. Sequencing of the amplified mtDNA samples was carried out by MWG Eurofins (MWG Biotech), samples were prepared as per MWG Eurofins purified PCR products requirements (www.eurofinsdna.com) for sequencing services.

3.2.3 Sequence Analysis

Quality Assessment of the Sequence Data

Visual analysis of the electropherograms for each sequence was performed to identify areas of concern according to the QV scores, for quality and accuracy of the nucleotide assignments in the sequence. Areas with low scores were eliminated from subsequent analysis or short regions were clipped from the end of the sequence to ensure the accuracy of the final sequence for use in comparative analysis. Areas of poor quality sequence data were re-sequenced or re-amplified where necessary to ensure the fidelity of nucleotide assignments within the genome sequence.

Confirmation of the Sequence Accuracy

All sequences were aligned prior to analysis; Forward and Reverse sequences were aligned using the BLASTn software, (http://blast.ncbi.nlm.nih.gov/Blast.cgi), to confirm the nucleotide assignment of the sequence results. The BLASTn software was also used to search for species identifying matches, the model searches the combined

NCBI nucleotide databases (GenBank, EMBL and DDBJ), to identify homologous sequences which may aid in confirmation of species identification. Due to the lack of published sequence data for this genus on the combined NCBI databases, homology of the sequences to published data for avians within the *Cracidae* order were accepted as identification of accuracy of the sequence origin and eliminating the possibility of contamination from other sources. Sequences were considered to be accurately identified if they fell within the accepted limits of the appropriate gene; homology to both the correct mitochondrial gene and the appropriate region of that gene, with an acceptable level of identification over the amplified region (according to the species of comparison within the database).

Electropherograms were examined for indicators of possible NUMTs; distinguishing features include double peaks from co-amplification, unexpected indels, frameshifts, and mismatches in regions of overlap between sequences (Haddrath & Baker, 2001; Mindell et al., 1999). Subject sequences were analysed in both BLAST and Geneious R8 software packages, with the specific aim of identifying features which may be indicative of NUMTs.

Genome Sequence Analysis

Multiple sequence alignments were performed using Geneious R8 software with subsequent visual verification. All sequences were aligned with the reference sequence for the *Gallus gallus* whole mitochondrial genome (accession number X52392), as a standard reference point for nomenclature of the sequences. Data from all three individual samples and the *Gallus gallus* reference were aligned to aid in the constructional elements of the genome structure.

Genome structure

Genome characteristics were identified by alignment with reference whole mitochondrial genomes from NCBI, including the standard avian reference *Gallus gallus* (accession number X52392) published by Desjardin and Morais (1990). The

mitochondrial genomes were annotated according to standard mitochondrial genome nomenclature, in comparison to reference sequences and where present indicative gene features (initiation and termination codons etc.).

Alignments were performed in Geneious R8.0, to identify the key indicative genome features; genes, non-coding regions, initiation and termination codons, intergenic gaps and conserved features. All features were annotated onto the genome for future reference. Subsequently genome sequences were annotated in the Sequin version 13.70 (Network Aware, 2015) software for publication to the NCBI database, under the following accession numbers K221021, K221052 and K221053.

Statistical analysis of mtDNA gene sequences involved the categorical identification or exclusion of the individual gene sequences from the published reference sequences, according to the observed sequence variations. Sequence comparisons were made between the samples and the species reference standards (NCBI). Analysis and identification was performed on genes from both the Heavy and Light strands, a total of 13 protein-coding genes, 2 rRNAs and 22 tRNAs. Initiation and termination (start and stop) codons were subsequently identified by translating the sequence using the standard vertebrate mitochondrial amino acid nomenclature in Geneious R8.0, and subsequently in Sequin version 13.70. Incomplete termination codons were identified as incomplete codons such as single nucleotides at the end of the gene, with reference to gene positions in the standard mitochondrial genome for comparison (*Gallus gallus* X52392).

The identification of tRNA genes and their secondary structures was performed using the software tRNAscan-SE 1.21 (Lowe & Eddy, 1997), under default search parameters with the vertebrate mitochondrial source reference. Where tRNA structures were undetected by tRNAscan-SE identification of the genes was carried out by visual identification and comparison with published Galliforme genomes in Geneious R8.0.

Using sequence alignments in Geneious R8.0, the single non-translated nucleotide at position 174 of the ND3 gene, common to most avian orders including Galliformes (Mindell, Sorenson, & Dimcheff, 1998), was identified for reference in subsequent sequence and genome analyses. This single nucleotide insertion results in a frameshift in the translation to amino acids for the gene and as a consequence this non-coding nucleotide was removed from the analysis in all subsequent data sets.

3.3 Results

3.3.1 Compositional Analysis of the Whole Mitochondrial Genome

The genome structure and gene order of the mitochondrial genome of the Trinidad Piping Guan conforms to the typical gene arrangement of other avians in the Galliforme order, including the standard reference *Gallus gallus* (Desjardins & Morais, 1990). The complete mitochondrial sequence was obtained for three individual Trinidad Piping Guans and all compositional features were identical in all individuals.

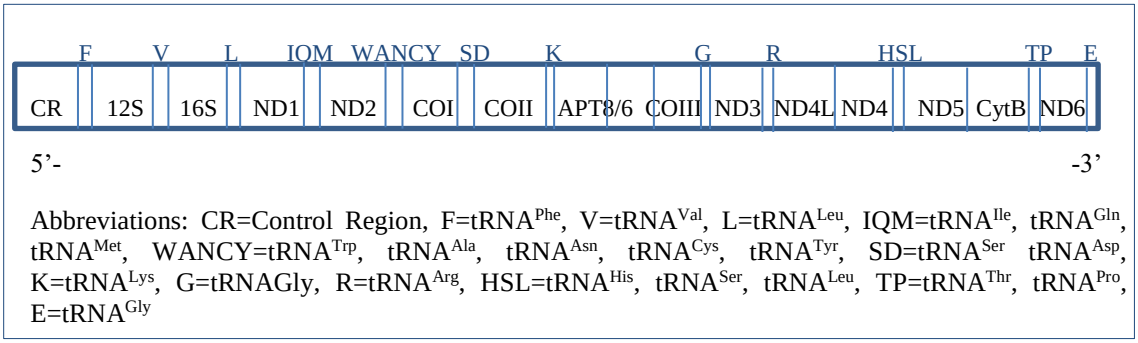


Figure 3.2: Mitochondrial gene organisation in the Trinidad Piping Guan, *Pipile pipile.*

The complete sequence for the Trinidad Piping Guan mitochondrial genome is 16,664bp±2 in length, and contains thirteen protein coding genes, two rRNA genes, twenty-two tRNA genes and a non-coding Control region, in conformation with standard vertebrate mitogenomic structure. The Trinidad Piping Guan mitochondrial genome structure and gene order is consistent with the standard avian genome arrangement as per Mindell et al., (1998), see Figures 3.2 and 3.3. Most protein coding genes are interspersed with at least one tRNA gene, and the standard tRNA packages (WANCY, IQM and HSL) are present in positions consistent with other Galliforme genomes.

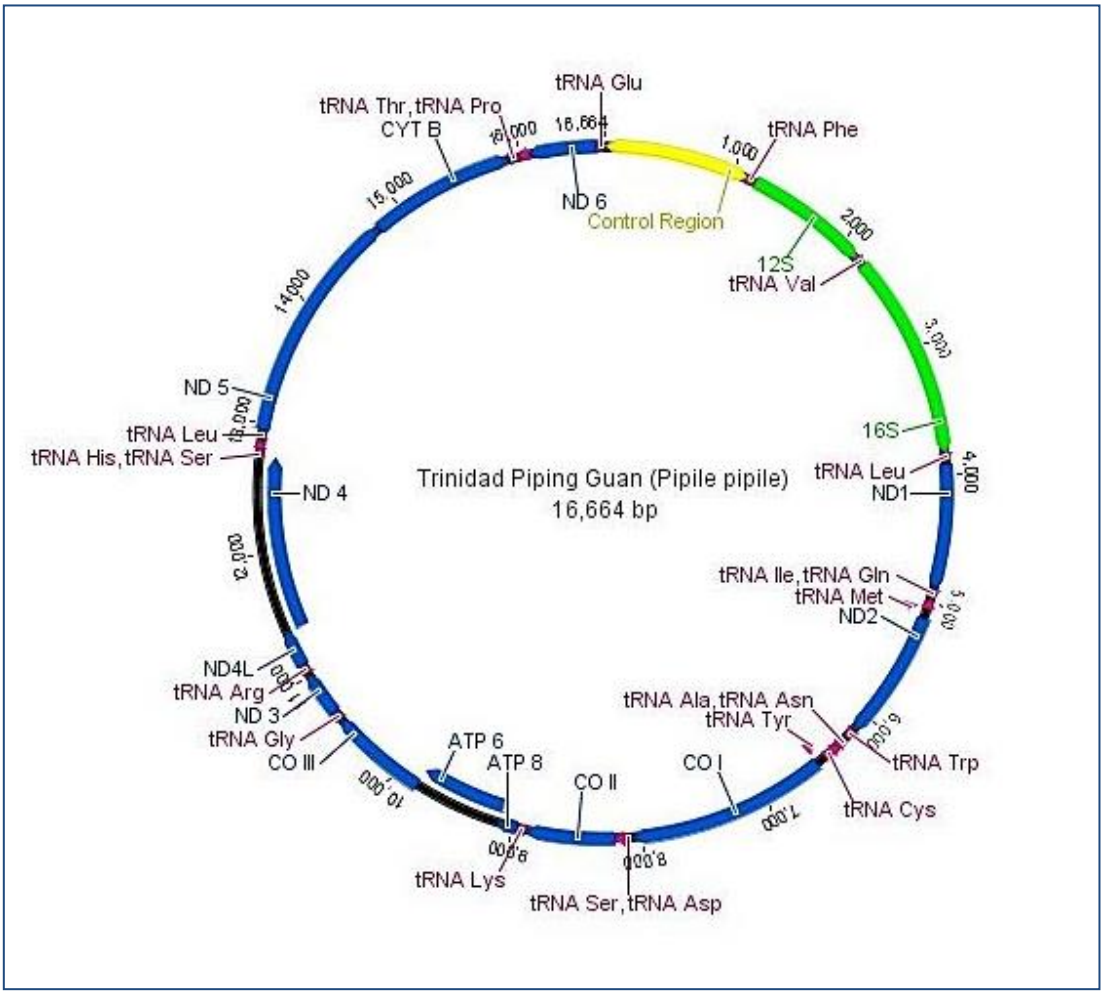


Figure 3.3: The genomic organisation of the Trinidad Piping Guan mitochondrial genome, Nucleotide position 1 is located at the 5' end of the Control region.

The majority of genes are contiguous within the sequence however gene overlaps occur in four locations; tRNAGln and tRNA Met, tRNA Cys and tRNATyr, ATP8 and ATP6, and ND4L and ND4 consistent with gene overlap features in other Galliformes. Additionally non-coding intergenic spacers of variable length (1-11 nucleotides) occur between genes throughout the genome, see Table 3.2.

3.3.2 Gene Identification & Features

Gene regions were identified through the comparison of the genome with reference standards including the *Gallus gallus* standard reference sequence X52392. Identification of the start and end features of the genes were identified through the translation of the genome sequence using the Geneious R8 software. Initiation codons

and termination codons were identified in translation using the standard vertebrate amino acid coding nomenclature.

Base composition of the Trinidad Piping Guan complete mitochondrial genomes demonstrate an A+T bias (53.8%), where overall base composition of the genome were as follows; A (30.1%), T (23.7%), C (32.3%) and G (13.9%).

3.3.3 Protein-coding Genes

All Protein coding genes were consistent in both size and organisational structure with other Galliforme mitochondrial genomes. Protein coding genes were identified through the observation of initiation and termination codons in the amino acid sequence (standard vertebrate mitochondrial genome definitions). The initiation codon ATG is the most frequently observed in the protein-coding genes, ATA and GTG were also observed in the ND1 and the COI and ND4L genes respectively. Most of the genes use the TAA or AGG termination codons however, the genes ND2, ATP8, COIII, ND4L and ND4 demonstrated incomplete stop codons in the form of either T or A codons consistent with observations in other avian genomes.

The additional non-translated nucleotide at position 174 of the ND3 gene was observed in all three individual Trinidad Piping Guans, consistent with the observation of this additional nucleotide in other Galliforme species (Mindell et al., 1998).

Table 3.2: Gene information for the complete mitochondrial genome of the Trinidad piping Guan (accession number KU221051).

	Strand*	Length	Position	Anti-codon	Initial codon	Terminal codon	Overlap (Number of nucleotides)	Intergenic spacer (Number of nucleotides)
Control Region		1124	1-1124					
tRNA Phe		69	1125-1194	GAA				
12S		977	1195-2172					
tRNA Val		70	2173-2242	TAC				TG (2)
16S		1593	2245-3838					
tRNA Leu		74	3839-3912	TAA				CCTCCCAACCA (11)
ND1		975	3924-4898		ATG	TAA		
tRNA Ile		72	4899-4965	GAT				AAGAAAC (7)
tRNA Gln	C	71	4965-5045	TTG			A/T (1)	
tRNA Met		69	5045-5116	CAT				ATGAACCCC (9)
ND2		1029	5126-6155		ATG	T#		
tRNA Trp		80	6156-6235	TCA				CAACACT (7)
tRNA Ala	C	72	6243-6311	TGC				CC (2)
tRNA Asn	C	72	6314-6385	GTT				AA (2)
tRNA Cys	C	66	6388-6452	GCA			C (1)	
tRNA Tyr	C	69	6452-6524	GTA				TG (2)
COI		1547	6527-8074		GTG	AGG		
tRNA Ser	C	77	8075-8141	TGA				TAT (3)
tRNA Asp		69	8145-8213	GTC				C (1)
COII		683	8215-8898		ATG	TAA		
tRNA Lys		72	8899-8971					
ATP 8		163	8972-9135		ATG	A#	ATGAACCTAA (10)	
ATP 6		683	9126-9809		ATG	TAA	A (1)	
COIII		783	9809-10592		ATG	T#		
tRNA Gly		69	10953-10661	TCC				
ND3		351	10662-11013		ATG	TAA		C (1)
tRNA Arg		70	11015-11084	TGC				
ND4 L		296	11085-11381		GTG		ATGTTAA (7)	
ND4		1377	11375-12752		ATG	T#		
tRNA His		69	12753-12811	GTG				
tRNA Ser		73	12813-12886					
tRNA Leu		70	12888-12958					
ND5		1817	12959-14776			TAA		
Cyt B		1143	14777-15919		ATG	TAA		
tRNA Thr		70	15923-15992	TGT				AC (2)
tRNA Pro	C	70	15995-16064	TGG				
ND6		529	16065-16594		AAC	CAT		
tRNA Glu	C	68	16595-16660	TTC				AACT (4)

* Strand coding C=Complement # Incomplete terminal codon

3.3.4 tRNA genes

tRNA genes were identified through comparison with standard reference genomes and the identification of probable secondary structures using the tRNAscan-SE 1.21 software (Lowe & Eddy, 1997). tRNA secondary structures were identified for twenty of the twenty-two tRNA genes using the prediction of secondary structures, see Figure 3.4. $tRNA^{Lys}$ and $tRNA^{Ser}$ were not compatible with the secondary structure software, suggesting variation in the structure in comparison to the standard vertebrate mitochondrial reference. Conformation of these tRNA sequences with other Cracid reference sequences using BLAST identified the probable start and termination points for the $tRNA^{Lys}$ and $tRNA^{Ser}$ (HSL) genes, as indicated in Table 3.2.

Eight of the twenty-two identified tRNAs were coded on the Complement strand. The average size of the tRNA genes was 70.9±3.06nt, the maximum length being 80nt for $tRNA^{Trp}$, the shortest being $tRNA^{Cys}$ at 66nt. Structure and size of the tRNA genes is consistent with that observed in other Galliforme mitochondrial genomes.

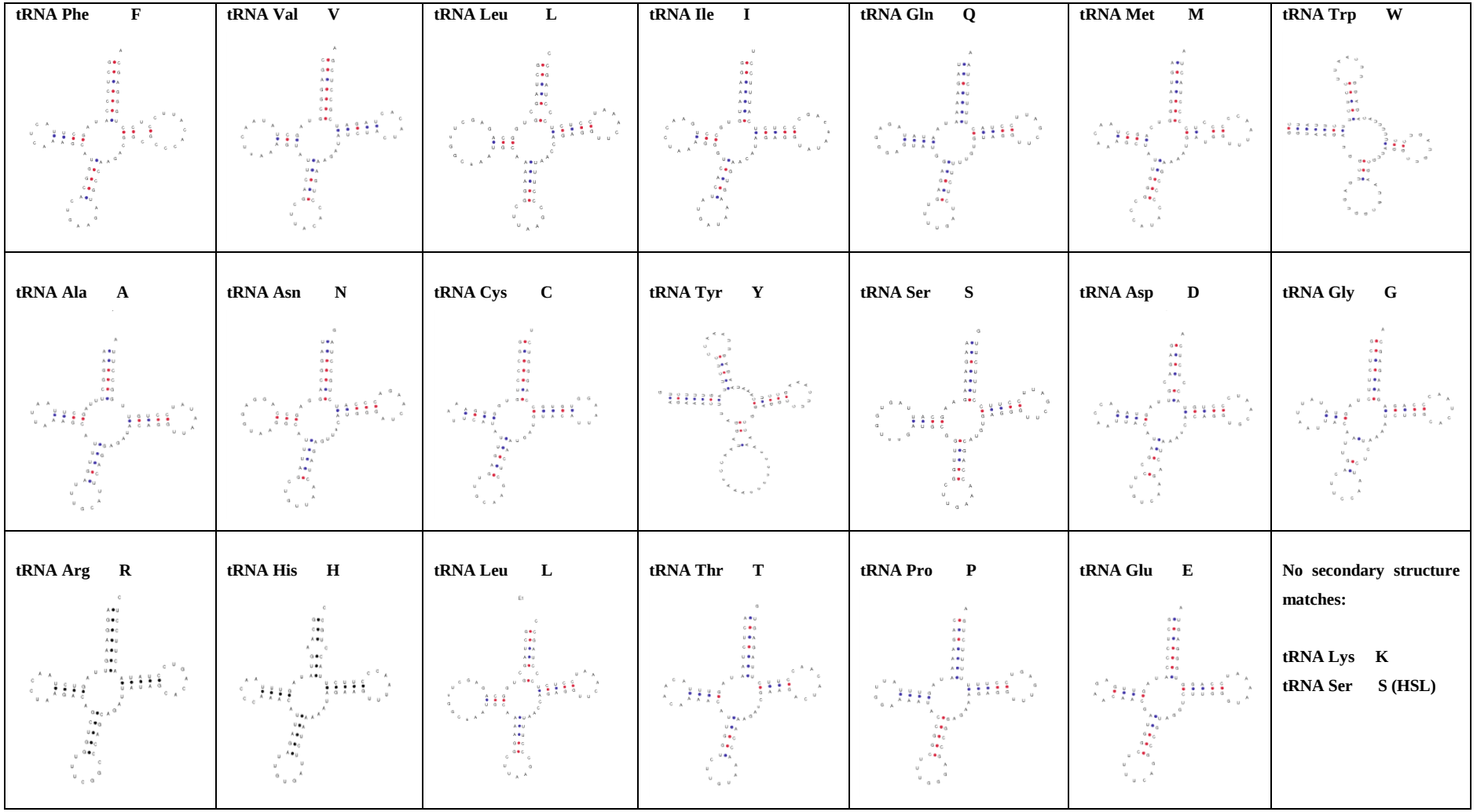


Figure 3.4: Secondary structures of tRNA genes of the Trinidad Piping Guan *Pipile pipile*, as constructed by tRNAscan-SE 1.21 software.

3.3.5 Non-coding Regions

The non-coding Control Region was identified using comparison with the standard reference sequences and the conserved features of the region including Conserved Sequence Blocks (CSBs) and structural features, as presented for Cracids by Pereira, Grau, & Wajntal (2004).

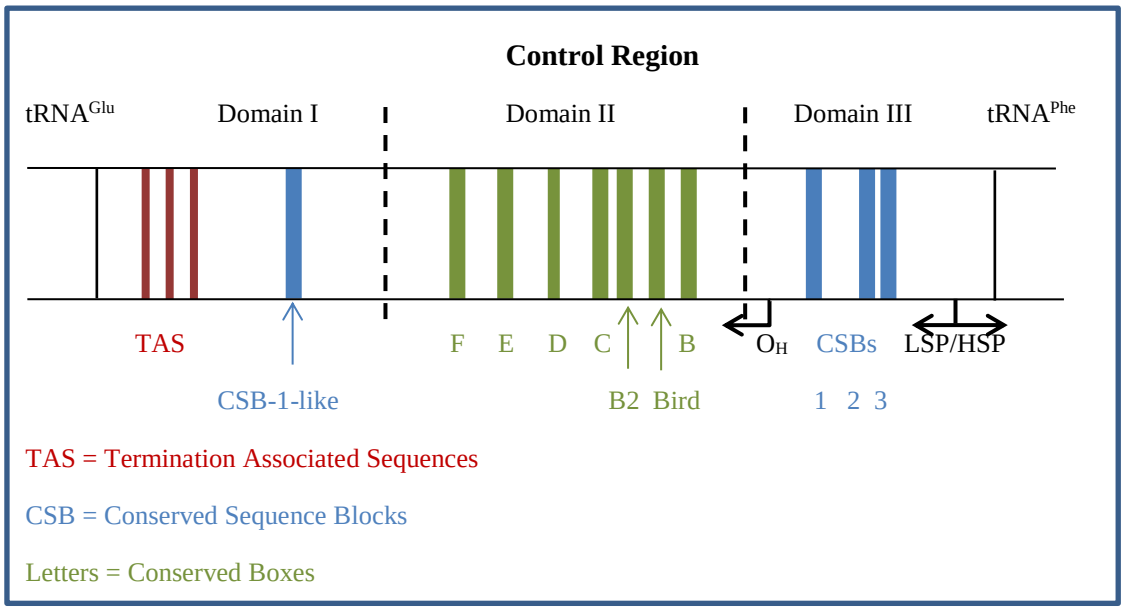


Figure 3.5: Conserved features of the Control Region of the Trinidad Piping Guan.

Domain regions within the Trinidad Piping Guan Control Region are consistent with other Cracids (Pereira et al., 2004), allowing the comparable identification of conserved features, see Figure 3.5. The sequence for domain I indicates the presence of several Termination Associated Sequences and a region which is like the CSB 1 block, all of which are observed in Cracids. Domain II contains seven of the known conserved boxes including the Saunders B2 box and the Bird box, which are strongly associated with avian genomes (Ruokonen & Kvist, 2002). The third domain contains the three common vertebrate Conserved Sequence Blocks as well as the Origin of Heavy strand replication (O_H) and both the Heavy and Light Strand Promotors (LSP/HSP). The Origin of Heavy strand replication is identifiable near the start of the third domain within the control region, and in common with other avian mitochondrial genomes, the Light strand origin of replication is absent from the Control Region (Desjardins & Morais, 1990; Marshall & Baker, 1997; Mindell et al., 1999; Ruokonen & Kvist, 2002).

3.3.6 Genomic Variation

Complete mitochondrial genomes were sequenced for the three individual Trinidad Piping Guans which were 16663±2bp in length, the shortest sequence being for Reference sample 1 at 16,661bp. The three individuals demonstrate 99.9% homology across the complete mitochondrial genome sequence (16,650 identical sites). Variations between the three individuals are outlined in Table 3.3. The length variation between individual mitochondrial genomes was due to indels in the Control Region at positions 58, 76, 734 and 1,032 (nucleotide insertions or deletions).

Table 3.3: Single Nucleotide Polymorphisms in the complete mitochondrial genomes of three individuals of the species *Pipile pipile*.

Nucleotide Position	58	76	734	1,032	4,737	5,808	8,385	9,068	9,147	11,132	14,814	15,781	15,915	15,999	16,143
Gene	**Control Region**				**ND1**	**ND2**	**COII**	**ATP8**	**ATP6**	**ND4L**	**CytB**			**tRNA^{Pro}**	**ND6**
TPG-R1	del	del	del	del	G	C	G	C	C	G	A	T	C	G	C
TPG-R2	C	del	T	A	C	T	C	T	T	T	A	T	C	A	T
TPG-E1	C	G	T	del	G	C	C	C	C	T	G	C	T	A	T

del = deletion from the sequence/ insertion of nucleotide at this position (indel)

There were 15 Single Nucleotide Polymorphisms (SNPs) observed in the three mito-genomes, indicative of a highly homologous sequence between individuals. The non-coding Control Region was most variable, with four indels observed between the three individuals. The most variable gene observed was the Cytochrome B gene, with three nucleotide variations, however the Reference samples 1 and 2 were 100% homologous throughout this gene, variation was only observed in the Embryo sample. The variation between the whole mitochondrial genomes of these three individual Trinidad Piping Guans is very limited (15/16,665nt), potentially indicative of a loss of genetic variation in the species.

3.4 Discussion

The complete mitochondrial genome was sequenced for three individuals of the Trinidad Piping Guan species (*Pipile pipile* synonym *Aburria pipile*), accession numbers for the genome entries on Genbank KU221051, KU221052 and KU221053 (Grass, Hosie, & McDowall, 2016). Publication of the genomes presented in Appendix 3. These mitochondrial genomes represent the first published sequences from contemporary wild Trinidad Piping Guans from the Island of Trinidad, available on GenBank, previous published voucher specimen sequence data being sourced from historical samples in the Smithsonian Institution collection.

Shed feathers of the two reference Trinidad Piping Guans were successfully used as the source of DNA for the sequencing of the whole mitochondrial genomes, demonstrating the utility of this DNA source for complete mito-genome molecular analyses. However the nature of the samples and environmental effects had a negative impact on the quality of the DNA recovered.

The mitochondrial genome of the Trinidad Piping Guan conforms to the standard Galliforme type, as published by Desjardin and Morais (1990) for the domestic Chicken *Gallus gallus,* and other Galliforme species. Gene organisation and size is consistent throughout the genome with the Galliforme type, as expected. Variation between the three individuals is characterised within the mitochondrial genomes based on single base substitutions or indels, otherwise referred to as single nucleotide polymorphisms (SNPs). Between the three individuals there is very little variation evident between the complete mitogenomes, indicative of a close homology between all individuals (99.9%).

This study uses three temporally distinct individuals, which originated from the wild at unknown locations. Thus, the individuals are unlikely to be direct sibling but given the small population size may share familial relationships and common ancestors.

Minimal genetic variation between individuals of such a small population may be indicative of a single dominant mitochondrial haplotype in the population, with only minor substitutional variation. This is potentially indicative of the loss of genetic variation under a number of circumstances, including founder population isolation, historical bottleneck events, recent significant population declines and possibly a contemporary population bottleneck situation at present. Evidence from further individuals randomly selected throughout the population is needed to assess further the full extent of the current genetic status of the species.

Loss of genetic variation is known to affect a species' risk of extinction, due to the effects on fitness and the ability of a species to adapt (Richard Frankham, 2005, 2008). Low mitochondrial variation within the Trinidad Piping Guan is not necessarily problematic in the long term, depending on the extent of population decline and the length of time the decline endures, populations are able to recover with minimal effects on fitness (Beissinger et al., 2008; Nichols, Bruford, & Groombridge, 2001; Weber, 2004).

3.5 Conclusions

This study has enabled the sequencing of three complete mitochondrial genomes from verified reference Trinidad Piping Guan individuals from the wild in Trinidad. The focus of this study has been the production of the complete mitochondrial genome for use in a number of applications including the further development of mitochondrial markers for the analysis of genetic variation, phylogenetic analysis and the inclusion of the genus within a Galliforme tree, as will be demonstrated in the subsequent chapters. These mitochondrial genomes may act as a standard reference for contemporary wild Trinidad Piping Guans in further work on the genus in numerous research areas.

Chapter 4 - Intra-specific Genetic Variation in the Trinidad Piping Guan

4.1 Introduction

The Trinidad Piping Guan

The Trinidad Piping Guan, known locally as the Pawi, is endemic to the island of Trinidad in the southern Caribbean. It is one of fifty-six recognised species of Cracid, which typify the Gallinaceous form in the South American Neotropics; large bodied, arboreal frugivoures found in tropical and sub-tropical forest habitat. The Trinidad Piping Guan, although once abundant throughout Trinidad, has become locally isolated to the Northern Range, anthropomorphic pressures being largely responsible for the species' decline.

Population density of the Trinidad Piping Guan has been estimated to be between 0.75-2.14 per km^2, based on a population estimate by Hayes et al. in 1994 of 70-200 individuals in an area of 150km^2 of suitable habitat. This is low in comparison to the estimated population density of Common Piping Guans in the Peruvian Amazon of 0.44-9.37 per km^2 (Begazo & Bodmer, 1998) and 3.38 per km^2 (Gonzalez, 1999). Black-fronted Piping Guans have been observed at population densities between 0.93-3.17 per km^2 in the Atlantic forests of Brazil, one of the most heavily human populated areas of South America (Galetti et al., 1997). The variation in population density of these *Pipile* species may be a reflection of the differences in available habitat and competition for resources in these different localities, and the influence of natural niche capacities within each environment.

The decline of Piping Guan species has been reported in a number of South American regions (Begazo & Bodmer, 1998; Brooks, 1999; Galetti et al., 1997; Pereira & Wajntal, 2001). There is a paucity of information and research into the Trinidad Piping Guan however, currently more information is available for the closely related,

mainland conspecific Common or Blue Throated Piping Guan (Pipile cumanensis). In comparison to the Common Piping Guan we may be able to infer life history traits which may be significant in future genetic trends in the Trinidad Piping Guan species.

The Trinidad Piping Guan is legally classified as an Environmentally Sensitive Species, awarding it the highest levels of conservation protection within Trinidad, and much of its habitat is protected as Environmentally Sensitive Areas (Alemu et al., 2005). Such protection in Trindadian legislation may be critical to the species' survival in its current habitat; Piping Guans are known to have limited dispersal capabilities due to their inability to fly large distances, and many species of the Cracid family are considered site faithful (Pereira & Wajntal, 2001). Site fidelity has been reported in a resident group of Trinidad Piping Guans in Grande Riviere since 1989, research focused on this group by Alemu et al. (2005) demonstrated consistent observations in a small habitat specific area of mature, regenerated forest.

Reproductive success is vital to population stability and recovery, and genetic variation within a population is known to play a key role in the hatching and mortality rates of juvenile avian species (Beissinger et al., 2008). Clutch size along with survivability of offspring has an important role in the ability of populations to recover, Begazo and Bodmer (1998) reported Common Piping Guan clutch sizes to be 2.6±1.3 eggs (nests observed between November and February). Small clutch size may be significantly detrimental to the sustainability of the Trinidad Piping Guan population, especially when considered in conjunction with illegal hunting pressures and habitat loss.

Pipile species and other Cracids are considered a significant source of protein to rural hunters, Latin Americans and Amerindians across South America (Brooks, 1999). Hunting of Cracids in the Amazon has been shown to be most abundant within 5km of native villages, and Common Piping Guans were heavily hunted within this area (approximately 59% of all Cracids hunted were Common Piping Guans), exceeding maximum sustainable harvest levels (Begazo & Bodmer, 1998). Hunting in Trinidad is a culturally ingrained behaviour linked to socio-economic trends and cultural

attitudes, despite increased awareness and improved perceptions of conservation and endangered species (Waylen, 2005). This demonstrates the impact of anthropomorphic pressures associated with the close proximity of human populations to critical habitat for large avians such as the Trinidad Piping Guan.

The relatively rapid decline of a species such as the Trinidad Piping Guan, with limited dispersal abilities, long generation times and low reproductive rates, will naturally result in the loss of diversity and increase in inbreeding which may result in a reduction of adaptive potential and an increased probability of extinction for the species. Determination of the genetic structure of such a population, with declining trends and endangered status, may be critical to its management and conservation.

The persistence of species requires that loss of individuals does not out-weigh gain; if the mortality or loss of a species through other means such as hybridisation or emigration is greater than the rate of births and immigration into the population, it is indicative of a species which is at a high risk of extinction. Failure to reproduce in sufficient numbers to keep up with rates of mortality is a great concern for the Trinidad Piping Guan. Critical life history traits of the Piping Guans restrict sustainability of this genus; low reproductive rates, long generation time, small clutch size, habitat dependence and poor dispersal strategies (Brooks, 1999). Population viability analysis requires substantial data resources from many sources including genetics, population structure, reproductive and mortality data, and habitat monitoring. Genetic analysis plays a significant role in monitoring wild populations, however it is also vital to establishing genetically viable captive populations in the future as both reserve populations and to supplement the wild populations. The Piping Guans are of great socio-economic and ecological importance to the Neotropical forest ecosystems of Trinidad, as previously discussed in Chapter 1.1 (Alemu et al., 2005; Brooks, Pando-Vasquez, & Ocmin-Petit, 2005; Waylen, 2005). Thus, the management and conservation of the Trinidad Piping Guans may become a necessity.

Captive populations have been highly successful in other Cracid species in terms of both re-introducing individuals to bolster wild populations and in providing a species reserve for a species that has gone extinct in the wild (Cavero & Pratolongo, 2011; Grau et al., 2003; Pereira & Wajntal, 1999). The White-winged Guan (*Penelope albipennis*) has been successfully bred in captivity to enable the re-introduction of individuals to the wild in Peru (Cavero & Pratolongo, 2011). DNA fingerprinting was used to genetically monitor captive populations of Dusky-legged Guans (*Penelope obscura*) and Rusty-margined Guans (*Penelope superciliaris*), to fascilitate the successful re-introduction of individuals to the wild in Brazil to bolster natural populations (Pereira & Wajntal, 1999). In a case study of the Agaloas Curassow *Mitu mitu* by Grau et al. (2003), it was estimated that the population declined from around sixty individuals to being extinct in the wild within a period of thirty years. A subsequent recovery strategy utilising captive birds (n=3 founders) resulted in a small captive population of the species. However, the *Mitu mitu* captive population demonstrates distinctly limited genetic variation in the Cytochrome *b* gene and Control Region of the mitochondrial genome (Grau et al., 2003), which highlights the need to establish breeding populations with a broad genetic ancestry and thus maximise genetic variation in the population for the future health and fitness of the species.

A key feature of any management strategy must be to preserve the genetic integrity of the species, with the aim to preserve the genetic constitution of the present line whilst encouraging intra-specific variation. The adverse effects of inbreeding must be considered within any management strategy and, as with the case of the Agaloas curassow, the negative effects of small founding population size may be significant in both the short term and long term (Grau et al., 2003). Most critically, avian species are susceptible to the effects of decreased fertility and survivability of offspring when inbreeding occurs (Beissinger et al., 2008; Strahl & Grajal, 1991), in combination with the slow reproductive and population recovery rates of these large avian species, genetic recovery may be severely restricted.

Species management will become increasingly important for the Trinidad Piping Guan as populations decline and long term conservation strategies may be needed. The

effects of reintroducing inbred individuals to re-establish or support the existing wild population may be unpredictable due to the negative influence of loss of adaptive ability due to the loss of genetic variation (Grau et al., 2003). Habitat availability is fundamental to the possibility of reintroduction, where problems continue with habitat loss and fragmentation, reintroduction of individuals to the wild will continue to raise issues of resource management and conflict with both other species and humans. A guarantee of protection is the minimum requirement for re-introduced individuals of a critically endangered species.

Small Populations, Founders and Bottlenecks

Genetic diversity is a core concept linked to both biological and ecological health, central to the concept is the theory that diversity is key to a populations ability to survive and thrive. Loss of genetic variation is a direct function of a populations size, growth rate and timing, and duration of past population events (Chan, Anderson, & Hadly, 2006). The genetic variation or diversity within a species has been linked directly with the risk of extinction faced by that species, critically so with small populations (Frankham, 2008). The reduction of variation caused by small population effects has notable negative influences on the species and individuals' fitness, ability to adapt to environmental or demographic change and the ability to respond or resist to physical challenges such as disease. Founder populations are usually restricted both in terms of population size and the available genetic variation. Genetic variation fuels the evolution of species through response to environmental change and adaptive potential of the population, therefore being critical to small populations such as island endemic species (Frankham, 1997).

Anthropogenic disturbances such as harvesting and habitat destruction directly contribute to population reductions and loss of genetic diversity within wild populations, the consequences of which may be a genetic bottleneck. Bottlenecks are of great concern to the conservation of a species; inbreeding, reduced fitness and survival and genetic drift can have a significant effect on a species and result in its failure to thrive and recover. The impact of bottlenecks is largely dependent on the

temporal persistence and severity of the loss within the population or species. The sources of bottleneck events may inform the effect it has on the species and its ability to recover from the event; stochastic events or catastrophes have immediate influences but may allow fast recovery, however long term anthropogenic events (overexploitation or hunting) may cause prolonged bottlenecking and restrict recovery and growth. Depending on the species, its recovery may be determined by its life history traits and its ability to counter mortality rates with reproduction rates. The Trinidad Piping Guan has significant disadvantages in its life history traits, in that it has low rates of reproduction, long generation time and large biomass, all of which are factors known to contribute to increased extinction risk in small populations (Frankham, 2005; Terborgh, 1974).

Combinations of genetic, demographic, environmental and catastrophic processes maintain bottleneck levels in populations and significantly limit population recovery (Beissinger et al., 2008). Beissinger et al. (2008), in studies of parrots noted that four processes maintain a bottleneck; inbreeding reduced nesting success, adults failure to nest due to lack of sites, lack of mates and skewed sex ratios, environmental nest failure, and critically reduced survival rates. The full effects of genetic bottlenecks are not easily observed in species with slow generation times and long lifespans, life history traits are key factors in the analysis of genetic variation within a population or species. Weber (2004) noted that a substantial loss of genetic variation generally requires a near extinction or prolonged period of low population abundance, as demonstrated within the mitochondrial genome of Guadalupe Fur Seals. The intensity of population bottlenecks and the duration of the bottleneck period is directly linked to genetic diversity and evolutionary potential in the contemporary population (Beissinger et al., 2008; England et al., 2003).

Inbreeding as a result of population bottlenecks, depending on the severity and duration of the population loss, is often expressed in a general reduction in the fitness of the offspring and commonly has a significant impact on the fertility of the species. Reduction in fitness or inbreeding depression is typically observed as a reduction in heterozygosity and an increase in genetic load (deleterious or rare alleles) (Weber,

2004). The purging of deleterious recessive alleles is also recognised as a consequence of population reductions, as a result of the loss of heterozygosity a function which may minimise reductions in fitness caused by inbreeding and genetic loss (Bouzat, 2010; England et al., 2003; Weber, 2004).

The negative effects of inbreeding both genetically and physiologically become more acute as the population size decreases and the duration of the bottleneck increases (Beissinger et al., 2008; Frankham, 2008). Delays in population recovery can result in an increase in genetic load of deleterious alleles and lethal recessive alleles, resulting in loss of fitness. Immigration from outside populations of the same species (geneflow) is the main source of immediate relief from bottleneck events, however where immigration is not a possibility variation will recover slowly through the natural processes of mutation (Keller et al., 2001). In research into the genetic recovery of Song Sparrows (*Melospiza melodia*) in Canada, Keller et al. (2001) demonstrated the allelic decline of the species through the population bottleneck (loss of 42.5% in a single winter) with rapid immigration induced recovery observed over three years to pre-bottleneck levels. However, Keller et al. acknowledge that isolated natural populations such as island endemic species are an exception to this type of recovery strategy.

Genetic diversity is a function of mutation rate and effective population size, therefore within small populations genetic drift maintains the low level of genetic diversity and haplotype variation (Fauvelot, Bernardi, & Planes, 2003). Variation in the mutation rate between species, along with other life history traits such as generation time, may influence the recovery of genetic variation in a small or bottleneck population. In a study on the recovering population of Whooping Cranes (*Grus americana*), Glenn, Stephan, & Braun (1999) noted that mitochondrial DNA haplotype variation was significantly lower post-bottleneck, requiring additional management efforts to preserve genetic diversity. Similar losses in mitochondrial haplotypes observed post multiple bottleneck events in Northern Elephant Seals represent a substantial decline in mitochondrial genetic diversity as a consequence of demographic events (Hoelzel, Fleischer, Campagna, Le Boeuf, & Alvord, 2002; Weber et al., 2000).

Extinction is potentially the ultimate consequence of population bottlenecks, however it is largely dependent upon the status of the population prior to the effects of the bottleneck; the number of founding lineages, time since population foundation, generation time and life history traits, biogeographic constraints and the degree of population density regulation. Importantly the current genetic structure of a population demonstrates evidence of past genetic and demographic events as well as the evolutionary process of the species. It is important therefore to consider historical as well as contemporary demographics when analysing the genetic structure of a population or species.

Islands, Inbreeding & Extinction

The isolation of species in island ecosystems has an important role in species diversification, speciation of immigrants result from localised adaptation to the island environment as they exploit ecological resources and niches in the absence of competitors (Biber, 2002). The diversity of species in a given environment is a dynamic balance of the rates of speciation and immigration against the rates of extinction and emigration. Island populations are subject to stronger influences of extinction processes; the natural rise and fall of species is biased by the limitations of species movements within a confined environment. Land-bridge islands, such as Trinidad, traditionally exhibit higher degrees of diversity than oceanic island systems, due to historical connections to the mainland and isolation of mainland species within the island system. Oceanic islands by contrast are critically limited by the immigration of species by colonisation of widely dispersing species (Emerson, 2002).

Species diversity in island systems is a consequence of the diversification of founder populations into adapted species from environmental niches (adaptive radiation), colonisation of neighbour islands or other land masses, diversification of founders to adapted species due to stochastic events (lava, erosion, floods etc.) and speciation through bottleneck and founder events (Emerson, 2002). Land-bridge islands, such as Trinidad, connected to the adjacent mainland during periods of lowered sea levels may be more diverse due to their historical connections to mainland species and subsequent

effects of speciation, by contrast colonisation of oceanic islands is less successful. Typically colonisation events are rare and stochastic, in species such as the Trinidad Piping Guan which are sedentary and have limited dispersal strategies this isolation is often unique and further colonisation events are unlikely. As such the restriction of geneflow in relation to oceanic barriers is a major contributor to speciation of island forms. Island endemic species are constrained by their environmental niche, and are therefore susceptible to environmental change, disease and introduced species.

Frankham (1997) proposes that genetic variation in island species is determined by the relationships of island size, immigration rate, mortality rate and mutation. Genetic variation in an isolated population, may only be increased through mutation, where immigration is not possible, therefore the adaptive potential and genetic variation within an island endemic species such as the Trinidad Piping Guan is restricted to that of the founder population. Adaptive selection may have significant influence on small populations which suffer anthropological and environmental pressures, the genetic potential for adaptation being critical for survival in new environments and niches.

In island populations it is expected that increased inbreeding due to bottlenecks at foundation and through low average population sizes will be observed (Frankham, 1995 & 1998, Habel and Zachos, 2013). Inbreeding results as a direct consequence of restricted population size over extended periods, therefore is a significant risk in island endemic species. In naturally outbreeding populations inbreeding produces genetic drift, inbreeding depression and the reduction of fitness within a species, thereby increasing extinction risk. The rate of loss of genetic diversity, may be directly linked to the genetically effective population size and the subsequent inbreeding coefficient (Frankham, 2005; Habel & Zachos, 2013). Evidence from natural and laboratory populations suggests inbreeding is directly associated with extinction risk (Frankham, 1998). Small and declining populations, where inbreeding and loss of genetic diversity is unavoidable, present an elevated risk of extinction (Frankham, 2005).

Extinction risk is considered to be higher in island populations than in mainland populations (Frankham, 2008). The major causes of extinctions in island species are predominantly anthropogenic; over exploitation, habitat destruction, introduced species and introduced diseases however, island populations are particularly susceptible to demographic and environmental stochasticity. Since 1600, 75% of recorded extinctions have been island species, Frankham (1997, 1998) reported that only 20% of bird species inhabit islands, however 90% of extinct birds have been island species (based on records since 1600). Island endemism has a strong association with extinction proneness, the vulnerability of a species to extinction may be linked with its ability to disperse and adapt to new environments and situations. Island environments have restricted dispersal availability, with species therefore being restricted to an evolutionary niche or defined habitats and vulnerable to habitat loss and hunting pressures. Catastrophic events have the potential to induce traumatic population declines or complete extinction events in endemic island species, due to a lack of reserve populations to re-establish the species. In addition non-traumatic declines where mortality rates out weigh reproduction rates produce slow, long term population extinctions in island species.

Endemic island birds have generally higher extinction rates than non-endemic island populations (also evidenced in reptiles) (Blackburn, Petchey, Cassey, & Gaston, 2005; Frankham, 1997). Endemic island species, such as the Trinidad piping Guan, are often amongst those most at risk of biogeographic extinction, and are therefore often of greatest concern for conservation. The limitations of island endemism mean these species are more vulnerable to anthropogenic pressures and biological insults, the island environment and within it habitat specialisation mean species such as the Trinidad Piping Guan are at higher risk of extinction. The Trinidad Piping Guan is also susceptible to many of the deterministic and stochastic factors associated with the processes of biological extinction including habitat loss, over exploitation or introduced species, and demographic, environmental, genetic or catastrophic events respectively (Frankham, 2005).

Biber (2002) suggests that human colonisation of islands has resulted in substantial impacts on native avian populations, most notably increased extinction levels and re-distributions of species due to the effects of human interactions with the landscape over time. The Trinidad Piping Guan is endemic to the island of Trinidad, and with anthropogenic loss of habitat the species is now confined to the north-eastern regions of Trinidad (Hayes, Sanasie, et al., 2009). Environmental stability is critical to complex species communities and species diversity (Brooks et al., 2005). Trends in habitat loss are often unsustainable, deforestation being significant in the decline of the Trinidad Piping Guan through the loss of primary habitat (Hayes, Sanasie, et al., 2009). The regeneration of such tropical forest habitats may take hundreds of years, or may never return to their original state (Brooks et al., 2005).

Sedentary species such as the Galliforme avians, are less resilient to the broader effects of habitat fragmentation, compared to those with stronger dispersal abilities. Bush et al. 2010 noted that those species with better dispersal strategies displayed improved genetic structure in relation to fragmentation patterns and spatial scales. Guans and Curassows are recorded as constituting one of the highest biomasses in hunting (Brooks, 1999; Strahl & Grajal, 1991), being susceptible to both subsistence hunting and sport hunting pressures. Cracids are widely accepted as a major protein source for Latin Americans and Amerindians, and hunting pressures are directly linked to human population densities and habitat loss (Brooks, 1999). Like other *Pipile* species the Trinidad Piping Guan is also subject to strong pressures of hunting by local human populations, however little information is available on the extent of hunting on Trinidad (Waylen, 2005). Vulnerability of the Trinidad Piping Guan is enhanced by its island endemic status, with no opportunity for immigration the species is reliant on reproduction to maintain population levels therefore hunting pressures exacerbate greatly the risks of extinction for this species.

Social structure among Trinidad Piping Guans is thought to revolve around a social group within an established territory, Hayes (2009) identified a single group which may have occupied the study site since 1963 or longer, however the extent of interactions with other groups is unknown. Limited territory size may be influential in

the extent of inbreeding occurring within the Trinidad Piping Guan population, home ranges were estimated to be around 19ha in the study group (Hayes, 2009). The population size of Neotropical frugivores is known to fluctuate with the seasonality of fruit blooms (Brooks et al., 2005) and therefore the availability of resources is critical to maintaining a stable population in these species. Terborgh (1974), suggests avians with large body mass for their trophic class, those that forage or nest on the ground, and those which demonstrate slow reproductive life history traits, are prone to extinction. Large avian species are more susceptible to the pressures of human influence such as hunting and habitat loss, as is the case with the Trinidad Piping Guan and other Cracids (Brooks, 1999; Brooks et al., 2005). Additionally, confinement to small areas, such as islands, increase the pressures of metabolic demand from large avian species in response to competition for resources, critically limiting population growth.

Island species such as the Trinidad Piping Guan are expected to demonstrate lower genetic variation due to founder effects and population restrictions through demographic and stochastic mechanisms, when compared to an equivalent mainland species (Frankham, 1997). Genetic analysis of endemic island species such as the Trinidad Piping Guan may elucidate both historical and contemporary trends in the population structure of a species, which may subsequently inform conservation strategies in the future.

Molecular Markers in Avian Genome Analysis

Mitochondrial DNA Markers

The genetic diversity within a species is influenced by many factors including species life history traits, mating systems, natural selection, population structure and periodic declines. The mitochondrial genome is a small, circular molecule approximately 16-17kb in length in most vertebrate species, with a relatively fast evolutionary rate, which makes it a good marker of both inter-specific and intra-specific genetic

variation. The rate of evolution for the mitochondrial genome has been determined as 0.02 substitutions per base per million year, approximately 10x that of nuclear DNA (Brown et al., 1979). Recent research demonstrates evolutionary substitution rates differ among codon sites, between genes and vary greatly among species, affecting the phylogenetic signal. Patton and Baker (2006) determined that genes with slower rates of evolution, such as COI, ND2 and ND5 produce more robust phylogenetic reconstructions. Mitochondrial DNA, being non-recombining, is often treated as nearly neutral evolutionarily and in this respect the diversity of a population should reflect the population size (Bazin et al., 2006).

Mitochondrial markers are ideally suited to the needs of conservation genetics in many respects, most prominently in the ability to recover mitochondrial DNA from environmentally degraded samples and biologically poor sources of DNA. As previously discussed, the mitochondrial genome is present within biological materials in much higher copy numbers than nuclear genome copies, and in addition is a much smaller and robust molecular structure, therefore it is more easily recovered and amplified from non-human biological materials (Hsieh et al., 2001). It has been noted that the recovery of mitochondrial markers, such as the Cytochrome *b* or Cytochrome c oxidase I genes, are relatively easily achieved with modern PCR techniques, even from highly degraded DNA, ancient DNA and degraded mixtures (Hsieh et al., 2001; Tobe & Linacre, 2009)

Cytochrome *b* and COI are standard species diagnostic markers utilised in a wide scope of analyses including phylogeography, population genetics, wildlife forensics and species identification (Ogden, Dawnay, & McEwing, 2009; Tobe & Linacre, 2010; Wilson-Wilde, Norman, Robertson, Sarre, & Georges, 2010). Both genes have received considerable attention in the last few decades, and can be used with confidence in many applications however reference data for less common species is limited. Kocher et al. in 1989, recognised the broad application of these genes, with prospects for the identification of genetic diversity in natural populations, the ability to track genetic changes through time using historical samples as well as enlightening the complexities of genetic structure and function (Kocher et al., 1989).

Multi-gene mitochondrial analysis has been widely used in the analysis of variation both within and between species of Cracidae (Hosner et al., 2016; Ibarguchi, 1999; Pereira et al., 2002; Pereira et al., 2004) and other large neotropical avians (Lutz et al., 2013; Tavares, Gonçalves, Miyaki, & Baker, 2011). The use of large DNA data sets and combined nuclear and mitochondrial gene sequences is well established as a technique to analyse relationships and evolutionary histories of families of avians including Cracids (Hosner et al., 2016; Paton & Baker, 2006; Pereira et al., 2002). The congruence between independent gene trees and combined gene trees may infer an increased confidence in the results of phylogenetic and evolutionary relationship hypotheses. The application of larger sequence data sets reduces the amount of stochastic error in substitutions which can be inferred by bootstrap support for nodes or likelihood distances at various depths within the trees (Pereira et al., 2002). Multi-gene analysis however may give rise to confounding information, and may be indicative of gene evolution rather than providing confidence in species evolutionary relationships (gene tree as opposed to species tree); heterogeneity among the gene data may therefore impede phylogenetic inference (Pereira et al., 2002).

The use of multiple genes from the mitochondrial genome has a disadvantage in that they share a single linkage, inherited as a single unit therefore yielding a single mito-genomic tree. Additionally phylogenies inferred from single genes may be misrepresentative of the entire molecule tree (Zink & Barrowclough, 2008). The limitations of mitochondrial multi-gene approaches have thus recommended the inclusion of unlinked nuclear DNA loci, as a means of improving confidence in the phylogenetic topology (Zink & Barrowclough, 2008). Commonly used are nuclear gene intron sequences; nuclear mutation rates being slower than that of mtDNA which is advantageous in defining older phylogenetic relationships but may however hide more recent divergences (Prychitko & Moore, 1997). In this instance it is advantageous to include both nuclear and mitochondrial genes in a multigene approach to the analysis of genetic structure, particularly within small or inbred populations. Mitochondrial DNA has demonstrated conflicting data when compared to nuclear markers of variation. Disparity in the evolution of the genomes results in conflicting phylogenetic patterns and evolutionary relationships (Hosner et al., 2016; Pereira et al., 2002). The mitochondrial DNA genome demonstrates lower levels of variation

than the nuclear genome, but recovers more quickly than the nuclear genome from genetic loss (Fay & Wu, 1996).

Cytochrome b

Cytochrome *b* (Cyt B) is a mitochondrially encoded gene involved in the respiratory chain complex III, used in electron transport and the generation of ATP, as such its structure and function are largely constrained. The Cytochrome *b* gene has been shown to have significant utility in the analysis of species identification, evolutionary and phylogenetic relationships, genetic variation and many other ecological analyses. Its utility and use over a prolonged period mean there are significant amounts of data available for this gene, including its use in avian species analysis. Cytochrome *b* has been widely used in the identification of species for many years, including avian speciation (An, Lee, Min, Lee, & Lee, 2007; Kocher et al., 1989; a. Linacre et al., 2011; Parson, Pegoraro, Niederstätter, Föger, & Steinlechner, 2000; Verma & Singh, 2003). In addition it has been widely used as a molecular marker in the analysis of phylogenetic relationships and evolutionary analysis (Armsrtong et al., 2001; Johns & Avise, 1998; Prychitko & Moore, 2000; Tobe, Kitchener, & Linacre, 2010).

Lee et al. (2008) identified interspecific diversity ranges from 5.65 to 27.68% between avian species in analysis of a short region of the Cytochrome *b* gene. The broad range of species included two Galliformes (*Gallus gallus* and *Phasianus colchicus*), however no Cracids were included in the analysis. Intra-specific analysis of avian species, by Lee et al. (2008) determined diversity within species of between 0-6.78% using the Cytochrome *b* gene.

Many comparisons have been made between the use of Cytochrome *b* and other mitochondrial genes including the Cytochrome c oxidase subunit 1 gene (Meiklejohn et al., 2014; Paton & Baker, 2006; Tobe et al., 2010), and conflicts between such genes have been well documented. In the analysis of Galliforme relationships the CytB gene did not resolve a well supported phylogeny as expected, other mitochondrial genes

(including COI and ND5) performing much better, with higher statistical support for nodes (Meiklejohn et al., 2014). Tobe, Kitchner and Linacre (2010) have analysed in detail the utility of the CytB and COI genes in the reconstruction of mammal phylogenies. They determined both mitochondrial genes have utility in the analysis of variation; CytB demonstrated higher congruence with taxonomy and a higher percentage of variable sites, however in analysis of intra-specific variation, COI showed similar levels of variation in comparison to CytB (Tobe et al., 2010).

Comparisons of the Cytochrome *b* gene with the nuclear gene Beta-Fibrinogen intron 7, determined that CytB evolves at a faster rate than the nuclear intron (Moyle, 2004; Prychitko & Moore, 2000), which is beneficial in the analysis of closely related avian species however saturation effects associated with high substitution rates mitigate resolution of deeper relationships within the Barbets (Moyle, 2004). Prychitko and Moore (2000) suggest that in the phylogenetic analysis of Woodpeckers, Cytochrome *b* has limited utility; two thirds of the gene lack phylogenetic informity, being highly homologous among species, and the high substitution rates of the gene obscure the evolutionary history of deeper relationships within the family.

Cytochrome c oxidase subunit 1

The Cytochrome c oxidase enzyme consists of thirteen subunits, ten encoded on the nuclear genome and three mitochondrially encoded. Cytochrome c oxidase is part of the electron transport chain in the process of ATP synthesis, as such its structure and function is constrained. The COI gene is currently one of the most common genes used in conservation applications, the Barcode of Life Database (BOLD) has made a significant contribution to research and provided an important resource for taxonomy and species identification. COI barcoding has provided a global standardised protocol for delineating species and analysing species relationships; with the application of universal primer sets for the amplification of a 650bp standard region of the 5'end of COI (Hajibabaei, Singer, Hebert, & Hickey, 2007; Hebert, Stoeckle, Zemlak, & Francis, 2004). However there is currently only a limited quality assurance, as with any public database, that the sequence data is representative of the current species wild

type, much of the sequence information being sourced from museum and zoological collections (Valentini, Pompanon, & Taberlet, 2009).

Species boundaries are commonly inferred through mitochondrial DNA sequence variation, with most studies suggesting a boundary of approximately 2% in vertebrate taxa is indicative of inter-species relationships, with intra-specific variation below this level. Hebert et al. (2004) postulated a divergence threshold based in the mean intra- and inter- specific variation of North American birds, however these thresholds have been much debated. The standard species threshold of tenfold the mean intra-specific variation for the group, suggested a threshold of 2.7% in North American birds, delineating 90% of species sequenced (n=260) (Hebert et al., 2004), higher than traditionally held species thresholds for general mitochondrial genes. Within species variation for the barcoding region of COI averaged 0.27% for North American birds, however Hebert et al. (2004) recognised that this may not be representative of global avifauna particularly in biodiverse tropical regions.

The utility of the COI barcode for taxonomic and species identification has been widely discussed, the gene region has shown strong promise for species identification in many taxa however, a number of studies have demonstrated severe limitations of the technique including paraphyletic sequences and congener homology (Elias et al., 2007; Craig Moritz & Cicero, 2004; Spooner, 2009; Will & Rubinoff, 2004). Tobe, Kitchener and Linacre (2009, 2010) directly compared the utility of the COI gene as a marker of species identification for mammal taxa, to the Cytochrome *b* gene, and demonstrated a higher degree of conservation within the COI gene between species (43.7% and 22.4% respectively). Intra-specific variation within mammal species was lower for the COI gene than CytB, 2-12% and 3-18% respectively, in cows, dogs and humans (Tobe, Kitchener, & Linacre, 2009). In avian species the intra-specific distance was found to be much lower; 0.24% in COI and 0.74% in Cytochrome b, compared to 0.48% in the 16S rRNA gene (Aliabadian et al. 2009). These results were similar to those observed in the initial analysis of the barcoding gene by Hebert et al. (2004), on North American avians (0.27% intra-specific variation).

The COI barcoding region has been much critiqued in recent years regarding the robustness, methodology and utility in ecology and phylogenetics. The COI barcoding region has been criticised for its failure to record an accurate phylogeny in complex and closely related species groups (Elias et al., 2007; Kwong, Srivathsan, & Meier, 2012; Craig Moritz & Cicero, 2004; Spooner, 2009). Additionally, issues surrounding the fitness of the region to test a clear hypothesis, inappropriate use of tree models and statistical analysis have all been raised as issues surrounding research into the COI barcoding gene region (Roe & Sperling, 2007; Tavares et al., 2011; Will & Rubinoff, 2004).

Control Region

The mitochondrial Control Region is a non-coding segment of the genome which varies greatly both between and within species. It contains a number of structural genomic features including; transcriptional promoters for the L-strand and H-strand (LSP and HSP), Heavy strand replication Origin (O_H), displacement loop (D-loop) and a number of conserved sequence motifs (CSB) (Ruokonen & Kvist, 2002). In most avian species the Control Region spans the region between the $tRNA^{Glu}$ and $tRNA^{Phe}$ genes, with the exception of some avian families which exhibit gene re-arrangements in this segment of the genome, see Chapter 3 Figure 3.2 (Marshall & Baker, 1997; Mindell et al., 1998). Domains within the Control Region are typically defined by the conserved motifs within each region, see Chapter 3 Figure 3.6. Domain I is associated with short Termination Associated Sequences (TAS), and a high adenine composition. The central domain (II) contains the displacement loop or D-loop and is associated with Conserved Sequence Blocks (CSB) some of which are class specific such as the avian Bird-Box CSB (Ruokonen & Kvist, 2002). Domain III is characterised by the presence of the Origin of H strand replication (O_H), the CSB-1 and transcription promotors LSP and HSP (in avians present as a single bi-directional promotor). In Galliformes and Anseriformes it has been noted that the CSB-1 structure is repeated in Domain I with a CSB-1-like region (Desjardins & Morais, 1990). The third domain typically demonstrates the highest degree of variation, predominantly due to repeat structures, and the central domain retains a higher degree of conservation (Marshall & Baker, 1997).

The avian Control region demonstrates higher rates of substitutions and higher incidence of indels in Domains I and III, compared to the more conserved Domain II (Ruokonen & Kvist, 2002). However conserved sequence motifs are observed throughout the region in all three domains; number, localisation, spacing and degree of conservation of these motifs has been shown to vary considerably among and within Class of vertebrates (Marshall & Baker, 1997; Ruokonen & Kvist, 2002). Wink (2006) estimated that the D-loop of the control region is typically 4 to 6 times more variable than other protein-coding mitochondrial genes. Szalanski, Church, Oates, Bischof, & Powers (2000) estimated Genetic variation in the third domain of the Dloop in Wild Turkeys (*Meleagris gallopavo*) to be 0.7% on average, with 16 variable sites in 502bp of DNA sequence. The control region varies greatly both in length and structure between species; significant differences in structure arising from simple substitutions, indels, large insertion and deletions, duplications and variation in tandem repeat copy number (Marshall & Baker, 1997).

The Control Region is commonly used as a marker for the analysis of population structure and individual variation due to the highly variable regions in Domains I and III. Typically the Control region of the mitochondrial genome has a much higher rate of mutation than other gene regions of the mitochondrial genome; in humans the Control region average mutation rate is 14-17.5% per million years compared to 2% per million years for mitochondrial protein-coding genes in mammals (Hewitt, 2000). Ruokonen and Kvist (2002) suggest the control region may be the most rapidly evolving part of the mitochondrial genome, with estimated mutation rates between 0.54 and 17.30% per million years in the hyper-variable regions. The higher rates of substitutions observed in the hyper-variable regions provide improved resolution of phylogenetic relationships (Wink, 2006). With more distant comparison between orders or families show significantly elevated levels of variation within the control region making it ideal for analysis between closely related species (Marshall & Baker, 1997).

Due to the highly variable nature of the Control Region alignment of sequences is challenging, depending upon the taxonomic level of analysis (Angleby & Savolainen,

2005; Pereira, Grau, & Wajntal, 2004; Wink, 2006), and may result in ambiguous alignments. High mutation rates resulting in large numbers of substitutions and the occurrence of tandem repeats and indels confound homology of alignments and may present problems in phylogenetic applications (Pereira, Grau, & Wajntal, 2004; Ruokonen & Kvist, 2002). The hyper-variability of nucleotides within this region may also result in saturation issues and bias substitution patterns. Conserved sequence blocks (CSB) and Termination associated sequences (TAS) demonstrate a much higher degree of conservation within genus. The variation in substitution rates throughout the different domains, features and blocks of the Control Region mean that calibration of the region as a whole is difficult, in addition to saturation issues at hypervariable sites (Lucchini & Randi, 1998).

The high degree of variability in some regions of the Control Region has resulted in its common use as a molecular marker in the analysis of aDNA or museum specimens, where degradation critically limits the size of DNA product attainable. Using a short, but highly variable marker from the mitochondrial genome is highly beneficial in the study of historical populations through the analysis of ancient DNA (Hoelzel et al., 2002; Weber, 2004). For example the use of the Control Region in the Rock Partridge (*Alectoris graeca*) demonstrated evidence of historical population and phylogeographic structure. Lucchini and Randi (1998) utilised the higher degree of variability of the Control Region to determine the genetic relationships of the species, in light of glacial historical effects. An in-depth study of the mitochondrial Control region in Cracids, by Pereira, Grau & Wajntal (2004), analysed twenty-seven Cracid species spanning eleven genera to identify conserved sequence motifs, estimate substitution rates and compare rates of substitution to other commonly used protein-coding genes. Comparison with the ND2 and CytB genes indicates that the complete Control Region in Cracids evolves at a slower rate than these protein coding genes, at rates consistent with those observed in other Galliformes (0.16% per site per million years). Randi et al. (2001) reported higher rates of evolution in the Control Region compared to CytB, in closely related species of the genus *Lophura* (Galliformes).

NADH dehydrogenase subunit 2

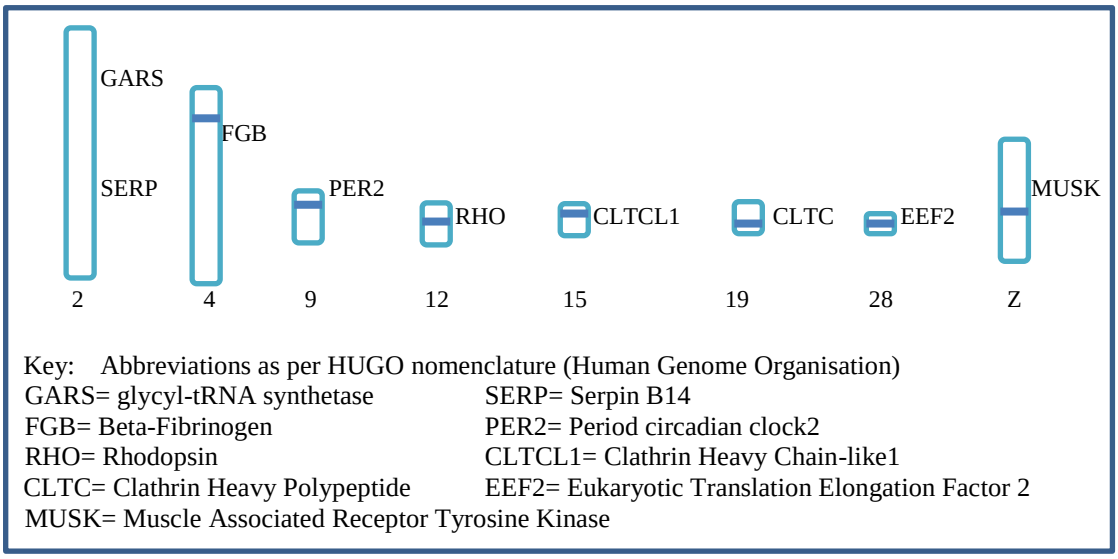
NADH dehydrogenase 2 (ND2) is a mitochondrially encoded gene, which is a subunit of the NADH dehydrogenase (ubiquinone) complex I involved in the electron transport chain. The ND2 gene is regarded as one of the most variable genes in the avian mitochondrial genome after ATPase8 and ND6. ATPase 8 is the shortest gene in the mitochondrial genome and so it has limited utility and ND6 demonstrates hyper-variable qualities between species making it difficult to resolve (Meiklejohn et al., 2014; Sorenson, 2003). In analysis of the Tetreoninae (of the Phasianidae) clade of Galliformes ND2 was significant in resolving relationships among the genera, supported by high bootstap values, however at a more basal level the gene was less successful at resolving Galliforme divergences (Dimcheff et al., 2002). The application of additional genes to the analysis of the Tetraoninae and older Galliforme divergences by Dimcheff et al. (2002) improved the resolution of more basal separations of Galliforme families.

Nuclear Genome Markers

Nuclear approaches to genetic variation often involve a frequency approach such as the use of microsatellites, however phylogenetic relationships are difficult to infer using this approach. The vertebrate nuclear genome varies greatly in terms of both size and structure between taxa and species. Typically in excess of 1 billion base pairs in length, only approximately 2% of the genome encodes for proteins with over 75% of the genome being extragenic DNA containing many repetitive elements (Wink, 2006). Nuclear genes are highly variable due to the bi-parental inheritance of alleles and the effects of recombination, therefore sequencing of extra-genic regions such as introns have been shown to be robust in phylogenetic analysis, and are now common markers in avian phylogenetics (Wink, 2006).

Nuclear DNA sequencing is increasingly being used to support mitochondrial sequence information to estimate relationships among taxa (García-Moreno, Sorenson, & Mindell, 2003; Morgan-Richards et al., 2008; Pereira et al., 2002; Zink & Barrowclough, 2008). The use of nuclear sequence information provides an

independent estimate of relationships and evolutionary history, where mitochondrial genes are linked by their common genome. In addition the differences in the evolutionary rates between mitochondrial and nuclear gene sequences allows for a more robust analysis of the evolutionary relationships. Nuclear sequence analysis of the avian class has utilised several regions of the genome (see Figure 4.1); exon analysis for example RAG1, CLTC and ZENK (Chojnowski et al., 2008; Miller et al., 2008) and intron analysis including the OTC, Beta-Fibrinogen and Ovomucoid genes (Crowe et al., 2006; Morgan-Richards et al., 2008; Nishibori, Shimogiri, Hayashi, & Yasue, 2005; Prychitko & Moore, 1997, 2000).



Key: Abbreviations as per HUGO nomenclature (Human Genome Organisation)

GARS= glycyl-tRNA synthetase	SERP= Serpin B14
FGB= Beta-Fibrinogen	PER2= Period circadian clock2
RHO= Rhodopsin	CLTCL1= Clathrin Heavy Chain-like1
CLTC= Clathrin Heavy Polypeptide	EEF2= Eukaryotic Translation Elongation Factor 2
MUSK= Muscle Associated Receptor Tyrosine Kinase	

Figure 4.1: Karyotype diagram of loci used in this study, based on the Chicken genome (Gallus gallus), adapted from Kimball et al. (2009).

Analysis by Kimball et al. (2009) identified numerous introns across the avian genome which demonstrated utility in a broad range of genetic analyses, including phylogenetics. Nuclear Intron sequences have been found to have utility in phylogenetic analysis as they evolve more rapidly than other nuclear gene regions (coding exons and untranslated regions) and there is a lack of constraint on the sequence evolution and enhancing their utility in resolving rapid radiations (Armsrtong et al., 2001; Jarvis et al., 2014; Wang et al., 2013). Introns have regulatory functions, therefore there is less constraint on the evolution of these regions, enabling high variability in the sequence within and between species (Prychitko & Moore,

2000). Introns demonstrate slower substitution rates than mitochondrial genes, meaning they enable reliable phylogenetic analysis at deeper taxonomic relationships (Moyle, 2004). Analysis of Galliforme phylogenetic relationships using an intron from the Ovomucoid gene demonstrated significant levels of parsimony informative sites comparable with that of the Cytochrome *b* gene (Armsrtong et al., 2001). The use of the Ovomucoid intron G in analysis by Crowe et al (2006) concluded that the addition of nuclear DNA sequencing to a combined mtDNA data set (CytB, ND2, CR and 12S) had no effect on phylogenetic structure but improved node support in analysis. The combination of nuclear and mitochondrial gene regions has been widely reported, and demonstrates a high degree of success in resolving relationships within genus in the context of biogeographical and speciation analysis, including analysis of the Cracidae (Hosner et al., 2016; Miller et al., 2008; Patel et al., 2011; Ribas, Miyaki, & Carcraft, 2009; Ribas, Moyle, Miyaki, & Cracraft, 2007).

Beta-Fibrinogen, Intron 7

The genetic analysis of β-fibrinogen gene (FGB) introns has been widely used in phylogenetic and biogeographical analysis in many species of avians including the *Ramphastidae* Toucans (Lutz et al., 2013; Patel et al., 2011), Barbets (Moyle, 2004) and Woodpeckers (Prychitko & Moore, 1997, 2000). Genetic analysis of this gene region in combination with mitochondrial genes has demonstrated strong links between genetic diversification of congeneric species and the major geological events in the Neotropics (Lutz et al., 2013; Patel et al., 2011; Ribas et al., 2009). The 7th intron of the β-fibrinogen gene (FGB) has been widely used in analysis of species relationships and intra-specific variation in a number of avian species including the Cracidae (Hosner et al., 2016; Pereira et al., 2002) and the Galliforme order (Wang et al., 2013). The slower rate of evolution of the FGB, compared to the Cytochrome *b* gene, enables robust estimates of phylogenies and greater phylogenetic resolution (Moyle, 2004). In comparison of tree topology between FGB and CytB analysis in Woodpeckers, topology was largely congruent with variation being in the branch nodes and lengths suggesting well established evidence of species relationships and utility of these regions in species level phylogenetics (Prychitko & Moore, 2000). Morgan-Richards et al. (2008) examined the β-fibrinogen intron 7 in relation to the reconstruction of large scale avian class phylogenies, in comparison to mitochondrial

reconstructions. The rapid evolution of the FGB intron presents difficulties in the reconstruction of broader avian phylogenies as a result of the limitations of alignment, phylogenies resolving well within genus but performing less well with increased genetic variations in the higher taxonomic classifications (Morgan-Richards et al., 2008).

Serpin B14

Serpins primary function is related to intracellular protein inhibitors in mammals and birds, the Serpin B14 gene is only found in avians (Kaiserman & Bird, 2005). The Serpin B14, intron c has been used in a number of studies of inter-species phylogenetics including the analysis of relationships within the Galliforme order (Wang et al., 2013). Analysis by Wang et al. (2013) suggests that the SerpinB14 has a strong phylogenetic signal, which differs from other genes and gene regions in comparative phylogenetic analyses.

Nuclear DNA Gender Markers

The use of markers on the sex-linked chromosomes has been widely published in avian research, most commonly used is the Chromobox-helicase-DNA binding gene or CHD (Dawson et al., 2001; Fridolfsson & Ellegren, 2000; Fridolfsson & Ellegren, 1999; Lee et al., 2010; Ong & Vellayan, 2008). The use of genes present on the avian W and Z chromosomes allows gender identification through the occurrence different chromosome composition of male and female individuals; the CHD-Z occurring in both male and female individuals and the CHD-W being present in only in females, presenting a bi-allelic profile of WZ in female and ZZ in males.

The CHD gene is highly conserved across the avian class, therefore it represents a useful molecular marker for gender identification in a broad range of avian species (Griffiths et al., 1998). A high degree of sequence evolution between species has resulted in a variation in product size between species, however utility of the marker is not compromised due to this variation (Griffiths et al., 1998). The total size of the

CHD gene on the W and Z chromosomes is approximately 5kb in the chicken, however numerous primers have been developed to amplify sexually indicative markers of around 300bp to 800bp (Chang et al., 2008; Griffiths et al., 1998). Different primer combinations have been used with varying degrees of success between species (Harvey et al., 2006; Lee et al., 2010). Species variation has been observed in several studies, in relation to the amplified fragment sizes of the CHD-Z and CHD-W regions and the success of primer annealing between species (Griffiths et al., 1998; Lee et al., 2010).

Study Objectives

The aim of this study is to establish a clear analysis of genetic variation in the contemporary wild Trinidad Piping Guan population, using both mitochondrial and nuclear genome markers. It is predicted that this analysis will demonstrate evidence of the small contemporary population size and current bottleneck influences, in addition to evidence of the historical effects of population foundation in an isolated island system. The current captive UK population of Piping Guans is used as a model of a known small population system within the genus, for comparison to the wild population. The use of a known small population model, may highlight the level of genetic variation in the wild Trinidad Piping Guan population in the context of small population effects such as inbreeding depression and genetic drift. In addition this study will develop markers for future analysis of genetic variation in this species and provide data for use in future analysis of the species and conspecifics of the genus *Pipile*.

4.2 Experimental Methodology

4.2.1 Genetic Samples

Wild Trinidad Piping Guans The eight Trinidad Piping Guan (*Pipile pipile*) samples imported from the wild in Trinidad were primarily selected for analysis of the intra-specific variation, see Appendix 1 for sample information.

Captive Piping Guan Samples Samples from nine Trinidad Piping Guans held in captivity in the UK were analysed for evidence of genetic variation in a population of known size, see Appendix 1 for sample information. Species360 (formerly the International Species Information Service) records taxonomically identify these individuals as Trinidad Piping Guans, however their species identity is questioned due to the lack of morphological characteristics conforming to the 'Trinidad type'. Initial species tests of Robinson (2011), identified these individuals as *Pipile cumanensis*, which is consistent with morphology. However this analysis is based on a SNP test using a single species-indicative nucleotide position in the mitochondrial ND2 gene and CytB gene. For the purposes of this study the species remain unclassified and samples are referred to as the 'captive Piping Guan population'.

4.2.2 Mitochondrial and Nuclear DNA Sequencing PCR

Primer sets were selected based on previous research or designed using conserved regions of Galliforme species genomes, including the standard Galliforme reference genome *Gallus gallus* X52392 (Desjardins & Morais, 1990).

Mitochondrial DNA Primers

Primer sets were developed for four mitochondrial genes, and overlapping primer sets utilised to ensure complete coverage of the entire gene, see Table 4.1. Three mitochondrial protein coding genes, Cytochrome *b*, Cytochrome c oxidase subunit 1

and NADH dehydrogenase subunit 2, and the non-coding Control Region were selected for full gene analysis.

Table 4.1: Primers used in amplification of the whole Control Region and whole Cytochrome B, Cytochrome Oxidase subunit I and NADH dehydrogenase 2 genes of the mitochondrial genome.

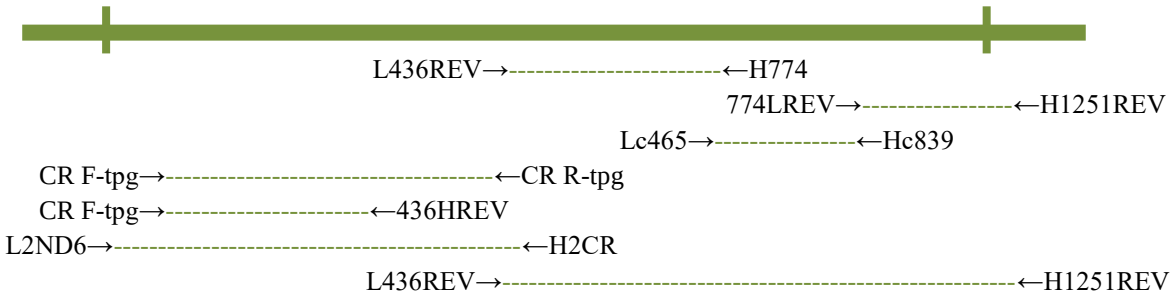
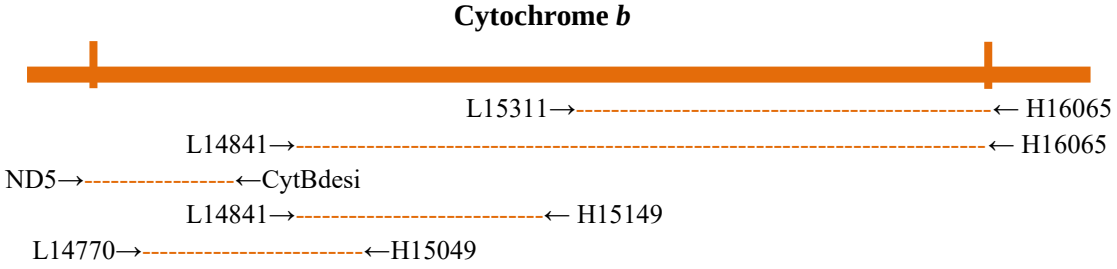
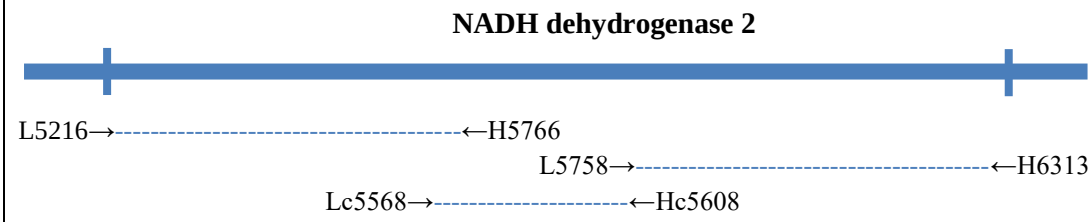
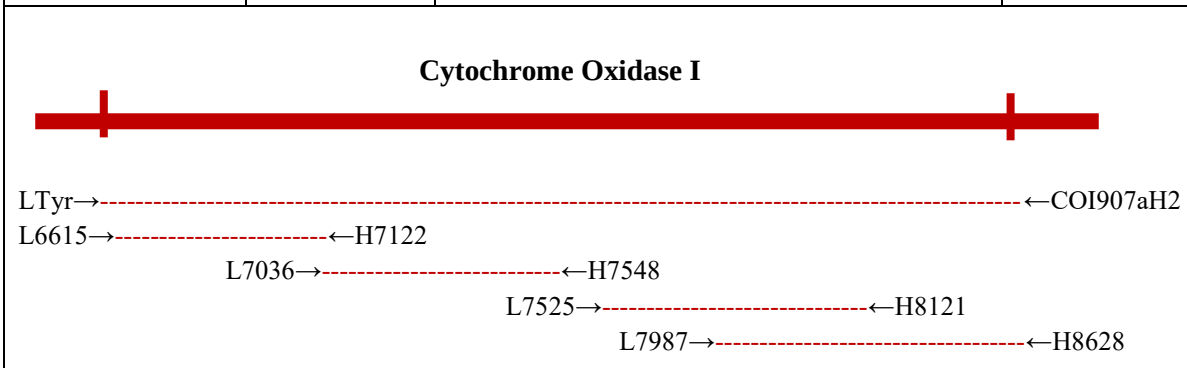
Gene	Primers (5'- 3')* #		Reference
Control Region	L436REV	CCT CAC GAG AAA YCA GCA AC	Sorrenson (2003)
	H774	CCA TAT ACG CCA ACC GTC TC	Sorrenson (2003)
	774LREV	GAG ACG GTT GGC GTA TAT GG	Sorrenson (2003)
	H1251REV	TCT TGG CAT CTT CAG TGC CRT GC	Sorrenson (2003)
	Lc465	GCG CCT CTG GTT CCT ATA TCA	Robinson (2011)
	Hc839	AAG CAT TCA CTA AAT AGC ACC A	Robinson (2011)
	CR F-tpg	CG TTG TTC TCA ACT ATG GGA GC	This research
	CR R-tpg	TGA TAT AAG AAC CAG AGG CGC	This research
	436HREV	GTT GCT GRT TTC TCG TGA GG	Sorrenson (2003)
	L2ND6	TAC CCC ATA ATA CGG AGA GG	Robinson (2011)
	H2CR	GAG ATG TCC AGA CCT AGA TT	Robinson (2011)
Control Region  			
Cytochrome *b*	L15311	CTC CCA TGA GGC CAA ATA TC	Kimball (1999)
	H16065	TTC AGT TTT TGG TTT ACA AGA C	
	L14841	AAA AAG CTT CCA TCC AAC ATC TCA GCA TGA TGA AA	Kocher (1989)
	H15149	AAA CTG CAG CCC CTC AGA ATG ATA TTT GTC CTC A	Kocher (1989)
	ND5	TAG CTA GGA TCT TTC GCC CT	Robinson (2011)
	CytBdesi	TGT AGG TTG CGG ATT AGC CAG C	Robinson (2011)
	L14770	TAG GNC CNG ARG GNY TNG C	Sorenson (2003)
	H15049	GTR TCN GCD GTR TAR TGY ATD GC	Sorenson (2003)
Cytochrome *b*  			

Table 4.1 continued.

Gene	Primers (5'- 3')* #		Reference
NADH dehydrogenase 2	L5216	GGC CCA TAC CCC GRA AAT G	Sorrenson (2003)
	H5766	RGA KGA GAA RGC YAG GAT YTT KCG	Sorrenson (2003)
	L5758	GGN GGN TGA ATR GGN YTN AAY CAR AC	Sorrenson (2003)
	H6313	ACT CTT RTT TAA GGC TTT GAA GGC	Sorrenson (2003)
	Lc5568	AGG ATC GCA CCT AAC CAC TGC C	Robinson (2011)
	Hc5608	GTA GGC TGA GTT TGG GGC TA	Robinson (2011)
			
Cytochrome Oxidase subunit 1	LTyr	TGT AAA AAG GWC TAC AGC CTA ACG C	Taveres (2008)
	COI907aH2	GTR GCN GAY GTR AAR TAT GCT CG	Taveres (2008)
	L6615	CCY CTG TAA AAA GGW CTA CAG CC	Sorenson (2003)
	H7122	ATN GTD GTR ATR AAR TTR ATD GCH CC	Sorenson (2003)
	L7036	GGN ACN GGN TGA ACH GTN TAY CC	Sorenson (2003)
	H7548	GTD GCN GAN GTR AAR TAD GCT CG	Sorenson (2003)
	L7525	GTN TGR GCH CAY CAY ATR TTY AC	Sorenson (2003)
	H8121	GGG CAG CCR TGR ATT CAY TC	Sorenson (2003)
	L7987	TCH GAY TAY CCW GAY GCN TAY AC	Sorenson (2003)
	H8628	TCR TAG STT CAR TAT CAY TGR TGN CC	Sorenson (2003)
			

* standard nucleotide coding system for degenerate base codes
# nomenclature of primer names may not be consistent with nucleotide position on the avian genome (primers designed in mammalian research

Nuclear Marker Primers

Primer sets for a total of nine nuclear gene regions were tested, as outlined in Figure 4.2, however only the Beta-Fibrinogen and SerpinB14 primers provided sufficient

amplifications for analysis. Primer Sets for the Beta-Fibrinogen intron 7 and SerpinB14 intron c gene regions are provided in Table 4.3, all other primer set information is outlined in Appendix 2.

Table 4.3: Primers used in amplification of the Beta-fibrinogen intron 7 and Serpin B14 intron c, regions of the nuclear genome.

Gene	**Primers (5'- 3')* #**		**Product Size (bp)**	**Reference**
FGB-intron7	FIB-BI7L FIB-BI7Uc FIB-U2 FIB-L2	TCCCCAGTAGTATCTGCCATTAGGGTT GGAGAAAACAGGACAATGACAATTCAC GTAACCCATAATGGGTCCTGAG CTTCTGAGTAGGCAGAACTT	~900	Prychitko & Moore, (1997)
SERP-intron c	SERP-F SERP-R	GTTCGCTTTGATAAACTTCCAGG GGTGATTTGGTTGAGNATGTC	655	Kimball et al., (2009)

PCR Amplification

Polymerase Chain Reaction was performed using Illustra™ puRe Taq Ready-To-Go PCR Beads (GE Healthcare), as per the standard method (Chapter 2). Amplifications with a final volume of 25µl using the puRe Taq reaction bead and 5µl of DNA extraction. Poor quality DNA samples (low DNA concentrations) were amplified using the Q5 high fidelity PCR method (Chapter 2), amplifications with a final volume of 25µl including 5µl of DNA extraction.

Standard three stage PCR cycles were used for the exponential amplification of the target DNA products. A basic protocol was used with the annealing cycle tailored to the specific T_m of the primers used, as per the general method, Chapter 2.

4.2.3 CHD Gender determination

Determination of the gender of an individual was observed through the presence of amplified products consistent with the CHD-W and CHD-Z fragments. Female avians

produce amplifications for both the CHD-W and CHD-Z fragments, producing two fragments (WZ). Male avians produce only the CHD-Z fragment, a single amplified fragment (ZZ).

Several CHD primers sets were tested on the Trinidad Piping Guan and captive Piping Guan samples, the following two primer sets were most effective during initial trials; P2/P8 (Griffiths et al., 1998) and CHD1F/CHD1R (Lee et al., 2010), see Table 4.4.

