Using ancient DNA to understand the evolution and biogeography of the Caribbean land mammal fauna

A thesis submitted to

Royal Holloway, University of London

for the degree of

DOCTOR OF PHILOSOPHY

Roseina Woods

Department of Biological Sciences

September 2017

Declaration of Authorship

I, Roseina Woods, hereby declare that this thesis and the work presented in it is entirely my own. Where I have consulted the work of others, this is clearly stated.

Signed:

Date:

Abstract

The insular Caribbean is a prefect arena for the study of evolution and biogeography. One exceptional feature of the Caribbean islands is the colonisation of this archipelago by terrestrial mammal fauna. Historic changes in climate, colonisation of new species and the arrival of humans has greatly affected the endemic fauna of the Caribbean. Due to the extinction of much of the Late Quaternary Caribbean mammal fauna during the Holocene, many aspects of taxonomy in this group remains unresolved, preventing a complete study of evolutionary questions surrounding speciation, colonisation and extinction and critically, hindering the conservation of endangered species. This study has incorporated three groups of endemic Caribbean mammals: eulipotyphlan insectivores, rodents and primates, in order to investigate patterns in colonisation and evolutionary histories. Advanced techniques in extraction and sequencing of degraded or ancient DNA were utilised as part of this study in order to successfully able to extract and sequence DNA from museum and zooarchaeological specimens, despite the high degradation indicative of DNA from tropical regions.

This study has found that whilst the Caribbean is home to ancient relic species, the islands are also hotspots for the generation of new species, as shown by the recent intra-island speciation of Hispaniolan *Nesophontes* species. This study has also been able to demonstrate patterns of insular evolution including accelerated evolution and the generation of unusual morphology, both exhibited by the Jamaican primate *Xenothrix* and extinct caviomorph rodents. Colonisation histories differ between groups, but both rodent and primate taxa looked at as part of this study colonised the Caribbean via over water dispersal during the mid-late Miocene, a period of low global sea level, when a riverine connection may have facilitated faunal movement between the Caribbean and South America. Previous taxonomy in many of the species looked as part of this study were based on morphology and needs to be re-assessed in light of the results of this molecular analysis. This is particularly important for the living Cuban hutia in the genera *Capromys* and *Mesocapromys* where the conservation of declining populations is vital for the continued survival of these unqiue taxa.

Acknowledgements

First and formost I would like to thank my supervisors and advisors, Prof. Ian Barnes, Dr Sam Turvey, and Prof. Mark Brown, for their patience and support throughout my PhD. I would also like to thank Dr Sam Turvey for allowing me to sample from many of the specimens included in this thesis, and Prof. Ian Barnes for providing me with an environment that allowed me to work independently, but also being around for discussion and advice when I most needed it.

A special mention must go to Dr Selina Brace who has been a massive influence on my PhD and whose advice, expertise and encouragement were indispensible to me throughout this project.

This studentship was funded by the National Environment Research Council (NERC), I am very grateful for the opportunity this funding provided me with. I would also like to thank the many museums that allowed me to use their collections for this project including the Natural History Museum (NHM), Instituto de Ecología y Sistemática (IES), The University of Oriente, Museo d'historia natural tomas romay, Royal Albert Memorial Museum (RAMM), Peabody Museum Yale, and Florida Museum of Natural History (FLMNH).

I would like to thank all the staff in the NHM Imaging and Analysis Centre (IAC) for their help with micro-CT scanning during this project. I would like to thank Roberto Miguez at the NHM for allowing me access to the collection. Additionally I would like to thank Holly Morgenroth at RAMM for taking the time to allow me to sample from *Geocapromys thoracatus* specimens for this thesis. I would particularly like to thank Prof. Ross MacPhee for providing me with samples for the *Xenothrix* chapter and allowing me access to his expertise. I would like to thank Prof. Pierre-Henri Fabre and Dr Nathan Upham for sharing their data, advice and expertise.

A special thanks goes out to all my fellow postgraduates and friends at Royal Holloway particularly Kate and Lauren, at the Natural History Museum all of the occupants of PA417, Dr Tom Booth, Dr Sophy Charlton and especially Dr Melissa Marr who took the time to help me when ever I needed it, and Leila D'Souza for sharing in my island woes.

My final thanks goes to my family for your continued support, which means so much to me. Charlie, Jay, Beagle and Calpyso, my Mum who first instilled in me my love of the natural world and Dave for technical support.

Contents

Title page **1**

Declaration of Authorship *2*

Abstract **3**

Acknowledgements **4**

List of figures and tables **12**

Chapter 1. Introduction **17**

1.1 Overview *17*

1.2 Island biogeography and evolution of insular species *17*

Introduction to island study 17

Island effects 18

Mechanisms of speciation and adaptive and non-adaptive radiations 19

Colonisation and extinction 20

1.3 The Caribbean islands *21*

Overview 22

A brief timeline of the formation of the Caribbean islands (Iturralde-Vinent, 2006) (Figure 2) 23

Climate in the wider Caribbean 25

Cuba 26

The Cayman Islands and the Cayman trench 27

Hispaniola 27

Puerto Rico 28

Jamaica 28

1.4 The Caribbean land mammal fauna *28*

Rodentia 28

Lipotyphla 29

Primates 29

Colonisation 29

The Aves ridge and the GAAR-landia hypothesis 30

Over-water dispersal 30

Extinction 30

1.5 Conservation archaeogenomics *31*

1.6 Taxonomy of Caribbean land mammal species *32*

1.7 Ancient DNA *33*
Ancient DNA degradation and preservation 33
Ancient DNA contamination 34
Ancient DNA in the tropics 35
Next Generation Sequencing 35
Target capture enrichment 36
1.8 Thesis Aims *36*
References *38*

Chapter 2. Methods 45
2.1 Overview *45*
2.2 Taxonomic sampling *45*
Capromyidae 45
Heptaxodontidae 45
Heteropsomyinae 46
Nesophontes 46
Xenothrix 46
2.3 Ancient DNA extractions *46*
2.4 Preliminary PCR and shotgun Sanger sequencing *47*
2.5 Next Generation Sequencing *47*
Library builds 47
Sequencing 48
Target capture enrichment 48
2.6 Post sequencing data processing *49*
Assessment of sequence data 50
Mapping ascertainment bias 51
Alignments: Programs and parameters 54
Alignments: Partitioning and concatenation 54
3.7 Phylogenetics *55*
Phylogenetics: Maximum Likelihood 55
Phylogenetics: Bayesian 56
2.8 Pairwise distance analyses *57*
2.9 Divergence dating *57*
References *58*
Appendix Chapter 2. Methods *61*

Chapter 3. Palaeogenomic analyses reveal recent intra-island evolutionary radiation in the extinct Caribbean "island shrew" (Nesophontes) 67
3.1 Abstract 67
3.2 Introduction 68
Island evolution and the Caribbean's endemic lipotyphlan insectivores 68
Nesophontes taxonomy and specimen preservation 69
Delimitation of species 72
Levels of intraspecific morphological and molecular variation in lipotyphlan insectivores 74
Recovery of molecular data from extinct Caribbean fauna 75
3.3 Study overview and aims 75
3.4 Materials and methods 76
Specimen information 76
Micro-CT scanning of Nesophontes specimens 76
Extractions, sequencing and phylogenetic analysis 76
3.5 Results 78
3.6 Discussion and conclusions 81
Overview 81
Interpretation of genetic distances and divergence dates in comparison with other lipotyphlan species 81
Interpretation of genetic distances and divergence dates in comparison with other mammals 83
Hispaniolan climate and geology 84
A comparison with other Hispaniolan vertebrate radiations 86
References 89
Appendix Chapter 3. Palaeogenomic analyses reveal recent intra-island evolutionary radiation in the extinct Caribbean "island shrew" (Nesophontes) 95

Chapter 4. Ancient DNA of the extinct Jamaican monkey *Xenothrix* reveals extreme insular change within a morphologically conservative primate radiation 109
4.1 Abstract 109
4.2 Introduction 110
Island evolution and extinctions: the insular Caribbean mammal fauna 110
The evolutionary history of the extinct Greater Antillean primates 112
Colonization, biogeographic origin and phylogenetic relationships of the Greater Antillean primates 113

Recovery of molecular data from extinct Caribbean fauna 113
4.3 Study overview and aims *114*
4.4 Materials and Methods *118*
Data Collection 118
Sequence analysis 119
4.5 Results *120*
4.6 Discussion and Conclusions *125*
Morphological versus molecular phylogenies for Caribbean primates 125
Implications for primate insular evolution and morphological conservatism 125
Colonisation and evolutionary history of the Caribbean primate fauna 127
References *129*
Appendix Chapter 4. Ancient DNA of the extinct Jamaican monkey Xenothrix reveals extreme insular change within a morphologically conservative primate radiation *135*
Chapter 5. Phylogeographic analysis of sympatric Cuban hutia reveal parallel patterns evolution in a mammalian insular radiation 147
6.1 Abstract *147*
Island evolution and the Caribbean 148
Cuba 148
Cuban hutias 149
Capromys 153
East-West split hypothesis 153
Mesocapromys 155
Mysateles 156
Extraction and sequencing of ancient DNA from the tropics 156
Conservation 157
6.3 Study aims and objectives *157*
6.4 Methods *158*
Specimen information 158
Extractions and sequencing 158
Post sequencing data processing and phylogenetic analysis 158
6.5 Results *159*
6.5 Discussion and conclusions *172*
Overview 172
Is the east-west split in Capromys supported? 172

Do Mesocapromys and Capromys populations show the same phylogeographic structure? 173
Divergence dates 173
Taxonomic implications 174
Conservation implications 175
Final conclusions 176
References 177
Appendix Chapter 5. Phyleogographic analysis of sympatric Cuban hutia reveal parallel patterns evolution in a mammalian insular radiation 181
Chapter 6. Evolutionary histories and biogeographic origins of the Greater Antillean rodent fauna 192
6.1 Abstract 192
6.2 Introduction 193
Caribbean island evolutions and extinctions 193
Greater Antillean caviomorph rodents 193
Wider colonisation histories and evolutionary relationships 194
Inter-island colonisation histories and evolutionary relationships 195
6.3 Capromyidae 196
Isolobodon and Hexolobodon 197
Capromys species from the Cayman Islands 197
Geocapromys columbianus and Geocapromys thoracatus 199
6.4 Heteropsomyinae 200
6.5 Heptaxodontidae 201
6.6 Study aims and objectives 202
6.7 Methods 202
Specimen information 202
Extractions and sequencing 202
Post sequencing data processing and phylogenetic analysis 205
6.8 Results 206
6.9 Discussion and conclusions 212
Taxonomic implications 213
Colonisation of the insular Caribbean by Greater Antillean caviomorph rodents 214
Wider inter-island evolutionary relationships 214
Inter-island evolutionary histories 215
Heteropsomyinae 215

Capromyidae ... 216
Insular evolution ... 216
References ... *218*
Appendix Chapter 6. Evolutionary histories and biogeographic origins of the Greater Antillean rodent fauna ... *224*
Chapter 7. Discussion and conclusions ... 231
7.1 Overview ... *231*
7.2 Palaeogenomic analyses reveal recent intra-island evolutionary radiation in the extinct Caribbean "island shrew" (Nesophontes) ... *231*
7.3 Ancient DNA of the extinct Jamaican monkey Xenothrix reveals extreme insular change within a morphologically conservative primate radiation ... *231*
7.4 Phylogeographic analysis of sympatric Cuban hutia reveal parallel patterns evolution in a mammalian insular radiation ... *232*
7.5 Evolutionary histories and biogeographic origins of the Greater Antillean rodent fauna ... *232*
7.6 Thesis aims ... *233*
7.7 Overall thesis summary, critical evaluation and future work ... *235*
Ancient DNA ... 235
Colonisation histories ... 236
Patterns of insular evolution ... 237
References ... *239*
Appendix: Chapter 8. (a) The small and the dead: ancient DNA studies using micromammal species ... 241
8a.1 Abstract ... *241*
8a.2 Introduction ... *241*
8a.4 Cave sites ... *243*
8a.5 Resolving phylogenetic placement of extinct species ... *244*
8a.6 Species identification ... *244*
8a.7 DNA barcoding ... *245*
8a.8 Owl pellets ... *245*
8a.9 Biogeographic hypotheses and island micromammals ... *246*
8a.10 Interaction with Humans ... *246*
8a.11 Population genetics ... *246*
8a.12 Conservation ... *247*
8a.13 Climate change ... *248*

8a.14 Conclusion, future directions ... 248

References ... 250

List of figures and tables

Chapter 1. Introduction

Figure 1. Map of the Caribbean region. Dashed lines indicate Greater and Lesser Antilles and larger Greater Antillean islands are labelled. 22

Figure 2. Maps of the Caribbean region during three time periods: i) Eocene-Early Oligocene, ii) Late Oligocene and iii) Early/Middle Miocene adapted from Iturralde-Vinent (2006). 24

Figure 3. Bathymetric map of the Caribbean region (Hastings & Dunbar 1999; Amante & Eakins 2009) 26

Chapter 2. Methods

Figure 1. Illustration of the mapping process for the limitation of mapping ascertainment bias. 53

Table 1a. Specimens sampled as part of this thesis 61

Table 2a. Mapping parameters used in CLC genomics workbench 66

Chapter 3. Palaeogenomic analyses reveal recent intra-island evolutionary radiation in the extinct Caribbean "island shrew" *Nesophontes*

Table 1. Nesophontes species previously described on the basis on morphology, and their known geographical ranges in the Caribbean. 71

Table 2. Sequencing results for N. zamicrus sample used for phylogenetic analysis 78

Figure 1. Maximum likelihood phylogeny generated using RAxML, including N. zamicrus and N. paramicrus and extant lipotyphlan insectivores for which whole mitochondrial genome data are available 79

Figure 2. A time-calibrated phylogeny generated in BEAST, showing estimated divergence dates for N. zamicrus and N. paramicrus. 80

Figure 3. Map of the Caribbean and the island of Hispaniola, showing geological features of interest including paleo-island boundaries. 86

Figure 1a. Nesophontes micro-CT scans 95

Figure 2a. Bayesian inference phylogeny inclusive of all out-group taxa. 96

Figure 3a. Graphs output from CLC genomics workbench (v.8). 97

Table 1a. Evolutionary models chosen for each gene in the alignment using the program Partitionfinder (Lanfear et al. 2012) 98

Table 2a. Species used in pairwise genetic distance analysis downloaded from Genbank NCBI database (Benson et al. 2005) *99*

Table 3a. Genetic pairwise distance matrix values for N. paramicrus and N. zamicrus for four nuclear and two mitochondrial genes *99*

Table 4.1a. Pairwise distances for lipotyphlan insectivore sister species pairs in the mitochondrial gene 12s (mitochondrially encoded 12S ribosomal RNA). *100*

Table 4.2a. Pairwise distances for lipotyphlan insectivore sister species pairs the protein coding nuclear gene CREM (cAMP responsive element modulator) *101*

Table 4.3a. Pairwise distances for lipotyphlan insectivore sister species pairs the protein coding nuclear gene BRCA1 (Breast Cancer 1 gene) *101*

Table 4.4a. Pairwise distances for lipotyphlan insectivore sister species pairs the protein coding nuclear gene RAG1 (recombination activating gene 1). *102*

Table 4.5a. Pairwise distances for lipotyphlan insectivore sister species pairs the protein coding nuclear gene BDNF (brain derived neurotrophic factor). *103*

Table 4.6a. Pairwise distances for lipotyphlan insectivore sister species pairs the protein coding nuclear gene CREM (cAMP responsive element modulator). *103*

Table 4.7a. Pairwise distances for lipotyphlan insectivore sister species pairs in the mitochondrial gene cytB (cytochome B). *104*

Table 5a. Genes used in phylogenetic analysis including GenBank NCBI database accession numbers for downloaded sequence data.. .... *105*

Table 6a Fossil constraints and priors used in divergence dating-analysis *108*

Chapter 4. Ancient DNA of the extinct Jamaican monkey Xenothrix reveals extreme insular change within a morphologically conservative primate radiation

Figure 1. Five alternative tree topologies derived from previous proposals about the evolutionary origins of Xenothrix, including South American extant genera.. .... *115*

Figure 2. Map showing Greater and Lesser Antillean islands including Jamaica. Somerville Cave, where specimens analysed as part of this study were excavated *118*

Table 1. Sequencing results including pre-target capture enrichment screening results and post-capture results. *121*

Figure 3. Genus-level maximum likelihood phylogeny generated using whole mitochondrial genomes and produced in RAxML, using data sequenced in this study for Xenothrix and data for 15 other primate genera from Genbank, and with Macaca fuscata selected as outgroup. *123*

Figure 4. Time-calibrated phylogeny showing estimated divergence dates for Xenothrix, callicebine species, and other platyrrhine genera. *124*

Table 1a. Genes used to generate baits for capture enrichment *135*

Table 2a. Genes used as reference sequences for alignment of raw reads during analysis 135

Figure 1a. (part 1) Coverage of the reference (whole mitochondrial genome) by mapped reads. Reads mapped using CLC genomics work bench v.8 (CLC Bio-Qiagen, Aarhus, Denmark) 136

Figure 1a. (part 2) Total number of reads mapped to the reference (whole mitochondrial genome). Reads mapped using CLC genomics work bench v.8 (CLC Bio-Qiagen, Aarhus, Denmark) 136

Table 3a. substitution models chosen for the partitioned dataset 137

Figure 2a. Bayesian phylogeny constructed using MrBayes (Ronquist & Huelsenbeck 2003). 139

Table 4a. Tests of alternative topologies suggested by previous studies made by submitting the sitewise log-likelihood values obtained from RAxML v.8 (Stamatakis 2014) to CONSEL (Shimodaira & Hasegawa 2001), to calculate the p-values for each tree topology by using the AU test. 139

Table 5a. Fossil constraints and priors used in divergence date analysis 140

Figure 3a. Specimens used for the extraction of aDNA. (left) A) femur: AMNH 268003 and (right) B) proximal ulna: AMNH 268010 141

Figure 4a. Consensus tree adapted from MacPhee and Horovitz (2004) including Caribbean taxa. 142

Table 6a. BLAST results from the NCBI database for reads mapped to reference sequences 143

Chapter 5. Phylogenetic analysis of sympatric Cuban hutia reveal parallel patterns of evolution in a mammalian insular radiation

Table 1. List of capromyid species and subspecies found on Cuba with IUCN threat status if accessed 150

Figure 1, Part 1 Map of Cuba (west) and offshore islands including Isla de la Juventud, with scale bar and labels for localities. Numbered localities indicate the locations of different species, subspecies and populations of Cuban Capromys and Mesocapromys (Appendix: Table 1-3a) 151

Figure 1, Part 2 Map of Cuba (east) and offshore islands including Isla de la Juventud, with scale bar and labels for localities. Numbered localities indicate the locations of different species, subspecies and populations of Cuban Capromys and Mesocapromys (Appendix: Table 1-3a) 152

Table 2: Sequencing results including pre-target capture enrichment screening results and post-capture results. 161

Figure 2. Divergence-dated Bayesian phylogeny constructed using MrBayes (Ronquist & Huelsenbeck 2003) 169

Figure 3. Map of Cuba showing Capromys and Mesocapromys sample locations included in Figure 1. .... *171*

Table 1a. Sample information from this study including sample identification number and location data as well as genes used in analysis *181*

Table 2a. Sample information from (Upham & Borroto-Páez 2017) including sample identification number, location data, and genes used in analysis *184*

Table 3a. Sample information from (Upham & Borroto-Páez 2017) including sample identification number and location data as well as genes used in analysis *185*

Table 4a. Estimates of evolutionary divergence between cytB sequence data from Mesocapromys and Mysateles individuals from different localities. *187*

Table 5a Estimates of evolutionary divergence between cytB sequence data from Capromys indervisuals from different localities. *188*

Table 6a. Substitution models chosen for the partitioned dataset *189*

Table 7a Genes used to generate baits for capture enrichment *189*

Figure 1a. Map of Cuba showing ecological regions adapted from WWF database *190*

Figure 2a. Maximum likelihood phylogeny produced using RAxML v.8 (Stamatakis 2014) which was implemented in CIPRES Science Gateway v.3.3 (Miller, Pfeiffer, & Schwartz, 2010). *191*

Chapter 6. Evolutionary histories and biogeographic origins of the Greater Antillean rodent fauna

Figure 1. Map of the Caribbean region highlighting the location of the Cayman Islands, with a more detailed scaled map showing the three islands that make up the Cayman Islands group: Grand Cayman, Little Cayman and Cayman Brac. *198*

Figure 2. Map of the Caribbean region including the geographic position of Little Swan Island. *199*

Table 1. Specimens sampled for aDNA as part of this study. Including: species, sample ID, material, locality of collection and Museum or collection *204*

Table 2: Sequencing results including pre-target capture enrichment screening results and post-capture results. *207*

Figure 3. Bayesian phylogeny constructed using MrBayes (Ronquist & Huelsenbeck 2003) *210*

Figure 4. Time-calibrated phylogeny showing estimated divergence dates for Greater Antillean caviomorph species and extant Neotropical echimyid genera *211*

Figure 1a. Maximum likelihood phylogeny produced using RAxML v.8 (Stamatakis 2014) which was implemented in CIPRES Science Gateway v.3.3 (Miller, Pfeiffer, & Schwartz, 2010). *224*

Table 1a. Genes used to generate baits for capture enrichment 225

Table 2a. Sequence data used in post-sequencing data processing as reference sequences in read mappings 226

Table 3a. Mapping parameters used in CLC genomics work bench 227

Table 3a. Echimyid outgroup taxa for whole mitochondrial genome and nuclear genes . 228

Table 4a. Substitution models chosen for the partitioned dataset 229

Table 5a. Fossil constraints and priors used in divergence dating analysis 230

Chapter 8a. The small and the dead: A review of ancient DNA studies analysing micromammal species

Figure 1a. Ancient DNA studies including micromammals from 1990-2017 257

Figure 2a. Ancient DNA studies including different micromammal orders (1990-2015) .. 257

Table 1a. All published studies from 1990 to 2017 in which ancient or degraded DNA was utilised to study micromammal fauna 258

Chapter 1. Introduction

1.1 Overview

The study of insular species has been key to the development and continued understanding of evolutionary theory. The Caribbean islands in particular have been vital in providing a backdrop for the continued study of species diversity. In this thesis I have examined the evolutionary and biogeographic histories of the Caribbean's rare and extinct endemic land mammal fauna by utilising specialist ancient DNA techniques. The following chapter supplies a background for this study; detailing the history of island studies in mammalian evolution, a geological and faunal review of the Caribbean region, and a brief introduction to phylogenetics and ancient DNA. I have also included a review of ancient DNA studies on micro-mammal fauna, which are the taxonomic focus of three of the four primary data chapters in this thesis. In the final section of this introductory chapter I set out the aims and expected outcomes of this research.