Table 4.4: CHD Primer sets and associated PCR Thermal Cycles

<table>
<tr><th>Gene</th><th colspan="2">Primers (5'- 3')* #</th><th>Product Size</th><th>Reference</th></tr>
<tr><td>CHD</td><td>P2
P8</td><td>CTC CCA AGG ATG AGR AAY TG
TCT GCA TCG CTA AAT CCT TT</td><td></td><td>Griffiths et al., (1998)</td></tr>
<tr><td colspan="5">P2/P8 Primers Thermal Cycle:
Initial Denaturation 5mins at 94ºC
30cycles of { Denaturation 30sec at 94ºC
Annealing 45sec at 48ºC±1-3ºC
Elongation 45sec at 72ºC }
Final Elongation 10mins at 72ºC
Samples held at 4ºC
(Griffiths et al., 1998; Harvey et al., 2006)</td></tr>
<tr><td>CHD</td><td>CHD1F
CHD1R</td><td>TATCGTCAGTTTCCTTTTCAGGT
CCTTTTATTGATCCATCAAGCCT</td><td>Z 320-610bp
W380-800bp</td><td>Lee et al., (2010)</td></tr>
<tr><td colspan="5">CHD1F/CHD1R Primers Thermal Cycle:
Initial Denaturation 5mins at 94ºC
30cycles of { Denaturation 30sec at 94ºC
Annealing 45sec at 48ºC
Elongation 45sec at 72ºC }
Final Elongation 10mins at 72ºC
Samples held at 4ºC
(Lee et al., 2010)</td></tr>
</table>

The use of the P2/P8 primers is demonstrated in Figure 4.3, in the identification of gender in the captive *Pipile* (samples from Lotherton L1-L6), and the use of two control individuals of known sex (sexually di-morphic *Phasianus colchicus*). The male positive control is identified by a single band, highlighted in the lane profile histogram

as a single peak, where as the female positive control produces two distinct bands and a double peak in the lane profile histogram (see Figure 4.2).

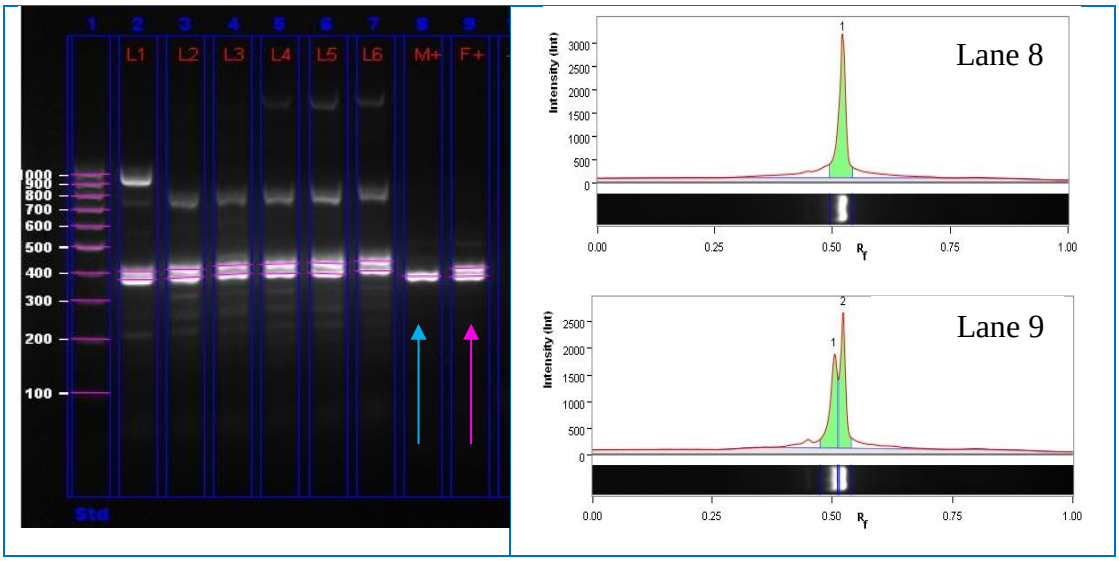


Figure 4.2: CHD Sex determination using P2/P8 primers (Griffiths et al., 1998). Gel electrophoresis image visualising male positive control sample in lane 8 and female positive control in lane 9. Histograms of lane 8 and 9 profiles showing sex determination of the CHD-Z and CHD-W fragments. Lane 1 100bp molecular ladder, and Lanes 2-7 captive *Pipile* sp. samples L1-L6.

4.2.4 Sequence Analysis

Sequencing of the amplified mtDNA samples was carried out by MWG Eurofins (MWG Biotech) using the customised A'la carte Sequencing or Value Read Services.

Purification of the PCR products was carried out using GenElute™ PCR Clean Up Kit (Sigma-Aldrich) or MinElute Gel Extraction Kit (Qiagen), using the manufacturers protocol as previously described in Chapter 2.

Prior to sequencing DNA amplifications were subjected to quantitative analysis and concentration or dilution as appropriate for the sequencing of the DNA sample, as recommended by the MWG Eurofins sample preparation guidelines. Purified DNA amplification products were quantified via gel electrophoresis image analysis in the ImageLab Software (Biorad). Samples were quantified from the standard curve, calculated from the molecular ladder using linear regression to determine absolute quantity, as per Chapter 2 general methods.

Samples were prepared as per MWG Eurofins requirements (www.eurofinsdna.com) for sequencing services as follows;

- **Value Read Service**

Individual micro-centrifuge tubes were prepared and labelled with a barcode, as follows; 10 µl of Purified DNA products at a concentration of 5 ng/µl, additionally separate tubes of forward and reverse Primers (10pmol/µl) were prepared.

- **Miq2Seq Service**

Mix2Seq pre-barcoded tubes were prepared as follows; 15 µl of Purified DNA products at a concentration of 5 ng/µl (fragments 300-1000bp), with 2 µl of either the forward or reverse Primers (10pmol/µl) added.

Quality Assessment of the Sequence Data

Visual analysis of the electropherograms for each sequence was performed to identify areas of concern according to the QV scores, for quality and accuracy of the nucleotide assignments in the sequence. Nucleotides with a low QV were clipped from the end of the sequence to ensure the accuracy of the final sequence for use in comparative analysis. Verification of the sequence was conducted through replication of the sequencing to confirm the accuracy of nucleotide assignment within the sequence, and where possible large fragment sequencing to confirm accuracy of overlapping fragment construction.

Confirmation of the Sequence Accuracy

All sequences were aligned prior to analysis; Forward and Reverse sequences were aligned using the BLAST Megablast program (http://blast.ncbi.nlm.nih.gov/Blast.cgi), to confirm the accuracy of the nucleotide assignment of the sequence results. Sequences were analysed using the BLASTn 2.2.25 software (Altschul et al., 1990), to determine the best local alignment between the sample sequence and the combined nucleotide databases, thus identifying homologous sequences which may aid in species identification. Sequences were assessed in the context of nearest species match and gene region match, in relation to nucleotide sequences on the NCBI combined databases.

4.2.5 Sequence Alignment & Construction

Complete genes were reconstructed from overlapping fragments, see primer diagram in Table 4.2. All fragments were designed to overlap by 50-100bp in order to ensure authenticity of the complete gene sequence.

Multiple sequence alignments were performed using Clustal X 2.0 with subsequent visual verification. All sequences were aligned with the reference sequence for the *Gallus gallus* whole mitochondrial genome (GenBank accession number X52392), as a standard reference point for nomenclature of the sequences. Data was partitioned primarily according to the sample and alignments were achieved to identify intra-species variations within the group and inter-species variations between sample groups.

4.2.6 Phylogenetic Analysis

Complete genes were analysed individually to establish variation within the gene and the gene tree, and in a combined gene analysis of the concatenated gene sequences to establish the overall genetic variation and phylogenetic relationships for the complete sequence data. The software jModelTest 2.1.6 (Darriba, Taboada, Doallo, & Posada, 2012), on the CIPRES Science Gateway was used to determine the best substitution model to construct phylogenetic trees for each dataset.

Molecular phylogenetic analysis of the individual gene trees were constructed by Maximum Likelihood method under the tested model in MEGA 7.0 (Kumar et al., 2016). Evolutionary history was inferred through analysis of the tree with the highest log likelihood, with bootstrap support analysis. Initial trees in the heuristic search (Maximum Likelihood method) were constructed under the MEGA algorithms for a Maximum Composite Likelihood search. All codon positions were included in the analysis (1^{st} + 2^{nd} + 3^{rd} + Noncoding), gaps and missing data were eliminated from the analysis. Maximum Likelihood methods analyse the likelihood of observing a given set of sequence data using phylogenetically informative sites, under a specific substitution model; each potential tree topology is analysed and the highest (maximum) likelihood tree is the topology selected. Maximum Likelihood trees have been shown to demonstrate high congruence with taxonomic identification. Bootstrap

Confidence Value (Felsenstein, 1985) was analysed for each topology using 500 replicate re-samples, statistical support was tested for each data set phylogenetic analyses to infer accuracy of the reconstructed tree. Bootstrap confidence values below 75% were excluded from the topology. Confidence in the reconstructed tree hypothesis is assumed with Bootstrap confidence values for internal nodes above 95%.

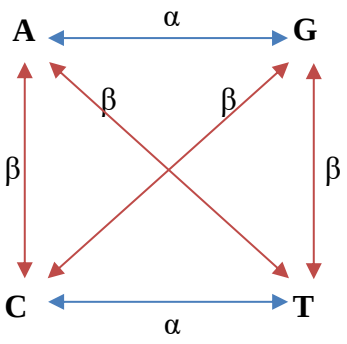
Substitution Models

The following substitution models were selected for data set analysis, as most appropriate for the given data by jModeltest.

Single complete gene analyses were constructed under the Hasegawa-Kishino-Yano (HKY) model of substitution, with bootstrap support.

HKY (Hasegawa-Kishino-Yano, 1985) Model

The HKY substitution model assumes unequal base frequencies ($\pi A \neq \pi C \neq \pi G \neq \pi T$), and that Transitions and Transversions occur at different rates (Ts/Tv bias), see Figure 4.3.



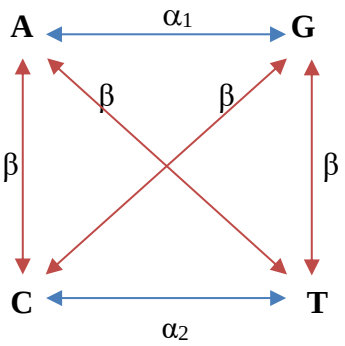
$$P_t = \begin{bmatrix} . & \beta\pi C & \alpha\pi G & \beta\pi T \\ \beta\pi A & . & \beta\pi G & \alpha\pi T \\ \alpha\pi A & \beta\pi C & . & \beta\pi T \\ \beta\pi A & \alpha\pi C & \beta\pi G & . \end{bmatrix} .$$

Figure 4.3: Hasegawa-Kishino-Yano substitution model matrix (Felsenstein, 2004; Nei & Kumar, 2000; Page & Holmes, 1998; Yang, 2014).

Combined multi-gene analyses were constructed under the Tamura-Nei (T93) model of substitution, with bootstrap support.

T93 (Tamura-Nei, 1993) Model

The Tamura-Nei substitution model is similar to that of HKY, it accounts for unequal base frequencies ($\pi A \neq \pi C \neq \pi G \neq \pi T$) assuming a GC bias, and accounts for Transitions and Transversions occur at different rates assuming a Ts/Tv bias, see Figure 4.4.



$$P_t = \begin{bmatrix} . & \beta\pi C & \alpha_1\pi G & \beta\pi T \\ \beta\pi A & . & \beta\pi G & \alpha_2\pi T \\ \alpha_1\pi A & \beta\pi C & . & \beta\pi T \\ \beta\pi A & \alpha_2\pi C & \beta\pi G & . \end{bmatrix} .$$

Figure 4.4: Tamura-Nei substitution model matrix (Felsenstein, 2004; Nei & Kumar, 2000; Page & Holmes, 1998; Yang, 2014).

4.2.7 Genetic Distance or Pairwise Distance Analysis

The degree of sequence dissimilarity was evaluated by calculating Pairwise Genetic Distances using the MEGA 7.0 program (Kumar et al., 2016); where the distance is an estimate of the number of nucleotide substitutions per nucleotide position between two sequences.

Genetic distance (p) between species or haplotype were estimated using the following equation;

$$p = n_d / n$$

where, n_d is the number of different nucleotides between two sequences and n is the total number of nucleotides (Nei & Kumar, 2000).

It should be noted that when the relationship between samples is large the p may be underestimated due to the effects of backward and parallel substitutions. Therefore, it

is necessary to use the correct model of nucleotide substitution (and substitution rate matrix) in phylogenetic analysis.

Transition/Transversion Ratio

Frequencies of different nucleotide pairs between sequences is calculated using the Transition / Transversion Ratio (R = Ts / Tv). Transitions usually occur more frequently than transversions, subject to the number of nucleotides sampled and the number of samples in the data set (Nei & Kumar, 2000).

Segregating Sites

Polymorphism may be measured as a function of the number of segregating sites per nucleotide site (P_s), using the following equation;

$$P_s = S / n$$

where, S is the number of segregating sites and n is the number of nucleotide sites (sequence length multiplied by the number of sequences) (Nei & Kumar, 2000).

Nucleotide Diversity

Nucleotide diversity (π) as a measure of polymorphism in a randomly mating population, indicates the "*average number of nucleotide differences per site between two sequences*" (Nei & Kumar, 2000).

$$\pi = \sum_{ij}^{q} \bar{X}_i \bar{X}_j \hat{d}_{ij}$$

where, q is total number of alleles (sequences), x_i is population frequency of the *i*th allele, and d_{ij} is the number of nucleotide differences per site between the *i*th allele and the *j*th allele.

(Nei & Kumar, 2000)

Mean Evolutionary Diversity is an estimated measure of the number of base differences per site from the mean diversity calculations for the entire population,

disregarding population structure (Kumar et al., 2016; Nei & Kumar, 2000), as per the following calculation;

$$\pi_T = \frac{q}{q-1}\sum^{q} \bar{X}_i \bar{X}_j \hat{d}_{ij}$$

(Nei & Kumar, 2000)

Mean Evolutionary Diversity is calculated using the above equation in MEGA7.0.

4.3 Results

4.3.1 Mitochondrial DNA Sequence Analysis

Identification of coding-gene regions was achieved using the Geneious R8 software (Kearse et al., 2012), in comparison to known reference sequence *Gallus gallus* (accession number X52392). Coding regions were identified using translated amino acid sequences and the identification of initiation and termination codons for the gene region. The Control Region, which is non-coding, was identified through recognition of conserved sequence blocks and features in comparison to the Galliforme references and the identification of the flanking tRNA genes ($tRNA^{Glu}$ at the 5'end and $tRNA^{Phe}$ at the 3'end).

No evidence of NUMTs was detected in the sequence data used in the following analyses.

Control Region

The Control Region was identified as in the standard Galliforme mitochondrial genome location between the flanking $tRNA^{Gly}$ and $tRNA^{Phe}$, at base locations 1 to 1124. The region is 1124bp ±2 in length in all *Pipile pipile* individuals, with variation in length attributed to indels in a poly-G/poly-C stretch between nucleotide positions 49 to 76 (the C-string).

Comparison of the Trinidad Piping Guan Control Region with other Cracids demonstrates a high degree of similarity in the presence of CSB features and location within the domains of the control region, as discussed in Chapter 3. The wild Trinidad Piping Guans sample group demonstrated genetic variation at five nucleotide positions in the Control Region (see Table 4.4 and Appendix 4, Figure 1). Nucleotide frequencies within the Control Region of the wild Trinidad Piping Guan population were A (26.2%), T (31.3%), C (27.2%) and G (15.3%), demonstrating an A+T bias (57.5%). Average nucleotide frequencies varied across the Domains of the Control Region, as noted in other Cracids by Pereira et al., (2004). Domain I demonstrates high

percentages of A, T and C as a result of the C-string repeat region and the repeat TA motifs in the Termination Associated Sequence (TAS).

Analysis of the Control Region in the wild population of Trinidad Piping Guans (1126bp) identifies four variable nucleotide sites, representing five haplotypes (see Table 4.4). Four of the identified single nucleotide polymorphisms represent indels (insertion/deletions) within the sequence, the remaining SNP was identified as a purine-purine A↔G Transition. Genetic distance between the five represented haplotypes in the population is 0.00089 (0.89%), see Supplementary Information, Table S1 (gaps ie. indels are excluded from pairwise genetic distance calculations).

Table 4.4: Analysis of single nucleotide variation in the Control Region in the wild Trinidad Piping Guan and UK captive Piping Guan populations.

Nucleotide Position (within Control Region)	59	77	192	235	243	305	322	461	503	554	735	926	1033	1106	Number of Individuals
Trinidad Piping Guans (wild)															
Haplotype 1	C	G		G							T		A		4
Haplotype 2	-	-		G							-		-		1
Haplotype 3	C	-		G							T		A		1
Haplotype 4	C	G		A							T		A		1
Haplotype 5	C	G		G							T		-		1
Captive Piping Guans															
Haplotype 1	C	-	A	T	T	G	C	-	T	C		A		G	1
Haplotype 2	-	-	A	T	T	G	C	-	T	C		A		G	5
Haplotype 3	-	G	A	T	C	A	T	-	C	T		G		A	1
Haplotype 4	-	-	G	C	T	A	C	-	T	C		A		A	1
Haplotype 5	-	-	A	T	T	G	C	A	T	C		A		G	1
Transition (Ts) / Transversion (Tv)	Indel	Indel	Ts	Ts	Ts	Ts	Ts	Indel	Ts	Ts	Indel	Ts	Indel	Ts	

Analysis of the Control Region in the captive Piping Guans (1126bp) identifies nine variable nucleotide sites, representing five haplotypes (see Table 4.4). An additional three of the identified single nucleotide polymorphisms represent indels within the

sequence (insertion/deletions). Genetic distance between the five represented haplotypes in the population is 0.00628 (0.62%), see Supplementary Information, Table S1 (gaps ie. indels are excluded from pairwise genetic distance calculations).

Nucleotide frequencies within the Control Region of the captive Piping Guan population were, on average, A (26.1%), T (31.3%), C (27.0%) and G (15.6%), demonstrating an A+T bias (57.4%). Nucleotide composition for the Control Region are similar for both the wild Trinidad Piping Guan population and captive Piping Guan populations.

Cytochrome b

The Cytochrome *b* gene (CytB) was identified at bases 14,777 to 15,919 in the mitochondrial genome (accession KU221053), the gene is 1143bp in length in all individuals, initiating with an ATG codon and terminating with TAA.

Nucleotide frequencies within the Cytochrome *b* gene of the wild Trinidad Piping Guan population were A (27.0%), T (25.1%), C (35.3%) and G (12.6%), demonstrating an A+T bias (52.1%). The wild Trinidad Piping Guans sample group demonstrated three single nucleotide polymorphism variations in the 1142bp of the Cytochrome *b* gene, pairwise distance between samples range from 0 to 0.003 (0-0.3%), see Supplementary Information, Table S2, representing three haplotypes (see Table 4.5). All single nucleotide polymorphisms are Transitions, which represent one first codon and two second codon non-synonymous variations which alter the amino acid code structure of the gene (see Table 4.5).

Nucleotide frequencies within the Cytochrome *b* gene of the captive Piping Guan population were, on average, A (27.1%), T (24.8%), C (35.6%) and G (12.5%), demonstrating an A+T bias (51.9%). Nucleotide composition for the CytB gene are similar for both the wild Trinidad Piping Guan population and captive Piping Guan

populations, the Cytochrome *b* gene demonstrates a low G content (12%), compared to the other genes analysed.

Table 4.5: Analysis of single nucleotide variation in the Cytochrome *b* gene in the wild Trinidad Piping Guan and UK Captive Piping Guan populations.

Nucleotide Position (within Cyt B gene)	37	736	783	876	1004	1138	Number of Individuals
Trinidad Piping Guans (Wild)							
Haplotype 1	A				T	C	5
Haplotype 2	G				C	T	1
Haplotype 3	A				T	C	2
Position in Amino acid codon	2nd				2nd	1st	
Captive Piping Guans							
Haplotype 1		T	T	C			
Haplotype 2		C	T	T			7
Haplotype 3		T	C	T			1
Position in Amino acid codon		1st	3rd	2nd			1
Transition (Ts) / Transversion (Tv)	Ts	Ts	Ts	Ts	Ts	Ts	
Amino acid variation							
Haplotype 1	K(AAA)	F(TTC)	N(AAT)	T(ACC)	I (ATC)	H(CAC)	
Haplotype 2	E(AGA)	L(CTC)	N(AAT)	I (ATC)	T(ACC)	Y(TAC)	
Haplotype 3	K(AAA)	F(TTC)	N(AAC)	I (ATC)	I (ATC)	H(CAC)	

Single nucleotide polymorphisms in the Cytochrome *b* gene of the captive population represent three haplotypes, which differ from those observed in the wild Trinidad Piping Guans, pairwise distances within the captive population range between 0-0.002 (0-0.2%), see Supplementary Information, Table S2. All of the polymorphic variations are Transitions, two of which are non-synonymous 1st and 2nd codon variations, altering the amino acid sequence in the gene (see Table 4.5).

Cytochrome c oxidase subunit 1

The Cytochrome c oxidase subunit 1 gene (COI) was identified at nucleotide position 6526 to 8076 in the mitochondrial genome (accession KU221053), the gene is 1,551bp

in length, 517 amino acid codons initiating with a GTG codon and terminating with an AGG stop codon, in all Trinidad Piping Guan individuals. Nucleotide frequencies within the Cytochrome Oxidase I gene of the wild population were A (26.4%), T (26.2%), C (31.3%) and G (16.1%), demonstrating an A+T bias (52.6%). The wild Trinidad Piping Guans sample group demonstrated no genetic variation in the Cytochrome c oxidase subunit 1 gene, all samples were 100% homologous, pairwise distance 0.0 (see Appendix 4, Figure 3).

Analysis of the COI gene in the UK captive population of Piping Guans (1551bp) identifies four variable nucleotide sites, representing two haplotypes (see Table 4.6). Genetic distance between the two represented haplotypes in the population is 0.003 (0.257%), see Supplementary Information, Table S3. All single nucleotide polymorphisms were identified as third codon Transitions, purine-purine A↔G and pyrimidine-pyrimidine C↔T, none represented a change in amino acid structure in the gene.

Table 4.6: Analysis of single nucleotide variation in the Cytochrome c oxidase subunit 1 gene in the UK captive Piping Guan population

Nucleotide Position (within COI gene)	459	546	876	1362	Number of Individuals
Haplotype 1	G	C	C	A	8
Haplotype 2	A	T	T	G	1
Position in Amino acid codon	3rd	3rd	3rd	3rd	
Transition (Ts) / Transversion (Tv)	Ts	Ts	Ts	Ts	
Amino acid variation – **Haplotype 1** **Haplotype 2**	 L (CTA) L (CTG)	 T (ACC) T (ACT)	 H (CAC) H (CAT)	 L (CTA) L (CTG)	

Nucleotide frequencies within the Cytochrome Oxidase I gene of the captive Piping Guan population were, on average, A (26.3%), T (26.5%), C (31.0%) and G (16.2%), demonstrating an A+T bias (52.8%). Nucleotide composition for the COI gene are similar for both the wild Trinidad Piping Guan population and captive Piping Guan populations.

NADH dehydrogenase subunit 2

The NADH dehydrogenase subunit 2 gene (ND2) was identified at nucleotide positions 5117 to 6155 in the mitochondrial genome (accession KU221053), the gene is 1039bp in length in all individuals, initiating with an ATG codon and terminating with T, an incomplete stop codon. Nucleotide frequencies within the ND2 gene were A (31.1%), T (24.1%), C (34.6%) and G (10.2%), demonstrating an A+T bias (55.2.%).

The wild Trinidad Piping Guans sample group demonstrated one single nucleotide polymorphism variation in the ND2 gene, pairwise distance 0.00096 (0.096%), representing two haplotypes (see Table 4.7 and Appendix 4, Figure 4). The single nucleotide polymorphisms identified represents a first codon Transitions, pyrimidine-pyrimidine C↔T, which demonstrates no change in amino acid structure in the gene.

Table 4.7: Analysis of single nucleotide variation in the NADH dehydrogenase subunit 2 gene in the wild Trinidad Piping Guan population

Nucleotide Position (within COI gene)	363	459	691	708	Number of Individuals
Trinidad Piping Guan					
Haplotype 1			C		7
Haplotype 2			T		1
Position in Amino acid codon			1st		
Captive Piping Guan					
Haplotype 1	A	G		G	8
Haplotype 2	G	A		A	1
Position in Amino acid codon	3rd	3rd		3rd	
Transition (Ts) / Transversion (Tv)	Ts	Ts	Ts	Ts	
Amino acid variation – Haplotype 1 Haplotype 2	G (GGA) G (GGG)	L (CTG) L (CTA)	L(CTA) L (TTA)	K (AAG) K (AAA)	

Analysis of single nucleotide polymorphism in the ND2 gene of the UK captive population of Piping Guans (1039bp) identifies three variable nucleotide sites,

representing two haplotypes (see Table 4.7). Genetic distance between the two represented haplotypes in the population is 0.003 (0.288%), see Supplementary Information, Table S4. All single nucleotide polymorphisms were identified as third codon Transitions, purine-purine C↔T, none represented a change in amino acid structure in the gene.

Combined mtDNA Sequence Variation

The combined mitochondrial gene analysis comprises 4879bp in length in all individuals, concatenated Cytochrome b, Cytochrome c oxidase subunit 1, NADH dehydrogenase 2 and Control Region sequences including start and stop codons.

The wild Trinidad Piping Guans sample group demonstrated nine single nucleotide polymorphism variations across the concatenated gene sequences, mean total within group pairwise distance of 0.0003 (0.18%), representing five haplotypes, see Supplementary information Table S5. Within the protein coding gene sequences single nucleotide polymorphisms identified four Transitions, pyrimidine-pyrimidine C↔T or purine-purine A↔G, with no Transversions evident. Of these three are non-synonymous 1st and 2nd codon transitions which alter the amino acid sequence of the gene and one is a nonsynonymous 3rd codon substitutions which demonstrates no change in amino acid structure in the gene.

The captive UK Piping Guan population demonstrated twenty-two single nucleotide polymorphism variations across the concatenated gene sequences, mean within group pairwise distance of 0.0008 (0.45%), representing three haplotypes. Within the protein coding gene sequences single nucleotide polymorphisms identified represent nineteen Transitions, no Transversions were observed. Of these two are non-synonymous 1st and 2nd codon transitions which alter the amino acid sequence of the gene and eight are synonymous 3rd codon substitutions which demonstrates no change in amino acid structure in the gene.

4.3.2 Nuclear DNA Sequence Analysis

Beta fibrinogen Intron 7

Analysis of the nuclear genome region for Beta fibrinogen intron 7 produced a sequence of 922bp in length, in four of the Trinidad Piping Guan samples (Reference 1, Reference 2, Pawi 4 and Pawi 5), nuclear sequence for this region failed to amplify successfully in the remaining four samples due to low samples quality and lack of nuclear DNA in the extraction. The gene region for intron 7 was confirmed by BLASTn analysis on the NCBI website; highly significant matches to other Cracidae species (99% homologous with *Pipile jacutinga* accession number AY140708).

Intra-specific variation for the four Trinidad Piping Guan individuals was low, p-distance 0.0007, in comparison the known group of nine related individuals from the UK captive population demonstrated a similarly low p-distance of 0.0029, for intron 7 of Beta-fibrinogen, see Supplementary Information, Table S6.

Table 4.8: Analysis of single nucleotide variation in the nuclear DNA Beta-fibrinogen intron 7 in the wild Trinidad Piping Guan population and captive Piping Guan population

Nucleotide Position (within Beta-fibrinogen intron 7)		99	268	276	390	642	718	771	Number of Individual
Trinidad Piping Guans									
	Haplotype 1	G	C	A	A	T	G	G	2
	Haplotype 2	G	C	A	A	T	G	A	2
Captive Piping Guans									
	Haplotype 1	G	C	A	A	T	G	A	1
	Haplotype 2	A	C	G	G	T	A	A	3
	Haplotype 3	A	C	G	A	C	G	A	
	Haplotype 4	A	T	A	G	T	A	A	5
Transition (Ts) / Transversion (Tv)		Ts	Ts	Ts	Ts	Ts	Ts	Ts	

SERPIN B14

A region 560bp in length of the SERPIN B14 intron c was sequenced from the nuclear genome (located on chromosome 2 of *Gallus gallus*), in three of the wild Trinidad Piping Guan individuals (Reference 1, Pawi 4 and Pawi 5) and nine captive Piping Guans, DNA from all other individuals failed to amplify in this region. In BLAST analysis, to demonstrate species and gene region identification all individuals in the sample sets produced highly significant homology with samples identified as *Pipile pipile* (accession number KX356185), with a 99% identity match.

Nucleotide frequencies within the SERPIN B14 intron of the wild population were A (33.8%), T (24.8%), C (18.9%) and G (22.4%), demonstrating an A+T bias (58.6%). In comparison nucleotide frequencies in the captive Piping Guan populations were A (34.3%), T (24.4%), C (19.2%) and G (22.2%), demonstrating a similar A+T bias (58.7%). In total five variable sites were observed in the entire data set, across 560bp of the SERPIN B14 intron. The wild Trinidad Piping Guans sample group demonstrated a single nucleotide polymorphism at base 509 of the sequence, representing a pairwise distance of 0.00179 (0.179%), see Supplementary Information, Table S7. This therefore resulting in two haplotypes within the Trinidad Piping Guan data set, see Table 4.9, however only three individuals were analysed.

Table 4.9: Analysis of single nucleotide variation in the nuclear DNA SERPIN B14 intron c in the wild Trinidad Piping Guan population

Nucleotide Position (within SERPIN B14)	186	432	435	475	509	Number of Individuals
Trinidad Piping Guans						
Haplotype 1	G	T	T	T	A	1
Haplotype 2	G	T	T	T	G	2
Captive Piping Guans						
Haplotype 1	G	A	A	T	A	1
Haplotype 2	C	A	A	T	A	3
Haplotype 3	C	A	A	C	A	5
Transition (Ts) / Transversion (Tv)	Tv	Tv	Tv	Ts	Ts	

Analysis in the UK captive population of Piping Guans identifies a further two variable nucleotide sites, representing three haplotypes (see Table 4.9). Genetic distance between the three represented haplotypes in the population is 0.00337-0.00893 (0.337-0.893%), see Supplementary Information, Table S7.

Combined nDNA Sequence Variation

The combined nuclear gene analysis comprises 1482bp in length in all individuals, concatenated Beta-Fibrinogen intron 7 and Serpin B14 intron c sequences. The wild Trinidad Piping Guans sample group demonstrated 2 single nucleotide polymorphism variations across the concatenated nuclear gene intron sequences (in three individuals), total pairwise distance of 0.001 (0.13%), representing two haplotypes. The captive UK Piping Guan population demonstrated 8 single nucleotide polymorphism variations across the concatenated gene sequences (in nine individuals), total pairwise distance range between 0.001-0.004 (0.53%), representing seven haplotypes. Pairwise distance between the wild Trinidad Piping Guan population and the captive Piping Guan population is 0.004, representing 99.47% homology between populations for the combined nDNA sequence data (1482bp), see Supplementary Information, Table S8.

4.3.3 Summary of Genetic Diversity

Genetic diversity within both the wild Trinidad Piping Guan population and captive Pipile population, within the given sample sets, is low with regard to measures of evolutionary change between sequences (Transition/Transversion Ratio) and genetic polymorphism (segregating sites and nucleotide diversity), see Tables 4.10 and 4.11.

Within the wild Trinidad Piping Guans the most variable individual mitochondrial DNA gene or region was the Control Region which contained the highest number of segregating sites, however Cytochrome b demonstrated the highest number of Transitions. In comparison the nuclear DNA gene introns demonstrated higher levels

of variation, Serpin B14 intron c showing the highest variation within the population. Overall measures of genetic variation were low within the wild Trinidad Piping Guan samples.

Table 4.10: Measures of genetic diversity within the wild Trinidad Piping Guan population

Gene	No. of Sequences	Segregating Sites (S)	Length (L)	n	P_s	π_T	Transitions	Transversions	R	Non-synonymous	Synonymous
Control Region	8	5	1126	9008	$5.55x10^{-4}$	0.0002	1	0	1	-	-
Cytochrome b	8	3	1142	9136	$3.28x10^{-4}$	0.0009	3	0	3	3	0
Cytochrome oxidase 1	8	0	1551	12408	0	0	-	-	0	-	-
ND2	8	1	1039	8312	$1.20x10^{-4}$	0.0002	1	0	1	0	1
Combined mtDNA	8	9	4879	39032	$2.30x10^{-4}$	0.0003	5	0	5	3	1
Whole Genome mtDNA	3	15	16663	49989	$3.00x10^{-4}$	-	8	3	2.66	-	-
β-fibrinogen intron 7	4	1	922	3688	$2.71x10^{-4}$	0.0007	1	0	1	-	-
Serpin B14 Intron c	4	1	560	2240	$4.46x10^{-4}$	0.0011	1	0	1	-	-
Combined nDNA	4	2	1482	5928	$3.37x10^{-4}$	0.0032	2	0	2	-	-

n = Number of nucleotide sites (sequence length x number of sequences)
P_s = Number of segregating sites per nucleotide (S/n)
π_T = Mean Evolutionary Diversity in the entire population – avg. number of nucleotide differences per site
R = Transition/Transversion Ratio

The use of whole mitochondrial genomes demonstrated the highest number of individual segregating sites, as would be expected by increasing the number of nucleotide sites used, however the number of segregating sites per nucleotide (P_s) was lower overall when using the whole genomes ($3.00x10^{-4}$) as compared to using the Control Region alone ($5.55x10^{-4}$).

Table 4.11: Measures of genetic diversity within the captive Piping Guan population

Gene	No. of Sequences	Segregating Sites (S)	Length (L)	n	P_s	π_T	Transitions	Transversions	R	Non-synonymous	Synonymous
Control Region	9	12	1126	10134	$1.18x10^{-3}$	0.0021	1	0	1	-	-
Cytochrome b	9	3	1142	10278	$2.91x10^{-4}$	0.0007	3	0	3	2	1
Cytochrome oxidase 1	9	4	1551	13959	$2.86x10^{-4}$	0.0005	4	0	4	0	4
ND2	8	3	1039	8312	$3.60x10^{-4}$	0.0007	1	0	1	0	3
Combined mtDNA	8	22	4879	39032	$5.63x10^{-4}$	0.0012	5	0	5	2	8
β-fibrinogen intron 7	9	6	922	8298	$7.23x10^{-4}$	0.0029	6	0	6	-	-
Serpin B14 Intron c	9	2	560	5040	$3.96x10^{-4}$	0.0013	1	1	1	-	-
Combined nDNA	9	7	1482	13338	$5.24x10^{-4}$	0.0023	7	1	7	-	-

n = Number of nucleotide sites (sequence length x number of sequences)
P_s = Number of segregating sites per nucleotide (S/n)
π_T = Mean Evolutionary Diversity in the entire population – avg. number of nucleotide differences per site
R = Transition/Transversion Ratio

In comparison the captive Pipile population demonstrated slightly higher genetic variation in all genes analysed, see Table 4.11. The Control Region, again demonstrated the highest number of segregating sites compared to any of the other mitochondrial DNA or nuclear DNA genes analysed. However the nuclear DNA beta-fibrinogen intron 7 demonstrated the highest number of segregating sites per nucleotide (P_s) and highest mean evolutionary diversity (π_T), in comparison to the other genes analysed. The lower levels of variation observed in the wild Trinidad Piping Guan population nuclear gene regions may be directly associated with the reduced sample size in comparison to the captive population (sampling bias).

Generally genetic diversity within the captive Pipile population may be considered as low, which is expected in a small captive population with a restricted genepool and high expected levels of inbreeding. However, the wild Trinidad Piping Guan

demonstrates lower genetic variation than the captive population, despite being a larger randomly mating population, which may be indicative of founder and bottleneck effects. Where the level of variation (number of substitutions per site) is high there is an increased error associated with backward and parallel substitutions which may affect the calculation of this measure of genetic variation.

Due to the low levels of genetic variation observed in both the wild Trinidad Piping Guan and captive UK Piping Guan populations, the number of Transitions and Transversion substitutions are very low. Most variations within both populations are Transitions, purine to purine (A↔G) or pyrimidine to pyrimidine (C↔T), which indicates a high Transition substitution frequency in the populations. The Transition/Transversion ratio (R), therefore is high for most gene regions (normal R values for mitochondrial genes being upto 15 and nuclear genes being 0.5-2 (Nei & Kumar, 2000)). However, the Transition/Transversion ratio is subject to sampling effects resulting in bias, as a consequence of sequence length and number of variable sites (Nei & Kumar, 2000). The sample size in this instance is small in terms of number of individuals and the length of sequence used therefore there is sampling bias to the calculation, additionally the number of Transitions and Transversions is low therefore the inferred error may be high. All models associated with phylogenetic reconstruction account for bias in the Transition and Transversion frequencies (HKY and T93).

In the wild Trinidad Piping Guan sample set most substitutions were non-synonymous substitutions of the 1st or 2nd codon nucleotide, which result in an amino acid change. This may suggest these substitutions are fixed within the population, as the substitution rates for 1st and 2nd codon sites is usually lower than that of 3rd codon sites (A. Drummond & Bouckaert, 2015). In comparison the captive Pipile population has a higher number of synonymous substitutions (3rd codon sites) than non-synonymous substitutions, which are silent and result in no change in amino acid sequence.

4.3.5 Phylogenetic Relationships

Mitochondrial DNA

Pairwise distance between the wild Trinidad Piping Guan population and the captive Piping Guan population is 0.01486 (1.48%), representing 98.52% homology between populations for the Control Region. In phylogenetic reconstruction using the Control region the wild Trinidad Piping Guans and captive Piping Guans branch as monophyletic sister populations, with high bootstrap support (<77%), see Figure 4.5.

Mean pairwise distance between the wild Trinidad Piping Guan population and the captive Piping Guan population is 0.005 (0.5%), representing 99.5% homology between populations for the Cytochrome *b* gene. In phylogenetic reconstruction using the Cytochrome *b* gene the wild Trinidad Piping Guans and captive Piping Guans branch as monophyletic sister populations, with moderate bootstrap support (>51%), see Figure 4.5.

Pairwise distance between the wild Trinidad Piping Guan population and the captive Piping Guan population is 0.011 (1.09%), representing 98.91% homology between populations for the COI gene. In phylogenetic reconstruction using the COI gene the wild Trinidad Piping Guans and captive Piping Guans branch as monophyletic sister populations, with high bootstrap support (<77%), see Figure 4.5.

Pairwise distance between the wild Trinidad Piping Guan population and the captive Piping Guan population is 0.005 (0.48%), representing 99.52% homology between populations for the ND2 gene. In phylogenetic reconstruction using the NADH dehydrogenase 2 gene the wild Trinidad Piping Guans and captive Piping Guans branch as monophyletic sister populations, with moderate bootstrap support (51%), see Figure 4.5.

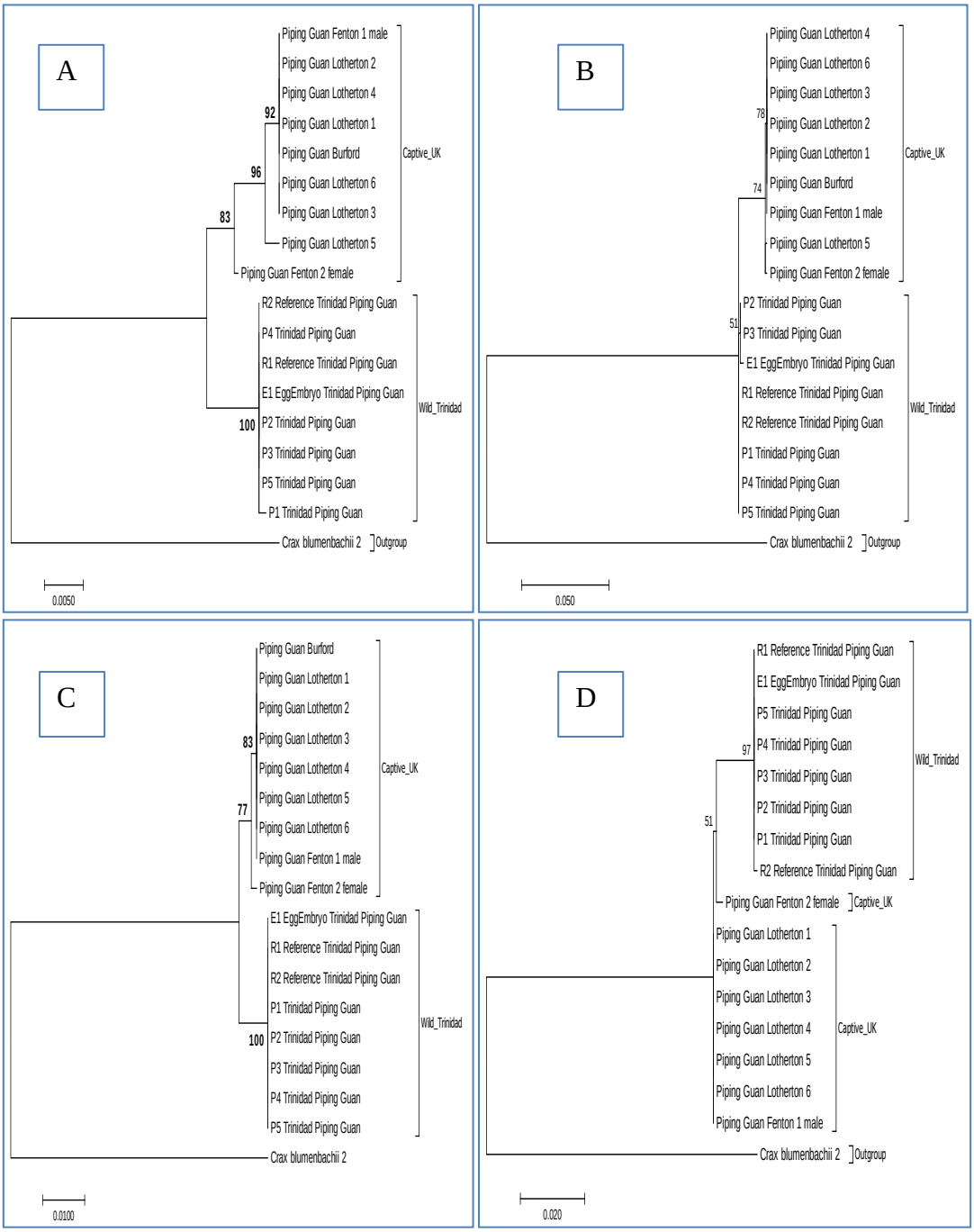


Figure 4.5: Molecular Phylogenetic analysis of the mitochondrial DNA Control Region (A), the Cytochrome b gene (B), Cytochrome oxidase 1 (C) and NADH dehudrogenase 2 (D). Using the Maximum Likelihood method using the Hasegawa-Kishino-Yano model, with 500 bootstrap replicates, demonstrating phylogenetic relationships between the wild Trinidad Piping Guan and the UK captive Piping Guan population. Bootstrap support (%) given at the nodes.

In phylogenetic reconstruction using the combined mitochondrial DNA sequence data the wild Trinidad Piping Guans and captive Piping Guans branch as monophyletic sister populations, with high bootstrap support (<74%), see Figure 4.6.

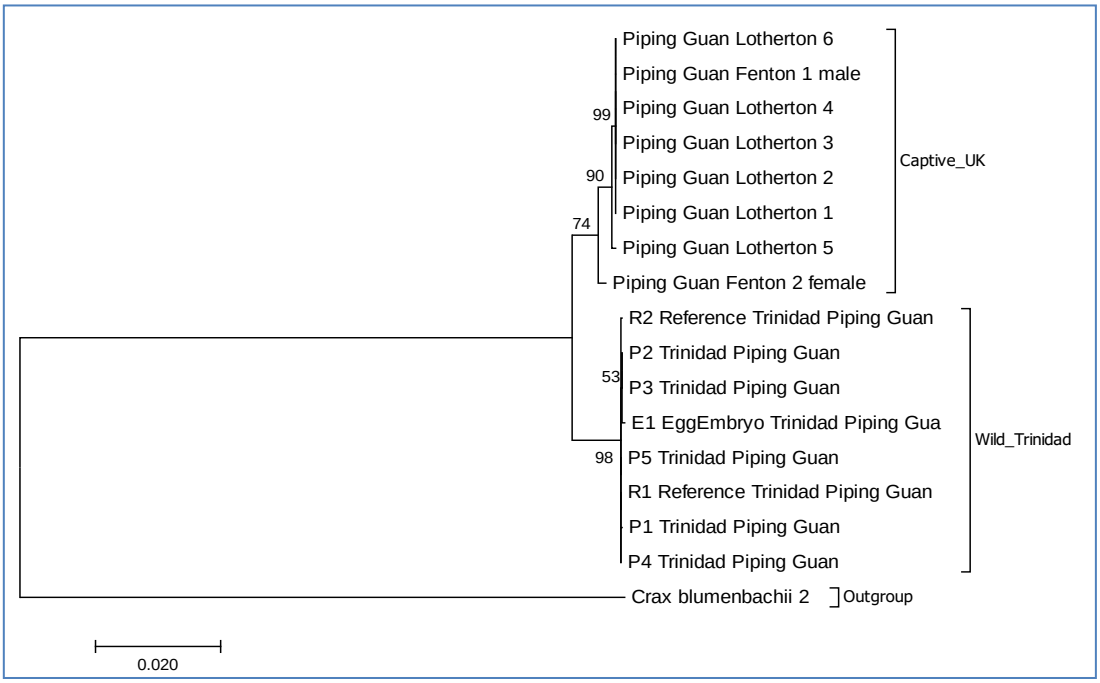


Figure 4.7: Molecular Phylogenetic analysis of the combined mitochondrial sequence analysis of the CytB, COI, ND2 and Control region (4879bp) by Maximum Likelihood method using the Tamura-Nei (T93) model with invariant sites (+I), and 500 bootstrap replicates, demonstrating phylogenetic relationships between the wild Trinidad Piping Guan and the UK captive Piping Guan population. Bootstrap support (%) given at the nodes.

Nuclear DNA

In phylogenetic analysis, using the Tamura-Nei model of nucleotide substitution, the Beta-fibrinogen intron 7 sequence fails to resolve the Trinidad Piping Guan from the Captive UK population of Piping Guans indicative of the close genetic relationship, see Figure 4.7.

Pairwise distance between the wild Trinidad Piping Guan population and the captive Piping Guan population is 0.0734 (7.34%), representing 92.66% homology between populations for the SERPIN B14 intron c. The Serpin B14 sequences demonstrated two distinctive segregating sites between the wild and captive populations, which resulted in the distinction of the two populations in phylogenetic analysis. In phylogenetic reconstruction using the SERPIN B14 nucleotide data, the wild Trinidad

Piping Guans and captive Piping Guans branch as monophyletic sister populations, with medium to high bootstrap support (<62%), see Figure 4.7.

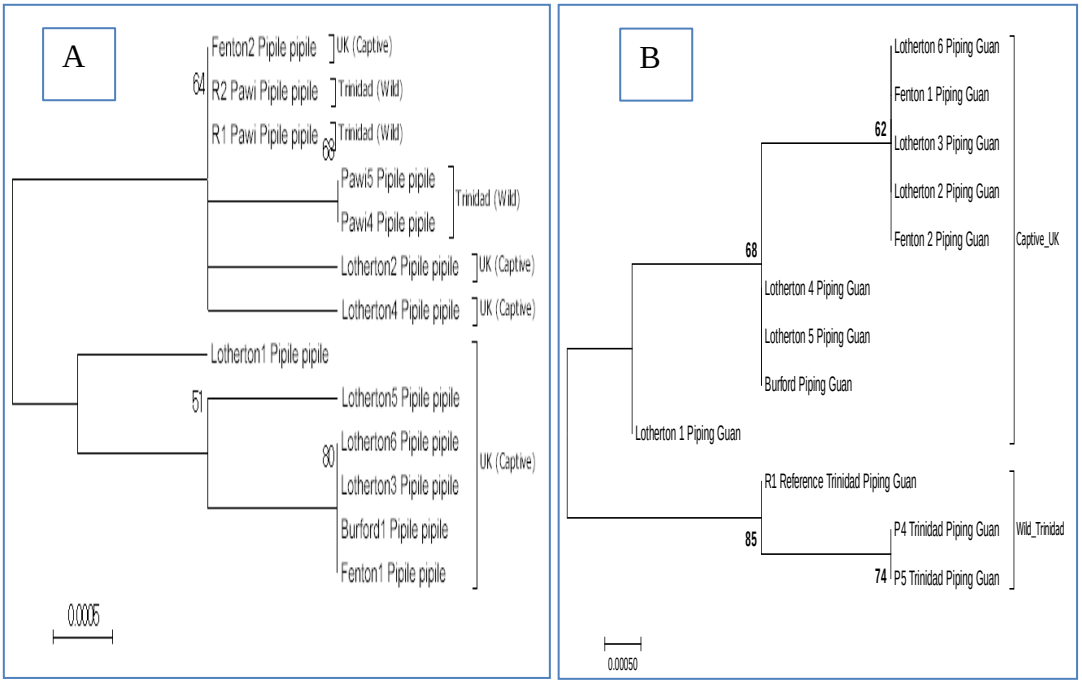


Figure 4.7. Molecular Phylogenetic analysis of the Beta-Fibrinogen intron 7 (A) and Serpin B14 intron c (B) by Maximum Likelihood method using the Jukes-Cantor model, with 500 bootstrap replicates, demonstrating phylogenetic relationships between the wild Trinidad Piping Guan and the UK captive Piping Guan population. Bootstrap support (%) given at the nodes.

In phylogenetic reconstruction using the combined nuclear DNA sequence data the wild Trinidad Piping Guans and captive Piping Guans branch as monophyletic sister populations, with high bootstrap support (93%), see Figure 4.8. However, the wild Trinidad Piping Guan population is closely allied with three of the captive Piping Guan individuals basal of the wild type, therefore indicative of a close genetic relationship.

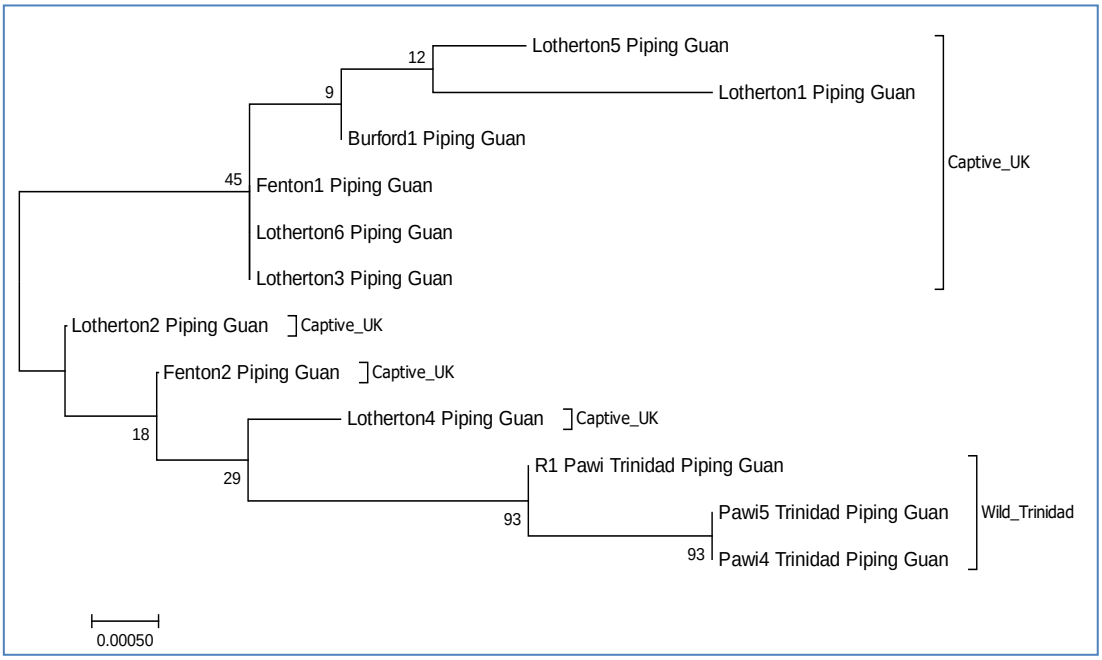


Figure 4.8: Molecular Phylogenetic analysis of the combined nuclear DNA sequence analysis of the Beta-Fibrinogen intron 7 and Serpin B14 intron c (1482bp) by Maximum Likelihood method using the Tamura-Nei (T93+Γ) with 500 bootstrap replicates, demonstrating phylogenetic relationships between the wild Trinidad Piping Guan and the UK captive Piping Guan population. Bootstrap support (%) given at the nodes.

4.3.6 CHD Sexing Analysis

Amplification using standard PCR was successful in most contemporary samples, however only a single fragment was amplified in low DNA concentrations. Biased amplification of the CHD-Z fragment has been observed in a number of CHD-sexing studies and is indicative of low quantities of nuclear DNA and the preferential amplification of fragment (Bantock, Prys-Jones, & Lee, 2008; Harvey et al., 2006). The amplification of the CHD-Z fragment is not sufficient for sex identification as it is present in both males and females. Q5 high fidelity taq reactions were more successful in the amplification of discrete CHD-Z and CHD-W fragments in sufficient yields as to be indicative of sex in many of the samples. Table 4.12 indicates the sex determination of all samples, where there is sufficient nuclear DNA yield to infer sex.

Table 4.12: Gel electrophoresis analysis results for the sex determination of wild Trinidad Piping Guans and captive Piping Guans using the P2/P8 and CHD1F/CHD1R CHD tests(Griffiths et al., 1998; Lee et al., 2010).

Sample	**P2/P8 CHD test**		**CHD1F/CHD1R CHD test**		**Sex**
	CHD-Z	**CHD-W**	**CHD-Z**	**CHD-W**	
Female Control	+	+	+	+	**Female**
Male Control	+		+		**Male**
TPG-E1	+	+	+	+	Female
TPG-R1	+	+	+	+	Female
TPG-R2	Insufficient nDNA		+		Unconfirmed
TPG-P1	Insufficient nDNA		Insufficient nDNA		
TPG-P2	Insufficient nDNA		Insufficient nDNA		
TPG-P3	Insufficient nDNA		Insufficient nDNA		
TPG-P4	+	+	+		Unconfirmed
TPG-P5	+	+	+		Unconfirmed
L1 Pipile	+	+	+	+	Female
L2 Pipile	+	+	+	+	Female
L3 Pipile	+	+	+	+	Female
L4 Pipile	+	+	+		Unconfirmed
L5 Pipile	+	+	+		Unconfirmed
L6 Pipile	+	+	+		Unconfirmed
F1 male Pipile	+	+	+		Unconfirmed
F2 female Pipile	+	+	+	+	Female
B Pipile	+	+	Insufficient nDNA		Unconfirmed
C1 male Crax	Insufficient nDNA		Insufficient nDNA		
C2 female Crax	+	+	+	+	Female

+ Positive = amplified CHD fragment present

A number of primer sets were tested to identify gender of the *Pipile* samples, two CHD primer sets were selected based on successful amplification profiles and used to identify gender of the samples; P2/P8 (Griffiths et al., 1998) and CHD1F/CHD1R (Lee et al., 2010). Analysis of both primer sets produced conclusive positive known gender controls in Pheasant samples (known male and female), however in *Pipile* specimens analysis was less consistent.

The analysis using the P2/P8primers of Grifiths et al. (1998) produced appropriately sized fragments in the known male and known female Pheasant samples, however in analysis of the *Pipile* specimens the primers amplified two fragments in all samples (wild Trinidad Piping Guans and captive *Pipile* individuals) which would be indicative

of all samples being female. This analysis was determined to be erroneous in concluding all samples were female, as some of the captive *Pipile* specimens are known males from successfully breeding groups. It was therefore determined the P2/P8 primers were producing spurious amplifications, which may erroneously be mis-identified as the CHD-W and CHD-Z fragments in gender determination.

The CHD1F/CHD1R analysis produced three amplified fragments within the molecular weight region of interest, indicative of the CHD-Z and CHD-W fragments plus a spurious amplified fragment, which may confound analysis. The spurious amplified fragment was not evident in the Pheasant known gender controls or Crax specimens indicating it may be specific to the genus *Pipile* (see Figure 4.9). In addition to the spurious amplified fragment, the putative CHD-W and CHD-Z fragments presented a reliable estimation of gender in the *Pipile* specimens; known male and female captive *Pipile* individuals (breeding pair from the Fenton collection) were reliably identified as male and female via fragment analysis.

To confirm the reliability of the gender determination using the CHD1F/CHD1R primer set of Lee et al. (2010) the putative CHD-W and CHD-Z fragments were sequenced in several individuals, both wild Trinidad Piping Guans and captive *Pipile*, to determine amplification of the correct regions of the sex chromosomes and genome. The CHD-W fragment was sequenced in three *Pipile* individuals and was determined in BLASTn (https://blast.ncbi.nlm.nih.gov/Blast.cgi) to be a sequence match to the Chromobox-helicase-DNA binding gene from chromosome W, with the closest match on the NCBI DNA database to be *Gallus gallus* with an 88% identity score (accession number GU132944). The CHD-Z fragment was sequenced in two *Pipile* individuals and was determined in BLASTn to be a match for the Chromobox-helicase-DNA binding gene from chromosome Z, with a match to *Ara militaris* with a 94% identity. The sequence matches for the CHD-W and CHD-Z chromosome fragments is sufficient to prove amplification of the correct gene regions, however the available CHD sequence data for Galliforme species in the NCBI database is limited and therefore the match scores are not significant.

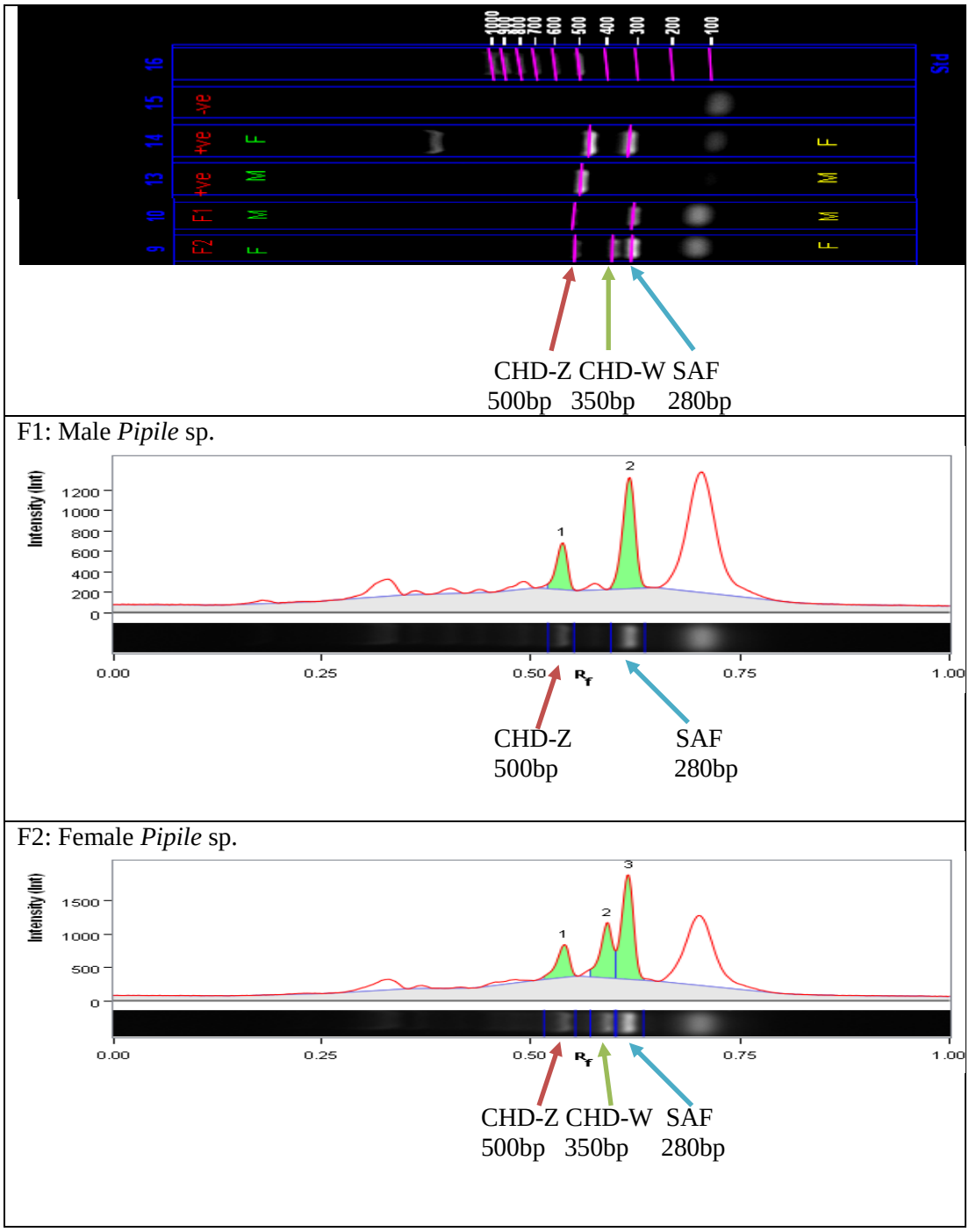


Figure 4.9: Gel electrophoresis analysis of the CHD sex marker in *Pipile* specimens of known sex using the CHD1F/CHD1R primer set (Lee et al., 2010): F1 known male *Pipile*, F2 known female *Pipile* (captive UK population), positive Male control and positive Female control known sex Pheasants (*Phasianus colchicus*) and negative control. Results indicative of three observed fragments CHD-Z, CHD-W and Spurious Amplified Fragment (SAF), male CHD profile ZZ, female CHD profile ZW.

4.4 Discussion

This study represents novel analysis of the genetic variation of the wild Trinidad Piping Guan endemic to the island of Trinidad, with reference to a small captive population of Piping Guans from the UK. It is necessary to consider the sample size in analysis of both the Trinidad Piping Guans and captive Piping Guans, which is limited, however based on current population estimates they represent 11% (Hayes, Sanasie, et al., 2009) and 25% (Species360, 2012) of the populations respectively.

Mitochondrial DNA Genetic Variation

Sequence variation in the mitochondrial genome (proportion of variable sites, nucleotide diversity and haplotype diversity) are widely used as measures of genetic variation both within and between populations or species. However, much variation is observed between genes and regions within the mitochondrial genome, notably between protein coding genes and the Control Region (Meiklejohn et al., 2014).

The four mitochondrial gene regions demonstrate distinctively low levels of genetic variation in the eight Trinidad Piping Guan individuals analysed in this study. It is important to note that when analysed in isolation some of the genes would indicate no genetic variation in the Trinidad Piping Guan population, COI shows zero variation between individuals. Had this gene been analysed in isolation, it may have been assumed that no variation was present between all individuals. In general variation observed in the wild Trinidad Piping Guan was lower than that observed in the captive Piping Guan population, potentially indicative of the genetic consequences of small population size and island effects, both contemporary and historical. Low levels of genetic variation are evident in many endemic island species, which demonstrate both historical founder or colonisation restriction and contemporary anthropogenic-fuelled population declines in population size, including; Hawaiian endemics such as the Hawaiian coot *Fulica alai*, and the Hawaiian gallinule *Gallinulla galeata sandvicensis* (Sonsthagen, Wilson, & Underwood, 2017); or the Sechelles Black Parrot *Coracopsis nigra barklyi*, endemic to the Seychelles (Jackson, Bunbury, Przelomska, & Groombridge, 2016).

The mitochondrial genome is generally treated as a selectively neutral measure of genetic variation within a population. Mitochondrial DNA is not wholly neutral, the influence of natural selection and life history traits such as mating behaviours, act on the evolution of the mitochondrial genome diversity within a species. Natural selection has a purifying influence which reduces deleterious mutation effects present in the genome and thus influences genetic diversity. Adaptive evolution of the mitochondrial genome also influences the structure of the genome as it maintains functionality of the respiratory mechanism, thereby restricting mutation within the genome, which could restrict diversity. Bazin et al. (2006) acknowledge these influencing factors in relation to the conservation of a species, where low mitochondrial genetic diversity is not necessarily detrimental to the species, but rather a function of natural selection and preservation of an adapted haplotype. The low genetic variation observed in the contemporary wild Trinidad Piping Guan population is thus comparable with other avian species which have undergone historical and contemporary bottleneck events, including the Wooping Crane (Glenn et al., 1999), Herring Gulls (Chen, Yauk, Hebert, & Hebert, 2001), and Galapogos Hawks (Bollmer, Kimball, Whiteman, Sarasola, & Parker, 2006).

Control Region

The Control Region of the Trinidad Piping Guans demonstrated the highest number of segregating sites per nucleotide (S/L), which conforms to the expectation of high nucleotide substitutions and length variations in the hypervariable regions of domains I and III (Ruokonen & Kvist, 2002). Genetic variation in the wild Trinidad Piping Guans was low with only five segregating sites (0.62%), only one of which was a SNP, the remainder being indels. All Control Region haplotypes varied by only a single base variation (0.08%), with the exception of haplotype 2 which demonstrates four indel sites (0.3%). In comparison, the captive population of UK Piping Guans demonstrate higher levels of variation, with five haplotypes across 12 segregating sites 1.06%. Variation were predominantly SNPs within domain I, and haplotypes varied by between two and eight segregating sites (0.1-0.7%). A number of the segregating sites are shared between the wild Trinidad Piping Guan and captive Piping Guan populations inferring that these may be strongly associated with genetic variation at

the genus level. Genetic variation in the Trinidad Piping Guans and captive Piping Guans is lower than that observed in many other species, including other Galliformes and Finches (Marshall & Baker, 1997; Ruokonen & Kvist, 2002; Szalanski et al., 2000). Higher levels of variation are exhibited in other Galliforme genera, including the *Bambusicola* partridges (Huang et al., 2010), which demonstrate significantly larger population sizes and wider distributions. Sonsthagen et al. (2017) demonstrated that in island endemic avians which had experienced recent severe population declines, analysis of the Control Region showed low levels of variation comparable to mitochondrial protein coding genes (ND2). The level of variation exhibited in the Trinidad Piping Guan Control Region is similar to that demonstrated in endemic Hawaiian coots and Hawaiian gallinule (five haplotypes and two haplotypes respectively), which have undergone significant recent declines in population size (Sonsthagen et al., 2017), similar to that experienced by the Trinidad Piping Guan.

The high degree of variation evident in the Control Region in most species may be indication of a higher rate of evolution compared to the rest of the genome (Ruokonen & Kvist, 2002), however in the Trinidad Piping Guan population substitution rates within the Control Region are comparable to that of transitions in the protein coding genes (CytB, COI and ND2). A slower rate of evolution in the mitochondrial Control Region has also been observed in other Galliformes (Lucchini & Randi, 1998; Randi et al., 2001) and between Cracidae species (Pereira et al., 2004). High mutation rates within the Control Region are difficult to analyse in that the effects of saturation and back-mutation in the hypervariable regions are not easily discerned (Pereira et al., 2004).

Cytochrome b

The Cytochrome *b* gene is one of the most highly used mitochondrial genes in the analysis of genetic variation and phylogenetic relationships. Genetic variation within the Cytochrome *b* gene is low in both the Trinidad Piping Guan and captive Piping Guan populations in each instance three SNPs are observed defining three haplotypes (0.08-0.17%). None of the segregating sites are shared between the Trinidad Piping

Guan and captive Piping Guan populations. In comparison to other species the level of variation in the Piping Guan populations is low, mean intra-specific variation in in avians for the Cytochrome *b* gene being reported at 0.74% (Aliabadian, Kaboli, Nijman, & Vences, 2009). In the island endemic Seychelles Black Parrot, genetic variation in the contemporary population demonstrated low genetic variation, with only a single observed haplotype in both the Control Region and Cytochrome b genes (Jackson et al., 2016). Thus, the low level of variation observed in the Trinidad Piping Guan is consistent with the levels of variation observed in other species with similar biogeographic and environmental histories.

Cytochrome c oxidase subunit 1

The Cytochrome c oxidase subunit 1 is a prominently used mitochondrial gene for the analysis of genetic variation, both intra-specific and inter-specific. The Cytochrome Oxidase I gene is 100% homologous between all wild individual Trinidad Piping Guans, used in this study, over the entire gene sequence (1,551bp). The mean intra-specific variation for avians using the COI gene is typically below 1% (Tavares et al., 2011), and has been reported in Neotropical avians at 0.24% (Kerr, Lijtmaer, Barreira, Hebert, & Tubaro, 2009) and 0.27% in North American avian species (Hebert et al., 2004). The captive Piping Guan population demonstrated four segregating sites defining two haplotypes, an intra-specific variation of 0.25%, which is consistent with published levels of variation in this gene.

NADH dehydrogenase subunit 2

The ND2 gene shows a single segregating site over the complete gene sequence (0.09%) in the wild Trinidad Piping Guan population, defining two haplotypes. In comparison the captive Piping Guan population demonstrates two haplotypes segregated by three variable nucleotide sites (0.28%). Comparable haplotype diversities were identified in Hawaiian avian endemic island species, in analysis of the ND2 gene (Sonsthagen et al., 2017), where lower genetic variation was observed in contemporary populations in comparison to historical pre-bottleneck populations.

Nuclear DNA Genetic Variation

Mutation rates differ between genes and genomes, protein coding genes in the mitochondrial genome evolve faster than protein-coding genes in the nuclear genome (Wink, 2006). Introns, due to their non-coding regulatory function, evolve at a faster rate than nuclear protein-coding gene regions, however evidence suggests they evolve slower than mitochondrial genes (Moyle, 2004; Prychitko & Moore, 2000). The analysis of nuclear DNA sequences was critically limited by the quality of the samples available, the samples collected from the wild being environmentally degraded meant that it was not possible to sequence nuclear gene regions in most of the Trinidad Piping Guan samples. Problems associated with the quality of avian samples has been widely reported in the literature (Gebhardt et al., 2009; Harvey et al., 2006; Hogan et al., 2008; Johannson et al., 2012; Leeton & Christidis, 1993; Ong & Vellayan, 2008; Segelbacher, 2002), and PCR techniques developed in this study such as the use of primerless PCR and touchdown PCR have enabled the recovery of nuclear DNA sequences from samples which failed to amplify under standard PCR conditions.

The Beta-fibrinogen intron 7 region demonstrated low variation in the wild Trinidad Piping Guan population (0.10%), indicative of two haplotypes. In comparison, the captive Piping Guan population demonstrated higher levels of variation in this gene intron (0.75%), with four haplotypes evident. None of the segregating sites were shared between the two populations indicating a distinctive genetic variation between the wild and captive populations.

In the wild Trinidad Piping Guan population the Serpin B14 intron c again demonstrated low variation with a single segregating site (0.17%) within the population. The captive Piping Guan population however demonstrated slightly higher levels of variation in the Serpin B14 intron (0.35%). The overall higher degree of variation in the nuclear gene introns compared to mitochondrial genes is evident in the analysis of the number of segregating sites per nucleotide; in the wild Trinidad Piping Guans there is a tenfold increase between mitochondrial and nuclear DNA sequence analysis (0.0003 and 0.0032 respectively). In the captive population the difference between mitochondrial and nuclear DNA sequence variation is less pronounced (0.0012 and 0.0023 respectively). This is indicative of the differences in substitution

rates between nuclear introns and mitochondrial genes. However, it may also indicate the effects of sampling bias, in terms of both sequence length and number of nucleotide sites in the analysis. The difference between the two populations may also represent sampling effects, as previously mentioned the sample numbers for both populations are low, but the number of Trinidad Piping Guan samples used in nuclear sequence analysis was approximately half of that used in mitochondrial sequence analysis therefore producing a sampling bias.

Nucleotide Frequencies

Avians typically demonstrate a compositional bias in the mitochondrial genome, low Guanine contents and an A-T bias throughout the genome. Kocher et al. (1989) reported a generally higher relative percentage of Guanine and lower percentage of Thymine in avians compared to mammals and fish. The average nucleotide frequencies across the mitochondrial genome of the Trinidad Piping Guan are consistent with that observed in other avians and Galliformes (Grass et al., 2016; Meiklejohn et al., 2014). The avian Cytochrome *b* gene is known to demonstrate low G content (Johns & Avise, 1998; Randi et al., 2001), consistent with the nucleotide frequency observation in the Trinidad Piping Guan (12.6%). The ND2 gene also demonstrated low G content (10.2%) which resulted in a high A-T bias of 55.2%.

Gender determination

The accurate determination of gender may be critical to the maintenance of healthy populations, both in captivity and the wild, and may play an important role in monitoring populations through time. This study has demonstrated a significant issue with gender determination using standard CHD gender markers in individuals of the genus *Pipile*. Conventional P2/P8 CHD markers by Griffiths et al. (1998) fail to adequately and reliably identify gender in the *Pipile* genus due to amplification of spurious DNA fragments inferring erroneous results. The analysis result of both CHD sex primer sets is inconsistent between primer sets in the *Pipile* species. The CHD1F/CHD1R primers of Lee et al. (2010) produce a more consistent result across the sexes, compared to the P2/P8 primers of Griffiths et al. (1998), which produces a biased result indicative of female markers in known male and known mixed sex groups

in the captive population. Results are conflicting between species, fragment sizes have been shown to vary between the species used in this study, consistent with published research (Fridolfsson & Ellegren, 2000; Griffiths et al., 1998; Lee et al., 2010; Ong & Vellayan, 2008). The amplification of spurious DNA fragments within the region of concern, using the CHD1F/CHD1R primers, may confound the analysis and potentially lead to incorrect gender determination in the *Pipile* species. Preferential amplification of the CHD-Z fragment and the spurious DNA fragment was observed, which may potentially result in erroneous sex classification if the CHD-W fragment fails to amplify in sufficient yields to be visualised. Griffiths et al. (1998) demonstrated an improved bi-allelic amplification may be achieved using a lower annealing temperature as a result of higher fidelity of priming (Harvey et al., 2006; Hecker & Roux, 1996), which may be beneficial in future analysis.

Genetic variation and the Trinidad Piping Guan

The levels of genetic variation may be indicative of population size effects and genetic ancestry, both historical and contemporary. The Trinidad Piping Guan population was historically widespread and abundant which promotes nuclear genetic variation through geneflow, whereas mitochondrial variation is largely constrained by maternal ancestry and therefore by founder levels of genetic variation. Non-synonymous substitutions and amino acid changes accrue and become fixed over a longer timescale, and may be evidence of more distant divergences (Wink, 2006). Mutations at non-coding sites or synonymous substitutions are neutral in that they are not selectively constrained sites and do not influence fitness (Wink, 2006), therefore there is often higher degrees of variation observed at these sites compared to coding positions.

Low genetic variation is not an uncommon phenomenon, experienced and observed in many species of fauna and flora (Galtier, Depaulis, & Barton, 2000; Habel & Zachos, 2013). The recovery of genetic variation in most species is largely reliant on immigration for the introduction of new alleles and variation (Tallmon, Luikart, & Waples, 2004), however island endemic species are not able to recover through the immigration of new individuals into the population. The rate of genetic recovery within a population is strongly associated with a number of factors including social

and breeding systems, dispersal strategies and inbreeding, loss of adaptive potential and fitness, and environmental and anthropogenic factors. The Trinidad Piping Guan population, as an island endemic species is restricted in its ability to recover from genetic loss or prolonged low genetic variation. The Trinidad Piping Guan demonstrates many traits which may inhibit population recovery including slow reproductive rates and small clutch sizes, which critically limit the species' ability to recover from genetic loss.

Founder effects may have a pronounced effect on mtDNA genetic distances and genetic variation (Wilson et al., 1985), the isolation of the Trinidad Piping Guan on the island of Trinidad will have had an impact on the genetic ancestry of the species. The availability of genetic variation will have been deterministically limited to the individuals present at the time of isolation, with a restricted genetic population the amount of variation is finite. The evolution of the genome through mutation is therefore temporally limited to the time since isolation, which in the case of the Trinidad Piping Guan is not long enough to accrue significant levels of variation. Small founder number and slow population growth rates may result in prolonged and severe genetic impoverishment in a population (Nichols et al., 2001), which may directly influence fitness and adaptive potential in this species. Habel and Zachos (2013) however, maintain that the historical view of founder and colonisation effects on limited genetic diversity in island endemics, are currently outweighed by more recent anthropomorphic pressures such as unsustainable hunting, which are resulting in a rapid decline in variation through the catastrophic loss of haplotypes.

Small refugial populations, such as island endemic species, are capable of sustaining high levels of genetic variation (Nichols et al., 2001), particularly when the founding population is genetically diverse and of sufficient population numbers (effective population size). Population densities are critically limited by the available resources including food, nest sites and territories (Nichols et al., 2001). Within the island system there are finite resources available for the Trinidad Piping Guan to utilise, therefore restricting population growth. The subsequent effects of anthropomorphic threats such as habitat destruction, agriculture and pollution will also impact on the population size

through restricting the environmental capacity and therefore restrict potential for the population to grow in the future.

The population density of Trinidad Piping Guans is considered to be lower than that of other Piping Guans (Gonzalez, 1999; Hayes, Sanasie, et al., 2009), and Cracid species (Donegan et al., 2001; Johnson et al. 1993). This may be indicative of many anthropogenic and ecological factors influencing the species and its population size, but may also be evidence of a contemporary population bottleneck in this once abundant species. The intensity and duration of population bottleneck events is instrumental in the potential recovery of a species. Genetic diversity is a requirement in order for populations to thrive and adapt under environmental pressures and anthropomorphic forces; species must be able to adapt and retain evolutionary potential and there must be sufficient variation to retain fitness (England et al., 2003).

Population bottleneck events may be demonstrated by the degree of variability in the mitochondrial and nuclear genomes of a species; recent bottleneck events may demonstrate significantly low or zero mitochondrial variation but high nuclear variation, ancient bottlenecks may result in low mitochondrial variation but high nuclear variation and prolonged bottleneck events would demonstrate low mitochondrial variation and low nuclear variation (Wilson et al., 1985). Populations experiencing significant bottleneck events could potentially lose most if not all of their mitochondrial DNA variability, elimitating all but a single matriline in several females, but maintain some of its nuclear variability in the biparental population. However, prolonged bottleneck events in conjunction with inbreeding may eliminate the nuclear variation in the population as well.

This study has demonstrated very low variation in the Trinidad Piping Guan population in the mitochondrial genome, which may be indicative of the recent decline in population size and potentially the effects of founder isolation with restricted maternal ancestry. Additionally, the study suggests that the variation in the nuclear gene intron sequences is also low, which may potentially indicate the effects of a history of prolonged or repeated bottlenecks. This study is only able to suggest that the low variation is a consequence of genetic bottlenecks, to establish the full extent of genetic

loss it would be necessary to examine the genetic history of the species from historical specimens, however there is insufficient record of this species in museum collections to establish any such analysis. Future analysis of genetic variation in the Trinidad Piping Guan species may incorporate Bayesian Skyline analysis and Ecological Niche Modelling to determine the inference of past population size on the contemporary population.

Bottlenecks have very variable effects on genetic diversity, fitness and risk of extinction (Bouzat, 2010), the outcomes of which may be largely determined by stochastic and chance effects of genetic and historical events. Research by Szabo, Khwaja, Garnett, & Butchart (2012) into global trends in avian extinctions highlighted over 150 recorded avian species have become extinct since 1500 (potentially more have been un-recorded). Their analysis demonstrates that the current rates of extinction are unprecedented in human history, two to three times higher than the background rate. The drivers of avian extinction are varied, however there are strong links to anthropogenic causes and habitat pressures, differing with each environment. Island endemic species have an evolutionary history of lower effective population sizes, which has a direct link to the increased risk of extinction through the genetic effects of decreased diversity including loss of fitness and reduced evolutionary potential (Bouzat, 2010). The Trinidad Piping Guan, as an island endemic species, has an increased risk of extinction due to many factors including its physical (metabolic demand and biomass) and reproductive traits (small clutch size and slow generation time). Being an island endemic it is also at greater risk of stochastic and catastrophic events including environmental catastrophes, disease, predation and competition by introduced species. Limited genetic variation is also known to increase the risk of extinction through reduced fitness, inbreeding effects and reduced adaptive potential. The low level of genetic variation in the Trinidad Piping Guan is therefore of significant concern, to the future survival of the species and its ability to adapt to a changing environment and climate.

Biogeographic Information

The samples collected from the wild Trinidad Piping Guan population was sampled from several regions within the Northern Range on the North-east of the island of

Trinidad. Those samples of known origins, where the geographic location of the sample collection has been recorded, may be associated with specific genetic haplotypes. The two distinctive regions associated with the sample collection are Grande Riviere on the North coast and Morne Bleu within the Northern Range mountains, see Figure 2.2 and Appendix 1. These two populations are well established, possibly resident in stable territories (Alemu et al., 2005; Hayes, Shameerudeen, et al., 2009), and regular sightings of Piping Guans in these locations are recorded over the past five years (eBird, 2012). The phylogenetic analysis of the Trinidad Piping Guan samples has established several closely related mitochondrial haplotypes within the population.

Within the Grande Riviere locale, samples were collected form two locations, the Plato plantation (Pawi 2 and Pawi 3) and Matlot (Pawi 4 and Pawi 5), genetic analysis indicates distinctive haplotypes associated with each region. Additionally, the sample from Morne Bleu (Pawi 1) is a different haplotype to those associated with Grande Riviere. This may be indicative of familial association between the sampled individuals and matrilineal types within each region, suggesting that females show limited dispersal from the territory. Hayes et al. (2009) and Alemu et al. (2005) suggests that Trinidad Piping Guans are site faithful and reside in territories of approximately 19ha, and genetic analysis in this study may show evidence to support this observation.

Critically, the analysis of haplotype association with geographic location is limited by the number of samples of known origins, the remaining samples are of unknown origins however, they can be associated with the known haplotypes. In order to establish a full analysis of intra-specific relationships within the Trinidad Piping Guan population it would be necessary to significantly increase the sample size and sample broadly across the geographic territory of the species to sample from all haplotypes within the population.

Conservation Implications

Active management of avian populations which have undergone bottleneck events have been shown to reduce the effects of genetic loss and provide a reserve population which protects the species from stochastic catastrophic events in the wild (Beissinger et al., 2008; Nichols et al., 2001). Effective genetic management and monitoring is needed to supplement wild populations with low genetic variation to facilitate genetic recovery (Beissinger et al., 2008). Captive management strategies may be highly beneficial to the Trinidad Piping Guan, in terms of establishing a reserve population and captive reintroductions to supplement the wild population.

Reintroduction programmes have been successfully established in several other critically endangered Cracid species, including the White-winged Guan *Penelope albipennis* (Cavero & Pratolongo, 2011), the Dusky-legged Guan *Penelope obscura bronzia*, the Rusty-margined Guan *Penelope supercilians jacupemba* (Pereira & Wajntal, 1999) and the extinct in the wild Alagoas Curassow *Mitu mitu* (Grau et al., 2003). The White-winged Guan has been captively bred since 1986, despite Guans being highly sensitive to stress in captivity, to establish a stable captive population and re-introduce White-winged Guans to conservation areas (Cavero & Pratolongo, 2011). This successful captive breeding programme presents precedent for establishing such programmes in other critically endangered Cracids, including the Trinidad Piping Guan. Genetic management plays a significant role in the development of a healthy reintroduction population, previous genetic monitoring in Penelope species has been performed by DNA profiling (Pereira & Wajntal, 1999), however techniques using microsatellites and DNA sequencing may be more valuable genetic markers. Captive populations are by definition founded by a small number of individuals therefore the genetic diversity within the population is limited. Genetic monitoring of managed pairings enables the preservation of the genetic diversity within the captive population, and minimises the loss of diversity which naturally occurs in small populations and through the effects of genetic drift and relaxed selection.

Conservation policies often incorporate a sustainable management of natural resources by local people, which can be detrimental to the structure of the forest system making dependant species prone to extinction (Galetti et al., 1997). The Piping Guans are of

significant socio-economic and biological importance within the Neotropical forest ecosystem, as they are one of a number of large avian frugivore species which have an important role in forest maintenance through seed dispersal and seed predation. The consequences of extinction may extend beyond the loss of an avian species to a significant impact within the forest ecosystem including its ability to naturally regenerate over time.

Exploitation of fruiting trees for wood harvests would be detrimental to large avain frugivore taxa, increasing competition for resources, and opening them to conflict with man through their use of agriculture sites for food resources and breeding sites. Some species of Cracid have been shown to utilise secondary forests in response to the loss of primary habitat, however others such as the *Crax* genus are site faithful and show little ability to adapt (Borges, 1999; Galetti et al., 1997). The Trinidad Piping Guans have been shown to adapt to the presence of humans within their environment, and have adapted their diet to suit the available resources (Hayes, Shameerudeen, et al., 2009). However, they have also been shown to demonstrate site fidelity (Alemu et al., 2005) which may critically affect their conservation and future survival.

4.5 Conclusions

This study has examined the genetic variation in DNA sequences observed in the wild and captive populations of the Trinidad Piping Guan. The analysis has demonstrated low levels of variation in both the mitochondrial and nuclear gene sequences, which may be attributed to both historical and contemporary population influences of genetic diversity. Development of techniques to examine the genetic variation in the Trinidad Piping Guan species may of great benefit to future in-situ and ex-situ conservation efforts.

Chapter 5. Evolution of the Trinidad Piping Guan

5.1 Introduction

The Avian Class & Evolutionary History of the Avians

The avian class is one of the major lineages which diverged from the Archosaurs; their closest extant relative being the Crocodilians as opposed to their more distant relationship with mammals. Early crown avians overlapped in evolutionary history with the archaic birds, including the Ichthyornis, and the flying reptiles, such as the Pterosaurs, the crown avians being the only group to survive into the Teritary (Slack et al., 2006). The modern avian class (Neorinthes) is clearly partitioned into two distinct clades; the Paleognaths including the ratites and tinamous, and the Neognaths which includes all other modern birds (Braun & Kimball, 2002; van Tuinen et al., 2000). There are approximately 10,500 extant species of avians across a diverse range of classes, the most expansive being the Neoaves. These clades have been well documented and defined by morphological, behavioural and genetic observations (Ibarguchi, 1999, Mindell et al., 1999, Carcraft, 2001, Paton, Haddrath and Baker, 2002, Jarvis et al., 2014, Zhang et al., 2014). Basal among the Neognaths is the Galloanserae, a species rich clade containing the Galliformes (landfowl) and Anseriformes (waterfowl) (Eo et al., 2009; van Tuinen et al., 2000).

The Cretaceous-Paleogene boundary (K-Pg) approximately 66 million years ago (Mya), also referred to as the Cretaceous-Tertiary boundary (K-T), was an instrumental point in vertebrate evolution globally, the Chicxulub meteorite impact mass extinction event had irreversible biological consequences (Carcraft, 2001; Cooper & Penny, 1997). Cretaceous avians suffered mass extinctions followed by a period of adaptive radiation in the Paleogene, however the extent of survival and rate of divergence of avian lineages is much debated (Cooper & Penny, 1997; Ibarguchi, 1999; Pacheco et al., 2011; Slack et al., 2006; Zhang et al., 2014; Zink et al., 2004). Primitive avian lineages are known from fossil records of the Jurassic and Triassic periods, however evidence from these periods is limited and incomplete. Cooper and

Penny (1997) estimated the earliest avian diversifications as pre K-Pg boundary, approximately 100Mya in the early Cretaceous period, based on molecular analysis of the mitochondrial 12S gene and nuclear C-mos proto-oncogene, using paleontological data to calibrate divergence times. Biogeographic and phylogenetic analyses of modern avians indicates a strong link between the dispersal of the southern continents and the diversification of avian lineages approximately 100Mya (Carcraft, 2001; Cooper et al., 2001; Harrison et al., 2004).

The Quaternary period (2.4Mya to present) has been defined by frequent global climatic oscillations, resulting in a series of ice ages the most recent occurring approximately 11,000 years ago (Avise & Walker, 1998; Hewitt, 2000). Global fluctuations in climate on this scale naturally produced significant changes in species distributions and survival, as evidenced by the fossil record and genetic legacy of contemporary species (Avise & Walker, 1998; Avise, Walker, & Johns, 1998). Major glacial periods were typified by polar ice sheet expansion, compression of the temperate zones, significant glaciation of the montane regions, reduced sea levels, restriction of tropical forests and the extension of deserts and savannahs (Cooper & Penny, 1997; Hewitt, 2000), resulting in major population fluctuations.

The exceptional levels of biodiversity in the Neotropics has been recognised since the 19th century (noted by early naturalists such as Darwin and VonHumbolt), driven by cycles of geographical and environmental stability and in-stability (Rull, 2011). Environmental shifts linked to Quaternary glaciations and tectonic reorganisation induced barriers and pathways for evolution of Neotropical biota through speciation linked to refugia and diversification through paleogeographical changes such as the Andean uplift (Rull, 2011). The release of ecological niche environments as a consequence of major biogeographic events and catastrophes enabled rapid diversification of the avian class (Jarvis et al., 2014). Both flora and fauna species of the South American continent diversified as a result of major geographical changes during this period. In the case of Trinidad, it is theorised that, the continental land bridge to Venezuela was exposed allowing the dispersal of species from mainland South America to the island of Trinidad (Grau et al., 2005).

Genetic relationships within the Galliforme order have been heavily influenced by temporal and geographical events. The genetic relationships and evolutionary patterns of Galliforme and avian species have been clearly linked to historical events on a global scale; the genetic diversification of species through successive range changes and population dispersal (Zink et al., 2004). Diversification of the avian class has been shown to slow and accelerate in response to stochastic and global events (Zink et al., 2004), the balance of speciation and extinction shifting in response to local changes. Unique genetic populations arose in light of these geographical changes, however contemporary mtDNA haplotypes originate in the pre-glacial past with little divergence through new mutation, due to the relatively low mutation rate of the genome (Hewitt, 2000; Zink et al., 2004). Current mitotypes may reveal both the demographic and geographic history of a species with genetic divergence proceeding over millions of years. The genetic imprints of successive glaciations and historical events may in theory be evident in the genome in response to evolutionary adaptation to environmental and geographical range changes (Zink et al., 2004).

The Galliforme Order

The Galliforme order is one of the most important groups of avians to human society, comprising 281species in 81 genera, the order is of significant socio-economic and ecological importance. Galliformes are one of the most widely distributed orders of avians, with global distribution and an immense variety of morphological adaptation allowing them to colonise almost every continent. The first published avian mitochondrial genome and avian nuclear genome sequences were of the domestic Chicken *Gallus gallus* (Desjardins & Morais, 1990), and it remains one of the most highly studied avian species. Despite being one of the most studied avian orders, taxonomic relationships remain unclear within the Galliformes due to the low morphological variation within families (Kimball, Braun, Zwartjes, Crowe, & Ligon, 1999; Van Tuinen & Dyke, 2004) and inconsistencies among molecular datasets. Poorly characterised species and uneven taxon sampling have resulted in a bias in available data for the Galliformes, which directly impacts on the ability to resolve a phylogenetic framework (Eo et al., 2009).

Galliformes are generally considered one of the more basal clades of the modern avians (Neognathes) closely associated with the Anseriformes (Dimcheff et al., 2002; Dyke et al., 2003; Kan et al., 2010; Meiklejohn et al., 2014; Shen et al., 2010; Wang et al., 2013). Morphological classification of the Galliforme order by Dyke et al., (2003) using osteological and anatomical features is broadly supportive of traditional classification and the initial molecular phylogenetic placement of the families using DNA-DNA hybridisation (Sibley & Ahlquist, 1990). Within the Galliforme order there are seven well accepted groups, however their true classification as families or sub-families is disputed and remains unresolved (Kan et al., 2010). The seven accepted suprageneric classifications within the Galliforme order comprise; Megapodiidae, Cracidae, Odontophoridae, Numudidae, Phasianidae, Meleagrididae and Tetraonidae. Of these classifications the most disputed groups are the Odontophoridae and Numididae (New World and Old World Quails) (Dimcheff et al., 2002; Dyke et al., 2003; Kan et al., 2010; Nishibori et al., 2001; Van Tuinen & Dyke, 2004), see Figure 5.1.

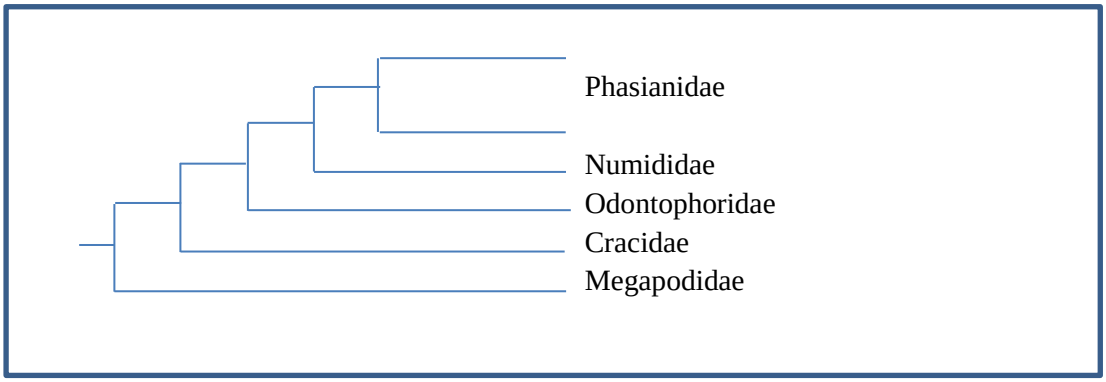


Figure 5.1: Current Consensus on Galliforme Relationships based on DNA hybridisation, Cytochrome B and ND2 partitions adapted from Dimcheff et al., (2002) and Van Tuinen & Dyke (2004)

The Cracidae by contrast are well established as one of the oldest lineages, closely allied with the Megapodiidae (Mound builders, Scrubfowl and Brushturkeys etc.) (Dimcheff et al., 2002; Dyke et al., 2003; Sibley & Ahlquist, 1990). The Phasianidae (Pheasants and Partridges) are the most diverse clade within the Galliformes and evidence is indicative of a rapid radiation post K-Pg resulting in high genetic diversity within this group which demonstrates cryptic morphological taxonomy (Kimball et al., 1999). The Phasianidae, due to their diversity, remain largely unresolved and sub-

families within the classification have been greatly disputed (Crowe et al., 2006; Kan et al., 2010). Discrepancies among the Phasianidae are thought to present due to incomplete representations of species in the phylogenetic analyses, cryptic divergences and multiple evolutionary paths, resulting in complex topologies within this clade (Kan et al., 2010; Meiklejohn et al., 2014). The turkeys represent another cryptic group, the Meleagridae often being phylogenetically positioned within the Phasianidae in locations which vary greatly between analysis (Crowe et al., 2006; Kimball et al., 2001; Meiklejohn et al., 2014; Pereira & Baker, 2006; Slack et al., 2007), this clade is indicative of the incomplete resolution of the Galliforme mito-genomic evolutionary tree.

In avian genomics the Cytochrome *b* and Cytochrome *c* oxidase I genes are the most common genes applied to phylogenetic and evolutionary research. Previous studies into the phylogenetics of Galliforme lineages has largely focused on single gene or multi-gene analyses, rather than whole genome analysis, and have resulted in significant inconsistencies in tree topology (Armsrtong et al., 2001; Dimcheff et al., 2002; Nishibori et al., 2001). Conflicting information from single gene trees may be misrepresentative of the complete mito-genomic relationships within the order, as indicated by the variation observed in some species (Odontophoridae, Meleagridae, etc.), the causes of which can be unclear (Meiklejohn et al., 2014). The use of whole mitochondrial genomes is generally accepted to provide a more accurate representation of phylogenies than single or multi-gene analysis (Kan et al., 2010; Meiklejohn et al., 2014). A paucity of complete mitochondrial genome information is currently available to confirm species relationships among the Galliforme order, or to resolve the relationship between Galloanserae (Galliformes and Anseriformes). Slack et al. (2003) used a single species from each of the Galloanserae orders within a broader avian analysis to infer their relationship as a sister group basal within the Neognaths. However further taxon sampling is necessary to clarify issues of Galliforme evolution and further resolve the avian tree. Figure 5.2 demonstrates cladograms of the phylogenetic relationships of Galliformes as proposed by a variety of single gene, multi-gene and whole mitochondrial genome techniques. The phylogenetic relationships within the Galliformes clearly identify the seven key suprageneric clades, as described previously, with some evidence of further sub-divisions of the Phasianids.

The early analysis of Kimball et al (1999) examines the phylogenetic relationships within the Phasianidae in the context of other Galliformes, including Cracids. Kimball et al. utilised single gene analysis of the Cytochrome *b* gene and Control Regions of the mitochondrial genome to infer phylogenetic relationships of 39 species of Galliformes, demonstrating evidence of the rapid speciation of the Phasianids. Phylogenetic reconstruction using the Cytochrome *b* gene confirmed the basal position of the Cracidae family in relation to the other Galliformes (Kimball et al., 1999).

Pereira and Baker (2006) used combined analysis of three mitochondrial genes (Cytochrome *b*, 12S and ND2) to resolve molecular timescales of the Galliforme order relationships. Pereira and Bakers research supports the early diversification of the Cracidae and Megapodidae, and relationships within the Galliformes, in agreement with other research (Dimcheff et al., 2002; Dyke et al., 2003). As with the research of Kimball et al (1999), Periera and Baker (2006) present a non-monophyletic Phasianidae clade incorporating successive sister relationships with the Tetraonidae, Meleagridae, Numididae and Odontophoridae.

Crowe et al. (2006) analysed extensively the Galliforme order utilising molecular and taxonomic data from 158 species across 65 genera, their conclusions support the sub-division of the order into the seven accepted suprageneric taxa. Molecular analysis using combined DNA sequence data of the CytB, ND2, CR, 12S and nuclear ovomucoid intron G (OVO-G), in conjunction with morpho-behavioural character state analysis, constructed a highly resolved and well supported phylogenetic tree (see condensed tree in Figure 5.2).

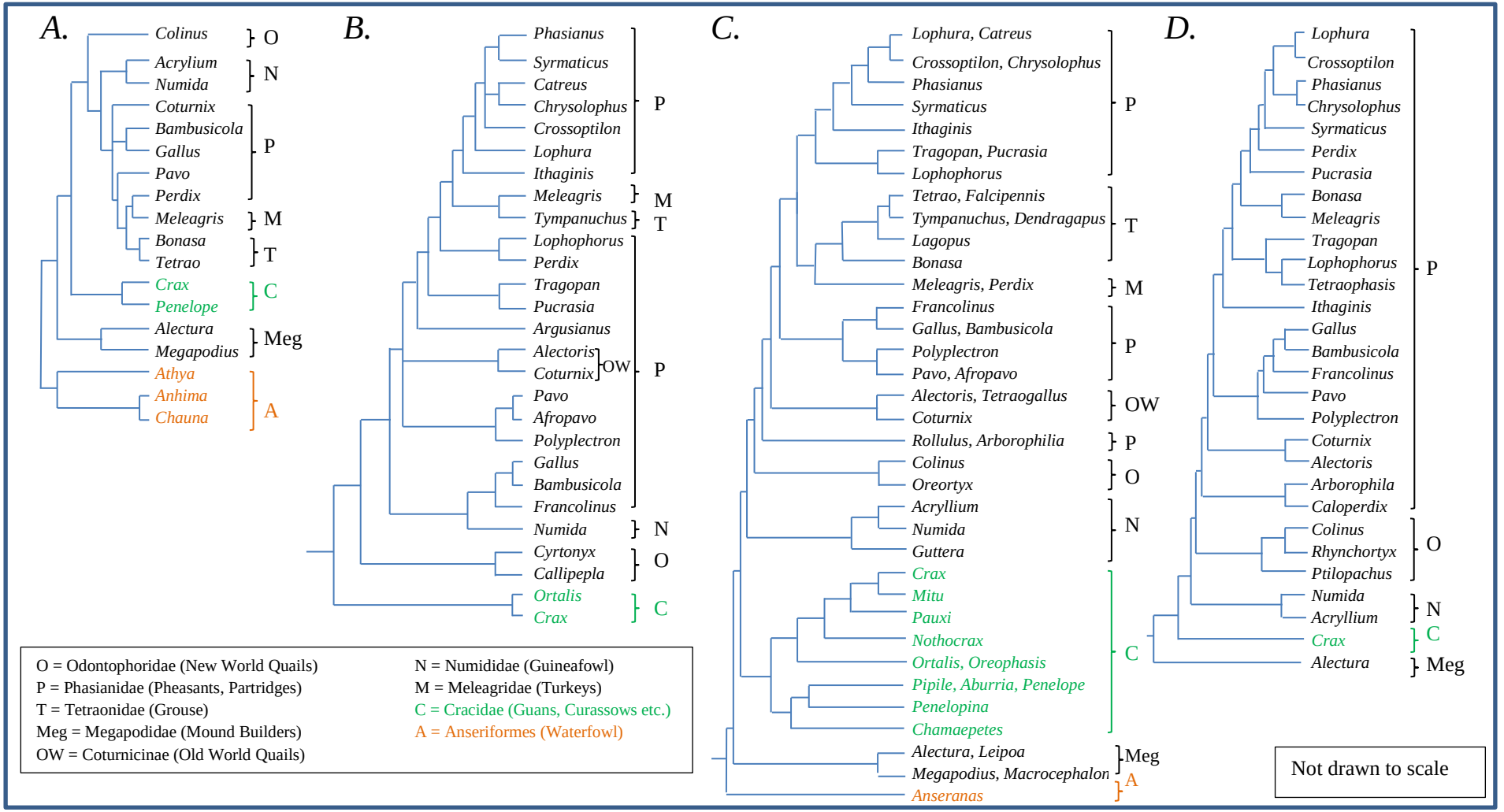


Figure 5.2: Cladograms from various molecular phylogenetic analyses of the Galliforme genera; **A**. Pereira & Baker (2006) – 12S, ND2 & CytB 1756bp ML, **B**. Kimball et al. (1999) – CytochromeB 1143bp ML, **C**. Crowe et al. (2006) – CR, CytB, ND2, 12S and Ovo-G combined data set (condensed tree), **D**. Meiklejohn et al (2014) – whole mtDNA 13genes ML.

The research of Meikeljohn et al (2014) investigated conflicting analysis within the Galliformes though the use of complete mito-genomes from 47 species of Galliformes, including the novel use of Cracid whole mito-genomes in analysis. The results of this study show that the use of whole mito-genomes to construct Galliforme phylogenies results in high degrees of support, and demonstrate trees congruent with other studies including Kimball and Braun (2008) and Wang et al., (2013). Comparison of single gene trees (individual protein-coding genes) and whole mito-genome trees showed congruence between tree topology however node support varied greatly. Conflicting signals were observed in the relationship between the Cracidae and Megapodidae, however this may be indicative of the narrow taxon sampling in this area of the tree or model effects (Meiklejohn et al., 2014).

The Cracidae is a primitive family within the Galliforme order, demonstrating a high level of divergence from other Galliformes, genetically and morphologically. The Cracidae family contains 11 genera with over 50 species, however species relationships are unresolved within this family, despite analysis of multiple genes from the mitochondrial and nuclear genomes (Grau et al., 2005; Grau et al., 2003; Hosner et al., 2016; Pereira et al., 2002). The cracid family is generally regarded as having diversified through a number of routes including parapatric and sympatric modes the result of both ecological and biogeographic causes. Significant geographical movements have been observed between historical and contemporary species distributions based on fossil evidence (Tordoff & MacDonald, 1957) and molecular analysis (Hosner et al., 2016). Ibarguchi (1999) analysed evolutionary divergences and relationships of the Cracids to other basal avians using 12S and Cytochrome *b* genes; analysis confirms the relationship of the Galloanserae, with Cracids basal among the Galliforme order. Analysis of the 12S and Cytochrome *b* genes estimate the divergence of Cracids from a common ancestor between ~54-36MYA, consistent with fossil evidence (Ibarguchi, 1999).

To date a small number of complete mitochondrial genomes have been described for *Cracidae* species, which includes two Cracid species (*Crax daubentoni* and *Crax rubra*, accession numbers KJ914544 and KJ914544 respectively), however these only

recently been utilised in a broader analysis of avian phylogenetics (Meiklejohn et al., 2014). In direct analysis of the Cracidae clade within the Galliforme tree Meiklejohn et al. (2014) acknowledges a distinct gap in the data set for the Cracidae family and the requirement for the addition of further cracid mitochondrial genomes to analyse the basal topology of the tree. The inclusion of a Trinidad Piping Guan whole mitochondrial genome, will to some extent further this aim, improving the resolution of the Cracid branches within the tree.

In depth analysis of the Cracidae, by Hosner et al. (2016), incorporating 38 of the accepted 55 Cracid species produced the most complete analysis of the family to date using multi gene analysis of both mitochondrial and nuclear DNA markers. This extensive study of the evolutionary history and diversifications of the Cracidae demonstrated strong support for the four main clades within the order; the Currasows, the Chachalacas, the Horned Guans and the Guans. The Guan sub-order reconstructs strong monophyletic branches of the *Pipile, Aburria, Penelope* and *Chamaepetes*, in analyses of individual mitochondrial genes (Hosner et al., 2016).

Piping Guans of the genus *Pipile* (*Aburria*) are represented in a number of phylogenetic studies, however their genetic status remains undefined. The analysis of Hosner et al. (2016) used a combined model approach with molecular data from single mitochondrial genes, nuclear gene introns and ultraconserved element sequences. This approach identifies some contradictory areas of analysis, including the nesting of the *Aburria* within the *Pipile*, initially proposed by Grau et al. (2005), which only occurs in some individual mitochondrial gene analyses and is not supported by nuclear markers (Hosner et al., 2016). Molecular divergence estimates for the genus *Aburria/Pipile* suggest significant influence of the glacial periods on the isolation and speciation of the genus and its ancestral populations (Grau et al., 2005). Biogeographic analysis of Hosner et al. (2016), indicates the Cracid ancestor diverged 48-68Mya from the Old World-Central American region, to diversify in tropical regions in response to declining global temperatures and biogeographic events on the American continents. This analysis is consistent with fossil evidence from North American Cracid ancestors and known biogeographic/climatic models. Hosner et al. (2016), determine the

colonisation of South America by the Guans is proposed ca. 3.9-7.0 Mya, inferred through a time-calibrated tree, with the diversification of Guan species during the Pliocene ca. 3.0-4.8 Mya consistent with major biogeographic changes on the South American continent.

As previously discussed, the Trinidad Piping Guan is theorised to have become isolated on the island of Trinidad off the coast of Venezuela during the last glacial period after the sea levels rose (Grau et al., 2005). The isolation of the Trinidad form of Piping Guan is therefore estimated to represent a more recent diversification than that observed in the mainland forms of the genus.

Avian Mito-genomics & Evolutionary Phylogenetics

The Avian class has a diverse genetic background which represents almost 10,000 species across a diverse and complex evolutionary history. Understanding the genetic associations within the macro and micro evolutionary relationships between extant species may resolve phylogenetic and biogeographic hypotheses for the Avian class. Understanding the complexities of the avian relationships through genomics may be instrumental in establishing conservation priorities both on an individual species basis but also on an ecological or global level.

Vertebrate mitochondrial genomes are well suited to the reconstruction of evolutionary relationships, producing robust well supported topologies (Meiklejohn et al., 2014), as well as providing insights into the genetic structure and function of the molecule, survival of lineages, ecological processes and biogeography (Harrison et al., 2004). The mitochondrial genome is largely considered to demonstrate selective neutrality, which is important in the evolution of the molecule (Shen, Shi, Sun, & Zhang, 2009). The prominent functional role of genes within the mitochondrial genome and small amount of non-coding sequence within the genome mean that functional constraints restrict the evolution of the genome. In addition its lack of recombination and maternal mode of inheritance make the mtDNA genome make it ideal for a number of molecular

analyses including phylogenetics and evolutionary research. The mitochondrial genome arrangement has been found to be conserved among many species of mammals, amphibians and some fish (Macey et al., 1997; Pereira, 2000), as previously discussed in Chapter 3. The use of long sequence data sets demonstrates a higher degree of accuracy in the reconstruction of phylogenetic hypotheses, and this technique has highlighted incongruences between the use of single genes or short DNA sequences in recovering a robust and representative phylogenetic tree (Paton, Haddrath and Baker, 2002).

The phylogenetic relationships among extant avian orders remains unresolved, general consensus implies that extant birds are divided into two groups, the Paleognaths (Ratites, Tinamous etc.) and the Neognaths (all other modern avian orders) (Cooper et al., 2001; Harlid & Arnason, 1999; Härlid, Janke, & Arnason, 1997; Mindell et al., 1999; Pacheco et al., 2011; Slack et al., 2003). Within the Neognaths several conflicting evolutionary trees have been proposed, in some orders the monophyly has been well established whilst in others paraphyly is evident; the status of some orders remains uncertain such as the Falconiformes and the Piciformes (Pacheco et al., 2011). In analysis of avian species identification Lee et al (2008) incorporated 331 individual avians across 40 species to perform cluster analysis of phylogenetic relationships for a short fragment of the Cytochrome B gene. Analysis demonstrated clustering of species congruent with morphological identifications however tree structure is inconsistent with typical genetic relationships. Pacheco et al. (2011) analysed the complete mitochondrial genomes of 80 avian species within the Neoaves, which demonstrated the timings of radiation and diversification within the modern avian tree, rooted with the Galloanserae. This analysis found that the COI gene demonstrated the slowest evolution, with low levels of variance among orders, which is supportive of the genes use as a Barcoding gene (Pacheco et al., 2011). Early research by Mindell et al. (1999), using whole mitochondrial genomes, found the ratites to be embedded within the avian tree using analysis of seven avian species (from 7 orders) with reptilian and mammal outgroups, which rejects traditional topology.

Paraphyly has been reported in several mitochondrial genome phylogenetic analyses (Johnson, 2001; Meiklejohn et al., 2014; Mindell et al., 1999). Research by Johnson (2001) suggests paraphyly may be a consequence of insufficient taxonomic representation producing erroneous relationships within avain topologies, single species representatives erroneously inserted in neighbouring species groups. Addition of increasing numbers of avian species in mito-genomic analyses has improved confidence in phylogenetic estimations and developed improved methods for more accurate, well resolved trees (Harrison et al., 2004; Meiklejohn et al., 2014). The use of large scale data sets (increased taxonomic coverage) has a stabilising effect on the phylogenetic reconstructions observed in the avian tree (Morgan-Richards et al., 2008). Additionally it has been noted that the inclusion of outgroups of increasing organismal divergence may disrupt species topologies of well established clades as a consequence of long branch effects (Johnson, 2001; Morgan-Richards et al., 2008; Slack et al., 2007). The avian class has no close common ancestor with other taxonomic groups, the closest extant group being the reptiles, therefore the effects of rooting trees with a reptile outgroup may produce erroneous topological relationships and long branch effects.

A large number of whole mitochondrial genomes of avians have now been published which allow a better understanding of the phylogenetic relationships among avian Orders (Slack et al., 2003). The use of complete mitochondrial genomes as opposed to single gene or multi-gene analysis may be able to resolve traditionally held views regarding the avian evolutionary tree. An increased taxon sampling using complete mitochondrial genomes should resolve topological issues, however it is limited greatly by the extinction of species and lack of evidence from ancestral states (Harrison et al., 2004). Jarvis et al. (2014) analysed a number of genome and mito-genome techniques to discern the relationships in the Neoavian tree, incongruences arose as a consequence of insufficient taxon sampling and long branch attraction effects. Nuclear genome analysis is highly supportive of the basal divergences within the extant avian tree, however incongruences with mitochondrial and DNA-DNA hybridisation analyses arise within the Neoaves (Jarvis et al., 2014), see Figure 5.3.

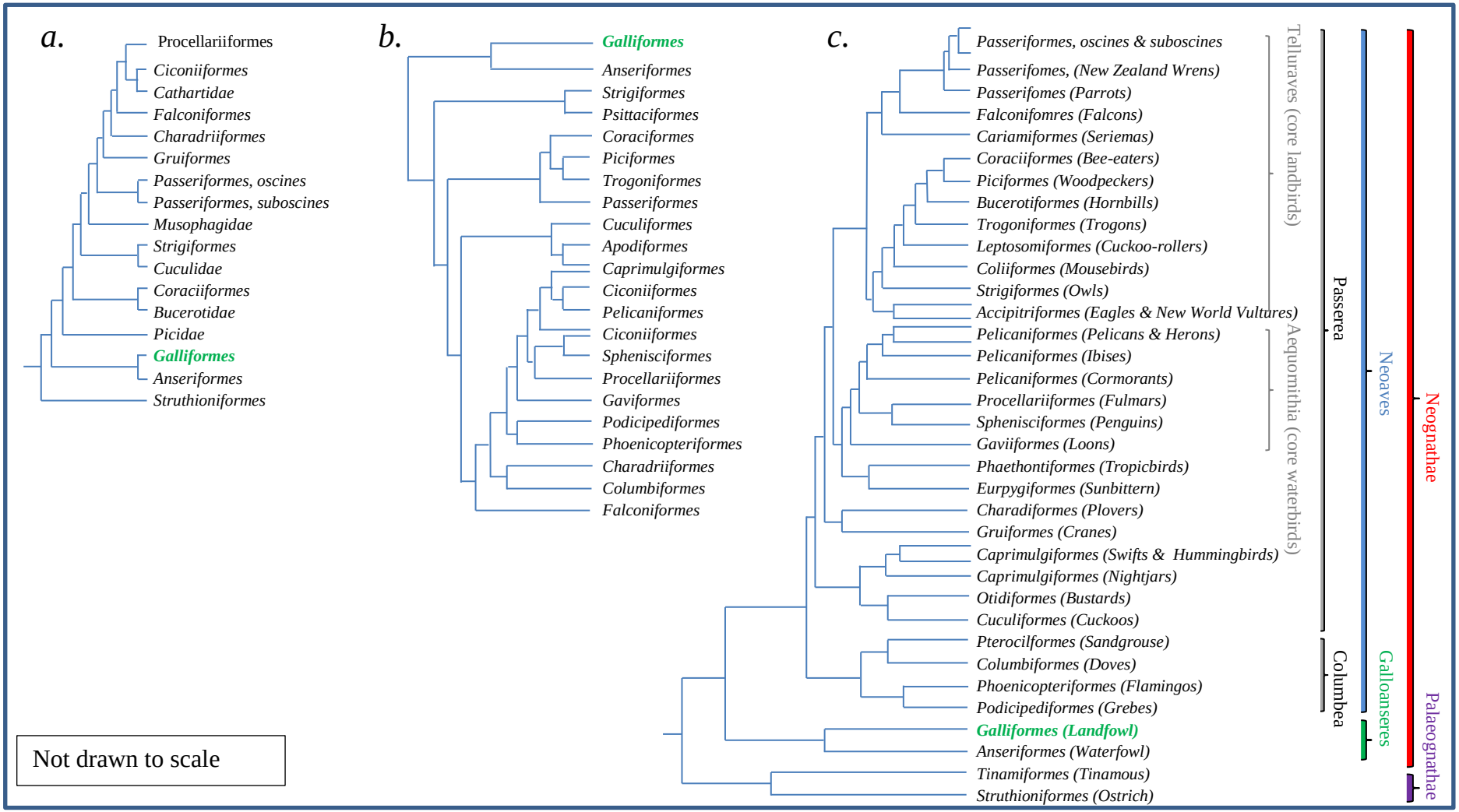


Figure 5.3: Cladograms from various molecular phylogenetic analyses of the Avian class, adapted from; **a.** Traditional Classification DNA-DNA hybridisation (Sibley & Ahlquist, 1990), **b**. Whole mitochondrial genomes (Pacheco et al., 2011), and **c.** Whole Genome data (nuclear), Next Generation Sequencing ~3.5% of average avian genome (Jarvis et al., 2014), condensed tree.

Avian Evolution & Molecular Timescales

The timing of the evolution and major diversification events of modern and paleognathe bird orders remains unresolved (Pacheco et al., 2011; Van Tuinen & Dyke, 2004). Indirect evidence from fossils is limited by the lack of complete avian fossil specimens, due to the nature of avian skeletal structure and fossilisation processes. Molecular clocks are typically used to infer the age of molecular divergence events, potentially dating ancient avian divergences within the phylogenetic tree, displacement events and ecological niche specialisation resulting from diversification. The rate of accumulated mutations is linked to a timescale inferred through the use of fossil or molecular calibration points, the accuracy of which is dependent upon the use of appropriate and well supported calibration points (Slack et al., 2006).

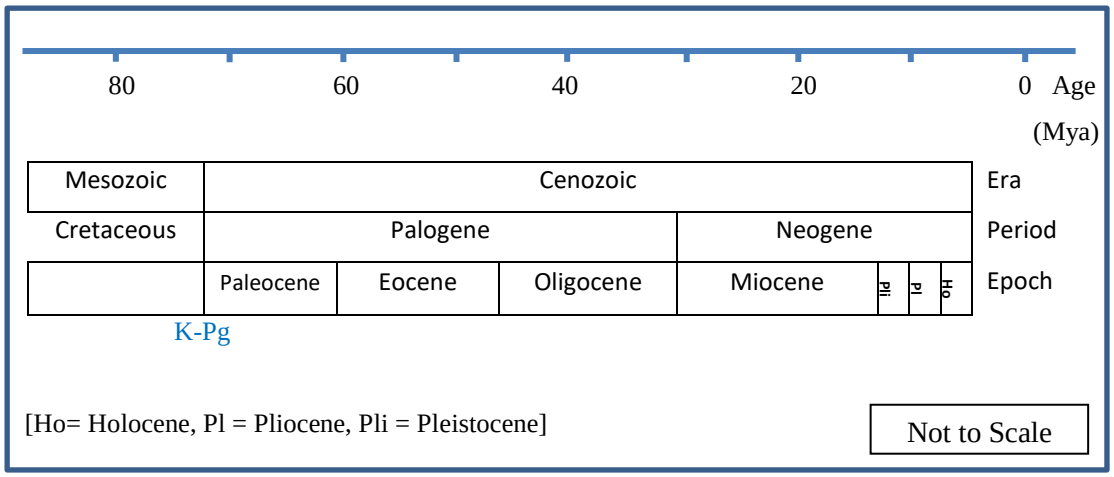


Figure 5.4: Representation of the scale of geological time, Age is represented in Millions of Years (Mya)

The main theories surrounding the radiation and diversification of the neognathe avian lineages revolve around the key period of the K-Pg boundary. Molecular data is unable to determine the extent of extinctions but may inform the biological consequences of the events of the K-Pg boundary and the subsequent diversification of extant species (Cooper & Penny, 1997). The palaeognath avians (ratites etc.) arose during the late cretaceous, genetic evidence supporting their divergence within the biogeological events of the breakup of Gondwana (Cooper et al., 2001; Haddrath & Baker, 2001). Much research supports the hypothesis outlining the mass extinction of most avian

lineages at the K-Pg boundary followed by explosive radiation and diversification of species post-K-Pg boundary. However alternate hypotheses support the diversification of avian lineages prior to the K-Pg boundary and the subsequent survival of many lineages post K-Pg, including previous molecular dating research (Cooper & Penny, 1997; Pacheco et al., 2011; van Tuinen & Hedges, 2001). Evidence from other taxonomic classes also supports the influence of the biogeographical events of the K-Pg on the diversification of species (Nilsson, Arnason, Spencer, & Janke, 2004; Nilsson, Gullberg, Spotorno, Arnason, & Janke, 2003).

Most research concurs in the estimation of the early diversification of the avian bird orders prior to the K-Pg boundary (65 Mya) (Cooper and Penny, 1997, Haddrath and Baker, 2001, van Tuinen and Hedges, 2001, Paton, Haddrath and Baker, 2002). The analysis of Pacheco et al (2011) conforms with that of Pereira & Baker (2006), Patton (2002) and Carcraft (2001), in the placement of the major orders pre-KPg with modern avians diverging in the mid-Cretaceous. This analysis supports the survival of a large number of avian lineages at the K-Pg boundary with subsequent diversification and radiation into the modern Neoaves. Much research into the estimated timescale for the diversification of modern avians conforms to the geological timeframe of the major continental divisions of Gondwana in the late Cretaceous (Carcraft, 2001; Haddrath & Baker, 2001). However the analysis of Jarvis et al., (2014) using whole genome data contradicts the popular assumptions of a major diversification pre-KPg, instead favouring a broader timing of avian diversification in the early Cenozoic to late Mesozoic (50-70 Mya).

Molecular evidence from complete mito-genome analysis of basal avians suggests the divergence of the Palaeognaths and Neognaths at around 110.1±19.8Mya (Haddrath & Baker, 2001), an older estimate than previous research established using only the Cytochrome b gene (97±12 Mya) (Härlid et al., 1997). The analysis of mito-genomes by Paton, Haddrath & Baker (2002) places the origin of modern birds at an earlier diversification time, 108.2-156.5 Mya, with most avian lineages diversifying pre-KPg.

The research of Van Tuinen and Hedges (2001) established the stem age of the Galliforme order at 89.8±7 Mya (GA_{90}), based on the Galliforme-Anseriforme divergence, a well supported node of divergence in all avian research. Van Tuinen and Dyke (2004) utilised this molecular calibration point for stem Galliformes in addition to multiple fossil calibrators to infer divergence estimates within the Galliforme order. The mito-genomic analysis of Haddrath and Baker (2001) estimates the divergence of the Galloanserae (Landfowl-Waterfowl) to have occurred at 85.4±16.9 Mya, which is consistent with the more detailed analysis of Galliforme mito-genomes (van Tuinen & Hedges, 2001). The combined morpho-behavioural and molecular cladistics of Crowe et al. (2006) supports the emergence of basal Galliformes prior to the K-Pg boundary, with subsequent continental evolution and diversification of species within the order.

Calibration with the stem Galliformes (GA_{90}) and fossils indicated a slower rate of divergence within the Galliformes resulting in crown Cracid divergence estimates at around 70 Mya, prior to the K-Pg boundary (Van Tuinen & Dyke, 2004). This estimation was found to be concordant with fossil record estimations, with the origins of the Cracids and Megapodes in the Cretaceous period, with much of the morphological diversification of the Galliformes in the Tertiary (Van Tuinen & Dyke, 2004). The analysis of molecular divergence times for the major Galliforme lineages using complete mitochondrial genomes, by Kan et al. (2010), produced a concordant estimate at around 78 Mya. However, the analysis of Kan et al. (2010) lacked representative taxa from the Cracidae family.

Pereira & Baker (2006) developed a multi-locus Bayesian estimation of molecular time scale for Galliformes, using the Cytochrome *b*, 12S rRNA and NADH dehydrogenase 2 genes. This approach, calibrated using fossil records and previous molecular timescales of Van Tuinen & Hedges (2001), inferred credibility intervals for key Galliforme orders however failed to resolve uncertainties at the genus level, particularly within the complex Phasianidae. Pereira & Baker's time estimates outlined ordinal diversification of the Galloanserae in the Cretaceous at 85-135 Mya, with crown group radiation of the Megapodidae between 77-123 Mya. Cracidae diverged from other Gallinaceous lineages in the late Cretaceous period 71-114 Mya,

concordant with van Tuinen and Hedges (2001) estimate of 76 Mya and Pereira et al.'s (2002) molecular time estimate of 90-64 Mya. The broad Bayesian credibility intervals method of Pereira and Baker (2006) accounts for uncertainties in substitution rates and branch lengths, and removed bias of anchored fossil calibration points, through the use of age priors. The morpho-molecular analysis of Crowe et al. (2006) infers an estimated divergence date of 60.2±4.7 Mya for the stem Cracidae, consistent with that of Hosner et al. (2016) which estimates the stem Cracidae divergence at ca. 47.5-68.6 Mya. The analysis of Hosner et al. (2016) comprehensively examines all current mitochondrial and nuclear sequence data available for the Cracidae order in a combined cladistics approach which uses well established fossil calibrations. Hosner et al. (2016) demonstrate evidence of the biogeographic and climatic influences on the diversification of the Cracids from ancient Galliforme lineages across the Americas. Their analysis indicates crown Cracids diversified ca. 20.5-26.0 Mya, with modern Cracid families emerging between 3.0-7.9 Mya (Guans and Curassows) congruent with more recent biogeographical events (Hosner et al., 2016).

Calibration of molecular clocks

Molecular phylogenies can be used as tools to analyse both evolutionary and ecological hypotheses. The application of a molecular timescale to phylogenetic divergences can be indicative of the age of those divergences from a common ancestor, however there are assumptions made regarding the rates of variation and evolution among groups which may infer error in the analysis. The use of multiple calibration points within the analysis reduces these erroneous effects; including fossil evidence and geographic or paleo-climatic events. Fossil evidence can provide important insights into temporal estimates of divergences among orders which may be valuable in the calibration of molecular clocks. Alternately geographical or historical events such as the age of an island or a temporal estimate for a geographic event can be used to calibrate a clock. The paucity of identifiable avian orders in the fossil record for the Cretaceous period make full molecular clock analysis problematic; a lack of late Cretaceous fossil horizons and lack of geographic sampling mean there is scarcity in the fossil evidence for this period (Cooper & Penny, 1997; Pacheco et al., 2011). Most Galliforme fossils are unsuitable for use as calibration points due to a lack of timescale

evidence or their fragmentary nature (Pereira & Baker, 2006); the age of fossils is subjective due to uncertainties in geological age of the strata within which the fossils are found. Several avian fossils which conform to the Cracidae type have been identified in the fossil record of North America (Tordoff & MacDonald, 1957), such fossils may be beneficial in the analysis of contemporary species of the order.

Fossil records from the Galloanserae are relatively well known due to the large body size and density of the bone structures, compared to other more fragile avian species which resist the fossilisation process. Suitable fossils are known for the Galliformes, Anseriformes and Cracids, including *Gallinuloides* a stem Galliforme and *Vegavis iaai* an ancestral Anseriforme (see Table 5.1). *Glainuloides* has been well established in the analysis of both Avian evolutionary relationships and more specifically Galliforme evolutionary phylogenetics (Crowe et al., 2006; Jiang, Wang, Peng, Peng, & Zou, 2014; Van Tuinen & Dyke, 2004), including analysis of the Cracidae (Hosner et al., 2016), due to its consistent molecular and palaeo-geological timeframe.

The fossil Cracid *Procrax brevipes,* which resembles the genus *Pipile,* from the late Eocene to early Oligocene period (Tordoff & MacDonald, 1957), support molecular time estimates of the fossil timescale at ~33 Mya for this stem Cracid (Pereira et al., 2002; Pereira & Baker, 2006; Van Tuinen & Dyke, 2004), see Table 5.1. The order Cracidae is therefore of importance to the calibration of Galliforme and Avian evolutionary trees, as the fossil *Procrax brevipes* is of great utility as a fossil calibration. The addition of Cracidae genome sequences will clarify the timing of the evolutionary relationships within the Galliformes and basal avian orders as they relate closely to one of the few avian fossils from the pre-Cretaceous period.

Table 5.1: Fossils & Geological events used in the calibration of avian Molecular Clocks.

Fossil name	Origin (node)	Temporal Period	Location	Reference
Procrax brevipes C_{35}	Cracidae (Pipile)	Oligocene ~33 Mya (P&B'06 35-53MYA)	South Dakota, USA	Tordoff & MacDonald, 1957; Van Tuinen & Dyke, 2004
Gallinuloides wyomingensis $P\text{-}N_{52}$	Galliforme-Anseriforme	Late Oligocene ~50 – 54 Mya (51.6-75.3)	Wyoming, USA	Tordoff & MacDonald, 1957; Crowe et al., 2006; Hosner et al., 2016; Jiang et al., 2014; van Tuinen & Dyke, 2004
Numida		>35-40 Mya		Helm-Bychowski & Wilson, 1986
Phasianoidea		<35-41 Mya		Helm-Bychowski & Wilson, 1986
Gallus bravardi $G_{14.5}$	Crown Gallus	~14.5 Mya		Lambrecht, 1993; van Tuinen & Dyke, 2004
Gallus gallus + *Gallus sonneratii*		8 Mya		He et al., 2009; Jiang et al., 2014
Palaeortyx gallica O_{40}	Numididae + Phasianidae	Late Oligocene ~35 Mya (H'16 28.0-51.7)	Enspel, Germany	Milne-Edwards, 1869; Hosner et al., 2016; Van Tuinen & Dyke, 2004
Amitabha urbinterdictensis	Phasianids-Odontophorids	50-52 Mya		Gulas-Wroblewski & Wroblewski, 2003; Crowe et al., 2006
Schaubortyx keltica $C\text{-}G_{35}$	Odontophoridae + Phasianids (Gallus)	Oligocene 32-38 Mya (H'16 23.0-47.7)	Armissan, France	Brodkorb, 1964; He et al., 2009; Hosner et al., 2016; Jiang et al., 2014; Kan et al., 2010; Van Tuinen & Dyke, 2004
Rhegminornis calobates	Meleagris + Tympanuchus	Early Miocene 18.0-41.7 Mya	Florida, USA	Hosner et al., 2016
Callipepla shotwelli	Callipepla + Colinus	Early Miocene 3.5-27.2 Mya	Oregon, USA	Hosner et al., 2016
Romainvillea	Anseriformes (Goose)	Middle Eocene ~40 Mya		Swinton, 1975
Paraortygoides GA_{52}	Galliforme-Anseriforme	Lower Middle Eocene ~52 Mya		Van Tuinen & Dyke, 2004
Quercymegspodius M_{35}	Crown Megapodidae	~35 Mya		Van Tuinen & Dyke, 2004
Vegavis iaai	Anseriforme	66 Mya		Pacheco et al., 2011
Galloanseriformes		90 Mya		Jiang et al., 2014; van Tuinen & Hedges, 2001
Waimanu sp. (Penguin fossils)	Stork + Penguin	62 Mya		Slack et al., 2006

Methodological differences in calibration may have significant influence on the determination of evolutionary divergence date estimates. Previous analyses have inferred different divergence timescales, according to the calibration techniques used in the analyses; the Anseriforme-Galliforme split has been recorded as 96 Mya by Sibley & Ahlquist, 1990 (DNA-DNA hybridisation), 68±5 Mya by Wadell et al., 1999 (12S rDNA) and 89.8 Mya by Van Tuinen & Hedges, 2001 (nuclear DNA). Pereira and Baker (2006) utilised a Bayesian credibility intervals method with an age priors approach, as opposed to fixed node ages, thereby removing bias associated with anchoring fossils at fixed nodes. This technique of age priors calculates a broader time estimate, however has a high level of confidence (95%).

The use of external calibration points for the inference of avian timescales has been demonstrated with both mammal and reptile calibration points. However, analysis using a bird-alligator calibration point revealed calibration errors associated with differing substitution rates between taxa (Paton, Haddrath and Baker, 2002). In addition the use of mammal and reptile outgroups has been shown to alter tree topology in several instances (Paton et al., 2002; Slack et al., 2007).

Study Objectives

The focus of this study is the novel inclusion of the complete mitochondrial genome of the Trinidad Piping Guan in the analysis of evolutionary relationships within the Galiforme order, and in a wider avian context. This is the first time a species of the genus *Pipile* has been incorporated into complete mito-genomic analysis of inter-specific relationships within the Galliforme order. The analysis will examine the phylogenetic placement of the genus *Pipile* using both individual mitochondrial genes and complete mitochondrial genomes, to determine a robust and reliable estimation of the relationships of the genus to other Galliformes and avian families. In addition, the analysis will examine the temporal relationships within the phylogenetic hypotheses, to establish an estimate of the temporal divergence of the genus and the Cracidae clade.

5.2 Experimental Methodology

5.2.1 Whole Mitochondrial Genome Sequence Data

Mitochodrial Genome Sequence Data

Multiple sequence alignments were performed using Clustal X 2.0 alignment software with subsequent visual verification and editing in Geneious R8. The reference sequence for the *Gallus gallus* whole mitochondrial genome (accession number X52392) was used as the standard reference point for nomenclature of all sequences. Analysis of alignments were performed in Geneious R8.0, to identify the key indicative genome features; genes, non-coding regions, initiation and termination codons, intergenic gaps and conserved features. Initiation and termination (start and stop) codons were identified by translating the sequence using the standard vertebrate mitochondrial amino acid nomenclature in Geneious R8. The identified gene features were used in subsequent data set editing.

5.2.2 Phylogenetic Analysis

Analysis of whole mitochondrial genomes and individual gene analysis was conducted using Geneious R8.0 (Kearse et al., 2012), MEGA 7.0 (Kumar et al., 2016) and MrBayes software (Huelsenbeck & Ronquist, 2001) for phylogenetic analysis. Multi-model analyses is standard in large data set analyses (Morgan-Richards et al., 2008; Shen et al., 2010), whereby use of multiple statistical approaches can be compared to enable comparison of topological reconstructions.

Whole mitochondrial genome data was collated from the NCBI database in Geneious R8.0. Data was partitioned into three distinct genome sets for separate analysis of Galliforme relationships and Avian relationships, see Table 5.2. A total of fifty-eight Galliforme mitochondrial genomes were sourced from the NCBI database for use in both the Galliforme and Avian phylogenetic reconstructions. The fifty-eight

Galliforme genomes were selected to represent a broad spectrum of species from the genus incorporating the major families (see Appendix 5, table 1). In addition, mitochondrial genomes from forty-five other Avian species were sourced to represent a broad spectrum of avian orders (see Appendix 5, table 2). Mitochondrial genomes for the Green Turtle (*Chelonia mydas*), Iguana (*Iguana iguana*) and the Mississippi Alligator (*Alligator mississippiensis*) were also sourced from the NCBI database for use as reference outgroups in the analysis.

Table 5.2: Data sets used for phylogenetic and evolutionary analysis of *Pipile pipile*

Data set	Taxon Level	Genome or gene	Alignment Length (nt)	Number of sequences	Outgroup
Galliformes Data Set	Order	Whole mtDNA (13 genes)	10,858	63	Anseriformes (n=4)
Avian Data Set	Class	Whole mtDNA (13 genes)	11,433	69	None
Avian + Outgroup Data Set	Class	Whole mtDNA (12 genes)	11,010	72	Reptiles (n=3)
Data Partitions by Individual Genes or Genome Region					
Galliforme CR Data Set	Order	Control Region	1,429	63	Anseriformes (n=4)
Galliforme CytB Data Set	Order	Cytochrome b	1,143	63	Anseriformes (n=4)
Galliforme COI Data Set	Order	Cytochrome c oxidase I	1,551	63	Anseriformes (n=4)
Galliforme ND2 Data Set	Order	ND2	1,038	63	Anseriformes (n=4)
Avian CytB Data Set	Class	Cytochrome b	1,143	72	Reptiles (n=3)
Avian COI Data Set	Class	Cytochrome c oxidase I	1,551	72	Reptiles (n=3)
Avian ND2 Data Set	Class	ND2	1,068	72	Reptiles (n=3)

The ‘Galliforme data set’ comprises fifty-nine Galliforme mitochondrial genomes and an outgroup of four Anseriforme sequences, similar to the configuration of data sets in Meiklejohn et al., (2014). The data set included the novel mitochondrial genome of the Trinidad Piping Guan reference sample (accession KU221051); produced in this study (see Chapter 3), this represents the first use of this species in complete mito-genomic phylogenetic reconstructions. Sequences in the Galliforme data set were concatenated to include only the protein-coding genes and rRNA genes, with the

removal of the tRNA genes, Control Region, Stop codons and Inter-genic nucleotides. A short region of the ND5 gene (~100bp) was removed from this data set to allow the inclusion of the mitochondrial genomes for *Aborophila brunneopectus* and *Aborophila gingica*, as the ND5 gene was incomplete and may introduce erroneous topology to the phylogenetic reconstruction. The total length of the Galliforme data set was 10,858 nucleotides.

The 'Avian data sets' incorporate a condensed Galliforme group of twenty-three genomes including the Trinidad Piping Guan and Crax genomes, with forty-six additional genomes representing a further thirty orders of the Aves class and three reptile outgroups, consistent with (Jarvis et al., 2014), see Appendix 5 table 2. This data set is designed to analyse the Avian class with a focus on the Galliforme order, to place the Galliformes in an evolutionary context. Two versions of this dataset were compiled for analysis; the 'Avian data set' consisting of the avian genomes without the reptile outgroup, and the 'Avian + Outgoup data set including the reptile outgroup.

The 'Avian data set' (no outgroup) contained a total of sixty-nine avian genomes, this data set was comprised of concatenated sequences of twelve protein coding genes and two rRNA genes, with stop codons and inter-genic gaps removed. The total length of the 'Avian data set' alignment was 11,433 nucleotides.

The 'Avian + Outgroup data set' contained the concatenated sequences of seventy-two mitochondrial genomes (including the three reptile outgroups). To attend to the issue of gene rearrangement between reptile and avian classes, and issues associated with low sequence homology between class, the ND6 gene was excluded from this analysis data set. The total length of the 'Avian + Outgroup data set' alignment was 11,010 nucleotides.

The whole genome data sets were primarily aligned in Geneious 8.0 using the ClustalW alignment algorithm. After alignment data was trimmed to exclude the

hypervariable Control Region, all tRNA genes, stop codons and inter-genic spacer nucleotides, as per Slack et al. (2003) and Pacheco et al. (2011). The elimination of this nucleotide data resulted in data sets containing only rRNA genes and protein-coding genes. Further to this in the Avian + Outgroup data set to account for gene re-arrangement problems with the Reptile out-group the ND6 gene was removed from the data set, as per Pacheco et al., (2011).

The variation in substitution rates throughout the different domains, features and blocks of the Control Region mean that molecular calibration of the region is difficult (Lucchini & Randi, 1998). In addition to this hypervariability of sites within the Control Region may heighten saturation effects in phylogenetic analysis, therefore the Control Region was excluded as a suitable region for evolutionary phylogenetics and the analysis of evolutionary timescales.

The ND3 gene in most avian orders contains a single inserted nucleotide at position 174 within the gene sequence, this insertion results in a frameshift in the translation to amino acids (Mindell, Sorenson, & Dimcheff, 1998). As a non-coding nucleotide this was removed from the analysis in all data sets.

Data sets was compiled in Geneious R8, and converted to Fasta and Nexus formats for analysis in MEGA and MrBayes respectively, manual editing of the .fas and .nex files was conducted in text editing software.

Preparation of Individual Gene Data Sets

Individual gene data sets were extracted from the original data set alignments for the complete Control Region, Cytochrome B, Cytochrome Oxidase I and ND2 genes. The complete gene data sets were re-aligned in Geneious 8.0 using ClustalW. Individual genes were analysed for both the ‘Galliforme + Outgroup data set’ and the ‘Avian + Outgroup data sets’ (see Table 5.2). The Control Region was excluded from individual

gene analysis of the 'Avian + Outgroup' data due to the high rates of variation at the higher taxonomic level and saturation effects, as previously discussed.

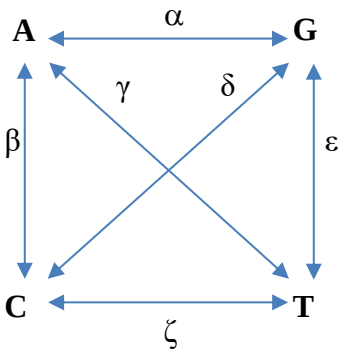
Phylogenetic re-constructions were performed using MEGA and MrBayes from these complete gene alignments, for comparison with the whole mitochondrial genome data sets to analyse the robustness of the use of individual genes compared to whole mitochondrial genomes at the Order and Class taxon levels. Individual gene trees were compared to the whole genome data set trees, to analyse of the accuracy and robustness of the use of single genes in comparison to whole genomes at several taxonomic levels. Comparison was made using the statistical support values for key nodes within the trees which support the various clades and inter-ordinal relationships.

Phylogenetic Models

Prior to Phylogenetic analysis in either MrBayes or MEGA software packages each data set was analysed in Modeltest to determine the most appropriate substitution model for the data.

General Time Reversible model (GTR)

The model designated in model tests for all whole genome data sets (Galliforme and Aves data sets) was the GTR+I+Γ or General Time Reversible model (Lanave et al., 1984), with a proportion of Invariable sites (I) and Gamma shaped distribution of rates across sites (Γ). The GTR model dictates that all substitution rates are different and base frequencies may be unequal ($\pi A \neq \pi C \neq \pi G \neq \pi T$), and therefore all possible substitutions have different probabilities. The GTR model has six parameters each of which has its own substitution probability, see Figure 5.5.



$$P_t = \begin{bmatrix} . & \alpha\pi G & \beta\pi C & \gamma\pi T \\ \alpha\pi A & . & \delta\pi C & \varepsilon\pi T \\ \beta\pi A & \delta\pi G & . & \zeta\pi T \\ \gamma\pi A & \varepsilon\pi G & \zeta\pi C & . \end{bmatrix} .$$

Figure 5.5: General Time Reversible substitution model matrix (Felsenstein, 2004; Nei & Kumar, 2000; Page & Holmes, 1998; Yang, 2014).

The MEGA and MrBayes software set parameter priors as default under the GTR substitution model criteria; topology, branch length (clock constraints), stationary frequencies of the nucleotides, nucleotide substitution rates, proportion of invariable sites and α shape parameter of the gamma distribution.

Maximum Likelihood Analysis in MEGA7.0

Molecular Evolutionary Genetics Analysis (MEGA) was used to estimate phylogenetic relationships under the Maximum Likelihood model, using Bootstrap re-sampling to provide statistical support for nodes (Felsenstein, 1985). Maximum Likelihood reconstructions, as previously discussed in Chapter 4.2, infer trees using the highest likelihood of branch lengths to inform topology. Phylogenetic analysis was performed under the Maximum Likelihood method using the General Time Reversible substitution model with Gamma distribution and Invariant sites (GTR+I+Γ).

The Maximum Likelihood tree was analysed using 1000 Bootstrap re-sampling replicates to provide statistical support for branch nodes. Bootstrap values are presented at the nodes as a percentage, bootstrap values below 75% were excluded from the trees. Bootstrap confidence values of 95% or more infer confidence in the node position (statistically significant), and therefore topology.

Bayesian Inference Analysis in MrBayes 3.2.6

The MrBayes phylogenetic software package is a Bayesian Inference analysis method. Bayesian methods are advantageous in the analysis of large data sets, and datasets where there is a temporal influence (past priors affect future distributions) (Drummond & Bouckaert, 2015). MrBayes analyses were performed on the CIPRES Science gateway (Di Fiore et al., 2014), due to the computational requirements of running multiple chains on large data sets.

Analysis in MrBayes was conducted using the Marcov Chain Monte Carlo statistical analysis method/algorithm, as previously discussed (Chapter 2.10) the MCMC uses sampling from prior distributions to construct the 'best' tree through a heuristic approach, creating a consensus tree using the greatest probabilities of a given node occurring at a specific position within a topology.

The number of generations and sampling parameters were determined through multiple analyses attempts, using generated data to inform further analyses. Data sets were run with between 1,250,000 to 50,000,000 generations, depending on the size of the data set and the complexity of relationships between taxa sequences. Initial analyses were conducted using 1.25 million generations, increasing as necessary according to the results of the statistical support for the tree reconstructions and convergence. The General Time Reversible model with Gamma distributions and Invariant sites was selected for the reconstruction of all trees in MrBayes under the MCMC method. The MrBayes default setting for number of simultaneous analyses was used (Nruns = 2); two independent analyses were performed simultaneously starting from different random trees until convergence is achieved. Convergence is determined through posterior probability distribution comparison of the two independent sample trees. Trees in the MCMC run were sampled and compared every 1000^{th} generation. Burn-in values for all analyses were set at 25% as standard, to discard the first 25% of generations. Chains were sampled every 500 generations as standard; diagnostics including log likelihood values of the chains and split frequencies recorded in the output log. Analysis is completed (or stopped) when the

Average Standard Deviation of Split Frequencies approaches zero, at which point the two tree samples (runs) converge. If on completion of the analysis the average standard deviation of split frequencies is greater than 0.05, further analysis is required with an increased generation sample size. Standard deviations between 0.01 and 0.05 indicate adequate convergence, standard deviations below 0.01 are a good indication of convergence and analysis may be stopped. Support for clades within the tree topology is indicated by posterior probabilities below 0.95, indicative of low support and possible error in clade structure. Log likelihood values are indicative of the completion point of the analysis, the analysis may be stopped once the log likelihood values stabilise (also indicated by a lack of trends evident in the log likelihood plot for the two runs). Increasing or decreasing trends in the log likelihood data indicate increased generation sampling is required.

Bayesian Evolutionary Analysis by Sampling Trees in BEAST 2.2

Data sets (Nexus format) were analysed in the BEAST 2.2 software (Drummond et al., 2012), to establish evolutionary relationships and molecular timescales. BEAST 2.2 uses the Marcov Chain Monte Carlo algorithm (MCMC) for Bayesian inference to characterise posterior distributions for time-tree construction. Marcov Chain Monte Carlo parameters and treatments were determined through the repeated analysis of each data set to achieve appropriate posterior values.

BEAST log files of the data sets were analysed in Tracer v1.6.0 (Drummond et al., 2012), to determine quality of the analysis. Tracer v1.6.0 provides statistical analysis of the data logs for MCMC analysis including Effective Sample Size analysis, 95% HPD posterior probabilities, and Marginal Distributions. Estimated Sample Size (ESS) of the analysis must exceeded 100, ESS values below 100 indicate the parameter is undersampled therefore an increase in generations is required. Posterior Probabilities of each bipartition between two taxa is analysed and sorted according to probability, in relation to the number of times the bipartition split is observed in all runs, and therefore the posterior probability of the bipartition (Split Frequency). Average Standard Deviation of Split Frequencies, should approach 0.0 for all bipartitions as the

MCMC runs converge. The convergence diagnostic, Potential Scale Reduction Factor (PSRF) should approach 1.0 as the MCMC runs converge (consensus is achieved). PSRF values above 1.0 indicate the data is under sampled therefore an increase in generations is required. Branch and Node Parameters of the reconstruction represent the 95% Highest Posterior Density (HPD) of branch length and nodes is equivalent to a 95% confidence interval for node support. The informative taxon bipartitions probability, ie. node probabilities, and PSRF should approach 1.0 as runs converge.

Fossil and molecular dates were used to inform priors in BEAST 2.2 evolutionary and molecular timescale analyses, see Table 5.3 for priors used in analyses. Calibration of the molecular clock correlates node height with an absolute timescale resulting in a time-tree (Drummond & Bouckaert, 2015). Calibration densities are compiled based on the geological timescale (age) of a fossil, representing the minimum and maximum ages of the fossil to account for dating error in the geological record. Molecular calibration points were selected from literature, to represent molecular age of a specific node within the topology.

Table 5.3: Priors included in analysis in BEAST 2.2 (BEAUti) Fossil and Molecular Calibration points.

Fossil / Species Prior	Minimum Date	Maximum Date	Type	Use	Reference
Galinuloides wymoingensis	51.6	75.3	Fossil prior	Avian data set Galliforme +outgroup data set	Hosner et al., 2016
Palaeortyx gallica	28.0	51.7	Fossil prior	Avian data set Galliforme +outgroup data set	Hosner et al., 2016
Neognathes	110.1±18.9		Molecular Prior	Avian data set	Haddrath & Baker, 2001
Procrax brevipes	44.0±6.0 (33-53)		Molecular Prior	Galliforme +outgroup data set	Pereira & Baker, 2006; Van Tuinen & Dyke, 2004

Phylogenetic and Molecular Clock Trees

Phylogenetic trees for Maximum Likelihood analyses were produced and edited within the MEGA 7.0 software package. Bootstrap support for nodes were added to tree diagrams (nodes with support lower than 75% were excluded).

Consensus Trees inferred by analysis in MrBayes (with a posterior probability of 95%) were visualised in FigTree v1.4.3 software. Posterior probability support values for nodes were added to the tree diagram, either as probabilities (0.0-1.0) or converted to the percentage equivalent (0-100%).

Consensus Trees inferred by analysis in BEAST were condensed in TreeAnnotator (BEAST software) and visualised in FigTree v1.4.3 software. Timescale legends were applied to tree diagrams in FigTree, and posterior probabilities applied to the nodes. Multiple individual tree sets produced by BEAST were visualised in DensiTree v2.2.5 (Bouckaert & Heled, 2014).

5.3 Results

5.3.1 Compositional Analysis of the Whole Mitochondrial Genome

The mitochondrial genome of the Trinidad Piping Guan conforms in both structure and gene order with that of other Galliforme avians, including the *Gallus gallus* reference sequence of Desjardin and Morais (1990). Sequences were determined for three individuals of the Trinidad Piping Guan species, as discussed in Chapter 3, of which all compositional features were identical for all individuals with only minimal variations observed between individuals. The Trinidad Piping Guan mitochondrial genome is consistent in structure with the standard avian genome structure, as per Mindell et al. (1998), a general structure which is observed in the majority of avian species (Macey et al., 1997; Mindell et al., 1998).

5.3.2 Individual Genes

The identification of genomic structure and gene characteristics is important in the analysis of genomic level data sets; non-coding nucleotides, protein-coding genes, start and stop codons. Translation of the mitochondrial genome sequences were determined in Geneious R8, for the isolation of the gene sequence data sets. The non-coding Control Region was identified using comparison with the standard reference sequences and the conserved features of the region including Conserved Sequence Blocks (CSBs) and structural features, as presented for Cracids by Pereira, Grau, & Wajntal (2004). The individual Cytochrome *b*, Cytochrome oxidase 1 and ND2 genes were isolated from the complete genomes through identification of the gene features in alignment with reference sequences.

5.3.3 Phylogenetic Analysis

MrBayes output files of the data sets were analysed to determine the quality of the analysis, and where necessary analyses treatments were modified to produce the final analysis. Statistical analysis of the generation logs for the MCMC analysis was recorded in the output files.

5.3.3.1 Galliforme Data Analysis

Individual Gene Analysis – Galliforme and outgroup Data Sets

The individual protein-coding genes Cytochrome c oxidase subunit 1, Cytochrome *b* and NADH dehydrogenase subunit 2 were individually analysed using the Galliforme plus outgroup data set of 59 Galliforme species and four Anseriforme outgroup species. In addition the non-coding Control region was also analysed using the Galliforme plus outgroup data set.

Each gene region data set was tested individually in jModeltest to determine the most appropriate model of substitution for the Galliforme plus outgroup data sets. It was determined that the General Time Reversible model with Gamma shape and Invariant sites parameters (GTR+I+Γ) is the most appropriate substitution model for all data sets.

Cytochrome b

The Cytochrome *b* (CytB) gene produced an alignment of 1143 nucleotides, of which 584 nucleotide sites (51.1%) were conserved across orders, 559 sites being variable (48.9%). Average base compositions within the Galliforme order were 29.0%, 33.7%, 11.5% and 25.8% for A, C, G and T respectively, demonstrating high C content and deficit G content. Phylogenetic reconstructions using both the Maximum Likelihood and Bayesian Inference methods for the CytB gene are presented in Appendix 6, Figure 16. The CytB gene analysis of the Galliforme data demonstrate the complex relationships of the Galliforme group. The Cracids present as a sister group to the Megapods (80-100% support) at the base of the Galliforme tree. The Numididae (Guineafowls) and Odontophoridae (New World Quails) families are basal to the remaining groups within the tree, which are combined in a paraphyletic Phasianidae (Pheasants & Partridges) clade which includes the Tetraonidae (Grouse), Meleagridae (Turkeys) and Coturnicinae (Old World Quails).

Cytochrome oxidase subunit 1

The Cytochrome oxidase c subunit 1 (COI) gene produced an alignment of 1551 nucleotides, of which 923 nucleotide sites (59.5%) were conserved across orders, 628 sites being variable (40.5%). Average base compositions within the Galliforme order were 26.9%, 30.1%, 16.4% and 26.6% for A, C, G and T respectively. Phylogenetic reconstructions using both the Maximum Likelihood and Bayesian Inference methods for the COI gene are presented in Appendix 6, Figure 17. Phylogenetic reconstruction using the COI gene demonstrate, in contrast to CytB analysis, a monophyletic Cracidae branch with over 70% support and a monophylectic Megapodidae (100%) branch. The Megapodidae is basal to all other families in the Galliforme tree. The remaining families including the Phasianids, Numididae and Odontophoridae remain unresolved with the main clades branching within the Pahsianidae clade.

NADH dehydrogenase subunit 2

The NADH dehydrogenase subunit 2 (ND2) gene produced an alignment of 1038 nucleotides, of which 400 nucleotide sites (38.5%) were conserved across orders, 638 sites being variable (61.5%). Average base compositions within the Galliforme order were 31.5%, 34.8%, 9.2% and 24.4% for A, C, G and T respectively, demonstrating distinctly low G content and high C content. Phylogenetic reconstructions using both the Maximum Likelihood and Bayesian Inference methods for the ND2 gene are presented in Appendix 6, Figure 18. The ND2 gene tree topology in contrast indicates monophyletic Cracidae and Megapodidae families with high degrees of support (89-100%), with the Megapodidae basal within the tree, additionally the Numididae and Odontophoridae families are monophyletic branches basal of the remaining Phasianidae clade (which includes Meleagridae and Tetraonidae).

Control Region

The Control Region (CR) produced an alignment of 1429 nucleotides, of which only 328 nucleotide sites (23.3%) were conserved across orders, 1101 sites being variable (76.7%). Average base compositions within the Galliforme order were 25.9%, 26.6%, 14.2% and 33.3% for T, A, C and G respectively, demonstrating a distinctly high T content and high A-T bias (59.2%). Phylogenetic reconstructions using both the Maximum Likelihood and Bayesian Inference methods for the CR are presented in

Appendix 6, Figure 19. Similar to the ND2 and COI gene trees the CR infers monophyletic Cracidae and Megapodidae families with high levels of support (98-99%), with Megapodidae basal in the Galliforme tree. The Numididae and Odontophoridae are represented in a sister clade, which is basal of other Phasianids.

Whole Mitochondrial Genome Analysis – Galliforme Data Set

The Galliforme data set analysed 13 protein-coding genes within the mitochondrial genome of 59 Galliforme species, plus four Anseriforme outgroups (n=63). The Galliforme plus outgroup data set (Galliforme+out n=63) produced a total alignment sequence length of 10, 858 nucleotides, of which 5,071 nucleotide sites (46.7%) were conserved across orders, 5,787 sites being variable (53.3%). Average base compositions within the Galliforme order were 29.3%, 33.8%, 12.3%, 24.6% for A, C, G and T respectively. Bayesian Inference analysis in MrBayes reconstructed phylogenetic topology using the GTR+I+Γ model, sampling consisted of 25 million generations sampled every 1000 generations, discarding the first 25% of the run as burn-in, with posterior probabilities to support node topologies, see Figure 5.6. The Maximum Likelihood analysis conducted in MEGA 7.0 reconstructed phylogenetic relationships among the Galliforme order and Anseriforme outgroup, using the non-uniform evolutionary rates with Gamma distribution and Invariable sites (GTR+G+I) model, as shown in Figure 5.7. Bootstrap replications were conducted to demonstrate support for bifurcating nodes (500 replicates).

Both Maximum Likelihood and Bayesian analyses produced a phylogenetic topology where the Megapodidae and Cracidae have a sister relationship in the basal clade of the Galliformes, with 100% support. The Numididae and Odontophoridae families are monophyletic branches, basal to the remaining Phasianidae, Tetreonidae and Meleagridae clade. As with the individual gene trees, the Phasianidae family is divided into distinct clades interspersed with the Turkey, Grouse and Old World Quails. The divisions in the Phasianidae represent clear morphological clades, which relate to the Pheasants, Partridges, Peafowl and Junglefowl clades. High statistical support is present throughout the Galliforme whole mito-genome tree, nodes supported at between 91 and 100% throughout.

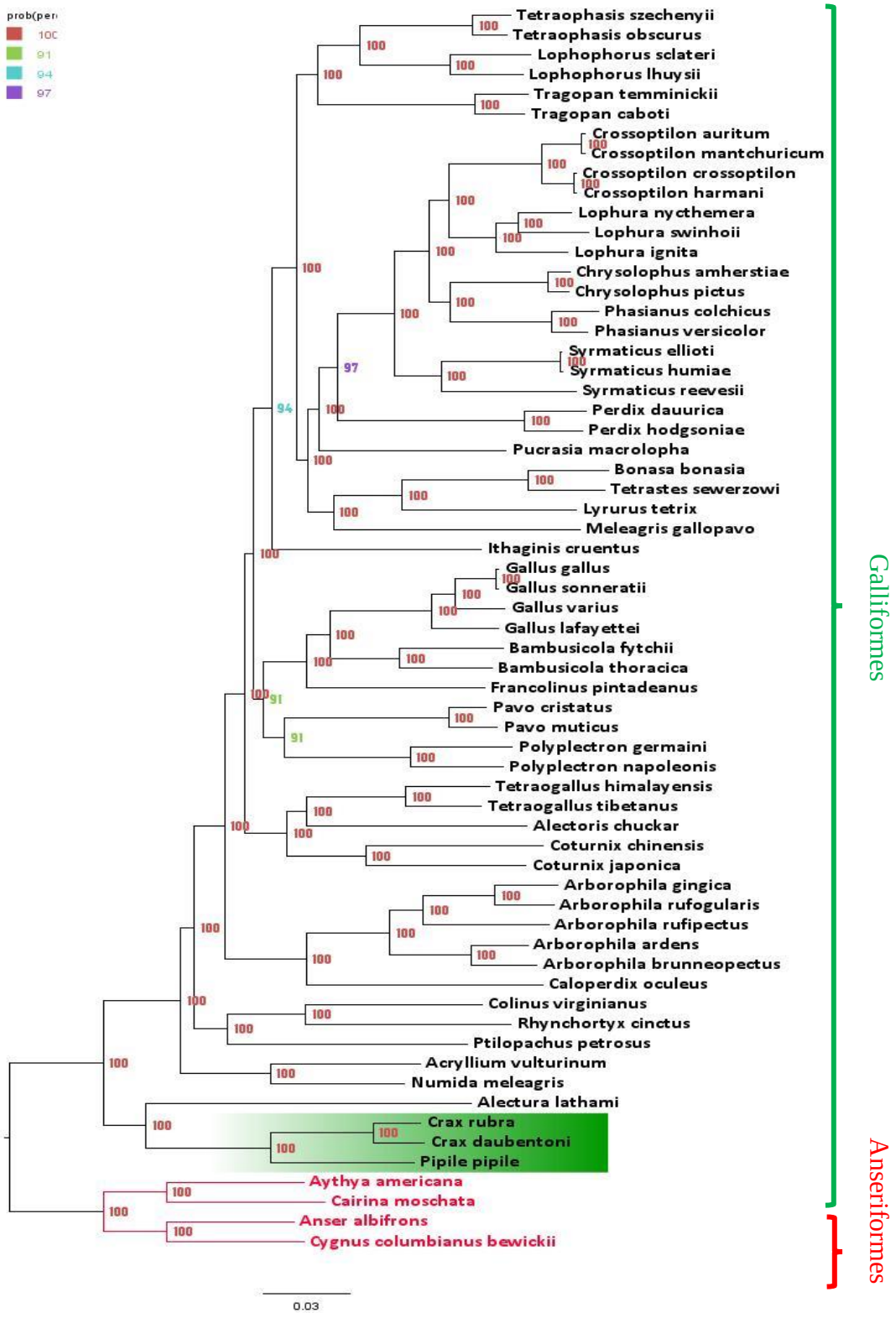


Figure 5.6: Phylogenetic relationships among Galliformes based on analysis of 10,858bp of Whole Mitochondrial Genome data (13 protein-coding genes) from 59 genera of the Galliforme order including the novel *Pipile pipile,* and 4 Anseriforme genera as a putative outgroup. Topology and branch lengths are based on General Time Reversible substitution models under non-uniform evolutionary rates with Gamma distribution and Invariable sites (GTR+G+I), through independent analysis of Bayesian inference in MrBayes. Bayesian analysis used MCMC with 25 million generations sampled every 1000 generations, discarding the first 25% of the run as burn-in, with posterior probabilities to support node topologies. Cracidae are highlighted in green.

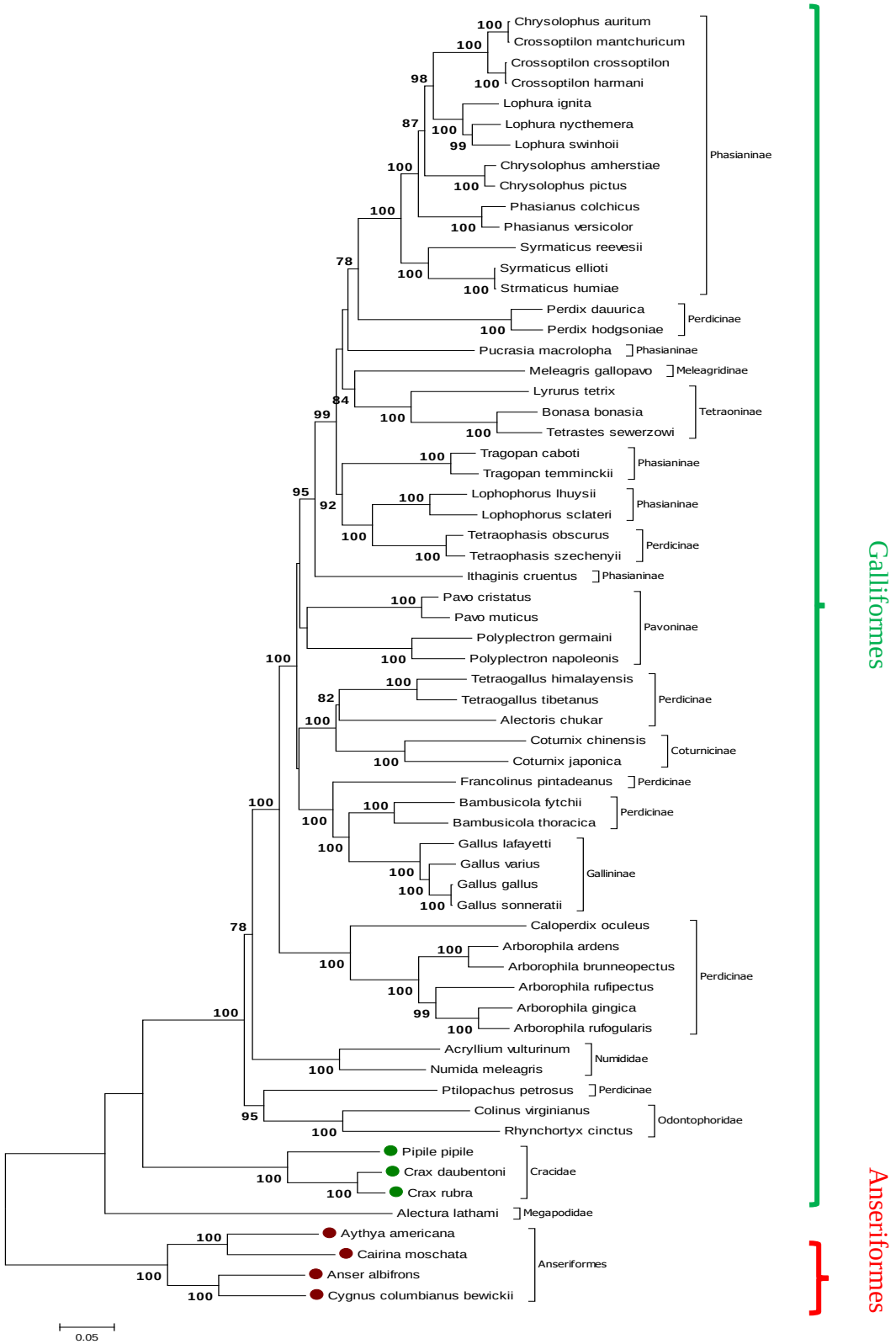


Figure 5.7: Phylogenetic relationships among Galliformes based on analysis of 10,858bp of Whole Mitochondrial Genome data (13 protein-coding genes) from 59 genera of the Galliforme order including the novel *Pipile pipile*, and 4 Anseriforme genera as a putative outgroup. Topology and branch lengths are based on Maximum Likelihood, General Time Reversible substitution models under non-uniform evolutionary rates with Gamma distribution and Invariable sites (GTR+G+I), analysis performed in MEGA7.0 with 1000 Bootstrap replicates for node statistical support (numbers at nodes in %, in bold), only nodes with >75% statistical support displayed. Cracidae are highlighted in green.

5.3.3.2 Avian Data Analysis

Whole Mitochondrial Genome Analysis

The Avian data set analysed 13 protein-coding genes within the mitochondrial genome of 68 avian species. The Avian data set (Avian n=68) produced a total alignment sequence length of 11,433 nucleotides (sequence lengths of 11,343bp±8bp), of which 3061 nucleotide sites were conserved across orders, 8372 sites being variable (73.2%). Average base compositions within the Aves class were 29.7%, 33.9%, 12.4% and 24.0% for A, C, G and T respectively. Bayesian Inference analysis in MrBayes was reconstructed for the Avian data set using the GTR+I+Γ model, sampling consisted of 25 million generations sampled every 1000 generations, discarding the first 25% as burn-in, with posterior probabilities to support node topologies, see Figure 5.8. The Maximum Likelihood analysis was conducted in MEGA 7.0 using the GTR+I+Γ model with 500 bootstrap replications to support node topologies, as shown in Figure 5.9.

Analysis of the Avian data set (without a reptile outgroup) by both Maximum Likelihood and Bayesian Inference methods roots the tree with the Palaeognathe clade (Ratites), with a high probability support for the nodes (99-100%). Analysis of both tree methods also branch the Galloanserae clade monophyletically, in a sister relationship to all other Neoaves (modern birds), with a high degree of support (95-100%). However, within the Galliforme family the Cracidae-Megapodidae relationship remains unresolved. Using Bayesian Inference analysis the Megapods are basal and both the Cracids and Megapods are monophyletic (100%). However in Maximum Likelihood analysis the Cracidae and Megapodidae branch in a sister clade, basal to other Galliformes (89%). Relationships within the Neoaves clade are also unresolved, several groups demonstrate polyphyletic relationships (Pelicaniformes, Ciconiformes and Passeriformes), and low statistical support for nodes is evident throughout. Bayesian Inference analysis of the Neoaves failed to resolve relationships clearly, resolving a paraphyletic topology, inconsistent with Maximum Likelihood analysis.

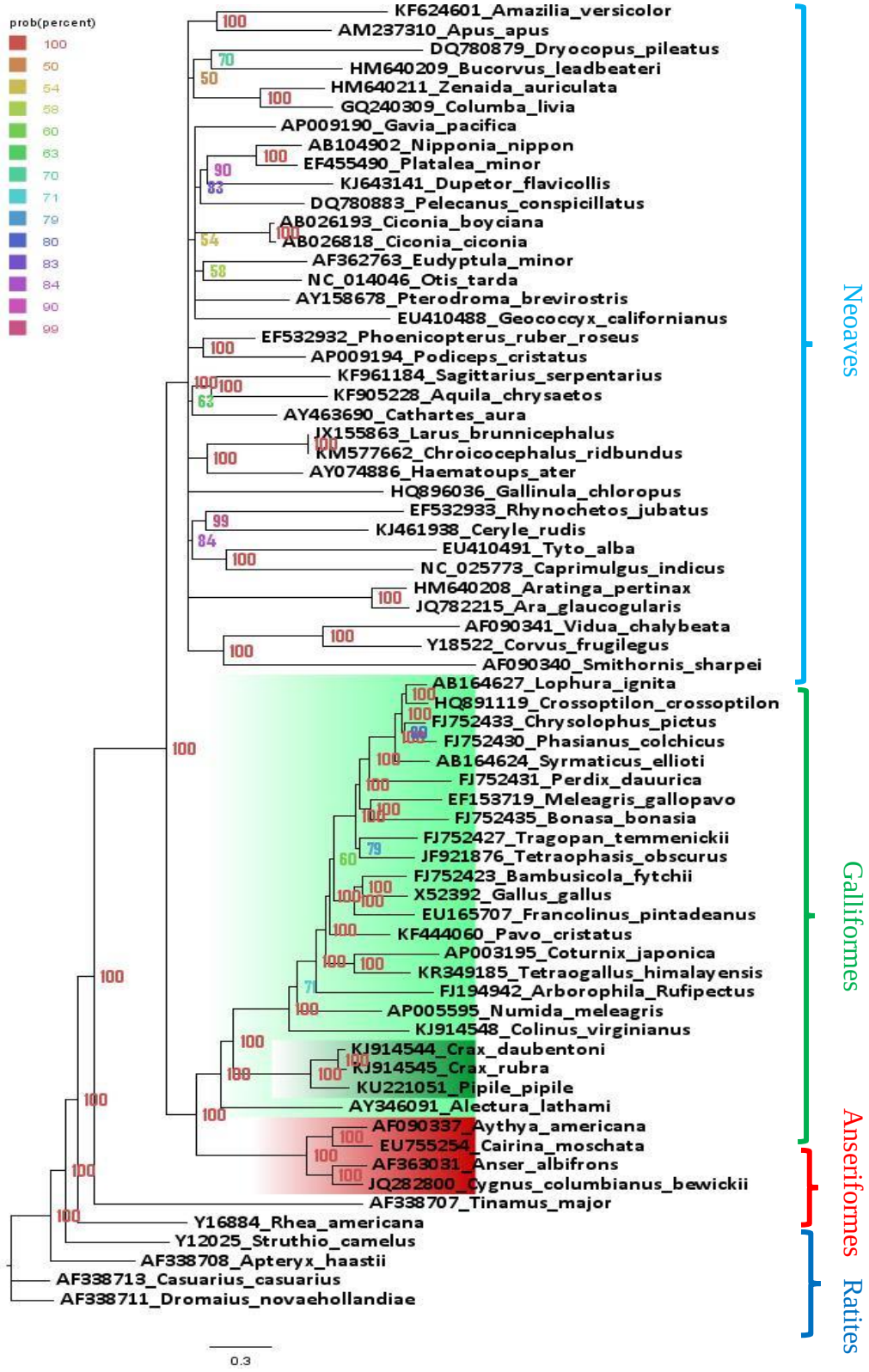


Figure 5.8: Phylogenetic relationships among Avians based on analysis of 11,433bp of Whole Mitochondrial Genome data (13 protein-coding genes) from 68 genera across the Aves class; 23 of the Galliforme order including the novel *Pipile pipile*. Topology and branch lengths are based on Maximum Likelihood General Time Reversible substitution models under non-uniform evolutionary rates with Gamma distribution and Invariable sites (GTR+G+I), through independent analysis of Bayesian inference in MrBayes. Bayesian analysis used MCMC with 25million generations sampled every 1000 generations, discarding the first 25% of the run as burn-in, with posterior probabilities to support node topologies. Cracidae are highlighted in dark green.

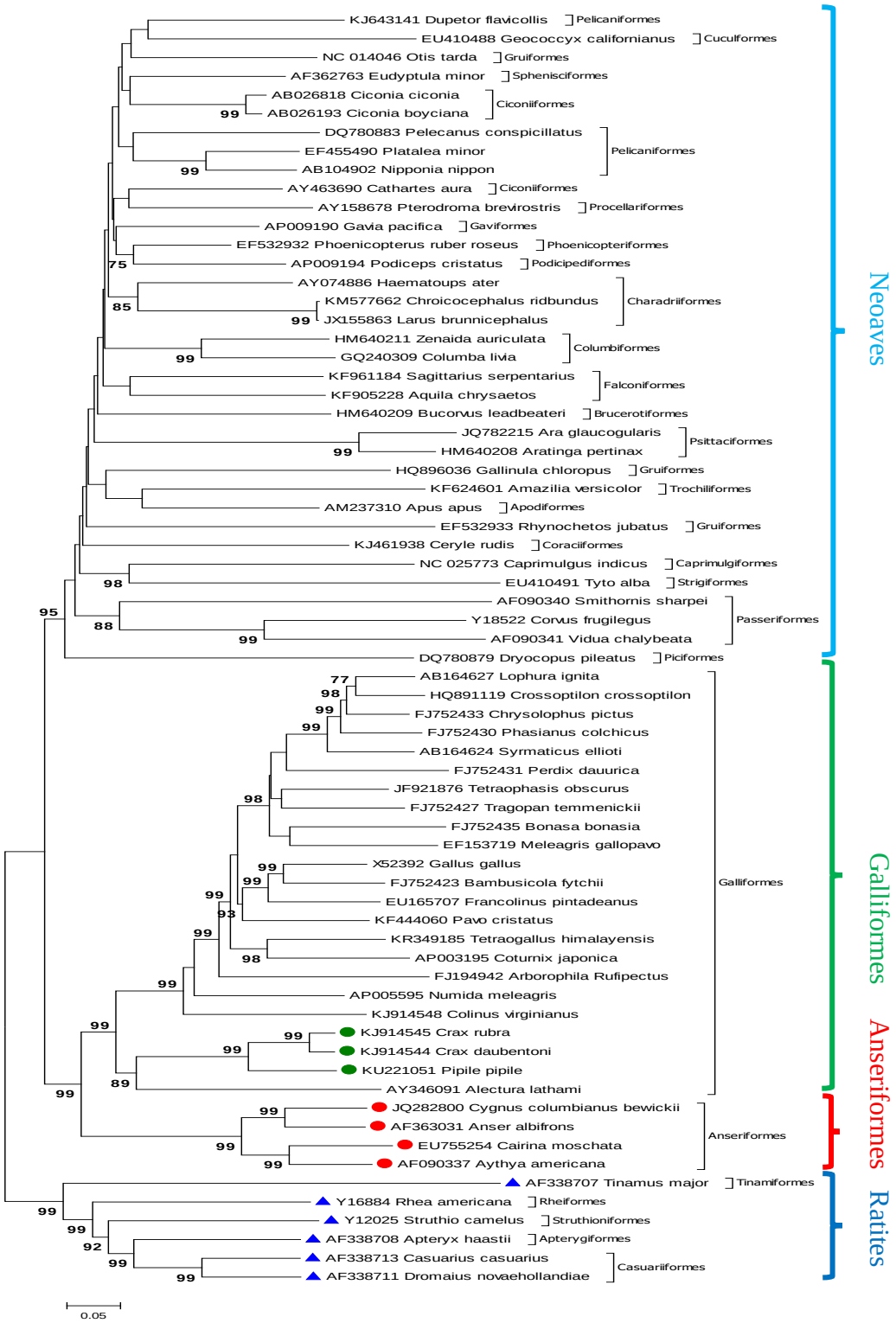


Figure 5.9: Phylogenetic relationships among Galliformes based on analysis of 11,433bp of Whole Mitochondrial Genome data (13 protein-coding genes) from 68 genera across the Aves class; 23 of the Galliforme order including the novel *Pipile pipile*. Topology and branch lengths are based on Maximum Likelihood General Time Reversible substitution models under non-uniform evolutionary rates with Gamma distribution and Invariable sites (GTR+G+I), in MEGA7.0 with Bootstrap statistical support. Maximum Likelihood analysis in MEGA with 500 bootstrap replicates for node statistical support (numbers at nodes in %, in bold), only nodes with >75% statistical support displayed. Cracidae are highlighted in green.

5.3.3.3 Avian with Outgroup Data

Individual Gene Analysis

The Avian plus outgroup data set analysed using the individual protein-coding genes COI, CytB and ND2 within the mitochondrial genome of 68 avian species and three Reptile/Squamata outgroup species. The Control Region was not analysed for this data set due to the significant bias inferred by inter-species heterogeneity within this region, analysis of broad taxonomic class data sets is difficult with the Control Region of the mitochondrial genome.

Cytochrome b

The Cytochrome *b* (CytB) gene produced an alignment of 1143 nucleotides, of which 523 nucleotide sites (45.8%) were conserved across orders, 620 sites being variable (54.2%). Average base compositions within the Aves class were A 27.7%, C 35.0%, G 12.5% and T 24.8%, for the CytB gene, demonstrating a distinct C bias and G deficit. The CytB gene tree, for the Avian plus outgroup data set, supports a sister relationship between the Cracidae and Megapodidae with 88% support, branching basally within the Galliforme order. The Galliforme and Anseriforme orders are well defined branching as a Galloanserae clade which is well supported by both methods of analysis (100%). However there are discrepencies between Bayesian Inference and Maximum Likelihood analyses in the deeper branches of the Avian CytB gene tree. The Bayesion Inference supports the Galloanserae clade is basal to all other Neoaves (modern avians), within the Neognathe group, and the division of the Palaeognathe and Neognathe as clear, monophyletic sister clades with high support for nodes (100%). However, the Maximum Likelihood method places the Galloanserae clade sister to the Palaeognathes (Ratites), with low support 53%, this Galloanserae-Ratite clade branching within the Neognathes.