1.2 Island biogeography and evolution of insular species

Introduction to island study

This thesis focuses on land mammals endemically situated on the islands of the Caribbean. The primary aim of this study is not only to resolve the taxonomy of these species but also to answer evolutionary questions that are intricately linked to our understanding of island evolutionary biology and biogeography as a whole. There is a long history of using islands to study evolutionary processes, biogeography and extinction dynamics. Isolated from the mainland, islands have been described as "natural laboratories" (Losos & Ricklefs, 2009). They often act as microcosms for the mainland, where colonizing species and the evolutionary pressures acting on them can be disentangled in a closed system. Island systems are attractive for the study of evolution because, due to their isolation from the mainland and their high level of habitat diversity considering their often small size, islands are associated with high levels of endemism. Early island studies include Darwin's voyage of the Beagle (1839) and works on natural selection and evolutionary theory (1859) and Wallace's synchronous works on island evolution (1869), both of which played a fundamental role in our continued understanding of evolutionary concepts. Islands have also helped us to increase our understanding of species diversities, and extinction and colonization dynamics within ecosystems. MacArthur and Wilson (1963) developed a quantitative mathematical model for prediction of species diversity in a given isolated area - the equilibrium theory – which suggests that species diversity on insular systems is made up from new arrivals, and newly evolved species, minus the number of extinction events. This theory has found support in a study of Caribbean *Anolis*

lizards (Losos & Schluter, 2000). Increasingly studies of island evolutionary processes have used molecular phylogenetic methods, these methods can help to establish the colonization histories of groups and uncover the processes that drive insular species diversity (Emerson 2002).

Many factors affect the faunal composition of islands, including the degree of isolation or distance from the mainland, and island size. The diversity of habitats within an island and the number of other islands in an archipelago also affect species richness (Heaney 2000). Two types of islands can be distinguished: *i)* oceanic islands: which are relatively distant from the nearest mainland (to which they have never been connected) and originally arose from the sea floor often via volcanic activity, and *ii)* continental islands, which are found on the continental shelf and may have been historically connected to the mainland via land-bridges when sea levels were lower. These definitions are general and there are many exceptions; for example Madagascar is a continental island, but as it separated from Africa in the Cretaceous many of its attributes are more comparable to true oceanic islands. In the case of the Caribbean, islands of various ages and origins exist within the archipelago. The geological history of the islands is complex, and this impedes the generation of a comprehensive geological timeline. The question of whether any or all of the Greater Antillean islands were once connected to the mainland, for example, is still debated (Hedges, 2006; Iturralde-Vinent & MacPhee, 1999). Due to the long length of isolation, even including hypothetical land bridges, the Greater Antillean islands have many of the attributes of true oceanic type islands.

Island effects

Predators or interspecific competitors are often absent from island systems, and the climate, vegetation and geology of islands can generate an environment that is very different to the mainland. A colonizing species therefore often finds itself within a rather different selective environment to that experienced on the mainland, with rather different ecological constraints. Niche partitioning in founder sub-populations, can then drive speciation through allopatry (Losos & Queiroz, 1997). Within this context, one of the most dramatic forms of island evolution is insular size change. Here, the optimal size of the species is shifted due to a number of island associated factors; including the lack of predators, resource deprivation and neoteny. There are examples of large mammals evolving to a size much smaller than mainland relatives and small mammals evolving to much larger forms (VanValen 1973).

Although insular size change is an often cited result of island evolution, island species often exhibit other morphological adaptations to island life, such as changes in dentition associated with diet change and in limb morphology, with adaptations for different kinds of locomotion (Sondaar 1977). As colonizing species evolve to fill new and empty niches they may develop morphological features that are shared with mainland species that fill similar ecological niches (Van Der Geer, 2014). This adaptation can complicate phylogenetic analysis of island species that converge morphologically

with lineages they are only distantly related to, or conversely drastically differ in morphology from their closest mainland ancestors.

Whilst interspecific competition on islands is often decreased, intraspecific competition within the successful colonizing species is often increased (González-Suárez et al. 2015). Increased intraspecific competition on island systems may explain the fact that whilst islands often contain species that have dramatically deviated from mainland forms, they also often simultaneously support species with more derived or ancestral features.

In addition to size and shape change, a handful of studies have suggested that the rate of evolution may be faster on island systems. In one case, island dwarfism in large herbivores is proposed to occur over only a few thousand years (Lister 1996). In a broader review of existing data, Millien (2006) investigated island and mainland rodents to show that rates of morphological evolution on insular systems are significantly greater than in comparable mainland fauna, and later found that these rates were associated with island size (2011).

Finally, isolation of colonisers does not always lead to speciation and dramatic adaptations. In some cases no or very little change occurs, often the protection isolated islands provide allows these "living fossil" taxa to persist long after mainland forms become extinct. Examples include the Hispaniolan and Cuban *Solenodon* and recently extinct *Nesophontes* (Brace et al., 2016), Madagascan lemurs (Kistler et al. 2015), and giant tortoises found in the Galapagos and the Seychelles (Hansen et al. 2010) .

Mechanisms of speciation and adaptive and non-adaptive radiations

Island biologists often focus their research on how observed species diversity arose: the mechanisms of speciation. With a lack of interspecific competitors and new insular specific habitats, islands provide new niches for colonizing species to exploit. When a species reaches the island the variety of new niches that are open can generate adaptive radiations; the process wherein a species separates in to multiple species after morphological or behavioural adaptations to new environments prevent gene flow between populations (Losos & Ricklefs, 2009; Schluter, 2000; Whittaker & Frenandez-Palacios, 2007). As well as classical adaptive radiations, island systems are also known to generate non-adaptive radiations. Non-adaptive radiations are the result of one ancestor species splitting in to multiple descendent species, that are separated by non-adaptive means; such as species colonizing an island chain that are separated geographically, but may continue to occupy the same ecological niche (Rundell & Price 2009).

One often cited factor impacting island colonizing species is the founder effect - described as the loss of genetic variation, due to the small number of individuals that often establish isolated island

populations (Wright 1945). Founder effects can often increase the chances and rapidity of speciation between island and mainland populations. In addition to founder effects island populations often have small population sizes increasing the likelihood of inbreeding and the effects of genetic drift (Frankham 1998, Runemark et al 2010). Colonization of islands can also result in population level processes such as introgression or hybridization (gene flow from the gene pool of the colonizing species to the gene pool of the endemic species, followed by repeated backcrossing), whilst the variation and change in the levels of isolation of populations, within and between, islands can affect the genetic relationships between island species and populations (McGuire et al 2007).

Radiations on island systems are most often the result of allopatric speciation, where an archipelago of islands causes geographic separation of populations that then speciate by adaptive or non adaptive means. A famous example of this phenomena is Darwin's finches from the Galapagos archipelago, where congeneric species each utilise a different part of the habitat on multiple islands (Lack 1947). Allopatric speciation can also occur on a single island if there are significant geographical barriers to gene flow between populations, such has mountain ranges or water ways. Species generated by allopatric speciation can become secondarily sympatric, if those geographic barriers are later removed. In an island system sympatric species can be representatives of the same radiation, or alternately represent multiple colonization events. Peripatric speciation can result in similar species diversity to allopatric speciation, and occurs when a small peripheral population is isolated from the main population preventing gene flow. Peripatric speciation after geographic isolation is more likely because genetic drift acts more quickly in small populations. This mode of speciation is particularly relevant in insular systems where endemic species often have smaller population sizes.

Colonisation and extinction

Islands often have disproportionately less species diversity when compared to mainland systems, this can be attributed to low rates of colonization and high rates of extinction (MacArthur 1965). Two forms of colonization are generally recognised: land-bridge colonization and over-water colonization. Colonization from a land bridge via the mainland may include a wide spread of species that are occupying the area adjacent to the island at the point in time that the bridge exists, and represent a "snapshot" in time. Land-bridge colonizers may also include species that are not usually good dispersers such as carnivores and fresh water organisms (Zhou et al. 2006). Over-water dispersal or sweepstake dispersal (Simpson 1977) colonization often generates a more random faunal assembly with a bias towards species that are predisposed to dispersal or are semi-aquatic. Reverse colonization of the mainland by island fauna was originally assumed to be rare due the restrictions on invasion of mainland faunal communities and loss of dispersal ability. However this

phenomenon is now considered more common (Bellemain & Ricklefs 2008), and has been demonstrated in Caribbean *Anolis* lizards (Nicholson et al. 2005).

The timing of a colonization event can be determined using molecular clocks. Molecular clocks use known mutation rates or mutation rates paired with fossil evidence to estimate the time of divergence between species in the phylogeny, providing an estimate of when a colonizing species splits from its mainland ancestor (Donoghue & Benton, 2007). Molecular clock analysis can be calibrated using fossils or geological or geographic data, although these different types of data can sometimes contradict each other (Thorpe et al. 2005). Island emergence in particular can be used to give an earliest date for colonization and a lower limit for molecular clock analysis. If DNA data are not available, dating of fossil material or sediment from the layer a fossil is found gives an estimated earliest occurrence date for taxa.

Extinctions on island systems may be caused by a variety of factors. Competition with other species, either new colonizers or island endemics, can cause extinction of insular populations (MacArthur 1965). Natural environmental factors such as changes in climate may affect the habitat of an island species. Climate change may result in islands or habitats decreasing its size after a rise in sea-level. Climate induced habitat change may also increase the likelihood of species extinction (Courchamp et al. 2014). Although this factor might also affect mainland habitats, island species are particularly vulnerable to habitat change as movement to local refugia may not be possible. Island species are often restricted to small population sizes, that make extinction more likely in the event of natural disasters such as storms or volcanic eruptions (Purvis et al. 2000). Many island species become extinct as a consequence of human activity, this may be directly (e.g.: hunting (Turney et al. 2008)) or indirectly (e.g.: introduction of invasive species (Sax & Gaines 2008) and habitat destruction (Brooks et al. 2002)) related to human colonizers.

1.3 The Caribbean islands

The Caribbean Islands have been included in extensive studies of island biogeography and evolution; in lizards (Thomas et al. 2009), snakes (Burbrink et al. 2012), birds (Mila et al. 2010), toads (Alonso et al., 2012), insects (Seal et al. 2011) and bats (Davalos 2005). The Caribbean archipelago's composition of old and new, and small and large islands makes it ideal for studying evolutionary processes such as speciation.

Figure 1. Map of the Caribbean region. Dashed lines indicate Greater and Lesser Antilles and larger Greater Antillean islands are labelled.

Overview

The Caribbean islands are a tropical archipelago made up of the Greater Antilles: larger islands in the west (Bahamas, Cuba, the Cayman Islands, Hispaniola and Puerto Rico) and the smaller islands of the Lesser Antilles in the east. The Greater Antilles comprising fragments of continental crust and has gone through fragmentation and re-joined in various combinations throughout their journey to the present day formation (James 2005). The Greater Antillean islands originated to the west of the present day Isthmus of Panama until plate movements carried them to their current position (Ricklefs & Bermingham 2008). Parts of present day Hispaniola, Cuba and Puerto Rico may have remained above water throughout the Tertiary, and at least since the middle Eocene (Donnelly 1988). The proto-greater Antilles moved through the gap between Central and South America at the end of the Cretaceous (over 60 Myr ago) and it is possible that the proto-greater Antilles abutted what is now southern Mexico (Iturralde-Vinent & MacPhee 1999). Previous studies have suggested past connections between islands, and mainland bridges (Iturralde-Vinent & MacPhee, 1999). The Bahamas form a platform that has been *in situ* throughout the Tertiary (Mullins & Lynts 1977). The Lesser Antillean islands are made up of a volcanic Island arc that may have formed less than 20 Myr ago (Hedges, 2006). There is debate with regards to whether all or some of the insular landmasses remained permanently above sea level between the formation of

the Antilles in the Cretaceous and the Eocene (Hedges, 2006; Iturralde-Vincent, MacPhee, 1999; MacPhee, Iturralde-Vincent, 2005).

A brief timeline of the formation of the Caribbean islands (Iturralde-Vinent, 2006) (Figure 2)

Early Mid Jurassic (200-175 Ma): During the Early Middle Jurassic the Caribbean sea emerged in the form of a basin that developed between South and North America.

Cretaceous (145-65 Ma): The arc that formed during the Cretaceous most likely originated in the Pacific (Pendell 1994) and moved through a gap between North and South America, in what is now the Isthmus of Panama. This arc was above water at some points during this period, although due to the small size of the landmasses was most likely submerged shortly after.

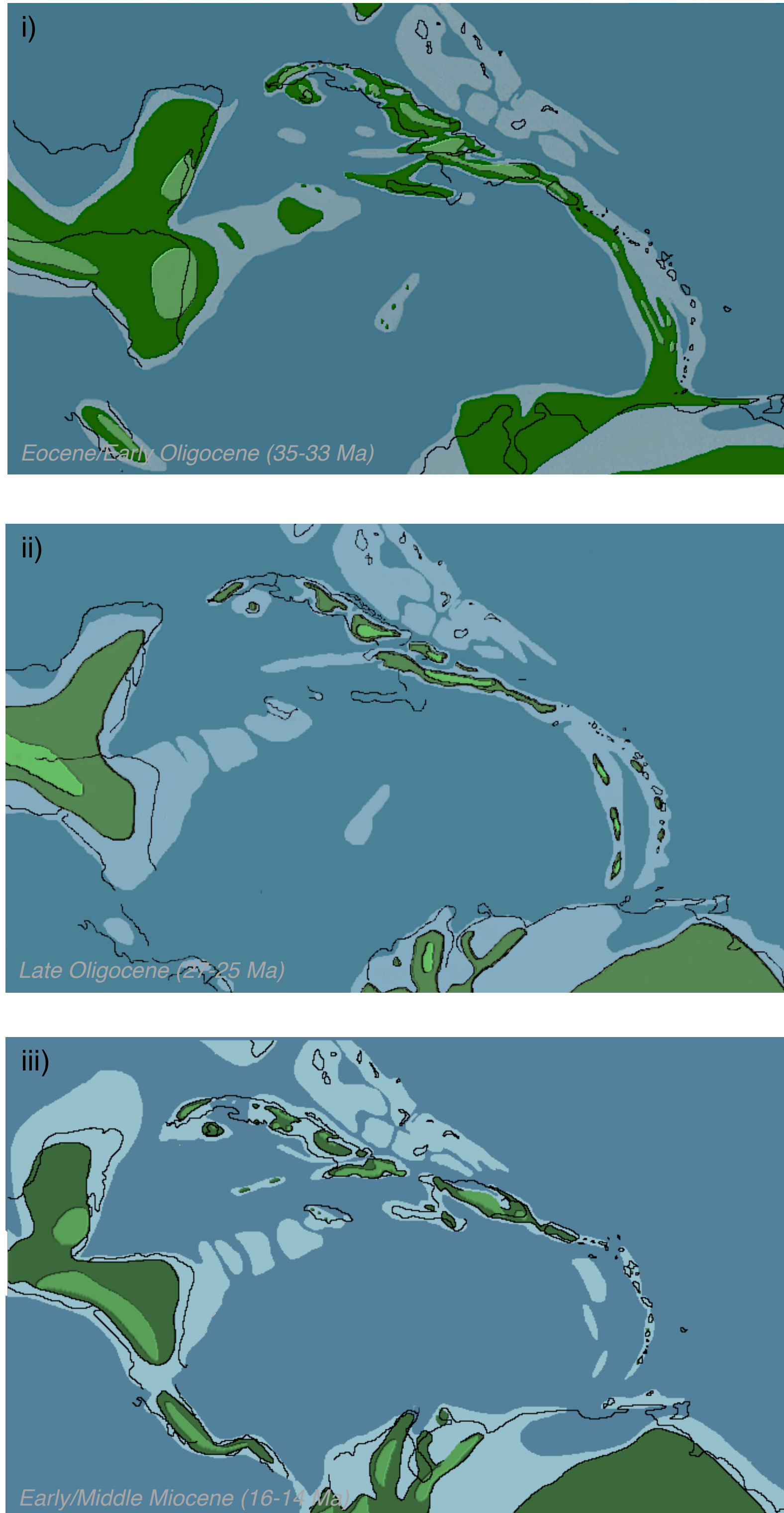
Palaeocene (66-56 Ma): Volcanic activity in the Palaeogene may have also produced above water landmasses, however these would have also likely submerged shortly after.

Eocene - Oligocene transition (~33 Ma): Tectonic uplift combined with low sea level, and caused above-water landmasses in the Caribbean sea to be at their most exposed. The now submerged Aves Ridge could have been emergent at this time possibly generating a land bridge between the Greater Antilles and South America. Also during this time west Cuba was separated from central and eastern Cuba by a seaway called the Havana-Matanza Channel.

Late Oligocene (28-23 Ma): During the late Oligocene sea levels rose and many landmasses exposed during the Eocene-Oligocene transition, including the proposed Aves Ridge, would have been submerged.

Early to Mid Miocene (23-13 Ma): During the Early to Mid Miocene disruption of the northern and southern Caribbean tectonic pate boundaries caused further isolation of the emergent land masses. Several present day trenches in the Caribbean sea formed during this time including: the Cayman Trench, Anagada Trench, Puerto Rico Trench and the Tortuga basin. Puerto Rico and Hispaniola may have still been connected at this point in time. During the Late to Mid Miocene the disappearance of the Havana-Matanza Channel would have allowed the reconnection of western Cuba with central and eastern Cuba.

Figure 2. Maps of the Caribbean region during three time periods: i) Eocene-Early Oligocene, ii) Late Oligocene and iii) Early/Middle Miocene adapted from Iturralde-Vinent (2006).



Climate in the wider Caribbean

Cyclical changes in climate and sea surface level during the Pleistocene affected species diversity and distributions worldwide, and are likely to have also affected the species that inhabited the Caribbean islands at this time. Oxygen isotopes from the Caribbean seafloor suggest sea surface temperature dropped by at least 3 or 4 degrees during glacial stages of the late Pleistocene (Donnelly 1994). The Caribbean islands were probably more arid during the late Pleistocene than they are at present. Evidence for this aridity in the Caribbean can be found in the now restricted distributions of species adapted to these dry habitats (such as: the red bellied woodpecker *Melanerpes superciliaris* (Olson & Hilgartner 1982)). Fossil specimens of some of these species are found in locations where living populations are no longer found. This suggests a reduction in the island species range, possibly mirroring the reduction of dry habitats as the climate of the Caribbean has changed (Olson & Hilgartner 1982). The same pattern is seen in South America, where due to a reduction in rainfall and humidity during the Pleistocene (Hewitt 2000) species adapted to wet tropical habitats were driven to refugia (Brown & Ab'Saber 1979). These refugia are thought to be important in Pleistocene speciation events. At 125,000 years ago, the end of the last interglacial sea surface levels were 8-10 m above the present level (Alt & Brooks 1965), meaning that some low laying islands or depressions may have been inundated with sea water during this period. Conversely at 17,000 years ago sea levels were 120 m below current levels possibly causing and increase in island size and removal of past biogeographic barriers (Gascoyne et al. 1979).

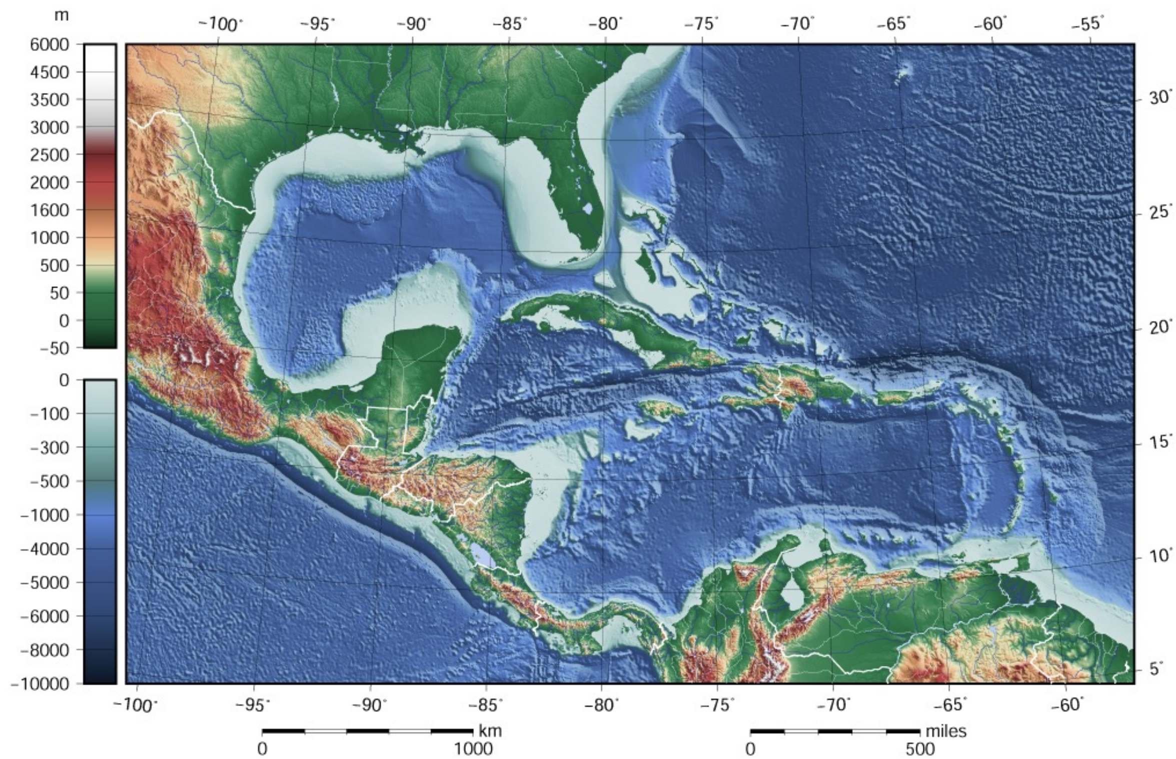


Figure 3. Bathymetric map of the Caribbean region (Hastings & Dunbar 1999; Amante & Eakins 2009)

Cuba

Many of the species looked at as part of this thesis are endemic to the Greater Antillean island of Cuba. The main island of Cuba is surrounded by smaller offshore islands; including the largest Isla de la Juventud (the Isle of Youth) or Isla de los Pinos (the Isle of Pines). The archipelago covers a total of 110,922 square km (Draper, 1994) and is the only part of the Greater Antilles that is situated on the North American (rather than Caribbean) plate (Iturralde-Vinent & Gahagan, 2002). The mountainous regions of the island are in the western province of Pinar Del Rio, the Escambray region in the centre of the island, and in the eastern province of Oriente. Several authors (Iturralde-Vinent & MacPhee 1999; Lewis & Draper 1990) have suggested that south-eastern Cuba and northern Hispaniola were once continuous, although the precise paleogeography of this theory remains unclear. The Canto Fault zone runs through the centre of the island and may represent the point of contact between eastern Cuba and northern Hispaniola (Gordon et al. 1997). Rapid uplift and erosion in the middle Eocene as part of the Caribbean plate's collision with the Bahamas platform may have been a trigger event for a separation of the Cuban and Hispaniolan palaeo-islands, if they were indeed connected. Geological reconstructions of Cuba have suggested that the island was previously made up of three separate archipelagos, between which there were deep seaways. These palaeo-islands were distinct through out the Cenozoic before coming together in the Pliocene (Iturralde-Vinent & MacPhee 1999; MacPhee et al 2003; Graham 2003). Fossil marine vertebrates found inland on central Cuba support the theory of historic submergence (MacPhee et

al 2003). Two palaeo-channels the Cardemas-Cochinos and Guacanayayabo-Nipe fault lines mark the historic boarder between western, central and eastern palaeo-islands. Pervious studies have looked at how partial submergence of these islands may have affected speciation events, one such study that looked at species and hybrids of green anoles lizards (Losos & Glor, 2003).

The Cayman Islands and the Cayman trench

Cayman Brac, Grand Cayman and Little Cayman make up the Cayman Islands situated on the Cayman ridge. This ridge forms the northern margin of the Cayman trench. The Cayman trench is approximately 1600km in length and km deep and rests north of Cuba, hypothetically acting as a faunal barrier between Cuba, Jamaica and Little Swan Island (Donnelly 1994). The timing of the opening of the Cayman trough is unresolved, although some studies suggest a mid-Eocene to early Miocene date for this geological event (Rosencrantz et al. 1986). The depth of the Cayman trench means that there were no land connections to the Cayman Islands during the Pleistocene and Pliocene, although it is possible that the Cayman islands were connected to one another and maybe even to the mainland of Cuba during the Miocene via the Cayman ridge (Fahlquist & Davies 1971). Grand Cayman and Little Cayman have similar surface topography and elevation above sea level, whereas the third island Cayman Brac dips to the southwest with a 50m rise at its north-eastern side. Both Grand Cayman and Little Cayman are low lying with approximately 50% of Grand Cayman presently covered in mangrove swamps (Parkinson et al. 1994). Peat layers provide a record of Holocene submergence of some areas of the island (Woodroffe 1981).

Hispaniola

The island of Hispaniola is divided politically into Haiti and the Dominican Republic and is the second largest landmass in the Caribbean with an area of about 80,00 square km (Lewis & Draper 1990). The island has four mountain ranges spreading from the northwest to the southeast these mountain ranges are separated by lower lying valleys. At 3,087m the twin peaks of Pico Duarte and La Pelona are the highest elevation in the Greater Antilles (Draper, Mann, & Lewis, 1994). In north-eastern and central Hispaniola the most prominent topographical structures were formed by middle to late Eocene collision between Hispaniola and the Bahamas platform. Hispaniola has a complex geological history culminating in three distinct biogeographic regions. Northern and southern palaeo-islands united in the late Miocene c- 10 Ma but remained at least partially separated by a depression known as the Neiba Valley that was periodically flooded to form a narrow sea channel until the Pleistocene. The southern region is further divided in to eastern and western regions by Bond’s line that forms a distinct biogeographic boundary (Maurrasse et al. 1980).

Puerto Rico

Puerto Rico was most likely in its present location by the late Oligocene to early Eocene (c- 21-13 Ma.) but only separated from Hispaniola later in the Eocene (Iturralde-Vinent, Gahagan, 2002). Subsidence occurred at several points between the middle Oligocene and middle Eocene and then between the middle Oligocene to Pliocene, submerging some parts of or the entire the island during these extended periods (Larue 1994). However, studies on vegetation on the island suggests that at least some areas were sub-aerial during these subsidence periods (Draper 1990).

Jamaica

Jamaica was formed as part of the Nicaraguan rise that extends eastwards from Honduras and Nicaragua, Jamaica and a few carbonate banks to the south of the island are the only parts of this feature that remain above water. Jamaica is the third largest of the Caribbean islands with and area of 11,264 square km plus 9,600 square km of offshore banks and shoals (Robinson 1994). However between the middle Eocene and middle Miocene differential subsidence caused the island to rise only a few meters above sea level (Draper, 1987). Late Pleistocene and Holocene rise in sea level has caused submergence the islands lower margins leading to the development of wetlands in coastal areas (Hendry & Digerfeldt 1989; Digerfeldt & Hendry 1987). There is no evidence that Jamaica was ever connected to any other landmass, at least after the Miocene, which precludes vicariance as a mechanism of distribution and in its production of species endemic to modern Jamaica.

1.4 The Caribbean land mammal fauna

Unlike many oceanic type archipelagos the Caribbean has been colonised by multiple groups of land mammals. This makes the Caribbean one of the few places where island biogeographic studies can incorporate mammals, and look at the way insular environments affect the evolution of these species. Much of the Caribbean's endemic mammal fauna is now extinct, obliterated during one of the largest mammalian extinction events of recent times (Davalos & Turvey, 2012). Therefore study of subfossil specimens is necessary to increase our understanding of the evolution of the Caribbean land mammals. This study focuses on three groups: rodents, eulipotyphlan insectivores and primates.

Rodentia

Rodents found in the Caribbean are part of what was once a diverse group containing several families. Oryzomyines are found in the Lesser Antilles (Brace, Turvey, Weksler, Hoogland, &

Barnes, 2015). Hystricognath caviomorph rodents, the hutia (Capromyidae) are found on the Greater Antillean islands. Whilst many species from this once diverse and speciose radiation are now extinct, Capromyidae contains some of the few survivors of the Caribbean's catastrophic Holocene extinction event. Living species in the genera *Capromys*, *Mesocapromys* and *Mysateles* are found on Cuba. Two species in the genus *Geocapromys*; *G. ingrahami* and *G. brownii* are found on two islands in the Bahamas and Jamaica respectively. The Hispaniolan hutia *Plagiodontia aedium* is the only living species found on the island of Hispaniola. Extinct taxa include: *Capromys* and *Geocapromys* species found on the Cayman islands, the Little Swan Island hutia *Geocapromys thoracatus*, and an extinct genus *Isolobodon* that was found on both Hispaniola and Puerto Rico. Recent studies have shown that hutia are part of the speciose neotropical spiny rat family Echimyidae (Fabre et al., 2014, 2016). This thesis also looks at extinct Cuban and Hispaniolan heterosomyines currently placed as a subfamily to the echimyids due to shared morphological features. *Elasmodontomys obliquus*, a single species from the now also entirely extinct group of giant hutias (Heptaxodontidae), is also examined.

Lipotyphla

Two families of lipotyphlan insectivores are known only from the insular Caribbean. Two surviving species are found within *Solenodon*, whilst the Holocene extinction event wiped out the entire family Nesophontidae. This thesis focuses on *Nesophontes* species found on Hispaniola. A recent study was able to resolve the relationship between *Solenodon* and *Nesophontes* using aDNA (Brace et al., 2016). However only a single *Nesophontes* species has ever been analysed at the molecular level, and intra-island evolutionary relationships between morphological species have yet to be explored.

Primates

Multiple extinct platyrrhine primate species have been described on the basis of morphology from the Greater Antillean islands in the Caribbean. This study focuses on the Jamaican *Xenothrix mcgregori*. Unique morphological traits including dental and postcranial features (MacPhee & Horovitz 2004; Horovitz & MacPhee 1999) have resulted in the attribution of *Xenothrix* and other Greater Antillean primates to several different mainland groups, and may be the result of island evolution (Cooke et al. 2011a).

Colonisation

Much of the previous work on the biogeography of the Caribbean has focused on the arrival of fauna. Two main competing theories have been discussed: *i)* over water dispersal via floating

vegetation (Darlington, 1938; Hedges, Hass, & Maxson, 1992), and *ii)* the arrival of fauna via one or multiple previously existing land bridges (Rosen 1985; Iturralde-Vinent & MacPhee 1999). Evidence for both theories relies upon the divergence dates between the Caribbean fauna and their closest mainland relatives that can be used as proxy earliest dates of possible colonization. Other forms of evidence for both theories include: the faunal composition of the Caribbean island endemics and the geographic origin of their mainland ancestors, and geological confirmation of the existence of a hypothetical land bridge.

The Aves ridge and the GAAR-landia hypothesis

The GAAR-landia (Greater Antilles + Aves Ridge) hypothesis suggests that land bridges such as the Aves Ridge could have connected South America with the Caribbean. The Aves Ridge is a now completely submerged ridge from Venezuela to the northern Lesser Antilles. Studies have postulated that the Aves Ridge may have formed a full land-bridge or "stepping stone" crossing at the Eocene-Oligocene transition (~34 Myr ago) for a 1-2 Myr interval (Rosen 1985; Iturralde-Vinent & MacPhee 1991). Molecular clock derived support for this geological feature is available in the form of Peltophryne toads from Cuba that were shown to have reached the Caribbean ~33 Ma (Alonso et al. 2012).

Over-water dispersal

The GAAR-landia hypothesis was disputed by Hedges (2001) who posited that over-water dispersal of fauna via flotsam rafts was a more likely explanation for the species diversity found in the Caribbean. The Caribbean islands are without large herbivores and carnivores. Fauna found on the islands are also characterised by high dispersal ability leading Hedges, among others to hypothesize that sweepstakes dispersal must be the mechanism of colonization. Additionally South America appears to be the source of many of the colonizers, which is consistent with recent Caribbean sea surface flow (Hedges, 2001). Present day over water dispersal and colonization of Caribbean islands has also been demonstrated in lizards (Censky, Hodge & Dudley., 1998). Recent molecular studies also seem to favour the over water dispersal theory, at least for the Caribbean's endemic rodents the capromyids (Fabre et al., 2014).

Extinction

The estimation of past species diversity in the Caribbean is made difficult by a combination of factors. Much of the Late Quaternary fossil record in the Caribbean is incomplete and the taxonomy uncertain (MacPhee et al. 2003). During the Holocene over 130 currently recognised species became extinct, leaving few survivors; ~14 surviving land mammal species (Turvey, Oliver, Narganes Storde, & Rye, 2007). A mixture of invasive species introductions (novel predators and

competitors) and habitat destruction (Steadman et al. 2007) are often cited as the cause of the most recent smaller mammal extinctions. Environmental change, either non-anthropogenic or following the arrival of Amerindians and early Europeans, are usually associated with the larger mammalian extinctions (Flemming & MacPhee, 1999; Turvey, 2009).

The late Pleistocene is associated with extreme environmental change. This dramatic change has been linked to regional faunal extinctions caused by factors characteristic of climate change: the switch from dry xeric habitats to more moist mesic ones, and decreases in island size and loss appropriate habitat due to sea level rise. One example of climate change induced extinction of an island endemic mammal, is the large rodent species *Amblyrhiza inudata*. This island giant is thought to have lost 93% of its island habitat during interglacial periods (MacPhee et al.,1998).

Human activity has been linked to many extinctions of island endemics (Diamond 1989). Although there is uncertainty surrounding the first colonization of the Caribbean by humans, there are thought to have been four waves, which can be distinguished by material culture: Lithic (~7000-4000 BP), Archaic (~5000-2000 BP), Ceramic (~2500 BP) and finally historical era colonization by Europeans (1492 CE) (Fitzpatrick & Keegan 2007).

The fundamental cause of extinctions can often be determined by comparing last occurrence dates of fossil taxa with possible causes such as the first occurrence dates of humans, invasive species or periods of dramatic climate change. A study by MacPhee (2009) showed that although the Caribbean mammal fauna started to decline when the first Amerindian colonizers arrived at the islands, many taxa continued to exist well after their arrival suggesting a drawn out “Sitzkrieg” extinction event, rather than the alternative hypothesis of “Blitzkrieg” style extinction. A study by Cooke *et al*., (2017) which combined new radiocarbon dating of fossil material and more accurate data on human colonisations also found that there was an overlap in the occurrence of endemic fauna and early human colonisation on most of the islands. This study also found that waves of human colonisations brought increasing extinction threats to the Caribbean land mammal fauna.

1.5 Conservation archaeogenomics

As described previously in this chapter the Caribbean region has seen unprecedented endemic biodiversity loss during the Holocene, with more mammalian species extinctions during this time period than any other region (Davalos & Turvey, 2012). Understanding the extinction dynamics in this region may provide useful insights that not only help to protect the surviving Caribbean mammal land fauna, but also increase our understanding of wider mammalian extinction patterns. This is particularly important in insular environments, where taxa are increasingly more vulnerable to extinction, due to extrinsic factors such as climate change and habitat loss. For extinction dynamics

to be properly understood species level taxonomy for the current and past diversity of the Caribbean needs to be accurately accessed.

This thesis looks at both extinct and extant species using ancient DNA methods. For many of the living Caribbean mammal taxa, sampling of live caught individuals is not possible as living populations are small or inhabit inaccessible locations. In these cases historic museum specimens, collected when population sizes were higher, can be a useful source of genetic information. Endangered, rare and isolated *Capromys* and *Mesocapromys* species and populations from specific localities can be found in the form of museum specimens (mounted specimens, skeletons and skins) allowing this study to sample from known populations, without invasive and expensive sampling of wild populations.

The endangered capromyids are not alone in this respect, as many endangered species lack comprehensive molecular taxonomic data, the collection of which is hindered by small population sizes and difficult to access species ranges. The development of increasing accurate and cost effective ancient DNA techniques will continue to allow molecular data to be extracted from degraded museum and zooacheological collections that can be used in conservation research. aDNA can be used for the purposes of estimation of historic population size, levels of gene flow and relationships between populations, all of which are vital to developing effective conservation and population management plans (Leonard 2008).

1.6 Taxonomy of Caribbean land mammal species

As previously mentioned in this chapter much of the Caribbean land mammal fauna is extinct, endangered or rare meaning that the description of species is often based on morphological traits described from museum or zooarchaeological specimens. As a result of the issues associated with morphological species descriptions based on fragmented remains and due to the ever changing nature of species delimitation, a reassessment of taxonomy, particularly inclusion of molecular analysis, is needed for many Caribbean mammal groups. Unfortunately due to the adverse preservation conditions of biomolecules in the tropics (see below), conventional techniques for extraction and sequencing (e.g. PCR) often can not provide enough the molecular data to inform for species definitions. Our understanding of Caribbean mammalian species diversity needs to be complete for the study of in-depth evolutionary questions surrounding speciation, colonization and extinction and for the conservation of endangered species. Of particular value in this endeavour is molecular data obtained using new ancient DNA techniques (e.g: NGS and target capture enrichment). These data can be used to contribute to our overall knowledge of species relationships, providing molecular analysis for species, that because of the state of specimens preservation, have previously only been studied in terms of morphology.

1.7 Ancient DNA

DNA sequence data extracted from blood, tissue or hair samples taken from modern organisms can be used for a variety of purposes from identifying species and providing a unified species concept, to understanding population dynamics such as abundance, distribution and population decline. However the identification of extinct or rare species using museum or zoo-archaeological specimens and answering the whole suite of questions which accompanies them, was until quite recently (late 1980s) (Higuchi et al. 1984) thought to be impossible. Ancient DNA refers to DNA recovered from specimens that have not been deliberately preserved or prepared for later use in molecular studies. This means that specimens used for the extraction of aDNA can range greatly in form and age; from 20th century museum collections of bones and skins to samples excavated from the permafrost. The generation of new techniques, in particular next-generation sequencing, has allowed progression from the analysis of short sections of genes from individual, recent (Holocene) samples to population level studies involving mitochondrial and even whole genomes (Willerslev & Cooper, 2005, Hagelberg *et al.,* 2015). The potential applications of aDNA (ancient DNA) studies are wide and varied from ancient human populations (Hofreiter, Serre, Poinar, Kuch, & Pääbo, 2001), to historical pathogens (Donoghue et al., 2004), to studies that contribute to conservation (Leonard 2008).

Ancient DNA degradation and preservation

As time passes damage to DNA increases. Two forms of damage to the molecular data recovered from specimens exists: *i*) damage caused by enzymatic processes (e.g.: microbial damage) and *ii*) damage caused by chemical reactions such as hydrolysis and oxidation, which can no longer be repaired after an organism has died. This means that the age of a specimen increases so does the damage to any potential aDNA fragments (Mitchell et al. 2005) although age is not the only predictor of damage. aDNA techniques attempt to recover nucleotide sequence information from DNA molecules that have not been completely destroyed by these processes. Particularly damaging are the release of cellular enzymes, that after death can cause degradation to DNA (Paabo et al., 2004). Whilst an organism is living, DNA molecules are conserved by enzymatic repair processes, after death these enzymes are broken down, with no counter-balance DNA molecules are degraded resulting in permanent loss of nucleotide sequence information. Post-mortem enzymatic processes and non-enzymatic hydrolytic cleavage of phosphodiester bonds in the phosphate sugar backbone can lead to the reduction of average fragment size in aDNA. aDNA exhibits other damage aside from fragmentation in the form of DNA modifications. These modifications may still allow for the amplification of the template molecules but incorporate incorrect nucleotide bases in the resulting sequence data. Hydrolytic reactions can cause the deamination, depurination and depyrimidination of bases (Hoss et al., 1996). Deamination of cytosine to uracil, 5-methyl-cytosine to thymine and adenine to hypoxathine cause transitions (C -> T and G -> A) to be incorporated in to DNA strands.

A large number of these C to T and G to A changes have been shown in aDNA studies (Gilbert et al., 2007).

The rate of DNA degradation over time is affected by a number of factors outside of the deceased organism, such as the environment a specimen decays in or specimen preservation and storage conditions. With the publication of various aDNA studies it has become clear that consistent low temperatures provide the best conditions for preservation of DNA (Willerslev et al., 2007) Sampling from polar environments or from inside caves are ways to lower potential damage caused by degradation.