Cytochrome oxidase c subunit 1

The Cytochrome oxidase c subunit 1 (COI) gene produced an alignment of 1548 nucleotides, of which 864 nucleotide sites (55.8%) were conserved across orders, 684 sites being variable (44.2%). Average base compositions within the Aves class were 27.1%, 30.6%, 16.3% and 25.9% for A, C, G and T respectively across the COI gene.

The COI gene tree for the Avian data supports the monophyly of the Cracidae and basal monophyly of the Megapodidae, with limited node support (65% and 50% respectively). The Galliforme order is well defined, being closely allied with the Anseriformes in A highly supported Galloanserae clade (100%), positioned basal in the Neognathes in relation to other Neoaves species. The division of the Neognathes (modern avians) from the Palaeognathes (Ratites), is well supported (99%) by the COI gene analysis.

NADH dehydrogenase 2

The NADH dehydrogenase subunit 2 (ND2) gene produced an alignment of 1,068 nucleotides, of which 235 nucleotide sites were conserved across orders, 833 sites being variable (67.6%). Average base compositions within the Aves class were 31.6%, 34.5%, 9.8% and 24.0% for A, C, G and T respectively. The analysis of the Avian plus Outgroup for the ND2 gene supports monophylies of the Cracidae and Megapodidae families, with the Megapods basal in the Galliforme order, with high levels of support (100%). The Galloanserae clade is well supported (96%) with a basal position in the Neognathes, sister to the remaining Neoaves clade. However the relationship between the Palaeognathes and Neognathes remains unresolved by the analysis of the ND2 gene (both Bayesian Inference and Maximum Likelihood methods).

Whole Mitochondrial Genome Analysis

The Avian plus outgroup data set analysed 12 protein-coding genes within the mitochondrial genome, ND6 was excluded due to high degrees of heterogeneity within the gene and its position on the Heavy strand within the genome, all other genes being encoded to the Light strand. The Avian plus outgroup data set (Avian+Out n=71) produced a total alignment sequence length of 11,010 nucleotides (sequence lengths of 10,826bp±15bp), of which 3822 nucleotide sites were conserved across orders, 7079 sites being variable (64.29%). Average base compositions within the Aves class were 24.6%, 29.1%, 33.8% and 12.5% for T, A, C and G respectively. Base composition within the Reptile/Squamata outgroup being 25.2%, 31.5%, 31.6% and 11.8% for T, A, C and G respectively, over an average sequence length of 10828bp±24bp.

The analysis of the Avian plus outgroup data set by Bayesian Inference methods in MrBayes (Figure 5.10), reconstructs a highly supported topology amongst avian species (70-100%). The diversification of the Neognathes from Paleognathes is clearly defined and highly supported (100%), with the Galloanserae monophyly basal amongst the modern avian Neognathes. High support of diversifications is observed throughout the Galliforme order, the Megapodidae being the basal monphyletic clade (100%) and the Cracidae branching subsequently in a monophyletic family (100%) basal of the remaining Galliforme families. Within the Neoaves the relationships are less well defined, and support for topology being lower. Maximum Likelihood analyses of the Avian plus outgroup data set (Figure 5.11) demonstrates a well supported diversification between the Neognathes and Palaeognathes (100%). The Galloanserae clade is basal amongst the Neoavian orders (100%), however within the Galliforme order the Cracidae and Megapodidae demonstrate a sister relationship with moderate support (87%) as opposed to monophyletic branches.

In both models of phylogeny the relationships within the Neoaves are not well supported by either Bootstrap or Posterior probabilities. The relationships between species in the Neoves remains unresolved in several orders, polyphyly being observed in a number of orders including the Gruiformes, Passeriformes, Ciconiformes and Pelicaniformes. Additionally, in comparison to Avian trees constructed without a Reptile outgroup (see Figures 5.25 and 5.26, Appendix 6) the relationships within the Neoaves are inconsistent and support for nodes are reduced when the reptile outgroup is removed.

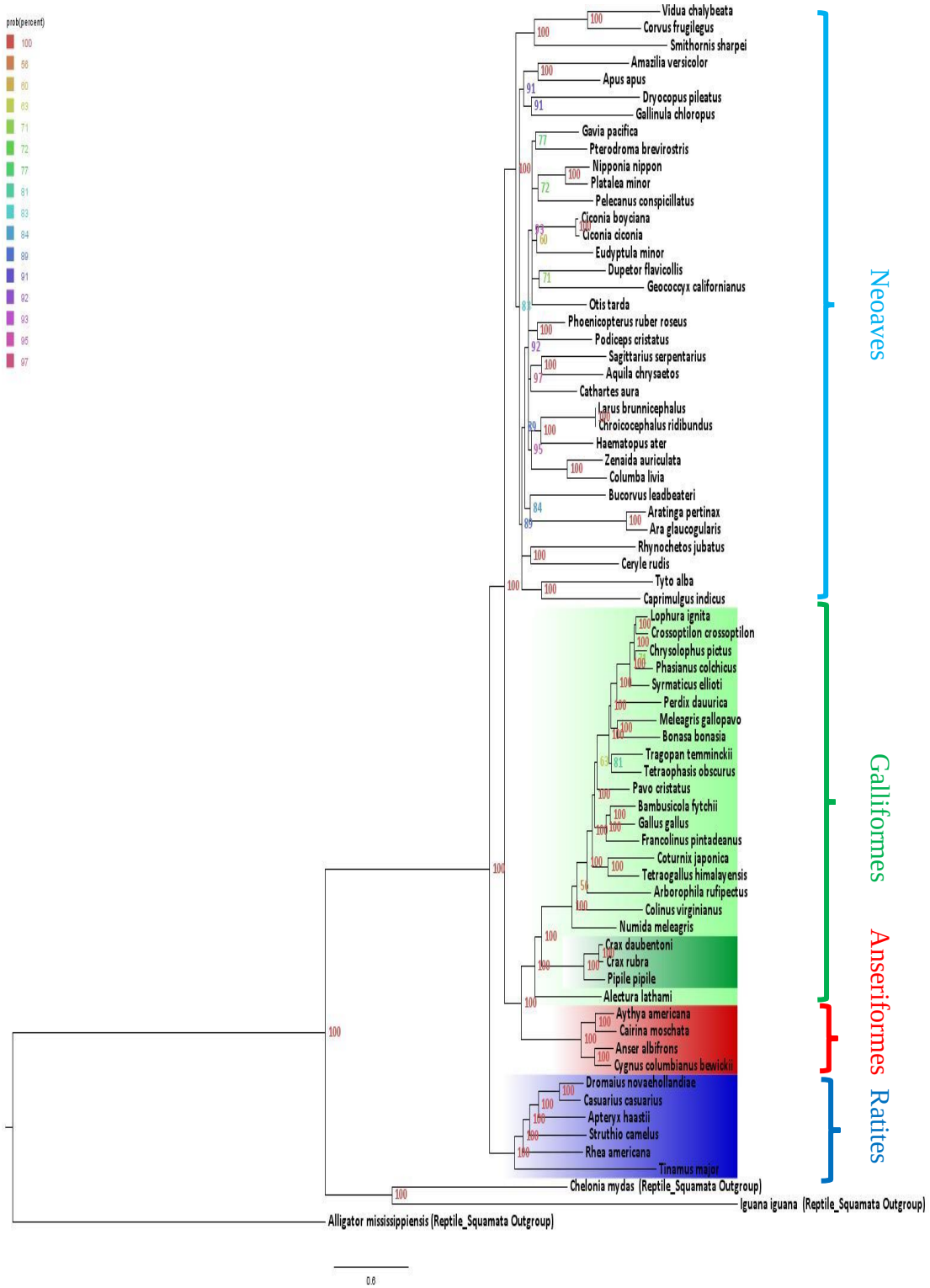


Figure 5.10: Phylogenetic relationships among Avians based on analysis of 11,010bp of Whole Mitochondrial Genome data (12 protein-coding genes) from 68 genera across the Aves class including the novel *Pipile pipile*, and 3 reptile outgroups. Topology and branch lengths are based on Maximum Likelihood General Time Reversible substitution models under non-uniform evolutionary rates with Gamma distribution and Invariable sites (GTR+G+I), through independent analysis of Bayesian inference in MrBayes. Bayesian analysis used MCMC with 25 million generations sampled every 1000 generations, discarding the first 25% of the run as burn-in, with posterior probabilities to support node topologies. Cracidae are highlighted in dark green.

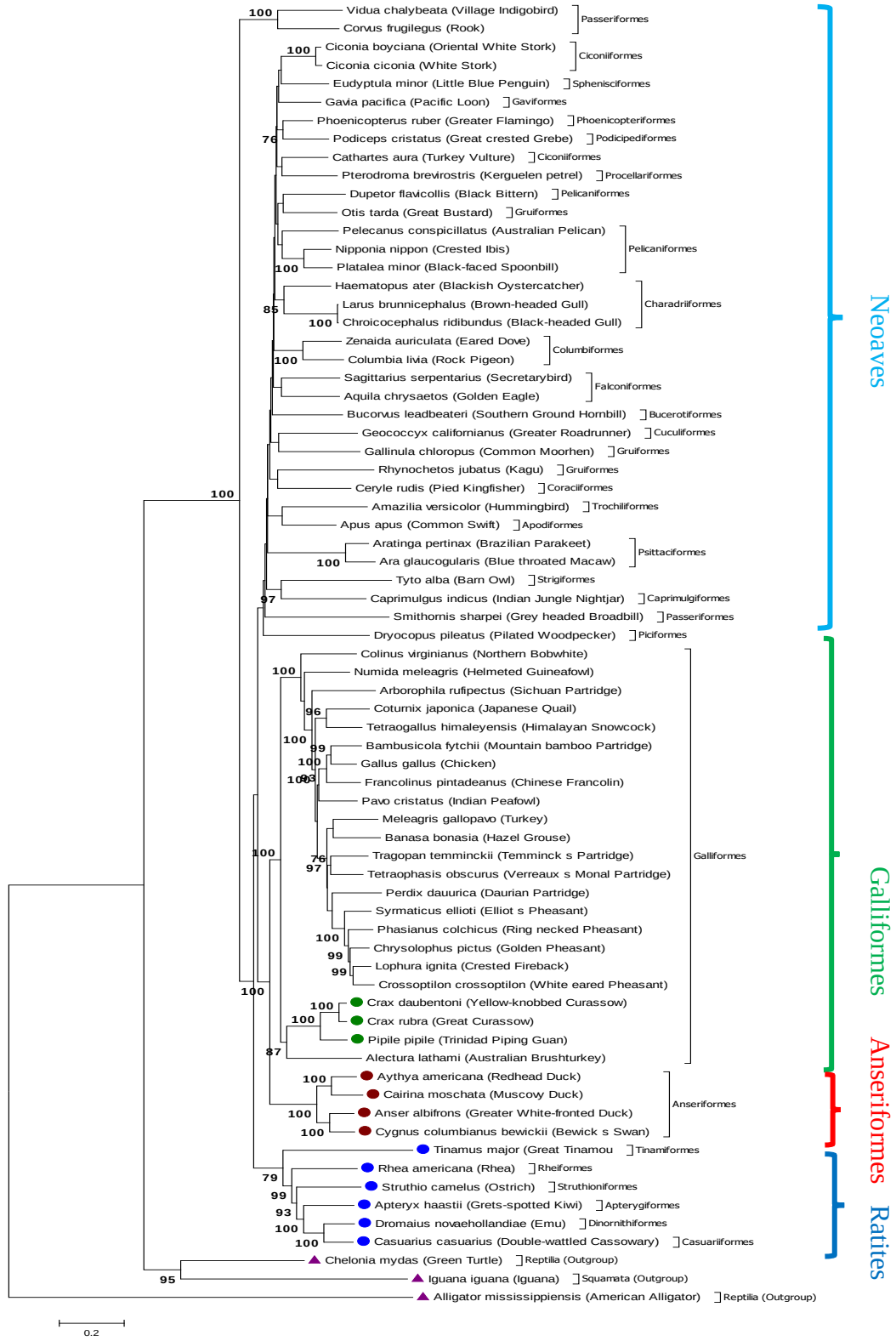


Figure 5.11: Phylogenetic relationships among Galliformes based on analysis of 11,010bp of Whole Mitochondrial Genome data (12 protein-coding genes) from 68 genera across the Aves class including the novel *Pipile pipile*, and 3 reptile outgroups. Topology and branch lengths are based on Maximum Likelihood General Time Reversible substitution models under non-uniform evolutionary rates with Gamma distribution and Invariable sites (GTR+G+I), in MEGA 7.0 with Bootstrap statistical support. Maximum Likelihood analysis in MEGA with 500 bootstrap replicates for node statistical support (numbers at nodes in %, in bold), only nodes with >75% statistical support displayed. Cracidae are highlighted in green.

5.3.4 Metadata Table Analysis

The use of multiple gene trees and both Bayesian Inference and Maximum Likelihood methods allow the exploration of inconsistencies in the phylogenetic reconstructions and conflicting signals at the taxonomic level. Metadata analysis indicates a number of topological discrepancies or inconsistencies in the phylogenetic placements of clades within the Galliforme and Avian trees, depending on the data set and partitioning used in analyses (individual gene, whole genomes and outgroups specifically), see Table 5.4. High degrees of branch and node support are observed throughout the Galliforme mito-genome tree (>91%), which is indicative of the robustness of the analysis. However the comparison of gene trees for the same taxon data, highlights inconsistencies between the genes in the recovery of a consistent topology and therefore inference of the 'true tree'.

Analysis of the Galliforme plus outgroup metadata indicates a clear inconsistency between the position of the Megapodidae (*Alectoris lathami*) in relation to the Cracidae and other Galliformes. The Megapodidae is consistently one of the basal clades within the Galliforme tree, however it presents as monophyletic in analysis of the Control Region and ND2 gene data sets, and contrastingly branched as sister to the Cracidae and separate from all other Galliformes in analysis of the Cytochrome *b* and Cytochrome oxidase 1 gene analysis. Whole genome analysis supports the sister relationship of the Megapodidae and Cracidae, see Table 5.4. In addition this inconsistency is observed between analysis by Maximum Likelihood and Bayesian Inference techniques, therefore indicative of a fundamental inconsistency likely due to the lack of representative samples from the Megapodidae and Cracidae clades.

Despite this inconsistency in the Megapodidae positioning, throughout all analyses there is a 100% support for the phylogenetic position of the Cracidae branch and the *Pipile* node, suggesting a high degree of accuracy in the topological position of this Family within the Galliformes.

Table 5.4: Analysis of bootstrap support for well accepted clades within avian phylogenetic trees, from this study and other published literature.

Analysis / Gene	Whole mtDNA Galliformes & outgroup	Whole mtDNA Avians	Whole mtDNA Avians & outgroups	Cyt b Galliformes & Outgroup	CO1 Galliformes & Outgroup	ND2 Galliformes & Outgroup	Control Region Galliformes & Outgroup	Cyt b Avians & Outgroup	CO1Avians & Outgroup	ND2 Avians & Outgroup	Whole mtDNA - Meikeljohn et al 2015	12S, ND2 & CytB – Pereira et al. 2002	CR, 12S, ND2, CytB & OVO-G - Crowe et al. 2006	Cytochrome b - Kimball et al., 1999	mtDNA & Introns - Hosner et al., 2016
Phylogenetic model	ML	ML	ML	ML	ML	ML	ML	ML	ML	ML	ML	ML	BI	ML	ML
Substitution Model	GTR +Γ+I	GTR +Γ+I	GTR +Γ+I	GTR +Γ+I	GTR +Γ+I	GTR +Γ+I	GTR +Γ+I	GTR +Γ+I	GTR +Γ+I	GTR +Γ+I	GTR +Γ+I	GTR +Γ+I	GTR +Γ+I	HKY85	GTR +Γ
Alignment length*	10,858	11,433	11,010	1143	1551	1038	1429	1143	1551	1038	~	1756	4437	1143	10,162
Figure number	5.7	5.9	5.11	5.16	5.17	5.18	5.19	5.22	5.23	5.24	5.2	5.2	5.2	5.2	n/a
Reptile-Aves															
Palaeognathe-Neognathae															
Galloanserae-Neoaves															
Galliformes-Anseriformes															
Megapodidae-other Galliformes															
Megapodidae – Cracidae															
Megapodidae/ Cracidae – other Galliformes															
Cracidae-other Galliformes															
Pipile-Crax															

*aligned base pairs, including indels

Key				
High Support 80-100%	Medium Support 50-79%	Low Support <49%	Un-supported	Not Applicable

Topologies in the basal region of the Avian mito-genome tree are consistent throughout analyses methods regardless of the use of a reptile outgroup; Palaeognathes (Ratites) branch basal of all other avian clades, which are represented in a single Neognathes (modern Avians) clade, with a high degree of support (99-100%). However individual gene trees are not consistent in the support of a Palaeognathe-Neognathe division.

The analysis of a condensed Galliforme taxon set in an Avian tree using complete mito-genomes consistently recovers a highly supported Galloanserae clades (99-100%), however support in individual gene trees is reduced (63-74%). Within the Galliformes the Megapodidae and Cracidae are consistently basal in both mito-genome and gene trees, however their relationship and position within the topologies vary between analyses.

5.3.5 Molecular Clock Analysis

Data Set Treatments in BEAST

The Galliforme data set and Avian data sets were analysed for molecular timescales and evolutionary relationships in BEAST2.2, using a Monte Carlo Marcov Chain method, under a General Time Reversible substitution model.

In analysis of the Avian data set the reptile outgroup was excluded to avoid inference of long branch effects and variation in taxonomic substitution rates. The Avian data set (68 species sequences) was analysed under three partitions (1^{st}, 2^{nd} and 3^{rd} coding sites), with linked clock model and trees. Data was analysed using a strict clock under the GTR model, with tree calibration priors; the fossils *Gallinuloides wyomingensis* (51.6 Mya Minimum, 75.3 Mya Maximum) and *Palaeortyx gallica* (28.0 Minimum, 51.7 Mya Maximum), and the molecular time calibration of the Neognathe-Palaeognathe diversification (110.1±19.8 Mya). The MCMC parameters were set at

50 million generations, sampled every 10,000 generations with trees every 10,000 generations, sampling a total of 5,000 trees with a 25% burn-in.

The Galliforme plus Outgroup data set (59 Galliforme and 4 Anseriforme species sequences) was analysed under three partitions (1st, 2nd and 3rd coding sites), with linked clock model and trees. Data was analysed using a strict clock under the GTR model, with tree calibration priors; the fossils *Gallinuloides wyomingensis* (51.6 Mya Minimum, 75.3 Mya Maximum) and *Palaeortyx gallica* (28.0 Minimum, 51.7 Mya Maximum), and the molecular time calibration of the *Procrax brevipes* fossil (44.0±6.0 Mya). The MCMC parameters were set at 50million generations, sampled every 10,000 generations with trees every 10,000 generations, sampling a total of 5,000 trees with a 25% burn-in.

The data sets were analysed in Tracer v2.2.4 to determine post-processing performance and treatments were modified accordingly for final analysis. Statistical analysis of the generation logs for the MCMC analysis is recorded in the output files, including Average Standard Deviation of Split Frequencies, and the Convergence diagnostics (Estimated Sample Size and Potential Scale Reduction Factor).

Marginal Posterior Densities

Data was partitioned into 1st, 2nd and 3rd codon positions, treated as coding data. Marginal Posterior Distribution plots, in Tracer v2.2.4, demonstrate differing rates of substitution between partitions. The Avian data set plot demonstrates relatively consistent mean substitution rates for the three coding partitions; 1.064, 1.181 and 0.754 for the 1st, 2nd and 3rd codon sites respectively (ESS >1000), 3rd codon substitution rate being marginally lower than 1st and 2nd codons. This is indicative of neutral or moderately positive selection in the protein-coding genes of the Avian mitochondrial genome data. The Galliforme data plots show 1st, 2nd and 3rd codon positions have relatively consistent substitution rates (0.704, 1.192 and 1.105 respectively). 1st codon substitution rates are marginally lower rates than 2nd and 3rd

codons. This is indicative of neutral or moderately positive selection in the protein-coding genes of the Galliforme mitochondrial genome data.

Molecular Timetrees

Analysis of the evolutionary timescales in BEAST2.2 produced molecular time-trees of the Avian and Galliforme data sets, with estimates of clade ages supports by posterior probabilities and 95% HPD intervals, see Figures 5.12 and 5.14 respectively. DensiTree tree data set analysis (Figures 5.13 and 5.15) demonstrate the spectrum of trees reconstructed by the BEAST2.2 MCMC analysis of the data sets. High concentration of tree plots in the DensiTree analyses is indicative of support for the consensus tree. Uncertainty in node heights is visualised in the variation in tree topologies. The consensus tree topology is that present in more than 99% of the tree samples, indicating a high support for topology in both of the data sets.

The molecular timetree analysis of the Galliforme data with an Anseriforme outgroup supports a stem age of Galliformes estimated at 134.15Mya (95% HPD interval 104.35-170.27Mya), which is an early estimation in comparison to other published research (Haddrath & Baker, 2001; van Tuinen & Hedges, 2001). The Cracidae clade (*Pipile* and *Crax* sp.) demonstrates a stem age estimate of 43.27Mya (95% HPD interval 33.8-54.35Mya), which is later in comparison to the analysis of Hosner et al. (2016) which placed the stem Cracidae at 58.2Mya (47.5-68.6Mya). This may however be representative of the limited Cracid taxon sampling used in this study, comparable to the analysis of Hosner et al. (2016). Other estimates, more recent divergences, are more consistent with previous research such as the stem *Gallus* estimate 8.08-13.5Mya, which is consistent with the analysis of van Tuinen & Dyke (2004).

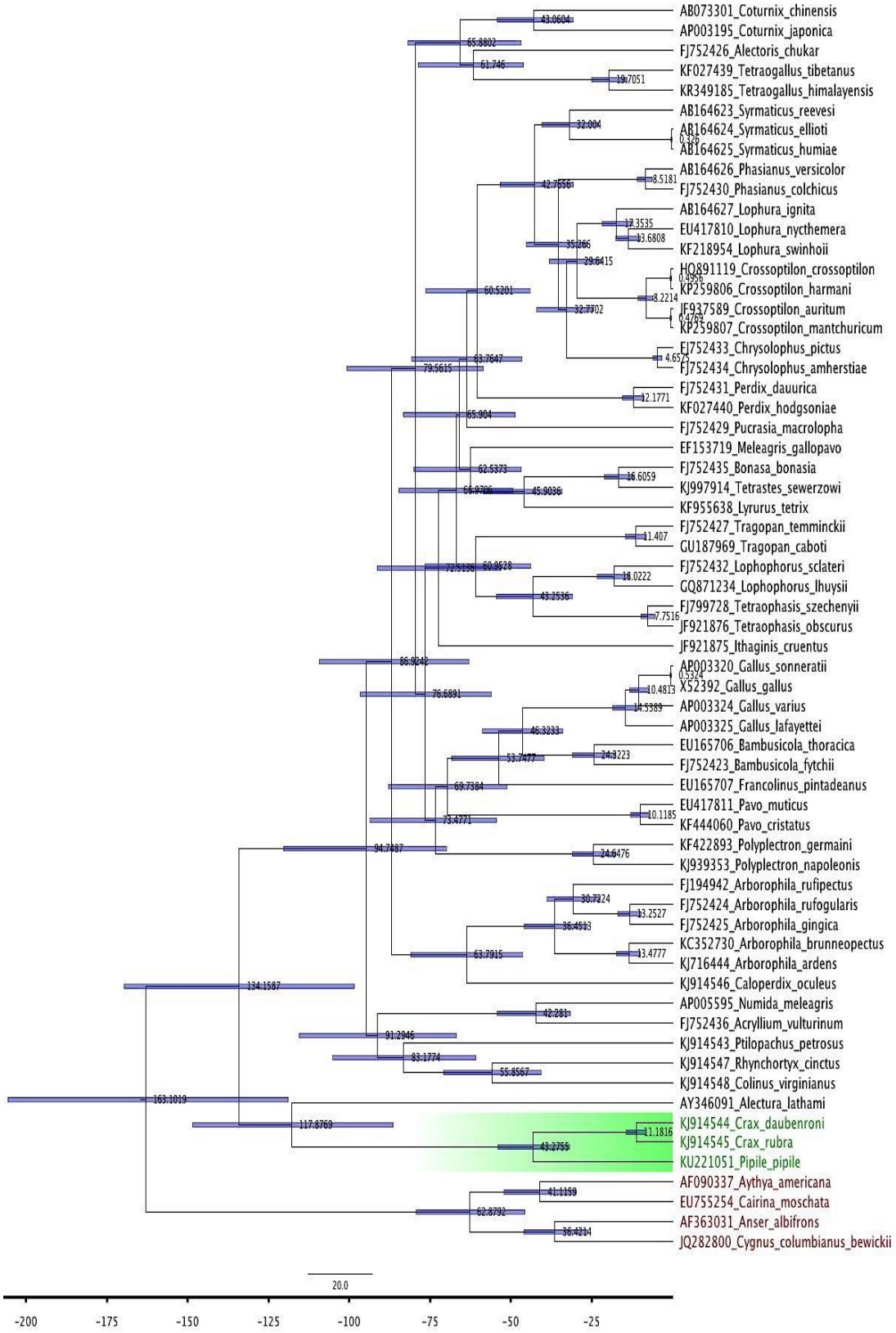


Figure 5.12: Evolutionary relationships among Galliformes based on analysis of 10,858bp of Whole Mitochondrial Genome data (12 protein-coding genes) from 59 genera of the Galliforme order including the novel *Pipile pipile*, and 4 Anseriforme genera as a putative outgroup. Node heights analysed in BEAST2.2 using a General Time Reversible models analysis and MCMC with 50 million generations sampled every 10,000 generations, discarding the first 25% of trees as burn-in. Node age indicated at the nodes, with error bars indicated in blue. Cracidae highlighted in green.

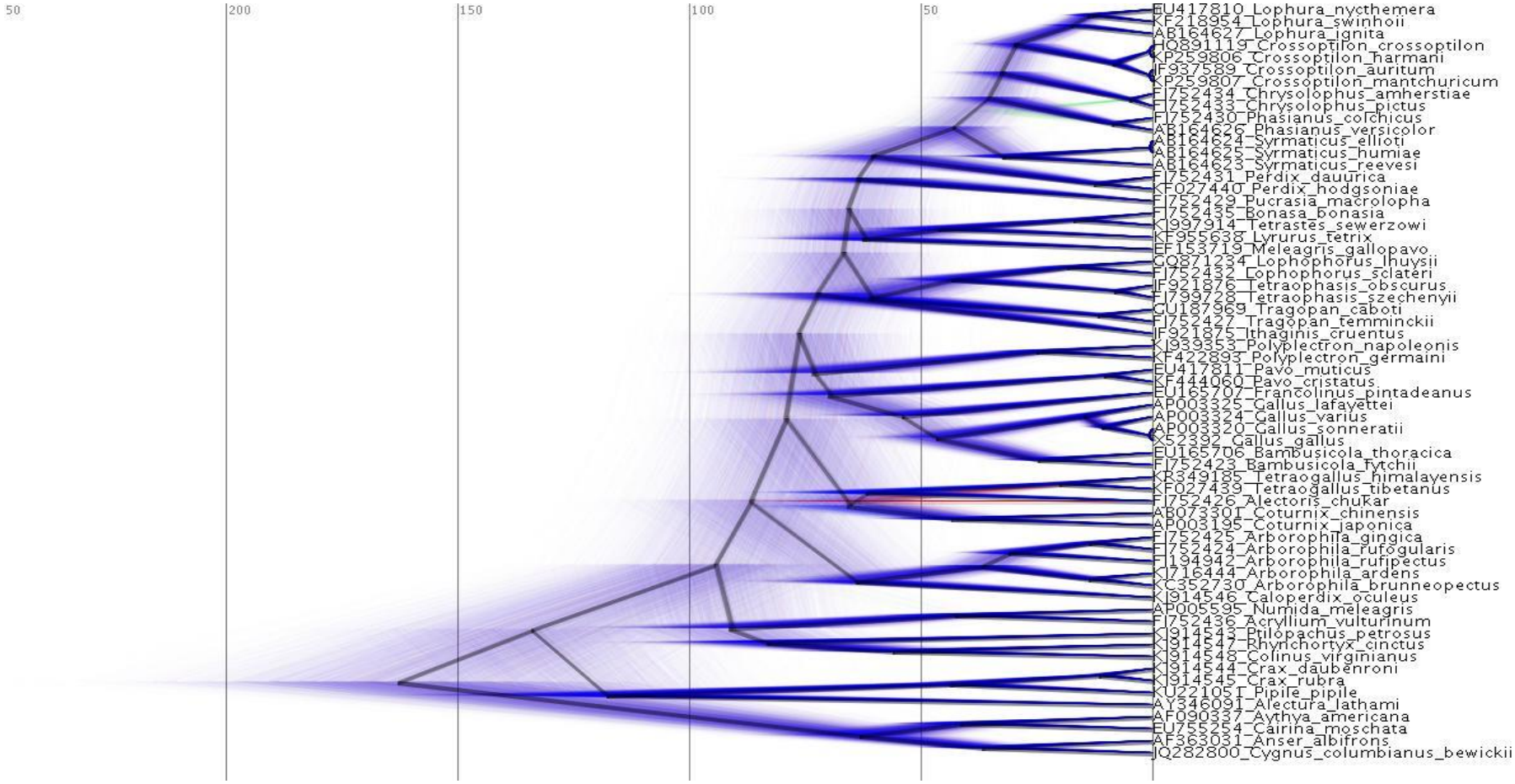


Figure 5.13: Evolutionary relationships among Galliformes based on analysis of 10,858bp of Whole Mitochondrial Genome data (12 protein-coding genes) from 59 Galliformes including the novel *Pipile pipile*, and 4 Anseriformes as an outgroup. Node heights analysed in BEAST2.2 using a General Time Reversible models analysis and MCMC with 50 million generations sampled every 10,000 generations, discarding the first 10% of trees as burn-in. Individual trees analysed in Densitree indicated by blue lines with consensus tree over-laid in black. Scale bars indicate increments of 50 million years.

The age of the root of the Avian tree is estimated as 117.62 Mya (95% HPD interval 81.99-151.64 Mya), this analysis produced a broad timescale estimate within the Cretaceous period, which is inclusive of many of the published time estimates (Haddrath & Baker, 2001; Härlid et al., 1997; Paton et al., 2002). This is consistent with the diversification of the major avian lineages prior to the K-Pg boundary (66 Mya), supported by the analysis of Cooper & Penny (1997), Haddrath & Baker (2001), Paton et al. (2002), and van Tuinen & Hedges (2001), and consistent with biogeographical hypotheses.

Within the Avian timetree the estimate of the stem Galliformes is 90.34 Mya (95%HPD interval 67.46-118.82 Mya), an estimate which is lower than in the Galliforme only analysis, which is however is more concordant with published research. Similarly, the stem Cracidae estimate is lower than in the Galliforme analysis, estimated at 37.02Mya (95%HPD interval 27.6-48.97Mya). The timescale of this estimate may be directly influenced by the limited number of Cracid taxa in the analysis, those present (*Crax* and *Pipile*) both diverging at a much later time period than the basal cracid lineages (Hosner et al., 2016). This analysis supports the expectation of the Cracid divergence after the K-Pg boundary in the Oligocene, with diversification of the family in an era appropriate to the major biogeographical and climatic disturbances of the South American continent.

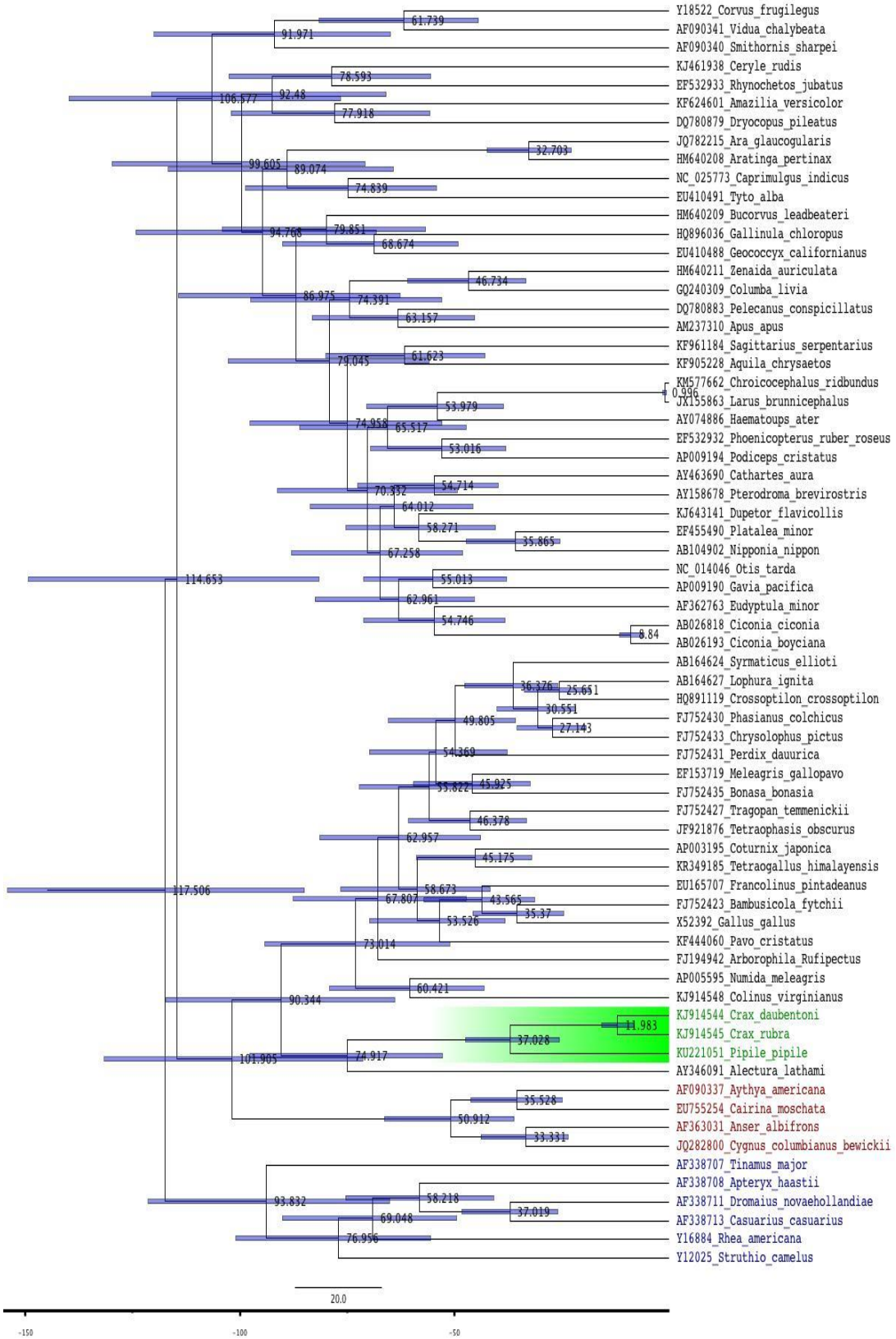


Figure 5.14: Evolutionary relationships among Galliformes based on analysis of 11,433bp of Whole Mitochondrial Genome data (12 protein-coding genes) from 68 genera of the Avian class including the novel *Pipile pipile*. Node heights analysed in BEAST2.2 using a General Time Reversible models analysis and MCMC with 50 million generations sampled every 10,000 generations, discarding the first 25% of trees as burn-in. Node age indicated at the nodes, with error bars indicated in blue. Cracidae highlighted in green.

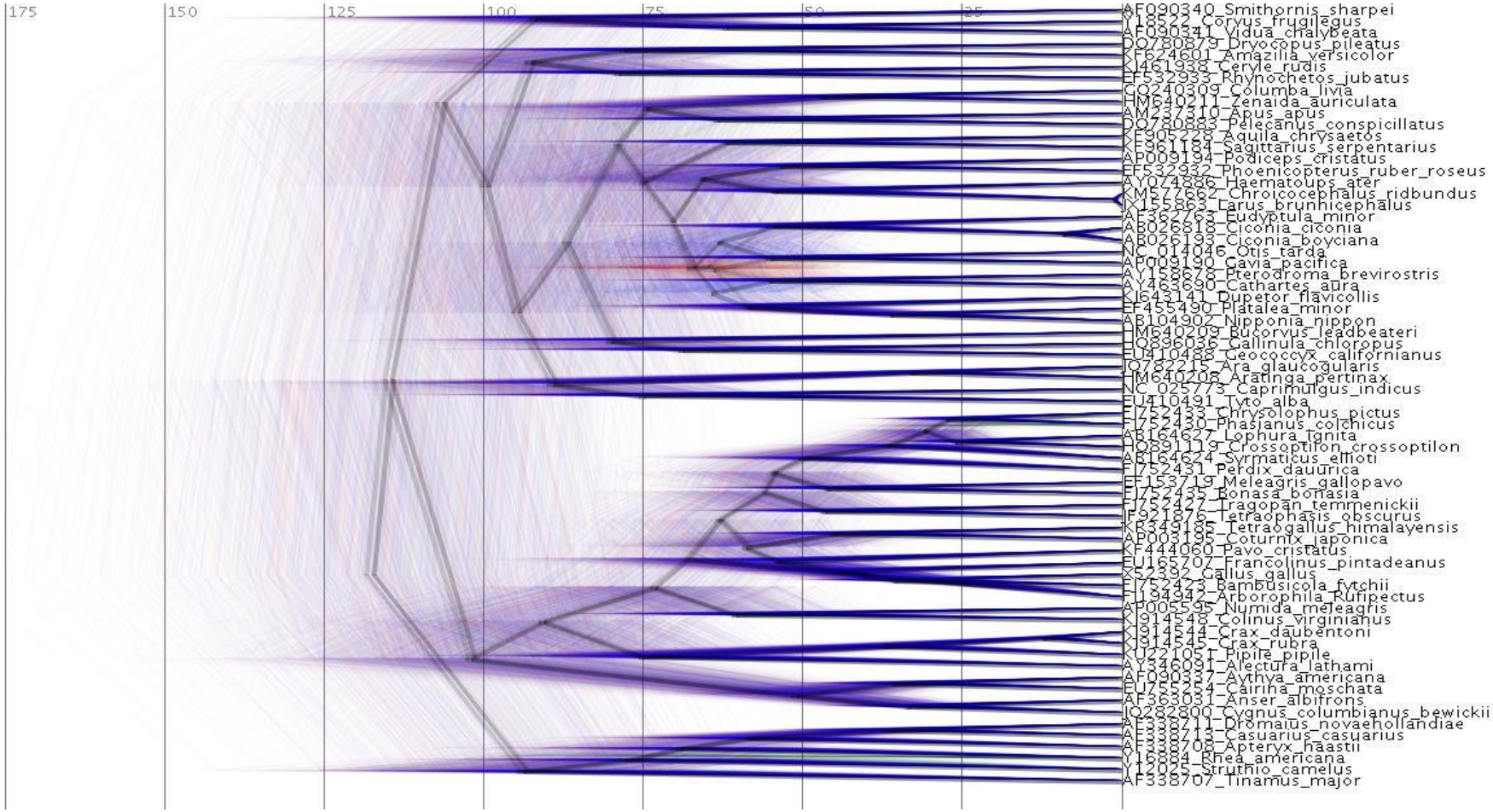


Figure 5.15: Evolutionary relationships among Galliformes based on analysis of 11,433bp of Whole Mitochondrial Genome data (12 protein-coding genes) from 68 Avian species including the novel *Pipile pipile*. Node heights analysed in BEAST2.2 using a General Time Reversible models analysis and MCMC with 50 million generations sampled every 10,000 generations, discarding the first 10% of trees as burn-in. Individual trees analysed in Densitree indicated by blue lines with consensus tree over-laid in black. Scale bars indicate increments of 50 million years.

5.4 Discussion

The focus of this study is placement of the Cracidae, specifically the genus *Pipile* within the Galliforme phylogenetic tree and its context within the Neognathae in the avian phylogenetic tree. The inclusion of more complete mitochondrial genome sequences helps resolve phylogenetic positions and relationships within both the Galliforme and Avian evolutionary trees.

Phylogenetic Reconstructions

Every phylogenetic tree reconstructed is a hypothesis of the 'true tree', under a given set of parameters with specific data, how realistic the topology is may be inferred by statistical tests but may not necessarily represent the actual evolutionary relationships between species. Phylogenetics aims to hypothesise the 'correct topology' of a group of species, representative of their 'true' evolutionary relationships. Congruence between the analyses of many different methods and models of phylogenetic reconstruction is a means of testing the 'correctness' of phylogenetic reconstruction of a given group of organisms. This analysis has demonstrated many congruencies with previously published phylogenies and between the analyses presented in this study, which is indicative of support for the relationships inferred by the analyses.

Phylogenetic reconstructions performed under two separate methods, Bayesian Inference and Maximum Likelihood, enable the comparison of topological reconstructions from the data sets. Data was partitioned in different ways to establish the most reliable reconstruction of the data. Larger amounts of sequence data improved the probability of recovering the 'correct' tree when the appropriate model of substitution is used, thus the use of concatenated mito-genome data sets minimise sampling errors produced by smaller data sets (Paton et al., 2002). In general, both the Galliforme phylogenetic trees and Avian phylogenetic trees conform broadly to traditional classifications of the species, within the sampling restrictions. The research of Slack et al., (2007) highlighted systematic bias in many previous studies as a consequence of incomplete sampling (evidenced by conflicting topologies). Slack et

al (2007) demonstrated that the inclusion of a more representative sampling effort resolved conflicting data through the breaking up of long branches within the avian tree, confirmed in broad sampling of the Galliformes by Meikeljohn et al. (2014). Long branch effects are noted by Braun and Kimball (2002), Slack et al (2008) and Pacheco et al (2011), directly as a consequence of limited taxon sampling. The issue of long branch attraction remains a problem in avian systematics; Jetz et al., (2012) report of the 9,993 avian species only 6,663 species are represented in molecular databases. In addition to the lack of representation of extant species in analysis, the issue of extinct species also has an impact on avian systematics and phylogenetic reconstruction. The limited scope of taxa for critical groups is apparent within the Galliforme order, more comprehensive sampling is needed for critical groups such as the Cracidae (Meiklejohn et al., 2014; Pacheco et al., 2011). Therefore, the inclusion of the mitochondrial genome of the Trinidad Piping Guan, and further mitochondrial genomes from other genera, has increased robustness of the analysis, as demonstrated by improved statistical support for nodes and clarity in the topological construction of phylogenetic relationships.

Comparative analysis of data sets with and without outgroups is indicative of the stability and robustness of these mito-genomic phylogenies. Outgroup effects are well documented in avian systematics (Meiklejohn et al., 2014; Pacheco et al., 2011; Paton et al., 2002; Slack et al., 2007); the inclusion of a reptile outgroup as the closest taxonomic class to avians has significant effects on the structure and topology. The inclusion of outgroups is known to have a significant effect on tree topology in the use of complete mitogenomes in phylogenetic analyses (Meiklejohn et al., 2014), error being induced through the long branch effects of mammal and reptile outgroups (Paton et al., 2002). This analysis incorporated a reptile outgroup consisting of the mitogenomes from the *Alligator mississippiensis* (Janke & Arnason, 1997), *Chelonia mydas* (Kumazawa & Nishida, 1999) and *Iguana iguana* (Janke, Erpenbeck, Nilsson, & Arnason, 2001). The inclusion of this outgroup, as with previous research by Slack et al. (2003) and Paton et al. (2001), roots the avian tree in the context of their relationship to the Archchosaurs (birds, reptiles & dinosaurs), avians being a sister clade to reptiles. Including this outgroup introduces a long branch basal of other taxa within the tree. Reptile genomes are known to demonstrate a similar mitochondrial

gene organisation to birds, however differing evolutionary rates among sites may interfere with phylogenetic conclusisons (Kumazawa & Nishida, 1995, 1999). The disruptive influence of outgroups has been established in avian phylogenies as a consequence of long branch effects and genetic distance in both mitochondrial and nuclear gene analysis (García-Moreno et al., 2003; Morgan-Richards et al., 2008; Pacheco et al., 2011; Slack et al., 2007). The use of an Anseriforme outgroup in the estimation of Galliforme phylogenetic relationships has a reduced effect on topology, the Anseriformes being closely related to the Galliformes, often observed as the Galloanserae clade (Dimcheff et al., 2002; Eo et al., 2009; Kan et al., 2010; Meiklejohn et al., 2014; van Tuinen et al., 2000). When removed from analysis the Anseriforme outgroup has little effect on the reconstructed topologies or node support, as observed in previous research (Meiklejohn et al., 2014). As such, use of the Anseriformes has limited influence on the phylogenetic reconstruction of the Galliforme tree, concordant with analysis of both rooted and unrooted tree topologies.

The alignment of long sequence data sets and genetically divergent species is challenging (Morgan-Richards et al., 2008), the broader the organismal range the more difficult the accurate alignment of sequence data. Other issues confound alignment and may mask true divergence relationships within the data, a significant issue is multiple substitutions at the same nucleotide position or backwards substitutions. The effects of such mutational substitution is the masking of phylogenetically informative variation regarding the divergence of species, which may result in erroneous parsimonious inferences.

Sequence length has been noted to have significant effects on the topology of avian phylogenetic reconstructions and the support for clade structures within the tree (Braun & Kimball, 2002). Increased sequence length has been shown to improve the proportion of supported nodes within avian phylogenetic trees, thereby inferring a higher probability of constructing the ‘correct’ topology (Braun & Kimball, 2002). Comparison of single gene trees and multi-gene trees in the Charadriiforme family determined variation in the utility and robustness between different genes and multi-gene analysis (Paton & Baker, 2006), disparities being more exaggerated in basal

divergences within the trees. Shorter genes and more highly conserved genes lack phylogenetic signal and therefore are less well suited to analysis of phylogenetic inference (Meiklejohn et al., 2014; Paton & Baker, 2006). Larger amounts of sequence data improve the estimation of phylogeny, the support for nodes and topology of the trees (Meiklejohn et al., 2014; Slack et al., 2007). The use of entire mitochondrial genomes is therefore advantageous in the analysis of diverse species groups and deep phylogenetic relationships, however presents increased issues of alignment difficulties.

An important inference in phylogenetics is the relationship between the type of analysis and species; the construction of gene trees as opposed to species trees. Where a single gene is analysed it is important to consider the degree of inference that can be gained from a single gene in relation to the evolution of a taxonomic class. The individual gene is under a constrained mode of evolution which is critically limited by function, failure of gene function being a defining factor in physiological function and therefore survival. Fixation of mutations within a gene must therefore still maintain the gene function, thereby limiting the capacity of the gene to evolve. A single gene used in phylogenetic reconstruction of a taxonomic class is therefore constrained within the genes function in that class, therefore the phylogenetic relationships are indicative of a gene tree rather than a species tree. The use of complete genome data in phylogenetic reconstruction is therefore a more accurate representation of evolution, as compared to a single gene, in the phylogenetic reconstruction of species evolutionary relationships.

The mean rate of divergence of the mitochondrial genome has been established as 2% per million year (Wilson et al., 1985), which is of use in establishing a temporal framework for the diversification of species and therefore phylogenetic inference. However, different genes evolve at different rates among lineages, due to their functional constraints, therefore influencing the substitution rate as applied to phylogenetic reconstructions (Paton et al., 2002). Where rate variation among nucleotide sites is large the use of Gamma parameter applied to the substitution model infers better phylogenetic reconstructions (Nei & Kumar, 2000).

Maximum Likelihood and Bayesian Inference methods adopt differing approaches to the reconstruction of phylogenies; Maximum Likelihood methods estimate the topologies based on the maximum probabilities of data, whereas Bayesian Inference estimates topology based on the posterior distribution probabilities (Drummond & Bouckaert, 2015). Bayesian methodologies also allow the analysis of the range of support for a node or branch position, as determined by the distribution of the position over the generation of trees analysed. In contrast, Maximum Likelihood methods only infer a single value for the maximum probability at a given node or branch position. The use of priors in Bayesian Inference analysis is beneficial in that known informative data is applied to the analysis, however the degree of accuracy of the priors used may have significant effects on hypotheses made based on the analyses. Drummond and Bouckaert (2015) suggest priors should not inform the posteriors, only the model and data should inform posterior estimations made in analysis. Comparison of the data using the two methodologies aids in the assessment of different techniques in establishing the correct or true tree, through the performance of different genes and mito-genome methods in recovering robust and statistically supported topologies.

Galliforme Phylogenies

Complete mito-genomic analysis further informs phylogenetic reconstruction through a greater abundance of informative sites therefore improving the ability of the analysis to resolve the 'correct' species topology. The Galliforme data is universally robust in all analyses, reconstruction of the major clades is broadly universal across Bayesian and Maximum Likelihood methods, for both rooted and unrooted methodologies. The inclusion of broader taxonomic sampling in whole genome analysis of this region of the Galliforme phylogeny demonstrates higher support rates and reduced long branch effects, as with previous research (Slack et al., 2007). This study is consistent with the analysis of Meikeljohn et al. (2014), with the inclusion of additional mito-genomes from the *Pipile pipile* species and eleven other Galliforme species.

Galliforme relationships remain a complex problem, incongruence between phylogenetic reconstructions, and morphological and molecular relationships has been

well documented (Dyke et al., 2003; Kan et al., 2010; Meiklejohn et al., 2014; Van Tuinen & Dyke, 2004). Complete mito-genomic analysis of the Galliforme tree is largely congruent with traditional classifications and morphological phylogenies, the key clades being well defined and highly supported statistically. The comparison of gene trees and mito-genome trees demonstrates a high level of support (<75%) for the major clades within the Galliforme order, which is comparable to the expected species tree, however some conflicting signals remain between the relationships of the key clades within the gene trees as compared to the mito-genome tree.

The divergence of the Megapodidae and Cracidae prior to other Galliforme families is universally accepted, consistent with other research (Armsrtong et al., 2001; Dimcheff et al., 2002; Meiklejohn et al., 2014; Sibley & Ahlquist, 1990). However, this study highlights the lack of resolution of the relationship between the Megapodidae and Cracidae, as either monophyletic or sister branches basal to other Galliforme families. Variation in the position of the Megapodidae and Cracidae is observed both as an effect of individual gene use and the method of analysis used. These two families are both under-represented taxonomically within the Galliforme tree, which may be having an adverse effect on the topological reconstruction. The inclusion of further mitochondrial genomes from these families should resolve the disparities observed in the relationship between the Cracidae and Megapodidae, and define their evolutionary relationships with a higher degree of consistency between genome and gene analyses.

The hypothesised phylogenetic trees through the genetic analysis of individual gene on mitochondrial genomes, confounds the analysis of relationships within the Galliforme lineages, including the Phasianidae. The partridges and pheasants remain unresolved, this diverse clade demonstrates some controversial relationships such as Turkey (*Meleagris gallopavo*) which consistently branches within the Phasianidae clade, however its exact relational position varies with type of analysis used. The complex relationships within the Phasianidae clade may be representative of rapid radiation and diversifications within the group, which is supported by both molecular and morphological reaearch (Jiang et al., 2014).

The position and relationship of the Odontophoridae and Numididae with relation to the other Phasianids remains unresolved, the use of single genes confounds analysis and the position within the tree is inconsistent between analyses. The complete mito-geneome analysis supports a basal position in relation to the remaining Phasianid clades with a high degree of support for the nodes, suggesting the relationship is better resolved through the use of complete mito-genomes than single genes.

The comparison of the Gene trees demonstrates lower nodal support in both methods of analysis, when compared to the use of complete mito-genomes in analysis, as has previously been reported (Braun & Kimball, 2002; Meiklejohn et al., 2014). The genes selected for this analysis (CytB, COI, ND2 and Control Region) are all widely used in species phylogenetics, however their reconstructed topologies varied greatly at the species level, despite recovering the key clades universally.

Avian Phylogenies

The analysis of Aves class level phylogenetics is highly sensitive to taxon sampling (Braun & Kimball, 2002), largely due to the effects of long branches and varied substitution rates between lineages critically influencing topological reconstructions. Many lineages are under-represented in genetic analysis, including the Cracidae, therefore there are limitations to the accuracy of reconstructing phylogenetic relationships within these families. Pacheco et al (2011) identify a number of clades which are in need of further resolution, through a broader taxon sampling, including the Coraciformes, Psittasiformes, Strigiformes, and Falconiformes. The long branch effects of limited sampling in avian analysis are therefore significant in the reconstruction of realistic phylogenies. Avian phylogenies are also challenging due to the effects of rapid radiation of avian lineages, as suggested by Barker et al. (2004), Pacheco et al. (2011) and Shen et al., (2010), and the influences of variation in substitution rates between lineages. Additionally the effects of convergent evolution within the avian class on the reconstruction of phylogenetic relationships have yet to be fully established (Fain & Houde, 2004; Morgan-Richards et al., 2008).

Early diversifications in the avian tree are well documented, the Paleognathe phylogeny has been established as basal to the Neognathe clade in the analysis of extant birds (Cooper et al., 2001; Haddrath & Baker, 2001; Harlid & Arnason, 1999; van Tuinen, Sibley, & Hedges, 1998). This study supports the basal location of the Palaeognathe clade, in relation to the evolution of all other modern avians. The division of the Palaeognathe and Neognathe lineages is well defined in the analysis of the complete mito-genomes. Within the Neognathae, the diversification of the Galloanserae basal to the rest of the Neoaves is well supported by mito-genome and gene trees, consistent with published mitochondrial and nuclear data (Jarvis et al., 2014; Pacheco et al., 2011; Paton et al., 2002; Slack et al., 2007).

Within the avian tree this study has focused on the inclusion of representative samples from the Cracidae, to test basal relationships within the Galliformes and disrupt long branch effects. Inclusion of the Trinidad Piping Guan as an 'older' representative of the Cracidae increases the robustness of the analysis throughout the models (99-100% support). The position of the Cracidae in relation to the Megapodidae family remains unresolved, as previously discussed, however both lineages are clearly established as basal amongst the remaining Galliforme families. The *Pipile* lineage is consistently basal of the *Crax* lineage throughout analysis in both gene and mito-genome trees, with high degrees of support (100%), indicating strong evidence of their earlier divergence within the Family.

The Passeriformes are one of the most speciose and diverse groups within the neovaes, as such sampling remains unrepresentative of the genetic diversity of this order of avians. This study highlights the lack of phylogenetic clarity in the placement of the Passeriformes, possibly as a result of insufficient taxon sampling and long branch effects. Complete mito-genome trees establishes the position of Passeriformes as basal amongst the Neoaves, however this analysis is not consistent throughout the gene and mito-genome trees. Inclusion of a broader representation of this speciose group may improve resolution and clarify their position in relation to other avian families.

The *Falconiformes* remain difficult to resolve, as has been acknowledged in previous research (Helbig et al. 1994; Pacheco et al., 2011; Slack et al., 2007). The divided position of the *Falconiformes* within the avian tree suggests possible evidence of convergent evolution (Fain & Houde, 2004), or may be the result of sampling biases on the topological reconstruction. Broader taxonomic sampling of this difficult group may enlighten phylogenetic histories and relationships within the Neoaves, of this cryptic group.

The analysis of the Avian data using single genes demonstrated significant variation from the complete mitochondrial gene analyses, in relation to internal topological relationships between species and the statistical support for topologies and node positions. Garcia-Moreno et al. (2003) indicated a congruence between analyses of avian phylogenies using the individual ND1, ND2, 12S and CytB genes, however this analysis demonstrates topological consistencies only within the major clade relationships, as opposed to the internal species relationships. The Cytochrome *b* and Cytochrome c oxidase 1 gene trees support topologies consistent with morphological expectations of the relationships within the taxonomic groups. However, ND2 fails to resolve a tree which is supportive of the Palaeognathe-Neognathe division or the Neoaves classification, which are both supported in CytB, COI and mito-genome analysis. The analysis of Meikeljohn et al. (2014) indicated that in Galliformes the individual gene analyses were largely incongruent, advocating the use of complete mito-genomes to provide the best phylogenetic signal.

Molecular Timescale

The analysis of molecular timescale in this study has produced broad estimates of divergence dates, with a wide distribution between lower and upper time estimates. This may be indicative of the effects of calibration information or errors in the prior dates used, or may also indicate sampling biases, as have been highlighted in the analysis. The inclusion of further prior calibration points from reliable and accurately dated sources and a broader sampling effort may result in a narrower molecular

timescale with smaller deviation limits, and will result in a more accurate and robust phylogenetic reconstruction.

The analysis of the Galliforme relationships in this study demonstrates rapid evolution and convergent evolutionary pathways within the Galliforme lineage, as has been reported in several analyses of the order (Jiang et al., 2014; Shen et al., 2010; Wang et al., 2013). The estimation of the stem Galliformes at 134.15Mya (104.35-170.27Mya), is early estimation in comparison to other published research (Haddrath & Baker, 2001; van Tuinen & Hedges, 2001). However the estimate for the diversification of the stem Cracids in the late Oligocene (33.8-54.35Mya) is consistent with the analysis of Pereira et al. (2002), but later in comparison to the analysis of Hosner et al. (2016). The diversification of the stem Cracids is largely concordant with the biogeographic origins hypothesis of Hosner et al., supported by diversifications of the Cracidae in response to the climatic and geographical events in South America of the Pliocene and Pleistocene epochs (Carcraft, 2001; Haddrath & Baker, 2001; Slack et al., 2007).

The addition of the *Pipile pipile* mito-genome sequence clarifies the timing of the evolutionary relationships within the Galliformes and basal avian orders as they relate closely to one of the few avian fossils from the pre-Cretacious period, *Procrax brevipes*. The fossil *Procrax brevipes* from the Oligocene (33-53Mya) which is associated with the order Cracidae is of particular importance to the calibration of Galliforme and Avian evolutionary trees, being closely associated with the basal lineages of the modern avian tree (Galloanserae). The disadvantage of using priors is that there are no set guidelines on selecting priors and they may heavily influence the estimates of posterior values. Additionally, the fossil is assumed to correspond to a specific node within the phylogeny, however this may not be an accurate representation (extinct off-shoot lineages or intermediary ancestors) (Drummond & Bouckaert, 2015). Divergence of the *Pipile* genus is acknowledged in phylogenetic analysis as being earlier than that of the *Crax* genus, with well supported node topologies (100% bootstrap support, 1.0 posterior probability). This earlier divergence point for the *Pipile* node alters previous molecular clock evidence based on the *Procrax brevipes* fossil calibration, as previous analysis has been based solely on the

Crax sp. mito-genomes alone (Meiklejohn et al., 2014). The addition of the *Pipile* mito-genome removes long branch effects within the Cracidae clade, and improves evidence to support the early divergence of the Cracidae. However, the addition of further Cracid mito-genomes would improve support for both phylogenetic analysis and molecular timescale estimates.

The estimation of avian timescales have indicated the major modern avians diversified in the Cretaceous, with the root of the avian tree estimated at 117.62 Mya. This estimate is concordant with the research of Haddrath & Baker, (2001) and Paton et al., (2002), indicative of diversification of the major avian lineages prior to the K-Pg boundary (66 Mya), and consistent with biogeographical hypotheses. Within this analysis the stem Cracidae estimate is lower than in the Galliforme analysis, estimated at 37.02 Mya (95%HPD interval 27.6-48.97 Mya), however, it supports the expectation of the Cracid divergence in the Oligocene, after the K-Pg boundary. This analysis therefore supports diversification of the family during the Neogene which is appropriate to the evolution of species in response to the major biogeographical and climatic events of the South American continent.

5.5 Conclusions

The sequencing of the complete mitochondrial genome of the Trinidad Piping Guan has enabled the inclusion of the genus *Pipile* in the mito-genomic evolutionary tree. Within the Galliformes the Cracids are one of the basal families, representative of the ancient evolutionary history of the clade. This analysis has placed the *Pipile* basal within the Cracids, however this study highlights the lack of mito-genomic data available for species of this family represented in the DNA databases. Phylogenetic reconstruction of the avian evolutionary tree supports the basal position of the Galliformes within the Neoaves, in close association with the Anseriformes. The use of complete mitochondrial genomes is highly advantageous in the analysis of evolutionary relationships between avain species, such as the Trinidad Piping Guan, reconstructing robust well-supported phylogenetic hypotheses.

Chapter 6. Taxanomic Identification and Inter-Specific Genetic Variation of the Genus *Pipile*

6.1 Introduction

The Piping Guans

Traditional classifications of the genus *Pipile* have altered greatly over the last 200 years, species of this genus have been documented since the late 1700s, with initial recognition of three forms *P. cumanensis, P. jacutinga* and *P. cujubi* (Vaurie, 1967). Species definitions have included many forms based on morphological characteristics and geographical locations. Vaurie (1967) in his systematic notes on the family Cracidae, defined six geographically linked forms within the genus *Pipile*;

- *Pipile pipile* (Jacquin 1784)
  A geographical type, range restricted to the Island of Trinidad
- *Pipile pipile cumanensis* (Jacquin 1784)
- *Pipile pipile grayi* (Pelzeln 1869)
- *Pipile cujubi cujubi* (Pelzeln 1858)
- *Pipile cujubi nattereri* (Reichenbach 1862)
- *Pipile jacutinga* (Spix 1825)

The relationships of the Piping Guans with the genus *Pipile* (synonym *Aburria*) is widely disputed, as previously discussed. Nuclear and mitochondrial gene sequences are indicative of a very close relationship (Pereira et al., 2002), with an estimated divergence time of 3.8 Mya, more recent that of the curassow genus (8-10 Mya) or the other guans (10-18 Mya) (Grau, Pereira, Silveira, Höfling, & Wajntal, 2005). The Piping Guans, excluding the Trinidad form, are widely distributed throughout South America from Venezuela and Colombia in the north, to Bolivia, Paraguay and Brazil in the south.

*Pipile pipile**

Pipile cujubi 89-6-1-254

Pipile grayi 1902-3-18-1853

Pipile cumanensis 1-W-4-2-84

Figure 6.1: Photographs of facial features of the Piping Guan genus *Pipile* (photographs reproduced with permission of *Trini Eco Warriors and the British Natural History Museum).

Morphologically the significant characteristics defining the Trinidad Piping Guan are the crest feathers and feathers of the nape which are *"…much darker than in any other form, the white area being restricted narrowly to the edges of the feathers.*" (Vaurie 1967), and extensive white areas on the upper wing coverts (Vaurie, 1967). This analysis is supported by the taxonomic research of Lopes (2009), in the documentation of museum specimens of Trinidad Piping Guans from the American Museum of Natural History in New York and The National Museum of Natural History, Smithsonian Institution in Washington DC.

There are six accepted forms of the Piping Guan, although species level classifications are still disputed by some, in relation to the classification as sub-species or species. Del Hoyo et al. (1994) recognises four species of *Pipile*, with a further two sub-species; *Pipile pipile*, *Pipile cumanensis*, *Pipile cumanensis grayi*, *Pipile cujubi*, *Pipile cujubi nattereri* and *Pipile jacutinga*. In addition to the relationships of the six *Pipile* species, molecular evidence suggests a close relationship with the Wattled Guan *Aburria*

aburria, which has led to the discussion surrounding the classification of the genus species as *Aburria* or *Pipile* (Grau et al., 2005; Remsen, 2007).

Lack of taxonomic clarity between species of the *Pipile/Aburria* genus does not alter the conservation priority of the critically endangered Trinidad Piping Guan. As an endemic island species with limited dispersal capability the Trinidad Piping Guan represents a unique, evolutionarily significant unit, the only species of the genus present on the island of Trinidad, which is genetically isolated from mainland conspecifics.

Genetic structure

Species of the genus *Pipile* (*Aburria*) may present complex genetic relationships and unique variation representative of their wide dispersal across the South American continent. As discussed in Chapter 5, the Cracidae have a long evolutionary history in comparison other modern avians, being a basal clade within the Galliforme evolutionary tree.

The unique genetic structure of this genus may result from the significant geographical dynamics of the South American continent, as previously discussed in Chapter 5. The South American continent underwent significant geographical and climatic shifts which resulted in many biodiversity hotspots and regions of high species endemism; the Andean uplift, arise of the Amazonian river basin and the Guyanan shield are examples of key geological shifts which resulted in changes in the taxonomic composition of the region (Vegas-Vilarrúbia et al., 2011). Altitudinal displacement, habitat loss generated migration, climate shifts, catastrophic extinctions, niche biome and hydrological movements are strongly linked to the diversification and evolution of species in the South Americas. The morphological and molecular similarities between species of the genus *Pipile* may therefore represent evidence of the ability of the species to retain function and identity in the event of extreme geographical and climatic disturbance and thus the ecological resilience of the genus.

Molecular divergence estimates for the *Pipile* genus suggest significant influence of the glacial periods on the isolation and speciation of the genus and its ancestral populations (Grau et al., 2005). Evidence from the analysis of mitochondrial and nuclear DNA markers indicates the Cracidae evolved in Mesoamerica, dispersing South in response to climatic shifts, becoming isolated as a consequence of the Panamanian Isthmus and diverging across South America in response to subsequent geological and climatic events (Hosner et al., 2016). This analysis is broadly supported by fossil evidence of Cracid ancestors in North America (Eastman, 1900; Tordoff & MacDonald, 1957; Wetmore, 1923, 1956), including the fossil *Pipile* ancestor *Procrax brevipes* (Tordoff & MacDonald, 1957). Diversification of the Pipile genus through the influence of biogeographical means, is supported by evidence in other Neotropical species including *Ramphastidae* Toucans (Lutz et al., 2013; Patel et al., 2011), *Pionopsitta* Parrots (Ribas et al., 2007), *Mionectes* Flycatchers (M. J. Miller et al., 2008) and *Brotogeris* Parakeets (Ribas et al., 2009). Additionally these genetic and biogeographical relationships are observed across the taxa including *Lepidoptera* and primates (Di Fiore et al., 2014; Matos-Maraví et al., 2013).

Biogeographically the *Pipile* species are widely distributed across the South American continent, see Figure 6.2. The most widely distributed species are the Common Piping Guans (or Blue-throated Piping Guan) *Pipile cumanensis cumanensis and P. cumanensis grayi,* which are widespread across the lowland forests of north and central South America. The Red-throated Piping Guans *Pipile cujubi cujubi* and *Pipile cujubi natterei* are centrally distributed in the lowland forests of Brazil and Bolivia. *Pipile jacutinga* is geographically isolated to the Atlantic forest regions, which may as a consequence have led to its diversification in isolation and hence its higher degree of morphological and molecular variation from other *Pipile* species.

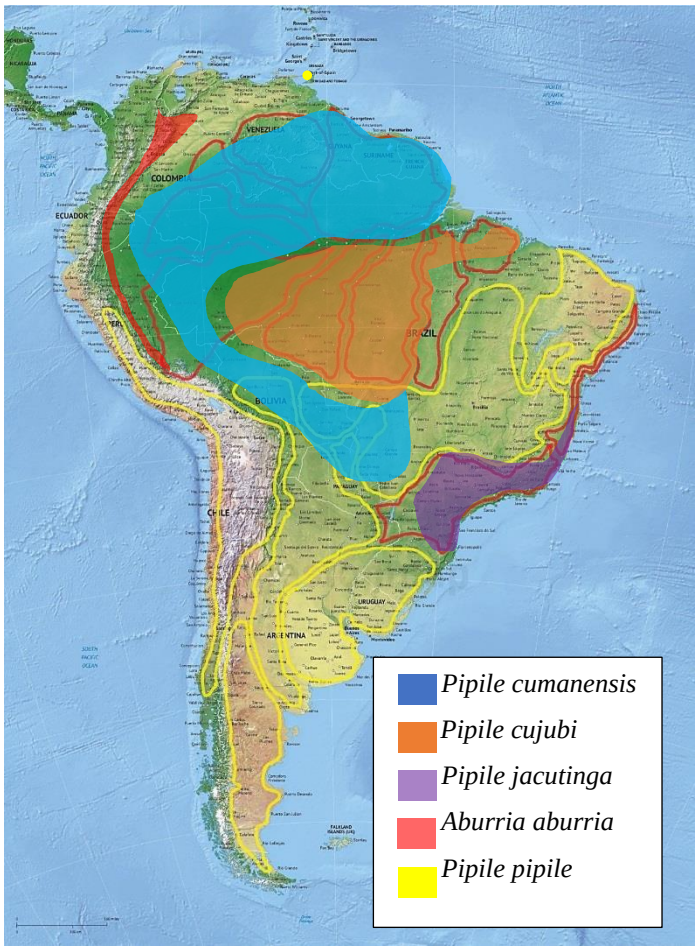


Figure 6.2: Estimated distribution of the Piping Guan species of the genus *Pipile* (synonym *Aburria*) and *Aburria aburria*, adapted from Grau et al. (2005). Areas of endemism and ecoregions outlined in yellow and red respectively, adapted from Carcraft (1985), Haffer (1985), and Tilston Smith et al. (2014).

Species

Species is a complex classification, there are many different definitions of species regulated by multiple factors including morphological characters, geographic distinction, and more recently genetics. The classification of species may be important for the management and conservation of taxa, traditional classifications of species have been predominantly based on morphological systematics, however more recent techniques such as genetic analysis have refined many species definitions. Contemporary methods of naming and classifying species need to recognise both evolutionary speciation and discrete biological groups, observable between organisms at both morphological and molecular levels (Helbig, Knox, Parkin, Sangster, &

Collinson, 2002). There have been over twenty-six different classifications of 'species' published, with differing criteria, which may have implications for the protection and conservation of species according to which concept definition is used. Some of the key concepts include the Biological Species Concept, the Evolutionary Species Concept and the Phylogenetic Species Concept (Haig et al., 2006), however the definition of Piping Guan species' under these differing concepts may have significant impact on current status and conservation of the contemporary species groups.

The 'Biological Species Concept' *"Groups of actually or potentially interbreeding natural populations, which are reproductively isolated from other such species"* – Mayer, 1942

This definition recognises the need of species to remain reproductively isolated in order to maintain evolutionary integrity (Helbig et al., 2002). This concept would delineate the Trinidad Piping Guan from other Piping Guans due to the isolation of the species on the island of Trinidad from its mainland conspecifics. However, Piping Guan species on mainland South America may be less well defined under this concept, should population boundaries overlap the potential for interbreeding and hybridisation remains.

The 'Evolutionary Species Concept' *"A species is a lineage of ancestral descent which maintains its identity from other such lineages and which has its own evolutionary tendencies and historical fate"* – Wiley, 1978

The evolutionary species concept requires that species maintain their integrity through time and space, and are distinguished as they diverge at ecological, behavioural, morphological and physiological levels (Helbig et al., 2002). The distinction of species is determined by evolutionary differentiation, which may vary between taxa (Richard Frankham et al., 2012). The Trinidad Piping Guan, due to its island isolation, is on a distinct evolutionary path from its mainland conspecifics. However, the degree of evolutionary differentiation between the species and its closest relative is undetermined. Morphologically few characteristics distinguish it from the Common Piping Guan (*Pipile cummanensis*) (Vaurie, 1967), and interbreeding would be

possible were the species to mix (Pereira & Wajntal, 1999). Genetically the variation between these species remains unclear, therefore the degree of evolutionary differentiation is difficult to estimate.

The 'Phylogenetic Species Concept' *"A species is the smallest diagnosable cluster of individual organisms within which there is a parental pattern of ancestry and descent"* – Carcraft, 1983

This concept requires species to be diagnosably different in phylogenetic hypotheses, however this does not account for future integrity of the species and reproductive isolation. Phylogenetics based on molecular analysis, implement the definition of species through diagnosable differences in their genetic makeup (DNA). In order to be diagnosably different sufficient generational time must have elapsed for the loss of shared alleles and the fixation of gene differences (Frankham et al., 2012). The Trinidad Piping Guans definition under these criteria is less well established than other concepts. If you are to define the species under morphological character phylogenetics, there are distinct phenotypic variations between the species, as described by Vaurie, (1967). However molecular phylogenetics has raised numerous issues regarding the degree of similarity and divergence between the species, and its relationship with the Wattled Guan *Aburria aburria* (Remsen, 2007).

Evolutionary Significant Units The use of a broader term of classification, Evolutionary Significant Units (ESU) has been proposed recently in a number of studies (Graham et al., 2004; Haig et al., 2006; Paplinska, Taggart, Corrigan, Eldridge, & Austin, 2011). This classification identifies groups of individuals or intra-specific lineages by their "reciprocally monophyloetic mtDNA haplotype" (Moritz, 1994), as a operational unit of evolutionary descent, for the purposes of conservation and population management (Paplinska et al., 2011).

In the definition of species it is important to extrapolate the future conservation implications, minimising harm caused through classification of species under differing

concepts and the legislative consequences. The classification of a species under these species concepts infer differences in the biological, conservation, financial and legal implications and considerations, which are highly subjective depending upon the definition used (Frankham et al., 2012; Helbig et al., 2002; Haig et al., 2006). Frankham et al. (2012) maintain that all species concepts are incongruent with the biological reality, species rarely 'fitting' a specific set of criteria which define its classification. The Trinidad Piping Guan species is a case where the definition of species may have significant conservation implications in terms of population management and legislative protection.

Classification of sub-species enables the management of vulnerable populations to be focused, rather than expending efforts on a broader species group. Historical classifications are now being assessed in the context of new evidence such as genetic analysis, traditional subspecies divisions based on morphology are not always concordant with genetics (Haig et al., 2006). Diagnosability should incorporate aspects of qualitative character state analysis both morphological and behavioural, but should also include evolutionary characters which are genetically defined (Helbig et al., 2002). Consideration of both biological and phylogenetic species concepts enables the classification of meaningful species divisions as evolutionarily definable entities (Haig et al., 2006).

Speciation

Speciation is a dynamic process, with many ecological, biological and geological causes. The direct observation of speciation in the natural world is usually impossible, however genetic analysis allows the inference of an indirect record of speciation through molecular phylogenetics. In addition molecular analysis enables the elucidation of temporal speciation events which may inform the causes of speciation. Species level phylogenetics using molecular information directly informs the relationships of species within a family or genera, the nodes reflecting common ancestral states and species divergences (Barraclough & Nee, 2001). However, molecular phylogenetics is often only a reflection of extant species and therefore may

be biased by the exclusion of missing or extinct con-specifics or ancestral species. Taxonomic species boundaries and traditional classifications of species may not correspond with genetic and evolutionary species within a clade, and may highlight systematic errors in species delineation (Barraclough & Nee, 2001).

Biogeographic structure of species groups are strongly linked to individual biotic responses to environmental shifts, observed over a temporal scale the evidence of biotic responses may be observed in the genetic variation between species. Organisms move directly as a response to environmental or climatic incursions in 'move or die' situations, and as a result of persistence and acclimation, adaptive forces strongly influence genomic evolution and therefore the diversification of species (Vegas-Vilarrúbia et al., 2011). The Trinidad form of the Piping Guans has been directly influenced by glacial oscillations of the last ice age; Grau et al. (2005) theorise the species diversified from the mainland form during periods of sea level change which isolated the continental island of Trinidad from the mainland coastal cordillera of Venezuela.

The observation of speciation on a temporal scale through molecular clock analysis provides an opportunity to analyse patterns of diversification within a family or genus, as discussed in Chapter 5. The physio-geography of the island of Trinidad (lowland peaks below 300m) suggest that diversification of the Trinidad form from the mainland species may be more recent than that observed in other members of the genus (Grau et al., 2005).

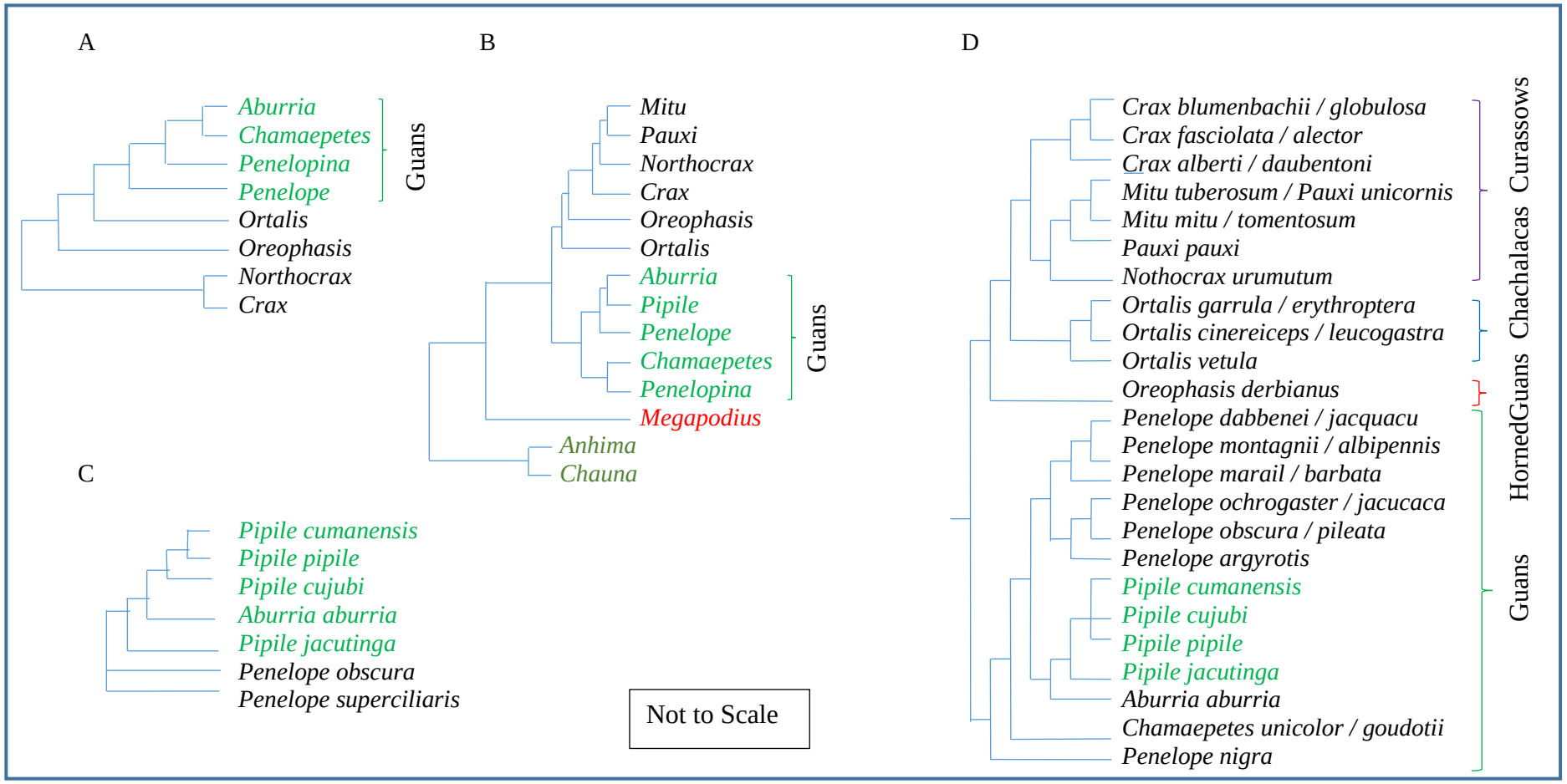


Figure 6.3: Inferred phylogenetic relationships of the Cracidae family, adapted from;
A. Morphological classification of Delacour & Amadon (1973) – *Aburria* and *Pipile* merged into *Aburria*, *Crax*, *Mitu* and *Pauxi* merged into *Crax*
B. Combined Nuclear and mitochondrial gene Maximum Likelihood tree of Pereira, Baker & Wajntal (2002)
C. Combined Control Region, Cytochrome b and ND2 mitochondrial gene tree (MP, ML and BA) of Grau et al. (2005)
D. Combined mtDNA, nDNA introns and UCE's Maximum Likelihood tree adapted from Hosner et al. (2016), condensed topology.

The diversification of other species within the genus *Pipile* is also likely to have resulted from major climatic and geographic changes during the last Ice Age. The Quaternary period was dominated by ice ages and periods of repeated global cooling at regular intervals, resulting from eccentricities in the earths orbit around the sun (Hewitt, 2000). These global climate changes vary according to latitude and topography; the equatorial regions of the Neotropics becoming temperate and arid, mountainous regions act as geographical barriers, high altitudes cooling and sea levels lowering. The genetic consequences of such geographical influences on populations are significant; populations encounter adaptive opportunities and population bottlenecks and founder influences have strong impacts on the genetic structures within populations. Fossil evidence suggests significant shifts in range changes of the Cracidae and Guans in particular from the North American continent to South America, which is indicative of a direct response to climatic shifts and specialist nature of the Cracids (Begazo & Bodmer, 1998; Hosner et al., 2016; Kattan, Muñoz, & Kikuchi, 2015; Tordoff & MacDonald, 1957).

Complete speciation progresses through several key stages; initiation of isolation of a population, increased genomic pressure of caused by isolation and complete reproductive isolation (Butlin, Galindo, & Grahame, 2008). Complete allopatry through extrinsic barriers is rarely observed, however island colonisations by non-migratory species is an example of complete isolation speciation (Butlin et al., 2008). The genomic effects of biogeographical speciation will naturally vary between species as each adapts to different influences and mutations accumulation through time. Repetitive cycles of separation and divergence of lineages, such as those produced by glacial oscillations, may additionally compound the effects on the genome. Understanding the genetic relationships within closely related species groups is of great importance to conservation of endangered species (Nishibori et al., 2005), including Galliforme species such as the Trinidad Piping Guan. The biogeographic hypothesis for the diversification of the *Pipile* genus is largely consistent with that observed in other Cracids and Neotropical avians (Grau et al., 2005; Lutz et al., 2013; Miller et al., 2008; Patel et al., 2011; Ribas et al., 2009, 2007).

Hosner et al. (2016) examined in detail the molecular relationships within the Cracidae, using both mitochondrial and nuclear DNA markers, to establish a biogeographical hypothesis for the predominance of the order on the South American continent. As discussed in Chapter 5, Hosner et al. (2016) determined the Cracidae originated in Mesoamerica (crown Cracids 20-26Mya), with a southerly distribution and diversification through the Americas in response to major climatic and geographic shifts, which is supported by fossil evidence (Tordoff & MacDonald, 1957). The biogeographic origins of the Galliforme order and contemporary Cracids (3-7.9Mya) is supported by molecular, morphological and fossil evidence (Carcraft, 2001; Haddrath & Baker, 2001; Hosner et al., 2016; Slack et al., 2007; Tordoff & MacDonald, 1957). However, in the investigation of the *Pipile–Aburria* genera relationship, the analysis of Hosner et al. (2016) demonstrates significant inconsistencies. Using only the mitochondrial DNA sequence information, as per the analysis of Grau et al. (2005), the *Aburria* genus is established within the *Pipile* genus, however the use of nuclear DNA markers places the *Aburria* genus outside of the *Pipile* clade in a basal relationship. This analysis therefore, demonstrates the unresolved nature of this relationship which needs further and more comprehensive genetic analysis.

Grau et al., (2005) determined that the current distribution of the genus *Pipile/Aburria* supports the theory that ancestral populations became isolated and speciated in elevated regions directly as a result of the major geographical events and climatic oscillations of the Pleistocene. The phylogeny of the genus can be difficult to define if insufficient molecular changes have accumulated to clarify relationships, where species remain close to their ancestral form. The diversifications of *Pipile cumanensis* and *Pipile cujubi*, and the isolation of *Pipile pipile* around 2.4-1.0MYA corresponds with the maximum sea transgression estimated at 2.5MYA, with the subsequent formation of sub-species during later climatic oscilations (Grau et al., 2005).

Genetic Isolation & Islands

Island environments offer a unique window into the evolution of species, providing a model system for the diversification of species from ancestral mainland forms. The geographic boundaries of islands critically restrict gene flow within and between species allowing species such as the Trinidad Piping Guan to evolve and adapt in complete isolation. The relationships between island species and their continental congeners or neighbouring island congeners can produce complex phylogenies due to the nature of island systematics and the ability of species to disperse, colonise and adapt (Emerson, 2002). The Piping Guans are a model example of a species with limited dispersal strategies, therefore should conform to patterns of speciation in relation to the isolation of the species via biogeographic means, including the isolation of the *Pipile pipile* on the island of Trinidad. However, the genetics of the genus are more complex in relation to historical genetic events which directly influence the contemporary genetic variation between the species. As discussed previously in Chapter 4, the Trinidad Piping Guan has demonstrated clear reductions in genetic variation likely as a result of both historical and contemporary bottleneck events. Isolation of the species in an island system critically restricts geneflow and removes the influence of immigration, which has the potential to improve genetic variation and fitness. The isolation of a species stronghold on a single island is of great conservation concern; endemic species are more susceptible to extinction level events and extinction rates are higher among endemic species than non-endemics (Blackburn et al., 2005).

Geographic isolation has had a significant impact on the genetic make-up of the Piping Guan species. Morphological variation may be a reflection of the dispersal history of ancestral species, a direct result of the selective pressures of the environment which is often highly evident in island species (Emerson, 2002). The Trinidad Piping Guan however, maintains significant characteristics of its mainland conspecifics, with little morphological or phenotypic variation from the Blue-throated Piping Guans (*Pipile cumanensis* and *P. cumanensis grayi*).

Value of Museum Samples

The acquisition of samples for genetic research of critically endangered species is notoriously difficult where populations are small or difficult to access. The use of invasive sampling techniques such as mist netting is both labour intensive and stressful to the animal, and impractical in large bodied arboreal species such as the Trinidad Piping Guan, which are notoriously difficult to locate due to their life history traits and the challenging geography of their habitat landscape (Alemu et al., 2005; Naranjit, 2010). Sample collections can therefore be easily supplemented by the use of specimens from museums and private collections, in addition to those collected in the field (Hogan et al., 2008). The importance and use of museum collections has increased substantially in recent decades with the advancement of scientific techniques to utilise these repositories of species information.

Museum collection specimens offer a unique historical perspective and biogeographic variety that is not available in many contemporary collections, this enables species to be studied on many levels. The maintenance of 'Species' and 'Evolutionary Significant Unit' references in collections enables a more thorough understanding of a species or populations spatial and temporal distribution, diversification and evolution. Historical collections allow the genetic ancestry of species to be temporally analysed, in the application of these collections to mitochondrial analysis of species the maternal lineages of current populations can be placed in a historical perspective (Chan et al., 2006; Glenn et al., 1999; Graham et al., 2004; Paplinska et al., 2011; Suarez & Tsutsui, 2004). Historical perspectives may be used to infer both ecological and evolutionary processes within and between species; patterns of distributions, niche modelling and phylogeographic hypotheses (Graham et al., 2004). Museum specimens are widely used to assess historical and ancestral levels of variations in species with low population sizes and those which have undergone significant population bottleneck events affecting contemporary levels of genetic variation within the population (Nichols et al., 2001; Weber, 2004).

A significant development in the use of museum collections has been to monitor population demographics over time, this is especially evident in the identification and monitoring of genetic bottlenecks and population declines. Hoelzel et al. (2002) and Weber et al. (2000) used museum collection specimens to identify multiple population bottleneck events in the Northern Elephant Seal, through the analysis of mitochondrial DNA haplotypes and genetic evidence of the declines. Analysis of haplotype diversity pre-bottleneck from historical samples is indicative of past levels of variation, with the subsequent recovery or status of the population may be evident in contemporary samples. The reconstruction of population history has been an important development in future conservation strategies, through the development of strategies to accurately reconstruct the timing and severity of past historical events on the extant genome thereby increasing the understanding of cause and consequence (Chan et al., 2006).