Ancient DNA contamination

Early studies in the field of aDNA highlighted an issue that was going to become one of the main hurdles in the extraction and sequencing of ancient molecular data. Due to the rarity and degradation of aDNA one of the main issues surrounding degraded molecular data is the contamination of samples. Environmental DNA can be thousands of times more amplifiable than an aDNA sample (Mitchell *et al.,* 2005). Bacterial colonisation of organisms after death mean that there is often a high level of bacterial DNA in aDNA samples. Other living species in the environment can also add to the contamination of the sample, including species found in cave sites where zoo-archaeological specimens are excavated and species occupying museums or specimen storage facilities. Other forms of contamination can take place during the extraction and sequencing of aDNA samples. Laboratory equipment and reagents used for extraction and library builds can be exposed to the same sources of contamination mentioned above. Extraction and pre-sequencing preparation of aDNA samples often takes place in universities or research institutions where molecular work on modern samples is also taking place, this introduces the risk of contamination of aDNA samples by the highly-amplified products of PCR reactions. The inclusion of an extremely small amount of PCR product in aDNA samples can be carried through the sequencing process and be included in the data analysed.

Specialist aDNA laboratories and protocols are used to limit the amount of contamination. These protocols include the separation of pre- and post-PCR laboratories. As well as being separated, pre-PCR laboratories used for aDNA research are not utilised for extraction of modern samples. Equipment and work benches in aDNA laboratories are cleaned before and after use using bleach and subjected to UV treatment. Controls are used in both extraction and pre-sequencing preparation of samples in order to detect any potential sources of contaminants (Knapp et al. 2012; Fulton 2012). In addition of anti-contamination measures taken pre-sequencing all sequencing data must be thoroughly checked for contamination using the BLAST database (Basic Local Alignment Search Tool; Altschul et al., 1997).

Ancient DNA in the tropics

As mentioned in the section above, the environment in which an organism dies greatly affects the condition of surviving DNA, leading to a taphonomic bias. Climate in particular appears to be a good predictor for the preservation of aDNA. Degradation of DNA is known to be greatly accelerated in warm tropical or subtropical regions (Hofreiter et al., 2015; Smith et al., 2001). The reason for this bias in DNA degradation lays in the chemical reactions that occur after death. The factors that most affect hydrolytic depurination in degraded DNA are pH and temperature. The differentiation between DNA preservation at different temperatures means that some sites are more likely to produce higher endogenous DNA content than others. In order to access the chances of successful aDNA extraction from specimens from different sites a thermal age is estimated. The thermal age adjusts the temporal age of different sites according to their individual thermal histories, using the known temperature dependence of DNA depurination (Smith et al. 2003).

The bias of higher DNA degradation in tropical regions poses a particular problem for this study as all species included in this research are from the tropical region of the Caribbean. In tropical environments certain factors can help to limit the damage done by increased temperature. For example cave sites are known to greatly increase the probability of endogenous DNA survival as they provide a more constraint cooler temperature in comparison with exposed landscapes (Gutiérrez-garcía et al., 2014; Hofreiter et al., 2015). DNA extracted from 1000 year old tortoise remains excavated from a sink-hole environment on the Bahamas suggests that anoxic, thermally buffered marine environments may also increase the survival of endogenous DNA from the tropics (Kehlmaier et al. 2016).

Although the tropical nature of this region results in difficulties extracting and sequencing DNA from even recently collected museum specimens (< 200 years), previous studies have been successful (Brace et al., 2015; Brace et al., 2016; Fabre, Vilstrup, Raghavan, Sarkissian, Willerslev, Douzery, & Orlando, 2014; Kehlmaier et al., 2016; Schroeder et al., 2015). The introduction of Next Generation Sequencing techniques (Guimaraes et al. 2016a) and target capture enrichment (Mohandesan et al. 2016) have been shown to increase the likelihood of successful extraction and sequencing of endogenous aDNA from tropical environments.

Next Generation Sequencing

Next-generation sequencing (NGS) is a general term for several different modern sequencing technologies that parallelize sequencing enabling them to produce thousands or millions of sequences at the same time (Mardis 2008). This has decreased the time and cost of sequencing from previously used Sanger sequencing exponentially over the last 10 years (Schuster 2007). The vast amounts of data produced by NGS along with the short read lengths produced by these

sequencing systems, make it an effective tool for aDNA sequencing. NGS has allowed the sequencing of whole mitochondrial genomes and the first functional genes from extinct species (Knapp & Hofreiter 2010).

NGS technologies have revolutionized the field of ancient DNA but also brought a whole host of new problems. One problem with using NGS technologies in the study of aDNA is the low quantity of endogenous DNA that is found in aDNA samples. Because of the large amount of data produced by NGS, endogenous DNA can sometimes be lost or diluted amongst the high amount of environmental contaminants. The large amount of data produced by NGS means that one of the largest hurdles for aDNA studies is retrieving endogenous aDNA from the output containing a high percentage of contaminants, this task is especially difficult if no molecular data for a close “reference” species exists. Another problem encountered when sequencing aDNA using NGS is the fact that the aDNA is very fragmented. This can mean that there are problems aligning the sequence data to reference sequences in post-sequencing data processing. NGS sequencing produces high volumes of data but is not targeted to particular regions of sequences. Because NGS is high output and non-targeting, another problem associated with the technology is low complexity associated with number of reads per position. Certain areas may yield high coverage and duplications, whereas other regions may not be sequenced at all.

Target capture enrichment

One way of ensuring higher endogenous aDNA yield from high output and non-targeting NGS is target capture enrichment. The basis of this technique is to target specific genes or multiple genes from the study species before sequencing takes place. This is achieved by selectively enriching sequences using probes representing regions of interest, enrichment is carried out prior to the sequencing run giving an increased depth of sequencing on target areas (Enk et al. 2014). Target capture enrichment has been utilised as part of this study through the use of MYbaits kits ordered from the company MYcroarray (www.mycroarray.com). These kits are used for in-solution hybridization capture of the targeted DNA, and use biotinylated RNA probes that are complementary to the target sequences. The ready-made capture kits include baits that are manufactured to order, using the most closely related species, and the target specific genes used in the analysis. These kits are routinely used in aDNA studies as an affective method of targeting the study species, thus decreasing problems in post-sequencing data processing and analysis associated with contamination

1.8 Thesis Aims

This thesis looks at the unique, endangered and extinct Caribbean land mammal fauna. The extraction and sequencing of molecular data for the purposes of this study is complicated by: the

rarity of some specimens of taxa that are extinct or have low population sizes, the degradation of DNA in tropical environments such as the Caribbean, and the lack of data already available for these species and closely related taxa. Newly developed techniques in the extraction and sequencing of aDNA will be used and trialed in order to combat the issues associated with aDNA extraction from the taxa included in this study. The data generated by this study will then be used to test biogeographic and evolutionary hypotheses in the Caribbean land mammal fauna. Below are the main aims of this thesis:

i) *To utilize recently-developed aDNA techniques, including NGS and target capture enrichment, to extract and sequence highly degraded DNA from museum and zooacheological specimens from the tropical Caribbean.*

ii) *To generate a phylogenetic placement for some recently extinct Caribbean species, where the evolutionary history is currently unresolved or highly debated. The use of aDNA is the only means for generating the first genetic phylogeny including these taxa. For living taxa included in this study, molecular data produced will contribute to redefining species relationships to better inform conservation efforts.*

iii) *To use aDNA data to test long standing biogeographic hypotheses relating to the colonisation of the Caribbean by various land mammal fauna. Molecular analyses are imperative here, as they not only provide a statistically robust framework to identify extant sister taxa, and therefore the mainland region of origin, molecular data also allows the use of molecular clocks to date the divergence between mainland and island species.*

iv) *This study will investigate intra-island and wider inter-island relationships among species to explore evolutionary histories and the mechanisms of speciation acting on the Caribbean's endemic land mammal fauna.*

The chapters in this thesis are unified by the common goal of uncovering the genetic affinities between Caribbean island mammal taxa. This information if vital for the continued study of evolutionary questions in the context of island mammals.

In order to achieve these aims I have studied multiple groups of endemic Caribbean land mammal fauna including both living and extinct taxa. The chapters which follow explore the biogeography and evolutionary relationships of the extinct Hispaniolan insectivore genus *Nesophontes*, the extinct primate *Xenothrix Mcgregori,* Cuban hutia (Capromyidae), and finally the wider Hystricognath rodent radiation within the Caribbean including hutia (Capromyidae), the extinct heterosomyines and the giant hutia: heptaxodontids.

References

Alonso, R., Crawford, A.J. & Bermingham, E., 2012. Molecular phylogeny of an endemic radiation of Cuban toads (Bufonidae: *Peltophryne*) based on mitochondrial and nuclear genes. *Journal of Biogeography*, 39(3), pp.434–451.

Alt, D. & Brooks, H.K., 1965. Age of Florida marine terraces. *Journal of Geology*, 73, pp.406–11.

Altschul, S.F. et al., 1997. Gapped BLAST and PSI-BLAST: a new generation of protien database search programs. *Nucleic Acids Research*, 25, pp.3389–3402.

Amante, C. & Eakins, B.W., 2009. *ETOPO1 1 Arc-Minute Global Relief Model Procedures, Data Sources and Analysis*, Technical Memorandom NESDIS NGDC-24 National Geophysical Data Center, NOAA.

Bellemain, E. & Ricklefs, R.E., 2008. Are islands the end of the colonization road? *Trends in Ecology and Evolution*, 23(8), pp.461–468.

Brace, S. et al., 2016. Evolutionary history of the Nesophontidae, the last unplaced Recent mammal family. *Molecular biology and evolution*, 33(12), p.msw186.

Brace, S. et al., 2015. Unexpected evolutionary diversity in a recently extinct Caribbean mammal radiation. *Proceedings of the Royal Society B: Biological Sciences*, 282(1807), pp.20142371–20142371.

Brooks, T.M. et al., 2002. Habitat loss and extinction in the hotspots of biodiversity. *Conservation Biology*, 16(4), pp.909–923.

Brown, K.S. & Ab'Saber, A.N., 1979. Ice-age forest refugia and evolution in the neotropics; correlaton of paleoclimatological, geomorphological and paleoecological data with modern biological endermism. *Paleoclimas*, 5, pp.1–30.

Burbrink, F.T., Ruane, S. & Pyron, R.A., 2012. When are adaptive radiations replicated in areas? Ecological opportunity and unexpected diversification in West Indian dipsadine snakes (Colubridae: Alsophiini). *Journal of Biogeography*, 39(3), pp.465–475.

Censky Hodge, K., Dudley, J., E.L., 1998. Over-water dispersal of lizards due to hurricanes. *Nature*, 395(October), p.556.

Cooke, S.B. et al., 2017. Anthropogenic Extinction Dominates Holocene Declines of West Indian Mammals. *Annual Review of Ecology, Evolution, and Systematics*, 48(1).

Cooke, S.B., Rosenberger, A.L. & Turvey, S., 2011. An extinct monkey from Haiti and the origins of the Greater Antillean primates. , 108(7).

Courchamp, F. et al., 2014. Climate change, sea-level rise, and conservation: Keeping island biodiversity afloat. *Trends in Ecology and Evolution*, 29(3), pp.127–130.

Darlington, P.J., 1938. The origin of the Fauna of the Greater Antilles, with discussion of the dispersal of animals over water and through the air. *The Quarterly Review of Biology*, 13(3), pp.274–300.

Darwin, C., 1859. *On the origin of species by natural selection*, New York: P. F. Collier & Son.

Darwin, C., 1839. *The voyage of the Beagle: Journal of reseraches into the natural history and*

geology of the countries visited during the voyage of HMS Beagle round the world, Wordsworth Editions.

Davalos, L.M., 2005. Molecular phylogeny of funnel-eared bats (Chitoptera: *Natalidae*), with notes on biogeography and conservation. *Molecular Phylogenetics and Evolution*, 37(1), pp.91–103.

Davalos, L.M. & Turvey, S.T., 2012. West Indian Mammals: The old the new and the recently extinct. In *Bones, Clones, and Biomes: The History and Geography of Recent Neotropical Mammals*. pp. 157–202.

Diamond, J.M., 1989. The present, past and future of human-caused extinctions. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 325, pp.469–477.

Digerfeldt, G. & Hendry, M.D., 1987. An 8000 year Holocene sea-level record from Jamaica: implications for interpretation of Caribbean reef and coastal history. *Coral Reefs*, 5(4), pp.165–169.

Donnelly, T.W., 1988. Geologic constraints on Caribbean biogeography. *Zoogeography of Caribbean insects*, 15, p.37.

Donnelly, T.W., 1994. The Caribbean Sea Floor. In *Caribbean Geology: An Introduction*. Kingston: U.W.I Publishers Association, pp. 41–60.

Donoghue, H.D. et al., 2004. Tuberculosis: from prehistory to Robert Koch, as revealed by ancient DNA. *The Lancet infectious diseases*, 4(9), pp.584–592.

Donoghue, P.C.J. & Benton, M.J., 2007. Rocks and clocks: calibrating the Tree of Life using fossils and molecules. *Trends in Ecology and Evolution*, 22(8), pp.424–431.

Draper, G., 1987. A revised tectonic model for the evolution of Jamaica. In *Proceedings workshop on the status of Jamaican geology*. Kingston: Geological Society of Jamaica, pp. 151–169.

Draper, G., 1994. Cuba. In *Caribbean Geology: An introduction*. Kingston: U.W.I Publishers Association, pp. 65–82.

Draper, G., Mann, P. & Lewis, J., 1994. Hispaniola. In *Caribbean Geology: An introduction*. Kingston: U.W.I Publishers Association, pp. 129–145.

Emerson, B.C., 2002. Evolution on oceanic islands: Molecular phylogenetic approaches to understanding pattern and process. *Molecular Ecology*, 11(6), pp.951–966.

Enk, J.M. et al., 2014. Ancient whole genome enrichment using baits built from modern DNA. *Molecular Biology and Evolution*, 31(5), pp.1292–1294.

Fabre, P.-H., Vilstrup, J.T., Raghavan, M., Sarkissian, C. Der, Willerslev, E., Douzery, E.J.P., Orlando, L., et al., 2014. Rodents of the Caribbean: origin and diversification of hutias unravelled by next-generation museomics. *Biology Letters*, 10(7).

Fabre, P.-H., Vilstrup, J.T., Raghavan, M., Sarkissian, C. Der, Willerslev, E., Douzery, E.J.P. & Orlando, L., 2014. Rodents of the Caribbean: origin and diversification of hutias unravelled by next-generation museomics. *Biology letters*, 10(7), p.20140266-.

Fabre, P.H. et al., 2017. Mitogenomic phylogeny, diversification, and biogeography of South American spiny rats. *Molecular Biology and Evolution*, 34(3), pp.613–633.

Fahlquist, D.A. & Davies, D.K., 1971. Fault-block origin of the western Cayman Ridge, Caribbean

Sea. *Deep-Sea Research and Oceanographic Abstracts*, 18(2), pp.243–253.

Fitzpatrick, S.M. & Keegan, W.F., 2007. Human impacts and adaptations in the Caribbean Islands: an historical ecology approach. *Earth and Environmental Science Transactions of the Royal Society of Edinburgh*, 98(1), pp.29–45.

Flemming, C. & MacPhee, R.D.E., 1999. Redetermination of holotype of *Isolobodon portoricensis* (Rodentia, Capromyidae):with notes on recent mammalian extinctions in Puerto Rico. *American Museum Novitates*, 3278.

Fulton, T.L., 2012. Setting Up an ancient DNA laboratory. *Methods in Molecular Biology*, 840, pp.1–11.

Frankham, R., 1998. Inbreeding and extinction: island populations. *Conservation biology*, *12*(3), pp.665-675.

Gascoyne, M., Benjamin, G.J. & Schwarz, H.P., 1979. Sea-level lowering during the Illinoian glaciation: evidence from a Bahama “blue hole.” *Science*, 205(806), p.8.

Van Der Geer, A. a E., 2014. Parallel patterns and trends in functional structures in extinct island mammals. *Integrative Zoology*, 9(2), pp.167–182.

González-Suárez, M., Bacher, S. & Jeschke, J.M., 2015. Intraspecific Trait Variation Is Correlated with Establishment Success of Alien Mammals. *The American Naturalist*, pp. 537-

Gordon, M.B., Mann, P., Cáceres, D., & Flores, R., 1997. Cenozoic tectonic history of North America-Caribbean plate boundary zone in western Cuba. *Journal of Geophysical Research*, 102, pp.10–55.

Guimaraes, S. et al., 2016. A cost-effective high-throughput metabarcoding approach powerful enough to genotype ~44 000 year-old rodent remains from Northern Africa. *Molecular Ecology Resources*.

Gutiérrez-garcía, T. a et al., 2014. Ancient DNA and the tropics : a rodent’s tale. *Biology letters*, 10(6), p20140224

Hagelberg, E., Hofreiter, M. & Keyser, C., 2015. Introduction. Ancient DNA: the first three decades. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 370(1660), p.20130371.

Hansen, D.M., Donlan, C. J., Griffiths, C. J., & Campbell, K. J., 2010. Ecological history and latent conservation potential: Large and giant tortoises as a model for taxon substitutions. *Ecography*, 33(2), pp.272–284.

Hastings, D. & Dunbar, P.K., 1999. *Global Land One-Kilometer Base Elevation (GLOBE) v. 1.*, National Geophysical Data Center, NOAA.

Heaney, L.R., 2000. Dynamic Disequilibrium:A Long-Term, Large-Scale Perspective on the Equilibrium Model of Island Biogeography. *Global Ecology and Biogeography*, 9, pp.59–74.

Hedges, B.S., 2006. Paleogeography of the Antilles and origin of West Indian terrestrial vertebrates. *Annals of the Missouri Botanical Garden*, pp.231–244.

Hedges, S.B., 2001. Biogeography of the West Indies: An overview. *Biogeography of the West Indies: Patterns and perspectives*, pp.15–33.

Hedges, S.B., 2006. Paleogeography of the Antilles and Origin of West Indian Terrestrial Vertebrates. *Annals of the Missouri Botanical Garden*, 93(2), pp.231–244.

Hedges, S.B., Hass, C. a & Maxson, L.R., 1992. Caribbean biogeography: molecular evidence for dispersal in West Indian terrestrial vertebrates. *Proceedings of the National Academy of Sciences of the United States of America*, 89(5), pp.1909–1913.

Hendry, M. & Digerfeldt, G., 1989. Palaeogeography and palaeoenviroments of a tropical coastal wetland and offshore shelf during Holocene submergence, Jamaica. *Palaeogeography, palaeoclimatology, palaeoecology*, 73(1), pp.1–10.

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

Higuchi, R. et al., 1984. DNA sequences from the quagga, an extinct member of the horse family. *Nature*, pp.282–284.

Hofreiter, M. et al., 2001. Ancient DNA. *Nature reviews. Genetics*, 2(5), pp.353–359.

Hofreiter, M. et al., 2015. The future of ancient DNA: Technical advances and conceptual shifts. *BioEssays*, 37(3), pp.284–293.

Horovitz, I. & MacPhee, R.D., 1999. The Quaternary Cuban platyrrhine *Paralouatta varonai* and the origin of Antillean monkeys. *Journal of human evolution*, 36(1), pp.33–68.

Iturralde-Vinent, M. A. Gahagan, L., 2002. Latest Eocene to Middle Miocene Tectonic Evolution of the Caribbean: Some Principles and their Implications for Plate Tectonic Modelling. In *Caribbean Geology into the third Millennium: Transactions of the Fifteenth Caribbean Geological Conference*. pp. 47–62.

Iturralde-Vinent, M.A., 2006. Meso-Caribbean paleogeography:implications for the historical biogeography of the region. *International Geology Review*, 48(9), pp.791–827.

Iturralde-Vinent, M.A. & MacPhee, R.D.E., 1999. Paleogeography of the Caribbean region: implications for Cenozoic biogeography. *American Museum of Natural history*.

James, K.H., 2005. A simple synthesis of Caribbean geology. *Caribbean Journal of Science*, 39, pp.69–82.

Kehlmaier, C. et al., 2016. Tropical ancient DNA reveals relationships of the extinct Bahamian giant tortoise *Chelonoidis alburyorum*. *Proceedings of the Royal Society B: Biological Sciences*, 21(16), pp.1–7.

Kistler, L. et al., 2015. Comparative and population mitogenomic analyses of Madagascar's extinct, giant "subfossil" lemurs. *Journal of Human Evolution*, 79, pp.45–54.

Knapp, M., Clarke, A. C., Horsburgh, K. A., & Matisoo-Smith, E. A., 2012. Setting the stage - Building and working in an ancient DNA laboratory. *Annals of Anatomy*, 194(1), pp.3–6.

Knapp, M. & Hofreiter, M., 2010. Next generation sequencing of ancient DNA: Requirements, strategies and perspectives. *Genes*, 1(2), pp.227–243.

Lack, D., 1947. *Darwin's finches*, Cambridge, UK: Cambridge Univ. Press.

Larue, D., 1994. Puerto Rico and the Virgin Islands. In *Caribbean Geology: An Introduction*. Kingston: U.W.I Publishers Association, pp. 151–161.