The development of advanced genetic techniques including methods of PCR and DNA extraction have significantly advanced the application of museum collections to phylogenetic analyses (Mundy et al., 1997). In avian species numerous techniques have been developed for the extraction of DNA from different tissue types in historical specimens including eggshells (Oskam et al., 2010), feathers (Bantock et al., 2008; Leeton & Christidis, 1993), footpad skin (Bantock et al., 2008; Mundy et al., 1997) and bones (Cooper et al., 1992). The extraction of DNA from historical specimens has many potential pitfalls, most notably the incursion of foreign DNA from other species including humans and the adverse effects of preservation and storage techniques used on the specimens (as discussed in Chapter 2). Despite the difficulties of amplifying ancient DNA from historical samples, both nuclear and mitochondrial genome analysis is widely being applied in the analysis of historical specimens.

Another significant advancement in the use of museum specimens is their application to global species identification systems such as the Barcoding of Life Data system (BOLD http://www.barcodinglife.com/). Significant efforts have been made in recent years, by Natural History institutions to genetically catalogue their collections, often incorporating the BOLD system of barcoding using the COI gene. The use of internationally recognised collections such as the British Natural History Museum and

the National Museum of Natural History (USA) as sources of genetic information, enables the integration of vouchered and holotype specimens into phylogenetic analyses, ensuring the integrity of interspecific analysis. However, the accuracy of the museum or collection record is critical to use of the specimens in the future and the impact of the conclusions which can be drawn from the use of such samples (Graham et al., 2004).

Pipile genus in Museum Collections

Samples from all species of the genus *Pipile* are represented in international museum collections of avian species. The geographic distribution of sample origin is diversely spread across South America which allows the biogeography and phylogeography of the genus to be analysed. In addition the sample collection is temporally distributed enabling the analysis of an historical perspective of genetic variation in comparison to contemporary populations. The genus *Pipile* (*Aburria*), is not widely represented in museum collections, Lopes (2009) recorded a total of 253 specimens in collections worldwide (81 *Pipile cumanensis*, 30 *Pipile cumanensis grayi*, 42 *Pipile cujubi*, 76 *Pipile cujubi nattereri* and 3 *Pipile pipile*). The Trinidad Piping Guan is particularly under-represented, there being only three specimens held in collections internationally (SMM12.2008 unknown museum location, and AMNH471557 & AMNH59510 in the American Museum of Natural History, USA) (Lopes, 2009). Lopes (2009) analysed phenotypic and morphological characteristics which define the species of the genus *Pipile* (*Aburria*), reporting on specimens from a number of museum collections world wide (see Table 6.1). The British Museum of Natural History, Ornithological collection at Tring in the UK, holds a total of 19 specimens; 13 *Pipile cumanensis*, 3 *Pipile cumanensis grayi* and 4 *Pipile cujubi* (no subspecies given).

Table 6.1: Holotype specimens of the genus *Pipile* (*Aburria*), as reported by Lopez (2009).

Species	Synonym	Collection number	Museum
Pipile pipile (Jacquin, 1784)	*Aburria pipile*	NMW22283	Naturhistorisches Museum Wein, Vienna
Pipile cumanensis (Jacquin, 1784)	*Aburria cumanensis*	NMW22287	Naturhistorisches Museum Wein, Vienna
Pipile cumanensis grayi (Pelzeln, 1869)	*Aburria cumanensis grayi*	BMNH1858.6.25.11	British Museum of Natural History, UK
Pipile cujubi cujubi (Pelzeln, 1858)	*Aburria cujubi cujubi*	NMW22309	Naturhistorisches Museum Wein, Vienna
Pipile cujubi nattereri (Reichenbach, 1862)	*Aburria cujubi nattereri*	NMW22283	Naturhistorisches Museum Wein, Vienna
Pipile jacutinga (Spix, 1825)	*Aburria jacutinga*	NWM22310	Naturhistorisches Museum Wein, Vienna

Ancient DNA Analysis

The main issue regarding the use of ancient and old DNA (aDNA) is that of authenticity (Gilbert, Bandelt, Hofreiter, & Barnes, 2005; Graham et al., 2004; Mundy et al., 1997; Paplinska et al., 2011), with regards to the production of a genuine DNA product, authentic to the specimen. Contamination with modern DNA is a significant issue in the use of aDNA, particularly prevalent in museum specimens is the contamination of specimens with human DNA from direct contact. Museum specimens are decaying with time; the age and storage of museum and historical samples has a direct impact on the ability to recover DNA from the specimen. Post mortem decay of DNA can result in numerous issues unique to aDNA including, physical destruction and sequence errors (Paplinska et al., 2011; Wandeler, Hoeck, & Keller, 2007). Gilbert et al. (2005) suggest criteria for maintaining authenticity and producing meaningful data which includes the isolation of work areas, the use of negative controls, appropriate molecular behaviour (PCR product size and species specificity), reproducibility and independent replication.

Temporal decomposition places limitations on the utility of museum specimens, under the current technological limitations. The improvement of PCR techniques and technological advances are making the amplification and analysis of historical DNA much easier, but currently there are limitations to the scope of analysis. There is a substantial decrease in PCR amplicon size in direct relation to the age of a specimen (Glenn et al., 1999), therefore the use of shorter amplicons is required to improve success. Wandeler, Hoeck, & Keller (2007) recommend an amplified fragment size of less than 200bp to facilitate DNA analysis of historical specimens, due to the high levels of DNA decay and high risk of contamination. However sample age and temporal decay is not necessarily the limiting factor in the quality of DNA analysis of historical specimens, sample preservation techniques, environmental degradation and contamination have significant effects on quality (Wandeler et al., 2007). In analysis of biological sample types, little difference was observed in the type of tissue used from avian museum specimens (Whooping Cranes, *Grus americana*), when comparing bone and footpad tissue (Glenn et al., 1999). However, footpad skin from historical avian specimens (proximal phalanx) is recommended by Mundy et al. (1997), due to its superior DNA extraction quality compared to feathers and other tissues. Sampling from the footpad skin is also advantageous in that in the majority of avian species it does not adversely affect the morphological characteristics which define taxonomic classifications, and is thus minimally invasive and destructive in terms of the tissue removed from the specimen (Mundy et al., 1997).

Mitochondrial Gene & Species Identification

Use of mitochondrial gene markers as the sole element in the analysis of evolutionary patterns and phylogenetic relationships has been critiqued and often deemed controversial on both theoretical and practical grounds (Roe & Sperling, 2007). The use of mitochondrial gene data dominates much of the molecular systematics research, largely as a consequence of the ease of use. Gene choice and fragment length vary throughout the research, with little standardisation; often markers are selected on the basis of previous use and existing taxon specific developments as opposed to selection for maximum informativeness. The substitution patterns within genes directly influence phylogenetic relationships in accordance with the degree of variability

and/or conservation within the sequence, a consequence of functional gene constraints. The degree of variation is understood to differ between lineages and taxon, and increases significantly as taxa become increasingly diverged from a common ancestral lineage. The accumulation of mutations or informative characters as well as the loss of characters through saturation, over time results in phylogenetically informative population and taxonomic structures.

The use of mitochondrial DNA is significantly valued in the analysis of ancient and historical DNA samples, where the extraction of low quantities of genomic material precludes the use of nuclear DNA markers. In this instance, mitochondrial DNA has significant benefits to the analysis of historical specimens, enabling the inclusion of temporal analysis of species and population structures. Cytochrome *b* and Cytochrome c oxidase subunit 1 are two of the most commonly used genes in species identification, as previously discussed in Chapter 4. However in recent research, individual gene analysis has often been superseded by multi-gene analytical approaches to phylogenetic analysis. Cytochrome *b* is commonly utilised in the analysis of genetic variation both within and between species (Kocher et al., 1989; Parson, Pegoraro, Niederstätter, Föger, & Steinlechner, 2000; Tobe & Linacre, 2010; Tobe & Linacre, 2009; Verma & Singh, 2003), including forensic species identification and taxonomic identification. Early developments of universal species primers, for the Cytochrome *b* gene (Kocher et al., 1989), were instrumental in broad application of the Cytochrome *b* gene to speciation research and phylogenetics. The Cytochrome c oxidase subunit 1 (COI) gene of the mitochondrial genome has been utilised extensively in molecular systematics, its role as the DNA barcoding gene by Hebert et al. (2003) expanded its use in species delimitation and phylogenetic analysis dramatically in the past decade. The barcoding region of COI, approximately 650bp of the 5'end of the gene, has demonstrated extensive use in the development of the understanding of character evolution, evolutionary rates and relationships, and the analysis of population structures and phylogenetic relationships. Hebert et al. (2003) extol the virtues of the COI barcode as a critical tool in modern taxonomy for the delimitation of species boundaries and the diagnosis of new or cryptic species, however there has been much criticism of its role in taxonomy and its varied success among taxonomic groups

(Aliabadian et al., 2009; Kwong et al., 2012; Craig Moritz & Cicero, 2004; Roe & Sperling, 2007; Spooner, 2009; Valentini et al., 2009).

The mitochondrial Control Region has been utilised in both inter- and intra- specific variation analysis of many species of avifauna. The Control Region demonstrates some of the highest levels of variation in the mitochondrial genome between related species and is elevated with evolutionary distance (Marshall & Baker, 1998). Genetic distance between individuals of the same species may also be higher than some protein-coding genes due to hypervariability demonstrated in two domains within the Control Region. Paplinska et al. (2011) recognises the use of the Control Region from museum specimens to establish historical perspectives on Evolutionary Significant Units across temporal populations of threatened species. The high degree of variation in the Control Region observed in most species facilitates its use in establishing evolutionarily significant haplotypes which may be indicative of both extant and extinct populations. However, the utility of the mitochondrial Control Region may be limited in part by hypervariable site saturation which may underestimate sequence divergence mutation events associated with speciation. In addition domains and nucleotides may evolve at heterogenous rates therefore resulting in bias in the analysis of these regions (Huang et al., 2009). The hyper-variability of the Control Region also presents difficulties in the alignment of sequences, and therefore may present issues in the reliability and robustness of phylogenetic reconstructions (Angleby & Savolainen, 2005)

Role of Species Identification in Conservation

Cracids, including the Piping Guans are consistently difficult to monitor in the wild due to the nature of the habitat, behavioural traits and increasing scarcity (Alemu et al., 2005; Brooks, 2006; Galetti et al., 1997; Jimenez, Londono, & Cadena, 2003; Waylen, 2005). The Neotropical forest environment is dense and difficult to traverse, therefore it is difficult to employ traditional visual monitoring techniques such as point counts or transects (Jimenez et al., 2003). Aural surveys used in a number of Cracid surveys demonstrate intrinsic error in terms of distance errors as vocalisations sample

a much larger area, and demographic sex and age bias of vocalisations however, the technique can be more efficient than traditional methods in monitoring infrequently sighted species (Jimenez et al., 2003). The utility of a simple robust molecular test which unambiguously identifies Cracid species from biological samples, using a small amount of sequence data from the mitochondrial genome that can be amplified from almost any tissue type, has enormous potential in conservation monitoring research. Samples opportunistically collected from the wild, such as moulted feathers, eggshell, or faecal samples could be used to identify species biodiversity within a geographic range, monitor numbers of individuals and genetic diversity within a population and track phylogeographic changes in response to conservation concerns such as climate change.

Further to the identification of 'species' or evolutionary significant units, molecular techniques play a significant role in the detection of hybridisation and introgression (Randi, 2008). Examples of hybridisation within the Galliforme order are not uncommon, and is often a natural occurrence in parapatric contact zones between species of the same genus such as the *Alectoris* partridges, *Coturnix* quail (Randi, 2008) and the Alagoas Curassow *Mitu mitu* (Grau et al., 2003). The limitations of this type of analysis lie in the uniparental inheritance of the mitochondrial genome; male introgression is masked, as only the maternal lineage is represented. Thus, true hybridisation and admixture assessment should incorporate nuclear bi-parental markers, for a complete analysis.

Wildlife DNA Forensics is another area of research which may benefit greatly from the development of speciation tests, in the context of conservation through law enforcement. Cracids, and the Trinidad Piping Guan particularly are under threat from illegal hunting and over-harvesting. The ability to prosecute the perpetrators of illegal hunting, through the use of molecular speciation as an investigative tool in law enforcement, may act as a significant deterrent to illegal practices. Analytical techniques with the capacity to provide genetic evidence of a species' identity is a well established tool in forensic casework (Ogden, Dawnay, & McEwing, 2009; Ogden, 2011). The value of molecular species identification techniques in forensics lies in the

ability to provide a probabilistic match to known reference samples; in this context molecular techniques can categorically exclude species from identification and provide the probabilistic inclusion of a species as a source of evidential material (Angleby & Savolainen, 2005).

Genetic markers may also inform many other areas of forensic interest regarding an 'unknown' sample, in addition to species identification, including geographic origins, familial ancestry (paternal and maternal), and individualisation (Ogden et al., 2009). One of the most important concepts with wildlife forensics is the establishment of robust and reliable references, such as genetic data from taxonomically recognised museum specimens of voucher standard or reference collection of known species. Confidence in the accurate molecular identification of species in forensic casework, is dependent upon the degree of variation between species and species level indicative markers within the sequence (Linacre, 2008). Forensic practice requires rigorous standards and validation, equivalent to those required in human forensics, in order that the integrity of the analysis can be maintained for evidence presentation. The legislative protection of species of conservation concern, such as the Trinidad Piping Guan, may benefit greatly from the ability to accurately and unambiguously determine species from samples of unknown origins.

6.2 Experimental Methodology

6.2.1 The Samples

Analysis of inter-specific variation incorporated all of the available specimens from the wild Trinidad Piping Guans and historical Pipile collections of the British Natural History Museum and Museo de Zoologia da Universidade de Sao Paulo, Brazil (see Table 6.2 and Appendix 1).

Table 6.2: Samples used in Taxonomic analysis of the genus *Pipile*.

Sample Identifier	Description	Comments	Geographic Origin
TPG-R1	Reference 1	accession KU221051	Trinidad
TPG-R2	Reference 2	accession KU221052	Trinidad
TPG-E1	Egg-embryo	accession KU221053	Trinidad
TPG-P1	Pawi Feather 1		Trinidad
TPG-P2	Pawi Feather 2		Trinidad
TPG-P3	Pawi Feather 3		Trinidad
TPG-P4	Pawi Feather 4		Trinidad
TPG-P5	Pawi Feather 5		Trinidad
BNHM AC1	*Pipile cumanensis*	1902.3.13.1852	Peru
BNHM AC2	*Pipile cumanensis*	1953.68.44	Oriene, Equador
BNHM AC3	*Pipile cumanensis*	1940.12.5.43	Rio Juno, Equador
BNHM AC4	*Pipile cumanensis*	1856.11.5.10	Colombia
BNHM AC5	*Pipile cumanensis*	1922.3.5.185	Pomeroon River, British Guiana
BNHM AC6	*Pipile cumanensis*	1892.1.16.138	Takutu River, British Guiana
BNHM ACG1	*Pipile cumanensis grayi*	1910.7.9.107	Alto, Paraguay
BNHM ACJ1	*Pipile cujubi*	1889.6.1.254	Lower Amazon
BNHM ACJ2	*Pipile cujubi*	1850.11.30.21	Unknown, South America
MZUSP AC1	*Aburria cumanensis*	MZUSP 76361	Rio Macaua, Brazil
MZUSP ACG1	*Aburria cumanensis grayi*	MZUSP 'no number'	Unknown, South America
MZUSP ACJ1	*Aburria cujubi*	MZUSP 20832	Rio Tapajos, Brazil
MZUSP ACJ2	*Aburria cujubi*	MZUSP 21947	Rio Tapajos, Brazil
MZUSP CF1	*Crax fasciolata*	MZUSP 79216	Unknown, South America
MZUSP PT1	*Pauxi tuberosa*	MZUSP 28037	Unknown, South America

Reference sequences for the genus *Pipile* and family *Cracidae* were sourced through the combined DNA databases of NCBI (https://www.ncbi.nlm.nih.gov), see Appendix 7.

6.2.2 Ancient DNA Extractions

The extraction of DNA from museum samples of between 50 to 200 years in age and unknown preservation techniques required use of molecular techniques adapted for historical or ancient DNA (aDNA). The Qiagen DNeasy™ Blood and Tissue Extraction Kit was used to extract DNA from the museum specimens, both footpad skin (of the proximal phalanges) and plucked feathers, as described in Chapter 2.3. The Qiagen DNeasy kit has been found to be well suited to the extraction of aDNA, as silica based extraction systems removes PCR inhibitors which have been associated with aDNA (Kemp, Monroe, & Smith, 2006; Rohland & Hofreiter, 2007; Rohland, Siedel, & Hofreiter, 2010; Yang et al., 1998).

As previously outlined in Chapter 2.7, precautionary measures to control contamination were used in the extraction and PCR amplification of aDNA, all equipment was sterilised with disinfectants and UV irradiation at 254nm for three hours (pipettes, labware, pipette tips, microcentrifuge tubes, reagents and PCR water). Surfaces and cabinets were cleaned prior to work to physically decontaminate and destroy surface DNA in the work station environment for use with the museum specimens. Throughout the extraction process controls were used to identify potential contamination and ensure unambiguous authenticity of the DNA recovered. Replicate extractions were produced for the Sao Paolo museum feather specimens, however sample size for the British Natural History Museum footpad skin samples precluded replications of the DNA extraction. Blank controls were extracted alongside the museum specimens as a direct quality control for the extraction method and to identify potential contamination of the extraction reagents.

Quantification of aDNA Extracts Extracted DNA from the museum specimens was quantified by Qubit 3.0 Fluorometry with the Qubit® dsDNA HS Assay Kit (Invitrogen), as described in Chapter 2.4.

6.2.3 PCR & Primers

Optimisation of the method for sequencing was achieved by PCR amplification of the samples using the universal and species specific oligonucleotide sets. Primer sets were designed from the existing sequence data from the Trinidad Piping Guan genome (see Chapter 3) and reference sequence data from GenBank, targeting the hypervariable domain I of the Control Region, Cytochrome *b* and the barcoding region of Cytochrome oxidase subunit 1. Primers were designed specifically for use across species of the genus *Pipile* and to assist in optimum recovery of ancient DNA, through short fragment amplifications. Standard degenerate bases (see Table 2.1, in Chapter 2) were included in primers designed using Primer3 to account for species specific variation in the priming sites. A target amplicon length of 150 to 250bp was identified as the optimum achievable fragment length for these historical specimen DNA extractions, see Table 6.3.

Table 6.3: Primers used in amplification of the domain I of the Control Region, Cytochrome B, and Cytochrome Oxidase subunit I genes of the mitochondrial genome.

Gene	Primers (5'- 3')* #		Reference
Cytochrome B	L14841#	AAA AAG CTT CCA TCC AAC ATC TCA GCA TGA TGA AA	Kocher et al., (1989)
	H15149#	AAA CTG CAG CCC CTC AGA ATG ATA TTT GTC CTC A	Kocher et al., (1989)
	MUS1 F	ACT ACA CYG CAG AYR YYA C	This research
	MUS1 R	TCC TCA TGG GAG RAC ATA YC	This research
	MUS2 F	TYG GAC AAA CCC TAG TAG A	This research
	MUS2 R	GGG GAT GAA TAT GAG TGA GA	This research
	MUS3 F	TCT TCG CCC TAC ACT TTT TA	This research
	L15330AV	GGA CAA ATA TCA TTC TGA GG	Lee et al., (2008)
	H15551AV	GGG TGG AAT GGG ATT TTG TC	Lee et al., (2008)
	MUS26 F	AAC TTG AAA CAC RGG AGT AA	This research
	MUS26 R	TCC YCC TCA AGY TCA TTC TA	This research
Cytochrome B			
L14841→----------←H15149 MUS1 F→-------←MUS1 R L15330AV→-------←H15551AV MUS2 F→------←MUS2 R MUS3 F→----←MUS2 R MUS26 F→-------←MUS26 R			
Control Region domain I	L2ND6	TAC CCC ATA ATA CGG AGA GG	Robinson, 2010
	H2CR	GAG ATG TCC AGA CCT AGA TT	Robinson, 2010
	IntL2ND6	ATG TCC TWC RTG ACT AGC TT	This research
	IntH2CR	GCT TAG AAY AGG ACG CAW AC	This research
	MUS4 F	CCC CAT ACA TTA TGG TRC AG	This research
	MUS4 R	AATCCATCTGATACGTCGAG	This research
	MUS18 F	CTC TGC TAT TGG TTC CCT TT	This research
	MUS18 R	AGT GTC AAR ATG ATW CCB SAT A	This research
	MUS23 F	GYR TCT TTC TAR YCA CAT TTC	This research
	MUS24 R	CCA TTC ACT GGT TAG GTG AC	This research
Control Region			
L2ND6→----------← H2CR IntL2ND6→----------←IntH2CR MUS4 F→----------← MUS4 R MUS18 F→----------←MUS18 R MUS4 F→-----← MUS24 R MUS23 F→------← MUS4 R			

Table 6.3 continued.

Gene	Primers (5'- 3')* #		Reference
Cytochrome Oxidase subunit 1	LCO1490	GGT CAA CAA ATC ATA AAG ATA TTG G	Hebert et al. (2005)
	HCO2198	TAA ACT TCA GGG TGA CCA AAA AATCA	This research
	MUS6 R	TTT ATY CGT GGG AAT GCT AT	This research
	MUS15 F	ACC TGT GAC CTT CAT CAA TC	This research
	MUS15 R	ATG AGR GGR ACT AGT CAG TT	This research
	MUS10 F	AGG AAC CYT ACY AGG AGA CG	This research
	MUS10 R	TTT ATY CGT GGG AAT GCT AT	This research
	MUS8 F	ATA GCA TTC CCA CGR ATA AA	This research
	MUS8 R	GAR AYR CCT GCT AGR TGA AG	This research
	MUS9 F	AAC TGA CTA GTY CCY CTC AT	This research
	MUS9 R	GCT AGR TGA AGG GAG AAG AT	This research

Cytochrome Oxidase I

LCO1490→----------←HCO2198 (Hebert et al. Barcoding Region)

LCO1490→----------←MUS6 R

MUS15 F→----------←MUS15 R

MUS8 F→------←MUS8 R

MUS9 F→---------←MUS9 R

MUS10 F→----------←MUS10 R

* standard nucleotide coding system for degenerate base codes

nomenclature of primer names may not be consistent with nucleotide position on the avian genome (primers designed in mammalian research)

Preparation of the PCR reaction mixture was carried out as previously described (see Chapter 2), with additional precautions and control measures appropriate to the amplification of aDNA. Touchdown PCR was selected as the primary PCR amplification technique due to its function in improving amplification of difficult target DNA, improved DNA yields and improved accuracy, as previously described in Chapter 2.

Primerless PCR (pPCR) reactions were prepared using Illustra™ puRe Taq Ready-To-Go PCR Beads (GE Healthcare) in a total reaction volume of 25µl, including 10µl of template DNA and Nuclease free water. Post pPCR the product is used in standard

PCR reactions with appropriate primers for the target gene region; 8µl of pPCR template aDNA in a 25µl standard PCR reaction.

The Q5® High Fidelity 2X Master Mix protocol was used in a 25µl reaction volume as follows; 12.5µl of Q5® High Fidelity 2X Master Mix, 10µM Forward primer, 10µM Reverse primer, 8µl of pPCR template aDNA and Nuclease Free Water.

6.2.4 Sequence Preparation

Sequencing of the amplified mtDNA samples was carried out by MWG Eurofins (MWG Biotech) using the Mix2Seq Sequencing service, as previously described in Chapter 2. Purification of the PCR products was carried out using MinElute Gel Extraction Kit (QIAGEN). Prior to sequencing DNA amplifications were subjected to quantitative analysis and concentration as recommended by the MWG Eurofins sample preparation guidelines (2-5ng/µL). Quantification was carried out by agarose gel electrophoresis and image analysis in the Image Lab software (Bio-Rad) as previously described in Chapter 2.4.

6.2.5 Sequence Analysis

Quality Assessment & Confirmation of the Sequence Accuracy

Replication of sequencing and amplification of multiple overlapping regions was carried out to confirm the homology of the sequence data in a number of samples, thereby ensuring the quality of nucleotide assignment in the sequences and the validity of the sequence data as a representation of the species. Heterozygous bases were identified visually from the electropherograms in a number of the museum specimen sequences, as observed previously in the analysis of DNA from museum specimens (Wandeler et al., 2007). Where heterozygous bases were identified, sequencing was repeated in triplicate to establish true nucleotide assignment for the position.

All sequences were aligned prior to analysis; Forward and Reverse sequences were aligned using the BLASTn 2.2.25 program (http://blast.ncbi.nlm.nih.gov/Blast.cgi), to confirm the nucleotide assignment of the sequence results. Sequences were additionally analysed using the BLASTn software to establish closest species matches within the combined NCBI DNA databases.

Sequence Alignment

Multiple sequence alignments were performed using Clustal X 2.0 and Geneious R8.0 software with subsequent visual verification. All sequences were aligned with the reference sequence for the *Gallus gallus* whole mitochondrial genome (GenBank accession number X52392), as a standard reference point for nomenclature of the sequences.

Data Set Partitions

Data sets were constructed from the museum specimens, wild Trinidad Piping Guans and reference sequences obtained from the NCBI DNA database, for the mitochondrial Control Region, Cytochrome b gene and Cytochrome c oxidase subunit 1 gene. Each gene region was analysed individually, as reference sequence data from the NCBI database may not represent a single individual from a given species, therefore a combined gene approach may infer errors in the analysis of representative species.

Data sets were constructed to maximise the number of nucleotides in the given gene region alignment, the number of samples in the data sets were thus restricted to the availability of the sequence data for the given gene region (see Table 6.4). Reduction in the nucleotide alignment length facilitated the inclusion of increased numbers of samples, therefore multiple analyses were performed for the given gene regions.

Table 6.4: Specimen information for the final data sets used in taxonomic analysis of the Pipile genus and Cracidae

Data Set	Number of BNHM specimens	Number of MZUSP specimens *	Number of reference sequences (NCBI)	Alignment Sequence length (nucleotides)
Cytochrome b gene				
Pipile Cyt B 448	9	2	0	448
Pipile Cyt B 314	9	2	0	314
Pipile Cyt B 221	9	2	0	221
Cracidae CytB 448	9	2	53 (+ 1 outgroup)	448
Cracidae CytB 221	9	2	83 (+ 1 outgroup)	221
Cytochrome oxidase subunit 1				
Pipile COI 380	9	2 (+ 1 outgroup)	0	380
Pipile COI 159	9	3 (+ 1 outgroup)	0	159
Cracidae COI 338	9	3	63	338
Cracidae COI 159	9	4	75	159
Control Region				
Pipile CR 631	7	1 (+ 1 outgroup)	0	631
Pipile CR 408	8	2 (+ 1 outgroup)	0	408
Pipile CR 329	9	2 (+ 1 outgroup)	0	329
Cracidae CR 633	7	3	29	633
Cracidae CR 408	8	3	39	408
Cracidae CR 329	9	3	39	329

* MZUSP outgroup from the Cracidae order – *Pauxi tuberosa* or *Crax fasciolata*

6.2.6 Statistical Analysis

Statistical analysis of mtDNA sequences involve the categorical identification or exclusion of the samples from the reference groups, according to the observed sequence variations (Single Nucleotide Polymorphisms and Indels) and match analysis against the NCBI DNA database using BLASTn (Altschul et al., 1990).

Pairwise Distance Analysis

The degree of sequence dissimilarity was evaluated through calculation of the Pairwise Genetic Distance in the MEGA 7.0 software, under the appropriate model of substitution, as per Chapter 4.2.

6.2.7 Phylogenetic Analysis

The sequence alignments were analysed using the jModelTest 2.1.6 software through the CIPRES Science Gateway (Darriba et al., 2012), to identify the most appropriate model for reconstruction of phylogenetic trees. Data sets were analysed under the Hasegawa-Kishino-Yano (HKY) substitution model, as described in Chapter 4.2.6, which accounts for Transition and Transversion rate variations ($\alpha \neq \beta$) and unequal nucleotide frequencies ($\pi A \neq \pi C \neq \pi G \neq \pi T$).

Maximum Likelihood Analysis in MEGA7.0

The sequence datasets were analysed using a Maximum Likelihood model in the MEGA7.0 software (Kumar et al., 2016). As previously discussed in Chapters 4 and 5, Maximum Likelihood methods analyse the likelihood of observing a given set of sequence data using phylogenetically informative sites, under a given model of substitution (HKY); each potential tree topology is analysed and the highest (maximum) likelihood tree is the topology selected (Nei & Kumar, 2000). Bootstrap re-sampling was applied to infer reliability of the tree topology, bootstrap support at the nodes below 75% confidence were excluded from the topology.

Bayesian Inference Analysis in MrBayes 3.2.6

Bayesian Inference models were executed in MrBayes 3.2.6 (Huelsenbeck & Ronquist, 2001), through the CIPRES gateway (Miller et al., 2011). The Bayesian Inference method estimates posterior probability distribution to hypothesise evolutionary relationships between the sequences, thus constructing a consensus tree,

as previously described in Chapter 2. A Marcov Chain Monte Carlo sampling strategy was employed in MrBayes, using five million iterations sampled every 1000 generations with 25% burn-in discarded, under the HKY substitution model. Analysis was stopped or completed at the point of convergence (average standard deviation of split frequencies approaching zero). The concensus tree was constructed and posterior probabilities were calculated at the nodes, posterior probabilies below 0.75 were excluded from the topology. Bayesian Inference trees were visualised in FigTree v1.4.2.

6.3 Results

6.3.1. Data sets

A maximum sequence length of 593bp of the Cytochrome *b* gene was sequenced analysed in 11 of the museum specimens. Ten of the samples failed to amplify the complete sequence region of interest, producing three final data sets of 448 nucleotides, 314 nucleotides and 221 nucleotides in length, see Figure 6.4 of the unclipped CytB sequence alignment.

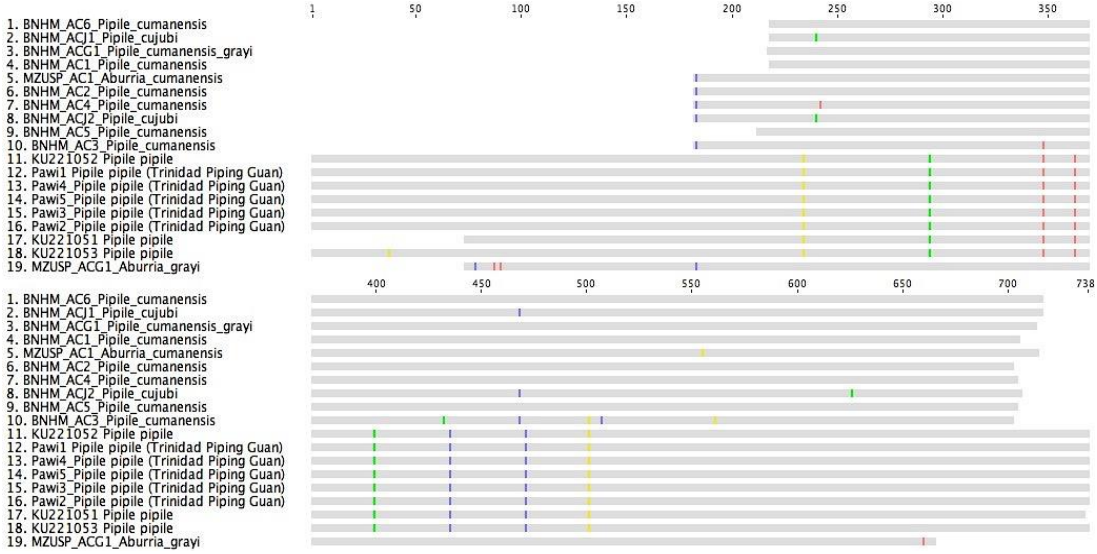


Figure 6.4: Alignment of *Pipile* species (*Aburria*) sequences of the Cytochrome *b* gene region, homologous nucleotides indicated in grey and single nucleotide polymorphisms indicated by colours, in Geneious R8.

Sequencing of the Cytochrome c oxidase 1 gene produced a maximum alignment length of 465 nucleotides of the barcoding region in 12 of the museum specimens. Three of the samples failed to amplify the complete 380 nucleotides sequence, producing a final alignment length of 159 bases in 13 specimens, see Figure 6.5 of the unclipped COI sequence alignment.

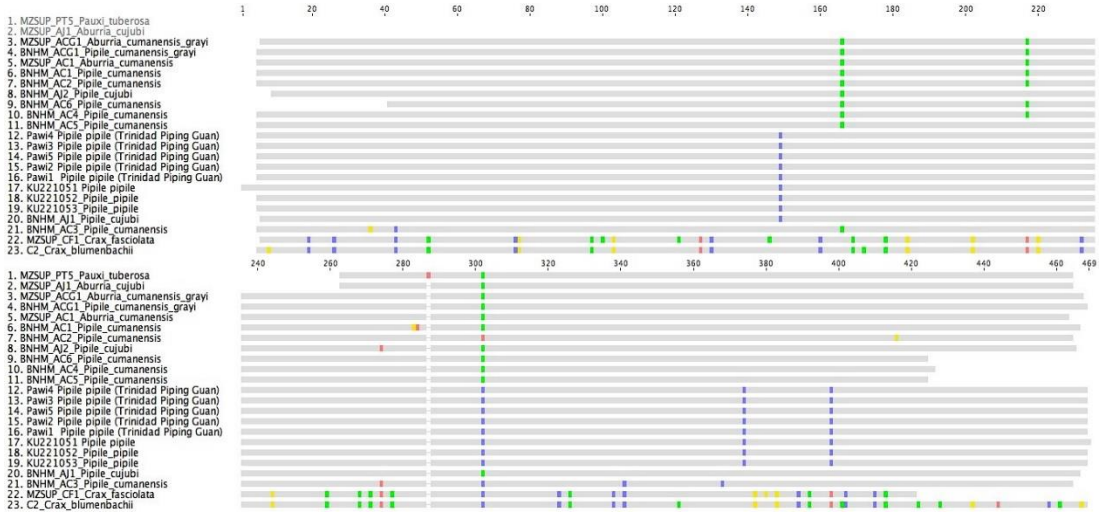


Figure 6.5: Alignment of *Pipile* species (*Aburria*) sequences of the Cytochrome oxidase subunit 1 gene region, homologous nucleotides indicated in grey and single nucleotide polymorphisms indicated by colours, in Geneious R8.

In analysis of the Control Region sequencing produced a maximum alignment length of 633 nucleotides in 11 specimens. Three of the museum specimens failed to amplify the complete region of interest, thus resulting in final alignment lengths of 408 and 329 nucleotides in 9 and 10 specimens respectively, see Figure 6.6 of the unclipped CR sequence alignment.

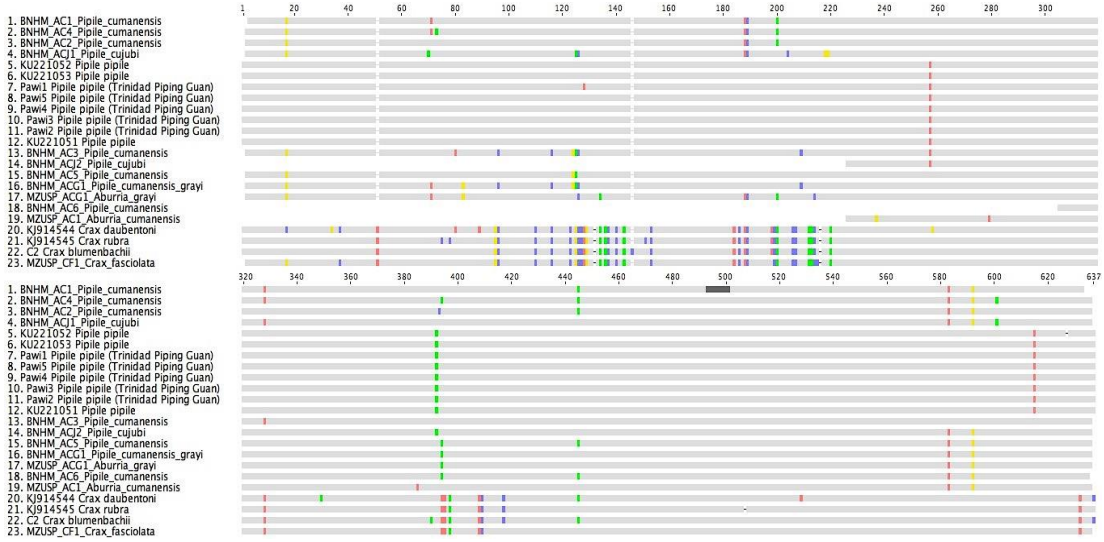


Figure 6.6: Alignment of Pipile species (Aburria) sequences of the Control Region, homologous nucleotides indicated in grey and single nucleotide polymorphisms indicated by colours, in Geneious R8.

6.3.2. Cytochrome *b* Gene Analysis

The complete Cytochrome *b* gene is 1143bp in length, at nucleotide position 14777 to 15919 in the reference *Pipile pipile* mitochondrial genome, accession number KU221051. The region of the CytB gene amplified in this study is located between nucleotides 14849 and 15492 of the reference KU221051 genome (positions 73 to 716 within the CytB gene).

Visual analysis of Single Nucleotide Polymorphisms (SNPs) for the CytB gene region, is indicative of species-specific nucleotides which are unique amongst the genus to *Pipile pipile*, in the available sequence data. A total of six Pipile pipile species-indicative SNPs were present within the 448 nucleotide alignment, see Figure 6.4. Pipile pipile specific SNPs were located at the following nucleotide positions; 15010-G, 15070-T, 15139-A, 15175-T, 15211-C and 15247-C. Additionally, a single species-specific SNP indicative of *Pipile cujubi,* was identified within this specific gene region, see Figure 6.4, at nucleotide position 15016-T. Phylogenetic reconstruction using the CytB gene demonstrates high posterior support for the separation of the Trinidad Piping Guan species basal of other mainland Blue-throated Piping Guans (*P. cumanensis*) and Red-Throated Piping Guans (*P. cujubi*), as indicated by the species-specific SNPs present in the alignment, see Figure 6.7. Within the phylogenetic analysis of the *Pipile* species, two representatives of the *Aburria aburri* (Wattled Guan) species were included in analysis from the NCBI DNA database (accession numbers AF165466 and AY354489). Concurrent with the previous analysis of Grau et al. (2005) and Hosner et al. (2016) using these sequences the *Aburria aburri* branches within the Pipile genus, basal of the blue throated Piping Guan species, indicative of a close genetic relationship between these species and a common genetic ancestry. Analysis of this gene region fails to differentiate between sub-species of the Blue-throated Piping Guans *P. cumanensis cumanensis* and *P. cumanensis grayi*, which branch together in a single clade.

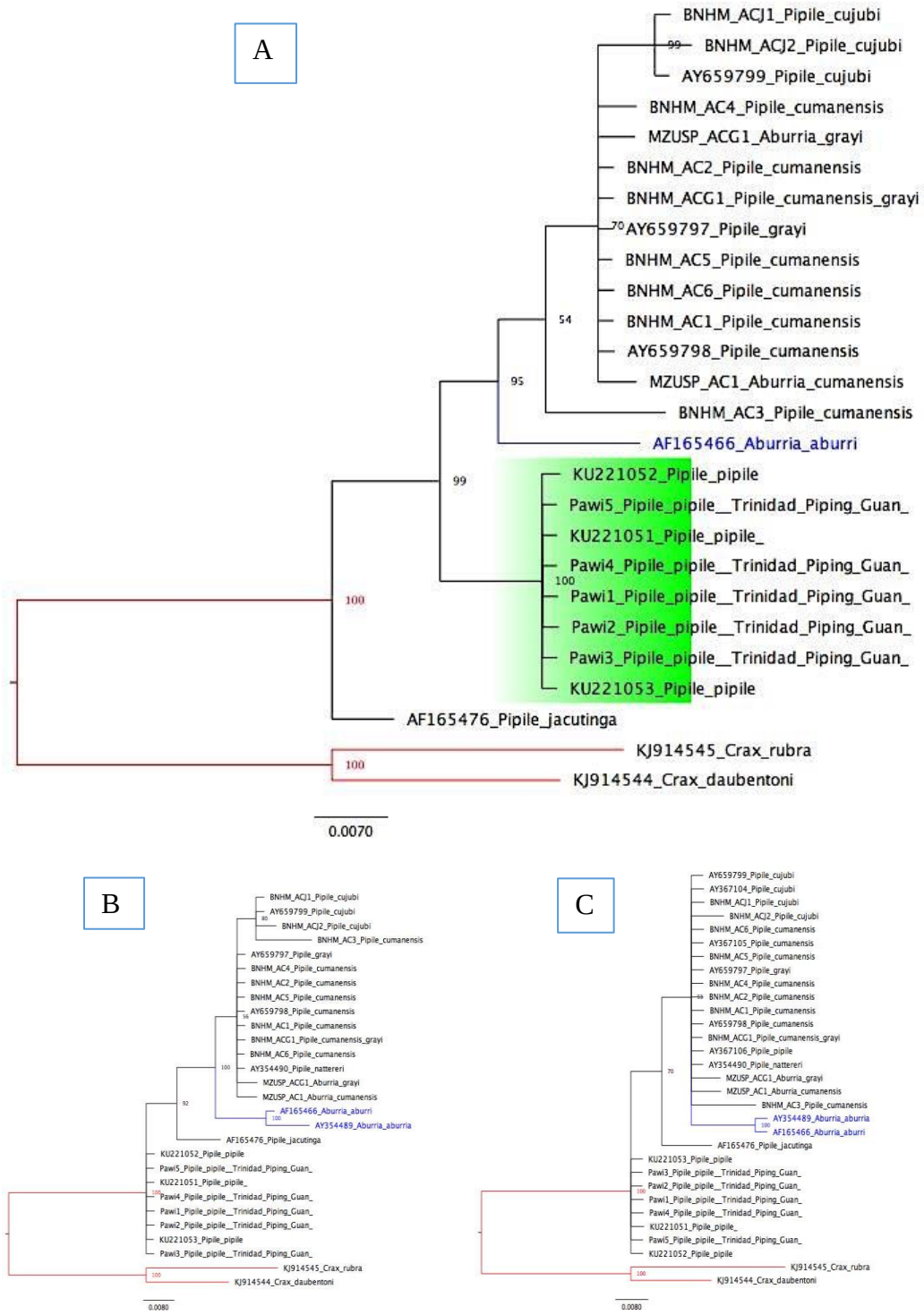


Figure 6.7: Phylogenetic analysis of *Pipile* species (*Aburria*) sequences of the Cytochrome b gene region, including specimens from the BNHM and MZUSP collections and wild Trinidad Piping Guans. Topology re-constructed under Bayesian inference MCMC method, using a HKY model of substitution with 5 million iterations sampled every 1000 generations in MrBayes. Posterior probabilities (%) to support topologies indicated at the nodes, Cracidae outgroup highlighted in red and Aburria aburri species within-clade inclusion indicated in blue. Analyses using an A.) 448 nucleotide alignment, B.) 314 nucleotide alignment, and C.) 221 nucleotide alignment of the Cytochrome b gene region.

Increasing the sequence length used in the alignment of CytB gene, improves the resolution of the analysis and provides a clearer definition of the species delimitations, as would be expected from the increased number of species-specific SNPs. The Red-throated Piping Guan (*P. cujubi ssp.*) branches in a close association with the Blue-throated Piping Guans, with the increased resolution of the 448 nucleotide CytB alignment the *P. cujubi* clade is distinguished from the other Piping Guan species. The reduced sequence length used also has a direct influence on the position of the Trinidad Piping Guan within the reconstructed tree; placing it basal of all other Piping Guans, as opposed to a sister relationship with the Blue-throated Piping Guans (see Figure 6.4). The shortened sequence alignment therefore fails to resolve relationships in a tree which is concordant with the known morphological relationships or the strongly supported evolutionary relationship between the mainland and Trinidad types. The phylogenetic analysis of the 448 nucleotide region of the CytB gene region for the Cracidae order included thirty-five Cracid species, including multiple individuals representing most species within the phylogenetic tree, see Figure 6.8. Analysis of the CytB gene region provides a well supported (>75%) topology which defines species-specific clades for the Crax, Northocrax, Ortalis and Oreophasis genus'.

The CytB gene fails to accurately resolve the Penelope, Chamaepetes, Mitu and Pauxi genus clades; the Chamaepetes branching within the Penelope genus clade and the Pauxi branching within the Mitu genus clade. In addition the Aburria genus is placed within the Pipile clade, as previously discussed. This analysis is indicative of limited accuracy in resolving relationships between the Cracidae species, using the 448 nucleotide region of the CytB gene alone.

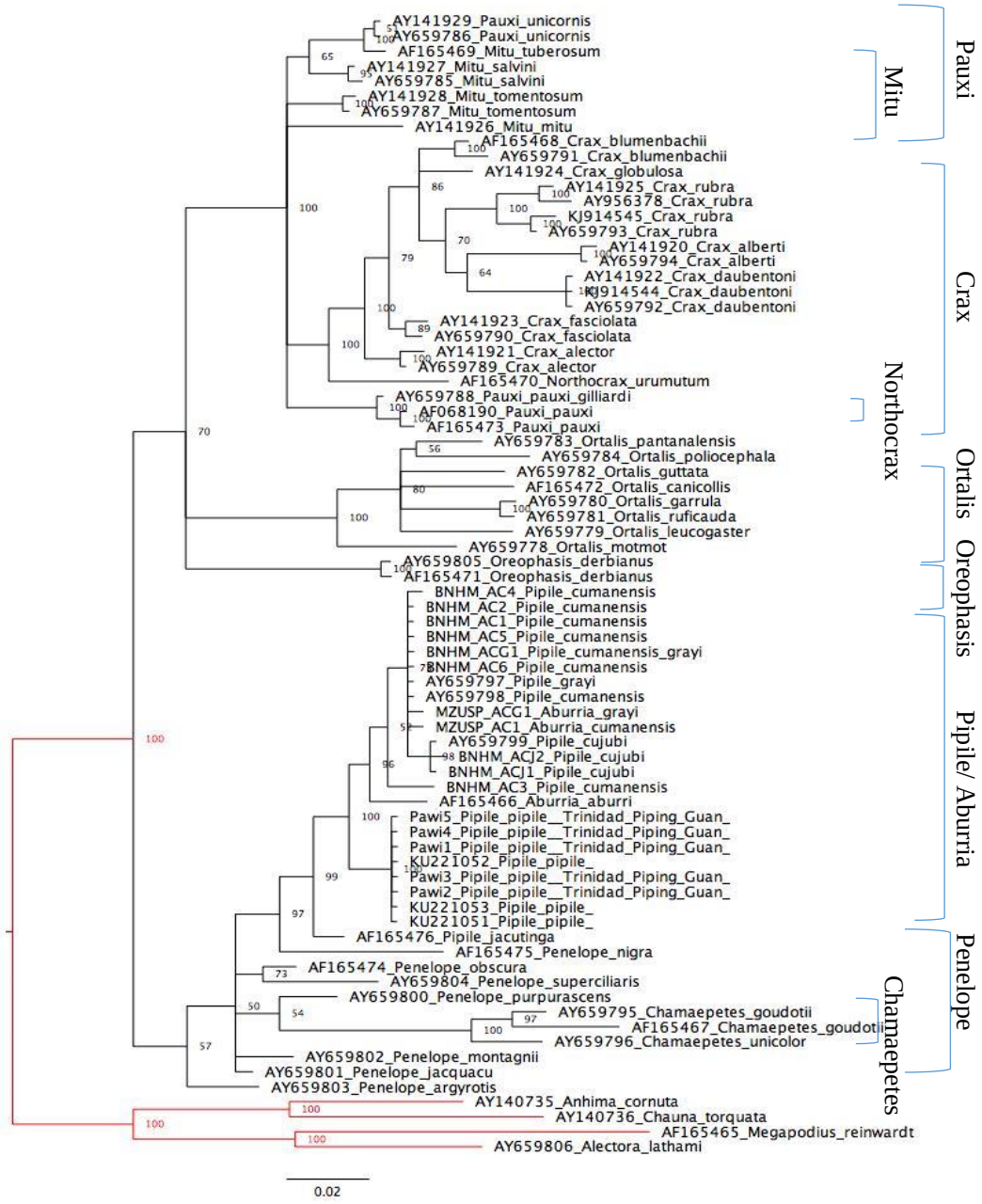


Figure 6.8: Phylogenetic analysis of *Cracidae* order, using 448 base sequences of the Cytochrome b gene, including specimens from the BNHM and MZUSP collections and wild Trinidad Piping Guans. Topology re-constructed under Bayesian inference MCMC method, using a HKY model of substitution with 5 million iterations sampled every 1000 generations in MrBayes. Posterior probabilities (%) to support topologies indicated at the nodes and outgroup indicated in red.

6.3.3. Cytochrome c oxidase subunit 1 Analysis

The Cytochrome c oxidase subunit 1 gene is 1551bp in length, positioned at nucleotides 6527 to 8074 in the reference *Pipile pipile* mitochondrial genome, accession number KU221051. Analysis of the 465 nucleotide region of the Cytochrome oxidase 1 sequence was located between nucleotides 6527 and 6991 of the reference sequence KU221051 (positions 5 to 468 of the Cytochrome oxidase 1 gene, part of the COI barcoding region (Hebert et al., 2004)).

Analysis of a 380 base segment of the COI barcoding region, failed to resolve clearly the genetic relationships between the species of the *Pipile* genus, see Figure 6.5. Within the Pipile clade the Trinidad Piping Guan species branch as a single group and are differentiated from other Piping Guans by two SNPs, see Figure 6.5 and Figure 6.9. Visual analysis of the alignment of *Pipile* species differentiates the Trinidad Piping Guan species with two species-specific SNP's at nucleotide positions 6896-C and 6921-C. Phylogenetic reconstruction of the gene region produced a topology which fails to differentiate between the Piping Guan species, with *Pipile cumanensis*, *Pipile cujubi* and *Pipile pipile* species branching within a single clade of closely associated sequences.

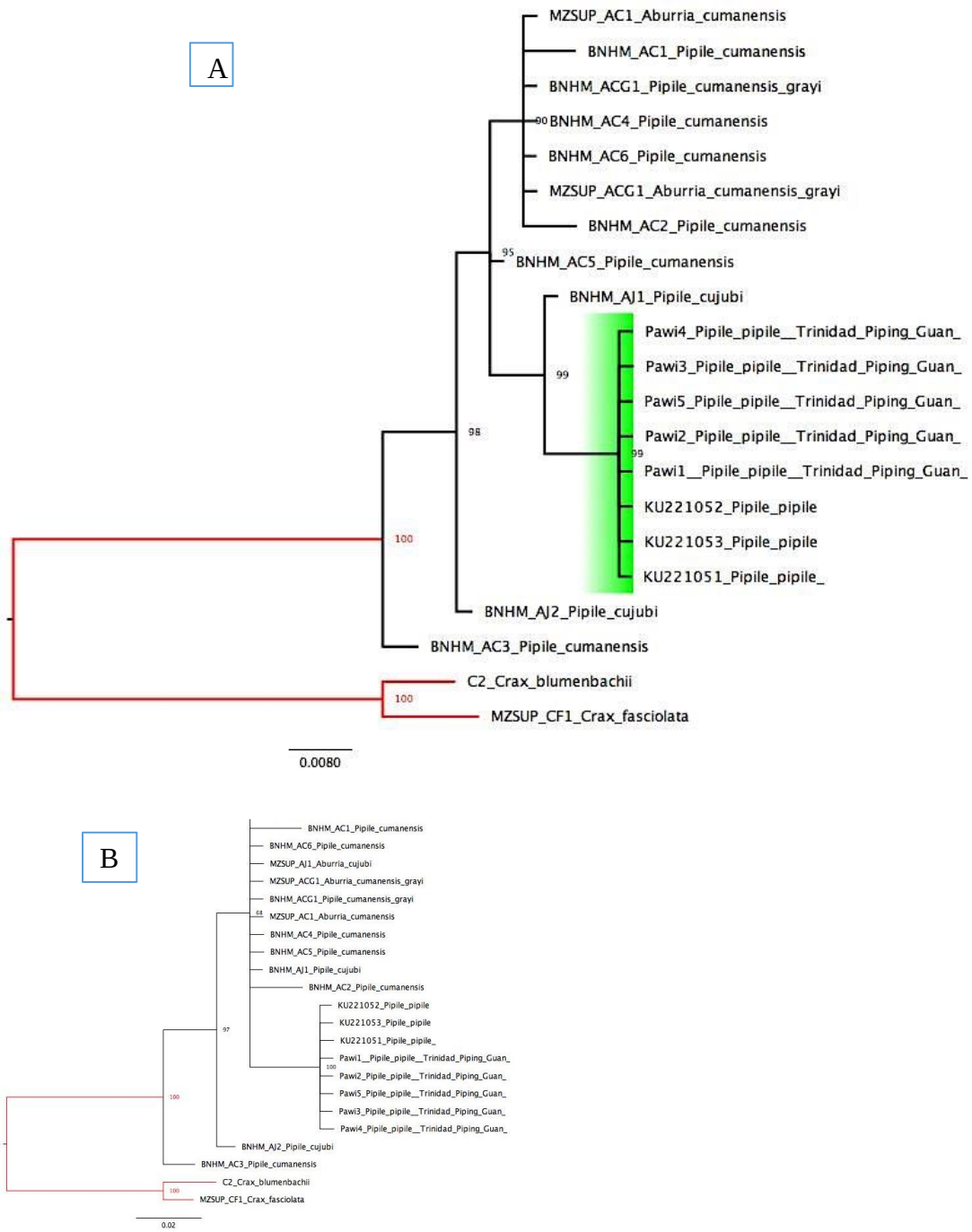


Figure 6.9: Phylogenetic analysis of *Pipile* species (*Aburria*) sequences of the Cytochrome c oxidase subunit 1 gene region, including specimens from the BNHM and MZUSP collections and wild Trinidad Piping Guans. Topology re-constructed under Bayesian inference MCMC method, using a HKY model of substitution with 5 million iterations sampled every 1000 generations in MrBayes. Posterior probabilities (%) to support topologies indicated at the nodes, and Cracidae outgroup highlighted in red. Analyses using an A.) 380 nucleotide alignment, and B.) 159 nucleotide alignment of the Cytochrome oxidase 1 gene region.

The Cracidae order is represented by COI gene sequences from twenty-six Cracid species, with multiple individuals representing each species in the NCBI DNA

database. The COI analysis of the Cracidae order provided a well resolved topology, with well supported and defined clades for the Chamaepetes, Ortalis, Crax, Pipile/Aburria and Penelope genus'. Structure within the Pipile genus remained consistent with that previously established in genus-specific analysis, see Figure 6.10.

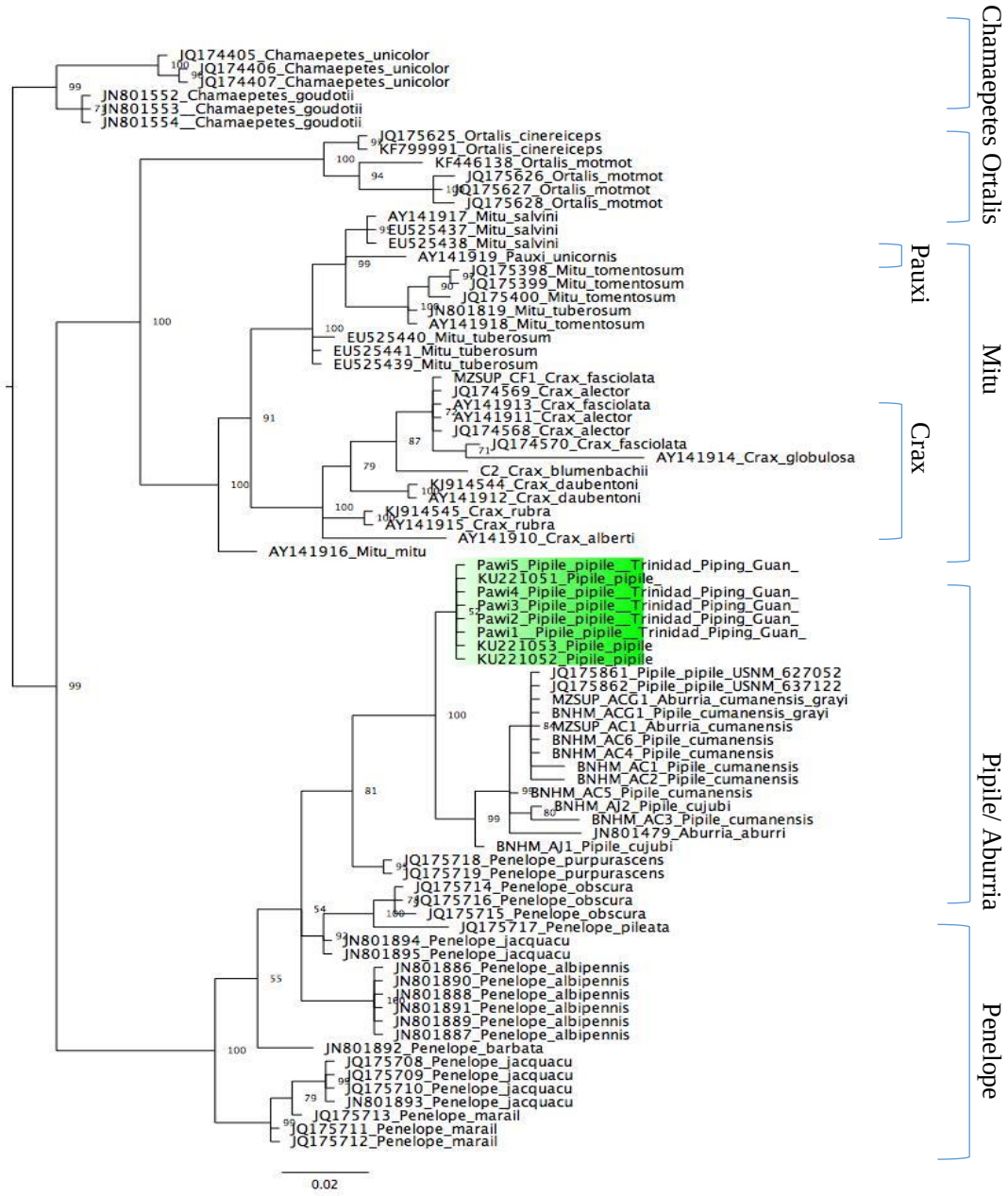


Figure 6.10: Phylogenetic analysis of *Cracidae* order using 338 base sequences of the Cytochrome c oxidase subunit 1 gene, including specimens from the BNHM and MZUSP collections and wild Trinidad Piping Guans. Topology re-constructed under Bayesian inference MCMC method, using a HKY model of substitution with 5 million iterations sampled every 1000 generations in MrBayes. Posterior probabilities (%) to support topologies indicated at the nodes (no outgroup is used).

The Crax, Mitu and Pauxi genus' remain unresolved, which may be indicative of bias due to the short sequence length used in analysis. Phylogenetic analysis of the COI gene region provides a good indication of species within the resolved genus', with high posterior support at the nodes.

6.3.4. Control Region Analysis

The Control Region of the Trinidad Piping Guan is 1124±2 nucleotides in length, shorter than that previously reported by Pereira et al (2004) at 1142bp in length, accession number AY145320. The Control Region is located at nucleotide positions 1 to 1124 in the reference *Pipile pipile* sequence, accession number KU221051. The amplified sequence region of the CR is located at nucleotide positions 110 to 742 in the reference sequence KU221051. Visual analysis of the *Pipile* alignment is indicative of a high degree of homology in the CR amongst the *Pipile* genus. Within the *Pipile* CR alignment, the *Pipile pipile* species is identifiable by a single unique SNP at nucleotide position 721-A. The CR alignment of the *Pipile* species provided a well resolved, highly supported phylogenetic tree which supported previous protein-coding gene analysis in isolating the Trinidad Piping Guan in a species-specific clade closely allied with the Blue-throated Piping Guans. This analysis however, failed to adequately resolve the relationship between the Blue-throated and Red-Throated Piping Guan species (*P. cumanensis ssp*. and *P. cujubi ssp*.). The increased resolution provided by longer nucleotide sequence length in the CR, provided an improved delimitation of the species within the Pipile genus. However, the *Pipile cujubi* sequence remains embedded within the *Pipile cumanensis* ssp. Clade, indicative of a limited ability to resolve the close relationships within the genus using the hypervariable CR.

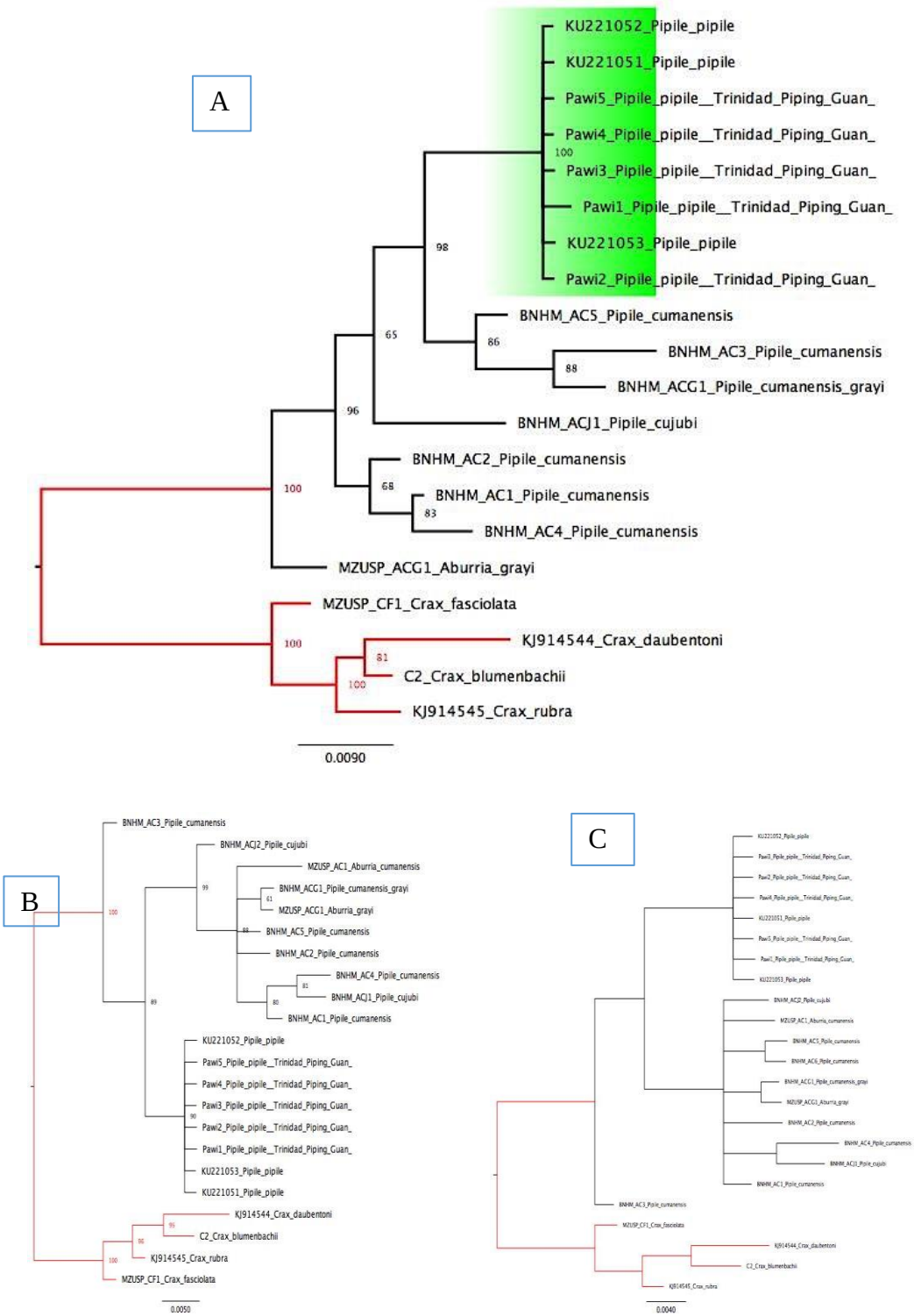


Figure 6.11: Phylogenetic analysis of *Pipile* species (*Aburria*) sequences of the Control Region, including specimens from the BNHM and MZUSP collections and wild Trinidad Piping Guans. Topology re-constructed under Bayesian inference MCMC method, using a HKY model of substitution with 5 million iterations sampled every 1000 generations in MrBayes. Posterior probabilities (%) to support topologies indicated at the nodes, and Cracidae outgroup highlighted in red. Analyses using an A.) 631 nucleotide alignment, B.) 408 nucleotide alignment, and C.) 329 nucleotide alignment of the Control Region.

The analysis of species relationships at the Cracidae order level was successful in resolving species relationships, using 633 nucleotides of the CR, see Figure 6.12. The Control Region phylogenetic analysis used twenty-five Cracid species, most of which are single individual representatives of the species.

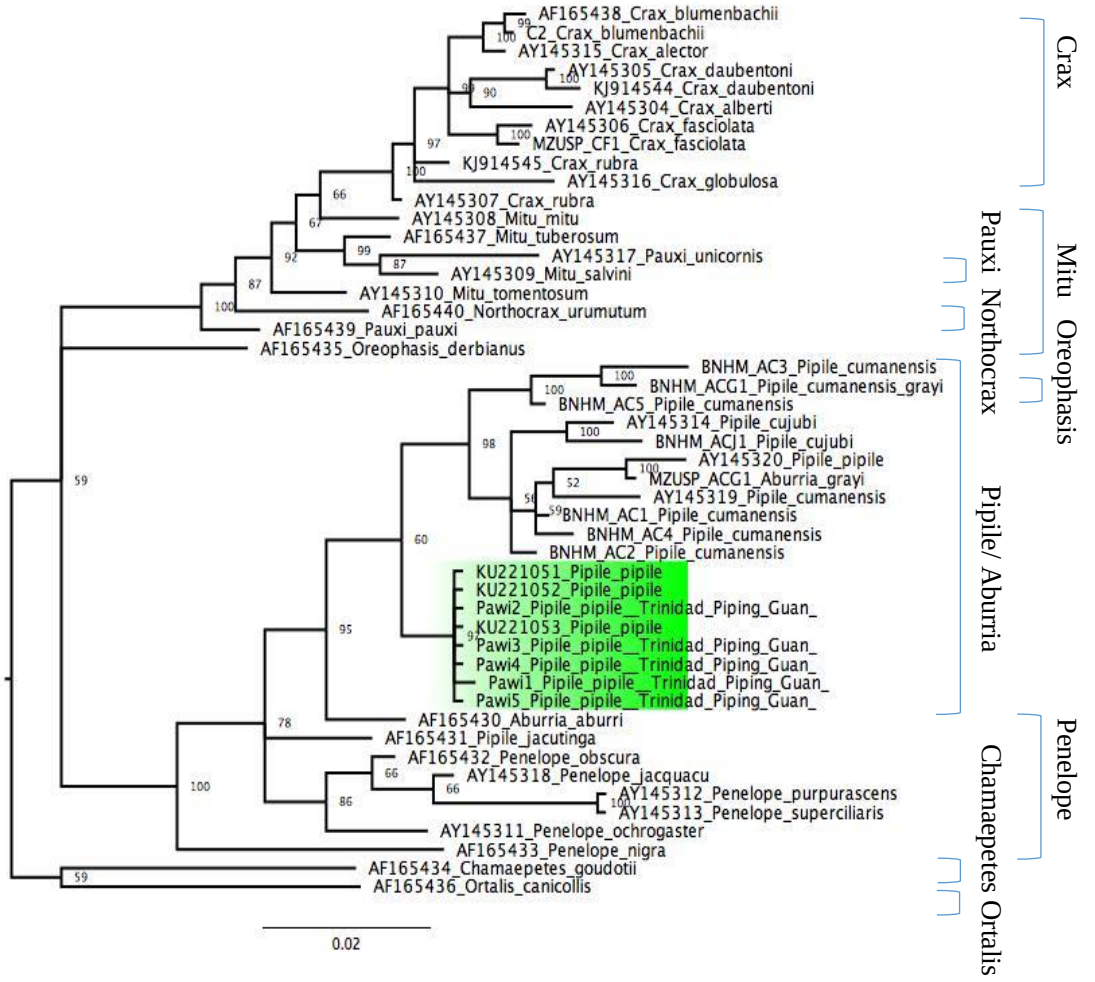


Figure 6.12: Phylogenetic analysis of *Cracidae* order using 633 base sequences of the Control Region, including specimens from the BNHM and MZUSP collections and wild Trinidad Piping Guans. Topology re-constructed under Bayesian inference MCMC method, using a HKY model of substitution with 5 million iterations sampled every 1000 generations in MrBayes. Posterior probabilities (%) to support topologies indicated at the nodes (no outgroup is used).

The 633 base section of the CR was successful in resolving the Crax, Penelope, Chamaepetes and Ortalis genus', with high posterior support for the nodes. However the Pauxi and Mitu genus' failed to resolve, branching in a clade with the Northocrax, support for these nodes was high indicative of a close relationship between these

species. This analysis used a relatively long section of the hypervariable (HV1) section of the CR, 633 bases, however it fails to clearly resolve species and genus level relationships amongst the Cracidae.

6.3.5 Genetic Distances

Genetic distances between species of the Pipile genus

In analysis of the genetic distances between the species of the *Pipile* genus the Trinidad Piping Guan is most closely associated with the *Pipile cumanensis* (Blue-Throated Piping Guan), see Table 6.5.

Table 6.5: Average genetic distances (p-distance) between species of the *Pipile* genus, for the 338 base region of the Cytochrome oxidase 1 gene, 448 base region of the Cytochrome b gene and 633 base section of the Control Region.

		Between Species Variation (p-distance)		
Species 1	**Species 2**	**Cyt B 448**	**COI 338**	**CR 631**
Pipile pipile	*Pipile cujubi*	0.02307	0.01842	0.02605
Pipile pipile	*Pipile cumanensis*	0.01897	0.02678	0.01926
Pipile pipile	*Pipile cumanensis grayi*	0.01860	0.02544	0.02443
Pipile pipile	*Pipile jacutinga*	0.02232	-	-
Pipile cujubi	*Pipile cumanensis*	0.00688	0.01053	0.01906
Pipile cujubi	*Pipile cumanensis grayi*	0.00595	0.00789	0.02262
Pipile cujubi	*Pipile jacutinga*	0.02753	-	-
Pipile cumanensis	*Pipile cumanensis grayi*	0.00298	0.00414	0.01745
Pipile cumanensis	*Pipile jacutinga*	0.02344	-	-
Pipile cumanensis grayi	*Pipile jacutinga*	0.02307	-	-

Analysis of the genetic distance between species of the genus Pipile demonstrates that the Trinidad Piping Guan, *Pipile pipile*, is closely associated with all of the *Pipile* species (p-distances 0.018-0.026). Using analysis of the short gene regions of the CytB and COI and CR, *Pipile pipile* is genetically closest to the *Pipile cumanensis* sub-species', as would be expected through geographical and morphological associations. The low genetic distances when compared to the Red-throated Piping Guans and Black-fronted Piping Guan (*Pipile cujubi* and *Pipile jacutinga*), demonstrates the close genetic relationship of these species. In analysis between the Blue-throated Piping Guan sub-species, the genetic distances are significantly lower than those observed at the species level (p-distance 0.002-0.01), indicative of the intra-specific relationship. This may therefore confirm the species level association of the Trinidad type *Pipile pipile* with the mainland conspecific Blue-throated Piping Guans.

It is important to note that the species definitions of the specimens used in this analysis are entirely based upon the classification of taxonomists (historical and contemporary), based on morphological characteristics which may be subject to classification error and may thus infer error in the analysis of closely related species.

Genetic distances within species of the Cracidae order

The comparison of genetic distances within the species of the Cracidae order in the mitochondrial gene sections is indicative of low levels of variation throughout the order. The Trinidad Piping Guan demonstrates a low level of variation in the CytB and CR, p-distances 0.005 and 0.006 respectively, and is 100% homologous in the COI gene. In the analysis of these short gene regions of the mitochondrial genome, the level of variation is comparative with that observed in other species of the order, However, the small number of sequences available for each species may significantly limit the analysis of variation for the majority of species, thus the analysis is only an indicator of variation at the species level.

In establishing the threshold genetic distance for 'species' definition, Hebert et al. (2004) indicated species level variation should be estimated at tenfold the intra-specific variation, in the COI barcoding region. Using this threshold, the general theory of a tenfold difference between inter-specific and intra-specific variation is generally not achieved in the analysis of species of the genus *Pipile* and in a broader context of Cracid species, see Tables 6.5 and 6.6. Intra-specific variation in the *Pipile* genus ranges between 0-0.008 in the COI and CytB gene regions, where as inter-specific variation is lower than expected by the tenfold threshold at between 0.01-0.026 (the CR was excluded as the non-protein coding hypervariable region of domain I has higher levels of intra-specific variation than protein coding genes). There is a distinct increase which differentiates between the inter- and intra- specific levels, however this increase is not in the region of a tenfold increase as predicted under the Barcoding theory of Hebert et al. (2004).

Table 6.6: Average genetic distances (p-distance) within species of the Cracidae order, for the 338 base region of the Cytochrome oxidase 1 gene, 448 base region of the Cytochrome b gene and 633 base section of the Control Region.

	Within Species Variation (p-distance)*		
Species	**COI 338**	**Cyt B 448**	**CR 633**
Aburria aburri	n/c	n/c	n/c
Chamaepetes goudotii	0	0.03125	n/c
Chamaepetes unicolor	0.00199	n/c	-
Crax alberti	n/c	0.00223	n/c
Crax alector	0	0.00446	n/c
Crax blumenbachii	n/c	0.00892	0.00163
Crax daubenotoni	0	0	0.00326
Crax fasciolata	0.00199	0.00669	0.00489
Crax globulosa	n/c	n/c	n/c
Crax rubra	0	0.01711	0.00489
Mitu mitu	n/c	n/c	n/c
Mitu salvini	0	0.00223	n/c
Mitu tomentosum	0.00499	0.00223	n/c
Mitu tuberosum	0.00898	n/c	n/c
Northocrax urumutum	-	n/c	n/c
Oreophasis derbianus	-	0.00223	n/c
Ortalis canicollis	-	n/c	n/c
Ortalis cinereiceps	0	-	-
Ortalis garrula	-	n/c	-
Ortalis guttata	-	n/c	-
Ortalis leucogastar	-	n/c	-
Ortalis motmot	0.01646	n/c	n/c
Ortalis pantanalensis	-	n/c	-
Ortalis poliocephala	-	n/c	-
Ortalis ruficauda	-	n/c	-
Pauxi pauxi	-	0.00446	n/c
Pauxi unicornis	n/c	0	n/c
Penelope argyrotis		n/c	-
Penelope albipennis	0	-	-
Penelope barbata	n/c	-	-
Penelope jacquacu	0.01277	n/c	n/c
Penelope marail	0.00199	-	-
Penelope montagnii	-	n/c	-
Penelope nigra	-	n/c	n/c
Penelope obscura	0.00199	n/c	n/c
Penelope ochrogaster	-	-	n/c
Penelope pileata	n/c	-	-
Penelope purpurascens	0	n/c	n/c
Penelope superciliaris	-	n/c	n/c
Pipile cujubi	0.00598	0.00148	0.01305
Pipile cumanensis	0.00812	0.00365	0.01707
Pipile cumanensis grayi	0	-	0.01468
Pipile jacutinga	-	n/c	n/c
Pipile pipile	0.00532	0	0.00616

*n/c single species representative, p-distance not calculable

Genetic distances within genera of the Cracidae order

Comparison of the level of variation observed within the genera of the Cracidae order indicates that the Pipile genus has relatively low levels of variation, see Table 6.7. The low level of variation observed is indicative of the close genetic relationship between the species of the *Pipile* genus. The genetic variation in the protein-coding genes CytB and COI within the *Pipile* genus is comparatively low in relation to other Cracid genera; approximately one third that of the other Guans (*Penelope* and *Chamaepetes*).

Table 6.7: Average genetic distances (p-distance) within genera of the Cracidae order, for the 338 base region of the mitochondrial Cytochrome oxidase 1 gene, 448 base region of the Cytochrome b gene and 633 base section of the Control Region.

	Within Genus Variation (p-distance)		
Genus	**Cyt B 448**	**COI 338**	**CR 633**
Aburria	n/c	n/c	n/c
Chamaepetes	0.03571	0.01596	n/c
Crax	0.03543	0.02648	0.01542
Mitu	0.02886	0.01497	0.01903
Northocrax	n/c	-	n/c
Oreophasis	0.00223	-	n/c
Ortalis	0.05213	0.02135	n/c
Pauxi	0.02455	n/c	0.03262
Penelope	0.04262	0.02655	0.03132
Pipile	0.01247	0.00970	0.01913

*n/c single species representative, p-distance not calculable

Analysis of genetic variation at the genus level is subject to the biases of the number of species representing that genus, in addition to the region of the gene or genome used in analysis. The *Ortalis* genus has a high level of variation (0.0052) in the CytB gene region, which may be due to the high number of representative species (n=8) compared to other genera in the analysis or may represent a greater evolutionary differentiation between the species of this genus, see Table 6.6.

6.4 Discussion

Natural History Museum Collections & Ancient DNA

The generation of sequences from *Pipile* reference samples from museum collections is of great utility to future analyses in many fields of research, and in the application of genetic information to the implementation of legislative protections and wildlife crime investigation. The use of historical samples is of interest in retrospective analysis of population structure, phylogenetic systematics and evolution, in relation to the biogeographic and environmental influences for species of concern. However, the value of museum collections is largely dependent upon the accurate taxonomic identification of the specimens within its catalogue, traditional classification systems were reliant on morphological and behavioural characteristics which may not be congruent with modern genetic classifications (Barraclough & Nee, 2001). Correct vouchering of specimens is critical to the observations which can be inferred in any phylogenetic analyses. Poorly kept records or badly preserved specimens in museum collections may result in the mis-identification of species and therefore the incorporation of errors into analyses (Olsson et al., 2008; Peterson et al., 2007). Voucher specimens from natural history collections enable reproducibility of the scientific method through preservation and provide a permanent record for future uses (Funk et al., 2005).

The extraction of DNA from historical specimens (50-200 years old) presents many difficulties in terms of sub-optimal sample sources, methodology, contamination, post mortem damage and PCR inhibitors (Kemp et al., 2006). The use of footpad skin from museum specimens was the most reliable source of DNA, in terms of both yield and quality, in comparison to cut feathers consistent with the analysis of Mundy et al., (1997). Extensive DNA degradation results in low yields of aDNA and low molecular weight DNA recovery. Mitchell et al. (2006) indicates DNA degradation is due to extensive damage to the phosphate sugar backbone through hydrolysis and destabilisation of the nucleotide bases (purines being lost more rapidly than pyrimidines). However, DNA degradation is temperature and environmentally dependant (Mitchell, Willerslev, & Hansen, 2005), therefore sensitive storage methods

may preserve DNA better and result in lower degradation. This study demonstrates the successful use of short PCR amplicons of 100-300bp from historical *Pipile* specimens between 64 and 167 years old, through the construction of large gene regions 465-635bp in length. The use of short overlapping sequences is a highly successful technique enabling a more robust analysis through improved sequence length and increased variable position nucleotides for phylogenetic reconstruction.

Preservation technique of museum skins has been linked to the occurrence of errors in DNA sequence, higher frequencies of sequence errors is associated with specimens preserved using formalin and arsenic techniques (Wandeler et al., 2007). C↔T transitions are a prominent post-mortem damage effect observed in aDNA (Paplinska et al., 2011; Wandeler et al., 2007). It was not known what techniques were used to preserve the museum skins used in this study, however to mitigate this issue the DNA sequencing was replicated to identify potential sequence errors and highlight variation which may be the result of post-mortem DNA damage. Techniques to 'boost' PCR have previously been described in methods to utilise aDNA, including primerless PCR, booster PCR and Touchdown PCR (Kemp et al., 2006; Weber, 2004). The use of such techniques in this study have proved to be highly successful in the amplification of DNA from *Pipile* museum specimens, where previous amplification using standard PCR techniques was unsuccessful. The use of museum specimens is distinctly limited by the degree of degradation exhibited by the individual sample, which may vary greatly between specimens, in general no DNA amplifications were achieved greater than 300 bases in length indicating a high degree of degradation in all samples.

Throughout this study it was paramount to ensure the integrity and validity of the DNA sequences produced from the museum collection specimens, of significant concern is the contamination of the data with modern DNA. Thus, it was important to consider contamination from both human and modern Piping Guan DNA sources, and therefore the authenticity of the sequence data considering the age, history and preservation of the individual specimens (Gilbert et al., 2005).

Species identification and the Pipile Genus

The classification of species is an important consideration in the conservation management of many species. Traditional classifications of the *Pipile* genus by Vaurie (1967) are largely consistent with modern taxonomy (del Hoyo et al., 2004; Lopes, 2009), based on the morphology of the genus. However, species delineation through morphological taxonomy has been linked to systematic errors and incongruences with genetic classifications in some species (Barraclough & Nee, 2001). Imperfect taxonomy was recognised by Funk and Omland (2014) as a major influencing factor on phylogenetic resolution, the consequences of which may be the failure of species to differentiate or over-splitting in the topology. Misidentification of samples and hybridisation are key issues in the question of taxonomy which would result in erroneous estimations of species level trees. As highlighted in Chapter 4, the introgression of hybrids in the captive Piping Guan population has significant effects on the genetic analysis of the genus. The need to sample broadly across the species is highlighted by this study, the sampling strategy should be representative of the needs of the research questions however the diversity of the sampling strategy should where possible sample multiple individuals across geographical populations as well as closely related congeners. The potential of hybridisation to influence genetic analysis is critical, closely related species which remain able to hybridise possess the potential to influence analysis through the introgression and transfer of alleles across species boundaries. The matrilineal mode of inheritance of the mitochondrial genome, means analysis of regions of this genome are thus highly susceptible to the influences of hybridisation.

The degree of variation observed between species is directly associated with the temporal distance between species and the substitution rates of the genome or gene being observed, as has been discussed in Chapter 5. Relatively ‘young’ species demonstrate lower degrees of variation, and therefore require markers with higher levels of resolution in order to observe species level variation (Wink, 2006). Nuclear genome markers are of a higher resolution than mitochondrial markers, due to their bi-parental inheritance and more rapid rates of substitution. However, the character traits of the mitochondrial genome and the wealth of mitochondrial sequence data available

on international DNA databases, mean that mitochondrial analysis of inter-specific variation is often highly successful. In addition spatial isolation has significant effects on the level of genetic diversity within species or genera, the genetic legacy of the species being associated with behavioural and biogeographic forces influencing evolutionary pathways (Dohms, Graham, & Burg, 2017; Habel & Zachos, 2013; Macfarlane, Natola, Brown, & Burg, 2016; Sonsthagen et al., 2017).

Inter-specific Variation

Early phylogeneticists recognised that phenetic similarities (character, phenotype and morphology) can be of limited value in the estimation of genetic similarities and in the reconstruction of phylogenetic hypotheses (Wiley, 1981). By sampling multiple individuals from multiple congeneric species a measure of true variation within a species be assessed (Funk & Omland, 2014). The research of Funk and Omland (2014) demonstrates the need for complete species level analysis within a genus to determine true genetic relationships in the context of paraphyly and polyphyly. The close genetic relationships both within and between species of a single genus may be complex and confound phylogenetic anlaysis.

The choice of mitochondrial marker used for species identification and phylogenetic analysis has been widely discussed (Tobe, Kitchener, & Linacre, 2010; Tobe et al., 2009). The analysis of Tobe et al. (2010), in the reconstruction of mammal phylogenies, demonstrated advantages to both the COI and CytB genes independently; CytB showed a higher percentage of variable sites and demonstrated a higher degree of congruence with taxonomy, however COI showed higher degrees of intra-specific variation than CytB. The use of CytB was supported over that of COI in the analysis of congeneric and closely related species (Tobe et al., 2010). COI barcoding and CytB sequences have been well documented as a source of species diagnostic markers (Dawnay, Ogden, McEwing, Carvalho, & Thorpe, 2007; Ogden et al., 2009; Rob Ogden, 2011; Tobe, Kitchener, & Linacre, 2009; Tobe & Linacre, 2009), of particular utility in wildlife DNA forensics.

Roe & Sperling (2007) advocate the use of maximum amounts of mtDNA sequence to infer species boundaries, as the limits of DNA barcoding fragment size make it vulnerable to the errors associated with short sequence including skewed data caused by localised regions of both high variation and labile sequence. This is particularly relevant in analysis such as this study, where complete mitochondrial gene analysis is difficult due to the historical nature of the samples and the restrictions of aDNA amplification. A reduced level of bootstrap or statistical support is common in the analysis of DNA from historical specimens as a consequence of the use of shorter sequence lengths (Graham et al., 2004; Paplinska et al., 2011). This study demonstrates a high degree of support within the individual gene trees for the analysis of *Pipile* and Cracidae relationships. The reduction in the sequence length, used to increase the sample numbers in the phylogenetic reconstructions, generally decreased the level of statistical support for nodes in the trees and additionally reduced nodal bifurcations within the trees thereby reducing resolution.

Additionally, temporal restrictions in the evolution of the mitochondrial genome have a significant influence on the degree of variation observed between recently diverged species such as the Trinidad Piping Guan and its mainland counterparts. Evidence from the analysis of the Galapagos Hawk and its closest mainland relatives, demonstrates restricted levels of mitochondrial variation between species (Bollmer et al., 2006), which is indicative of recent colonisation events. Thus, low levels of variation between the Trinidad Piping Guan and its mainland conspecifics is concordant with a relatively recent colonisation and isolation history, which is demonstrated in the evolution of the mitochondrial genome.

This study demonstrates the utility of the Cytochrome *b* gene in the delineation of species boundaries in the closely related *Pipile* genus. The Trinidad Piping Guan and the Red-fronted Guan (*Pipile cujubi*) are differentiated from the remaining *Pipile* species using a 448 nucleotide section of the gene. The inclusion of specimens from the museum collections has enabled the re-construction of a more robust phylogenetic tree, populated with a more representative sample set for the genus, which has aided in the genetic definition of these species. The analysis of Grau et al. (2003)

demonstrated species divergences between Curassow species of between 2.6-3.0%, across a 760bp sequence from Cytochrome *b* and the Control Region, slightly higher than that reported in other Cracids including the *Penelope* Guans (1.4%) and Curassows (1.3-2.6%). The analysis of the Cytochrome *b* gene in this study demonstrates species divergences of between 0.2-2%, for the 448 base sequence alignment, similar to those observed in other Cracid genera.

The research of Hebert et al. (2003) proposed a threshold for delimitation of species using the barcoding region of the Cytochrome oxidase 1 gene of tenfold the intra-specific variation, in North American avian species this equates to a species threshold of 2.7% inter-specific variation based on an average intra-specific variation of 0.27%. This intra-specific level of variation is similar to that observed in other published research into variation in the COI barcoding region in avians generally, 0.27% intra-specific variation observed by Aliabadian, Kaboli, Nijman, & Vences (2009) and 0.24% intra-specific variation in Neotropical avians demonstrated by Kerr, Lijtmaer, Barreira, Hebert, & Tubaro (2009). Analysis of the COI gene in this study used approximately half of the barcoding gene region (380 nucleotides), which demonstrated inter-specific variation within the *Pipile* genus between 0.4-2%, lower than that observed in general analysis of the avian barcoding region (Aliabadian et al., 2009; Hebert et al., 2004; Kerr et al., 2009).

The Control Region is largely advocated in the analysis of species delimitation, due to its hypervariable regions which infer a higher degree of variation in a small region of the genome which is beneficial in the analysis of closely related species. In depth analysis of the Control Region in Cracid species by Pereira et al. (2004), demonstrated the utility of this region in the differentiation of Cracid species. This research utilised the hypervariable region of domain I, within the Control Region to observe species relationships within the *Pipile* genus. Observation of inter-specific variation between *Pipile* species in this hypervariable region was between 1.7-2.6%, which is not significantly higher than that observed in the protein-coding regions analysed in this study. In comparison with Parrots of the genus *Coracopsis*, divergence estimates between species of the genus *Pipile* are concordant with other island endemic species

and their associated mainland conspecifics (0.6-3% inter-specific divergence) (Jackson et al., 2016).

The combined analysis of the Cytochrome b, Cytochrome oxidase 1 and Control Region sequences is strongly indicative of close genetic relationships within the *Pipile* genus, despite a long and diverse biogeographic and evolutionary history (Grau et al., 2005; Hosner et al., 2016). The close association between species of the genus may be associated with behavioural and biogeographic factors influencing dispersal and geneflow, thus producing distinct mitochondrial groups congruent with the geographic structure of the genus.

Taxonomic Identification

One of the main objectives of this analysis was to determine the ability to genetically differentiate the Trinidad Piping Guan from its mainland conspecifics, with particular attention to the relationship with the *Pipile cumanensis* species as the most morphologically and genetically similar species (del Hoyo et al., 2004; Grau et al., 2005; Pereira et al., 2004; Vaurie, 1967). This study emphasises the close genetic relationships within the *Pipile* genus which are indicative of niche specialism and genetic resilience, through retention of the genetic form and function despite broad geographic distribution.

Throughout the analysis in this study the Trinidad Piping Guan mitochondrial DNA sequence has demonstrated a distinctive relationship with congeneric species of the genus Pipile. The use of specimens from contemporary wild Trinidad Piping Guans has clearly demonstrated species indicative unique genetic markers in the mitochondrial genome of the modern Trinidad Piping Guan type. Vaurie's (1967) definition of the Trinidad form as a 'geographical type' is supported by genetic evidence of the species' historical geographic isolation and the contemporary influences of genetic isolation on the mitochondrial genome (as discussed in Chapter 4). The Trinidad Piping Guans consistently branch in close relation to the other Blue-

throated Piping Guans, *Pipile cumanensis* species, indicative of a shared common ancestry which supports morphological observations. The distinction of the contemporary wild population of Trinidad Piping Guans from other historical *Pipile pipile* references in the NCBI DNA database has been highlighted in this study. This analysis indicates a significant variation from the historical Cytochrome *b* mito-type which is closer allied with *Pipile cumanensis* ssp. This may be indicative of the loss of historical mitochondrial haplotypes in the modern Trinidad Piping Guan, as supported by the low genetic variation observed between individuals of the modern population.

Additionally, in clarification of the Trinidad Piping Guan relationship to NCBI records for *Pipile pipile* voucher sequences JQ175861 and JQ175862, from NMNH[2] specimen numbers USNM-627052 and USNM-637122 respectively, the species record on the NCBI database may be incorrectly recorded.

NMNH collection records list the specimens as follows (Smithsonian Institute, 2017);

- USNM 627052 – *Aburria pipile*, origin Guyana
- USNM 637122 – *Aburria pipile cumanensis*, origin Upper Takutu, Guyana

NMNH classify 23 specimens under the *Aburria pipile* identification (two of which are specified as *Aburria pipile cumanensis* and one *Aburria pipile grayi*), all of which originate in South America, and none of which are from Trinidad. Phylogenetic analysis of these specimens using Cytochrome oxidase 1, is congruent with the biogeographic information of the NMNH collection records in associating the sequences with *Pipile cumanensis ssp*. (synonym *Aburria cumanensis ssp.*) specimens from the BNHM and MZUSP collections, see Figure 6.10. This disparity is likely a reflection of the differing classification nomenclature used;

[2] NMNH – National Museum of Natural History, Smithsonian Institute, Washington DC, USA

Vaurie (1967)	BNHM & MZUSP	NMNH
• *Pipile cumanensis*	↔ *Aburria cumanensis*	↔ *Aburria pipile*
• *Pipile cumanensis grayi*	↔ *Aburria grayi*	↔*Aburria pipile grayi*
• *Pipile pipile*	↔ *Aburria pipile*	↔ *Aburria pipile pipile* (assumed)

The complexity of classifications within this genus are clearly reflected in the disparity between sample collection records, known geographic origins, traditional classifications and molecular evidence. The importance of biological, descriptive and genetic congruence in resolving such issues of taxonomic interpretation should be highlighted as it may have significant consequences to future research and analysis of this genus. The accuracy of the museum record is crucial to the conclusions being made (Paplinska et al., 2011). It has been suggested that improved annotation of sequences and molecular data being entered into Genbank may be advantageous (Olsson et al., 2008); improved and more informative analyses may be performed if information such as geographical and temporal information was included with molecular data.

The inclusion of additional specimens of these species from museum collections has added substantially to the current genetic knowledge of these species. The Blue-throated Piping Guan clades are well supported in the gene trees of this study, with high posterior probabilities, however definition of the sub-species groups has failed to resolve using the gene regions in this study.

The sub-species of the *Pipile cujubi* taxonomic group are less well defined by this analysis, specimens used in this analysis are not defined to the sub-species level therefore analysis of this study cannot define theses classifications. Additionally, specimens of these species are fewer, therefore the resolution of the group is less well defined in phylogenetic analysis. The Cytochrome *b* gene region used in this study was highly effective in genetically distinguishing the Red-throated Piping Guans from other *Pipile* species, as a distinctive clade within the phylogenetic tree. Improved

resolution of these species may be achieved through the acquisition of further samples for analysis and increased sequence length use in phylogenetic reconstruction.

The Black-fronted Piping Guan consistently occurs basal of the other Piping Guan species in phylogenetic analyses, consistent with other research (Grau et al., 2005; Hosner et al., 2016). This species is less well represented in the DNA databases than other *Pipile* species, therefore the analysis is critically limited the availability of sequence information.

The relationship of the Wattled Guan *Aburria aburri* to the other Piping Guans of the genus *Pipile*, has been closely debated in recent years (Grau et al., 2005; Remsen, 2007). The molecular analysis of Grau et al. (2005) and that of Hosner et al (2016) are indicative of the close genetic relationship between the species; the *Aburria aburri* species phylogenetically being placed within the *Pipile* genus topology. The analysis of this study supports this phylogenetic structure, however all of these analyses are conducted using the same *Aburria aburri* reference sequences from the NCBI DNA database (accession numbers AY354489, AF165466, JN801479 and AF165430). A more complete and thorough analysis of the species relationships would require further sampling from contemporary wild Wattled Guans, historical museum specimens and other Piping Guans to establish genetic relatedness and population variations.

The proposal to merge the *Pipile* species into the *Aburria* genus relates to genetic evidence of the Wattled Guan (*Aburria aburri*) being embedded within the Pipile species phylogenetic topology (Grau et al., 2005). However, a sister relationship between the Wattled Guan (*Aburria aburri*) and the Piping Guans (*Pipile sp*.) has also previously been supported by genetic and morphological analyses (Pereira et al., 2002; Pereira & Baker, 2006). The merge of these species has been analysed by the South American Classification Committee, however it remains subjective and therefore lacks the support of the committee on the grounds of a lack of physiological support, phylogenetic inconsistencies and low statistical support for reconstructions, and the use of non-vouchered specimens in analysis (Remsen, 2007).

6.5 Conclusions

The use of museum reference collections has many benefits in the analysis of evolutionary, species, and population level relationships, both in temporal and geographic contexts. The Trinidad Piping Guan, *Pipile pipile* has been clearly defined by its contemporary mito-type as a genetically, geographically and morphologically isolated species under the Biological, Evolutionary and Phylogenetic Species Concept definitions, and may be clearly defined as an Evolutionary Significant Unit as a consequence of its unique mitochondrial genome. The close genetic relationship between the *Pipile* genus has been highlighted by this study, and the application of historical specimens to genetic analysis of this genus. This research provides a basis for future analysis of the species relationships and the evolution of the genus *Pipile* (*Aburria*).

Chapter 7. General Discussion

The Trinidad Piping Guan is a critically endangered Cracid endemic to the island of Trinidad, thus it is a species of significant import to the ecology and biodiversity of the region. Little is known of the ecology, behaviour, biology, and current status of this cryptic species (Alemu et al., 2005; Hayes, Shameerudeen, et al., 2009; Waylen, 2005), however the species is considered to be in ongoing decline. Prolonged unsustainable hunting, destruction of primary habitat and slow reproductive rates has had a significant impact on the Trinidad Piping Guan population, and has resulted in the decline of the species (Hayes, Sanasie, et al., 2009).

This study represents the most comprehensive account of the genetics of the Trinidad Piping Guan to date and has established for the first time reference mitochondrial genomes for three contemporary reference representatives of the species. Non-invasive sampling techniques including the use of moulted feathers and opportunistically recovered biological samples from wild Trinidad Piping Guans and captive individuals of the genus *Pipile,* were sourced as a means to analyse the genetic variation evident within the wild Trinidad Piping Guan population and captive conspecifics. In addition, invasive, but minimally destructive tissue and feather samples were obtained from museum specimens of the genus *Pipile*, in order to explore the use of old or ancient DNA (~50 to 200 years old) from historical samples provide contextual genetic information on historical populations and species relationships. The aims of this research were to establish the use of genetic markers in the analysis of this critically endangered species, with particular reference to identifying within species variation, evolutionary and taxonomic relationships within the genus *Pipile*.

Whole mitochondrial genomes

The use of non-invasive sampling strategies has been highlighted in this study, through the use of opportunistically collected moulted feathers to achieve complete mitochondrial genome analyses. Non-invasive feather samples traditionally yield lower DNA concentrations than invasive techniques such as blood and tissue sampling

(Bush et al., 2005; Hogan et al., 2008; Johannson et al., 2012). However, the use of non-invasive techniques such as moulted feather collection is highly beneficial in that there are no adverse effects to individuals of an endangered species, such as stress or increased mortality rates. Traditionally, whole mitochondrial genomes have only been achieved through blood or tissue analysis which result in a high DNA yields. This study has shown that it is possible to complete analysis of whole mitochondrial genomes from DNA extracted from moulted feathers of varying quality. Establishing such techniques may broaden the scope of sampling strategies for analysis of genetic variation and evolutionary genomics, and to facilitate analysis of cryptic or rare species which are not easily observed in the wild, such as the Trinidad Piping Guan. The techniques developed and used in this study have great potential for use in many other species, particularly where samples are temporally or environmentally degraded. Advancement in techniques such as those used in this study will have significant benefits in many reaserch areas in the future including conservation genetics, particularly for species which are rare or endangered. This has broad implications for conservation and forensic applictions where species identification is an important factor in analysis.

The mitochondrial genome is ideally suited to the analysis of genetic variation and genomic evolution, within the boundaries of the mito-genomic mode of inheritance. The mitochondrial genome offers insights into both inter- and intra- specific variation, providing evidence of ancestral evolution, historical and contemporary genetic patterns. However, as previously discussed, the uni-parental matrilineal mode of inheritance demonstrates a lower discriminatory power than the bi-parental analysis of nuclear genomic markers such as microsatellites or nuclear genome sequencing. The use of whole mitochondrial genomes in avian phylogenetics is widely acknowledged as producing a robust and reliable phylogenetic hypothesis (Jarvis et al., 2014; Kan et al., 2010; Meiklejohn et al., 2014; Slack et al., 2007; Sorenson et al., 1999; Zhang et al., 2014). Mito-genomic phylogenetic trees represent a statistically supported reflection of evolutionary hypotheses, which can be prescriptive of species boundaries when the rates of evolution are consistent between lineages and founder populations are monomorphic. However, the evolutionary relationships within the Aves class remain largely unresolved and inconsistencies between morphological, molecular,

mitochondrial DNA and nuclear DNA phylogenies highlight these discrepancies (Braun & Kimball, 2002). Sampling bias has been shown to have considerable implications for resolution of the avian phylogeny (Eo et al., 2009; Yang, 2014), inclusion of broader taxonomic sampling and the inclusion of less well studied species such as the Trinidad Piping Guan, is highly beneficial to reconstructions. Additionally, the use of complete genome data and larger data sets has demonstrated compelling improvements in the resolution of the avian phylogeny and evolutionary relationships, increasing statistical support and offering higher degrees of congruence with morphological phylogenies. Thus, this study has been able to provide a clearer analysis of the genus *Pipile* in relation to the Galliforme order and the wider Aves class, reconstructing phylogenetic trees with higher levels of statistical support.

Genetic variation

Genetic diversity within a species is critical to enable populations to adapt to future environmental challenges, through evolutionary process and natural selection (Franklin & Frankham, 1998). The Piping Guan species' of the genus *Pipile*, including the Trinidad Piping Guan have demonstrated pronounced declines in recent decades largely as a consequence of habitat loss and unsustainable hunting pressures (Begazo & Bodmer, 1998; Brooks, 1999; Galetti et al., 1997; Pereira & Wajntal, 2001). The reduction in population size is compounded by slow reproductive rates and the potential influences of the loss genetic diversity on reproductive success and fitness. It is understood that avian species with depleted genetic variation demonstrate evidence of reduced reproductive success through the effects of inbreeding and reduced fitness, influencing clutch size, survivability of offspring and hatching rates (Beissinger et al., 2008), factors which are vital for population recovery and stability.

As previously discussed in Chapter 4, the genetic diversity of a population is a balance between the inflow of new variation or alleles through mutation and geneflow, and the loss of variation through selection pressures and mortality, within the context of a species' breeding system and biology. Typically, large out-breeding populations are expected to demonstrate high levels of genetic variation. Migration and geneflow

between populations is an important influence on genetic diversity within populations; immigration and emigration can have dramatic effects on allele frequencies and overall genetic diversity. The Trinidad Piping Guan is considerably restricted in this capacity; the species is an island endemic therefore immigration is not possible and the reduction in population size has critically restricted geneflow within the species which may result in inbreeding effects. Inbreeding is the natural consequence of small closed populations, where matings among related individuals are inevitable, the genetic result of which is a reduction in heterozygosity and a loss of allelic forms. If the Trinidad Piping Guan continues to decline in numbers the effective population size may become too low to withstand the effects of inbreeding depression and genetic drift, therefore increasing its susceptibility to extinction. Loss of genetic diversity in the Trinidad Piping Guan is therefore a natural consequence of the rapid decline of population numbers compounded by critical life history traits and environmental specialism. Small populations such as that of the Trinidad Piping Guan may be critically linked to higher risks of extinction as a consequence of the loss of genetic diversity. Notably the effects of genetic decline and the increased extinction rates observed in small populations are significantly exacerbated in island endemic species (Frankham, 2008; Keller et al., 2001).

Small populations behave differently from large populations, in terms of genetic effects; inbreeding is more pronounced in small populations therefore the effects are exaggerated, and the influence of genetic drift and the loss of allelic diversity is more pronounced in small populations. Both long term and short term periods of reduced population size have significant impacts on the genetic diversity within a population and the extent to which diversity is lost. The evidently restricted mitochondrial diversity observed in this study potentially demonstrates the influences of small population size and may be indicative genetic evidence of historical founder effects and contemporary bottlenecks in the Trinidad Piping Guan.

In the future, were the genetic diversity and population numbers to continue to decline, the physiological effects of reduced genetic diversity may become apparent in the species, as have been observed in other avians (Beissinger et al., 2008; Nichols et al.,

2001). Evolutionary potential and the ability of a species to adapt to change over time requires a significantly higher effective population size, therefore enabling a species to retain evolutionary potential whilst maintaining the species unit (Frankham, 2005; Franklin & Frankham, 1998). The capacity of a species, such as the Trinidad Piping Guan, to adaptively evolve under selective and environmental pressures may be maintained through genetic diversity. However, this ability is often lost in small populations where there is a reduction in genetic diversity and an increase in deleterious effects (Frankham, 2005), which may have serious long term consequences for the Trinidad Piping Guan. Population viability analysis may be an important step in the management of conservation efforts for the Trinidad Piping Guan species, incorporating both demographic and genetic information. Critical analysis of deterministic factors for both the direct and indirect causes of decline in the species, such as habitat loss, over exploitation, pollution, etc., is important for strategic conservation planning in order to implement future management programmes. In addition, the analysis of life history traits and genetic factors play a significant role in viability analysis (birth and survival rates, mortality rates, susceptibility and genetic rates, population size, habitat capacity and resource limits). As previously discussed, genetic variation within a population is crucial to persistence and long term survival, short term depletion through inbreeding and genetic drift result in a loss of fitness, and prolonged reduction in variation reduce adaptability, therefore increasing extinction risk. Population persistence and genetic recovery are intertwined within natural population dynamics, the demographic factors which allow populations to recover directly influence the genetic health of the species. Genetic recovery is acknowledged as being largely dependant upon the immediate impacts of immigration (Tallmon et al., 2004), however the slower processes of mutation, genetic drift and selection play a significant role where immigration is not possible, as with the Trinidad Piping Guan species.

Evolutionary mito-genomics

The analysis of the evolutionary history of the Neoaves has significantly advanced in the last decade, through technological advances in molecular techniques and computational analyses. Despite these advances the phylogenetic reconstruction of the

relationships between modern birds remains unresolved, with many families and orders presenting cryptic genetic relationships or failing to resolve, as discussed in Chapter 5. Thus, the inclusion of additional genomes to improve species and taxonomic coverage, is a major step forward in resolving relationships within the avian class.

The avian class has been sculpted by the major biogeographical and climatic shifts since it arose in the Cretaceous, diversifying extensively through adaptive radiation. Galliformes are one of the most important orders of avians to human society, with a global distribution of species some of which have been highly domesticated. The Galliformes are one of the basal orders amongst the neognathes, closely allied with the Anseriformes, and the Cracidae represent one of the oldest Galliforme families, which pre-dominate in the Neotropical forests of South America. Despite the import of the Galliforme family, complete mitochondrial genome data has been determined only for approximately one fifth of the known species, therefore a clear representation of the phylogenetic and evolutionary relationships between the species of this order has yet to be resolved. The Cracidae are particularly under-represented in genomic analyses, therefore the inclusion of the mitochondrial genome of the Trinidad Piping Guan is a significant advancement in the analysis of this group of species.

The Cracids are one of the most primitive families within the Galliforme order, with ancient origins supported by fossil evidence (del Hoyo et al., 2004; Hosner et al., 2016; Tordoff & MacDonald, 1957). Phylogenetic relationships within the Cracidae have been demonstrated using morphological and molecular techniques, establishing the close genetic ancestry of species within this family, as outlined in Chapter 5.1. In this study the genus *Pipile* has been represented in complete mitochondrial genome analysis for the first time, with two *Crax* species of the Cracidae order. This study therefore has enabled the Cracids to be established within the Galliforme order and Avian class phylogenies, in support of the analysis of Meiklejohn et al. (2014). Bayesian Inference and Maximum Likelihood methods of phylogenetic reconstruction produced highly supported trees which were similar in topological construction, therefore indicative of a well resolved data set. Additionally, the analyses produced in

this study are highly concordant with that of other research (Jarvis et al., 2014; Kan et al., 2010; Meiklejohn et al., 2014; Slack et al., 2007). The use of increased taxon sampling within the phylogenetic reconstructions, reducing the distance between sequences, produces a more reliable and realistic estimate of the genetic phylogeny of this group of species. Thus, the inclusion of the Trinidad Piping Guan mitochondrial genome and other novel sequences within the analysis of the Galliforme order establishes a robust phylogenetic hypothesis, highly concordant with morphological and molecular findings.

The Trinidad Piping Guan and the species of the genus *Pipile* represents an important group in the analysis of evolutionary relationships. The Trinidad form of Piping Guan represents an island endemic form of the genus *Pipile*, which has been genetically isolated since the island of Trinidad became disociated from the South American continent. Island environments provide important systems of speciation and evolution which are defined by geographic boundaries, and species such as the Piping Guans which are not able to migrate long distances represent models of relatively recent and contemporary in-situ evolution. As outlined previously, island species are highly susceptible to the impacts of reduced geneflow, adaptive and selective pressures, which may manifest in reduced genetic structure, increased inbreeding and fitness effects. Additionally, island systems are subject to higher levels of extinction due to environmental constraints, catastrophic incident risk, anthropomorphic pressures and the genetic effects of isolation (Emerson, 2002; Frankham, 1997; Frankham, 2008). Thus, the geographical and anthropogenic barriers which influence the Trinidad Piping Guan represent a long-term obstruction to geneflow within the species, however also enable independent evolution of the genome in its island form. The evolution of genetically significant units is heavily influenced by the effects and rates of emigration, immigration, fecundity and mortality. Environmental and biological impediments of the Trinidad Piping Guan species sustantially restrict spatial dispersal and therefore geneflow is highly restricted in this species. The impact of population and demographic effects has had significant influence on the mitochondrial genome of the Trinidad Piping Guan, founder and bottleneck effects reducing genetic variation within the population. A dominant mitochondrial haplotype is therefore indicative of

a constrained evolutionary unit, unique to this species, which has evolved in isolation on the island of Trinidad.

Taxonomic identification

The taxonomic status of a species or sub-species directly influences the extent of legislative protection that species receives. As discussed in Chapter 6, the accuracy of taxonomic units and species delineation is essential in the recognition of endangered species or populations. Similarly, incorrectly delineated species are vulnerable through a lack of protection and may be adversely allowed to hybridise with closely related taxa, disrupting unique evolutionary lineages. Resources and conservation efforts incorrectly allocated to erroneously diagnosed species may thus endanger efforts to conserve evolutionarily unique species.

Evidence in other avian species is supportive of the biogeographical divergence and distribution of the *Pipile* species. Similar patterns of biogeographical diversification are strongly evident in the *Ramphastidae* Toucans (Lutz et al., 2013; Patel et al., 2011), *Pionopsitta* Parrots (Ribas et al., 2007), *Mionectes* Flycatchers (M. J. Miller et al., 2008) and *Brotogeris* Parakeets (Ribas et al., 2009). These species diversifications have been linked to major geographical events such as the last glacial maximum of the Pleistocene 19-26 Mya, the Andean uplift 6-10 Mya and connectivity of Amazonian and Atlantic forest lowlands 2-3.6 Mya (Lutz et al., 2013). These genetic and biogeographical diversification relationships are not restricted to avian species, but is evident across the taxa for example in *Lepidoptera* and primates (Di Fiore et al., 2014; Matos-Maraví et al., 2013). These long-term barriers to geneflow and geographical boundaries exerted a strong influence on species diversification across taxonomic groups (da Silva & Patton, 1998; Di Fiore et al., 2014; Lutz et al., 2013; Matos-Maraví et al., 2013).

Historical collections have traditionally been an essential tool of taxonomists and systematisists, however have been increasingly used in a wider field of scientific

applications including monitoring global climate change, species adaptation, biodiversity trends, biological invasions and phylogeographical changes, along with applications in forensics, public health and agriculture (Graham et al., 2004; Mundy et al., 1997; Suarez & Tsutsui, 2004). Natural history museum collection samples have been used in genetic research since 1984, when DNA was extracted from an extinct equid the Quagga (Higuchi, Bowman, Freiberger, Ryder, & Wilson, 1984), since then it has become common practice to utilise historical collections in the analysis of both extinct and extant species. Museum collections act as functional references of vouchered specimens (Funk, Hoch, Prather, & Wagner, 2005; Graham et al., 2004; Olsson, Sundberg, Alström, Gelang, & Ericson, 2008; Peterson et al., 2007); a requirement of any scientific process is that specimens are correctly identified and available as a resource should the analysis be questioned in the future. However, one of the limitations of museum specimens is the bias towards the aesthetic, samples being predominantly collected from 'good looking' specimens infers a bias on the collection which may not be fully representative of the species.

This study has highlighted the need for the use of correctly vouchered and representative species samples in many aspects of the genetic analysis of the genus *Pipile*. It is crucial that the data from these analyses be made available for future use in research within both the confines of the genus and the broader extent of the avian class. Thus, the data generated by these studies can be used in future research, and can be independently verified where necessary by the long-term storage of sspecimens and DNA from the samples used in this analysis. By maintaining collections such as these it is possible that these samples can be used in future research as technology improves and techniques become more advanced enabling higher levels of analysis. The collection of specimens used in this analysis may thus become increasingly important in the future, particularly in light of the decline of many species in the Neotropics.

The definition of species is varied depending upon the assessed criteria, over twenty definitions for delineating species exist utilising morphological, genetic and behavioural criterion, as discussed in Chapter 6.1. The subjective nature of species definition is further complicated by groupings within a species describing individual

sub-species, many of which require legislative and conservation protection in their own right. The genus *Pipile* comprises six closely related, morphologically distinct species and sub-species of Piping Guan; the Trinidad Piping Guan *Pipile pipile*, the Blue-throated or Common Piping Guans *Pipile cumanensis cumanensis* and *Pipile cumanensis grayi,* the Red-throated Piping Guans *Pipile cujubi cujubi* and *Pipile cujubi nattereri,* and the Black-fronted Piping Guan *Pipile jacutinga* (del Hoyo et al., 2004; Grau et al., 2005; Vaurie, 1967).

Genetic distance as a measure of genetic, character based differentiation is commonly used as a significant factor in contemporary delineation of species or evolutionary separated groups (sub-species, genus, order etc.). Therefore, genetic techniques may offer a definitive diagnostic method in taxonomic identification, depending greatly on the degree of homology and variation between the species or population of concern. Conspecific lineages should represent recognisable branches on the phylogenetic tree, sharing common ancestral states, with the possible extinction of intermediate or ancestral genotypes. As previously discussed, mitochondrial DNA is commonly utilised in the delineation of species, including DNA Barcoding of the COI gene and CytB gene analysis within the mitochondrial genome. In analysis of individual mitochondrial genes (12S, ND1, ND2 and CytB), García-Moreno, Sorenson, & Mindell (2003) demonstrated congruence in the ingroup topology in broad analyses of the avian class, indicative of stability across taxa in the mitochondrial genome. Segelbacher & Piertney, (2007) utilised the Control Region and Cytochrome *b* regions of the mitochondrial genome to determine sub-specific links to biogeography within the Capercailie (*Tetrao urogallus*), using both contemporary and museum samples. Analysis of short mitochondrial gene regions may establish high levels of species indicative diversity and historical phylogeographic patterns of distribution. In a multi-gene approach Pereira, Baker & Wajantal (2002) used combined analysis of three mitochondrial genes (CytB, 12S and ND2) and the nuclear DNA sequences of three exons (RAG-1, RAG-2 and c-mos) and an intron (beta-fibrinogen) to resolve relationships of the Cracidae, demonstrating the utility of such multi-gene approaches within the Cracidae. The degree of evolutionary substitution within a gene directly influences the ability of a gene or gene region in the analysis of taxonomic

relationships, and can limit the utility of some regions in the genetic differentiation of species.

This study has for the first time utilised *Pipile* species specimens from historical ornithological collections to further understand the genetic relationships within the genus and establish genetic markers for taxonomic identification of the species' of this important genus. Comprehensive analysis by Hosner et al. (2016) utilised all of the available genetic data (nuclear and mitochondrial DNA) to classify four highly supported clades within the Cracidae; the Guans (*Pipile, Aburria, Penelope* and *Chamaepetes*), the Horned Guans (*Oreophasis*), the Chachalacas (*Ortalis*) and the Curassows (*Crax, Northocrax, Mitu* and *Pauxi*). This study demonstrates phylogenetic reconstructions highly concordant with the analysis of Hosner et al. (2016), in the delineation of the major clades within the Cracidae. However resolution at the genus level fails to completely resolve the Cracid tree, indicative of a complex genetic ancestry. The use of museum specimens may be restrictive in this respect, limiting the analysis to mitochondrial genome inference and restricting the quantity of sequence data used in the analysis, thus restricting phylogenetic resolution.

Within analysis of the genus *Pipile*, the inclusion of multiple species representatives (in contrast to all previous analyses which use single species representatives) has enabled an improved resolution of species relationships. The Trinidad Piping Guan is clearly defined as a unique genetic species unit, most closely allied to the Blue-throated Piping Guans (*P. cumanensis* ssp.) congruent with morphological similarities and consistent with biogeographical speciation theory (Grau et al., 2005). This analysis also isolates the Red-throated Piping Guans (*P. cujubi* ssp.) as a distinct clade, consistent with morphological expectations.

The close genetic relationship established between the *Aburria* and *Pipile* genera is as yet unresolved. This genetic association has been disputed by the South American Classification Committee, on the grounds of poor phylogenetic resolution (Remsen, 2007). Additional evidence from the use of nuclear DNA excludes the *Aburria* from

the *Pipile* genus placing it basal of the *Pipile* species (Hosner et al., 2016), therefore being indicative of some level of error associated with the original analysis of Grau et al. (2005). Further genetic analysis is required to resolve this relationship, through a comprehensive analysis of multiple individuals of both the *Pipile* and *Aburria* genus. The inclusion of sequence data from the *Pipile* genus in this study, goes some way towards such an analysis, providing multiple vouchered representatives of the *Pipile* genus for comparison to the *Aburria* genus. This research has thus highlighted significant limitations in the analysis of this genome in relation to the correct vouchering of samples and the publication of reliable data to international DNA databases. Analysis of rare species and species which are not well documented, such as the Trinidad Piping Guan, is heavily influenced by the availability of accurate and reliable information on international databases. Thus, significant inferences regarding the species status and conservation management may be reliant on such data. Where information is incorrect or unreliable it has significant implications for the species in terms of the conclusions which are being asserted. The provision of reliable reference data for species in the genus *Pipile* is critical to future research, as this study has demonstrated significant limitations associated with the current available data.

Further analysis with improved species and individual representation of the *Pipile* and *Aburria* genera may better resolve the complex relationships of the Piping Guans. Museum collections may act as important sources for such specimens which may enable a more thorough analysis of conspecific relationships at an evolutionary, population, biogeographic and temporal level providing unique insights into the genetic ancestry of these important species.

Conservation Management

The biodiversity of the Neotropical regions of South America are some of the most diverse terrestrial environments globally with high degrees of species endemism, thus they are areas of key concern for conservation (Rull, 2011; Vegas-Vilarrúbia et al., 2011). A significant proportion of neotropical avifauna is endangered or vulnerable to extinction, including many species of the Cracidae order (Strahl & Grajal, 1991). As

introduced previously, the Cracidae are important arboreal frugivores and insectivores, which play a significant role in the forest ecosystem, including seed predation and seed dispersal (Pereira & Wajntal, 2001). Cracids have been described as key indicator species, species' within this order being highly sensitive to hunting pressures, anthropomorphic pressures, climatic shifts and habitat quality (Galetti et al., 1997; Grau et al., 2003; Hosner et al., 2016; Strahl & Grajal, 1991). Additionally, Cracids play a significant role in rural communities, providing a source of protein and economic benefits to native populations (Brooks et al., 2005; Strahl & Grajal, 1991; Waylen, 2005).

The Trinidad Piping Guan is a protected species under both national and international legislation, and has been designated as a recognised Environmentally Sensitive Species in Trinidad. However, despite legislative protection since the 1970's the species remains in a state of ongoing decline. As previously discussed, conservation of the Trinidad Piping Guan has broad implications socially and economically, the species' decline being strongly linked to anthropomorphic pressures of loss of primary habitat and unsustainable hunting. Long term conservation strategies need to consider the needs of rural communities as well as the status of the species, strong traditional and cultural values directly influence the sustainability of many species (Abensperg-Traun, 2009), in relation to the socio-economic characteristics of the region or country. Preservation of primary forest habitats has economic implications as Trinidad has significant agro-forestry, quarrying and agricultural needs in the current Trinidad Piping Guan species stronghold in the Northern Range (Government of the Republic of Trinidad and Tobago, 2010a). The role of Piping Guans as protein sources for rural communities is well documented (Brooks, 1999; Galetti et al., 1997; Gonzalez, 1999; Redford & Robinson, 2009), however subsistence hunting is no longer a major threat to Trinidad Piping Guans (Hayes, Sanasie, et al., 2009). Hunting is declining in Trinidad, through educational and conservation initiatives in the region (Waylen, McGowan, & Milner-Gulland, 2009), however it represents a recognised method of supplementing incomes in rural demographic groups (Government of the Republic of Trinidad and Tobago, 2010a). Support for ecotourism in Trinidad is improving, in relation to specific species of interest including the Trinidad Piping Guan, due to economic benefits to small communities such as Grande Riviera which support small

recognised populations of the species (Waylen, 2005; Waylen et al., 2009). The Trinidad Piping Guan is thus an important species culturally to the island of Trinidad, and is of increasing importance economically. This study has highlighted the need to increase research related to this species and, through increased knowledge of the species, enhance conservation efforts locally. Through a greater understanding of the species, the status and cultural importance of the Trinidad Piping Guan may be improved affording it better protection on the island of Trinidad.

Institution of captive breeding is a significant step in the management of a declining species; establishing a breeding population insulated from environmental change, catastrophic events and anthropogenic threats. Supplemental and supportive breeding in captive environments have shown great success in many species recoveries, including critically endangered avians species such as the Puerto Rican Parrot (Beissinger et al., 2008) and Mauritius Kestrel (Nichols et al., 2001). However, there have been many captive breeding programmes which have shown much lower success rate or have failed entirely, success in such programmes is generally considered to be low (11-38% of cases) (Robert, 2009). Ex-situ conservation techniques often represent a theoretical ideal of a stable self-sustaining population, however the evolutionary implications of captivity on the species may be significant. As previously mentioned the genetic effects of small populations cannot be underestimated, particularly with reference to closed captive populations; high levels of inbreeding, fixation of deleterious alleles and loss of diversity. In addition, the impacts of adaptation to begnin captive conditions may be highly detrimental to future reintroduction; the processes of artificial selection, unconscious selection (anthropogenic) and the relaxation of natural selection processes directly influencing species physiologically and behaviourally (McDougall, Réale, Sol, & Reader, 2006; Williams & Hoffman, 2009). Notably behavioural adaptations to temperament such as boldness, exploration, stress response and anti-predator traits, have strong associations with survival in the wild post-reintroduction (McDougall et al., 2006; Merrick & Koprowski, 2017), and may be highly detrimental to the reintroduction of species such as the Trinidad Piping Guan from captive populations.

Captive breeding may have many benefits to the Trinidad Piping Guan species, not least the establishment of a secure ex-situ population to supplement wild populations. Education of the public is essential in the establishment of protection for endangered species, and access to a captive population provides an important educational opportunity. Captive populations also provide research opportunities, not otherwise available, to study behaviour, the ecology and biology of species which are rarely observed in their wild situation, such as the Trinidad Piping Guan. Piping Guans may potentially benefit from ex-situ conservation efforts in controlled captive breeding situations. Augmentation of wild populations is an instrumental step in the recovery of significantly declining populations. Evidence from wild and experimental populations indicates introduction of individuals from outside a declining population improve reproductive fitness through the restoration of genetic diversity (Beissinger et al., 2008; Nichols et al., 2001; Tallmon et al., 2004). Galliformes generally are considered to be well suited to captive breeding and domestication, and evidence from Common Piping Guans (*Pipile cumanensis*) being bred in captivity in Argentina suggests that Piping Guans develop well in captive environments under the correct breeding conditions (Maragliano, Tartara, & Montalti, 2008).

Captive population management aims to maintain genetic integrity within a small stable populations therefore, there are significant genetic considerations to the management of a captive population. This study highlights appreciable genetic differences between the contemporary wild Trinidad Piping Guan population and the current captive 'Trinidad' Piping Guan population in the UK (supported by morphological evidence), indicative of introgression from outside the species, as discussed in Chapter 4. The importance of such a finding in the current captive population, to the conservation of the wild Trinidad Piping Guan should not be underestimated. The genetic integrity of the wild Trinidad Piping Guan is fundamental to the species, as it represents a uniquely isolated evolutionary form of the Piping Guan genus. Introduction of genetics from outside this evolutionarily unique species from a compromised captive population, could therefore be highly detrimental to the future of the species in its current form.

Selection of a genetically diverse founder population is critically important to the growth and maintenance of an ex-situ population, however the removal of individuals from an already small population size has significant implications for the wild population, the Trinidad Piping Guans are already demonstrating low levels of genetic diversity as demonstrated in Chapter 4.3. Retention of heterozygosity within an established ex-situ population is important in maintaining a healthy outbred population and the subsequent utility of the individuals when reintroduced to the wild. Harvesting for captive breeding may have a critical effect on the wild population, selective hunting based on phenotypic traits, sex or physical size strongly influence population structure; altered sex ratios, reduced mean population age, behavioural modification, as well as restriction of gene flow. Removal of a specimen from the wild without adversely affecting the conservation status of that species is of critical importance to vulnerable species, unsustainable harvesting is a significant issue for all species of Piping Guans, including the Trinidad Piping Guan (Brooks, 1999, 2006; Hayes, Sanasie, et al., 2009). Founding a captive population from individuals with limited genetic variation would result in higher degrees of inbreeding effects, and subsequent reintroduction of 'inbred' individuals may be detrimental to the health of the wild population (Bradshaw et al., 2007). Outbreeding depression is also a consideration with captive population reintroduction, where captive individuals have selectively adapted to captive conditions. Conservation efforts must consider the full impacts of genetic erosion in captivity as well as in the wild when considering reintroduction programmes.

The aim of in-situ conservation management is to increase population size through the reduction of declining influences and alleviating the effects of stochastic threats, in order to preserve the species for the future. Establishment of multiple populations in separate locations limits the risk of stochastic catastrophic events to a vulnerable population. Translocations and artificial dispersals may re-establish geneflow through the introduction of new individuals to augment populations in the wild. Natural geneflow and dispersal may be encouraged through the improvement of habitat quality, nesting sites, and the management of dispersal restrictions such as fragmentation. Therefore, habitat assessment is critical to recovery as the environment must be able to support its current and future populations, and must facilitate adequate geneflow for maintenance of the genetic stability within the population in the future

alleviating deleterious genetic consequences. Thus it is critical for all aspects of in-situ and ex-situ conservation management, for which genetics plays a significant role, to be considered in efforts to 'future-proof' the Trinidad Piping Guan in its natural environment.

The link between low genetic diversity and risk of extinction in wild populations has been well documented and the consequences of low genetic diversity for the Trinidad Piping Guan may be predicted through a comprehensive genetic analysis of the species. Sustained small population size in the Trinidad Piping Guan will have a direct impact on long term species survival and its ability to thrive in the face of future environmental and biological threats. A clear conservation management strategy needs to consider all aspects of the both in-situ and ex-situ conservation, and genetic analysis may be fundamental in strategic management for the future persistence of the Trinidad Piping Guan species. The information generated by this study may go some way to assist in the development of a clear conservation strategy in terms of genetic management, however the future of the species may be reliant on its prioritisation at both a local and international level as a species of conservation concern. Funding of conservation projects is critically limited and the prioritisation of the species for future research efforts may be of significant importance to the persistence of the species in the coming years.

Chapter 8. Further Research

This research has provided an insight into the genetics of the Trinidad Piping Guan and its mainland conspecifics of the genus *Pipile*. Throughout this study and publication of the whole mitochondrial genome, the focus has been on developing techniques to analyse the genetic diversity and ancestry of this critically endangered species. This analysis has clearly identified both the strengths and weaknesses of analysis using the mitochondrial genome to analyse the genetics of species of the genus *Pipile*, and the main focus for the future analysis should be the development of suitable nuclear markers to better define genetic variation and population structure in this species. Nuclear DNA analysis has the potential to provide a more in-depth analysis of genetic variation within this species and its relationship to continental species of the genus.

Recovery of DNA from Avian samples

This study has largely been restricted by the quality of samples recovered from the wild in Trinidad and the difficulties of extracting historical DNA from museum specimens. However, technological advancements and improvements in molecular techniques, such as those used in this study, may enable the analysis of poor quality samples which are temporally and environementally degraded. This study has clearly demonstrated that avian biological samples are of significant value to conservation and evolutionary research through the potential to recover genomic information non-invasively from critically endangered species. The amplification of the complete mitochondrial genome from feathers of this rare and endangered species, was a novel approach to mito-gomonic analysis, and has considerably advanced the current knowledge of the Trinidad Piping Guan species and the *Pipile* genus. Additionally, this study has utilised museum skins as a valuable source of DNA in the analyses of these little-studied species, which has provided a valuable source of reference data for future analysis. Further development of molecular techniques such as those which have been used in this study, may enable the use of sample types previously considered ‘unusable’ or ‘difficult’, such as museum skins, poor quality moulted feathers and

eggshells. Advancement in these techniques opens up possibilities for research which have previously been unavailable to us, such as the use of museum specimens and old or ancient DNA providing temporal genetic information which may elucidate population structure and evolution histories of these species. This research has highlighted the significant utility of museum specimens in enhancing the analysis of the genetic structure and evolution of species. A significant international collection of specimens from the genus *Pipile* exist which could be of great potential benefit to the contemporary species of the genus.

This study has laid the foundations for such analysis in the future, providing a highly successful set of primers specifically designed for use on historically degraded DNA of species in this genus. If a collaboration of natural history museum collection specimens were established, there would be the potential for the analysis of over two hundred specimens of this genus, this could potentially be the most significant DNA collection of any of the Cracid species, and would represent significant advancement in the genetic information potential of this important family. Further sampling of the Trinidad Piping Guan is vital to improving analysis of the contemporary population in Trinidad and analysis would benefit significantly from the inclusion of more historical specimens from museum collections around the world. The limitation of a small sample size has meant that specific analysis techniques such as genetic structure, migration, haplotype cluster analysis are not able to be performed. Increasing the sample size would enable a broader analysis of the contemporary population structure and haplotype diversity and distribution throughout the region. Future analysis of this species would therefore benefit significantly from a wider sampling strategy, incorporating more samples from wild Piping Guan species, captive Piping Gaun collections and historical specimens from museum collections.

One area of research which may be expanded upon is the analysis of the contemporary genetic variation within the species, in relation to historical perspectives. Bayesian Skyline analysis, is one technique which may assist in determining influential historical events in the genetic history of the species, such as historical bottleneck and founder influences. This type of analysis would link contemporary levels of genetic

variation in the species with demographic factors through prior distribution models, to determine how biogeographic events have influenced genetic variation in the Trinidad Piping Guan species.

Other nuclear DNA markers & nuclear genomic analyses

nDNA Microsatellite Development

This study has analysed extensively the mitochondrial genome of the Trinidad Piping Guan, in many formats (complete mito-genomic, evolutionary history, inter-specific relationships and intra-specific variation analyses). However, the analysis of the mitochondrial genome is critically limited by its mode of inheritance and restricted model of evolution. Thus, the mitochondrial genome does not enable a compete measure of the genetic diversity within the Trinidad Piping Guan species and its mainland conspecifics. Further analysis of the nuclear genome and nuclear genomic markers in future research should facilitate a more complete examination of variation within this species and its relationships to other species of the genus *Pipile*.

To date there is currently no microsatellite marker set for species of the genus *Pipile*, and few studies have researched microsatellite profiling of the *Cracidae*. Hughes and Larson, (2000) developed a set of six microsatellite markers for the Wattled Curassow, however this marker set was determined to be monophyletic in Piping Guans. Initial tests of the Hughes and Larson (2000) microsatellite set on the Trinidad Piping Guan samples proved to be inconclusive, most samples failed to amplify with the primers and only single fragments were amplified, consistent with Hughes & Larsons findings that most of the markers are monophyletic in Piping Guans, see Figure 8.1. Analysis of the current sample set proved difficult due to the poor quality of some of the samples, however the development of techniques in future research for the recovery of nuclear DNA from feather specimens may significantly improve this area of analysis. Further research into microsatellites is needed for a more complete analysis of genetic diversity in the Trinidad Piping Guan, either through species-specific microsatellite development or cross-species microsatellite analysis. Cross-species microsatellite assessment has been very effective in many species and there is great potential for the

development of a microsatellite panel from other Galliforme species, due to the significant quantity of research in this area (Barbará et al., 2007; Galbusera et al., 2000; Primmer et al., 1996, 2005). Commercial kits for *Gallus* may have utility in the design of future microsatellite analysis through cross-species development (Linacre, 2008) , however the evolutionary differences between the *Gallus gallus* as a modern Galliforme and the Trinidad Piping Guan as a basal or ancient Galliforme lineage may represent a barrier in such a development strategy. The analysis of microsatellite genotypes within the Trinidad Piping Guan population will benefit the species significantly, establishing information on genotype frequency and kinship for the first time in this species. Kinship within the Trinidad Piping Guan is of great significance in monitoring both current and historical population trends and levels of inbreeding within the contemporary population, and may be invaluable to the development and monitoring of a genetically diverse captive breeding population.

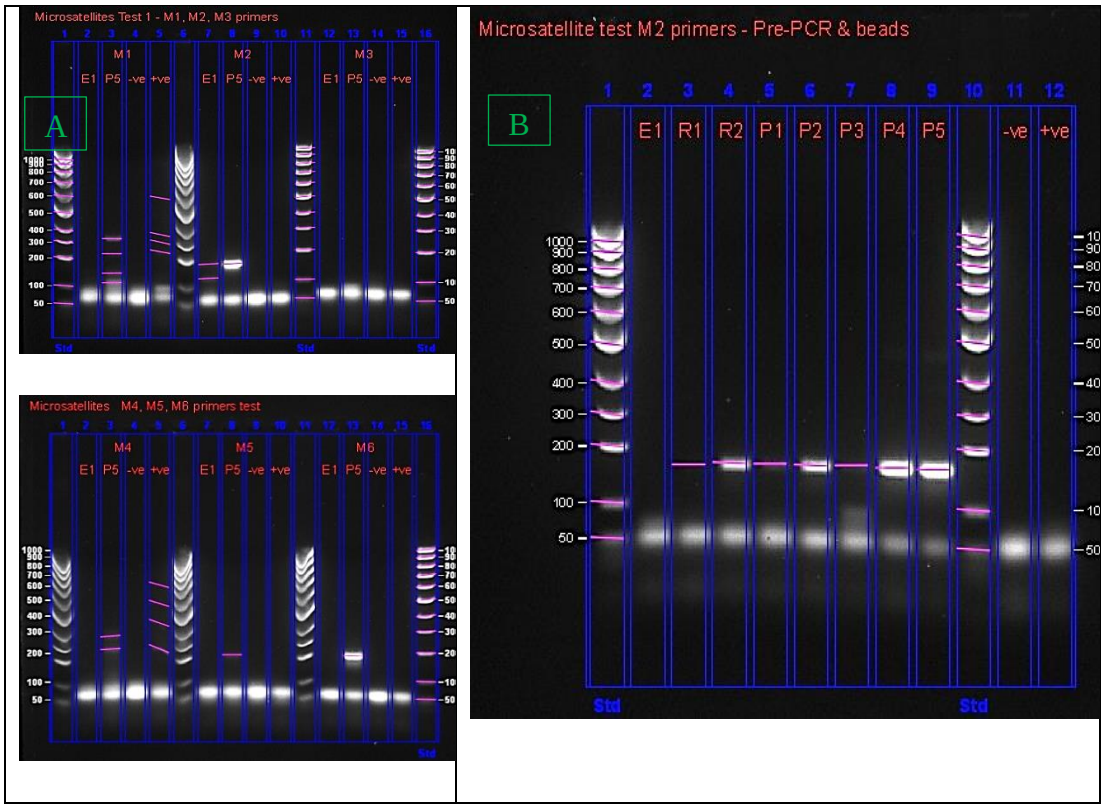


Figure 8.1: Gel electrophoresis image of the amplification of the Hughes & Larson (2000) Wattled Curassow microsatellite primer sets, using the Trinidad Piping Guan Pawi 5 sample. A). Gel 1 (top): Lane 1, 6, 11 & 16 Marker (DNA Ladder 100-1kb), Lanes 2-5, 7-10 & 12-15 microsatellite tests, Gel 2 (bottom): Lane 1, 6, 11 & 16 Marker, Lanes 2-5, 7-10 & 12-15 microsatellite tests. B). Lane 1 & 10 Marker, Lanes 2-9 Trinidad Piping Guan samples, microsatellite CgAAT32 (M2), Lanes 10 & 11 controls.

Total Genomic ISSRs Development

Alternative methods for the analysis of genetic variation and species level relationships have been used for many years to great effect including Inter Simple Sequence Repeats. Such techniques may be of great benefit to measuring genetic diversity in the Trinidad Piping Guan where conventional molecular techniques are not suitable or have yet to be developed, such as microsatellites. ISSRs use common microsatellite motifs, which are found widely across the nuclear genome, to amplify loci between repetitive elements within the genome. This type of analysis records variation across the genome at 10 to 80 loci in a single amplification (Wink, 2006), depending upon the presence of repetitive elements. As such it is a highly useful marker in population and species level analysis, and can be universally applied across most flora and fauna. Implementation of extragenic repetitive elements of the nuclear genome reveals high levels of variation both at the species and population level (Wink, 2006).

Initial experiments with ISSRs produced some interesting observations in variation between samples, predominantly in the variation between species, see Figure 8.2, indicative of low variation within the Trinidad Piping Guans, however further development of this techniques is needed to clarify the results. One issue that is raised with ISSR analysis is the inequality nuclear genome DNA in the Trinidad Piping Guan samples. As shown in figure 8.3, the quality of DNA has significant effects on the production of the ISSR loci; sample E1 in lane 4 produced only two bands indicative of low nuclear DNA yield, whereas a complete profile for this ISSR marker produced up to 21 bands in the Pawi4 sample in lane 11, indicative of high nuclear DNA content. Analysis with this type of nuclear marker has great potential to provide fast and effective markers of population variation, but it is entirely reliant upon a good quality DNA yield from all of the sample. Additionally, there is also potential for the analysis of inter-specific variation between closely related taxa, and potential for the development of a highly effective and quick species indicative test.

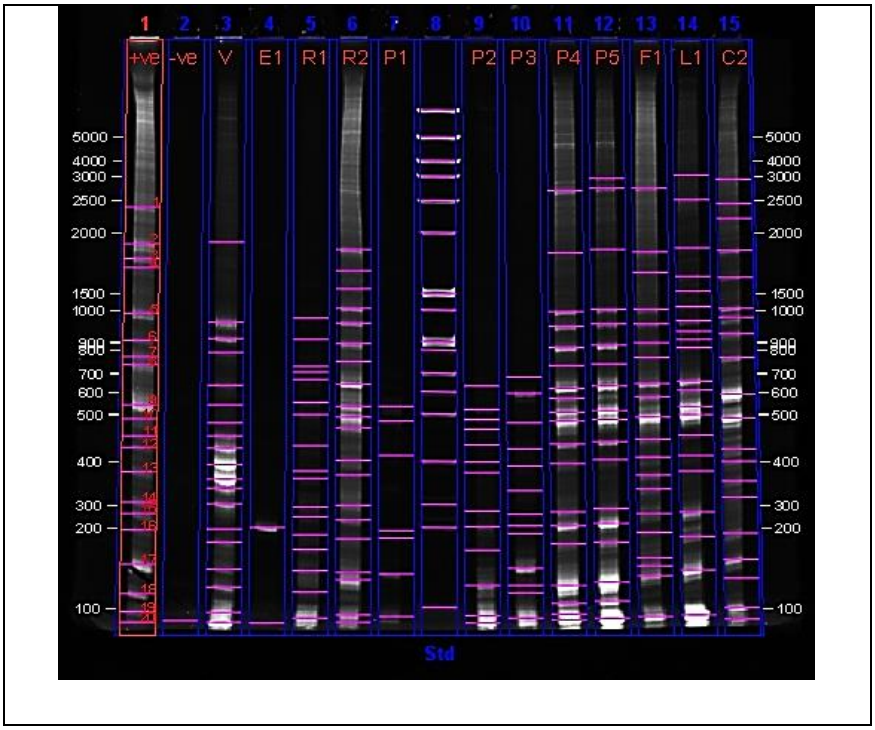


Figure 8.2: Polyacrylamide gel electrophoresis of ISSR $GACA_4$ amplifications in the Trinidad Piping Guan samples. Lane 8 Marker (DNA Ladder 100bp-2.5kb), Lane 1, 2 & 3 controls, Sample lanes = Lane 4 E1, Lane 5 R1, Lane 6 R2, Lane 7 P1, Lane 9 P2, Lane 10 P3, Lane 11 P4 and Lane 12 P5, Lanes 13-15 outgroup samples F1, L1 and C2 respectively (F1 & L1 captive *Pipile* sp. and C2 *Crax* sp.).

Next Generation Techniques & Whole Genome Analysis

The future of molecular biology and conservation genetics lies largely in the next generation techniques currently being developed. Jarvis et al. (2014) and Zhang et al. (2014) have demonstrated the utility of whole genome data in the analysis of avian phylogenetics, biogeography and evolutionary relationships. Whole Genome Shotgun sequencing techniques have demonstrated a significant capacity for inferring evolutionary traits within the avians (Zhang et al., 2014), despite the limitations of taxonomic sampling restrictions. The inclusion of more species and improved taxonomic sampling strategies will greatly advance avian genomics research, with potential benefits to species such as the Trinidad Piping Guan. Further research in this field may enlighten and enrich knowledge in a diverse range of research including conservation genetics, evolutionary genomics, functional adaptation, population dynamics and phylogenetics.

Conservation Assessments & Field Analysis

The link between the analysis of the genetics of a species and field assessment cannot be undervalued, without analysis of the species demographics and ecology, genetic data has significant limitations. Further conservation efforts and field monitoring of the Trinidad Piping Guan are ongoing in Trinidad, however the difficulties of field analysis as outlined in the previous chapters remain. In order for a comprehensive analysis of this species to move forward in the future, a greater understanding of all aspects of the Trinidad Piping Guan species biology should be developed, incorporating genetic analysis with biological, ecological, behavioural and environmental research. In the future demographic and population data should be collected which will enable Population Viability Analysis and Ecological Niche Modelling to be performed to help inform conservation decisions. The protection of this important island endemic avian may be better served through the collaboration of knowledge and understanding of the species as a whole, with genetic information to better predict the future responses of the species to a changing world.

Chapter 9. Conclusions

The aims of these studies were to establish and use genetic information from contemporary wild Trinidad Piping Guans in novel analysis of the species and its mainland counterparts of the genus *Pipile*. This study has for the first time examined the genetic diversity and evolutionary relationships of the Trinidad Piping Guan *Pipile pipile* and its conspecifics o. The analysis has incorporated the use of Trinidad Piping Guan specimens collected from the wild, representing approximately 4-11% of the contemporary wild population based on the most recent population size estimates (70-200 individuals) (Hayes, Sanasie, et al., 2009). In addition, reference specimens from *Pipile* species sourced from museum collections were used for the first time to analyse the relationships between the contemporary wild Trinidad Piping Guans and other species of the genus *Pipile* in relation to the biogeographic and evolutionary origins of the species.

This research has successfully produced reference sequences for the whole mitochondrial genomes of three individuals which represent the Trinidad Piping Guan species in its contemporary wild form, which have been published on the NCBI DNA databases. These mitochondrial genomes have the potential to inform many areas of further research into the Galliforme order, but also have the potential to be used in the protection of the species through the development of an understanding of genetic diversity within the species and the implementation of genetic analyses in conservation strategies. Further to the reference genomes demonstrated in this research numerous other key sources of genetic information have been generated for the genus *Pipile* and the Trinidad Piping Guan, including nuclear gene sequences, COI barcodes and ISSR fingerprint profiles. These sources, will be published on the NCBI nucleotide databases, and will therefore be available for the implementation of numerous genetic analyses in the future including species diagnostic markers, forensic references, individualising markers for parentage or kinship analysis and the identification of individuals within the population.

The analysis of genetic variation within the wild Trinidad Piping Guan population was an important component of this study, as it is an important measure of the currentstatus and health of the population. The Cytochrome *b*, Cytochrome c oxidase 1, NADH dehydrogenase 2 genes and Control Region of the mitochondrial genome, and the Beta-fibrinogen intron 7 and Serpin B14 intron c gene regions were analysed to examine the current genetic status observed in the wild Trinidad Piping Guan population in the Northern Range of Trinidad and the captive Piping Guan population in the UK. This analysis demonstrated low levels of variation in both sample collections, indicative of small population size and founder effects observed by both the wild and captive populations. The limited genetic diversity in the wild Trinidad Piping Guan may therefore represent the genetic legacy of historical bottlenecks and founder events as a consequence of biogeographic and climatic shifts, in addition to the current ongoing decline in population size within the species. The analysis was critically limited by the number of available samples for this species and future efforts should focus on the broadening of the analysis to include more samples from the contemporary wild population. This study has however benefitted future efforts in this area, by establishing markers and techniques which are successful in demonstrating genetic variation within this species, thus making future analyses more cost effective and easier. Establishing effective molecular methods for species such as the Trinidad Piping Guan are important for future conservation efforts, particularly in such species which are of low priority internationally but are regionally important.

In this study the complete mitochondrial genome of the Trinidad Piping Guan was used for the first time in genome level phylogenetic reconstruction of the Galliforme evolutionary tree. This analysis placed the *Pipile* genus basal of the *Crax* species in the Cracidae family, with high levels of statistical support (bootstrap and posterior probabilities). The Cracidae family resolved in a basal position within the Galliforme evolutionary tree, sister to the Megapodidae family. The high degree of congruence of the phylogenetic hypotheses and high levels of statistical support for the topologies for both Maximum Likelihood and Bayesian Inference phylogenetic methods indicate consistency in the topology and reliability of the technique. The addition of new mito-genomes in the analysis has been shown to improve resolution of the phylogenetic hypotheses, and construct phylogenetic topologies which are highly concordant with

previous morphological and molecular research, and taxonomic expectations. The novel inclusion of the *Pipile* genus in a Galliforme species tree is highly beneficial in establishing the evolutionary position of the genomically under-represented Cracidae family. This study has successfully established the use of reference mitochondrial genomes for the Trinidad Piping Guan, providing a contemporary reference for both the species and the genus for future analyses and research into these important avians.

The use of museum collection specimens of the genus *Pipile* has for the first time enabled the construction of a mitochondrial DNA reference collection of vouchered species representatives, which have been used to examine the inter-specific relationships within the genus. The development of molecular techniques and primers for the analysis of historical specimens in this genus has the potential to enable significant development in the analysis of relationships within the genus *Pipile* and establish molecular taxonomic identification of these important species. This analysis has highlighted specific taxonomic indicative markers of the Trinidad Piping Guan, which establish the species for the first time as an evolutionary significant unit, genetically and biogeographically isolated from its mainland conspecifics. In this respect establishing the Trinidad form of the Piping Guans as a genetically identifiable species for the first time may have significant consequences to the management of both wild and captive populations of this species, and emphasises the need to maintain the genetic integrity of this evolutionarily distinct species in its contemporary form.

This study has been highly successful in addressing the study aims, however it has left several areas of research open for future analysis and has raised numerous questions regarding the genetic status of the Trinidad Piping Guan and phylogenetic relationships within the genus *Pipile*. Therefore, it is recommended that the analysis of this species continues and efforts are made to clarify many of these problems. Publication of the data generated in this anlysis will have a significnant impact on the future resreach into the Cracidae family and the genus *Pipile*, as well as raising the profile of this little known species which is of significant impotance to the island of Trinidad.

References

Abensperg-Traun, M. (2009). CITES, sustainable use of wild species and incentive-driven conservation in developing countries, with an emphasis on southern Africa. *Biological Conservation*, *142*(5), 948–963. https://doi.org/10.1016/j.biocon.2008.12.034

Alemu, J., Davis, E., Keane, A., Mahabir, K., Mohais, S., Naranjit, A., … Stoker, B. (2005). *Project Pawi: Recovery of the Trinidad Piping-Guan (Pipile pipile)*. *The BP Conservation Programme*.

Aliabadian, M., Kaboli, M., Nijman, V., & Vences, M. (2009). Molecular identification of birds: Performance of distance-based DNA barcoding in three genes to delimit parapatric species. *PLoS ONE*, *4*(1). https://doi.org/10.1371/journal.pone.0004119

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

An, J., Lee, M.-Y., Min, M.-S., Lee, M.-H., & Lee, H. (2007). A molecular genetic approach for species identification of mammals and sex determination of birds in a forensic case of poaching from South Korea. *Forensic Science International*, *167*(1), 59–61. https://doi.org/10.1016/j.forsciint.2005.12.031

Angleby, H., & Savolainen, P. (2005). Forensic informativity of domestic dog mtDNA control region sequences. *Forensic Science International*, *154*(2–3), 99–110. https://doi.org/10.1016/j.forsciint.2004.09.132

Armsrtong, M., Braun, E., & Kimball, R. T. (2001). Phylogenetic utility of avian Ovomucoid Intron G: A comparison of nuclear and mitochondrial phylogenies in Galliformes. *The Auk*, *118*(3), 799–804.

Avise, J. C., Arnold, J., Ball, R., Bermingham, E., Lamb, T., Neigel, J., … Saunders, N. (1987). Intraspecific Phylogeography: The mitochondrial DNA bridge bewteen population genetics and systematics. *Annual Review of Ecology and Systematics*, *18*, 489–522.

Avise, J. C., & Walker, D. (1998). Pleistocene phylogeographic effects on avian populations and the speciation process. *Proceedings. Biological Sciences / The Royal Society*, *265*(1395), 457–463. https://doi.org/10.1098/rspb.1998.0317

Avise, J. C., Walker, D., & Johns, G. C. (1998). Speciation durations and Pleistocene effects on vertebrate phylogeography. *Proceedings. Biological Sciences / The Royal Society*, *265*(1407), 1707–1712. https://doi.org/10.1098/rspb.1998.0492

Backström, N., Fagerberg, S., & Ellegren, H. (2008). Genomics of natural bird populations: a gene-based set of reference markers evenly spread across the avian genome. *Molecular Ecology*, *17*(4), 964–80. https://doi.org/10.1111/j.1365-294X.2007.03551.x

Ballard, J. W., & Dean, M. D. (2001). The mitochondrial genome: mutation, selection and recombination. *Current Opinion in Genetics & Development*, *11*(6), 667–72. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11682311

Bantock, T. M., Prys-Jones, R. P., & Lee, P. L. M. (2008). New and improved molecular sexing methods for museum bird specimens. *Molecular Ecology*

Resources, *8*(3), 519–528. https://doi.org/10.1111/j.1471-8286.2007.01999.x

Baratti, M., Alberti, A., Groenen, M., Veenendaal, T., & Fulgheri, F. D. (2001). Polymorphic microsatellites developed by cross-species amplifications in common pheasant breeds. *Animal Genetics*, *32*, 222–225.

Barbará, T., Palma-Silva, C., Paggi, G. M., Bered, F., Fay, M. F., & Lexer, C. (2007). Cross-species transfer of nuclear microsatellite markers: potential and limitations. *Molecular Ecology*, *16*(18), 3759–67. https://doi.org/10.1111/j.1365-294X.2007.03439.x

Barker, F. K., Cibois, A., Schikler, P., Feinstein, J., & Cracraft, J. (2004). Phylogeny and diversification of the largest avian radiation. *PNAS*, *101*(30), 11040–11045.

Barraclough, T. G., & Nee, S. (2001). Phylogenetics and speciation. *Trends in Ecology & Evolution*, *16*(7), 391–399. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/22897956

Bazin, E., Glémin, S., & Galtier, N. (2006). Population size does not influence mitochondrial genetic diversity in animals. *Science (New York, N.Y.)*, *312*(5773), 570–2. https://doi.org/10.1126/science.1122033

Begazo, A. J., & Bodmer, R. E. (1998). Use and conservation of Cracidae (Aves : Galliformes) in the Peruvian Amazon. *Oryx*, *32*(4), 301–309.

Beissinger, S. R., Wunderle, J. M., Meyers, J. M., & Engen, S. (2008). Anatomy of a Bottleneck : Diagnosing Factors Limiting Population Growth in the Puerto Rican Parrot. *Ecological Monographs, ESA*, *78*(2), 185–203.

Bellard, C., Bertelsmeier, C., Leadley, P., Thuiller, W., & Courchamp, F. (2012). Impacts of climate change on the future of biodiversity. *Ecology Letters*, *15*(4), 365–377. https://doi.org/10.1111/j.1461-0248.2011.01736.x

Bellis, C., Ashton, K. J., Freney, L., Blair, B., & Griffiths, L. R. (2003). A molecular genetic approach for forensic animal species identification. *Forensic Science International*, *134*(2–3), 99–108. https://doi.org/10.1016/S0379-0738(03)00128-2

Bensasson, D., Zhang, D. X., Hartl, D. L., & Hewitt, G. M. (2001). Mitochondrial pseudogenes: Evolution's misplaced witnesses. *Trends in Ecology and Evolution*, *16*(6), 314–321. https://doi.org/10.1016/S0169-5347(01)02151-6

Biber, E. (2002). Patterns of endemic extinctions among island bird species. *Ecography*, *25*(March), 661–676.

BirdLife International. (2015). BirdLife International Species Information Pipile pipile. Retrieved October 10, 2015, from http://www.birdlife.org/datazone/speciesfactsheet.php?id=87

Blackburn, T. M., Petchey, O. L., Cassey, P., & Gaston, K. J. (2005). Functional diversity of mammalian predators and extinction in island birds. *Ecology*, *86*(11), 2916–2923. https://doi.org/10.1890/04-1847

Blake, J. G., & Loiselle, B. A. (2000). Diversity of Birds Along an Elevational Gradient in the Cordillera Central, Costa Rica. *The Auk*, *117*(3), 663–686. https://doi.org/10.1642/0004-8038(2000)117[0663:DOBAAE]2.0.CO;2

Bollmer, J. L., Kimball, R. T., Whiteman, N. K., Sarasola, J. H., & Parker, P. G. (2006). Phylogeography of the Galapagos hawk (Buteo galapagoensis): A recent arrival to the Galapagos Islands. *Molecular Phylogenetics and Evolution*, *39*(1),

237–247. https://doi.org/10.1016/j.ympev.2005.11.014

Borges, S. H. (1999). RELATIVE USE OF SECONDARY FORESTS BY CRACIDS IN CENTRAL AMAZONIA. *Ornitologia Neotropical*, *10*, 77–80.

Bouckaert, R., & Heled, J. (2014). DensiTree 2: Seeing Trees Through the Forest. *bioRxiv*. https://doi.org/http://dx.doi.org/10.1101/012401

Bouzat, J. L. (2010). Conservation genetics of population bottlenecks: the role of chance, selection, and history. *Conservation Genetics*, *11*(2), 463–478. https://doi.org/10.1007/s10592-010-0049-0

Bradshaw, C. J. a, Isagi, Y., Kaneko, S., Brook, B. W., Bowman, D. M. J. S., & Frankham, R. (2007). Low genetic diversity in the bottlenecked population of endangered non-native banteng in northern Australia. *Molecular Ecology*, *16*(14), 2998–3008. https://doi.org/10.1111/j.1365-294X.2007.03365.x

Branicki, W., Kupiec, T., & Pawlowski, R. (2003). Validation of cytochrome b sequence analysis as a method of species identification. *Journal of Forensic Sciences*, *48*(1), 83–7. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12570204

Braun, E. L., & Kimball, R. T. (2002). Examining Basal avian divergences with mitochondrial sequences: model complexity, taxon sampling, and sequence length. *Systematic Biology*, *51*(4), 614–625. https://doi.org/10.1080/10635150290102294

Brooks, D. M. (1999). Pipile as a protein source to rural hunters and Amerindians. In D. M. Brooks, A. J. Begazo, & F. Olmos (Eds.), *Biology and Conservation of the Piping Guans (Pipile)* (pp. 42–50). Cracid Specialist Group.

Brooks, D. M. (2006). *Conserving cracids: the most threatened family of birds in the Americas*. *Miscelaneous Publications of Houston Museum of Natural Sciences (6)*. Retrieved from http://www.cracids.org/AP_Engl_ebook.pdf

Brooks, D. M., Pando-Vasquez, L., & Ocmin-Petit, A. (2005). The relationship between environmental stability and avian population changes in Amazonia. *Ornothologia Neotropical*, *16*(3), 289–296.

Brooks, D. M., & Strahl, S. D. (2000). *Curassows, Guans and Chachalacas: Status survey and conservation action plan for Cracids 2000-2004*. Cambridge: IUCN / SSC Cracid Specialist Group.

Brown, W. M., George, M., & Wilson, A. C. (1979). Rapid evolution of animal mitochondrial DNA. *Proc. Natl. Acad. Sci. USA*, *76*(4), 1967–1971.

Bush, K. L., Dyte, C. K., Moynahan, B. J., Aldridge, C. L., Sauls, H. S., Battazzo, A. M., … Coltman, D. W. (2010). Population structure and genetic diversity of greater sage-grouse (Centrocercus urophasianus) in fragmented landscapes at the northern edge of their range. *Conservation Genetics*, *12*(2), 527–542. https://doi.org/10.1007/s10592-010-0159-8

Bush, K. L., Vinsky, M. D., Aldridge, C. L., & Paszkowski, C. a. (2005). A comparison of sample types varying in invasiveness for use in DNA sex determination in an endangered population of greater Sage-Grouse (Centrocercus uropihasianus). *Conservation Genetics*, *6*(5), 867–870. https://doi.org/10.1007/s10592-005-9040-6

Butchart, S. H. M., Walpole, M., Collen, B., van Strien, a., Scharlemann, J. P. W.,

Almond, R. E. a., … Watson, R. (2010). Global Biodiversity: Indicators of Recent Declines. *Science*, *328*(5982), 1164–1168. https://doi.org/10.1126/science.1187512

Butlin, R. K., Galindo, J., & Grahame, J. W. (2008). Review. Sympatric, parapatric or allopatric: the most important way to classify speciation? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *363*(1506), 2997–3007. https://doi.org/10.1098/rstb.2008.0076

Carcraft, J. (1985). Historical Biogeography and Patterns of Differentiation within the South American Avifauna: Areas of Endemism. *Ornithological Monographs*, *36*, 49–84.

Carcraft, J. (2001). Avian evolution, Gondwana biogeography and the Cretaceous–Tertiary mass extinction event. *Proc. R. Soc. Lond. B*, *268*, 459–469. https://doi.org/10.1098/rspb.2000.1368

Cavero, T., & Pratolongo, F. (2011). Health of the Critically Endangered white-winged guan Penelope albipennis and implications for its reintroduction and conservation in Peru. *International Journal of Galliformes Conservation*, (2), 43–53.

Chan, Y. L., Anderson, C. N. K., & Hadly, E. A. (2006). Bayesian Estimation of the Timing and Severity of a Population Bottleneck from Ancient DNA. *PLoS Genetics*, *2*(4), e59. https://doi.org/10.1371/journal.pgen.0020059

Chang, H.-W., Cheng, C.-A., Gu, D.-L., Chang, C.-C., Su, S.-H., Wen, C.-H., … Cheng, C.-C. (2008). High-throughput avian molecular sexing by SYBR green-based real-time PCR combined with melting curve analysis. *BMC Biotechnology*, *8*, 12. https://doi.org/10.1186/1472-6750-8-12

Chen, J. Z., Yauk, C. L., Hebert, C., & Hebert, P. D. N. (2001). Genetic Variation in Mitochondrial DNA of North American Herring Gulls, Larus argentatus. *Journal of Great Lakes Research*, *27*(2), 199–209. https://doi.org/10.1016/S0380-1330(01)70633-8

Chojnowski, J. L., Kimball, R. T., & Braun, E. L. (2008). Introns outperform exons in analyses of basal avian phylogeny using clathrin heavy chain genes. *Gene*, *410*(1), 89–96. https://doi.org/10.1016/j.gene.2007.11.016

Chubb, A. L. (2004). New nuclear evidence for the oldest divergence among neognath birds: The phylogenetic utility of ZENK (i). *Molecular Phylogenetics and Evolution*, *30*(1), 140–151. https://doi.org/10.1016/S1055-7903(03)00159-3

CITES. (2017). CITES Appendices - Species Database. Retrieved January 2, 2017, from https://cites.org/eng/app/appendices.php

Coghlan, M. L., White, N. E., Parkinson, L., Haile, J., Spencer, P. B. S., & Bunce, M. (2012). Egg forensics: an appraisal of DNA sequencing to assist in species identification of illegally smuggled eggs. *Forensic Science International. Genetics*, *6*(2), 268–73. https://doi.org/10.1016/j.fsigen.2011.06.006

Cooper, A., Lalueza-Fox, C., Anderson, S., Rambaut, A., Austin, J., & Ward, R. (2001). Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. *Nature*, *409*, 704–707. https://doi.org/10.1038/35055536

Cooper, A., & Penny, D. (1997). Mass survival of birds across the Cretaceous-Tertiary boundary: molecular evidence. *Science (New York, N.Y.)*, *275*, 1109–1113. https://doi.org/10.1126/science.275.5303.1109

Crowe, T. M., Bowie, R. C. K., Bloomer, P., Mandiwana, T. G., Hedderson, T. A. J., Randi, E., … Wakeling, J. (2006). Phylogenetics, biogeography and classification of, and character evolution in, gamebirds (Aves: Galliformes): effects of character exclusion , data partitioning and missing data. *Cladistics*, *22*, 495–532.

da Silva, M. N., & Patton, J. L. (1998). Molecular phylogeography and the evolution and conservation of Amazonian mammals. *Molecular Ecology*, *7*(4), 475–86. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9628001

Darriba, D., Taboada, G., Doallo, R., & Posada, D. (2012). jModelTest2: more models, new heuristics and parallel computing. *Nature Methods*, *9*(8), 772.

Dawnay, N., Ogden, R., McEwing, R., Carvalho, G. R., & Thorpe, R. S. (2007). Validation of the barcoding gene COI for use in forensic genetic species identification. *Forensic Science International*, *173*(1), 1–6. https://doi.org/10.1016/j.forsciint.2006.09.013

Dawson, D. a, Darby, S., Hunter, F., Krupa, A. P., Jones, I., & Burke, T. (2001). A critique of avian CHD -based molecular sexing protocols illustrated by a Z-chromosome polymorphism detected in auklets. *Molecular Ecology Notes*, *1*, 201–204.

Delacour, J., & Amadon, D. (1973). *Curassows and Related Birds* (First). New York: The American Museum of Natural History.

del Hoyo, J., Ellliot, A., & Sargatal, J. (1994). *Handbook of the Birds of the World: Volume 2, New World Vultures to Guineafowl*. (J. del Hoyo, A. Elliott, & J. Sargatal, Eds.). Barcelona: Lynx Edicions.

del Hoyo, J., Motis, A., Delacour, J., & Amadon, D. (2004). *Curassows and Related Birds* (Second). Barcelona and New York: Lynx Edicions and The National Museum of Natural History.

Desjardins, P., & Morais, R. (1990). Sequence and gene organisation of the chicken mitochondrial genome: a novel gene order in higher vertebrates. *Journal of Molecular Biology*, *212*, 599–634.

Di Fiore, A., Chaves, P. B., Cornejo, F. M., Schmitt, C. a., Shanee, S., Cortes-Ortiz, L., … Pacheco, V. (2014). The rise and fall of a genus: Complete mtDNA genomes shed light on the phylogenetic position of yellow-tailed woolly monkeys, Lagothrix flavicauda, and on the evolutionary history of the family Atelidae (Primates: Platyrrhini). *Molecular Phylogenetics and Evolution*, *82*, 495–510. https://doi.org/10.1016/j.ympev.2014.03.028

Dieffenbach, C. W., Lowe, T. M., & Dveksler, G. S. (1993). General concepts for PCR primer design. *Genome Research*, *3*(3), S30--37. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8118394%5Cnhttp://img.dxycdn.com/tradem d/upload/userfiles/file/2011/11/1322031004.pdf

Dimcheff, D. E., Drovetski, S. V, & Mindell, D. P. (2002). Phylogeny of Tetraoninae and other galliform birds using mitochondrial 12S and ND2 genes. *Molecular Phylogenetics and Evolution*, *24*(2), 203–15. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12144757

Dohms, K. M., Graham, B. A., & Burg, T. M. (2017). Multilocus genetic analyses and spatial modeling reveal complex population structure and history in a widespread resident North American passerine (Perisoreus canadensis). *Ecology and Evolution*, (April), 9869–9889. https://doi.org/10.1002/ece3.3478

Don, R. H., Cox, P. T., Wainwright, B. J., Baker, K., & Mattick, J. S. (1991). "Touchdown" PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Research*, *19*(14), 4008. https://doi.org/10.1093/nar/19.14.4008

Drummond, A., & Bouckaert, R. (2015). *Bayesian Evolutionary Analysis with BEAST* (Second). Cambridge University Press.

Drummond, A. J., Suchard, M. A., Xie, D., & Rambaut, A. (2012). Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, *29*(8), 1969–1973. https://doi.org/10.1093/molbev/mss075

Duraes, R., Carrasco, L., Smith, T. B., & Karubian, J. (2013). Effects of forest disturbance and habitat loss on avian communities in a Neotropical biodiversity hotspot. *Biological Conservation*, *166*, 203–211. https://doi.org/10.1016/j.biocon.2013.07.007

Dyke, G. J., Gulas, B. E., & Crowe, T. M. (2003). Suprageneric relationships of galliform birds (Aves , Galliformes): a cladistic analysis of morphological characters. *Zoological Journal of the Linnean Society*, *137*, 227–244.

Eastman, C. (1900). New fossil bird and fish remains from the middle Eocene of Wyoming. *Geol. Mag., New Series*, *7*(54–58).

eBird. (2012). eBird: An online database of bird distribution and abundance [web application]. Retrieved March 23, 2016, from http://ebird.org

Eckert, K. A., & Kunkel, T. A. (1990). High fidelity DNA synthesis by the Thermus aquaticus DNA polymerase. *Nucleic Acids Research*, *18*(13), 3739–3744. https://doi.org/10.1093/nar/18.13.3739

Egloff, C., Labrosse, a, Hebert, C., & Crump, D. (2009). A nondestructive method for obtaining maternal DNA from avian eggshells and its application to embryonic viability determination in herring gulls (Larus argentatus). *Molecular Ecology Resources*, *9*(1), 19–27. https://doi.org/10.1111/j.1755-0998.2008.02214.x

Elias, M., Hill, R. I., Willmott, K. R., Dasmahapatra, K. K., Brower, A. V. ., Mallet, J., & Jiggins, C. D. (2007). Limited performance of DNA barcoding in a diverse community of tropical butterflies. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1627), 2881–2889. https://doi.org/10.1098/rspb.2007.1035

Ellegren, H. (2004). Microsatellites: simple sequences with complex evolution. *Nature Reviews. Genetics*, *5*(6), 435–45. https://doi.org/10.1038/nrg1348

Emerson, B. C. (2002). Evolution on oceanic islands : molecular phylogenetic approaches to understanding pattern and process. *Molecular Ecology*, *11*, 951–966.

England, P. R., Osler, G. H. R., Woodworth, L. M., Montgomery, M. E., Briscoe, A., & Frankham, R. (2003). Effects of intense versus diffuse population bottlenecks on microsatellite genetic diversity and evolutionary potential. *Conservation Genetics*, *4*, 595–604.

Eo, S. H., Bininda-Emonds, O. R. P., & Carroll, J. P. (2009). A phylogenetic supertree of the fowls (Galloanserae, Aves). *Zoologica Scripta*, *38*(5), 465–481. https://doi.org/10.1111/j.1463-6409.2008.00382.x

Fain, M. G., & Houde, P. (2004). Parallel radiations in the primary clades of birds. *Evolution; International Journal of Organic Evolution*, *58*(11), 2558–2573. https://doi.org/10.1554/04-235

Fauvelot, C., Bernardi, G., & Planes, S. (2003). Reductions in the mitochondrial DNA diversity of coral reef fish provide evidence of population bottlenecks resulting from Holocene sea-level change. *Evolution; International Journal of Organic Evolution*, *57*(7), 1571–83. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12940362

Fay, J. C., & Wu, C. (1996). Letter to the Editor A Human Population Bottleneck Can Account for the Discordance Between Patterns of Mitochondrial Versus Nuclear DNA Variation. *Molecular Biology and Evolution*, *16*(7), 1003–1005.

Felsenstein, J. (1985). Use of Bootstrap in phylogenetic calculations. *Evolution*, *39*(4), 783–791.

Felsenstein, J. (2004). *Inferring Phylogenies*. Sinauer Associates Inc., Massachusetts.

Ferrero, M. E., González-Jara, P., Blanco-Aguiar, J. a., Sánchez-Barbudo, I., & Dávila, J. a. (2007). Sixteen new polymorphic microsatellite markers isolated for red-legged partridge (Alectoris rufa) and related species. *Molecular Ecology Notes*, *7*(6), 1349–1351. https://doi.org/10.1111/j.1471-8286.2007.01877.x

Frankham, R. (1997). Do island populations have less genetic variation than mainland populations? *Heredity*, *78 (Pt 3)*, 311–27. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9119706

Frankham, R. (2005). Genetics and extinction. *Biological Conservation*, *126*(2), 131–140. https://doi.org/10.1016/j.biocon.2005.05.002

Frankham, R. (2008). Inbreeding and Extinction: Island Populations. *Conservation Biology*, *12*(3), 665–675. https://doi.org/10.1111/j.1523-1739.1998.96456.x

Frankham, R., Ballou, J. D., Dudash, M. R., Eldridge, M. D. B., Fenster, C. B., Lacy, R. C., … Ryder, O. a. (2012). Implications of different species concepts for conserving biodiversity. *Biological Conservation*, *153*, 25–31. https://doi.org/10.1016/j.biocon.2012.04.034

Franklin, I., & Frankham, R. (1998). How large must populations be to retain evolutionary potential? *Animal Conservation*, *1*, 69–73.

Fridolfsson, A.-K., & Ellegren, H. (1999). A simple and universal method for molecular sexing of non- ratite birds. *Journal of Avian Biology*, *30*(1), 116–121.

Fridolfsson, a K., & Ellegren, H. (2000). Molecular evolution of the avian CHD1 genes on the Z and W sex chromosomes. *Genetics*, *155*(4), 1903–12. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1461215&tool=pmcentrez&rendertype=abstract

Funk, D. J., & Omland, K. E. (2014). Species-Level Paraphyly and Polyphyly: Species-level Causes , and Consequences , with Frequency , Insights from Animal Mitochondrial DNA. *Annual Review of Ecology, Evolution and Systematics*, *34*(2003), 397–423. https://doi.org/10.1146/132421

Funk, V. A., Hoch, P. C., Prather, L. A., & Wagner, W. L. (2005). The importance of vouchers. *Taxon*, *54*(1), 127–129.

Galbusera, P., van Dongen, S., & Matthysen, E. (2000). Cross-species amplification of microsatellite primers in passerine birds. *Conservation Genetics*, *1*, 163–168.

Galetti, M., Martuscelli, P., Olmos, F., & Aleixo, A. (1997). Ecology and conservation of the jacutinga Pipile jacutinga in the Atlantic forest of Brazil. *Biological*

Conservation, *82*(1), 31–39. https://doi.org/10.1016/S0006-3207(97)00004-9

Galtier, N., Depaulis, F., & Barton, N. H. (2000). Detecting bottlenecks and selective sweeps from DNA sequence polymorphism. *Genetics*, *155*(2), 981–7. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1461106&tool=pmcentrez&rendertype=abstract

García-Moreno, J., Sorenson, M. D., & Mindell, D. P. (2003). Congruent avian phylogenies inferred from mitochondrial and nuclear DNA sequences. *Journal of Molecular Evolution*, *57*(1), 27–37. https://doi.org/10.1007/s00239-002-2443-9

Gebhardt, K. J., Brightsmith, D., Powell, G., & Waits, L. P. (2009). Molted feathers from clay licks in Peru provide DNA for three large macaws (Ara ararauna , A. chloropterus , and A. macao). *Journal of Field Ornithology*, *80*(2), 183–192. https://doi.org/10.1111/j.1557-9263.2009.00221.x

Gibb, G. C., Kardailsky, O., Kimball, R. T., Braun, E. L., & Penny, D. (2007). Mitochondrial genomes and avian phylogeny: Complex characters and resolvability without explosive radiations. *Molecular Biology and Evolution*, *24*(1), 269–280. https://doi.org/10.1093/molbev/msl158

Gilbert, M. T. P., Bandelt, H. J., Hofreiter, M., & Barnes, I. (2005). Assessing ancient DNA studies. *Trends in Ecology and Evolution*, *20*(10), 541–544. https://doi.org/10.1016/j.tree.2005.07.005

Glenn, T. C., Stephan, W., & Braun, M. J. (1999). Effects of a Population Bottleneck on Whooping Crane Mitochondrial DNA Variation. *Conservation Biology*, *13*(5), 1097–1107.

Gonzalez, J. (1999). Hunting of Cracids in Varzea Forests of the Northeastern Peruvian Amazon: A Sustainability Analysis. *Bulletin of the IUCN/BirdLife/WPA Cracid Specialist Group*, *9*, 12–15.

Government of the Republic of Trinidad and Tobago. (2010a). *FOURTH NATIONAL REPORT OF TRINIDAD AND TOBAGO TO THE CONVENTION ON BIOLOGICAL*.

Government of the Republic of Trinidad and Tobago. (2010b). Proceedings of the 1 st Research Symposium on Biodiversity in Trinidad and Tobago. In A. Lawrence & H. Nelson (Eds.). St. Augustine, Trinidad: University of West indies and Government of the Republic of Trinidad and Tobago.

Graham, C. H., Ferrier, S., Huettman, F., Moritz, C., & Peterson, a T. (2004). New developments in museum-based informatics and applications in biodiversity analysis. *Trends in Ecology & Evolution*, *19*(9), 497–503. https://doi.org/10.1016/j.tree.2004.07.006

Grass, A., Hosie, C., & McDowall, I. (2016). The complete mitochondrial genome and phylogenetic position of the critically endangered Trinidad Piping Guan, Pipile pipile synonym Aburria pipile (Aves: Galliformes). *Mitochondrial DNA Part B*. 1(1), 649-650.

Grau, E. T., Pereira, S. L., Silveira, L. F., Höfling, E., & Wajntal, A. (2005). Molecular phylogenetics and biogeography of Neotropical piping guans (Aves: Galliformes): Pipile Bonaparte, 1856 is synonym of Aburria Reichenbach, 1853. *Molecular Phylogenetics and Evolution*, *35*(3), 637–45. https://doi.org/10.1016/j.ympev.2004.12.004

Grau, E. T., Pereira, S. L., Silveira, L. F., & Wajntal, A. (2003). Molecular markers contribute to a breeding programme of the extinct-in-the-wild Alagoas Curassow Mitu mitu and confirm the validity of the species. *Bird Conservation International*, *13*(2), 115–126. https://doi.org/10.1017/S0959270903003101

Griffiths, R., Double, M., & Dawson, R. (1998). A DNA test to sex most birds. *Molecular Ecology*, *7*, 1071–1075.

Guan, X., Silva, P., Gyenai, K. B., Xu, J., Geng, T., Tu, Z., … Smith, E. J. (2009). The mitochondrial genome sequence and molecular phylogeny of the turkey, Meleagris gallopavo. *Animal Genetics*, *40*(2), 134–141. https://doi.org/10.1111/j.1365-2052.2008.01810.x

Habel, J. C., & Zachos, F. E. (2013). Past population history versus recent population decline - founder effects in island species and their genetic signatures. *Journal of Biogeography*, *40*(1), 206–207. https://doi.org/10.1111/jbi.12017

Hackett, S. J., Kimballl, R. T., Reddy, S., Bowie, R. C. K., Braun, E. L., Braun, M. J., … Yuri, T. (2008). A Phylogenomic Study of Birds Reveals Their Evolutionary History. *Science*, *320*, 1763–1768. https://doi.org/10.1126/science.1157704

Haddrath, O., & Baker, A. J. (2001). Complete mitochondrial DNA genome sequences of extinct birds: ratite phylogenetics and the vicariance biogeography hypothesis. *Proceedings. Biological Sciences / The Royal Society*, *268*(January), 939–945. https://doi.org/10.1098/rspb.2001.1587

Haffer, J. (1985). Avian Zoogeography of the Neotropical Lowlands. *Ornithological Monographs*, *36*, 113–146.

Haig, S., Beever, E., Chambers, S., Draheim, H., Dugger, B., Dunham, S., … Sheffield, L. (2006). REVIEW : TAXONOMIC CONSIDERATIONS IN LISTING SUBSPECIES UNDER THE U . S . ENDANGERED SPECIES ACT. *Conservation Biology*, *20*(6), 1584–1594.