Leonard, J.A., 2008. Ancient DNA applications for wildlife conservation. *Molecular Ecology*, 17(19),

pp.4186–4196.

Lewis, J.F. & Draper, G., 1990. Resume of the geology of Hispaniola. In *Field Guide, Ninth Caribbean Geological conference*. Santo Domingo, Dominican Republic, pp. 5–31.

Lister, A.M., 1996. Dwarfing in island elephants and deer: Processes in relation to time of isolation. *Symposium of the Zoological Society of London*, 69, pp.277–292.

Losos, J.B. & Glor, R.E., 2003. Phylogenetic comparative methods and geography of speciation. *Trends in Ecology & Evolution*, 18, pp.220–227.

Losos, J.B. & Queiroz, K. De, 1997. Evolutionary consequences of ecological release in Caribbean *Anolis* lizards. *Biological Journal of the Linnean Society*, 61(4), pp.459–483.

Losos, J.B. & Ricklefs, R.E., 2009a. Adaptation and diversification on islands. *Nature*, 457(7231), pp.830–6.

Losos, J.B. & Ricklefs, R.E., 2009b. Adaptation and diversification on islands. *Nature*, 457(7231), pp.830–836.

Losos, J.B. & Schluter, D., 2000. Analysis of an evolutionary species-area relationship. *Nature*, 408(6814), pp.847–850.

MacArthur .R, 1965. Patterns of Species Diversity. *Biological Reviews of the Cambridge Philosophical Society*, 40(4), p.510-.

MacArthur & Wilson, 1963. An equilibrum theory of insular zoogeography. *Evolution*, pp.373–387.

MacPhee, R.D.E. & Horovitz, I., 2004. New Craniodental Remains of the Quaternary Jamaican Monkey Xenothrix mcgregori (Xenotrichini, Callicebinae, Pitheciidae), with a Reconsideration of the Aotus Hypothesis. *American Museum Novitates*, 3434, pp.1–51.

MacPhee, R.D.E., Iturralde-Vinent, M.A. & Gaffney, E.S., 2003. Domo de Zaza, an Early Miocene vertebrate locality in South-Central Cuba, with notes on the tectonic elolution of Puerto Rico and the Mona Passage. *American Museum Novitates*, pp.1–42.

Mardis, E.R., 2008. The impact of next-generation sequencing technology on genetics. *Trends in Genetics*, 24(3), pp.133–141.

Maurrasse, F., Pierre-Louis, R. & Rigaud, J., 1980. Cenozoic facies distribution in the southern peninsula of Haiti and the Barahona Peninsula, Dominican Republic, and it's relations concerning tectonic evolution of the La Selle-Baoruco block. *Caribbean Geology, Collected Contributions*, (9), pp.1–24.

McGuire, J. A., Linkem, C. W., Koo, M. S., Hutchison, D. W., Lappin, A. K., Orange, D. I., Lemos-Espinol, J., Riddle, B. R., Jaeger, J. R. (2007). Mitochondrial introgression and incomplete lineage sorting through space and time: phylogenetics of crotaphytid lizards. *Evolution* 61: 2879-2897

Mila, B.W. et al., 2010. The geographic scale of diversification on islands: Genetic and morphometric divergence at a very small spatial scae in the Mascarene grey white-eye (Aves: *Zosterops bordonicus*). *BMC evolutionary biology*, 10(1), p.158.

Millien, V., 2011. Mammals evolve faster on smaller islands. *Evolution*, 65(7), pp.1935–1944.

Millien, V., 2006. Morphological evolution is accelerated among island mammals. *PLoS Biology*,

4(10), pp.1863–1868.

Mitchell, D., Willerslev, E. & Hansen, A., 2005. Damage and repair of ancient DNA. *Mutation research*, 571(1–2), pp.265–276.

Mohandesan, E. et al., 2016. Combined hybridization capture and shotgun sequencing for ancient DNA analysis of extinct wild and domestic dromedary camel. *Molecular Ecology Resources*.

Mullins, H.T. & Lynts, G.W., 1977. Origin of the northwestern Bahama Platform: Review and reinterpretation. *Geological Society of America Bulletin*, 88(10), pp.1447–1461.

Nicholson, K.E. et al., 2005. Mainland colonization by island lizards. *Journal of Biogeography*, 32(6), pp.929–938.

Olson, S.L. & Hilgartner, W., 1982. Fossil and subfossil birds from the Bahamas. *Smithsonian Contributions to Paleobiology*, 48.

Parkinson, R.W., Delaune, R.D. & White, J.R., 1994. Holocene Sea-Level Rise and the Fate of Mangrove Forests within the Wider Caribbean Region. *Journal of Coastal Research*, 10(4), pp.1077–1086.

Purvis, A., Gittleman, J. L., Cowlishaw, G., & Mace, G. M., 2000. Predicting extinction risk in declining species. *Proceedings of the Royal Society B: Biological Sciences*, 267(1456), pp.1947–1952.

Ricklefs, R. & Bermingham, E., 2008. The West Indies as a laboratory of biogeography and evolution. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 363(1502), pp.2393–2413.

Robinson, E., 1994. Jamaica. In *Caribbean Geology: An introduction An In*. Kingston: U.W.I Publishers Association, pp. 111–122.

Rosen, D.E., 1985. Geological hierarchies and biogeographic congruence in the Caribbean. *Annals of the Missouri Botanical Garden*, 72, pp.636–659.

Rosencrantz, E., Ross, M.I. & Sclater, J.G., 1986. Age and spreading of the Cayman Trough as determined from depth, heat flow and magnetic anomalies. *Journal of Geophysical Research*, 93, pp.2141–2157.

Rundell, R.J. & Price, T.D., 2009. Adaptive radiation, nonadaptive radiation, ecological speciation and nonecological speciation. *Trends in Ecology and Evolution*, 24(7), pp.394–399.

Runemark, A., Hansson, B., Pafilis, P., Valakos, E.D. and Svensson, E.I., 2010. Island biology and morphological divergence of the Skyros wall lizard Podarcis gaigeae: a combined role for local selection and genetic drift on color morph frequency divergence?. *BMC evolutionary biology*, *10*(1), p.269.

Sax, D.F. & Gaines, S.D., 2008. Species invasions and extinction: The future of native biodiversity on islands. *Proceedings of the National Academy of Sciences*, 105(Supplement 1), pp.11490–11497.

Schluter, D., 2000. *The Ecology of Adaptive radiation*, OUP Oxford.

Schroeder, H. et al., 2015. Genome-wide ancestry of 17th-century enslaved Africans from the Caribbean. *Proceedings of the National Academy of Sciences of the United States of America*,

112(12), pp.3669–73.

Schuster, S.C., 2007. Next-generation sequencing transforms today's biology. *Nature*, 200(8), pp.16–18.

Seal, J.N. Kellner, K., Trindl, A., & Heinze, J., 2011. Phylogeography of the partenogentic ant Platythrea punctata: highly successful colonization of the West Indies by a poor disperser. *Journal of biogeography*, 38(5), pp.868–882.

Smith, C.I. et al., 2001. Not just old but old and cold? *Nature*, 410(April), pp.771–772.

Smith, C.I. et al., 2003. The thermal history of human fossils and the likelihood of successful DNA amplification. *Journal of Human Evolution*, 45(3), pp.203–217.

Sondaar, P.Y., 1977. Insularity and its effect on mammal evolution. In *Major patterns in vertebrate evolution*. Springer US, pp. 671–7007.

Steadman, D.W. et al., 2007. Exceptionally well preserved late Quaternary plant and vertebrate fossils from a blue hole on Abaco, The Bahamas. *Proceedings of the National Academy of Sciences of the United States of America*, 104(50), pp.19897–19902.

Thomas, G.H., Meiri, S. & Philimore, A.B., 2009. Body size diversification in *Anolis*: novel enviroment and island effects. *Evolution*, 63(8), pp.2017–2030.

Thorpe, R.S., Leadbeater, D.L. & Pook, C.E., 2005. Molecular clocks and geological dates: Cytochrome b of Anolis extremus substantially contradicts dating of Barbados emergence. *Molecular Ecology*, 14(7), pp.2087–2096.

Turney, C.S.M. et al., 2008. Late-surviving megafauna in Tasmania, Australia, implicate human involvement in their extinction. *Proceedings of the National Academy of Sciences*, 105(34), pp.12150–12153.

Turvey, S.T., 2009. *Holocene extinctions*, OUP Oxford.

Turvey, S.T., Oliver, J. R., Storde, Y. N., & Rye, P. , 2007. Late Holocene extinction of Puerto Rican native land mammals. *Biology letters*, 3(2), pp.193–196.

VanValen, L., 1973. Pattern and the balance of nature. *Evolutionary Theory*, pp.31–49.

Wallace, A.R., 1869. *The Malay Archipelago: the land of the orangutan and the bird of paradise; a narrative of travel, with studies of man and nature*, London, UK: Macmillan.

Whittaker, R.J. & Frenandez-Palacios. J. M., 2007. *Island biogeography: ecology, evolution and conservation*, Oxford University Pess.

Willerslev, E. et al., 2007. Ancient biomolecules from deep ice cores reveal a forested southern Greenland. *Science*, 317(5834), pp.111–114.

Willerslev, E., & Cooper, A., 2005. Ancient DNA. *Proceedings. Biological sciences / The Royal Society*, 272(1558), pp.3–16.

Woodroffe, C.D., 1981. Mangrove swamp stratigraphy and Holocene transgression, Grand Cayman Island, West Indies. *Marine Geology*, 41(3–4), pp.271–294.

Wright, S., 1945. Tempo and mode in evolution: a critical review. *Ecology*, *26*(4), pp.415-419.

Zhou, Z., Yang, X. & Yang, Q., 2006. Land bridge and long-distance dispersal - Old views, new evidence. *Chinese Science Bulletin*, 51(9), pp.1030–1038.

Chapter 2. Methods

2.1 Overview

This chapter provides general information on the Methods used each of the studies that make up this thesis. As all chapters presented in this thesis deal with the extraction, sequencing and analysis of degraded DNA from sub fossil and museum specimens of Caribbean origin; as such, much of these methods have remained the same between chapters. This chapter includes the information on the specimens collected and sampled as well as an overview of molecular and bioinformatics techniques and specific ancient DNA protocol. Tables and figures included here provide information on specimens and sequence data, molecular markers, evolutionary models and results from trialled techniques and sequencing attempts.

2.2 Taxonomic sampling

This thesis covers several taxonomic groups of Caribbean mammals. Specimens from which aDNA was extracted varied widely in age and preservation quality and included: museum skins and skeletal collections, spirit preserved specimens and sub-fossil and zoo archaeological remains. All samples were treated as aDNA despite the young age of some specimens (< 200 years) due to the high levels of DNA degradation found in the tropics. The section below details the taxonomic sampling of this study (Appendix: Table 1a)

Capromyidae

The Capromyidae are composed of approximately eight genera. For this study I sampled individuals from all extant genera (32x *Capromys*, 16x *Mesocapromys*, 2x *Mysateles* and 10x *Geocapromys*) except *Plagiodontia*, where data were already publically available. I additionally sampled eight specimens from the extinct genus, *Isolobodon* and a single specimen from the extinct genus *Hexalobodon*. Samples were taken from zooarchaeological remains and museum skeletal and skin specimens.

Heptaxodontidae

In the extinct family Heptaxodontidae I sampled six specimens from only one genus: *Elasmodontomys*. Specimens sample from were sub-fossil specimens of bone and tooth.

Heteropsomyinae

Two species of extinct Caribbean spiny rats were sampled (3x *Brotomys offela* and 1x *Boromys voratus*) all were zooarchaeological bone specimens. The Hispaniolan *B. voratus* specimen was collected in the field by Dr Samuel Turvey and the three samples of *B. offella* originally from Cuba, were sourced from maxilla and mandible specimens in the Peabody Museum, Yale University.

Nesophontes

Dr Samuel Turvey collected the *Nesophontes zamicrus* sample used in this study from a cave in Cue de Bosque Humido, Los Haitises National Park, Hato Mayor Province, Dominican Republic. The specimen was collected from the surface of this cave in situ with other small vertebrate skeletal elements. The remains were most likely the product of an ancient barn owl roost. A previous study (Brace et al. 2016) had successfully extracted and sequenced DNA from a *Nesophontes paramicrus* specimen from this site. The material collected from this cave site was subjected to AMS dating (including a *Nesophontes* specimen not used as part of this or previous ancient DNA studies) and provides an estimate of a late Holocene age (see Chapter 3 for discussion on dating) (MacPhee, Flemming & Lunde, 1999).

Xenothrix

Two *Xenothrix mcgregori* specimens from Jamaica were sampled for this study, both supplied by Professor Ross MacPhee from The American Museum of Natural History. The first sample with museum reference number AMNH 268003 was previously sampled for 14C dating, dating this specimen at ~1500 BP. The second sample with the museum reference AMNH 268010 has not been dated but is suspected to be of similar age. Both samples had been recovered from Somerville Cave, Clavendon, Jackson's Bay, Jamaica.

2.3 Ancient DNA extractions

All DNA extractions followed ancient DNA (aDNA) procedures and protocols. These differ from modern DNA extraction methods due to the fragmented nature of surviving endogenous DNA and issues surrounding contamination. DNA extracted from these samples is classified as "ancient" as the specimens are of tropical origin and most are over 200 years old (see the ancient DNA section of Chapter 1. for further information on ancient DNA and DNA preservation in tropical environments). Due to the probable highly degraded nature of DNA from these specimens, all DNA extractions were conducted in a dedicated ancient DNA laboratory at the Natural History Museum London.

Bone specimens were sampled by drilling in to the bone using a dremel hand held drill with a 2-3mm drill bit. The surface of the bone was first cleaned using the drill to minimise contamination with surface bone powder discarded. Drill bits were changed between specimens and all equipment used was sterilized with bleach and UV treated before and after use. In order to keep possible drill-induced heat damage of DNA to a minimum, drill speeds were kept low (below 1000 RPM) (Adler et al. 2011). Larger pieces of bone were powdered using a Mikro Dismembrator prior to extraction.

Bone extraction method follows the Dabney et al. (2013) protocol. This protocol is designed to retrieve short fragment lengths associated with aDNA, it uses proteinase K to digest bone powders and silica spin columns to purify the DNA. In this method the powdered bone sample is left in EDTA and proteinase K for a 20 hour overnight digestion. A binding buffer solution is then added to bind the DNA to a silica membrane (Rohland & Hofreiter 2007). Purification involves repeated ethanol washes before the elution of the DNA sample in a low-salt TET buffer (Tris-HCL, EDTA and Tween-20; Dabney et al. 2013).

Tissue was broken up using a scalpel or a pestle and mortar prior to extraction. QIAGEN DNeasy blood and tissue kits, that also use proteinase K for lysis and silica spin columns for purification, were used for extraction from tissue and hair. As with bone extractions all equipment was sterilised with bleach and UV treated before and after use.

2.4 Preliminary PCR and shotgun Sanger sequencing

Prior to the use of expensive Next Generation Sequencing techniques two other techniques namely: PCR and shotgun Sanger sequencing, were utilised. This allows us to screen samples assessing the presence of endogenous DNA extracted from each specimen screened. Several PCRs were run following PCR protocol previously adapted for aDNA amplification (Brace et al. 2012) using primers for two mitochondrial genes: mitochondrially encoded 12S RNA (12s) and cytochrome B (cytB), both having been used previously in studies of echimyids and capromyids (Fabre et al. 2014). Gel electrophoresis and subsequent Sanger sequencing indicated the presence of DNA in some samples and therefore the most appropriate samples to take forward to library building for Next Generation Sequencing.

2.5 Next Generation Sequencing

Library builds

Single index double stranded DNA libraries were built following a modified protocol based on Meyer & Kircher (2010). This library build protocol has several steps, the first transforms the raw DNA

fragments in the sample to blunt end DNA. Once the ends are blunt universal adapters can be ligated to the DNA, this allows the DNA to attach to the sequencing platform. In addition, each sample is given a unique index sequence embedded in the adapter sequence. The index allows multiple samples to be pooled and run together on the sequencer as individual samples are distinguished bioinformatically by their unique index post sequencing. Multiple clean up steps were incorporated using mini elute spin columns. AmpliTaq Gold® DNA Polymerase was used for hot-start PCR amplification. As with extractions, pre PCR library builds took place in a dedicated aDNA laboratory that is physically separated from the PCR laboratory.

Sequencing

All samples used in this investigation underwent shotgun sequencing on either Illumina MiSeq or NextSeq 500 platforms in order to assess the endogenous content and the complexity of the sample. If the sample showed endogenous content during screening but had low yield (coverage < 1X) from shotgun methods, were then subjected to target capture enrichment in order to increase endogenous DNA yield. Where DNA concentration in libraries was considered too low for capture enrichment (< 7µL at 14-72 ng/µL) extra PCR steps were included. These extra PCR rounds were kept to a minimum (< 2 rounds in total) and only used when strictly necessary to restrict loss of library complexity due to the amplification of duplicate reads. All Sequencing was performed on a single-lane using a 2 x 75bp mid-output kit, on an Illumina NextSeq 500 and MiSeq platforms at sequencing facilities in the Natural History Museum, London.

Target capture enrichment

One way of ensuring higher endogenous DNA yield is in-solution hybridisation capture enrichment. The basis of this technique is to target specific genes or regions of DNA. This is achieved by selectively enriching sequences by using baits representing the regions of interest. Enrichment is carried out prior to the sequencing run giving an increased depth of sequencing of on target areas (Enk et al. 2014). The latest version (v3) of the capture enrichment kits available from the company MYcroarray (www.mycroarray.com) was designed and purchased specifically for this study. The ready-made capture kits include baits that are manufactured to order. They were designed to target genes from the most closely related species for which there were data available to produce the baits. These kits are designed for in-solution hybridisation capture of the targeted DNA, and use biotinylated RNA baits that are complimentary to the target sequences. After hybridisation the baits are removed using streptavidin magnetic beads, the libraries are then subjected to amplification and shotgun sequencing. This protocol was new to the group and this study has been the first to use this protocol to look at ancient DNA in Caribbean land mammals. The protocol was adapted by diluting the baits to 0.5µl per sample to increase the amount of samples enriched per kit (Li et al. 2013), samples with low endogenous content during screening we enriched with undiluted baits.

2.6 Post sequencing data processing

Data from Illumina Nextseq 500, MiSeq and HiSeq X runs were demultiplexed (by their unique index) and uploaded to the high throughput bioinformatics software CLC Genomics Workbench v.8 (CLC Bio-Qiagen, Aarhus, Denmark). Adapters were removed from the paired-end Illumina reads which were trimmed using quality scores (Phred quality score = 0.05), ambiguity criteria (maximum number per read = 2) and read length (discard below 25 bp). Trimmed reads were then merged using the following parameters: mismatch cost: 2, gap cost: 3.

In order to extract relevant genes from the Next Generation Sequencing data, sequences were aligned or mapped to a reference sequence The NCBI database Genbank (Benson et al. 2005) was used to search for mitochondrial genomes and nuclear genes for reference sequences and taxa used as outgroups in the phylogenetic analysis. Read mappings of paired-end Illumina reads to reference sequences taken from the NCBI database Genbank (Benson et al. 2005) were also performed in CLC Workbench software v.8 (CLC Bio-Qiagen, Aarhus, Denmark). Several sets of parameters for mapping raw reads to the reference sequences were trialled. Two sets of parameters were ultimately utilised on these data i) categorised as 'relaxed' (=M) and ii) as 'strict' (=H) (Appendix: Table 2a). The parameters that differ between sets involve two aspects of the mapping. Firstly "Length fraction", this is the minimum percentage of total read alignment length that must match the reference sequence. The second parameter is the minimum percentage identity between aligned region of the read, and the reference sequence, know as the "Similarity fraction". Both 'relaxed' and 'strict' sets of parameters have different merits. Where chances of contamination from humans were high (e.g.: in the Caribbean primate *Xenothrix*) strict parameters were used to try and prevent both the incorporation of this contamination into the assembled sequences and the likelihood of ascertainment bias. More relaxed parameters were utilised when analysing data from taxa phylogenetically further removed from sources of contamination (i.e. in the case of a human contamination source and rodent study species). The use of more relaxed parameters allows the inclusion of genetically distinct regions in to the final sequence, that may otherwise have been excluded due the phylogenetic distance between reference sequence used and the subject species. In cases where no close reference sequence was available or where ascertainment bias was an issue, denovo mapping was attempted using the default parameters. However this technique of generating sequences requires a higher percentage of endogenous DNA than is often available for these highly degraded samples. After mapping PCR duplicates were then removed. Basic Local Alignment Search Tool (BLAST; Altschul et al. 1990) was utilised to ensure correct taxonomic identification of the samples.

Assessment of sequence data

Depth of coverage. Depth of coverage in the context of this thesis is a term used to describe the number replicate or overlapping reads at any given region. The average depth of coverage provided in sequencing results tables or given in the text in this thesis is therefore the average number of reads overlapping across the region of interest.

Total reads mapped. Total number of reads mapped indicates the total number of paired trimmed reads which are mapped to the reference sequence during any given mapping, this may vary with differing reference sequences or mapping parameters.