Hajibabaei, M., Singer, G. a C., Hebert, P. D. N., & Hickey, D. a. (2007). DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *Trends in Genetics : TIG*, *23*(4), 167–72. https://doi.org/10.1016/j.tig.2007.02.001

Harlid, A., & Arnason, U. (1999). Analyses of mitochondrial DNA nest ratite birds within the Neognathae: supporting a neotenous origin of ratite morphological characters. *Proceedings of the Royal Society B: Biological Sciences*, *266*(November 1998), 305. https://doi.org/10.1098/rspb.1999.0638

Harlid, A., Janke, A., & Arnason, U. (1998). The complete mitochondrial genome of Rhea americana and early avian divergences. *Journal of Molecular Evolution*, *46*(6), 669–679.

Härlid, A., Janke, A., & Arnason, U. (1997). The mtDNA sequence of the ostrich and the divergence between paleognathous and neognathous birds. *Molecular Biology and Evolution*, *14*(7), 754–761.

Harrison, G. L., McLenachan, P. A., Phillips, M. J., Slack, K. E., Cooper, A., & Penny, D. (2004). Four new avian mitochondrial genomes help get to basic evolutionary questions in the late cretaceous. *Molecular Biology and Evolution*, *21*(6), 974–983. https://doi.org/10.1093/molbev/msh065

Harvey, M. G., Bonter, D. N., Stenzler, L. M., & Lovette, I. J. (2006). A comparison of plucked feathers versus blood samples as DNA sources for molecular sexing.

Journal of Field Ornithology, *77*(2), 136–140. https://doi.org/10.1111/j.1557-9263.2006.00033.x

Hayes, F., Sanasie, B., & Samad, I. (2009). Status and conservation of the critically endangered Trinidad piping-guan Aburria pipile. *Endangered Species Research*, *7*(May), 77–84. https://doi.org/10.3354/esr00184

Hayes, F., Shameerudeen, C., Sanasie, B., Hayes, B., Ramjohn, C., & Lucas, F. (2009). Ecology and behaviour of the critically endangered Trinidad piping-guan Aburria pipile. *Endangered Species Research*, *6*(March), 223–229. https://doi.org/10.3354/esr00153

He, L., Dai, B., Zeng, B., Zhang, X., Chen, B., Yue, B., & Li, J. (2009). The complete mitochondrial genome of the Sichuan Hill Partridge (Arborophila rufipectus) and a phylogenetic analysis with related species. *Gene*, *435*(1–2), 23–28. https://doi.org/10.1016/j.gene.2009.01.001

Hebert, P. D. N., Stoeckle, M. Y., Zemlak, T. S., & Francis, C. M. (2004). Identification of Birds through DNA Barcodes. *PLoS Biology*, *2*(10), e312. https://doi.org/10.1371/journal.pbio.0020312

Hecker, K. H., & Roux, H. (1996). High and Low Annealing Temperatures Increase Both Specificity and Yield in Touchdown and Stepdown peR. *BioTechniques*, *20*(3), 478–485.

Helbig, A.J;Seibold, I;Bednarek, W;Gaucher, P;Ristow, D;Scharlau, W;Schmidl, D;Wink, M. (1994). Phylogenetic relationship among Falcon species (genus Falco) according to DNA sequence variation og the cytochrome b gene. *Raptor Conservation Today Proceedings of the lV World Conference on Birds of Prey and Owls*. Retrieved from http://www.uni-heidelberg.de/institute/fak14/ipmb/phazb/pubwink/1994/4. 1994.pdf

Helbig, A. J., Knox, A. G., Parkin, D. T., Sangster, G., & Collinson, M. A. R. T. I. N. (2002). Guidelines for assigning species rank †. *Ibis*, *144*, 518–525.

Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. *Nature*, *405*, 907–913. https://doi.org/10.1038/35016000

Higuchi, R., Bowman, B., Freiberger, M., Ryder, O., & Wilson, A. (1984). DNA sequences from the quagga, an extinct member of the horse family. *Nature*, *312*, 282–284.

Hoelzel, a. R., Fleischer, R. C., Campagna, C., Le Boeuf, B. J., & Alvord, G. (2002). Impact of a population bottleneck on symmetry and genetic diversity in the northern elephant seal. *Journal of Evolutionary Biology*, *15*(4), 567–575. https://doi.org/10.1046/j.1420-9101.2002.00419.x

Hogan, F. E., Cooke, R., Burridge, C. P., & Norman, J. a. (2008). Optimizing the use of shed feathers for genetic analysis. *Molecular Ecology Resources*, *8*(3), 561–7. https://doi.org/10.1111/j.1471-8286.2007.02044.x

Hosner, P. A., Braun, E. L., & Kimball, R. T. (2016). Rapid and recent diversification of curassows, guans, and chachalacas (Galliformes: Cracidae) out of Mesoamerica: Phylogeny inferred from mitochondrial, intron, and ultraconserved element sequences. *Molecular Phylogenetics and Evolution*, *102*, 320–330. https://doi.org/10.1016/j.ympev.2016.06.006

Höss, M., & Pääbo, S. (1993). DNA extraction from Pleistocene bones by a silica-based purification method. *Nucleic Acids Research*, *21*(16), 3913–3914.

https://doi.org/10.1093/nar/21.16.3913

Hsieh, H., Chiang, H., Tsai, L., Lai, S., Huang, N., Linacre, A., & Lee, J. C. (2001). Cytochrome b gene for species identi ® cation of the conservation animals. *Forensic Science International*, *122*, 7–18.

Huang, Z., Liu, N., Liang, W., Zhang, Y., Liao, X., Ruan, L., & Yang, Z. (2010). Phylogeography of Chinese bamboo partridge, Bambusicola thoracica thoracica (Aves: Galliformes) in south China: Inference from mitochondrial DNA control-region sequences. *Molecular Phylogenetics and Evolution*, *56*(1), 273–280. https://doi.org/10.1016/j.ympev.2010.01.028

Huang, Z., Liu, N., Xiao, Y., Cheng, Y., Mei, W., Wen, L., … Yu, X. (2009). Phylogenetic relationships of four endemic genera of the Phasianidae in China based on mitochondrial DNA control-region genes. *Molecular Phylogenetics and Evolution*, *53*(2), 378–83. https://doi.org/10.1016/j.ympev.2009.07.002

Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics (Oxford, England)*, *17*(8), 754–755. https://doi.org/10.1093/bioinformatics/17.8.754

Hughes, C. R., & Larson, E. D. (2000). Characterisation of microsatellite loci developed for the wattled curassow, Crax globulosa. *Molecular Ecology*, *9*, 632–633. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10792705

Ibarguchi, G. (1999). *Estimates of evolutionary divergence and phylogenetic relationships within Cracidae and among other basal birds*. Ottawa.

IUCN. (2017). The IUCN Red List of Threatened Species. Retrieved January 2, 2017, from http://www.iucnredlist.org/

Jackson, H. A., Bunbury, N., Przelomska, N., & Groombridge, J. J. (2016). Evolutionary distinctiveness and historical decline in genetic diversity in the Seychelles Black Parrot Coracopsis nigra barklyi. *Ibis*, *158*(2), 380–394. https://doi.org/10.1111/ibi.12343

Janke, A., & Arnason, U. (1997). The complete mitochondrial genome of Alligator mississippiensis and the separation between recent archosauria (birds and crocodiles). *Molecular Biology and Evolution*, *14*, 1266–1272.

Janke, A., Erpenbeck, D., Nilsson, M., & Arnason, U. (2001). The mitochondrial genomes of the iguana (Iguana iguana) and the caiman (Caiman crocodylus): implications for amniote phylogeny. *Proceedings. Biological Sciences / The Royal Society*, *268*(July 2000), 623–631. https://doi.org/10.1098/rspb.2000.1402

Jarvis, E. D., Mirarab, S., Aberer, A., Li, B., Houde, P., Li, C., … Zhang, G. (2014). Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science*, *346*(6215), 1126–1138.

Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K., & Mooers, A. O. (2012). The global diversity of birds in space and time. *Nature*, *491*, 1–5. https://doi.org/10.1038/nature11631

Jiang, L., Wang, G., Peng, R., Peng, Q., & Zou, F. (2014). Phylogenetic and molecular dating analysis of Taiwan Blue Pheasant (Lophura swinhoii). *Gene*, *539*(1), 21–29. https://doi.org/10.1016/j.gene.2014.01.067

Jimenez, I., Londono, G. A., & Cadena, C. D. (2003). Efficiency , bias , and consistency of visual and aural surveys of curassows (Cracidae) in tropical

forests. *Journal of Field Ornithology*, *74*(3), 210–216.

Johannson, M., McMahon, B., Hoglund, J., & Segelbacher, G. (2012). Amplification success of multilocus genotypes from feathers found in the field compared with feathers obtained from shot birds. *Ibis*, *154*, 15–20.

Johns, G. C., & Avise, J. C. (1998). A comparative summary of genetic distances in the vertebrates from the mitochondrial cytochrome b gene. *Molecular Biology and Evolution*, *15*(11), 1481–1490. https://doi.org/10.1093/oxfordjournals.molbev.a025875

Johnson, K. P. (2001). Taxon sampling and the phylogenetic position of Passeriformes: evidence from 916 avian cytochrome b sequences. *Systematic Biology*, *50*(1), 128–136. https://doi.org/10.1093/sysbio/50.1.128

Kaiserman, D., & Bird, P. I. (2005). Analysis of vertebrate genomes suggests a new model for clade B serpin evolution. *BMC Genomics*, *6*, 167. https://doi.org/10.1186/1471-2164-6-167

Kan, X.-Z., Yang, J.-K., Li, X.-F., Chen, L., Lei, Z.-P., Wang, M., … Yang, Z.-Y. (2010). Phylogeny of major lineages of galliform birds (Aves: Galliformes) based on complete mitochondrial genomes. *Genetics and Molecular Research : GMR*, *9*(3), 1625–33. https://doi.org/10.4238/vol9-3gmr898

Kan, X., Li, X., Lei, Z., Chen, L., Gao, H., Yang, Z., … Qian, C. (2010). Estimation of divergence times for major lineages of galliform birds : Evidence from complete mitochondrial genome sequences. *Journal of Biotechnology*, *9*(21), 3073–3078. https://doi.org/10.5897/AJB10.093

Kan, X. Z., Li, X. F., Lei, Z. P., Wang, M., Chen, L., Gao, H., & Yang, Z. Y. (2010). Complete mitochondrial genome of cabot's Tragopan, Tragopan caboti (Galliformes: Phasianidae). *Genetics and Molecular Research*, *9*(2), 1204–1216. https://doi.org/10.4238/vol9-2gmr820

Kattan, G. H., Muñoz, M. C., & Kikuchi, D. W. (2015). Population densities of curassows, guans, and chachalacas (Cracidae): Effects of body size, habitat, season, and hunting. *The Condor*, *118*(1), 24–32. https://doi.org/10.1650/CONDOR-15-51.1

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., … Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics (Oxford, England)*, *28*(12), 1647–1649.

Keller, L. F., Jeffery, K. J., Arcese, P., Beaumont, M. a, Hochachka, W. M., Smith, J. N., & Bruford, M. W. (2001). Immigration and the ephemerality of a natural population bottleneck: evidence from molecular markers. *Proceedings. Biological Sciences / The Royal Society*, *268*(1474), 1387–94. https://doi.org/10.1098/rspb.2001.1607

Kemp, B. M., Monroe, C., & Smith, D. G. (2006). Repeat silica extraction: a simple technique for the removal of PCR inhibitors from DNA extracts. *Journal of Archaeological Science*, *33*(12), 1680–1689. https://doi.org/10.1016/j.jas.2006.02.015

Kerr, K. C. R., Lijtmaer, D. A., Barreira, A. S., Hebert, P. D. N., & Tubaro, P. L. (2009). Probing evolutionary patterns in neotropical birds through DNA barcodes. *PLoS ONE*, *4*(2). https://doi.org/10.1371/journal.pone.0004379

Kimball, R. T., & Braun, E. L. (2008). A multigene phylogeny of Galliformes supports a single origin of erectile ability in non-feathered facial traits. *Journal of Avian Biology*, *39*, 438–445. https://doi.org/10.1111/j.2008.0908-8857.04270.x

Kimball, R. T., Braun, E. L., Barker, F. K., Bowie, R. C. K., Braun, M. J., Chojnowski, J. L., … Yuri, T. (2009). A well-tested set of primers to amplify regions spread across the avian genome. *Molecular Phylogenetics and Evolution*, *50*(3), 654–660. https://doi.org/10.1016/j.ympev.2008.11.018

Kimball, R. T., Braun, E. L., Ligon, J. D., Lucchini, V., & Randi, E. (2001). A molecular phylogeny of the peacock-pheasants (Galliformes: *Biological Journal of the Linnean Society*, *73*, 187–198.

Kimball, R. T., Braun, E. L., Zwartjes, P. W., Crowe, T. M., & Ligon, J. D. (1999). A molecular phylogeny of the pheasants and partridges suggests that these lineages are not monophyletic. *Molecular Phylogenetics and Evolution*, *11*(1), 38–54. https://doi.org/10.1006/mpev.1998.0562

Kocher, T. D., Thomas, W. K., Meyer, a, Edwards, S. V, Pääbo, S., Villablanca, F. X., & Wilson, a C. (1989). Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proceedings of the National Academy of Sciences of the United States of America*, *86*(16), 6196–200. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=297804&tool=pmcentrez&rendertype=abstract

Korbie, D., & Mattick, J. S. (2008). Touchdown PCR for Increased Specificity and Sensitivity in PCR Amplification. *Nature Protocols*, *3*(9), 1452–1456.

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.*

Kumazawa, Y., & Nishida, M. (1995). Variations in mitochondrial tRNA gene organization of reptiles as phylogenetic markers. *Molecular Biology and Evolution*, *12*, 759–772.

Kumazawa, Y., & Nishida, M. (1999). Complete mitochondrial DNA sequences of the green turtle and blue-tailed mole skink: statistical evidence for archosaurian affinity of turtles. *Molecular Biology and Evolution*, *16*, 784–792.

Kwong, S., Srivathsan, A., & Meier, R. (2012). An update on DNA barcoding: Low species coverage and numerous unidentified sequences. *Cladistics*, *28*(6), 639–644. https://doi.org/10.1111/j.1096-0031.2012.00408.x

Larkin, M., Blackshields, G., Brown, N., Chenna, R., McGettigan, P., McWilliam, H., … Higgins, D. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, *23*, 2947–2948.

Lee, J. C.-I., Tsai, L.-C., Huang, M.-T., Jhuang, J.-A., Yao, C.-T., Chin, S.-C., … Hsieh, H.-M. (2008). A novel strategy for avian species identification by cytochrome b gene. *Electrophoresis*, *29*(11), 2413–8. https://doi.org/10.1002/elps.200700711

Lee, J. C.-I., Tsai, L.-C., Hwa, P.-Y., Chan, C.-L., Huang, A., Chin, S.-C., … Hsieh, H.-M. (2010). A novel strategy for avian species and gender identification using the CHD gene. *Molecular and Cellular Probes*, *24*(1), 27–31. https://doi.org/10.1016/j.mcp.2009.08.003

Lee, J., Ryu, S., Kang, S., & Hwang, U. (2012). Complete mitochondrial genome of

the Bewick's swan Cygnus columbianus bewickii (Aves, Anseriformes, Anatidae). *Mitochondrial DNA*, *23*(2), 129–130.

Leeton, P., & Christidis, L. (1993). Feathers from museum bird skins - a good source of DNA for phylogenetic studies. *The Condor*, *95*, 465–466.

Li, B., Zhu, C., Ding, P., Bai, S., & Cui, J. (2016). Complete mitochondrial genome of black grouse (Lyrurus tetrix). *Mitochondrial DNA*, *27*(1), 134–135.

Li, X., Huang, Y., & Lei, F. (2015). Comparative mitochondrial genomics and phylogenetic relationships of the Crossoptilon species (Phasianidae, Galliformes). *BMC Genomics*, *16*, 42. https://doi.org/10.1186/s12864-015-1234-9

Linacre, A. (2008). The use of DNA from non-human sources. *Forensic Science International: Genetics Supplement Series*, *1*, 605–606. https://doi.org/10.1016/j.fsigss.2007.10.108

Linacre, a., Gusmão, L., Hecht, W., Hellmann, a. P., Mayr, W. R., Parson, W., … Morling, N. (2011). ISFG: Recommendations regarding the use of non-human (animal) DNA in forensic genetic investigations. *Forensic Science International: Genetics*, *5*(5), 501–505. https://doi.org/10.1016/j.fsigen.2010.10.017

Lopes, P. (2009). *Taxonomia Alfa e distribuição dos representantes do gênero Aburria Reichenbach , 1853 (Aves : Cracidae)*. Universidade de Sao Paulo.

Lowe, T. M., & Eddy, S. R. (1997). tRNAscan-SE: A program for inproved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research*, *25*, 955–964.

Lucchini, V., & Randi, E. (1998). Mitochondrial DNA sequence variation and phylogeographical structure of rock partridge (Alectoris graeca) populations. *Heredity*, *81 (Pt 5)*(December 1997), 528–36. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9881452

Lutz, H. L., Weckstein, J. D., Patané, J. S. L., Bates, J. M., & Aleixo, A. (2013). Biogeography and spatio-temporal diversification of Selenidera and Andigena toucans (Aves: Ramphastidae). *Molecular Phylogenetics and Evolution*, *69*(3), 873–83. https://doi.org/10.1016/j.ympev.2013.06.017

Ma, L., Zhang, X., Yue, B., & Ran, J. (2010). Complete mitochondrial genome of the Chinese Monal pheasant Lophophorus lhuysii, with phylogenetic implication in Phasianidae. *Mitochondrial DNA*, *21*(1), 5–7.

Macey, J., Larson, A., Ananjeva, N., Fang, Z., & Papenfiss, T. (1997). Two Novel Gene Orders and the Role of Light-Strand Replication in Rearrangement of the Vertebrate Mitochondrial Genome. *Mol. Biol. Evol.*, *14*(1), 91–104.

Macfarlane, C. B. A., Natola, L., Brown, M. W., & Burg, T. M. (2016). Population genetic isolation and limited connectivity in the purple finch (Haemorhous purpureus). *Ecology and Evolution*, *6*(22), 8304–8317. https://doi.org/10.1002/ece3.2524

Maragliano, R., Tartara, A., & Montalti, D. (2008). Post-hatching development of the Common Piping Guan (Aburria cumanensis) in La Plata Zoological Garden, Argentina. *BOLETIM DO GRUPO DE ESPECIALISTAS EM CRACÍDEOS BULLETIN OF THE CRACID SPECIALIST GROUP*, *25*, 39–42.

Marshall, H. D., & Baker, a J. (1997). Structural conservation and variation in the

mitochondrial control region of fringilline finches (Fringilla spp.) and the greenfinch (Carduelis chloris). *Molecular Biology and Evolution*, *14*(2), 173–84. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9029795

Marshall, H. D., & Baker, a J. (1998). Rates and patterns of mitochondrial DNA sequence evolution in fringilline finches (fringilla spp.) and the greenfinch (Carduelis chloris). *Molecular Biology and Evolution*, *15*(6), 638–46. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9615445

Martín-Gálvez, D., Peralta-Sánchez, J. M., Dawson, D. a, Martín-Platero, A. M., Martínez-Bueno, M., Burke, T., & Soler, J. J. (2011). DNA sampling from eggshell swabbing is widely applicable in wild bird populations as demonstrated in 23 species. *Molecular Ecology Resources*, *11*(3), 481–93. https://doi.org/10.1111/j.1755-0998.2010.02953.x

Matos-Maraví, P. F., Peña, C., Willmott, K. R., Freitas, A. V. L., & Wahlberg, N. (2013). Systematics and evolutionary history of butterflies in the "Taygetis clade" (Nymphalidae: Satyrinae: Euptychiina): towards a better understanding of Neotropical biogeography. *Molecular Phylogenetics and Evolution*, *66*(1), 54–68. https://doi.org/10.1016/j.ympev.2012.09.005

Matthee, C. A., Eick, G., Willows-Munro, S., Montgelard, C., Pardini, A. T., & Robinson, T. J. (2007). Indel evolution of mammalian introns and the utility of non-coding nuclear markers in eutherian phylogenetics. *Molecular Phylogenetics and Evolution*, *42*(3), 827–837. https://doi.org/10.1016/j.ympev.2006.10.002

McDougall, P. T., Réale, D., Sol, D., & Reader, S. M. (2006). Wildlife conservation and animal temperament: Causes and consequences of evolutionary change for captive, reintroduced, and wild populations. *Animal Conservation*, *9*(1), 39–48. https://doi.org/10.1111/j.1469-1795.2005.00004.x

Meiklejohn, K. a., Danielson, M. J., Faircloth, B. C., Glenn, T. C., Braun, E. L., & Kimball, R. T. (2014). Incongruence among different mitochondrial regions: A case study using complete mitogenomes. *Molecular Phylogenetics and Evolution*, *78*, 314–323. https://doi.org/10.1016/j.ympev.2014.06.003

Merrick, M. J., & Koprowski, J. L. (2017). Should we consider individual behavior differences in applied wildlife conservation studies? *Biological Conservation*, *209*(February), 34–44. https://doi.org/10.1016/j.biocon.2017.01.021

Miller, M. A., Pfeiffer, W., & Schwartz, T. (2011). The CIPRES Science Gateway : A Community Resource for Phylogenetic Analyses. In *2011 TeraGrid Conference: Extreme Digital Discovery*.

Miller, M. J., Bermingham, E., Klicka, J., Escalante, P., do Amaral, F. S. R., Weir, J. T., & Winker, K. (2008). Out of Amazonia again and again: episodic crossing of the Andes promotes diversification in a lowland forest flycatcher. *Proceedings. Biological Sciences / The Royal Society*, *275*(1639), 1133–42. https://doi.org/10.1098/rspb.2008.0015

Mindell, D. P., Sorenson, M. D., & Dimcheff, D. E. (1998). An extra nucleotide is not translated in mitochondrial ND3 of some birds and turtles. *Molecular Biology and Evolution*, *15*, 1568–1571.

Mindell, D. P., Sorenson, M. D., & Dimcheff, D. E. (1998). Multiple independent origins of mitochondrial gene order in birds. *Proc. Nat. Acad. Sci. USA*, *95*, 10693–10697.

Mindell, D. P., Sorenson, M. D., Dimcheff, D. E., Hasegawa, M., Ast, J. C., & Yuri, T. (1999). Interordinal relationships of birds and other reptiles based on whole mitochondrial genomes. *Systematic Biology*, *48*(1), 138–152. https://doi.org/10.1080/106351599260490

Mitchell, D., Willerslev, E., & Hansen, A. (2005). Damage and repair of ancient DNA. *Mutation Research*, *571*(1–2), 265–76. https://doi.org/10.1016/j.mrfmmm.2004.06.060

Mooney, H., Larigauderie, A., Cesario, M., Elmquist, T., Hoegh-Guldberg, O., Lavorel, S., … Yahara, T. (2009). Biodiversity, climate change, and ecosystem services. *Current Opinion in Environmental Sustainability*, *1*(1), 46–54. https://doi.org/10.1016/j.cosust.2009.07.006

Morgan-Richards, M., Trewick, S. a, Bartosch-Härlid, A., Kardailsky, O., Phillips, M. J., McLenachan, P. a, & Penny, D. (2008). Bird evolution: testing the Metaves clade with six new mitochondrial genomes. *BMC Evolutionary Biology*, *8*, 20. https://doi.org/10.1186/1471-2148-8-20

Moritz, C. (1994). Applications of mitochondrial DNA analysis in conservation: a critical review. *Molecular Ecology*, *3*(4), 401–411.

Moritz, C., & Cicero, C. (2004). DNA barcoding: promise and pitfalls. *PLoS Biology*, *2*(10), e354. https://doi.org/10.1371/journal.pbio.0020354

Moura, N. G., Lees, A. C., Andretti, C. B., Davis, B. J. W., Solar, R. R. C., Aleixo, A., … Gardner, T. A. (2013). Avian biodiversity in multiple-use landscapes of the Brazilian Amazon. *Biological Conservation*, *167*, 339–348. https://doi.org/10.1016/j.biocon.2013.08.023

Moyle, R. G. (2004). Phylogenetics of barbets (Aves: Piciformes) based on nuclear and mitochondrial DNA sequence data. *Molecular Phylogenetics and Evolution*, *30*(1), 187–200. https://doi.org/10.1016/S1055-7903(03)00179-9

Mundy, N., Unitt, P., & Woodruff, D. (1997). Skin from Feet of Museum Specimens as a Non-Destructive Source of DNA for Avian Genotyping. *The Auk*, *114*(1), 126–129.

Naranjit, K. (2010). Knowledge of the critically endangered pawi continues to increase. *Annual Review of the World Pheasant Association*, 31.

Nei, M., & Kumar, S. (2000). *Molecular Evolution and Phylogenetics*. Oxford University Press, Inc.

Network Aware. (2015). Sequin Application Version 13.70, National Centre for Biotechnology Information.

Nichols, R. a., Bruford, M. W., & Groombridge, J. J. (2001). Sustaining genetic variation in a small population: Evidence from the Mauritius kestrel. *Molecular Ecology*, *10*(3), 593–602. https://doi.org/10.1046/j.1365-294X.2001.01204.x

Nilsson, M. a., Arnason, U., Spencer, P. B. S., & Janke, A. (2004). Marsupial relationships and a timeline for marsupial radiation in South Gondwana. *Gene*, *340*, 189–196. https://doi.org/10.1016/j.gene.2004.07.040

Nilsson, M. a., Gullberg, A., Spotorno, A. E., Arnason, U., & Janke, A. (2003). Radiation of Extant Marsupials after the K/T Boundary: Evidence from Complete Mitochondrial Genomes. *Journal of Molecular Evolution*, *57*, 3–12. https://doi.org/10.1007/s00239-003-0001-8

Nishibori, M., Hayashi, T., Tsudzuki, M., Yamamoto, Y., & Yasue, H. (2001). Complete sequence of the Japanese quail (Coturnix japonica) mitochondrial genome and its genetic relationship with related species. *Animal Genetics*, *32*(6), 380–5. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11736810

Nishibori, M., Hayashi, T., & Yasue, H. (2004). Complete nucleotide sequence of Numida meleagris (Helmeted Guineafowl) Mitochondrial Genome. *Journal of Poultry Science*, *41*, 259–268.

Nishibori, M., Shimogiri, T., Hayashi, T., & Yasue, H. (2005). Molecular evidence for hybridization of species in the genus Gallus except for Gallus varius. *Animal Genetics*, *36*(5), 367–375. https://doi.org/10.1111/j.1365-2052.2005.01318.x

Nishibori, M., Tsudzuki, M., Hayashi, T., Yamamoto, Y., & Yasue, H. (2002). Complete nucleotide sequence of the Coturnix chinensis (blue-breasted quail) mitochondrial genome and a phylogenetic analysis with related species. *The Journal of Heredity*, *93*(6), 439–44. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12642645

Ogden, R. (2011). Unlocking the potential of genomic technologies for wildlife forensics. *Molecular Ecology Resources*, *11*(SUPPL. 1), 109–116. https://doi.org/10.1111/j.1755-0998.2010.02954.x

Ogden, R., Dawnay, N., & McEwing, R. (2009). Wildlife DNA forensics—bridging the gap between conservation genetics and law enforcement. *Endangered Species Research*, *9*, 179–195. https://doi.org/10.3354/esr00144

Ogden, T. H., & Rosenberg, M. S. (2006). Multiple Sequence Alignment Accuracy and Phylogenetic Inference. *Syst. Biol*, *55*(2), 314–328. https://doi.org/10.1080/10635150500541730

Olsson, U., Sundberg, P., Alström, P., Gelang, M., & Ericson, P. G. P. (2008). What is proper vouchering in phylogenetic studies of birds?-A reply to Peterson et al. (2007). *Molecular Phylogenetics and Evolution*, *48*(1), 383–385. https://doi.org/10.1016/j.ympev.2008.03.031

Omeire, D., Abdin, S., Brookes, D., & Miranda, H. (2015). Complete mitochondrial genome of Germain's Peacock-Pheasant Polyplectron germaini (Aves, Galliformes, Phasianidae). *Mitochondrial DNA*, *26*(2), 319–320.

Ong, A. H. K., & Vellayan, S. (2008). An evaluation of CHD-Specific primer sets for sex typing of birds from feathers. *Zoo Biology*, *27*(1), 62–9. https://doi.org/10.1002/zoo.20163

Oskam, C. L., Haile, J., McLay, E., Rigby, P., Allentoft, M. E., Olsen, M. E., … Bunce, M. (2010). Fossil avian eggshell preserves ancient DNA. *Proceedings. Biological Sciences / The Royal Society*, *277*(1690), 1991–2000. https://doi.org/10.1098/rspb.2009.2019

Pacheco, M. A., Battistuzzi, F. U., Lentino, M., Aguilar, R. F., Kumar, S., & Escalante, A. a. (2011). Evolution of modern birds revealed by mitogenomics: Timing the radiation and origin of major orders. *Molecular Biology and Evolution*, *28*(6), 1927–1942. https://doi.org/10.1093/molbev/msr014

Page, R., & Holmes, E. (1998). *Molecular Evolution: A Phylogenetic Approach*. Blackwell Publishing, Oxford.

Paplinska, J. Z., Taggart, D. a., Corrigan, T., Eldridge, M. D. B., & Austin, J. J. (2011). Using DNA from museum specimens to preserve the integrity of evolutionarily

significant unit boundaries in threatened species. *Biological Conservation*, *144*(1), 290–297. https://doi.org/10.1016/j.biocon.2010.09.005

Parmesan, C. (2006). Ecological and Evolutionary Responses to Recent Climate Change. *Annu.Rev. Ecol. Evol. Syst.*, *37*, 637–669. https://doi.org/10.2307/annurev.ecolsys.37.091305.30000024

Parson, W., Pegoraro, K., Niederstätter, H., Föger, M., & Steinlechner, M. (2000). Species identification by means of the cytochrome b gene. *International Journal of Legal Medicine*, *114*(1–2), 23–28. https://doi.org/10.1007/s004140000134

Patel, S., Weckstein, J. D., Patané, J. S. L., Bates, J. M., & Aleixo, A. (2011). Temporal and spatial diversification of Pteroglossus araçaris (AVES: Ramphastidae) in the neotropics: constant rate of diversification does not support an increase in radiation during the Pleistocene. *Molecular Phylogenetics and Evolution*, *58*(1), 105–15. https://doi.org/10.1016/j.ympev.2010.10.016

Paton, T. a, & Baker, A. J. (2006). Sequences from 14 mitochondrial genes provide a well-supported phylogeny of the Charadriiform birds congruent with the nuclear RAG-1 tree. *Molecular Phylogenetics and Evolution*, *39*(3), 657–67. https://doi.org/10.1016/j.ympev.2006.01.011

Paton, T., Haddrath, O., & Baker, A. J. (2002). Complete mitochondrial DNA genome sequences show that modern birds are not descended from transitional shorebirds. *Proceedings. Biological Sciences / The Royal Society*, *269*(1493), 839–46. https://doi.org/10.1098/rspb.2002.1961

Pereira, S. L. (2000). Mitochondrial genome organization and vertebrate phylogenetics. *Genetics and Molecular Biology*, *23*(4), 745–752.

Pereira, S. L., & Baker, A. J. (2006). A molecular timescale for galliform birds accounting for uncertainty in time estimates and heterogeneity of rates of DNA substitutions across lineages and sites. *Molecular Phylogenetics and Evolution*, *38*(2), 499–509. https://doi.org/10.1016/j.ympev.2005.07.007

Pereira, S. L., Grau, E. T., & Wajntal, A. (2004). Molecular architecture and rates of DNA substitutions of the mitochondrial control region of cracid birds. *Genome*, *47*, 535–545. https://doi.org/10.1139/G04-009

Pereira, S. L. P., Baker, A. J., & Wajntal, A. W. (2002). Combined Nuclear and Mitochondrial DNA Sequences Resolve Generic Relationships within the Cracidae (Galliformes , Aves). *Syst. Biol.*, *51*(6), 946–958. https://doi.org/10.1080/10635150290156051

Pereira, S. L., & Wajntal, A. (2001). Estimates of the genetic variability in a natural population of Bare-faced Curassow Crax fasciolata (Aves , Galliformes , Cracidae). *Bird Conservation International*, *11*, 301–308.

Pereira, S. L., & Wajntal, A. W. (1999). Reintroduction of guans of the genus Penelope (Cracidae, Aves) in reforested areas of Brazil: assessment by DNA fingerprinting. *Biological Conservation*, *87*, 31–38.

Peterson, A. T., Moyle, R. G., Nyári, A. S., Robbins, M. B., Brumfield, R. T., & Remsen, J. V. (2007). The need for proper vouchering in phylogenetic studies of birds. *Molecular Phylogenetics and Evolution*, *45*(3), 1042–4. https://doi.org/10.1016/j.ympev.2007.08.019

Ping-ping, J., Ping, D., & Sheng-Guo, F. (2006). Isolation and characterization of microsatellite markers in Elliot ’ s pheasant (Syrmaticus ellioti). *Molecular*

Ecology Notes, *6*, 1160–1161. https://doi.org/10.1111/j.1471-8286.2006.01472.x

Pratt, R., Gibb, G., Morgan-Richards, M., Phillips, M., Hendy, M., & Penny, D. (2009). Toward resolving deep neoaves phylogeny: data, signal enhancement, and priors. *Molecular Biology and Evolution*, *26*(2), 313–326.

Primmer, C. R., Moller, A., & Ellegren, H. (1996). A wide ranging survey of cross-species microsatellite amplification in birds. *Molecular Ecology*, *5*, 365–378.

Primmer, C. R., Painter, J. N., Koskinen, M. T., Palo, J. U., & Merila, J. (2005). Factors affecting avian cross-species microsatellite amplification. *Journal of Avian Biology*, *36*, 348–360.

Prum, R. O., Berv, J. S., Dornburg, A., Field, D. J., Townsend, J. P., Moriarty Lemmon, E., & Lemmon, A. R. (2015). A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature*, *526*(7574), 569–573. https://doi.org/10.1038/nature15697

Prychitko, T. M., & Moore, W. S. (1997). The utility of DNA sequences of an intron from the beta-fibrinogen gene in phylogenetic analysis of woodpeckers (Aves: Picidae). *Molecular Phylogenetics and Evolution*, *8*(2), 193–204. https://doi.org/10.1006/mpev.1997.0420

Prychitko, T. M., & Moore, W. S. (2000). Comparative evolution of the mitochondrial cytochrome b gene and nuclear beta-fibrinogen intron 7 in woodpeckers. *Molecular Biology and Evolution*, *17*(7), 1101–11. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10889223

Randi, E. (2008). Detecting hybridization between wild species and their domesticated relatives. *Molecular Ecology*, *17*(1), 285–293. https://doi.org/10.1111/j.1365-294X.2007.03417.x

Randi, E., Lucchini, V., Hennache, a, Kimball, R. T., Braun, E. L., & Ligon, J. D. (2001). Evolution of the mitochondrial DNA control region and cytochrome b genes and the inference of phylogenetic relationships in the avian genus Lophura (Galliformes). *Molecular Phylogenetics and Evolution*, *19*(2), 187–201. https://doi.org/10.1006/mpev.2001.0922

Redford, K. H., & Robinson, J. G. (2009). The game of choice: Patterns of indian and colonist hunting in the Neotropics. *American Anthropologist, New Series*, *89*(3), 650–667.

Reed, K. M., Mendoza, K. M., & Beattie, C. W. (2000). Comparative analysis of microsatellite loci in chicken and turkey. *Genome / National Research Council Canada = Génome / Conseil National de Recherches Canada*, *43*(5), 796–802. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11081969

Remsen, J. V. (2007). *Proposal 277 to the South American Classification Committee: Argument for the merge of Pipile into Aburria*.

Remsen, J. V., Cadena, C. D., Jaramillo, A., Nores, M., Pacheco, J. F., Pérez-Emán, J., … Zimmer, K. J. (2012). A classification of the bird species of South America. Retrieved July 11, 2012, from http://www.museum.lsu.edu/~Remsen/SACCBaseline.html

Rhymer, M., & Simberloff, D. (1996). EXTINCTION BY HYBRIDIZATION AND INTROGRESSION. *Annual Review of Ecology and Systematics*, *27*, 83–109.

Ribas, C. C., Miyaki, C. Y., & Carcraft, J. (2009). Phylogenetic relationships,

diversification and biogeography in Neotropical Brotogeris parakeets. *Journal of Biogeography*, *36*(9), 1712–1729.

Ribas, C. C., Moyle, R. G., Miyaki, C. Y., & Cracraft, J. (2007). The assembly of montane biotas: linking Andean tectonics and climatic oscillations to independent regimes of diversification in Pionus parrots. *Proceedings. Biological Sciences / The Royal Society*, *274*(1624), 2399–2408. https://doi.org/10.1098/rspb.2007.0613

Robert, A. (2009). Captive breeding genetics and reintroduction success. *Biological Conservation*, *142*(12), 2915–2922. https://doi.org/10.1016/j.biocon.2009.07.016

Robinson, L. (2011). *Genetic analysis of the critically endangered Trinidad Piping-guan (Pipile pipile): Implications for phylogenetic placement and conservation strategies*. University of Chester.

Roe, A. D., & Sperling, F. a H. (2007). Patterns of evolution of mitochondrial cytochrome c oxidase I and II DNA and implications for DNA barcoding. *Molecular Phylogenetics and Evolution*, *44*(1), 325–45. https://doi.org/10.1016/j.ympev.2006.12.005

Rohland, N., & Hofreiter, M. (2007). Comparison and optimization of ancient DNA extraction. *BioTechniques*, *42*(3), 343–52. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17390541

Rohland, N., Siedel, H., & Hofreiter, M. (2010). A rapid column-based ancient DNA extraction method for increased sample throughput. *Molecular Ecology Resources*, *10*(4), 677–83. https://doi.org/10.1111/j.1755-0998.2009.02824.x

Rompre, G., Robinson, W. D., Desrochers, A., & Angehr, G. (2009). Predicting Declines in Avian Species Richness under Nonrandom Patterns of Habitat Loss in a Neotropical Landscape. *Ecological Applications*, *19*(6), 1614–1627.

Ronquist, F., van der Mark, P., & Huelsenbeck, J. P. (2011). Bayesian phylogenetic analysis using MrBayes. In P. Lemey, M. Salemi, & A. Vandamme (Eds.), *The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing* (Second Edi, pp. 210–266). Cambridge University Press.

Rull, V. (2011). Neotropical biodiversity: timing and potential drivers. *Trends in Ecology & Evolution*, *26*(10), 508–13. https://doi.org/10.1016/j.tree.2011.05.011

Ruokonen, M., & Kvist, L. (2002). Structure and evolution of the avian mitochondrial control region. *Molecular Phylogenetics and Evolution*, *23*(3), 422–32. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12099796

Segelbacher, G. (2002). Noninvasive genetic analysis in birds : testing reliability of feather samples. *Molecular Ecology Notes*, *2*, 367–369. https://doi.org/10.1046/j.1471-8286

Segelbacher, G., & Piertney, S. (2007). Phylogeography of the European capercaillie (Tetrao urogallus) and its implications for conservation. *Journal of Ornithology*, *148*(S2), 269–274. https://doi.org/10.1007/s10336-007-0153-1

Sgrò, C. M., Lowe, A. J., & Hoffmann, A. a. (2011). Building evolutionary resilience for conserving biodiversity under climate change. *Evolutionary Applications*, *4*(2), 326–337. https://doi.org/10.1111/j.1752-4571.2010.00157.x

Shen, Y.-Y., Liang, L., Sun, Y.-B., Yue, B.-S., Yang, X.-J., Murphy, R. W., & Zhang,

Y.-P. (2010). A mitogenomic perspective on the ancient, rapid radiation in the Galliformes with an emphasis on the Phasianidae. *BMC Evolutionary Biology*, *10*(132), 1–10. https://doi.org/10.1186/1471-2148-10-132

Shen, Y. Y., Shi, P., Sun, Y. B., & Zhang, Y. P. (2009). Relaxation of selective constraints on avian mitochondrial DNA following the degeneration of flight ability. *Genome Research*, *19*(10), 1760–1765. https://doi.org/10.1101/gr.093138.109

Sibley, C. G., & Ahlquist, J. (1990). *Phylogeny and Classification of the Birds*. New Haven: Yale University Press.

Slack, K. E., Delsuc, F., Mclenachan, P. a., Arnason, U., & Penny, D. (2007). Resolving the root of the avian mitogenomic tree by breaking up long branches. *Molecular Phylogenetics and Evolution*, *42*, 1–13. https://doi.org/10.1016/j.ympev.2006.06.002

Slack, K. E., Janke, A., Penny, D., & Arnason, U. (2003). Two new avian mitochondrial genomes (penguin and goose) and a summary of bird and reptile mitogenomic features. *Gene*, *302*(1–2), 43–52. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12527195

Slack, K. E., Jones, C. M., Ando, T., Harrison, G. L., Fordyce, R. E., Arnason, U., & Penny, D. (2006). Early penguin fossils, plus mitochondrial genomes, calibrate avian evolution. *Molecular Biology and Evolution*, *23*(6), 1144–1155. https://doi.org/10.1093/molbev/msj124

Smithsonian Institute. (2017). Division of Birds Collection. Retrieved February 28, 2017, from http://collections.nmnh.si.edu/search/birds/

Sonsthagen, S. A., Wilson, R. E., & Underwood, J. G. (2017). Genetic implications of bottleneck effects of differing severities on genetic diversity in naturally recovering populations: An example from Hawaiian coot and Hawaiian gallinule. *Ecology and Evolution*, (September), 9925–9934. https://doi.org/10.1002/ece3.3530

Sorenson, M. D. (2003). Avian mtDNA primers. Retrieved from http://people.bu.edu/msoren/Bird.mt.Primers.pdf

Sorenson, M. D., Ast, J. C., Dimcheff, D. E., Yuri, T., & Mindell, D. P. (1999). Primers for a PCR-based approach to mitochondrial genome sequencing in birds and other vertebrates. *Molecular Phylogenetics and Evolution*, *12*(2), 105–14. https://doi.org/10.1006/mpev.1998.0602

Sorenson, M. D., & Quinn, T. W. (1998). Numts : A Challenge for Avian Systematics and Population Biology. *The Auk*, *115*(1), 214–221.

Spooner, D. M. (2009). DNA barcoding will frequently fail in complicated groups: An example in wild potatoes. *American Journal of Botany*, *96*(6), 1177–1189. https://doi.org/10.3732/ajb.0800246

Stemmer, W. (1994). DNA shuffling by random fragmentation and reassembly: In vitro recombination for molecular evolution. *Proc. Nat. Acad. Sci. USA*, *91*(October), 10747–10751.

Strahl, S. D., & Grajal, A. (1991). Conservation of large avian frugivores and the management of Neotropical protected areas. *Oryx*, *25*(1), 50–55.

Suarez, A. V, & Tsutsui, N. D. (2004). The Value of Museum Collections for Research

and Society. *Bioscience*, *54*(1), 66–74.

Szabo, J. K., Khwaja, N., Garnett, S. T., & Butchart, S. H. M. (2012). Global patterns and drivers of avian extinctions at the species and subspecies level. *PloS One*, *7*(10), e47080. https://doi.org/10.1371/journal.pone.0047080

Szalanski, A. L., Church, K. E., Oates, D. W., Bischof, R., & Powers, T. O. (2000). Mitochondrial DNA variation within and among wild turkey.pdf. *Transactions of Th eNebraska Academy of Sciences*, *26*, 47–53.

Szibor, R., Michael, M., Plate, I., & Krause, D. (2000). Efficiency of forensic mtDNA analysis. Case examples demonstrating the identification of traces. *Forensic Science International*, *113*(1–3), 71–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10978604

Tallmon, D. a., Luikart, G., & Waples, R. S. (2004). The alluring simplicity and complex reality of genetic rescue. *Trends in Ecology and Evolution*, *19*(9), 489–496. https://doi.org/10.1016/j.tree.2004.07.003

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6:Molecular Evolutionary Genetics Analysis version 6. *Molecular Biology and Evolution*, *30*, 2725–2729.

Tavares, E. S., Gonçalves, P., Miyaki, C. Y., & Baker, A. J. (2011). DNA barcode detects high genetic structure within neotropical bird species. *PloS One*, *6*(12), e28543. https://doi.org/10.1371/journal.pone.0028543

Terborgh, J. (1974). Preservation of Natural Diversity: The Problem of Extinction Prone Species. *Bioscience*, *24*(12), 715–722.

Tian, H., Hühmer, a F., & Landers, J. P. (2000). Evaluation of silica resins for direct and efficient extraction of DNA from complex biological matrices in a miniaturized format. *Analytical Biochemistry*, *283*, 175–191. https://doi.org/10.1006/abio.2000.4577

Tilston Smith, B., McCormack, J., Cuervo, A., Hickerson, M., Alexio, A., Cadena, C., … Brumfield, R. (2014). The drivers of tropical speciation. *Nature*, *515*, 406–409. https://doi.org/10.1038/nature13687

Tobe, S., & Linacre, A. (2010). DNA typing in wildlife crime: Recent developments in species identification. *Forensic Sci Med Pathol.*, *6*, 195–206.

Tobe, S. S., Kitchener, A. C., & Linacre, A. M. T. (2010). Reconstructing mammalian phylogenies: A detailed comparison of the cytochrome b and cytochrome oxidase subunit i mitochondrial genes. *PLoS ONE*, *5*(11). https://doi.org/10.1371/journal.pone.0014156

Tobe, S. S., Kitchener, A., & Linacre, A. (2009). Cytochrome b or cytochrome c oxidase subunit I for mammalian species identification—An answer to the debate. *Forensic Science International: Genetics Supplement Series*, *2*(1), 306–307. https://doi.org/10.1016/j.fsigss.2009.08.053

Tobe, S. S., & Linacre, A. (2009). Identifying endangered species from degraded mixtures at low levels. *Forensic Science International: Genetics Supplement Series*, *2*(1), 304–305. https://doi.org/10.1016/j.fsigss.2009.08.052

Tobe, S. S., & Linacre, A. M. T. (2008). A technique for the quantification of human and non-human mammalian mitochondrial DNA copy number in forensic and other mixtures. *Forensic Science International. Genetics*, *2*(4), 249–56.

https://doi.org/10.1016/j.fsigen.2008.03.002

Tordoff, H., & MacDonald, J. R. (1957). A new bird (Family Cracidae) from the early Oligocene of South Dakota. *Auk*, *74*, 174–184.

Trimbos, K. B., Broekman, J., Kentie, R., Musters, C. J. M., & Snoo, G. R. (2009). Using eggshell membranes as a DNA source for population genetic research. *Journal of Ornithology*, *150*(4), 915–920. https://doi.org/10.1007/s10336-009-0422-2

Tu, J., Si, F., Wu, Q., Cong, B., Xing, X., & Yang, F. (2014). The complete mitochondrial genome of the Muscovy duck, Cairina moschata (Anserifomes, Anatidae, Cairina). *Mitochondrial DNA*, *25*(2), 102–103.

Valentini, A., Pompanon, F., & Taberlet, P. (2009). DNA barcoding for ecologists. *Trends in Ecology & Evolution*, *24*(2), 110–7. https://doi.org/10.1016/j.tree.2008.09.011

Van Tuinen, M., & Dyke, G. J. (2004). Calibration of galliform molecular clocks using multiple fossils and genetic partitions. *Molecular Phylogenetics and Evolution*, *30*(1), 74–86. https://doi.org/10.1016/S1055-7903(03)00164-7

van Tuinen, M., & Hedges, S. B. (2001). Calibration of avian molecular clocks. *Molecular Biology and Evolution*, *18*(2), 206–213. https://doi.org/10.1093/oxfordjournals.molbev.a003794

van Tuinen, M., Sibley, C. G., & Hedges, S. B. (1998). Phylogeny and biogeography of ratite birds inferred from DNA sequences of the mitochondrial ribosomal genes. *Molecular Biology and Evolution*, *15*(4), 370–6. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10723745

van Tuinen, M., Sibley, C. G., & Hedges, S. B. (2000). The early history of modern birds inferred from DNA sequences of nuclear and mitochondrial ribosomal genes. *Molecular Biology and Evolution*, *17*, 451–457.

Vandamme, A. (2011). Basic concepts of molecular evolution. In P. Lemey, M. Salemi, & A. Vandamme (Eds.), *The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing* (Second, pp. 3–30). Cambridge University Press.

Vaurie, C. (1967). Systematic Notes on the Bird Family Cracidae No. 7: The Genus Pipile. *American Museum Novitates*, (2296), 1–16.

Vegas-Vilarrúbia, T., Rull, V., Montoya, E., & Safont, E. (2011). Quaternary palaeoecology and nature conservation: a general review with examples from the neotropics. *Quaternary Science Reviews*, *30*(19–20), 2361–2388. https://doi.org/10.1016/j.quascirev.2011.05.006

Verma, S., & Singh, L. (2003). Novel universal primers establish identity of an enormous number of animal species for forensic application. *Molecular Ecology Notes*, *3*(1), 28–31. https://doi.org/10.1046/j.1471-8286.2003.00340.x

Wandeler, P., Hoeck, P. E. a, & Keller, L. F. (2007). Back to the future: museum specimens in population genetics. *Trends in Ecology & Evolution*, *22*(12), 634–42. https://doi.org/10.1016/j.tree.2007.08.017

Wang, N., Kimball, R. T., Braun, E. L., Liang, B., & Zhang, Z. (2013). Assessing Phylogenetic Relationships among Galliformes: A Multigene Phylogeny with Expanded Taxon Sampling in Phasianidae. *PLoS ONE*, *8*(5).

https://doi.org/10.1371/journal.pone.0064312

Watanabe, M., Nikaido, M., Tsuda, T., Kobayashi, T., Mindell, D., Cao, Y., … Hasegawa, M. (2006). New candidate species most closely related to penguins. *Gene*, *378*, 65–73.

Waylen, K. (2005). *Turtles & Tourism , Perceptions & Pawi*. University of London.

Waylen, K. a., McGowan, P. J. K., & Milner-Gulland, E. J. (2009). Ecotourism positively affects awareness and attitudes but not conservation behaviours: a case study at Grande Riviere, Trinidad. *Oryx*, *43*(3), 343. https://doi.org/10.1017/S0030605309000064

Weber, D. S. (2004). Genetic Consequences of a Severe Population Bottleneck in the Guadalupe Fur Seal (Arctocephalus townsendi). *Journal of Heredity*, *95*(2), 144–153. https://doi.org/10.1093/jhered/esh018

Weber, D. S., Stewart, B. S., Garza, J. C., & Lehman, N. (2000). An empirical genetic assessment of the severity of the northern elephant seal population bottleneck. *Current Biology*, *10*(20), 1287–1290. https://doi.org/10.1016/S0960-9822(00)00759-4

Wetmore, A. (1923). Avian fossils from the Miocene and Pliocene of Nebraska. *Bull. Amer. Mus. Nat. Hist.*, *48*, 483–507.

Wetmore, A. (1956). A fossil guan from the Oligocene of South Dakota. *Condor*, *58*, 234–235.

Wiley, E. O. (1981). *Phylogenetics: The Thoery and Practice of Phylogenetic Systematics*. Library of Congress Cataloging in Publication Data.

Will, K. W., & Rubinoff, D. (2004). Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. *Cladistics*, *20*(1), 47–55. https://doi.org/10.1111/j.1096-0031.2003.00008.x

Williams, S. E., & Hoffman, E. A. (2009). Minimizing genetic adaptation in captive breeding programs: A review. *Biological Conservation*, *142*(11), 2388–2400. https://doi.org/10.1016/j.biocon.2009.05.034

Williamson, A. L. (2012). Touch DNA : Forensic Collection and Application to Investigations. *Journal of Assoc Crime Scene Reconstruction*, *18*(1), 1–6.

Wilson-Wilde, L., Norman, J., Robertson, J., Sarre, S., & Georges, A. (2010). Current issues in species identification for forensic science and the validity of using cytochrome oxidase 1 (CO1) gene. *Forensic Sci Med Pathol.*, *6*, 233–241.

Wilson, A. C., Cann, R. L., Carr, S. M., George, M., Gyllensten, U. B., Helm-Bychowski, K. M., … Stoneking, M. (1985). Mitochondrial DNA and two perspectives on evolutionary genetics. *Biological Journal of the Linnean Society*, *26*(January), 375–400. https://doi.org/10.1111/j.1095-8312.1985.tb02048.x

Wink, M. (2006). Use of DNA markers to study bird migration. *Journal of Ornithology*, *147*(2), 234–244. https://doi.org/10.1007/s10336-006-0065-5

Yang, D. Y., Eng, B., Waye, J. S., Dudar, J. C., & Saunders, S. R. (1998). Technical note: improved DNA extraction from ancient bones using silica-based spin columns. *American Journal of Physical Anthropology*, *105*(4), 539–43. https://doi.org/10.1002/(SICI)1096-8644(199804)105:4<539::AID-AJPA10>3.0.CO;2-1

Yang, R., Wu, X., Yan, P., Su, X., & Yang, B. (2010). Complete mitochondrial genome of Otis tarda (Gruiformes: Otididae) and phylogeny of Gruiformes inferred from mitochondrial DNA sequences. *Mol. Biol. Rep.*, *37*(7), 3057–3066.

Yang, Z. (2014). *Molecular Evolution: A Statistical Approach*. (O. Oxford University Press, Ed.).

Zeng, T., Tu, F., Ma, L., Yan, C., Yang, N., Zhang, X., … Ran, J. (2013). Complete mitochondrial genome of blood pheasant (Ithaginis cruentus). *Mitochondrial DNA*, *24*(5), 484–486.

Zhang, D., & Hewitt, G. (1996). Nuclear Integrations: challenges for mitochondrial DNA markers. *TREE*, *11*(6), 247–251.

Zhang, G., Li, C., Li, Q., Li, B., Larkin, D. M., Lee, C., … Ganapathy, G. (2014). Comparative genomics reveals insights into avain genome evolution and adaptation. *Science*, *346*(6215), 1311–1321.

Zhou, T., Sha, T., Irwin, D., & Zhang, Y. (2015). Complete mitochondrial genome of the Indian peafowl (Pavo cristatus), with phylogenetic analysis in phasianidae. *Mitochondrial DNA*, *26*(6), 912–913.

Zink, R. M., & Barrowclough, G. F. (2008). Mitochondrial DNA under siege in avian phylogeography. *Molecular Ecology*, *17*(9), 2107–21. https://doi.org/10.1111/j.1365-294X.2008.03737.x

Zink, R. M., Klicka, J., & Barber, B. R. (2004). The tempo of avian diversification during the Quaternary. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *359*(1442), 215-9-20. https://doi.org/10.1098/rstb.2003.1392

Appendix

Appendix 1, Table 1: Complete list of all biological samples used in this study.

Sample Abbreviation	Sample Name / Stud No.	Taxonomic Status	Source	Sex	Comments
R1	Reference 1	*Aburria/ Pipile pipile*	Trinidad, Emperor Valley Zoo, Port of Spain		Wild caught, Deceased 2007 CITES permit 1053 Species confirmed by veterinary examination & DNA
R2	Reference 2, Phoenix	*Aburria/ Pipile pipile*	Trinidad, Emperor Valley Zoo, Port of Spain	F	Captive, Originated in the wild CITES permit 1264 Species confirmed by veterinary examination & DNA
E1	Egg-embryo	*Aburria/ Pipile pipile*	Northern Range, Trinidad		Wild CITES permit 1116 Species confirmed by veterinary examination & DNA
Pawi 1	Quill	*Aburria/ Pipile pipile*	Morne Bleu, Northern Range, Trinidad		Wild CITES permit 1103 Species confirmed by DNA
Pawi 2	Feather 1	*Aburria/ Pipile pipile*	Plato plantation, Grande Riviere, Trinidad		Wild CITES permit 1264 Species confirmed by DNA
Pawi 3	Feather 2	*Aburria/ Pipile pipile*	Plato plantation, Grande Riviere, Trinidad		Wild CITES permit 1264 Species confirmed by DNA
Pawi 4	Feather 3	*Aburria/ Pipile pipile*	Matlot, Grande Riviere, Trinidad		Wild CITES permit 1264 Species confirmed by DNA
Pawi 5	Feather 4	*Aburria/ Pipile pipile*	Matlot, Grande Riviere, Trinidad		Wild CITES permit 1264 Species confirmed by DNA
AC1	1902. 3. 13. 1852	*Aburria/ Pipile cumanensis*	British Natural History Museum, UK	M	Pipile cumanensis Peru, The Andes, S Am. P.O. Simons, 12th March 1900
AC2	1953. 68. 44	*Aburria/ Pipile cumanensis*	British Natural History Museum, UK	M	Pipile cumanensis Cerro Galera, Orienie, Equador Leopoldo Gomez, 12th July 1940
AC3	1940.12.5.43	*Aburria/ Pipile cumanensis*	British Natural History Museum, UK	M	Pipile cumanensis Rio Juno, E. Equador L. Gomez, 8th Dec 1938
AC4	1856. 11. 5. 10	*Aburria/ Pipile cumanensis*	British Natural History Museum, UK		Pipile cumanensis Columbia Verreaux, <1856?
AC5	1922. 3. 5. 185	*Aburria/ Pipile cumanensis*	British Natural History Museum, UK	M	Pipile cumanensis Pomeroon River, Brit Guiana McConnell collection (Described in 'Birds of British Guiyana')
AC6	1892. 1. 16. 138	*Aburria/ Pipile cumanensis*	British Natural History Museum, UK	M	Pipile cumanensis (Jacq) Takutu River, Brit. Guiana H. Whitely, 1888
ACG1	1910. 7. 9. 107	*Aburria/ Pipile grayi*	British Natural History Museum, UK	M	Pipile cumanensis (A. p. grayi) Concireucia, Alto, Paraguay G. W. Tudor, 9th Sept 1909
ACJ1	1889. 6. 1. 254	*Aburria/ Pipile cujubi*	British Natural History Museum, UK		Amazonian Guan, Pipile cujubi Lower Amazon Zoological Society of London, 25th May 1873 (died 9th May 1875)
ACJ2	1850. 11. 30. 21	*Aburria/ Pipile cujubi*	British Natural History Museum, UK		Penelope atrofurpurea? S America Zoological Society of London, Burton 19. V. 97

Appendix 1, Table 1 continued.

Sample Abreviation	Sample Name / Stud No.	Taxonomic Status	Source	Sex	Comments
SPM-AC1	MZUSP 76361	*Pipile cumanensis*	Museo de Zoologia da Universidada de Sao Paulo, Brazil		DEFRA import no. P0A0/2009/050 Rio Macaua, Brazil
SPM-ACJ1	MZUSP 20832	*Pipile cujubi*	Museo de Zoologia da Universidada de Sao Paulo, Brazil	M	DEFRA import no. P0A0/2009/050 Rio Tapajos, Brazil
SPM-ACG1	MZUSP	*Pipile cumanensis grayi*	Museo de Zoologia da Universidada de Sao Paulo, Brazil		DEFRA import no. P0A0/2009/050 Mato Grosso
SPM-CF1	MZUSP 79216	*Crax fasciolata*	Museo de Zoologia da Universidada de Sao Paulo, Brazil		DEFRA import no. P0A0/2009/050
SPM-PT1	MZUSP 28037	*Pauxi tuberosa*	Museo de Zoologia da Universidada de Sao Paulo, Brazil		DEFRA import no. P0A0/2009/050
C1	Red-faced Curassow	*Crax blumanbachii*	Chester Zoological Gardens, UK	M	Bred in captivity
F1	Piping Guan	*Pipile pipile*	Fenton, UK	F	Bred in captivity DNA indicates error in taxonomic record supported by morphology (*Pipile cumanensis* not *Pipile Pipile*)
F2	Piping Guan	*Pipile pipile*	Fenton, UK	M	Bred in captivity DNA indicates error in taxonomic record supported by morphology
B	Piping Guan	*Pipile pipile*	Cotswold Wildlife Park, UK		Bred in captivity DNA indicates error in taxonomic record supported by morphology (*Pipile cumanensis* not *Pipile Pipile*)
L1	Piping Guan	*Pipile pipile*	Lotherton Hall Bird Garden, UK		Bred in captivity DNA indicates error in taxonomic record supported by morphology (*Pipile cumanensis* not *Pipile Pipile*)
L2	Piping Guan	*Pipile pipile*	Lotherton Hall Bird Garden, UK		Bred in captivity DNA indicates error in taxonomic record supported by morphology
L3	Piping Guan	*Pipile pipile*	Lotherton Hall Bird Garden, UK		Bred in captivity DNA indicates error in taxonomic record supported by morphology
L4	Piping Guan	*Pipile pipile*	Lotherton Hall Bird Garden, UK		Bred in captivity DNA indicates error in taxonomic record supported by morphology
L5	Piping Guan	*Pipile pipile*	Lotherton Hall Bird Garden, UK		Bred in captivity DNA indicates error in taxonomic record supported by morphology
L6	Piping Guan	*Pipile pipile*	Lotherton Hall Bird Garden, UK		Bred in captivity DNA indicates error in taxonomic record supported by morphology
Control 1	Chicken	*Gallus gallus*	Domestic, UK	F	
Control 2	Pheasant	*Phasianus colchicus*	UK	F&M	

Appendix 2: Table 1: Complete list of oligonucleotide primers used in the of sequencing of the complete mitochondrial genome and analysis of genetic variation.

Complete Mitochondrial Genome							
Gene	**Genome Position**	**Size bp**	**Forward**	**Sequence 3'-5'**	**Reverse**	**Sequence 3' -5'**	**Reference(s)**
Control Region	0-1227		L2ND6	TAC CCC ATA ATA CGG AGA GG	H2CR	GAG ATG TCC AGA CCT AGA TT	Robinson (2011)
			L436REV	CCT CAC GAG AAA YCA GCA AC	H1251REV	TCT TGG CAT CTT CAG TGC CRT GC	Sorenson (1999)
tRNA Phe	1228-1296		CR Phe L-ag	CTC CAC TAC CAA AAA CAA CC	CR Phe H-ag	TTA GGG CAT TTA CAC TGG AG	This research
12S	1297-2272	975	L1267	YAA AGC ATG RCA CTG AAG HYG	H1859	TCG DTT RYA GRA CAG GCT CCT CTA	Sorenson (2003)
			L1754	TGG GAT TAG ATA CCC CAC TAT G	H2294	TYT CAG GYG TAR GCT GAR TGC TT	Sorenson (2003)
tRNA Val	2273-2345		L2260	CAW GGT AAG YRT ACY GGA ARG TG	H2891	TGR TGG CTG CTT DAR GGC CHA C	Sorenson (2003)
16S	2346-3966		L2725	CGA GCY KGG TGA TAG CTG G	H3292	TGA TTR YGC TAC CTT YGC ACG G	Sorenson (2003)
			L3218	CGM CTG TTT ACC AAA AAC ATM RCC	H3748	CGG TCT GAA CTC AGA TCA CG	Sorenson (2003)
tRNA L	3967-4040		L3722	GGT TTA CGA CCT CGA TGT TGG	H4170	CCW CAN AYR TTD GGN CCY TTD CG	Sorenson (2003)
ND1	4050-5024		L3827		ND1 H-ag		This research & Sorenson (2003)
IQM tRNA Ile, Gln, Met	5025-5096	215	L4500	GTN GCM CAA CAN ATY TCH TAY GAA G	H5201	CCA TCA TTT TCG GGG TAT GG	Sorenson (2003)
	5102-5172 C		L5143	GAA CCT ACA CAR AAG RGA TCA AAA C	H5766	RGA KGA GAA RGC YAG GAT YTT KCG	Sorenson (2003)
	5172-5240						

Appendix 2, Table 1 continued.

Gene	Genome Position	Size bp	Forward	Sequence 3'-5'	Reverse	Sequence 3' -5'	Reference(s)
ND 2	5241-6281	1031	L5216	GGC CCA TAC CCC GRA AAT G	H5766	RGA KGA GAA RGC YAG GAT YTT KCG	Sorenson (2003)
			Lc5568	GTA GGC TGA GTT TGG GGC TA	Hc6508	GTA GGC TGA GTT TGG GGC TA	Robinson (2011)
			L5758	GGN GGN TGA ATR GGN YTN AAY CAR AC	H6313	ACT CTT RTT TAA GGC TTT GAA GGC	Sorenson (2003)
WANCY tRNA Trp,Ala,Asn, Cys,Tyr	6280-6355	371	L6182	TGG TAT CTC CGT AAA ACA CC	H6735	GCC TAC TAT GCC TGC TCA TGT A	Sorenson (2003)
	6362-6430 C						
	6434-6506 C						
	6508-6573 C						
	6573-6643 C						
CO I	6645-8192	1548	Ltyr	TGT AAA AAG GWC TAC AGC CTA ACG C	COI907aH2	GTR GCN GAY GTR AAR TAT GCT CG	Tavares (2008)
			L6615	CCY CTG TAA AAA GGW CTA CAG CC	H7122	ATN GTD GTR ATR AAR TTR ATD GCH CC	Sorenson (2003)
			L7036	GGN ACN GGN TGA ACH GTN TAY CC	H7548	GTD GCN GAN GTR AAR TAD GCT CG	Sorenson (2003)
			L7525	GTN TGR GCH CAY CAY ATR TTY AC	H8121	GGG CAG CCR TGR ATT CAY TC	Sorenson (2003)
			L7987	TCH GAY TAY CCW GAY GCN TAY AC	H8205	GGT TCG ATT CCT TCC TTT CTT G	Sorenson (2003)
SD tRNA Ser,Asp	8124-8254 C		L7987	TCH GAY TAY CCW GAY GCN TAY AC	H8628	TCR TAG STT CAR TAT CAY TGR TGN CC	Sorenson (2003)
	8261-8329						
CO II	8331-9041		L8386	GCN TCV TCN CCH ATY ATR GAA GA	H9235	TCR AAG AAG CTT AGG TTC ATG GTC A	Sorenson (2003)

Appendix 2, Table 1 continued.

Gene	Genome Position	Size bp	Forward	Sequence 3'-5'	Reverse	Sequence 3' -5'	Reference(s)
tRNA Lys	9016-9083		L8929	GGH CAR TGY TCA GAR ATY TGY GG	H9726	AGR TGN CCD GCT GTD AGR TTN GC	Sorenson (2003)
ATP 8	9085-9249						
ATP 6	9240-9923		COII F2-ag	CCT ACT AAT CGG CCT ACG AAA	H10343 REV	TGA GCT CAT GTT ACG GTG ACA CC	This research & Sorenson (2003)
CO III	9923-10706		L10236REV	TTC TGA GCA TTC TTC CAC TCA AGC	H10884REV	GGG TCG AAA CCG CAT TCG TAC CC	This research
tRNA Gly	10708-10775		L10635	CAY CAY TTY GGN TTY GAA GCH GC	H11100	TCT GCY CAY TCT ARK CCT CCY TG	Sorenson (2003)
ND3	10776-11126		L10906REV	CCN TAY GAR TGY GGN TTY GAY CC	ND3 R-ag	TAG AAG CTG AAG TGG GAG AG	This research
tRNA Arg	11128-11195		L11122	CAR GGA GGM YTA GAR TGR GCA GA	H11837	AGR GTD GCY TCR AAT GHR ATR TAR AA	Sorenson (2003)
ND 4 L	11196-11492		L11458REV	TCC ACA CGA ACA CAT GGC TCC GA	ND4 R-ag	CTC ATA GGG CTA GGG TGA G	This research & Sorenson (2003)
ND 4	11486-12863		L12156	CCH AAA GCH CAY GTW GAA GCH CC	H12488	ATT CGG CTG TGT GTY CGY TC	Sorenson (2003)
			ND4 F-ag	TAA TAA TCT CCC ACG GGT TA	H13047	CTT YYA YTT GGA KTT GCA CCA A	This research & Sorenson (2003)
HSL tRNA His, Ser, Leu	12864-12932		L12912	TAG AYT GTG AYT CTR AAA AYA GRA G	H13563	TGN AGD GCD GCD GTR TTD GC	Sorenson (2003)
	12934-12998						
	13000-13070						

Appendix 2, Table 1 continued.

Gene	Genome Position	Size bp	Forward	Sequence 3'-5'	Reverse	Sequence 3' -5'	Reference(s)
ND 5	13071-14888		ND5 L-ag	AAC CTA CAG CCT ACG CAT AAC	H13563	TGN AGD GCD GCD GTR TTD GC	This research & Sorenson (2003)
			L13525REV	GCT GGG AAG GAG TAG GAA TCA TAT CC	H14127	CCT ATT TTT CGR ATR TCY TGY TC	This research & Sorenson (2003)
			ND5 L2-ag	CTT TTC TAC GTC AAG CCA AC	ND5 H-ag	CGG ATG AGT AGG AAG ATT CC	This research
			ND5 L3-ag	AGC CTT CCT ACA CAT YTC AA	ND5 H2-ag	ART AGN CCT GCT RTR ATG CTC G	This research
			ND5 L5-ag	AAC CTA CAG CCT ACG CAT AAC	ND5 H3-ag	GGG AGG TAG ATT AGA GAG TT	This research
			ND5	TAG CTA GGA TCT TTC GCC CT	cytbDesi	TGT AGG TTG CGG ATT AGC CAG C	Robinson (2011)
Cyt B	14893-16035	1140	L14770	TAG GNC CNG ARG GNY TNG C	H15049	GTR TCN GCD GTR TAR TGY ATD GC	Sorenson (2003)
			L14841	AAA AAG CTT CCA TCC AAC ATC TCA GCA TGA TGA AA	H15149	AAA CTG CAG CCC CTC AGA ATG ATA TTT GTC CTC A	Kocher (1989)
			L14841	AAA AAG CTT CCA TCC AAC ATC TCA GCA TGA TGA AA	H16065	TTC AGT TTT TGG TTT ACA AGA C	Kocher (1989) & Kimball (1999)
			CytB F2	GGC TTC TAC TAC GGC TCA TA	CytB R2	CTT AAT ATG TGG AGG GGT GA	This research
TP tRNA Thr, Pro	16039-16107	159	L15311	CTC CCA TGA GGC CAA ATA TC	H16065	TTC AGT TTT TGG TTT ACA AGA C	Kimball (1999)
	16108-16177 C						
ND 6	16184-16705 C	519	LcytB	CAT GAG TGG GCA GCC AAC CAG T	Hdloop	GCT CCC ATA GTT GAG AAC AAC G	Robinson (2011)
E tRNA Glu	16708-16775 C	70	L2ND6		H2CR		Robinson (2011)

Appendix 2, Table 1 continued.

Nuclear Genome Sequencing							
Gene	**Chromosome**	**Size**	**Forward**	**Sequence 3'-5'**	**Reverse**	**Sequence 3' -5'**	**Reference(s)**
CHD	Z & W		CHD1F	TATCGTCAGTTTCCTTTTCAGGT	CHD1R	CCTTTTATTGATCCATCAAGCCT	Lee et al (2010)
CHD	Z & W		P2	TCT GCA TCG CTA AAT CCT TT	P8	CTC CCA AGG ATG AGR AAY TG	Griffiths et al (1998)
FGB intron 7	4		FIB-BI7L	TCC CCA GTA GTA TCT GCC ATT AGG GTT	FIB-BI7Uc	GGAGAAAACAGGACAATGACAATTCAC	Prychitko & Moore (1997)
			FIB-BI7L	TCC CCA GTA GTA TCT GCC ATT AGG GTT	FIB U2	GTA ACC CAT AAT GGG TCC TGA G	Prychitko & Moore (1997)
			FIB L2	CTT CTG AGT AGG CAG AAC TT	FIB-BI7Uc	GGA GAA AAC AGG ACA ATG ACA ATT CAC	Prychitko & Moore (1997)
SERPINB14 intron c	2	655	SERP-F	GTTCGCTTTGATAAACTTCCAGG	SERP-R	GGTGATTTGGTTGAGNATGTC	Kimball et al (2009)
GARS intron 11	2	303	GARS-11F	GTTCCTCTYATAGCTGAGAAGC	GARS-12R	CGTCCTTCTTRTAAGCTTTGCC	Kimball et al (2009)
PER2 intron 9	9	510	PER-9F	CATCTTCAYCCAAATGACAGACC	PER-10R	CCTGATTGGTGAATAGTCAAAAGG	Kimball et al (2009)
RHO	12	1152	RHO-F	GAACGGGTACTTTGTCTTTGGAGTAAC	RHO-R	CCCATGATGGCGTGGTTCTCCCC	Wang et al (2013)
CLTCL1 intron 7	15	605	CLTCL1-F	CACCAATGTTCTGCAGAATCCTGA	CLTCL1-R	CCAGCTTATCTTCCTTNAGCCATTTCTC	Wang et al (2013)
CLTC intron 7	19	817	CLTC-F	CAGAATCCTGATCTAGCTTTACGAATGGC	CLTC-R	CATTTCTCCAGAAGTTGTTTGCGTCC	Wang et al (2013)
EEF2	28	1125	EEF2-F	GAAACAGTTTGCTGAGATGTATGTTGC	EEF2-R	GGTTTGCCCTCCTTGTCCTTATC	Wang et al (2013)
MUSK intron 3	Z	700	MUSK-13F	CTTCCATGCACTACAATGGGAAA	MUSK-13R	CTCTGAACATTGTGGATCCTCAA	Wang et al (2013)

Appendix 2, Table 1 continued.

Museum Specimen Primers (Mitochondrial DNA)					
Gene	**Forward**	**Sequence 3'-5'**	**Reverse**	**Sequence 3' -5'**	**Reference(s)**
Control Region	IntL2ND6	ATG TCC TWC RTG ACT AGC TT	IntH2CR	GCT TAG AAY AGG ACG CAW AC	This research
Control Region	MUS4 F	CCC CAT ACA TTA TGG TRC AG	MUS4 R	AATCCATCTGATACGTCGAG	This research
Control Region	MUS4 F	CCC CAT ACA TTA TGG TRC AG	MUS24 R	CCA TTC ACT GGT TAG GTG AC	This research
Control Region	MUS23 F	GYR TCT TTC TAR YCA CAT TTC	MUS4 R	AATCCATCTGATACGTCGAG	This research
Control Region	MUS18 F	CTC TGC TAT TGG TTC CCT TT	MUS18 R	AGT GTC AAR ATG ATW CCB SAT A	This research
Cytochrome oxidase 1	LCO1490	GGTCAACAAATCATAAAGATATTG G	MUS6 R	TTT ATY CGT GGG AAT GCT AT	This research
Cytochrome oxidase 1	MUS15 F	ACC TGT GAC CTT CAT CAA TC	MUS15 R	ATG AGR GGR ACT AGT CAG TT	This research
Cytochrome oxidase 1	MUS8 F	ATA GCA TTC CCA CGR ATA AA	MUS8 R	GAR AYR CCT GCT AGR TGA AG	This research
Cytochrome oxidase 1	MUS9 F	AAC TGA CTA GTY CCY CTC AT	MUS9 R	GCT AGR TGA AGG GAG AAGAT	This research
Cytochrome oxidase 1	MUS10 F	AGG AAC CYT ACY AGG AGA CG	MUS10 R	TTT ATY CGT GGG AAT GCT AT	This research
Cytochrome b	MUS1 F	ACT ACA CYG CAG AYR YYA C	MUS1 R	TCC TCA TGG GAG RAC ATA YC	This research
Cytochrome b	MUS2 F	TYG GAC AAA CCC TAG TAG A	MUS2 R	GGG GAT GAA TAT GAG TGA GA	This research
Cytochrome b	MUS3 F	TCT TCG CCC TAC ACT TTT TA	MUS2 R	GGG GAT GAA TAT GAG TGA GA	This research
Cytochrome b	MUS26 F	AAC TTG AAA CAC RGG AGT AA	MUS26 R	TCC YCC TCA AGY TCA TTC TA	This research
Cytochrome b	L15330AV	GGA CAA ATA TCA TTC TGA GG	H15551AV	GGG TGG AAT GGG ATT TTG TC	Lee et al., (2008)

Appendix 2, Table 1 continued.

Nuclear Genome Other Markers						
Gene	**Type**	**Forward**	**Sequence 3'-5'**	**Reverse**	**Sequence 3' -5'**	**Reference(s)**
ISSR 1	ISSR	(GACA)4	GACAGACAGACAGACA	N/A	N/A	Wink (2006)
ISSR 2	ISSR	(CA)10	CACACACACACACACACACA	N/A	N/A	Wink (2006)
ISSR 3	ISSR	(CAC)5	CACCACCACCACCAC	N/A	N/A	Wink (2006)
ISSR 4	ISSR	(GAA)5	GAAGAAGAAGAAGAA	N/A	N/A	Wink (2006)
ISSR 5	ISSR	(AC)8T	ACACACACACACACACT	N/A	N/A	Wink (2006)
AAT13a	Microsatellite	AAT13a-F	GGCCATTTGTTGCACAGTAG	AAT13a-R	GATCTGGAGCTGCTTTTATTA	Hughes (2000)
AAT14	Microsatellite	AAT14-F	GTGCCCCAGCAGTAATAATA	AAT14-R	CCATTGTACCAAAGTCACAGTA	Hughes (2000)
AAT6	Microsatellite	AAT6-F	CTGCTGGCCACGATTCTTC	AAT6-R	TCACGTACTTTTGTCCGTATTGTA	Hughes (2000)
ATT16	Microsatellite	ATT16-F	GGTCCCTTCCAAGTTGAATCAT	ATT16-R	AAGCCAAGCATGGAAAGAAAATA	Hughes (2000)
AAT16	Microsatellite	AAT16-F	CCATAGGTGGGTTGTATTA	AAT16-R	AGCAGAGCCAATATGAAGTAA	Hughes (2000)
AAT13b	Microsatellite	AAT13b-F	TCACCACCATTTCCCAACAG	AAT13b-R	ATGAGATTTACCTTCAGTTCT	Hughes (2000)

Appendix 3:

Mitogenome Announcements: The complete mitochondrial genome and phylogenetic position of the critically endangered Trinidad Piping Guan, *Pipile pipile* synonym *Aburria pipile* (Aves: Galliformes)

Amelia A. Grass, Charlotte Hosie & Ian L. McDowall

Pages 649-650 | Received 13 Jun 2016, Accepted 29 Jul 2016, Published online: 04 Sep 2016

Download citation http://dx.doi.org/10.1080/23802359.2016.1219622

Abstract

The complete mitochondrial genome of the Critically Endangered Trinidad Piping Guan, *Pipile pipile* (Jacquin 1784) synonym *Aburria pipile* was sequenced for the first time in this study. The genome is 16,665 bp in length with overall base compositions of 30.1, 23.7, 32.3 and 13.9% for A, T, C, and G, respectively. Structurally, the *P. pipile* mitogenome is comparable to that of other Galliformes, thereby demonstrating typical avian gene organization. The mitogenome was subsequently used to produce a revised phylogenetic placement of *P. pipile* within the Galliforme order, positioning the *Pipile* genus basal within the Cracidae family. It is further envisaged that this novel genomic data will contribute to a wider understanding of genetic relationships within the genus *Pipile* and the analysis of the evolutionary relationships of the Galliforme order in a wider avian context.

Keywords: Mitochondrial, *Pipile*, *Aburria*, Pawi, Cracidae, Galliformes

Mitogenome announcement

The Trinidad Piping Guan *Pipile pipile* is a Critically Endangered Cracid (Galliforme) endemic to the island of Trinidad, with an estimated population of less than 200 individuals (Hayes et al. 2009a). The population has been in ongoing decline largely due to anthropogenic pressures (Hayes et al. 2009a, 2009b). The IUCN and BirdLife International recognize Trinidad Piping Guan as a species of concern, classified as Critically Endangered, protected under Appendix I of the Convention on the International Trade in Endangered Species (BirdLife International 2015; CITES 2015; IUCN 2015). This mitochondrial genome represents the novel use of genetic data sourced from contemporary Trinidad Piping Guans from the wild on the island of Trinidad.

Samples were sourced from wild Trinidad Piping Guans (*P. pipile*) in the Matura Forest Reserve, Northern Range, Trinidad. *P. pipile* mitochondrial genomes were amplified from the feathers of live birds and liver tissue from a nonviable embryo.

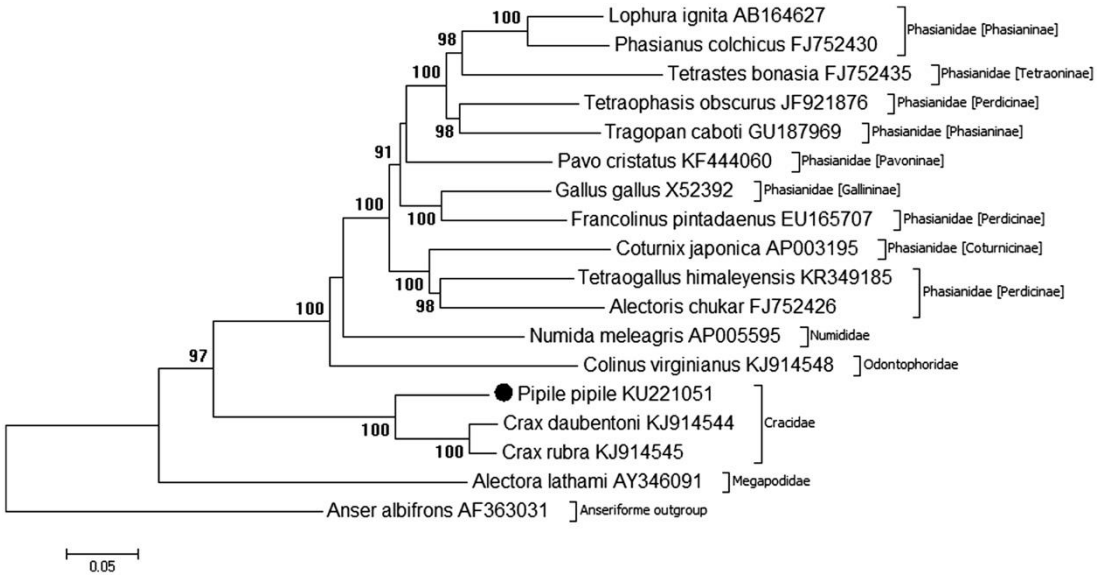
The complete mitochondrial genome of *P. pipile*, sequenced for both heavy and light strands, was 16,665 bp ±2 bp in length (Genbank: KU221051, KU221052, KU221053), 13 protein-coding genes, 2 rRNAs, and 22 tRNAs were identified demonstrating organization consistent with Galliforme mito-genomes. Base composition of the three complete genomes demonstrate an A + T bias (53.8%), where

overall base composition of the genome were as follows; A (30.1%), T (23.7%), C (32.3%), and G (13.9%).

All compositional features of the genome were consistent between samples, with length variation occurring due to indels in the hypervariable sections of the control region. The majority of genes are encoded on the heavy strand, however, the ND6 gene and eight tRNA genes are light strand encoded, as in other Galliformes. The majority of genes are contiguous within the sequence, however, gene overlaps occur in four locations; $tRNA^{Gln}$ and $tRNA^{Met}$, $tRNA^{Cys}$ and $tRNA^{Tyr}$, ATP8 and ATP6, and ND4L and ND4. Additionally, noncoding intergenic spacers of variable length (1–13nts) occur between some genes. The additional nontranslated nucleotide at position 174 of the ND3 gene was observed consistent with the observation of this additional nucleotide in other Galliformes (Mindell et al. 1998).

Phylogenetic analysis of the mitochondrial genome of *P. pipile* and other Galliforme species is shown in Figure 1, constructed using MEGA7.0 (http://www.megasoftware.net/) (Kumar et al. 2016). The *Pipile* genus resolves basally within the Cracidae order of the Galliforme phylogenetic tree, with a high degree of bootstrap support (100%). Position of the Cracidae family within the Galliforme tree is consistent with initial mitogenome analysis (*Crax* sp.) in previous research (Meiklejohn et al. 2014). Basal divergence of the *Pipile* genus in relation to the *Crax* is consistent with individual mitochondrial gene analysis illustrated by Crowe et al. (2006), however, full resolution of the Cracidae family is restricted by a lack of complete mitochondrial genome information for Cracid species. The complete mitogenome of the Trinidad Piping Guan, *Pipile pipile*, is used for the first time in phylogenetic analysis, providing essential molecular data and evolutionary information for further analysis of both the *Pipile* genus and the Cracidae family.

Figure 1. Phylogenetic relationships of *Pipile pipile* within the Galliforme order, inferred from whole mitochondrial genomes. Phylogenetic tree constructed using complete mitogenomes from Genbank in MEGA 7.0 using the Maximum-Likelihood method with 1000 bootstrap replicates. Genbank accession numbers are given after the species name, bootstrap support is given at each node.



Acknowledgements

The authors would like to acknowledge Prof. J. Cooper and the Pawi Study Group for their assistance in acquisition of verified reference samples from wild Trinidad Piping Guans.

Disclosure statement

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of this paper.

References

Birdlife International. 2015. BirdLife International Species Information Pipile pipile [Internet]. [cited 2015 Oct 10]. Available from: http://www.birdlife.org/datazone/speciesfactsheet.php?id =87.

CITES. 2015. CITES Species Database [Internet]. [cited 2015 Oct 10]. Available from: http://speciesplus.net/#/taxon_concepts/7109/legal.

Crowe TM, Bowie RCK, Bloomer P, Mandiwana TG, Hedderson TAJ, Randi E, Pereira SL, Wakeling J. 2006. Phylogenetics, biogeography and classification of, and character evolution in, gamebirds (Aves: Galliformes): effects of character exclusion, data partitioning and missing data. Cladistics. 22:495–532.

Hayes F, Sanasie B, Samad I. 2009a. Status and conservation of the critically endangered Trinidad Piping-Guan Aburria pipile. Endanger Species Res. 7:77–84. [Internet]. [cited 2012 May 16]. Available from: http://www.int-res.com/abstracts/esr/v7/n1/p77-84/.

Hayes F, Shameerudeen C, Sanasie B, Hayes B, Ramjohn C, Lucas F. 2009b. Ecology and behaviour of the critically endangered Trinidad Piping-Guan Aburria pipile. Endanger Species Res. 6:223–229.

IUCN. 2015. The IUCN Red List of Threatened Species. Red List [Internet]. [cited 2015 Oct 10]. Available from: http://www.iucnredlist.org/details/22678401/0.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870–1874.

Meiklejohn KA, Danielson MJ, Faircloth BC, Glenn TC, Braun EL, Kimball RT. 2014. Incongruence among different mitochondrial regions: a case study using complete mitogenomes. Mol Phylogenet Evol. 78:314–323.

Mindell DP, Sorenson MD, Dimcheff DE. 1998. An extra nucleotide is not translated in mitochondrial ND3 of some birds and turtles. Mol Biol Evol. 15:1568–1571.