Estimated endogenous content. Endogenous content values are often given by ancient DNA studies as a way of assessing the success of extraction, sequencing and post sequencing data analysis and to quanitify contamination. Endogenous values can be calculated as the percentage of reads mapping to reference sequences divided by the total number of reads for each sample. Due to the fact that reference sequences used for mapping raw reads to generate a consensus sequence, were often not from the target species true endogenous content could not be calculated. However as a way to quantify the results of extraction and sequencing during this study the same calculation was performed using the chosen non-species specific reference sequences to provide an estimated value for endogenous content.

Enrichment factor. The degree to which target capture enrichment improved the recovery of endogenous DNA for target regions was quantified by calculation of enrichment factors as well as comparing depth of coverage and total reads mapped in both shotgun and target capture methods. Enrichment factors were calculated as the proportion of reads mapping to target regions out of the total reads when compared to shotgun sequencing methods.

Sequence damage

Sequence data generated as part of this project is likely to contain errors resulting from sequencing, and damage inherent in ancient DNA (e.g.: Deamination of cytosine to uracil, 5-methyl-cytosine to thymine and adenine to hypoxathine and the resulting C -> T and G -> A transitions). The program MapDamage (Ginolhac et al 2011) is now often used in ancient DNA studies in order to quantify damage patterns associated with ancient DNA and verify the authenticity of ancient samples. This study did not include humans or closely related species, additionally the laboratory used for extraction and library builds of ancient samples was not used for modern molecular analysis of South American rodents, primates or eulipotyplan insectivores. All specimens sampled as part of this study had also been identified morphologically as the subject species and were part of either museum or private collections. Due to these facts the authenticity of ancient samples was unlikely to be incorrect on the basis of either misidentification or contamination during mapping to reference sequences. However ancient DNA associated sequence errors may still have affected our post-

sequencing data processing and phylogenetic analysis. We detected and corrected for mapping ascertainment bias in our mapping to reference sequences (see the following section). Additionally errors associated with ancient DNA damage may have affected phylogenetic analysis contributing to incorrect estimates of branch lengths and the associated divergence dates. Transitions erroneously incorporated in sequence data generated from ancient samples may artificatly inflate substutions, thus increasing the lengths of branches and influencing the level of divergence between species. These potential sources of error can be mitigated in a number of ways. For example parameters used for mapping of raw reads to a chosen reference genome can be altered, strict parameters increase the influence of mapping asertainment bias whilst relaxed parameters may allow the incorporation of more errors. The use of multiple molecular markers, for example the whole mitogenome and multiple nuclear genes, could help to dilute errors in the sequence data, conversely the use of few genetic markers will increase the effect of any errors on phylogenetic analysis.

Mapping ascertainment bias

Many species looked at as part of this study are extinct, critically endangered or understudied using molecular techniques; this means that often there is no reference sequence of the same or closely related species available to map raw reads to. In extinct species where there is still controversy over the taxonomic placement of the subject species, the closest related species may not be known. Due to the low number of endogenous reads recovered during sequencing of these samples de novo assembly (assembly of sequence data without a reference sequence) is not an option. However mapping to a reference sequence from a genetically far removed species can cause some issues for phylogenetic reconstructions. Mapping ascertainment bias occurs when mapping raw reads to a specific reference sequence results in bias towards that reference sequence, artificially increasing the genetic similarities between the resulting consensus of the raw reads to the original reference. Variable base calls in the reads coverage can be due to multiple factors: including sequencing errors, DNA damage and contamination. This variability can result in mapping ascertainment bias if there is a disagreement between two nucleotides, and one is chosen over the other based on its similarity to the reference sequence. This study used methods based on Westbury et al., (2017) in order to combat ascertainment bias issues with sequence data generated. This method involved generating multiple mappings of raw reads. Raw reads were mapped to four reference sequences (three ingroup reference sequences and one outgroup reference sequence) using the different mapping parameters mentioned above, six different levels of parameters “mismatch values” were implemented ranging from relaxed to strict mapping (Appendix: Table 2a). For each mapping a consensus sequence was extracted, resulting in a consensus sequence for different reference sequences for each mismatch value. Consensus sequences from each mismatch value (with differing reference sequences) were aligned to produce one sign consensus

sequence for each mismatch value. Any variable regions in this alignment were removed. Finally mismatch value consensus sequences were aligned to produce a final consensus sequence (Figure 1).

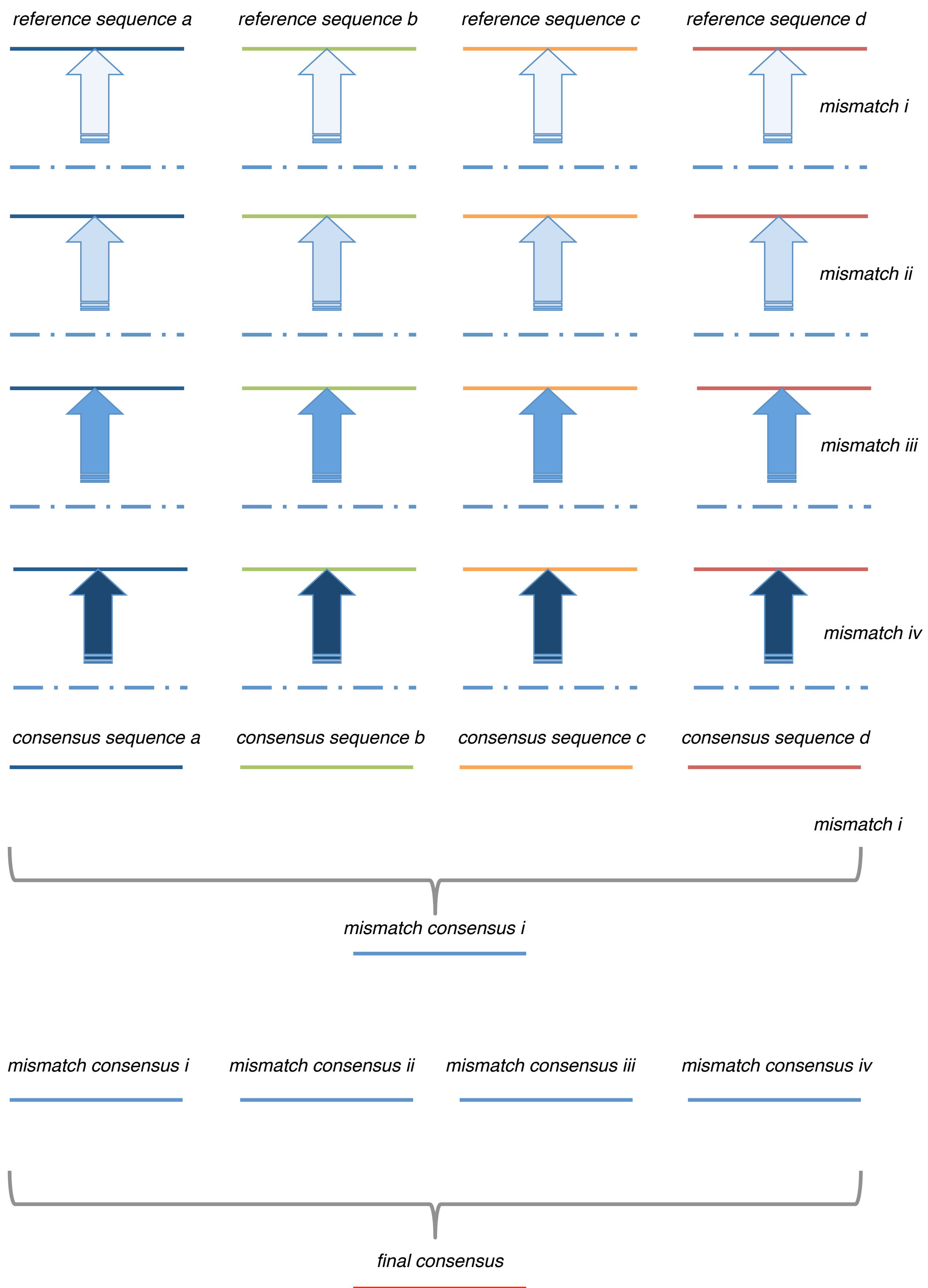


Figure 1. Illustration of the mapping process for the limitation of mapping ascertainment bias. Dotted line represents sequence data, coloured lines represent different alternative reference sequences to which the sequence data is mapped (indicated by arrows). Differently shaded arrows correspond to different mismatch (sequence mapping parameters) used to map sequence data to the reference sequence.

Alignments: molecular markers

For different aspects of this project alternative molecular markers were used to infer phylogeny and genetic divergence. In some cases the small amount of data available from old or degraded specimens limited the choice of marker. The aim to recover partial or whole mitochondrial genomes was achieved in many of the subject species included in this study. Where mitochondrial genome recovery was not possible several mitochondrial genes, for which reference sequences of closely related species were available on the NCBI database Genbank (Benson et al. 2005), were used. Mitochondrial genomes (and genes) have been used previously in ancient DNA studies looking to infer phylogenetic position of extinct or extant species (Meyer et al. 2014, Dabney et al. 2013b & Thalmann et al. 2013). Several factors including the cellular high copy number and physical structure of mitochondria increase chances of recovery from ancient samples. Mitochondrial DNA only represents the female line of the individual as it is inherited maternally. Mitochondrial DNA is more informative at shallow phylogenetic levels because it has shorter coalescent time than any given nuclear locus, making it more able to resolve relationships between recently diverged species or populations. However this shorter coalescent time also gives mitochondrial DNA poor resolving power for deep relationships. This is because mutations arise and accumulate over time, meaning that the sequences will be randomised, this phenomenon is referred to as saturation (Rubinoff & Holland 2005, Ballard & Whitlock 2004 & Funk & Omland 2003). The addition of nuclear genes with slower mutation rate are paramount in studies hoping to recover deep splits in phylogeny (Springer et al. 2001). Particular nuclear genes were chosen on the basis of available reference genes.

Alignments: Programs and parameters

For each project consensus sequences for each sample and publically available sequences from other relevant taxa that were to be included in the phylogeny and outgroup were then aligned using ClustalW (Larkin et al. 2007) alignment tool implemented in the program Geneious v. 8.0.5 (http://www.geneious.com, Kearse et al., 2012). Alignments were then manually edited and translated to protein alignments to check for non-sense base calls and correct alignment of stop codons. In some cases the program G-blocks (Talavera & Castresana 2007) was utilised in order to remove poorly aligned positions and divergent regions of the alignment. These positions are removed either because they are not homologous or because they are saturated by multiple substitutions, making it necessary to eliminate them prior to phylogenetic analysis.

Alignments: Partitioning and concatenation

The use of maximum likelihood (ML) and Bayesian algorithms in developing phylogenetic hypotheses from the above alignments requires a model of evolution or nucleotide substitution. These models describe the rate at which one nucleotide replaces another during evolution. Point

mutations such as these form two categories: transitions (purine to purine = A to G, or pyrimidine to pyrimidine = C to T) or transversions (purine to pyrimidine = e.g.: A to T or G to C). Transitional changes are known to be more common (Kimura, 1980). Models termed as 'corrected' take in to account the fact that because there are only four possible character states, it is expected that as genetic distance increases, some sites will undergo multiple superimposed substitutions, and in that process some sites may revert to the state they were originally in. 'Uncorrected' models which do not factor in multiple substitutions often underestimate the genetic distance. More complex models often fit data better than the more simplistic ones. However more complex models are more computationally intensive and thus consume more time, and as more parameters need to be estimated in complex models the margin for error also increases. Therefore the best practise is to choose a model complex enough to explain the data, but not too complex as to increase the two disagreeable factors mentioned above. Best-fit models of nucleotide substitution can be selected using statistical testing in programs such as Partitionfinder (Lanfear et al. 2012) utilised here.

Mitochondrial genomes were initially partitioned by annotating the entire mitogenome using the online software MITOS (Bernt et al. 2013), and then the program Partitionfinder (Lanfear et al. 2012) was used to asses which evolutionary model was most appropriate for each of the partitions. Partitionfinder was also used to assess which evolutionary model should be applied to the nuclear genes used in this study. The alignments for different genes to be included in the phylogeny were then concatenated to generate one alignment file using the program Seaview (Gouy et al. 2010).

3.7 Phylogenetics

Molecular phylogenetic studies reconstruct the evolutionary history of species by using differences in DNA sequence data. These studies are informed by the observation that all life is related by common decent; that closely related species share a common ancestor. As the accumulation of mutations in the genome is generally continuous over time, the evolutionary relationships of species can be accessed, by comparing multiple genes or whole genomes. Phylogenies can be dated by using molecular clock models and by adding fossil or geological data in order to include absolute dates for some nodes (Rutschmann, 2006).

Phylogenetics: Maximum Likelihood

Likelihood is the probability of observing the data, in this case the sequence data, in light of a particular model; here this is the nucleotide substitution model. The tree with the highest likelihood is the best estimate of the true phylogeny.

Bootstrapping analysis provides a way to quantify the strength of support for clades on Maximum likelihood tree estimations. Bootstrapping is a form of resampling analysis that takes columns of

characters out of the dataset, rebuilding the tree, and testing, through multiple iterations, if the same nodes are recovered each time. If the same node is recovered through 95 of 100 iterations, the node is well supported, giving a high bootstrap value of 0.95 or 95%. If the bootstrap value is low for a particular node, this suggests that only a few characters support the node.

Maximum Likelihood (ML) methods were implemented in the programs Geneious v. 8.0.5 (http://www.geneious.com, Kearse et al., 2012) and Seaview (Gouy et al. 2010) in order to generate preliminary trees. Final trees with bootstrap support values were generated in RAxML (Randomized Axelerated Maximum Likelihood) (Stamatakis, Hoover & Rougemont, 2008) through the CIPRES Science Gateway V. 3.3 (Miller, Pfeiffer & Schwartz, 2010).

Phylogenetics: Bayesian

The Bayesian method of tree generation also involves a likelihood function but where Maximum likelihood methods are the probability of the data given the model, Bayesian methods use the probability of the model (nucleotide substitution models and tree topology) given the sequence data. This method combines the prior probability of a phylogenetic tree with the likelihood of the data that produces the posterior probability on different trees. The higher the posterior probability of a tree the more likely it is to be correct.

Bayesian methods used in this study implement a MCMCMC (Metropolis-coupled Markov chain Monte Carlo) algorithm. MCMC (Monte Carlo Markov Chain) is used to approximate the posterior probabilities of trees by drawing (dependent) samples from the posterior distribution. In the MCMC algorithm several chains are run in parallel, these chains move in a ‘random walk’ through the tree landscape sampling values in proportion to their probability distribution. High values or peaks in the landscape are almost always accepted as are small steps down from peaks, however low values or troughs in the landscape are less likely to be accepted. Chains start with a burn in period, after this period samples congregate in a plateau, samples form this plateau are then used to construct the posterior distribution of all the parameters and all of the trees. Sampled trees can then be used to estimate a consensus tree, the proportion of sampled trees provide Bayesian posterior probability estimates of clade credibility. The addition of Metropolis-coupled algorithm to the Markov chain Monte Carlo algorithm allows chains to mix better, after the first cycle, two chains are chosen at random in order to swap states (with the probability of a swap being determined by the Metropolis-coupled equation). This allows the chain to potentially jump a valley in a single bound (Yang & Rannala 1997, Altekar et al. 2004).

The program MrBayes v. 3 (Ronquist & Huelsenbeck 2003) was used to estimate phylogeny using the above methods. For the majority of cases the following parameters were utilised using a MrBayes block: the MCMCMC algorithm incorporated in the MrBayes format using four chains

(three heated, one cold) that were run for 1×10^6 generations. Sampling initiated every 1×10^3 generations with a burn-in period of 250 trees. An appropriate outgroup was included for each analysis. For differences in parameter values across study groups, including nucleotide substitution models please see the individual chapter concerned.

2.8 Pairwise distance analyses

As well a phylogenetic analyses pairwise genetic distances were also calculated to estimate the evolutionary distance between individuals. Like Maximum Liklihood and Bayesian molecular phylogenetic methods distance estimation also requires a nucleotide substitution model as chosen in the program ParitionFinder (Lanfear et al. 2012). Evolutionary distances between sequences were estimated in the program MEGA v. 5 (Tamura et al. 2011) by computing the proportion of nucleotide differences between each pair of sequences. This technique provides a matrix of pairwise genetic distance values that can be compared with values from other species pairs. For evolutionary models used for each data set see chapters concerned.

2.9 Divergence dating

Bayesian analysis of molecular sequences for divergence dating was achieved using the software BEAST v 1.8.3 (Drummond & Rabat, 2007). This software uses MCMC (markov chain monte carlo) to generate rooted, time-measured phylogenies. Phylogeny and diversification times were jointly estimated under an uncorrelated relaxed lognormal clock. The partitioning scheme and best-fit models chosen in PartitionFinder were implemented and a speciation birth-death process was used for all analyses. The sampling distributions of each run were visualised using Tracer v. 1.6 (Rambaut et al. 2014) to assess convergence and to verify that the effective sample size was>200 for all parameters after a burn-in of 10 %. To obtain the posterior distribution of the estimated divergence times, fossil calibration points were used with lognormal priors to set a hard minimum and soft maximum bound (see specific chapters for fossil calibrations used). Finally maximum credibility trees were generated in TreeAnnotator v. 1.8.3 (Rambaut and Drummond 2010) using trees sampled in the prior distribution.

References

Adler CJ, Haak W, Donlon D, Cooper A & Genographic Consortium (2011) Survival and recovery of DNA from ancient teeth and bones. Journal of Archaeological Science, 38(5), 956-964.

Altekar G, Dwarkadas S, Huelsenbeck, JP & Ronquist F (2004) Parallel metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. Bioinformatics, 20(3), 407-415.

Altschul SF, Gish W, Miller W, Myers EW & Lipman DJ (1990) Basic local alignment search tool. Journal of molecular biology, 215(3), 403-410.

Ballard JWO & Whitlock MC (2004) The incomplete natural history of mitochondria. Molecular ecology, 13(4), 729-744.

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J & Wheeler DL (2005) GenBank. Nucleic Acids Research 33 (Database Issue), D34–D38.

Bernt M, et al., (2013) MITOS: Improved de novo metazoan mitochondrial genome annotation. Molecular phylogenetics and evolution, 69(2), pp.313-319.

Brace S, Barnes I, Powell A, Pearson R, Woolaver LG, Thomas MG & Turvey ST (2012) Population history of the Hispaniolan hutia Plagiodontia aedium (Rodentia: Capromyidae): testing the model of ancient differentiation on a geotectonically complex Caribbean island. Molecular Ecology, 21(9) 2239-2253.

Brace S, Thomas JA, Dalén L, Burger J, MacPhee RD, Barnes I & Turvey ST (2016) Evolutionary history of the Nesophontidae, the last unplaced recent mammal family. Molecular Biology and Evolution, 33(12), 3095-3103.

Dabney J, Knapp M, Glocke I, Gansauge MT, Weihmann A, Nickel B, Valdiosera C, García N, Pääbo S, Arsuaga JL and Meyer M (2013) Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. Proceedings of the National Academy of Sciences, 110(39), pp.15758-15763.

Drummond AJ & Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC evolutionary biology, 7(1), 1.

Enk JM, Devault AM, Kuch M, Murgha YE, Rouillard JM, & Poinar HN (2014) Ancient whole genome enrichment using baits built from modern DNA. Molecular Biology and Evolution, 31(5), 1292-1294.

Fabre PH, Vilstrup JT, Raghavan M, Der Sarkissian C, Willerslev E, Douzery EJ & Orlando L (2014) Rodents of the Caribbean: origin and diversification of hutias unravelled by next-generation museomics. Biology letters, 10(7), 20140266.

Funk DJ & Omland KE (2003) Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. Annual Review of Ecology, Evolution, and Systematics, 397-423.

Gouy M, Guindon S & Gascuel O (2010) SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Molecular biology and evolution, 27(2), 221-224.

Ginolhac, A., Rasmussen, M., Gilbert, M.T.P., Willerslev, E. and Orlando, L., 2011. mapDamage: testing for damage patterns in ancient DNA sequences. *Bioinformatics*, *27*(15), pp.2153-2155.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Mentjies P & Drummond A (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics, 28(12), 1647-1649.

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. Journal of molecular evolution, 16(2), 111-120.

Lanfear R, Calcott B, Ho SY & Guindon S (2012) PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Molecular biology and evolution, 29(6), 1695-1701.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ and Higgins DG (2007) ClustalW and ClustalX version 2, Bioinformatics 23(21): 2947-2948.

MacPhee RD, Flemming C & Lunde DP (1999) Last occurrence of the Antillean insectivoran Nesophontes: new radiometric dates and their interpretation. American Museum novitates, 3261.

Meyer M & Kircher M (2010) Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harbour Protocols, 2010(6), t5448.

Meyer M, Fu Q, Aximu-Petri A, Glocke I, Nickel B, Arsuaga JL, Martínez I, Graci A, de Castro JMB, Carbonell E and Pääbo S (2014) A mitochondrial genome sequence of a hominin from Sima de los Huesos. Nature, 505(7483), pp.403-406.

Miller MA, Pfeiffer W & Schwartz T (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Gateway Computing Environments Workshop (GCE), 1-8.