Appendix 4:

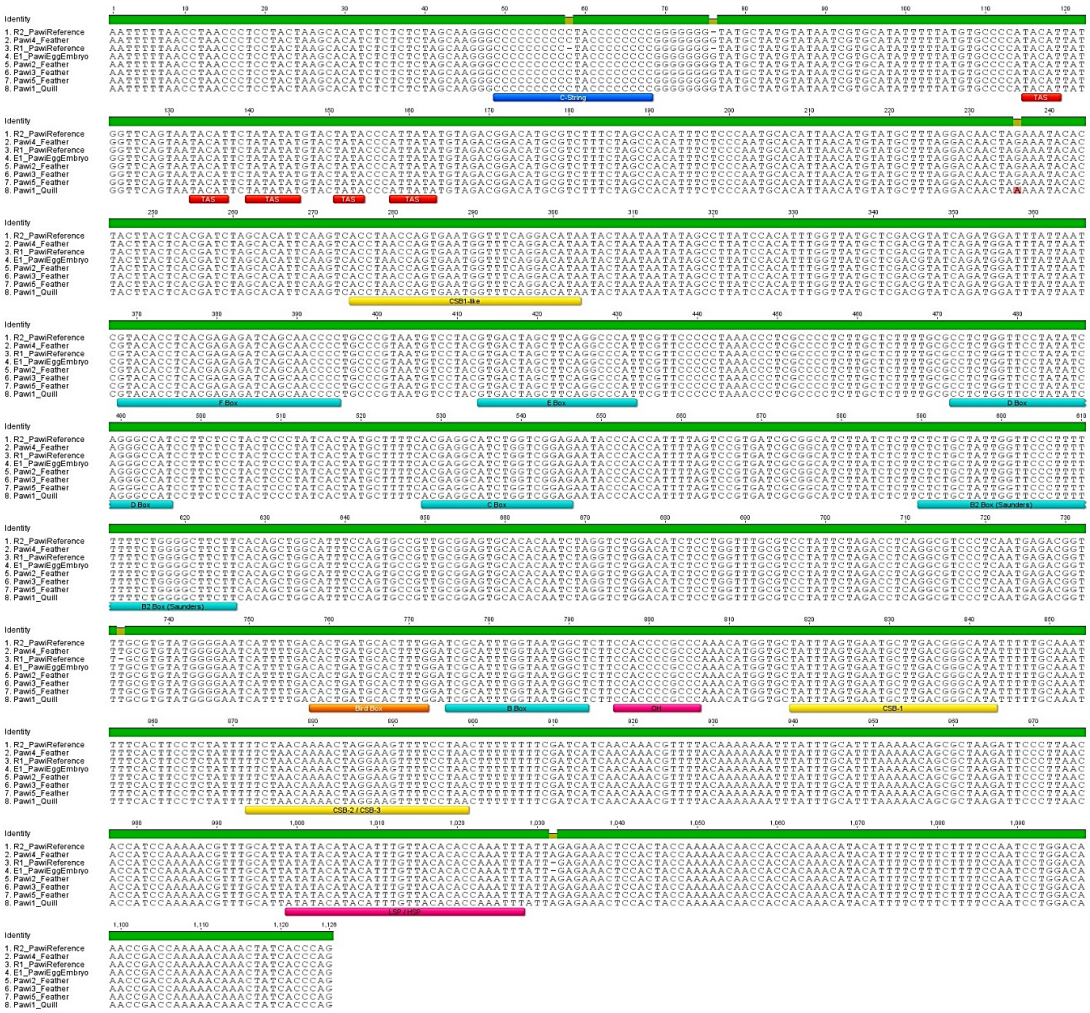


Figure 1: Alignment of the mitochondrial Control Region sequences of eight Trinidad Piping Guan individuals, using Clustal W in Geneious R8.0. Sequence homology indicated by green bar above the sequence, variations highlighted within the sequence* and Conserved sequence features identified below the sequence.
* inserts/deletions (indels) indicated by '-', nucleotide substitutions indicated by highlighting

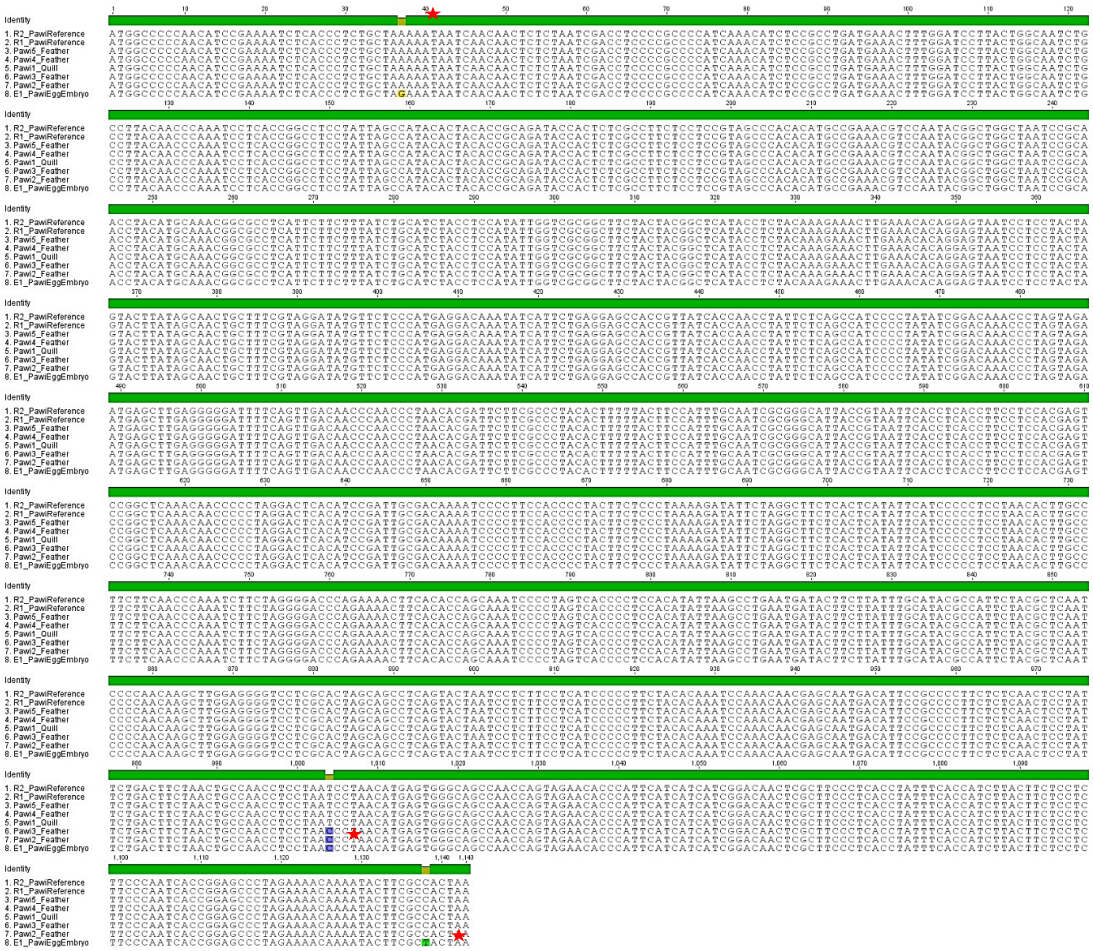


Figure 2: Alignment of the mitochondrial Cytochrome *b* gene sequences of eight Trinidad Piping Guan individuals, using Clustal W in Geneious R8.0. Sequence homology indicated by green bar above the sequence and variations highlighted within the sequence.

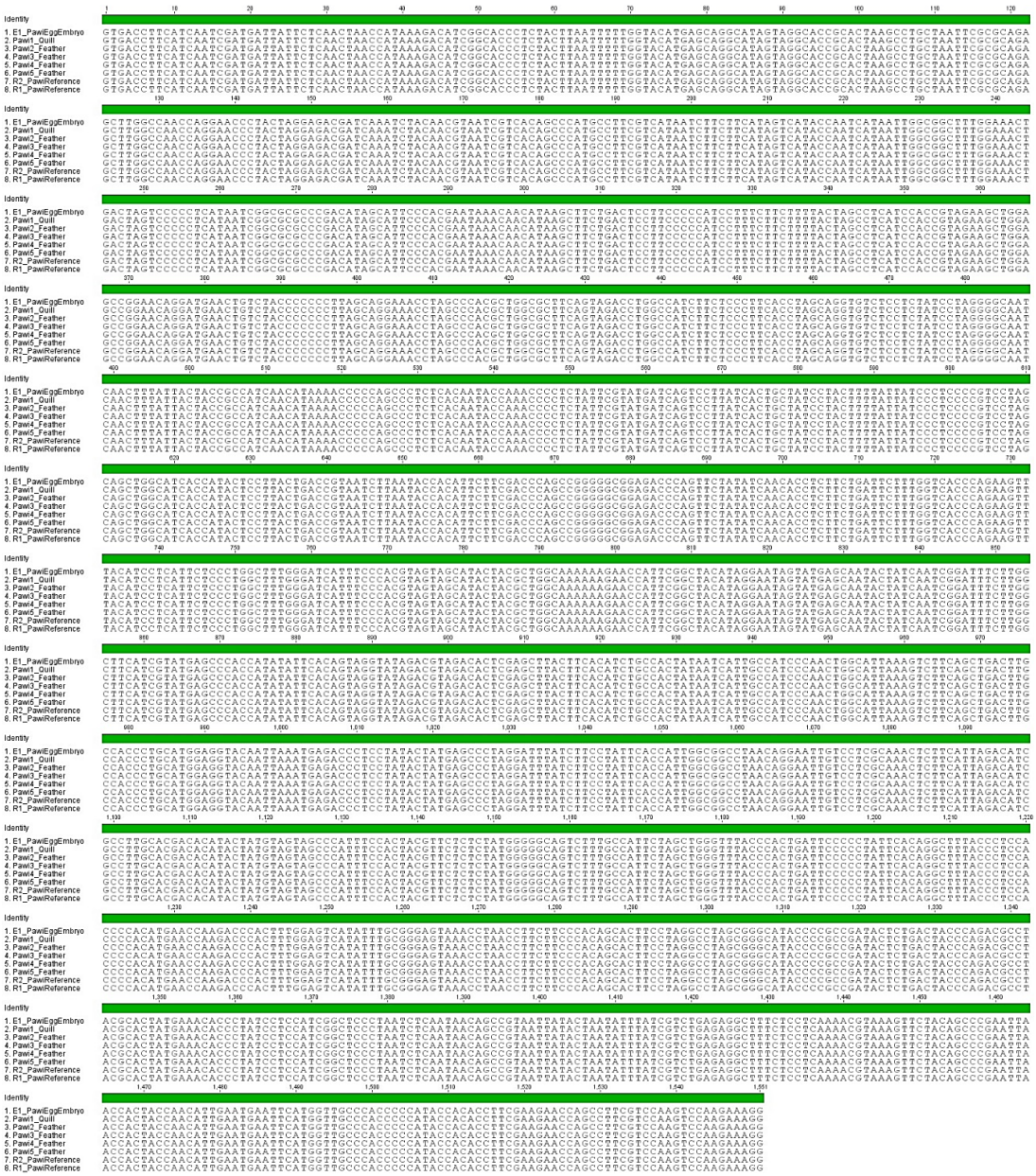


Figure 3: Alignment of the mitochondrial Cytochrome c oxidase subunit 1 gene sequences of eight Trinidad Piping Guan individuals, using Clustal W in Geneious R8.0. Sequence homology indicated by green bar above the sequence and variations highlighted within the sequence.

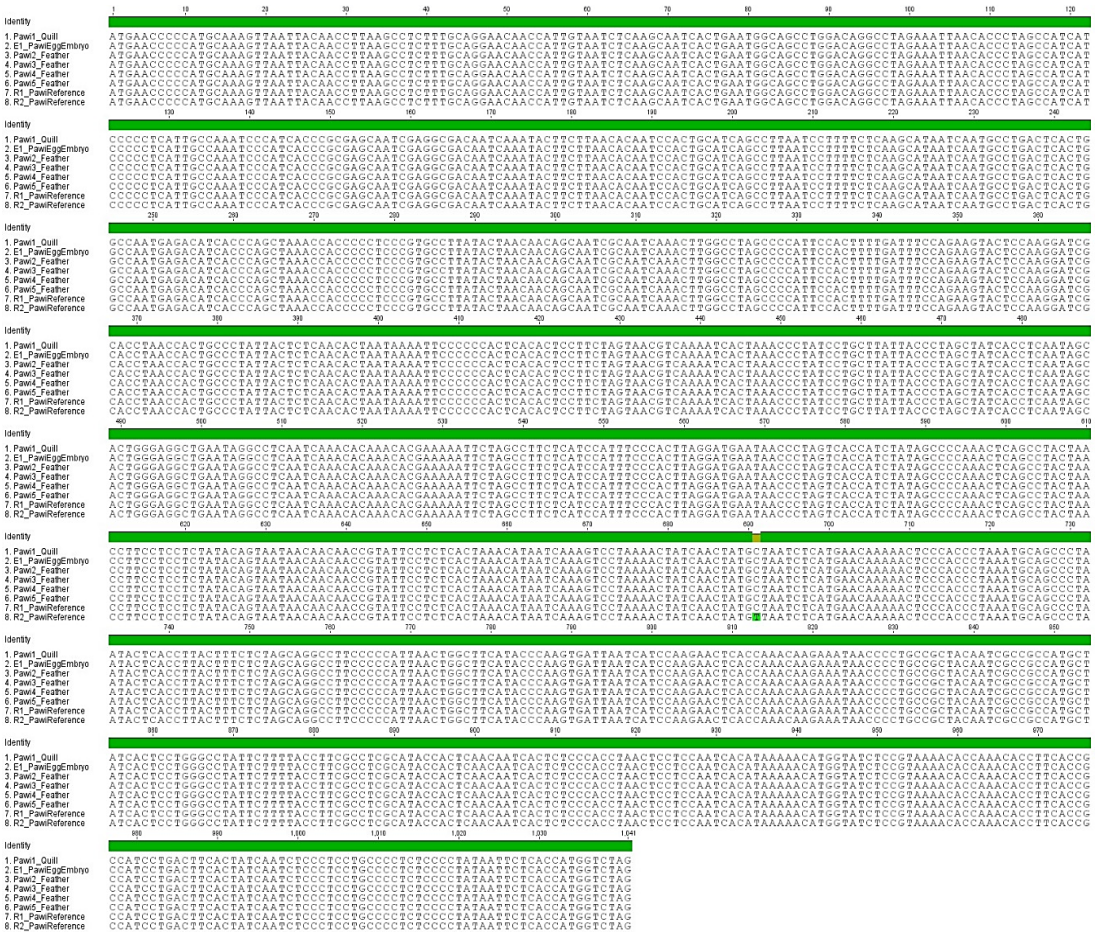


Figure 4: Alignment of the mitochondrial NADH dehydrogenase subunit 2 gene sequences of eight Trinidad Piping Guan individuals, using Clustal W in Geneious R8.0. Sequence homology indicated by green bar above the sequence and variations highlighted within the sequence.

Figure 5: Alignment of the nuclear DNA Beta-fibrinogen intron 7 sequences of four Trinidad Piping Guan individuals and nine UK captive Piping Guans, using Clustal W in Geneious R8.0. Sequence homology indicated by green bar above the sequence and variations highlighted within the sequence.

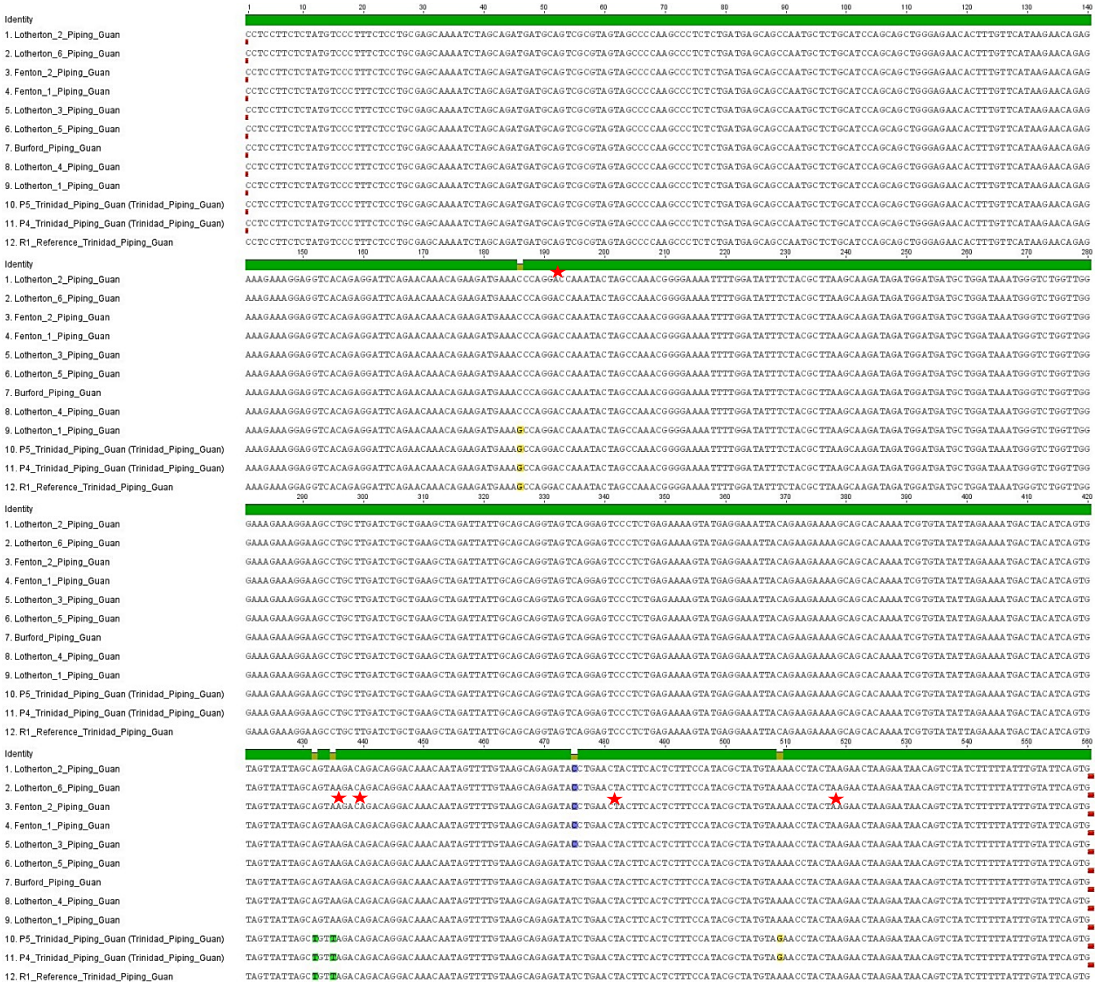


Figure 6: Alignment of the nuclear DNA SERPIN B14 intron c sequences of three Trinidad Piping Guan individuals and nine UK captive Piping Guans, using Clustal W in Geneious R8.0. Sequence homology indicated by green bar above the sequence and variations highlighted within the sequence.

Appendix 5: Table 1: Whole mitochondrial genome sequences used in the phylogenetic analysis of the genetic relationships between Galliformes.

Species	Common Name	Family	Order	Accession Number	Reference *only published to NCBI/GenBank database
Pipile pipile	Trinidad Piping Guan	Cracidae	Galliformes	N/A	This research
Crax rubra	Great Curassow	Cracidae	Galliformes	KJ914545	Meiklejohn et al., 2014
Crax daubentoni	Yellow-knobbed Curassow	Cracidae	Galliformes	KJ914544	Meiklejohn et al., 2014
Alectura lathami	Australian Brushturkey	Megapodidae	Galliformes	AY346091	Slack, Delsuc, Mclenachan, Arnason, & Penny, 2007
Numida meleagris	Helmeted Guineafowl	Numididae	Galliformes	AP005595	Nishibori, Hayashi, & Yasue, 2004
Acryllium vultuninum	Vulturine Guineafowl	Numididae	Galliformes	FJ752436	Shen et al., 2010
Colinus virginianus	Northern Bobwhite	Odontophoridae	Galliformes	KJ914548	Meiklejohn et al., 2014
Rhynchortyx cinctus	Tawny-faced Quail	Odontophoridae	Galliformes	KJ914547	Meiklejohn et al., 2014
Phasianus versicolor	Green pheasant	Phasianidae	Galliformes	AB164626	Kato et al 2004*
Phasianus colchicus	Ring-necked pheasant	Phasianidae	Galliformes	FJ752430	Shen et al., 2010
Gallus gallus	Chicken	Phasianidae	Galliformes	X52392	Desjardin & Morais 1990
Gallus varius	Green junglefowl	Phasianidae	Galliformes	AP003324	Nishibori, Shimogiri, Hayashi, & Yasue, 2005
Gallus lafayetii	Ceylon junglefowl	Phasianidae	Galliformes	AP003325	Nishibori, Shimogiri, Hayashi, & Yasue, 2005
Gallus sonneratii	Gray junglefowl	Phasianidae	Galliformes	AP003320	Nishibori, Shimogiri, Hayashi, & Yasue, 2005
Lophura ignita	Crested Fireback	Phasianidae	Galliformes	AB164627	Kato, Nishibori & Yasue 2004*
Lophura nycthemra	Silver pheasant	Phasianidae	Galliformes	EU417810	Shen, Shi, Sun, & Zhang, 2009
Lophura swinhoii	Swinhoe's pheasant	Phasianidae	Galliformes	KF218954	Jiang, Wang, Peng, Peng, & Zou, 2014
Pavo cristatus	Indian Peafowl	Phasianidae	Galliformes	KF444060	Zhou, Sha, Irwin, & Zhang, 2015
Pavo muticus	Green peafowl	Phasianidae	Galliformes	EU417811	Shen et al., 2009
Coturnix japonica	Japanese quail	Phasianidae	Galliformes	AP003195	Nishibori et al., 2001
Coturnix chinensis	King quail	Phasianidae	Galliformes	AB073301	Nishibori, Tsudzuki, Hayashi, Yamamoto, & Yasue, 2002
Meleagris gallopavo	Turkey	Phasianidae	Galliformes	EF153719	Guan et al., 2009
Francolinus pintadeanus	Chinese Francolin	Phasianidae	Galliformes	EU165707	Shen et al., 2009
Alectoris chukar	Chukar partridge	Phasianidae	Galliformes	FJ752426	Shen et al., 2010
Tetraogallus himaleyensis	Himalayan snowcock	Phasianidae	Galliformes	KR349185	An et al., 2015*
Tetraogallus tibetanus	Tibetan snowcock	Phasianidae	Galliformes	KF027439	Zhang, Zhou & Shen, 2013*
Polyplectron germaini	Germain's peacock-pheasant	Phasianidae	Galliformes	KF422893	Omeire, Abdin, Brookes, & Miranda, 2015
Polyplectron napoleonis	Palawan peacock-pheasant	Phasianidae	Galliformes	KJ939353	Quach, Brooks & Miranda, 2014*
Bonasa bonasia	Hazel grouse	Phasianidae	Galliformes	FJ752435	Shen et al., 2010
Tetrastes sewerzowi	Chinese grouse	Phasianidae	Galliformes	KJ997914	Li, Huang & Lei, 2014

Appendix 5, Table 1 contd.

Species	Common Name	Family	Order	Accession Number	Reference *only published to NCBI/GenBank database
Ptilopachus petrosus	Stone partridge	Phasianidae	Galliformes	KJ914543	Meiklejohn et al., 2014
Arborophila ardens	Hainan partridge	Phasianidae	Galliformes	KJ716444	Li, Huang & Lei, 2014*
Arborophila brunneopectus	Bar-backed partridge	Phasianidae	Galliformes	KC352730	Yan et al., 2013*
Arborophila rufipectus	Sichuan partridge	Phasianidae	Galliformes	FJ194942	He et al., 2009
Arborophila gingica	White necklaced partridge	Phasianidae	Galliformes	FJ752425	Shen et al., 2010
Arborophila rufogularis	Rufous throated partridge	Phasianidae	Galliformes	FJ752424	Shen et al., 2010
Caloperdix oculeus	Ferruginous partridge	Phasianidae	Galliformes	KJ914546	Meiklejohn et al., 2014
Syrmaticus reevesii	Reeve's pheasant	Phasianidae	Galliformes	AB164623	Kato, Nishibori & Yasue, 2004*
Syrmaticus ellioti	Elliot's pheasant	Phasianidae	Galliformes	AB164624	Kato, Nishibori & Yasue, 2004*
Syrmaticus humiae	Hume's pheasant	Phasianidae	Galliformes	AB164625	Kato, Nishibori & Yasue, 2004*
Bambusicola thoracia	Chinese bamboo partridge	Phasianidae	Galliformes	EU165706	Shen et al., 2009
Bambusicola fytchii	Mountain bamboo partridge	Phasianidae	Galliformes	FJ752423	Shen et al., 2010
Tragopan temminckii	Temminck's partridge	Phasianidae	Galliformes	FJ752427	Shen et al., 2010
Tragopan caboti	Cabot's tragopan	Phasianidae	Galliformes	GU187969	Kan et al., 2010
Perdix dauurica	Daurian partridge	Phasianidae	Galliformes	FJ752431	Shen et al., 2010
Perdix hodgsoniae	Tibetan partridge	Phasianidae	Galliformes	KF027440	Zhang, Zhou & Shen, 2013*
Pucrasia macrolopha	Koklass pheasant	Phasianidae	Galliformes	FJ752429	Shen et al., 2010
Lophophorus sclateri	Sclater's monal	Phasianidae	Galliformes	FJ752432	Shen et al., 2010
Lophophorus lhuysii	Chinese monal	Phasianidae	Galliformes	GQ871234	Ma, Zhang, Yue, & Ran, 2010
Chrysolophus pictus	Golden pheasant	Phasianidae	Galliformes	FJ752433	Shen et al., 2010
Chrysolophus amherstiae	Lady Amherst's pheasant	Phasianidae	Galliformes	FJ752434	Shen et al., 2010
Tetraophasis szechenyii	Szecheny's monal-partridge	Phasianidae	Galliformes	FJ799728	Meng, He & Yue, 2009*
Tetraophasis obscurus	Verreaux' monal partridge	Phasianidae	Galliformes	JF921876	Ma & Ran, 2011*
Crossoptilon crossoptilon	White eared pheasant	Phasianidae	Galliformes	HQ891119	Zhao & Zou, 2011*
Crossoptilon auritum	Blue eared pheasant	Phasianidae	Galliformes	JF937589	Li & Kan, 2011*
Crossoptilon harmani	Tibetan eared pheasant	Phasianidae	Galliformes	KP259806	Li, Huang, & Lei, 2015
Crossoptilon mantchuricum	Brown eared pheasant	Phasianidae	Galliformes	KP259807	Li, Huang, & Lei, 2015
Ithaginis cruentus	Blood pheasant	Phasianidae	Galliformes	JF921875	Zeng et al., 2013
Lyrurus tetrix	Black Grouse	Phasianidae	Galliformes	KF955638	Li, Zhu, Ding, Bai, & Cui, 2016

Appendix 5, Table 2: Whole mitochondrial genome sequences used in the phylogenetic analysis of the genetic relationships between Avian orders, in addition to those of the Galliforme order and including reptile outgroup sequences.

Species	Common Name	Family	Order	Accession Number	Reference *only published to NCBI/GenBank database
Eudromia elegans	Elegant Crested Tinamou	Tinamidae	Tinamiformes	AF338710	Haddrath & Baker, 2001
Tinamus major	Great Tinamou	Tinamidae	Tinamiformes	AF338707	Haddrath & Baker, 2001
Anser albifrons	Greater White-fronted Goose	Anatidae	Anseriformes	AF363031	Slack et al., 2003
Aythya americana	Redhead Duck	Anatidae	Anseriformes	AF090337	Mindell et al., 1998
Cairina moschata	Muscovy Duck	Anatidae	Anseriformes	EU755254	Tu et al., 2014
Cygnus columbianus bewickii	Bewick's Swan	Anatidae	Anseriformes	JQ282800	Lee, Ryu, Kang, & Hwang, 2012
Casuarius casuarius	Double-Wattled Cassowary	Casuariidae	Casuariiformes	AF338713	Haddrath & Baker, 2001
Dromaius novaehollandiae	Emu	Dromaiidae	Casuariiformes	AF338711	Haddrath & Baker, 2001
Emeus crassus (Extinct)	Eastern Moa	Emeidae	Dinornithiformes	AY016015	Cooper et al., 2001
Rhea americana	Rhea	Rheidae	Rheiformes	Y16884	Harlid, Janke, & Arnason, 1998
Pterocnemia pannata	Lesser Rhea	Rheidae	Rheiformes	AF338709	Haddrath & Baker, 2001
Struthio camelus	Ostrich	Struthionidae	Struthioniformes	Y12025	Härlid, Janke, & Arnason, 1997
Ciconia boyciana	Oriental White Stork	Ciconiidae	Ciconiiformes	AB026193	Yamamoto, 2000*
Ciconia ciconia	White stork	Ciconiidae	Ciconiiformes	AB026818	Yamamoto, 2000*
Cathartes aura	Turkey vulture	Cathartidae	Ciconiiformes	AY463690	Slack et al., 2007
Apteryx haastii	Great-spotted Kiwi	Apterygidae	Apterygiformes	AF338708	Haddrath & Baker, 2001
Eudyptula minor	Little Blue Penguin	Spheniscidae	Sphenisciformes	AF362763	Slack et al., 2003
Apus apus	Common Swift	Apodidae	Apodiformes	AM237310	Slack et al., 2007
Corvus frugilegus	Rook	Corvidae	Passeriformes	Y18522	Harlid & Arnason, 1999
Vidua chalybeate	Village Indigobird	Passeroidea	Passeriformes	AF090341	Mindell et al., 1999
Smithornis sharpei	Grey-headed Broadbill	Neognathae	Passeriformes	AF090340	Mindell et al., 1999
Nipponia nippon	Crested ibis	Threskiornithidae	Pelicaniformes	AB104902	Kodama, et al., 2003*
Platalea minor	Black-faced spoonbill	Threskiornithidae	Pelicaniformes	EF455490	Lee et al., 2007*
Pelecanus conspicillatus	Australian pelican	Pelecanidae	Pelecaniformes	DQ780883	Gibb, Kardailsky, Kimball, Braun, & Penny, 2007
Dupetor flavicollis	Black bittern	Ardeidae	Pelecaniformes	KJ643141	Wang et al., 2014*
Gavia pacifica	Pacific Loon	Gaviidae	Gaviformes	AP009190	Watanabe et al., 2006
Podiceps cristatus	Great Crested grebe	Podicipedidae	Podicipediformes	AP009194	Watanabe et al., 2006
Haematopus ater	Blackish oystercatcher	Haematopodidae	Charadriiformes	AY074886	Paton et al., 2002

Appendix 5, Table 2 contd.

Species	Common Name	Family	Order	Accession Number	Reference *only published to NCBI/GenBank database
Larus brunnicephalus	Brown headed gull	Laridae	Charadriiformes	JX155863	Yang, Wang, Huang & Xiao, 2012
Chroicocephalus ridibundus	Black headed gull	Laridae	Charadriiformes	KM577662	Dong, Zhou, Li & Zhao, 2014*
Pterodroma brevirostris	Kerguelen petrel	Procellariidae	Procellariformes	AY158678	Slack et al., 2006
Phoenicopterus ruber	Greater flamingo	Phoenicopteridae	Phoenicopteriformes	EF532932	Morgan-Richards et al., 2008
Dryocopus pileatus	Pilated woodpecker	Picidae	Piciformes	DQ780879	Gibb, Kardailsky, Kimball, Braun, & Penny, 2007
Rhynochetos jubatus	Kagu	Rhynochetidae	Gruiformes	EF532933	Morgan-Richards et al., 2008
Gallinula chloropus	Common moorhen	Rallidae	Gruiformes	HQ896036	Kan & Li, 2011*
Otis tarda	Great bustard	Otididae	Gruiformes	NC_014046	Yang, Wu, Yan, Su, & Yang, 2010
Geococcyx californianus	Greater roadrunner	Neomorphidae	Cuculiformes	EU410488	Pratt et al., 2009
Tyto alba	Barn owl	Tytonidae	Strigiformes	EU410491	Pratt et al., 2009
Columba livia	Rock pigeon	Columbidae	Columbiformes	GQ240309	Tsai et al., 2009 *
Zenaida auriculata	Eared dove	Columbidae	Columbiformes	HM640211	Pacheco et al., 2011
Aratinga pertinax	Brazilian brown throated parakeet	Psittacidae	Psittaciformes	HM640208	Pacheco et al., 2011
Ara glaucogularis	Blue throated macaw	Psittacidae	Psittaciformes	JQ782215	Urantowka, 2012*
Bucorvus leadbeateri	Southern ground hornbill	Bucorvidae	Bucerotiformes	HM640209	Pacheco et al., 2011
Ceryle rudis	Pied Kingfisher	Cerylidae	Coraciiformes	KJ461938	Zou, Bi, Huang & Jing, 2014*
Amazilia versicolor	Versicoloured emerald hummingbird	Trochilidae	Trochiliformes	KF624601	Prosdocimi et al., 2015*
Aquila chrysaetos	Golden eagle	Accipitridae	Falconiformes	KF905228	Doyle et al., 2014
Sagittarius serpentarius	Secretarybird	Sagittariidae	Falconiformes	KF961184	Mahmood, McLenachan, Gibb & Penny, 2014
Caprimulugus indicus	Indian jungle nightjar	Caprimulgidae	Caprimulgiformes	NC_025773	Zhou & Zhao, 2014
Outgroups					
Alligator mississippiensis	American Alligator	Alligatorinae	Alligatoridae	Y13113	Janke & Arnason, 1997
Iguana iguana	Iguana	Iguaninae	Iguanidae	AJ278511	Janke, Erpenbeck, Nilsson, & Arnason, 2001
Chelonia mydas	Green Turtle	Chelonia	Cheloniidae	AB012104	Kumazawa & Nishida, 1995

Appendix 6: Individual Gene Phylogenetic Inferences

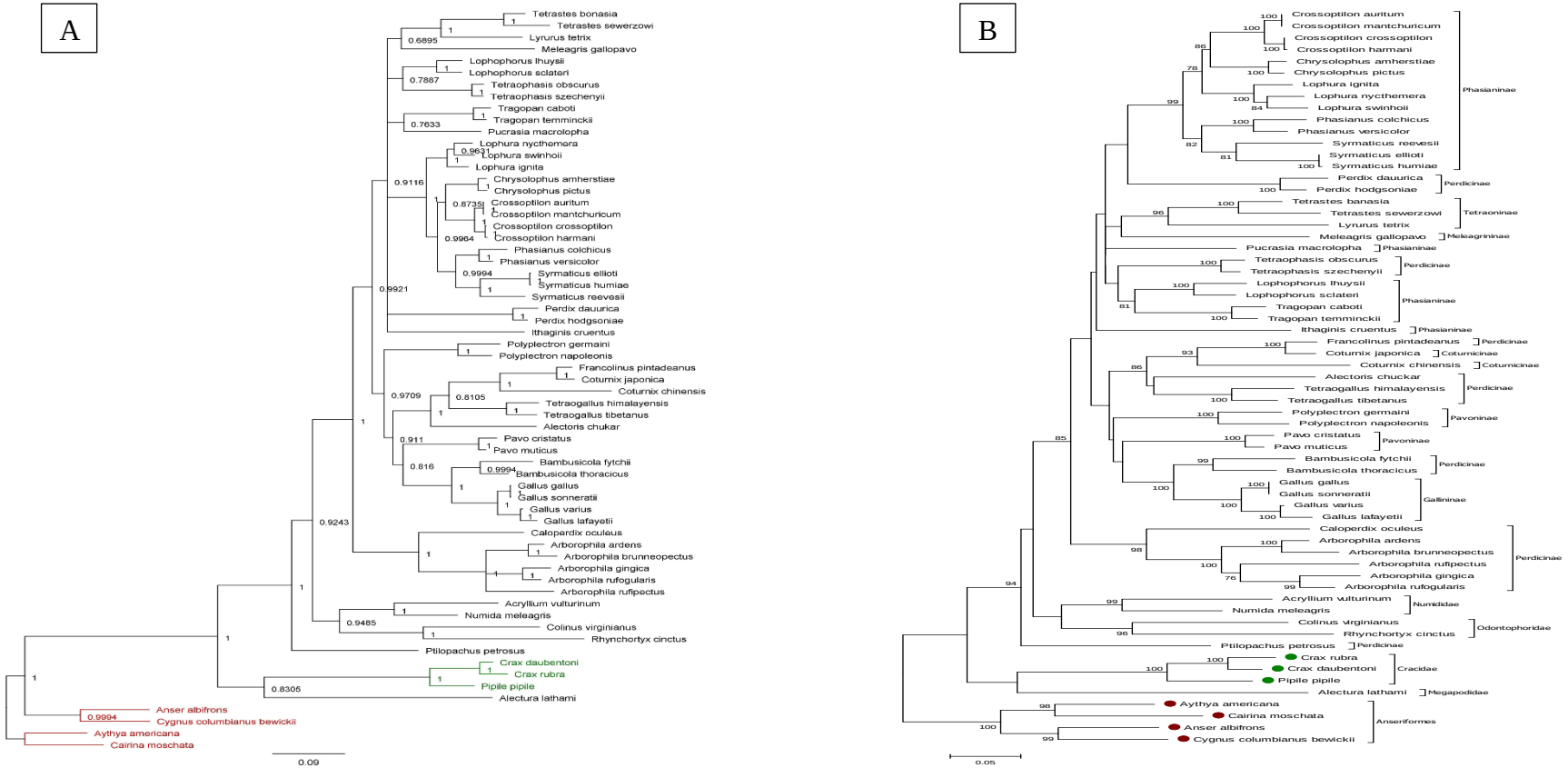


Figure 5.16: Phylogenetic relationships among Galliformes based on analysis of Cytochrome b gene data (1143bp) from 59 genera of the Galliforme order including the novel *Pipile pipile*, and 4 Anseriforme genera as a putative outgroup. Topology and branch lengths are based on Maximum Likelihood General Time Reversible substitution models under non-uniform evolutionary rates with Gamma distribution and Invariable sites (GTR+G+I), through independent analysis of Bayesian inference in MrBayes (A) and Maximum Likelihood inference in MEGA with Bootstrap statistical support (B). Bayesian analysis used MCMC with 1,100,000 generations sampled every 1000 generations, discarding the first 10% of the run as burn-in, with posterior probabilities to support node topologies. Maximum Likelihood analysis in MEGA with 500 bootstrap replicates for node statistical support (numbers at nodes in %), only nodes with >75% statistical support displayed.

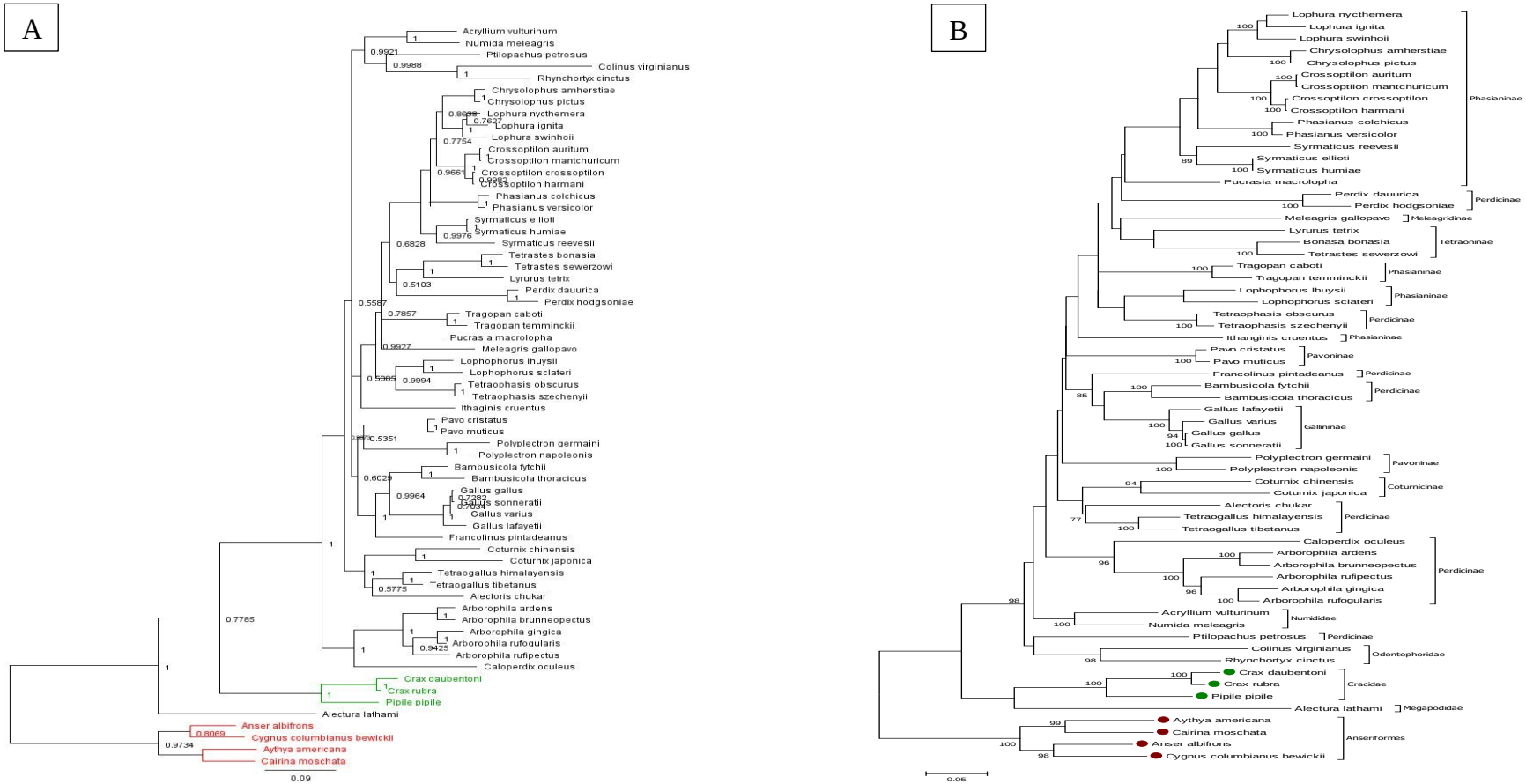


Figure 5.17: Phylogenetic relationships among Galliformes based on analysis of Cytochrome c oxidase subunit 1 (COI) gene data (1551bp) from 59 genera of the Galliforme order including the novel *Pipile pipile*, and 4 Anseriforme genera as a putative outgroup. Topology and branch lengths are based on Maximum Likelihood General Time Reversible substitution models under non-uniform evolutionary rates with Gamma distribution and Invariable sites (GTR+G+I), through independent analysis of Bayesian inference in MrBayes (A) and Maximum Likelihood inference in MEGA with Bootstrap statistical support (B). Bayesian analysis used MCMC with 1,100,000 generations sampled every 1000 generations, discarding the first 10% of the run as burn-in, with posterior probabilities to support node topologies. Maximum Likelihoood analysis in MEGA with 500 bootstrap replicates for node statistical support (numbers at nodes in %), only nodes with >75% statistical support displayed.

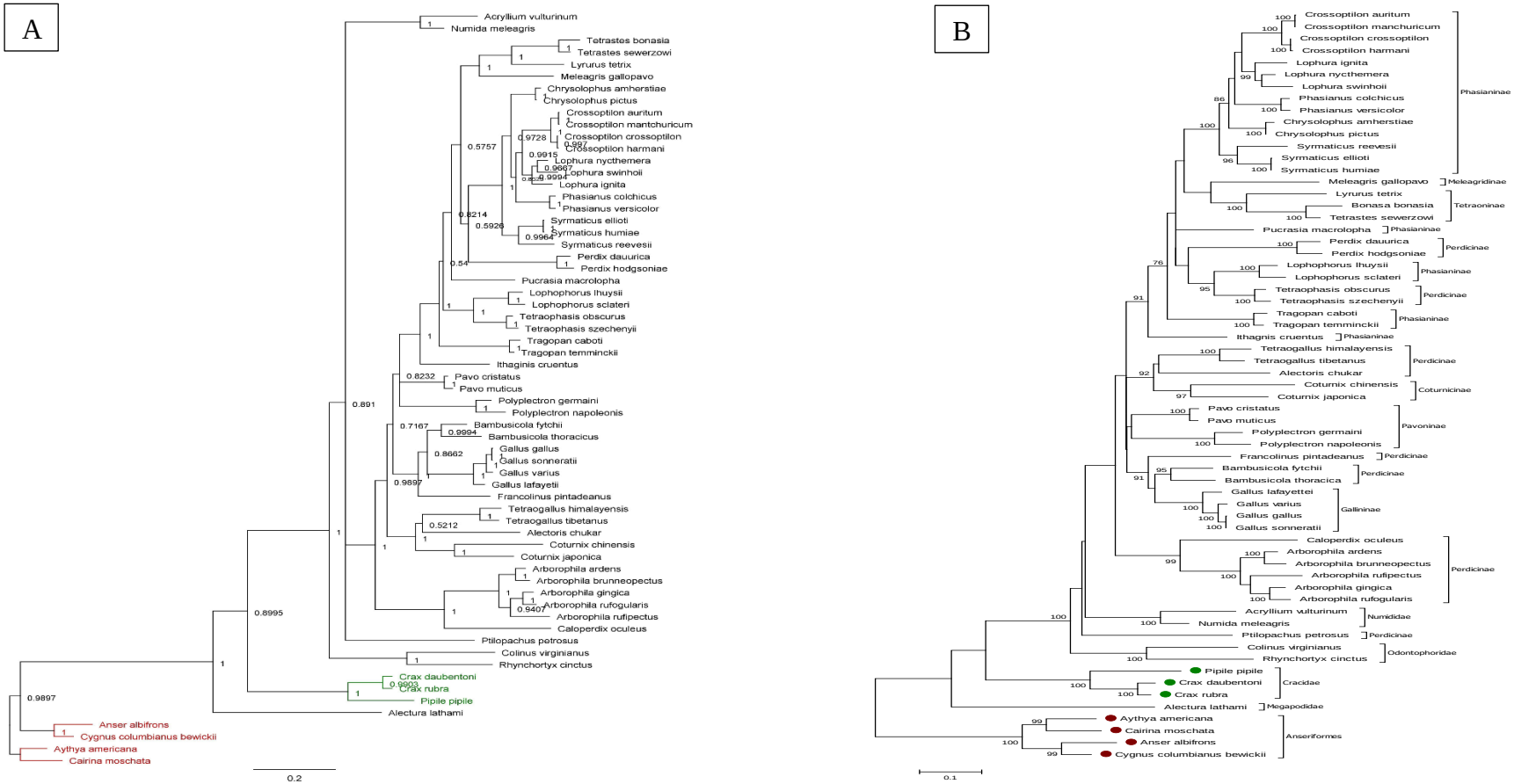


Figure 5.18: Phylogenetic relationships among Galliformes based on analysis of NADH dehydrogenase subunit 2 (ND2) gene data (1038bp) from 59 genera of the Galliforme order including the novel *Pipile pipile*, and 4 Anseriforme genera as a putative outgroup. Topology and branch lengths are based on Maximum Likelihood General Time Reversible substitution models under non-uniform evolutionary rates with Gamma distribution and Invariable sites (GTR+G+I), through independent analysis of Bayesian inference in MrBayes (A) and Maximum Likelihood inference in MEGA with Bootstrap statistical support (B). Bayesian analysis used MCMC with 1,100,000 generations sampled every 1000 generations, discarding the first 10% of the run as burn-in, with posterior probabilities to support node topologies. Maximum Likelihood analysis in MEGA ith 500 bootstrap replicates for node statistical support (numbers at nodes in %), only nodes with >75% statistical support displayed.

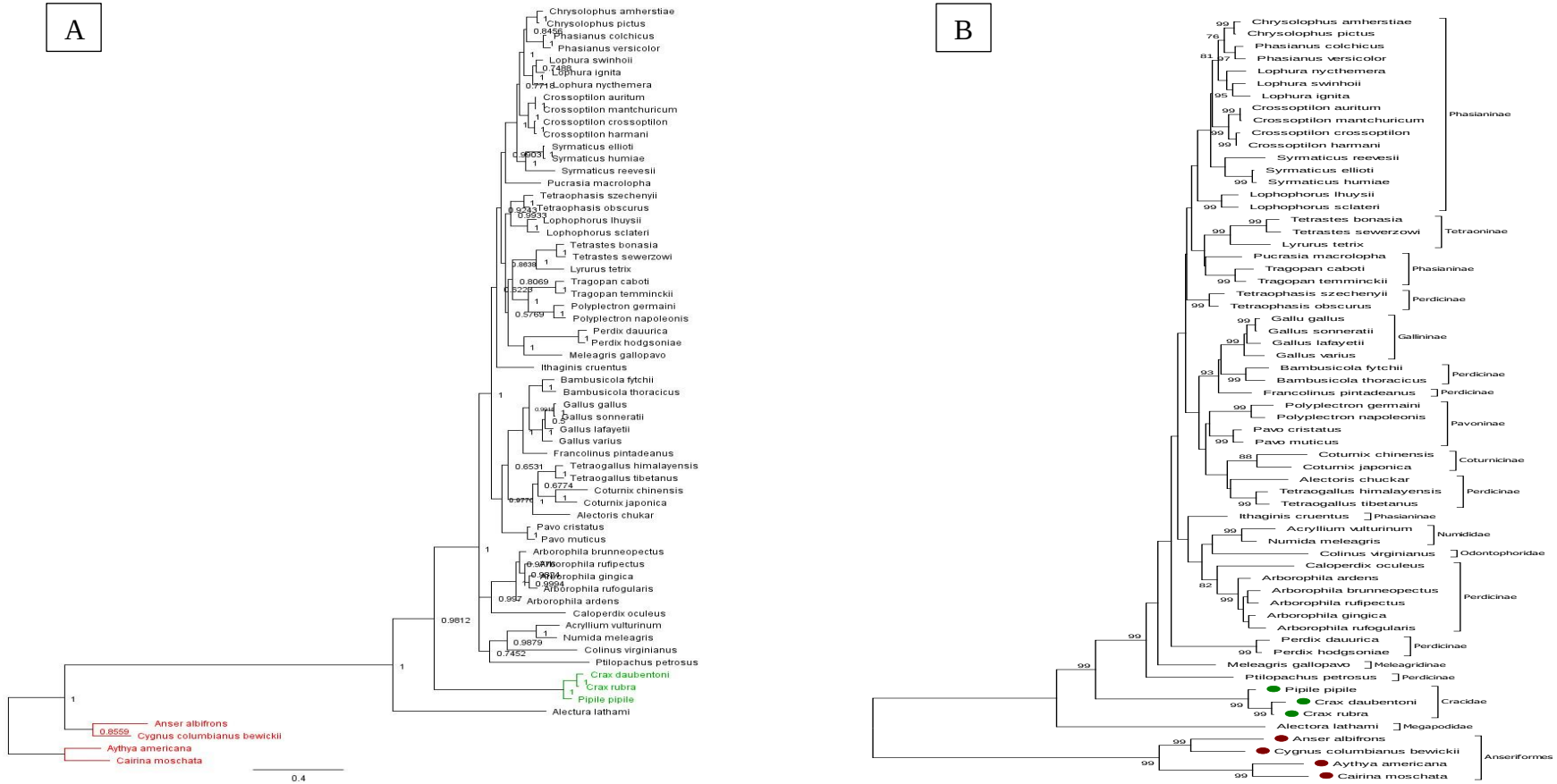


Figure 5.19: Phylogenetic relationships among Galliformes based on analysis of Control Region gene data (1429bp) from 59 genera of the Galliforme order including the novel *Pipile pipile*, and 4 Anseriforme genera as a putative outgroup. Topology and branch lengths are based on Maximum Likelihood General Time Reversible substitution models under non-uniform evolutionary rates with Gamma distribution and Invariable sites (GTR+G+I), through independent analysis of Bayesian inference in MrBayes with posterior probabilities (A) and Maximum Likelihood inference in MEGA with Bootstrap statistical support (B). Bayesian analysis used MCMC with 1,100,000 generations sampled every 1000 generations, discarding the first 10% of the run as burn-in, with posterior probabilities to support node topologies (<1). Maximum Likelihood analysis in MEGA used 500 bootstrap replicates for node statistical support (numbers at nodes in %), only nodes with >75% statistical support displayed.

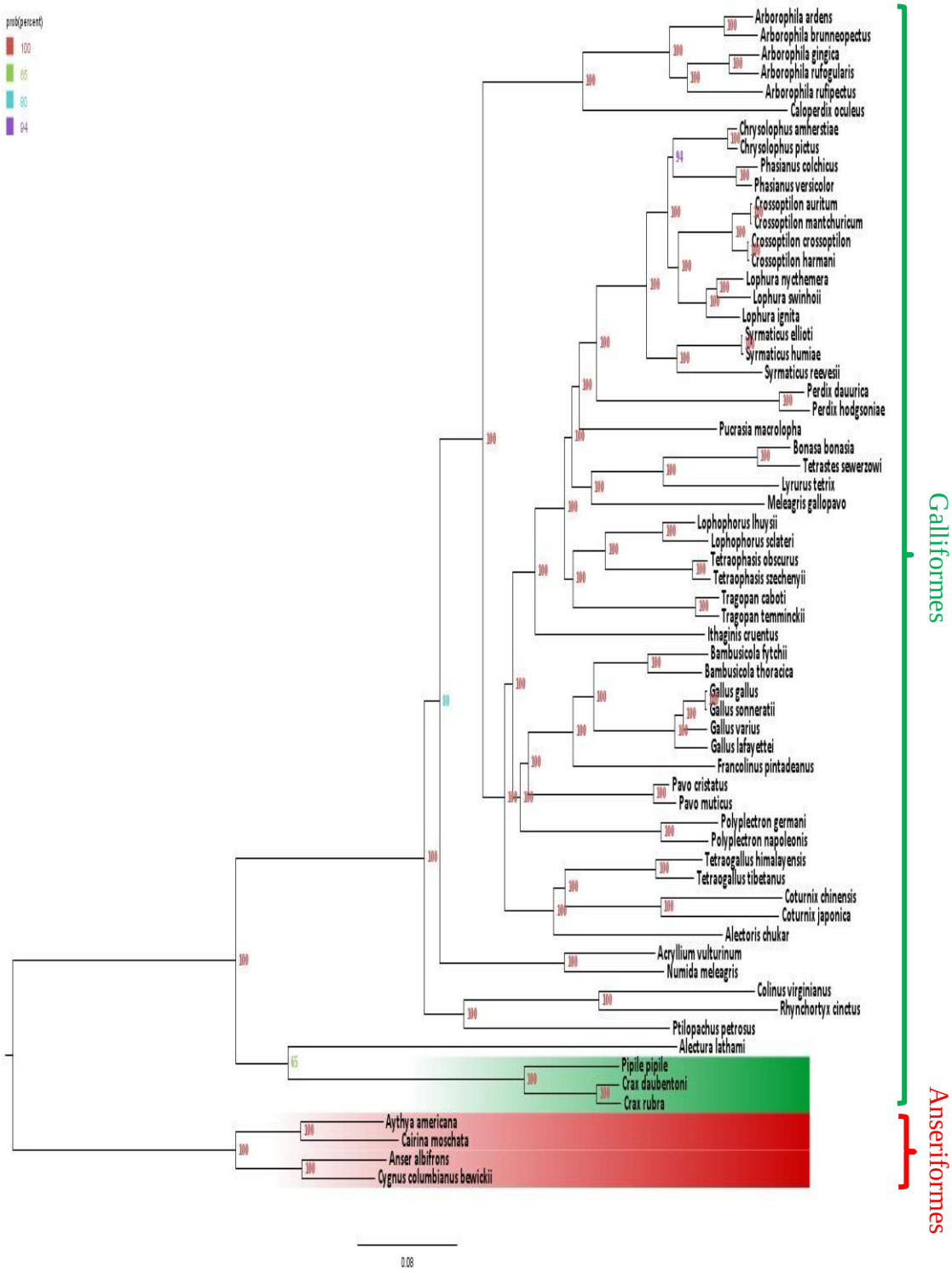


Figure 5.20: Phylogenetic relationships among Galliformes based on analysis of 13,365bp of Whole Mitochondrial Genome data (13 protein coding genes and 2 rRNA genes) from 59 genera of the Galliforme order including the novel *Pipile pipile*, and 4 Anseriforme genera as a putative outgroup. Topology and branch lengths are based on Maximum Likelihood General Time Reversible substitution models under non-uniform evolutionary rates with Gamma distribution and Invariable sites (GTR+G+I), through independent analysis of Bayesian inference in MrBayes (CIPRES Science Gateway). Bayesian analysis used MCMC with 1.25 million generations sampled every 1000 generations, discarding the first 25% of the run as burn-in, with Posterior probabilities to support node topologies (numbers at nodes in %).

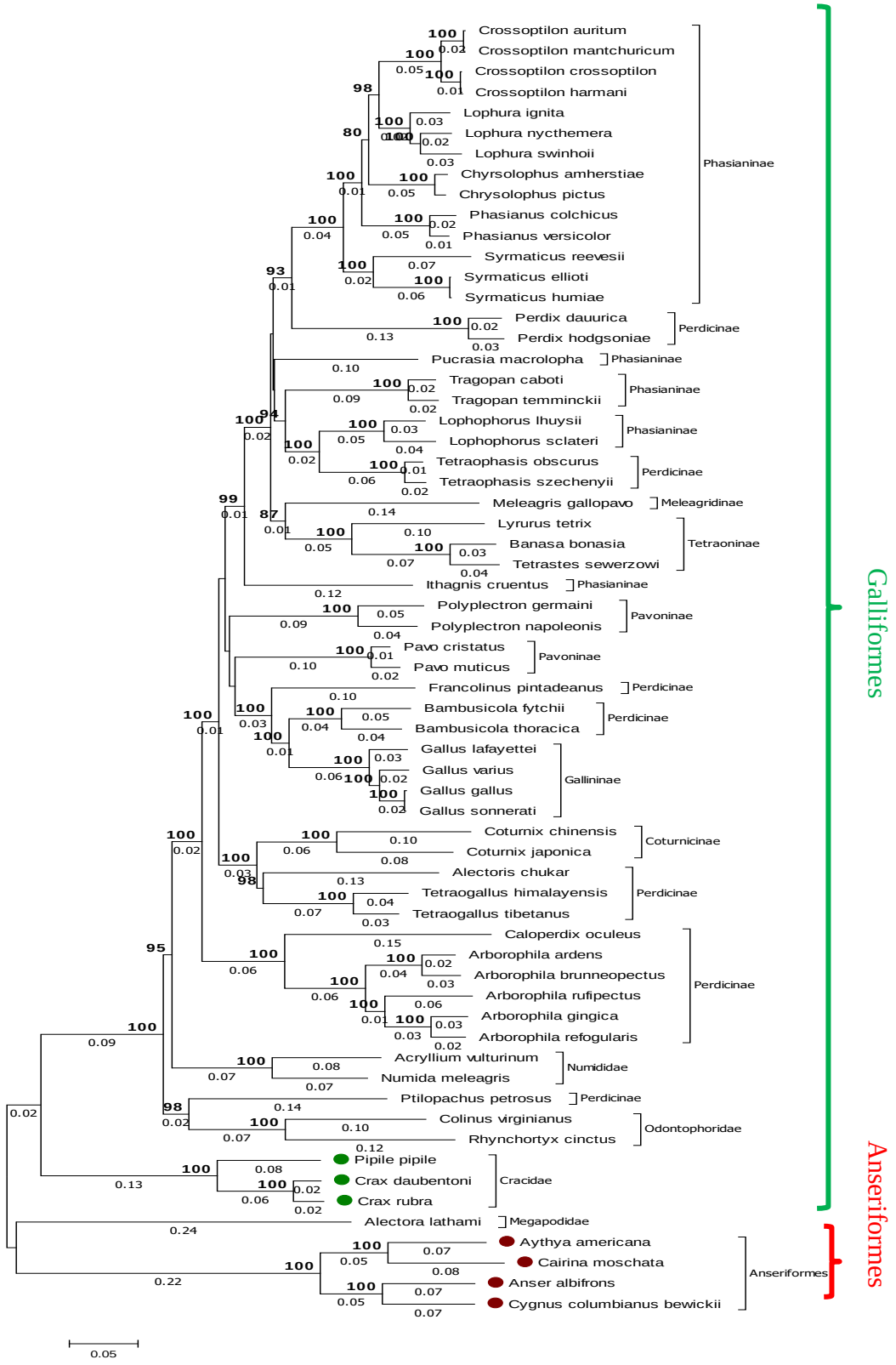


Figure 5.21: Phylogenetic relationships among Galliformes based on analysis of 13,365bp of Whole Mitochondrial Genome data (13 protein-coding genes and 2 rRNA genes) from 59 genera of the Galliforme order including the novel *Pipile pipile*, and 4 Anseriforme genera as a putative outgroup. Topology and branch lengths are based on Maximum Likelihood General Time Reversible substitution models under non-uniform evolutionary rates with Gamma distribution and Invariable sites (GTR+G+I), analysis in MEGA6.0 with Bootstrap statistical support. Maximum Likelihood analysis in MEGA with 500 bootstrap replicates for node statistical support (numbers at nodes in %, in bold), only nodes with >75% statistical support displayed.

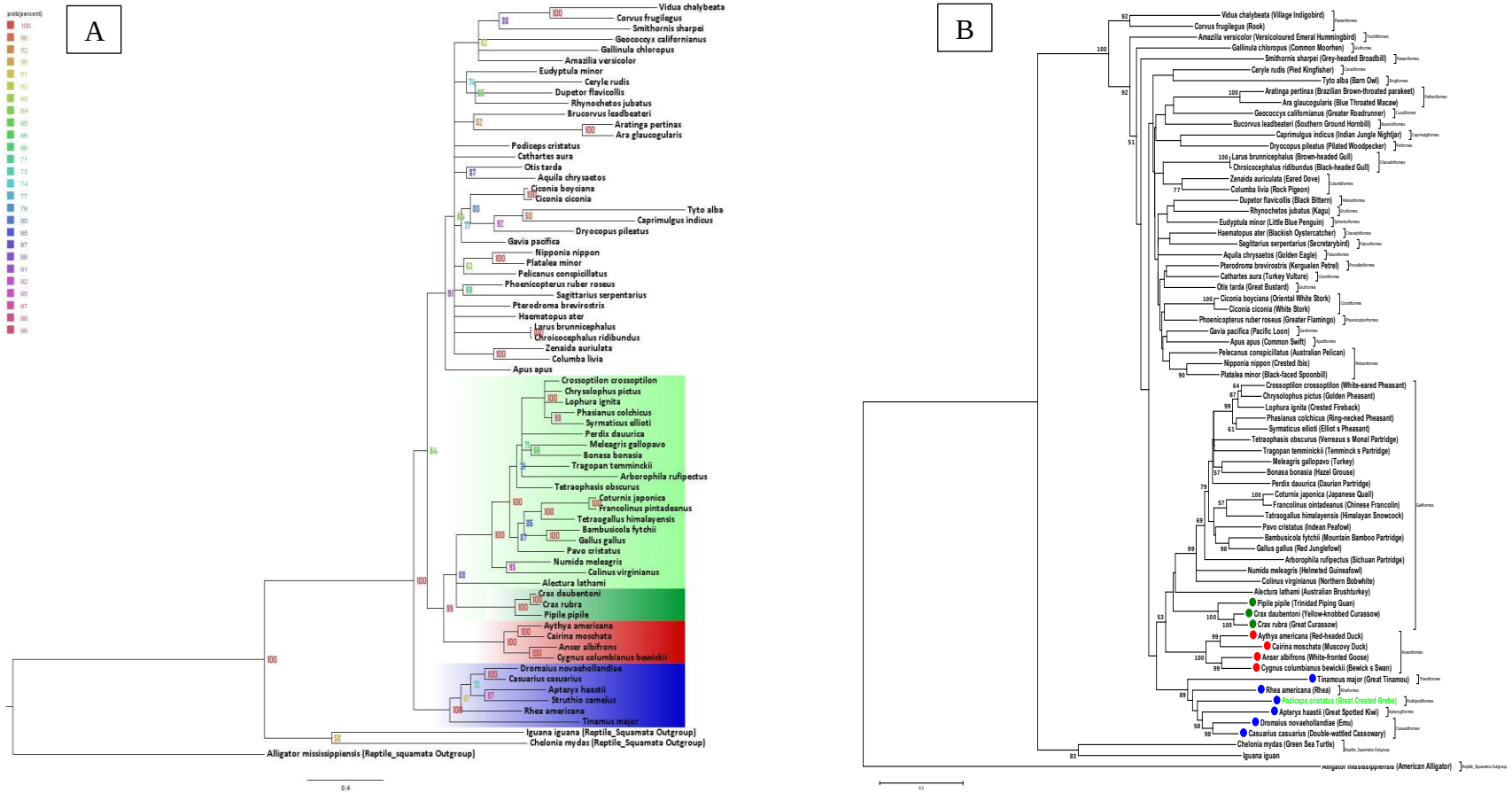


Figure 5.22: Phylogenetic relationships among Avian class based on analysis of Cytochrome b gene data (1143bp) from 71 genera; 23 of the Galliforme order including the novel *Pipile pipile*, and 3 reptile outgroups. Topology and branch lengths are based on Maximum Likelihood General Time Reversible substitution models under non-uniform evolutionary rates with Gamma distribution and Invariable sites (GTR+G+I), through independent analysis of Bayesian inference in MrBayes (A) and Maximum Likelihood inference in MEGA with Bootstrap statistical support (B). Bayesian analysis used MCMC with 5,000,000 generations sampled every 1000 generations, discarding the first 10% of the run as burn-in, with posterior probabilities to support node topologies. Maximum Likelihood analysis in MEGA with 500 bootstrap replicates for node statistical support (numbers at nodes in %), only nodes with >75% statistical support displayed.

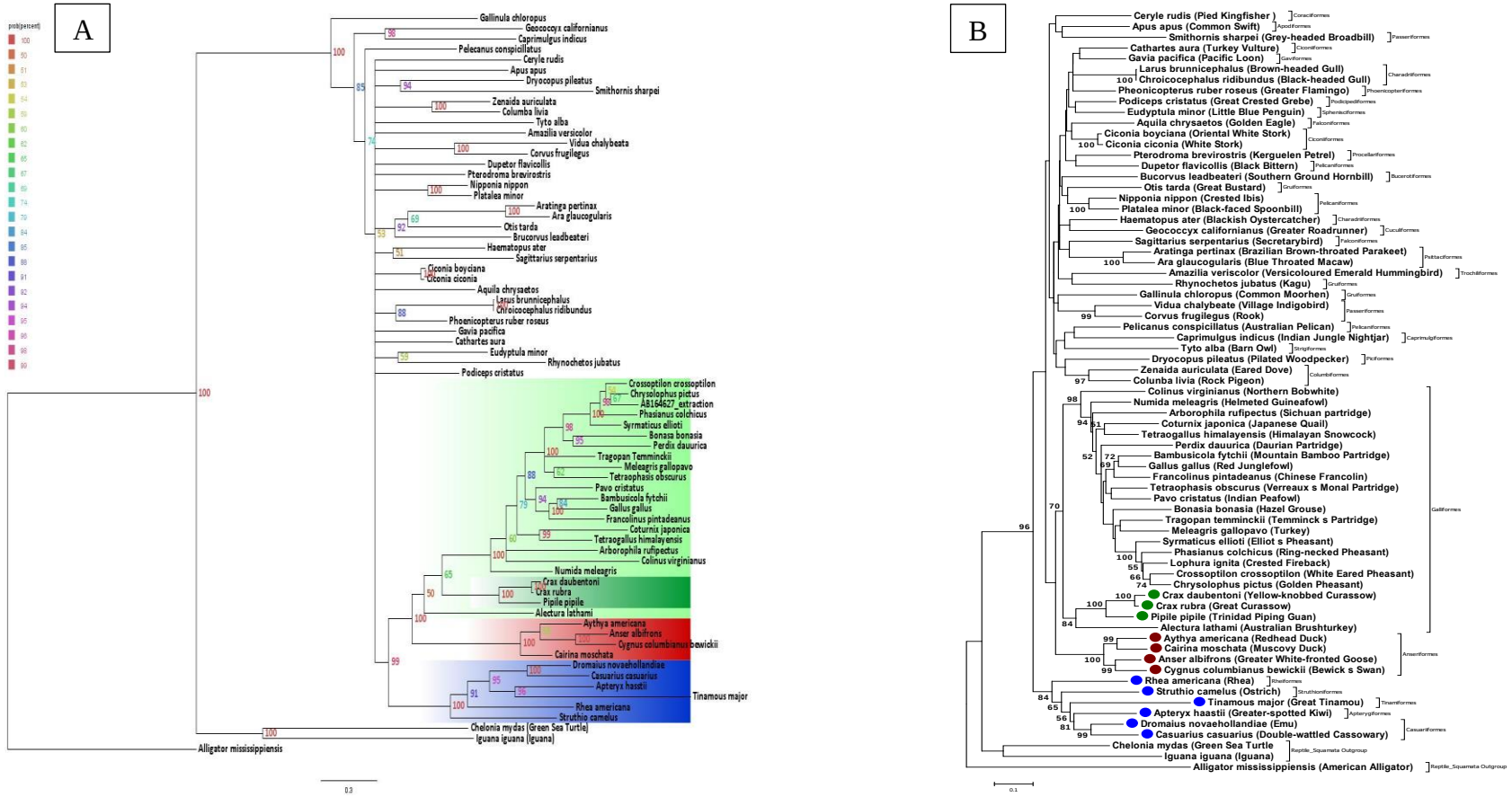


Figure 5.23: Phylogenetic relationships among Avian class based on analysis of Cytochrome c oxidase subunit 1 gene data (1548bp) from 71 genera; 23 of the Galliforme order including the novel *Pipile pipile*, and 3 reptile outgroups. Topology and branch lengths are based on Maximum Likelihood General Time Reversible substitution models under non-uniform evolutionary rates with Gamma distribution and Invariable sites (GTR+G+I), through independent analysis of Bayesian inference in MrBayes (A) and Maximum Likelihood inference in MEGA with Bootstrap statistical support (B). Bayesian analysis used MCMC with 7,500,000 generations sampled every 1000 generations, discarding the first 10% of the run as burn-in, with posterior probabilities to support node topologies. Maximum Likelihood analysis in MEGA with 500 bootstrap replicates for node statistical support (numbers at nodes in %), only nodes with >75% statistical support displayed.

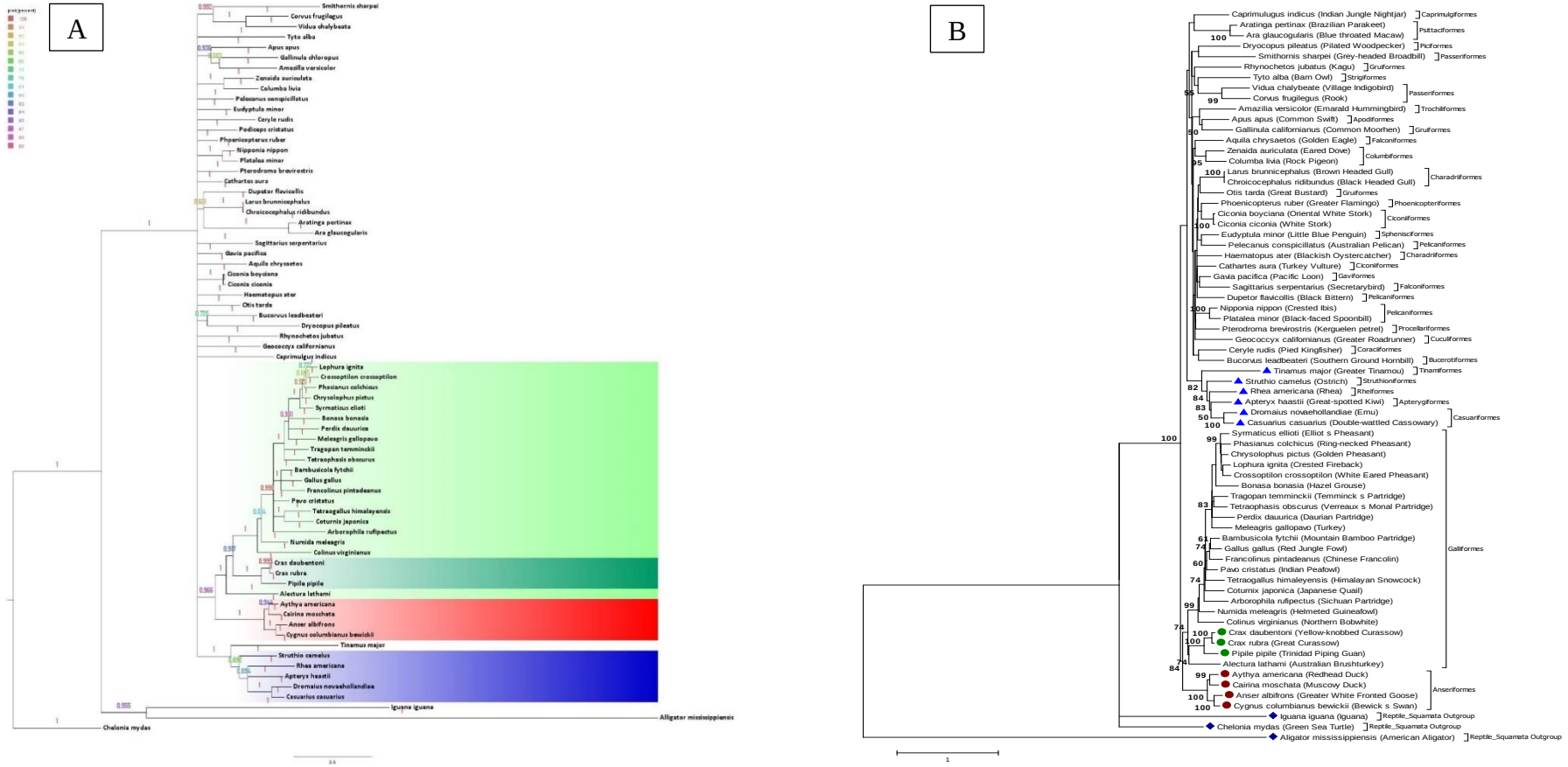


Figure 5.24: Phylogenetic relationships among Avian class based on analysis of NADH dehydrogenase subunit 2 gene data (1038bp) from 71 genera; 23 of the Galliforme order including the novel *Pipile pipile*, and 3 reptile outgroups. Topology and branch lengths are based on Maximum Likelihood General Time Reversible substitution models under non-uniform evolutionary rates with Gamma distribution and Invariable sites (GTR+G+I), through independent analysis of Bayesian inference in MrBayes (A) and Maximum Likelihood inference in MEGA with Bootstrap statistical support (B). Bayesian analysis used MCMC with 10,000,000 generations sampled every 1000 generations, discarding the first 10% of the run as burn-in, with posterior probabilities to support node topologies. Maximum Likelihood analysis in MEGA with 500 bootstrap replicates for node statistical support (numbers at nodes in %), only nodes with >75% statistical support displayed.

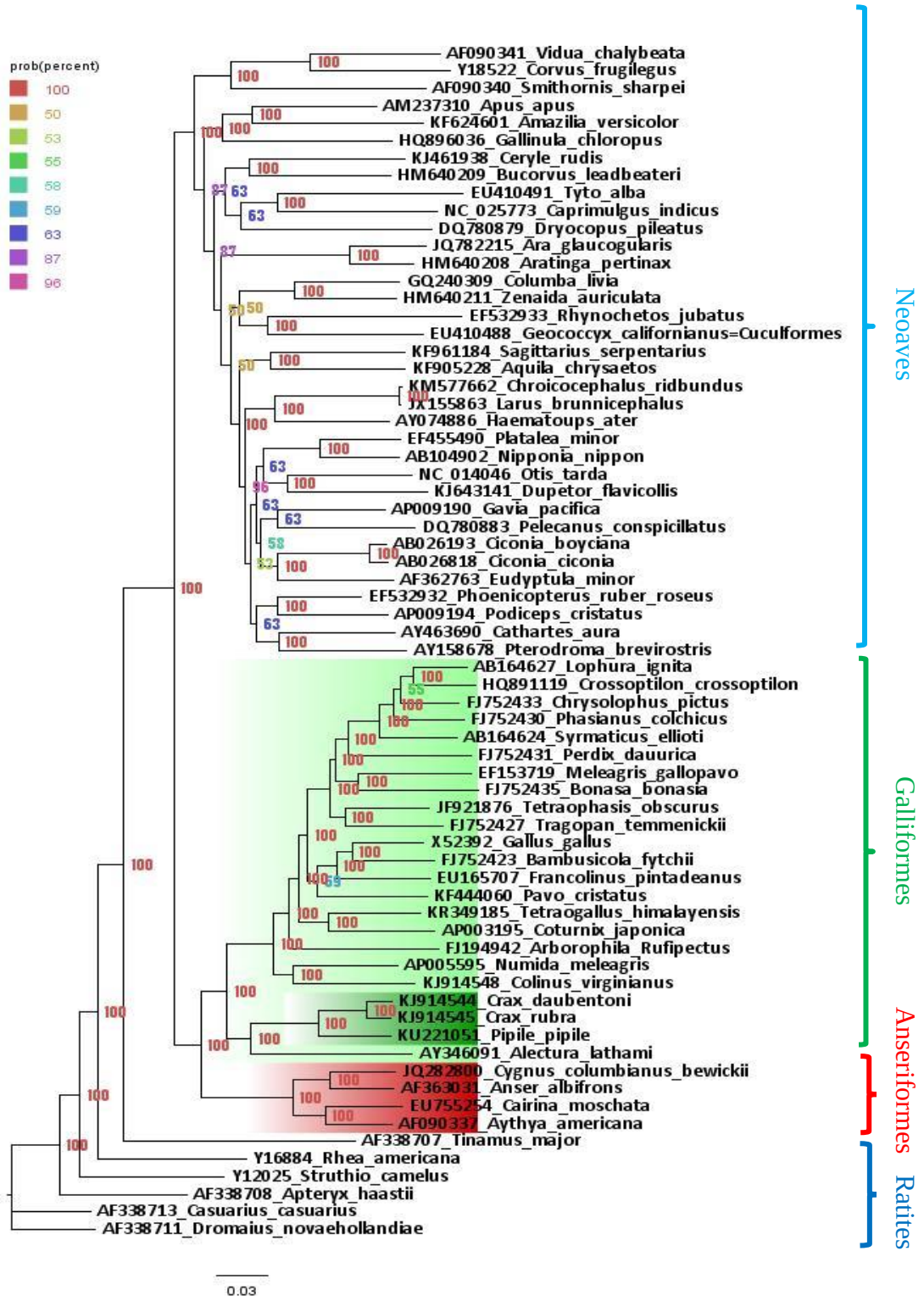


Figure 5.25: Phylogenetic relationships among Avians based on analysis of 11,433bp of Whole Mitochondrial Genome data (2 rRNA genes and 13 protein-coding genes) from 68 genera across the Aves class; 23 of the Galliforme order including the novel *Pipile pipile*. Topology and branch lengths are based on Maximum Likelihood General Time Reversible substitution models under non-uniform evolutionary rates with Gamma distribution and Invariable sites (GTR+G+I), through independent analysis of Bayesian inference in MrBayes. Bayesian analysis used MCMC with 20 million generations sampled every 1000 generations, discarding the first 25% of the run as burn-in, with posterior probabilities (%) to support node topologies.

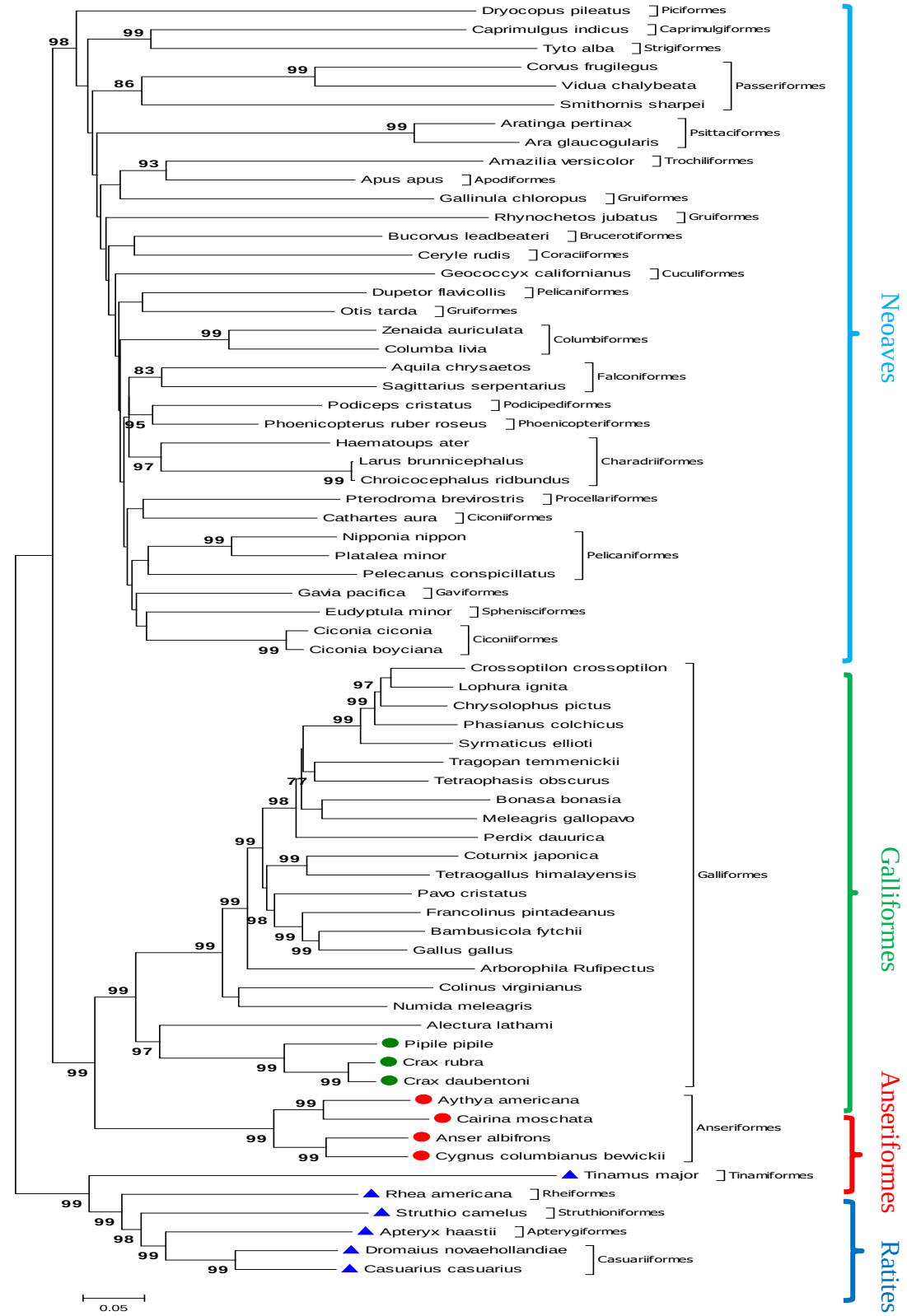


Figure 5.26: Phylogenetic relationships among Galliformes based on analysis of 11,433bp of Whole Mitochondrial Genome data (2 rRNA genes and 13 protein-coding genes) from 68 genera across the Aves class; 23 of the Galliforme order including the novel *Pipile pipile*. Topology and branch lengths are based on Maximum Likelihood General Time Reversible substitution models under non-uniform evolutionary rates with Gamma distribution and Invariable sites (GTR+G+I), in MEGA7.0 with Bootstrap statistical support. Maximum Likelihood analysis in MEGA with 500 bootstrap replicates for node statistical support (numbers at nodes in %, in bold), only nodes with >75% statistical support displayed.

Appendix 7: Table 1: Species Identification and NCBI DNA database accession numbers for sequences used in phylogenetic analysis.

English common name	Scientific name	CO1	Cytb	Control Region
Plain Chachalaca	*Ortalis vetula*		AY354494	
Grey-headed Chachalaca	*Ortalis cinereiceps*	JQ175625		
		KF799991		
Chestnut-winged Chachalaca	*Ortalis garrula*		AY659780	
Rufous-vented Chachalaca	*Ortalis ruficauda*		AY659781	
West Mexican Chachalaca	*Ortalis poliocephala*		AY659784	
Chaco Chachalaca	*Ortalis canicollis pantanalensis*		AY659783	
	Ortalis canicollis		AF165472	AF165436
White-bellied Chachalaca	*Ortalis leucogastra*		AY659779	
Speckled Chachalaca	*Ortalis guttata*	KF446138	AY659782	
Little Chachalaca	*Ortalis motmot*	JQ175626	AY659778	
		JQ175627		
		JQ175628		
Band-tailed Guan	*Penelope argyrotis*		AY659803	
Bearded Guan	*Penelope barbata*	JN801892		
Andean Guan	*Penelope montagnii*		AY659802	
Marail Guan	*Penelope marail*	JQ175711		
		JQ175712		
		JQ175713		
Rusty-margined Guan	*Penelope superciliaris*		AY659804	AY145313
Crested Guan	*Penelope purpurascens*	JQ175718	AY659800	AY145312
		JQ175719		
White-winged Guan	*Penelope albipennis*	JN801886		
		JN801887		
		JN801888		
		JN801889		
		JN801890		
		JN801891		
Spix's Guan	*Penelope jacquacu*	JN801893	AY659801	AY145318
		JN801894		
		JN801895		
		JQ175708		
		JQ175709		
		JQ175710		
Dusky-legged Guan	*Penelope obscura*	JQ175714	AF165474	AF165432
		JQ175715		
		JQ175716		
White-crested Guan	*Penelope pileata*	JQ175717		
Chestnut-bellied Guan	*Penelope ochrogaster*		AY367101	AY145311
White-browed Guan	*Penelope jacucaca*	JN801894		
		JN801895		
Highland Guan	*Penelopina nigra*	AF165499	AF165475	AF165433
Trinidad Piping Guan	*Pipile pipile*	KU221051	KU221051	KU221051
		KU221052	KU221052	KU221052
		KU221053	KU221053	KU221053
		JQ175861		
		JQ175862		
			AY367106	AY145320
Blue-throated Piping Guan	*Pipile cumanensis*		AY659798	AY145319
Gray's Blue-throated Piping Guan	*Pipile cumanensis grayi*		AY659797	
Red-throated Piping Guan	*Pipile cujubi*		AY367104	AY145314

Appendix 7, Table 1 contd.

English common name	Scientific name	CO1	Cytb	Control Region
Black-fronted Piping Guan	*Pipile jacutinga*		AF165476	AF165431
Wattled Guan	*Aburria aburri*		AF165466	AF165430
Black Guan	*Chamaepetes unicolor*	JQ174405	AY659796	
		JQ174406		
		JQ174407		
Sickle-winged Guan	*Chamaepetes goudotii*	JN801552	AF165467	AF165434
		JN801553	AY659795	
		JN801554		
Horned Guan	*Oreophasis derbianus*	AF165495	AF165471	AF165435
			AY659805	
Nocturnal Curassow	*Nothocrax urumutum*	AF165494	AF165470	AF165440
Crestless Curassow	*Mitu tomentosum*	AY141918	AY141928	AY145310
		JQ175398	AY659787	
		JQ175399		
		JQ175400		
Salvin's Curassow	*Mitu salvini*	AY141917	AY141927	
		EU525437	AY659785	
		EU525438		AY145309
Razor-billed Curassow	*Mitu tuberosum*	JN801819	AF165469	AF165437
		EU525439		
		EU525440		
		EU525441		
Alagoas Curassow	*Mitu mitu*	AY141916	AY141926	AY145308
Helmeted Curassow	*Pauxi pauxi*	AF165497	AF165473	AF165439
			AF068190	
			AY659788	
Horned Curassow	*Pauxi unicornis*	AY141919	AY141929	AY145317
			AY659786	
Great Curassow	*Crax rubra*	AY141915	AY141925	AY145307
			AY956378	
		KJ914545	KJ914545	KJ914545
			AY659793	
Blue-billed Curassow	*Crax alberti*	AY141910	AY141920	AY145304
			AY659794	
Yellow-knobbed Curassow	*Crax daubentoni*	AY141912	AY141922	
		KJ914544	KJ914544	KJ914544
			AY659792	AY145305
Black Curassow	*Crax alector*	AY141911	AY141921	AF145315
		JQ174568	AY659789	
		JQ174569		
Wattled Curassow	*Crax globulosa*	AY141914	AY141924	AY145316
Bare-faced Curassow	*Crax fasciolata*	AY141913	AY141923	AY145306
		JQ174570	AY659790	
Red-billed Curassow	*Crax blumenbachii*	AF165492	AF165468	AF165438
			AY659791	
Outgroups				
Horned Screamer (Anseriforme)	*Anhima cornuta*		AY140735	
Northern Screamer (Anseriforme)	*Chauna torqata*		AY140736	
Orange-footed Scrubfowl (Galliforme)	*Megapodius reinwardt*		AF165465	
Australian Scrubfowl (Galliforme)	*Alectora lathami*		AY659806	