Rambaut A, Suchard MA, Xie D & Drummond AJ (2014) Tracer v1.6

Roland N & Hofreiter M (2007) Ancient DNA extraction from bones and teeth. Nature protocols, 2(7), 1756-1762.

Ronquist F & Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics, 19(12), 1572-1574.

Rubinoff D & Holland BS (2005) Between two extremes: mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. Systematic biology, 54(6), 952-961.

Rutschmann F (2006) Molecular dating of phylogenetic trees: a brief review of current methods that estimate divergence times. Diversity and Distributions, 12(1), 35-48.

Springer MS, DeBry RW, Douady C, Amrine HM, Madsen O, de Jong WW & Stanhope MJ (2001) Mitochondrial versus nuclear gene sequences in deep-level mammalian phylogeny reconstruction. Molecular Biology and Evolution, 18(2), 132-143.

Stamatakis A, Hoover P & Rougemont J (2008) A rapid bootstrap algorithm for the RAxML web servers. Systematic biology, 57(5), 758-771.

Talavera G & Castresana J (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Systematic biology, 56(4), 564-577.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M & Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular biology and evolution, 28(10), 2731-2739.

Thalmann O, Shapiro B, Cui P, Schuenemann VJ, Sawyer SK, Greenfield DL, Germonpré MB, Sablin MV, López-Giráldez F, Domingo-Roura X and Napierala H (2013) Complete mitochondrial genomes of ancient canids suggest a European origin of domestic dogs. Science, 342(6160), pp.871-874.

Yang Z & Rannala B (1997) Bayesian phylogenetic inference using DNA sequences: a Markov Chain Monte Carlo method. Molecular biology and evolution, 14(7), 717-724.

Appendix Chapter 2. Methods

Table 1a. Specimens sampled as part of this thesis. Collection abbreviations are provided at the bottom of the table.

Family/sub family	Genus	Species/sub species	Material	Location	Collection	Museum ID	Extract ID
Capromyidae	*Capromys*	*Cp. pilorides gundlachianus*	Tissue	Cayo Leguas, Cuba	IES	1.414	RW023
Capromyidae	*Capromys*	*Cp. pilorides gundlachianus*	Tissue	Cayo Fragoso, Cuba	IES	1.320.	RW025
Capromyidae	*Capromys*	*Cp. pilorides gundlachianus*	Tissue	Cayo Fragoso, Cuba	IES	1.384	RW028
Capromyidae	*Capromys*	*Cp. pilorides gundlachianus*	Tissue	Cayo Fragoso, Cuba	IES	1.290.	RW030
Capromyidae	*Capromys*	*Cp. pilorides doceleguas*	Tissue	Cayos y Laberinto de las Doce Leguas, Cayo Caguana, Cuba	IES	1.466	RW026
Capromyidae	*Capromys*	*Cp. pilorides doceleguas*	Tissue	Cayos y Laberinto de las Doce Leguas, Cayo Caguana, Cuba	IES	1.433	RW027
Capromyidae	*Capromys*	*Cp. pilorides doceleguas*	Tissue	Golfo de Ana Maria, Jucaro, Cuba	IES	1.300.	RW029
Capromyidae	*Capromys*	*Cp. pilorides relictus*	Tissue	North Isla Juventud, Cuba	IES	1.76	RW033
Capromyidae	*Capromys*	*Cp. pilorides relictus*	Tissue	Cayo los Monos, North Isla Juventud, Cuba	IES	1.85	RW036
Capromyidae	*Capromys*	*Cp. pilorides relictus*	Tissue	Cayo los Monos, North Isla Juventud, Cuba	IES	1.87	RW042
Capromyidae	*Capromys*	*Cp. pilorides ciprianoi*	Tissue	South Isla Juventud, Cuba	IES	1.46	RW034
Capromyidae	*Capromys*	*Cp. pilorides ciprianoi*	Tissue	Hato de Milian, South Isla Juventud, Cuba	IES	1.31	RW041
Capromyidae	*Capromys*	*Cp. pilorides ciprianoi*	Tissue	South Isla Juventud, Cuba	IES	1.9	RW037

Table 1a. (continued) Specimens sampled as part of this thesis. Collection abbreviations are provided at the bottom of the table.

Family/sub family	Genus	Species/sub species	Material	Location	Collection	Museum ID	Extract ID
Capromyidae	*Capromys*	*Cp. pilorides ciprianoi*	Tissue	Estero las Piedras, South Isla Juventud, Cuba	IES	1.25	RW038
Capromyidae	*Capromys*	*Cp. pilorides pilorides*	Tissue	Guanahacabibes Peninsula, Cuba	IES	1.426	RW035
Capromyidae	*Capromys*	*Cp. pilorides pilorides*	Tissue	Guanahacabibes Peninsula, Cuba	IES	1.388	RW039
Capromyidae	*Capromys*	*Cp. pilorides pilorides*	Tissue	Guanahacabibes Peninsula, Cuba	IES	1.389	RW040
Capromyidae	*Capromys*	*Cp. pilorides pilorides*	Tissue	E. Sierra Maestra, Santiago de Cuba Province, Cuba	MHNTR		RW032
Capromyidae	*Capromys*	*Cp. pilorides pilorides*	Tissue	E. Sierra Maestra, Santiago de Cuba Province, Cuba	MHNTR		RW031
Capromyidae	*Capromys*	*Cp. pilorides pilorides*	Tissue	Guanahacabibes Peninsula, Cuba	IES	1.559.5	RW044
Capromyidae	*Capromys*	*Cp. pilorides pilorides*	Tissue	Managua, Cuba	IES	1.372	RW045
Capromyidae	*Capromys*	*Cp. pilorides pilorides*	Tissue	Cuba mainland	IES		RW024
Capromyidae	*Capromys*	*Cp. pilorides pilorides*	Tissue	Cuba mainland	SU		RW047
Capromyidae	*Capromys*	*Cp. pilorides pilorides*	Tissue		SU		RW048
Capromyidae	*Capromys*	*Cp. pilorides pilorides*	Tissue	Jatales de manati Los Comos Estate, Cuba	SU	99	RW049
Capromyidae	*Capromys*	*Cp. pilorides pilorides*	Tissue	Cuba mainland	SU		RW050
Capromyidae	*Capromys*	*Cp. pilorides pilorides*	Tissue		SU		RW051
Capromyidae	*Capromys*	*Cp. pilorides sp.*	Bone	Cayman Brac	FLMNH	UF 61292	RW057
Capromyidae	*Capromys*	*Cp. pilorides sp.*	Bone	Cayman Brac	FLMNH	UF 18588	RW058

Table 1a. (continued) Specimens sampled as part of this thesis. Collection abbreviations are provided at the bottom of the table.

Family/sub family	Genus	Species/sub species	Material	Location	Collection	Museum ID	Extract ID
Capromyidae	*Capromys*	*Cp. pilorides sp.*	Bone	Cayman Brac	FLMNH	UF 18671	RW059
Capromyidae	*Capromys*	*Cp. pilorides sp.*	Bone	Cayman Brac	FLMNH		RW066
Capromyidae	*Capromys*	*Cp. pilorides sp.*	Bone	Cayman Brac	FLMNH		RW067
Capromyidae	*Mesocapromys*	*Angelcabrerai*	Tissue	Cayos de Ana Maria			RW001
Capromyidae	*Mesocapromys*	*Angelcabrerai*	Tissue	Cayos de Ana Maria			RW002
Capromyidae	*Mesocapromys*	*Angelcabrerai*	Tissue	Cayos de Ana Maria			RW003
Capromyidae	*Mesocapromys*	*Angelcabrerai*	Tissue	Cayos de Ana Maria			RW004
Capromyidae	*Mesocapromys*	*Auritus*	Tissue	Cayo Fragoso, Cuba			RW005
Capromyidae	*Mesocapromys*	*Auritus*	Tissue	Cayo Fragoso, Cuba			RW006
Capromyidae	*Mesocapromys*	*Auritus*	Tissue	Cayo Fragoso, Cuba			RW007
Capromyidae	*Mesocapromys*	*Auritus*	Tissue	Cayo Fragoso, Cuba			RW008
Capromyidae	*Mesocapromys*	*Nanus*	Tissue	Zapata Peninsula	NHM	1938.5.24 .1	RW015
Capromyidae	*Mesocapromys*	*Nanus*	Tissue	Zapata Peninsula	NHM	1938.5.24 .2	RW016
Capromyidae	*Mesocapromys*	*Nanus*	Tissue (wet)	Zapata Peninsula	NHM	1938.5.24 .2	RW022
Capromyidae	*Mesocapromys*	*Nanus*	Tissue	Zapata Peninsula	IES		RW043
Capromyidae	*Mesocapromys*	*Melanurus*	Tissue	Santiago de Cuba Province, Gran Piedra	SU		RW046
Capromyidae	*Mesocapromys*	*Melanurus*	Tissue	Eastern Cuba			RW009
Capromyidae	*Mesocapromys*	*Melanurus*	Tissue	Eastern Cuba			RW010
Capromyidae	*Mesocapromys*	*Sanfelipensis*	Tissue	Cayo Juan Garcia	IES	1.198	RW052
Capromyidae	*Mysateles*	*Prehensilis*	Tissue	Cuba mainland	NHM	58.8.31.2	RW020
Capromyidae	*Mysateles*	*Prehensilis*	Tissue	Cuba mainland	NHM	46.10.3.8	RW021
Capromyidae	*Geocapromys*	*Brownii*	Tissue	Jamaica	NHM		RW013
Capromyidae	*Geocapromys*	*Brownii*	Tissue	Jersey Wildlife trust	NHM		RW014
Capromyidae	*Geocapromys*	*Ingrahami*	Tissue	Bahamas	NHM		RW017

Table 1a. (continued) Specimens sampled as part of this thesis. Collection abbreviations are provided at the bottom of the table.

Family/sub family	Genus	Species/sub species	Material	Location	Collection	Museum ID	Extract ID
Capromyidae	*Geocapromys*	*Ingrahami*	Tissue	Bahamas	NHM		RW018
Capromyidae	*Geocapromys*	*Ingrahami*	Tissue	Bahamas	NHM		RW019
Capromyidae	*Geocapromys*	*Thoracatus*	Tissue	Little Swan Island	RAMM (Exeter)		RW053
Capromyidae	*Geocapromys*	*Thoracatus*	Tissue	Little Swan Island			RW054
Capromyidae	*Geocapromys*	*columbinanus*	Bone	Cuba	PMYU		SB298
Capromyidae	*Geocapromys*	*columbinanus*	Bone	Cuba	PMYU		SB299
Capromyidae	*Geocapromys*	*columbinanus*	Bone	Cuba	PMYU		SB300
Capromyidae	*Isolobodon*	*Portoricensis*	Bone	Puetro Rico	ZSL		SB165
Capromyidae	*Isolobodon*	*Portoricensis*	Tooth (incisor)	Guana island	ZSL		RWI78
Capromyidae	*Isolobodon*	*Portoricensis*	Tooth (molar)		ZSL		RWI79
Capromyidae	*Isolobodon*	*Portoricensis*	Tooth (incisor)	"Two caves", Parque del este, DR.	ZSL		RWI80
Capromyidae	*Isolobodon*	*Portoricensis*	Tooth (molar)	Cueva De La Vaca, Puerto Rico	ZSL		RWI81
Capromyidae	*Isolobodon*	*Portoricensis*	Bone	Cueva De La Vaca, Puerto Rico	ZSL		RWI82
Capromyidae	*Isolobodon*	*Portoricensis*	Tooth (incisor)	Cueva De La Vaca, Puerto Rico	ZSL		RWI83
Capromyidae	*Isolobodon*	*Montananus*	Bone (mandible)	"Two caves", Parque del este, DR.	ZSL		RWI84
Capromyidae	*Hexolobodon*	*sp.*	Tooth (molar)	Hispaniola	NHM		RWH77
Heteropsomyinae	*Brotomys*	*Voratus*	Bone	Hispaniola	ZSL		SB180
Heteropsomyinae	*Boromys*	*Offella*	Bone	Cuba	ZSL		SB293
Heteropsomyinae	*Boromys*	*Offella*	Bone	Cuba	ZSL		SB295
Heteropsomyinae	*Boromys*	*Offella*	Bone	Cuba	ZSL		SB294
Heptaxodontidae	*Elasmodontomys*	*Obliquus*	Bone (rib)	Puerto Rico	ZSL		RW060
Heptaxodontidae	*Elasmodontomys*	*Obliquus*	Bone (rib)	Puerto Rico	ZSL		RW061

Table 1a. (continued) Specimens sampled as part of this thesis. Collection abbreviations are provided at the bottom of the table.

Family/sub family	Genus	Species/sub species	Material	Location	Collection	Museum ID	Extract ID
Heptaxodontidae	*Elasmodontomys*	*Obliquus*	Bone	Puerto Rico	ZSL		RW062
Heptaxodontidae	*Elasmodontomys*	*Obliquus*	Tooth	Puerto Rico	ZSL		RW063
Heptaxodontidae	*Elasmodontomys*	*Obliquus*	Bone	Puerto Rico	ZSL		RW064
Heptaxodontidae	*Elasmodontomys*	*Obliquus*	Tooth	Puerto Rico	ZSL		RW065
Nesophontidae	*Nesophontes*	*Zamicrus*	Bone (cranial)	Cueva de Bosque Humido, Hato Mayor Province, DR	ZSL		RWN70
Nesophontidae	*Nesophontes*	*Zamicrus*	Bone (cranial)	Cueva de Bosque Humido, Hato Mayor Province, DR	ZSL		RWN71
Nesophontidae	*Nesophontes*	*Zamicrus*	Bone (cranial)	Cueva de Bosque Humido, Hato Mayor Province, DR	ZSL		RWN72
Nesophontidae	*Nesophontes*	*Zamicrus*	Bone (cranial)	Cueva de Bosque Humido, Hato Mayor Province, DR	ZSL		RWN73
Nesophontidae	*Nesophontes*	*Zamicrus*	Bone (cranial)	Cueva de Bosque Humido, Hato Mayor Province, DR	ZSL		RWN74
Nesophontidae	*Nesophontes*	*Paramicrus*	Bone (cranial)	Cueva de Bosque Humido, Hato Mayor Province, DR	ZSL		RWN75
Nesophontidae	*Nesophontes*	*Paramicrus*	Bone (cranial)	Cueva de Bosque Humido, Hato Mayor Province, DR	ZSL		RWN76
Callicebinae	*Xenothrix*	*Mcgregori*	Bone (femur)		AMNH	268003	RW055
Callicebinae	*Xenothrix*	*Mcgregori*	Bone (proximal ulna)	Somerville Cave, Clavendon, Jackson's Bay, Jamaica	AMNH	268010	RW056

Museum abbreviations: IES: Instituto de Ecología y Sistemática, Cuba, MNHTR: Museo d'historia natural tomas romay, Cuba, SU: Santiago University, FLMNH: Florida Museum of Natural History, US, NHM: Natural History Museum, UK, RAMM: Royal Albert Memorial Museum, UK, PMYU: Peabody Museum, Yale University, US, ZSL: Zoological Society of London, UK.

Table 2a. Mapping parameters used in CLC genomics workbench

Mapping parameter	Length fraction	Similarity fraction
default	0.5	0.8
L	0.7	0.7
M	0.8	0.8
M1	0.85	0.85
H	0.9	0.9
H1	0.95	0.95
XH	1	1

Chapter 3. Palaeogenomic analyses reveal recent intra-island evolutionary radiation in the extinct Caribbean "island shrew" (*Nesophontes*)

3.1 Abstract

The Caribbean islands offer a unique opportunity to study evolutionary processes in insular faunas. However, the recent extinction of the majority of the Caribbean's endemic mammal fauna, combined with poor DNA preservation typical of tropical environments, has obstructed study of species relationships and evolutionary history in this island system. The Quaternary Caribbean fauna contained two endemic families of lipotyphlan insectivores, the Solenodontidae (solenodons) and Nesophontidae ("island shrews"), but the entire nesophontid radiation is now extinct and the evolutionary history of this family is poorly understood. Recent ancient DNA analysis of *Nesophontes paramicrus* from Hispaniola has determined the wider phylogenetic relationship between the Solenodontidae and Nesophontidae, but the relationships and taxonomic validity of the radiation of now-extinct species within the genus *Nesophontes* remain unclear. In this study, palaeogenomic methods were used to explore the evolution of the genus *Nesophontes* in greater depth and to clarify species relationships within Hispaniolan *Nesophontes*. We report the first ancient DNA sequence data for the extinct island shrew species *Nesophontes zamicrus*, which were analysed to determine the phylogenetic relationship between Hispaniolan *Nesophontes* species. These new data support a congruent lipotyphlan phylogeny to the previous molecular study of nesophontid higher-order relationships, and place *N. zamicrus* and *N. paramicrus* as closely related species that diverged as recently as 0.6 Ma (95% HPD lower/upper = 0.232–1.037 Ma). This study has utilised ancient DNA evidence to delimit species relationships in Hispaniolan *Nesophontes* species that were previously determined on morphology alone. Whereas both Solenodontidae and Nesophontidae can be interpreted as ancient "living fossil" lineages within the Mammalia, our results show that Hispaniolan nesophontids also experienced evolutionarily recent diversification. These findings suggest that Hispaniola was not only an insular sanctuary preserving ancient biodiversity, but also continued to be a hotspot for the generation of new mammal species up until the recent human-caused extinction of most of the Caribbean mammal fauna.

3.2 Introduction

Island evolution and the Caribbean's endemic lipotyphlan insectivores

Islands are often described as "natural laboratories" (Ricklefs & Bermingham 2008), as their isolation from continental mainland ecosystems can promote species radiations that evolve in a discrete system (Darwin 1859; Wallace 1869). The study of insular faunas has typically focused on the reconstruction of colonisation histories of endemic taxa, and on post-colonisation adaptive radiations (Schluter 2000; Losos & Ricklefs 2009). Island faunas are often unusual in their phylogenetic and ecological composition, size and morphology. There is also evidence that evolution is accelerated in island mammals (Millien 2006), causing morphological changes such as body size shifts to occur more rapidly in insular species than in mainland taxa. Larger islands provide a much more complex arena for the study of evolutionary pattern and process, as current or past barriers to gene flow such as mountain ranges or valleys can also drive intra-island allopatric speciation events.

The Caribbean islands have the ability to play an important role in our understanding of species relationships and patterns of evolutionary diversification, particularly because these islands are unusual in that they were colonised by several lineages of land mammals. However, our understanding of the relationships between these island mammal species is limited by the lack of many surviving taxa in the wake of Holocene extinction events that eliminated over 100 Caribbean mammal species or separate island populations (Davalos & Turvey 2012; Carstensen et al. 2013) (Cooke et al 2017). The poor preservation of DNA in tropical environments means that Caribbean mammal species designations and reconstruction of evolutionary histories has relied almost entirely upon morphological assignments that are often made from fragmented or limited skeletal material.

The Caribbean islands are home to one particularly unusual extant mammal family, the shrew-like lipotyphlan insectivore family Solenodontidae (Brandt 1833). This family contains two living species, which are found on the Greater Antillean islands of Hispaniola (Hispaniolan solenodon *Solenodon paradoxus*) and Cuba (Cuban solenodon *Atopogale cubana*). The remaining extant families within Lipotyphla include shrews (Soricidae), moles and desmans (Talpidae), and hedgehogs and gymnures (Erinaceidae) (Douady & Douzery 2003). Solenodons were, until recently, accompanied in the Caribbean by another family of smaller lipotyphlans, the Nesophontidae, which comprised the single genus *Nesophontes* (Anthony 1916a). Referred to as "island shrews", *Nesophontes* species were distributed across Cuba, Hispaniola, Puerto Rico, and the Cayman Islands (Miller 1929; Morgan 1994; Orihuela 2014; Turvey et al. 2007). (Table 1). All *Nesophontes* species are thought to have become extinct around 500 years ago, probably due to the accidental introduction of black rats (*Rattus rattus*) by the first European visitors to the Caribbean (MacPhee et al. 1999). This extinction event makes *Nesophontes* one of the few family-wide mammalian extinctions to have taken place in recent history (Turvey et al. 2007).

Recently, Brace *et al*. (2016) were able to use ancient DNA (aDNA) techniques to resolve the question of the relationship between the Solenodontidae and Nesophontidae, by recovering a near-complete mitochondrial genome and sequences for 17 nuclear genes from a 750-year-old Hispaniolan *Nesophontes paramicrus* specimen. This study was able to show that despite their morphological differences, these two families are sister taxa within the order Solenodonota, but diverged more than 40 million years ago.

Nesophontes taxonomy and specimen preservation

Nesophontes was described in 1916 (Anthony 1916), nearly 100 years after the description of *Solenodon paradoxus* (Brandt 1833). The first species described in the genus, *N. edithae*, was reported from Puerto Rico. Allen (1917, 1918) described a further two species, *N. longirostris* and *N. micrus*, from material found on Cuba and Cuba's largest offshore island, at the time called Isle of Pines (Isla de Pinos) and now known as Isle of Youth (Isla de la Juventud). A further three species *N. major*, *N. submicrus* (Arrendondo 1970) and *N. superstes* (Fischer 1977) were also described from Cuba. Fernandez et al (2005) and Silva Taboada et al. (2007) considered these three species to be synonyms, recognising only two species (*N. longirostris* and *N. micrus*) from Cuba and suggesting that the size difference between previously described species is due intraspecific ecotypic variation and sexual dimorphism. The main difference between Cuban Nesophontes species described on the basis of morphology is in size. Three *Nesophontes* species, *N. hypomicrus*, *N. paramicrus,* and *N zamicrus* have also been described from the island of Hispaniola (politically divided into the Dominican Republic and Haiti); these species were originally described from caves near St Michel, Haiti, but have since been collected from numerous cave deposits across the island (Miller 1929; Miller 1930; Brace et al. 2016). Like the Cuban *Nesophontes* species Hispaniolan *Nesophontes* species described on the basis of morphology are also mainly distinguished by differences in size. *N. paramicrus* was described as similar in size and characters to the Cuban *Nesophontes micrus* (Allen 1917), but differences in dental morphology supported its distinction at the species level. *N. hypomicrus* was described as resembling *N. paramicrus*, but was smaller in size, and again distinguished by significant difference in dental morphology. *N. zamicrus* specimens were smaller still; in fact the reduction in size is so significant that these specimens are considerably more fragile, and so most specimens only consist of the anterior portion of the skull. Descriptions of the Hispaniolan *Nesophontes* species are based on adult specimens, which can be morphologically distinguished from juvenile individuals due to patterns of tooth eruption and shape and the fusion of cranial bones, indicating that the differences between the three Hispaniolan *Nesophontes* species do not represent ontogenetic variation. However, currently recognised species-level variation in sympatric *Nesophontes* could actually reflect intraspecific variation, although this has been reassessed in Cuban species (Condis Fernandez et al. 2005; Silva Taboada et al. 2007), no study has used molecular data to re-evaluate the similar levels of variation also exhibited in Hispaniolan species.

Previous work on the taxonomy and evolution of *Nesophontes* has focused on using morphological, and particularly dental, characters to support the group's association with possible closest relatives, and to separate species within the genus. Species descriptions of the wider Caribbean mammal fauna reveal high

species diversity for this archipelago, of the kind often associated with island radiations. Although these kinds of insular radiations are documented on many island systems [5], including among other extant Caribbean vertebrate taxa (Glor et al. 2003), oceanic-type islands are rarely colonised by mammals and so comparative data on mammalian insular evolutionary patterns are largely lacking, making it difficult to infer whether Caribbean mammal species concepts based on limited morphological data alone are reliable. It is also not possible to determine the timing of divergence between different *Nesophontes* species on the basis of morphological data. In light of the recent aDNA evidence that provides an estimated 40 million year divergence date between Nesophontidae and Solenodontidae, the relationship between Hispaniolan *Nesophontes* species may represent a deep split, perhaps even at the genus level. However, our understanding of mammalian species relationships in the Caribbean needs to be better-resolved in order for in-depth evolutionary questions surrounding speciation, colonization and extinction to be addressed. Further evidence, such as molecular data, are required to test whether hypotheses of species diversity in island mammals based on morphology alone are likely to be correct. Molecular data obtained using ancient DNA techniques can therefore be used to contribute to our overall knowledge of species relationships and evolutionary processes in the Caribbean's land mammal fauna.

Table 1. Nesophontes species previously described on the basis on morphology, and their known geographical ranges in the Caribbean. Not all described taxa included in this table are currently considered to be valid.

Taxon	Range
Nesophontes edithae Anthony, 1916	Puerto Rico, Vieques, St. John, St Thomas
Nesophontes hypomicrus Miller, 1929	Hispaniola, Gonave Island
Nesophontes micrus Allen, 1917	Cuba, Isla de Juventud
Nesophontes paramicrus Miller, 1929	Hispaniola
Nesophontes zamicrus Miller, 1929	Hispaniola
Nesophontes longirostris Allen, 1917	Cuba
Nesophontes major Arredondo, 1970	Cuba
Nesophontes submicrus Arredondo, 1970	Cuba
Nesophontes superstes Fischer, 1977	Cuba
Nesophontes sp. Morgan, 1994	Grand Cayman
Nesophontes sp. Morgan, 1994	Cayman Brac

Delimitation of species

Species can also be defined by the phylogenetic or cladistic species concept, by generating a phylogeny via morphological or molecular methods to test for evolutionarily divergent lineages (Donoghue 1985), however species delimitation itself is notoriously controversial. Multiple species concepts give different criteria for species definitions (De Queiroz 2007). Defining extinct species known only from sub-fossil material is an even more complex issue, as although species identification is often achieved using phenotypic data, much of the data are lost in fossil specimens, where typically only some of the hard parts of the organism are preserved, and many distinct characteristics such as soft issue are lost (Mayr 2000). Efforts to define extinct species must therefore rely on the identification of preserved characters that provide a useful taxonomic signal within the intraspecific and ontogenetic variation expected for distinct species. Accurately identifying species using only scattered and fragmented fossilised remains is a difficult task, and several previous studies have highlighted the possibility that fossil specimens identified as different species may in fact be different morphs, sexes, or juveniles of the same species (Knell et al. 2013).

One of the main distinguishing features between previously described spatially and temporally co-occurring Cuban and Hispaniolan Nesophontes species is difference in size. In an early description of *N. edithae* from Puerto Rico, however, Anthony (1925) also suggested that individuals could be divided into two groups on the basis of size, but concluded that this size difference was due to sexual dimorphism. Sexual dimorphism can be defined as a significant behavioural or morphological difference between sexes, and is typically explained as the result of sexual selection (Andersson & Norberg 1981). Study of the “textbook” adaptive radiation of the Caribbean *Anolis* lizards has suggested that sexual dimorphism and interspecific divergence are alternative means of ecological diversification; sexual dimorphism may therefore be another way for species colonising island systems to exploit new or empty niches (Butler, Sawyer, & Losos, 2007). If this is the case, we might expect to find sexual dimorphism in insular genera that do not display multiple sympatric species. In the case of the Hispaniolan *Nesophontes* specimens, if morphological differences were purely an attribute of sexual dimorphism we would expect no significant genetic difference between two species. However, there is evidence for sexual size dimorphism in very few, if any living insectivores (Lindenfors et al. 2007), which indicates that sexual dimorphism is unlikely to account for observed variation in within-island samples of *Nesophontes*.

McFarlane (1999) later suggested that the observed *Nesophontes* size morphs from Puerto Rico may not be explained by sexual dimorphism, but instead represented different body size classes that were separated in time. This possibility illustrates a further difficulty of distinguishing and defining fossil species, because such taxa may be related to extant species in different ways; fossil species might represent an entirely extinct lineage sister to modern species, or they may represent fossil populations of extant modern species. Fossil specimens from different time intervals that differ in morphology (such as size) may represent morphological variation within a single evolving lineage, and can be categorised as “chronospecies” (Nadachowski 1993). There is evidence for within-species body size change across many vertebrate taxa during periods of Late

Quaternary climate change, including shrinking in large mammals (Schmidt & Jensen 2003) and birds (Stewart 1999). Because morphology can change over time within a single non-diverging species lineage, studies that designate species based purely on morphological evidence should be supported by additional analysis to confirm species status. If specimens are found in the same layer during excavation or if samples can be dated to confirm temporal sympatry, the possibility that they represent chronospecies can be ruled out. In situations where the above is not possible, molecular data can be used to help distinguish species by looking at genetic distances between sequence data and dating the divergence time between species. Unlike the Puerto Rican *Nesophontes* specimens, where morphological variation is spread across time, in Hispaniola different *Nesophontes* size morphs have been recovered from the same Late Quaternary horizons, meaning they occupied the same area when they were alive (Brace et al. 2016; Miller 1929; Miller 1930).

As well as chronospecies, two additional phenomena, insular size change (dwarfism and gigantism) (Van Valen, 1973) and insular radiations, may account for much of the diversity found on island systems in the recent fossil record. After the colonisation of an island system, species may radiate via allopatric speciation. Allopatric speciation occurs where populations are geographically separated, either by seaways between islands, or within larger islands where mountain ranges or valleys can drive intra-island speciation events (Schluter 2000; Losos & Ricklefs 2009).

Island dwarfism or gigantism occurs when a species arrives at an island, and because of the change in selective pressures in the surrounding environment, there is a change in the optimal size of the species. Island species are often smaller or larger than mainland forms (VanValen 1973). Whilst large mammals tend to shrink in size on islands (Lister, 1996), smaller mammals tend to increase in size leading to island gigantism. One example of an island giant is Amblyrhiza inundata, a large extinct caviomorph rodent, that inhabited the Lesser Antillean islands of Anguilla and St. Martin (Biknevicius et al. 1993). There is also evidence that evolution is accelerated in island mammals (Millien 2006) causing morphological changes such as size to occur more rapidly in insular species than in mainland forms.

Although multiple insular processes detailed above may have been acting on the recently extinct Caribbean mammal fauna, the extent to which certain processes generated the diversity exhibited in sub-fossil Hispaniolan species can be assessed, because Hispaniolan *Nesophontes* species were sympatric in both space and time. The size difference between Hispaniolan species therefore cannot be explained by either chronospecies or island dwarfism and gigantism. This does not exclude the possibility of allopatric speciation as the cause of intra-island diversity in Hispaniolan *Nesophontes*, however, as species separated by allopatric barriers may later find themselves in the same location if geological, geographical or ecological barriers are lifted (Schluter 2000). In Cuban green *Anolis* three species co-occur, but divergence dates suggest these secondarily sympatric species separated when past marine barriers isolated populations (Glor et al. 2004). Allopatric speciation is one process that may have generated species level diversity in Hispaniolan *Nesophontes* species, however morphological variation also occurs within temporally co-

occurring species, often artificially inflating the diversity of taxa described purely on the basis of morphological features. Limited hybridization is known to occur between sympatric populations of Cuban green *Anolis* (Glor et al 2004), as Hispaniolan *Nesophontes* are also sympatric, interspecific hybridization; as a result of secondary contact after the removal of allopatric barriers, is also a possibility. Mitochondrial DNA is often used as a marker for lineage diagnosis as due to the smaller effective population size of mitochondrial DNA, isoloated populations are likely to achieve reciprocal monophyly at a faster pace for mitochondrial DNA than for nuclear DNA (Wiens & PenKrot 2002). However mitochondrial DNA is affected by population processes due to hidden introgression events, incomplete lineage sorting, site saturation and the matrilineal pattern of inheritance; therefore particular caution must been taken with the molecular markers used for analysis.

Levels of intraspecific morphological and molecular variation in lipotyphlan insectivores

Within-species morphological variation caused by ontogenetic variation, sexual dimorphism and polymorphism is reported in almost all vertebrate systems (Hallgrimsson & Hall 2005). Some studies have suggested that this variation is increased on insular environments (Van Valen 1965). Incorrect understanding of intraspecific morphological variation can result in an erroneous estimation of species diversity, including within other insular lipotyphlans. For example, a new species of extinct Balearic insular shrew, *Nesiotites rafelinensis*, was described on the basis of a single specimen by Rofes & Cuenca-Bescós (2012). However, Furio & Pons-Monjo (2013) examined interspecific and intraspecific variability within *Nesiotites* and found that this genus displays a high degree of variability in monospecific assemblages, and demonstrated that the morphological differences used to describe *N. rafelinensis* fell within the intraspecific variation for the previously described species *N. ponsi*.

Whilst morphological differentiation between geographically separated populations and species is well documented (Gould & Johnston 1972), morphologically differentiated Hispaniolan *Nesophontes* species looked at as part of this study were sympatric. Within-habitat species richness is high in shrews, with as many as five separate sympatric species occupying one area (Kirkland, Jr. 1991). Size differentiation between recently diverged sympatric species has been suggested as a major evolutionary mechanism generating a reduction in interspecific competition, particularly when size difference coincides with trophic differentiation (Lack 1971). In a study of North American shrews, Fox & Kirkland (1992) found that species diversity in an area was not random, and that assemblages appeared to follow a rule whereby three species size classes were present. These three size classes mirror the species diversity found in Hispaniolan *Nesophontes* species. Fox & Kirkland (1992) found that there must be one species of each size class present in an area before a second species of one of any of the size classes was also found. The most likely cause of this non-random distribution of species is thought to be interspecific competition. A study of European shrews has found that crania were larger in allopatric populations: skull size was larger when only a single species was present, and smaller when more than one species co-occurred (Malmquist 1985).

In order to assess the accuracy of species status designated on the basis of morphology, molecular variation between purported species should also be investigated. For example, in the recently described Japanese mole *Mogera kanoana*, the genetic distance from the sympatric *M. insularis* was 7.5% for the mitochondrial gene cytochrome b (cytB) and 2.7% for 12s ribosomal RNA (12S). Critically, this study was also able to compare these genetic distance values with other living *Mogera* species and with intraspecific variation within *M. wogera* populations (cytB: 5.3%, 12s: 1.9%) (Kawada et al. 2007), to place the observed level of genetic variation seen by *M. kanoana* into wider taxonomic context. We are unfortunately currently unable to make similar comparisons across and within other extinct *Nesophontes* species to assess taxonomy against patterns of genetic variation. However, we can compare *Nesophontes* genetic divergence values with those found for other closely related mammal species pairs. Sequence divergence values for a sample of over 20 mammalian sister species have been shown to range from 4% to 11% (Bradley & Baker 2001). More specifically, molecular data from *Nesophontes* specimens could be compared to genetic distances and divergence dates recorded between different lipotyphlan insectivores, and between other small-bodied island mammal species such as Caribbean insular rodents and Madagascan tenrecs.

Recovery of molecular data from extinct Caribbean fauna

Molecular data, are required to test whether hypotheses of species diversity in Hispaniolan *Nesophontes* based on previous morphology-based studies are likely to be correct. However the extraction of ancient DNA from the tropics is a notoriously difficult task (Gutiérrez-garcía et al. 2014) Owl pellet remains of *Nesophontes* found at the surface of caves in Hispaniola were described by Miller (1930) to include tissue and hair, which led him to suggest that these species were still extant well into the 20th century. However, *N. hypomicrus* specimens with tissue and hair collected from Cueva Jurg, Parque Nacional Sierra de Baoruco, Dominican Republic, provided an uncalibrated radiocarbon date on bone collagen of 710 ± 50 before present (bp), suggesting extinction around the time of European arrival (MacPhee, Flemming & Lunde, 1999). Although the presence of organic material on *Nesophontes* specimens found on Hispaniola has proved not to be a reliable indicator of age, this unusual state of preservation may increase the chance of successful extraction of ancient DNA. This study also utilises Next Generation Sequencing techniques in order to increase the chances of successfully to extracting and sequencing endogenous aDNA from tropical environments, where DNA preservation is inhibited by high temperature and humidity (Gutiérrez-garcía et al. 2014).

3.3 Study overview and aims

This study aims to use specialist ancient DNA techniques in order to achieve four key aims: i) to extract and sequence ancient DNA for the first time from the extinct Hispaniolan lipotyphlan insectivore species *Nesophontes zamicrus* and *Nesophontes hypomicrus*; ii) to use these ancient DNA data to generate a genetic phylogeny including *N. hypomicrus*, *N. zamicrus* and *N. paramicrus*; iii) to date the divergences between these species and look at genetic distances for sequence data from each species, in order to

assess the extent to which these previously designated species represent genetically distinct species; iv) to assess the degree of sequence divergence between *Nesophontes* species in the context of data for other mammal species, in order to gain insight into the evolutionary processes that led to the generation of Late Quaternary mammalian diversity on Hispaniola.

3.4 Materials and methods

Specimen information

Seven individual *Nesophontes* partial crania, representing the two species *N. hypomicrus* and *N. zamicrus* that had not previously been analysed for ancient DNA, were obtained from a sub-fossil vertebrate collection from Cueva de Bosque Humido, Los Haitises National Park, Hato Mayor Province, Dominican Republic (19.077389 N, 69.477389 W, 44 m above sea level). Other sub-fossil specimens were also collected from the same deposit in this cave site, the majority of which were endemic small rodents, insectivores, bats, birds and reptiles, probably representing a prey accumulation made by the extinct giant Haitian barn owl. AMS dates for specimens excavated from this vertebrate assemblage (including a *Nesophontes* specimen not used for ancient DNA analysis) indicate a late Holocene, immediately pre-European age (*Nesophontes hypomicrus* skull: 734 ± 24 bp).

Micro-CT scanning of Nesophontes specimens

Micro-CT scanning was conducted, as specimens were so small that complete destruction was needed in order to extract aDNA (Appendix: Figure 1a). Micro-CT scans work by building up a series of 2D dimensional x-ray 'slices' into a full 3D representation of the object. This technique allows a virtual 3D representation of the specimen scanned, allowing visualisation of the inside of an object without destruction. These scans indicated the presence of soft tissue, confirming the unusual state of preservation of Hispaniolan *Nesophontes* species as described by Miller (1930). These scans not only preserve morphometric information about these rare specimens, but also ensure that future morphological study can be carried out.

Extractions, sequencing and phylogenetic analysis

The specimens were sampled by powdering the bone using a Mikro Dismembrator. Extractions (Dabney et al. 2013) and Next Generation Sequencing library builds (Meyer & Kircher 2010) took place in a dedicated aDNA laboratory at the Natural History Museum, London, physically separated from the post PCR laboratory, where all equipment used was sterilized with bleach and UV treated before and after use. A negative control was included in the set of samples during the entire process.

Screening of the *Nesophontes* samples was conducted using the Next Generation Sequencing platform Illumina NextSeq (Table 2.). To increase the amount of endogenous DNA sequenced, one sample with the

highest amount of reads mapped from screening was sent to the Department of Bioinformatics and Genetics at the Swedish Museum of Natural History for sequencing on an alternative Next Generation Sequencing platform Illumina HiSeq X. The Illumina NextSeq has an output of 120 Gb and produces around 400 million reads per run; in comparison, the Illumina HiSeq X is a system dedicated to large whole-genome sequencing and has an output of 1800 GB and can produce up to 6 billion reads per run.

Data from this sequencing run were mapped to the previously sequenced *N. paramicrus* whole mitochondrial genome and nuclear genes: ADRB2 (Adrenoceptor Beta 2), BDNF (brain derived neurotrophic factor), RAG1 (recombination activating gene 1), RAG2 (recombination activating gene 2), BCHE (butyrylcholinesterase), CREM (cAMP responsive element modulator), APP (Amyloid Beta Precursor Protein), BRCA1 (Breast Cancer 1 gene), TTN (titin), ADORA3 (Adenosine A3 Receptor), ATP7A (ATPase Copper Transporting Alpha), GHR (Growth Hormone Receptor), VWF (Von Willebrand Factor), BMI1 (BMI1 proto-oncogene, polycomb ring finger), EDG1 (endothelial differentiation G-protein coupled receptor 1) using CLC genomics workbench v.8 (CLC Bio-Qiagen, Aarhus, Denmark). Mapping parameters were as follows: Length fraction: 0.8, Similarity fraction: 0.8 (Table 2 and Appendix: Figure 3a). The resulting whole mitochondrial genome and nuclear genes were aligned and concatenated (Kearse et al. 2012; Larkin et al. 2007) and the partitioning scheme and best-fit models were chosen in PartitionFinder (Lanfear et al. 2012).

A maximum likelihood tree (Figure 2) with bootstrap support values was generated using RAxML v.8 (Stamatakis 2014) as implemented in CIPRES Science Gateway v.3. (Stamatakis et al. 2008; Miller et al. 2010). Bayesian trees were constructed using MrBayes (Ronquist & Huelsenbeck 2003) with four chains (three heated, one cold) that were run for 1×10^6 generations, sampling every 1×10^3 generations with a burn-in period of 250 trees (Appendix: Figure 2a).

Divergence dating analysis was completed using fossil calibrations and priors taken from Benton, Donoghue & Asher (2009) and dos Reis *et al.* (2012) (Appendix: Table 6a). Phylogeny and diversification dates were jointly estimated under an uncorrelated relaxed lognormal clock. The partitioning scheme and best-fit models chosen in PartitionFinder (Lanfear et al. 2012) were implemented and a Yule model of speciation was used; a birth-death model was run for comparison and generated identical topology. All other priors were left as default values in BEAUti v. 1.8.3 (Drummond et al. 2012). The analysis was run for 25 million generations, sampling every 1000 generations. Tracer v. 1.6.0 (Rambaut et al. 2014) was used to access convergence and effective sample size for all parameters after a burn-in of 10%. A maximum credibility tree was generated in TreeAnnotator v. 1.8.3 (Rambaut & Drummond n.d.) using trees sampled in the prior distribution.

A pairwise genetic distance analysis between *Nesophontes* species was also performed in MEGA v4 (Tamura et al. 2007) to assess the validity of previous morphology-based species delimitation. Separate nuclear and mitochondrial genes were used for genetic pairwise distance analysis. Pairwise genetic sequence divergences (calculated as Kimura two-parameter distances) were generated for *Nesophontes*

species as well as for species from other lipotyphlan insectivore genera taken from Genbank (Benson *et al.*, 2005) for four nuclear and two mitochondrial genes (Appendix: Tables 4.1a-4.7a).

3.5 Results

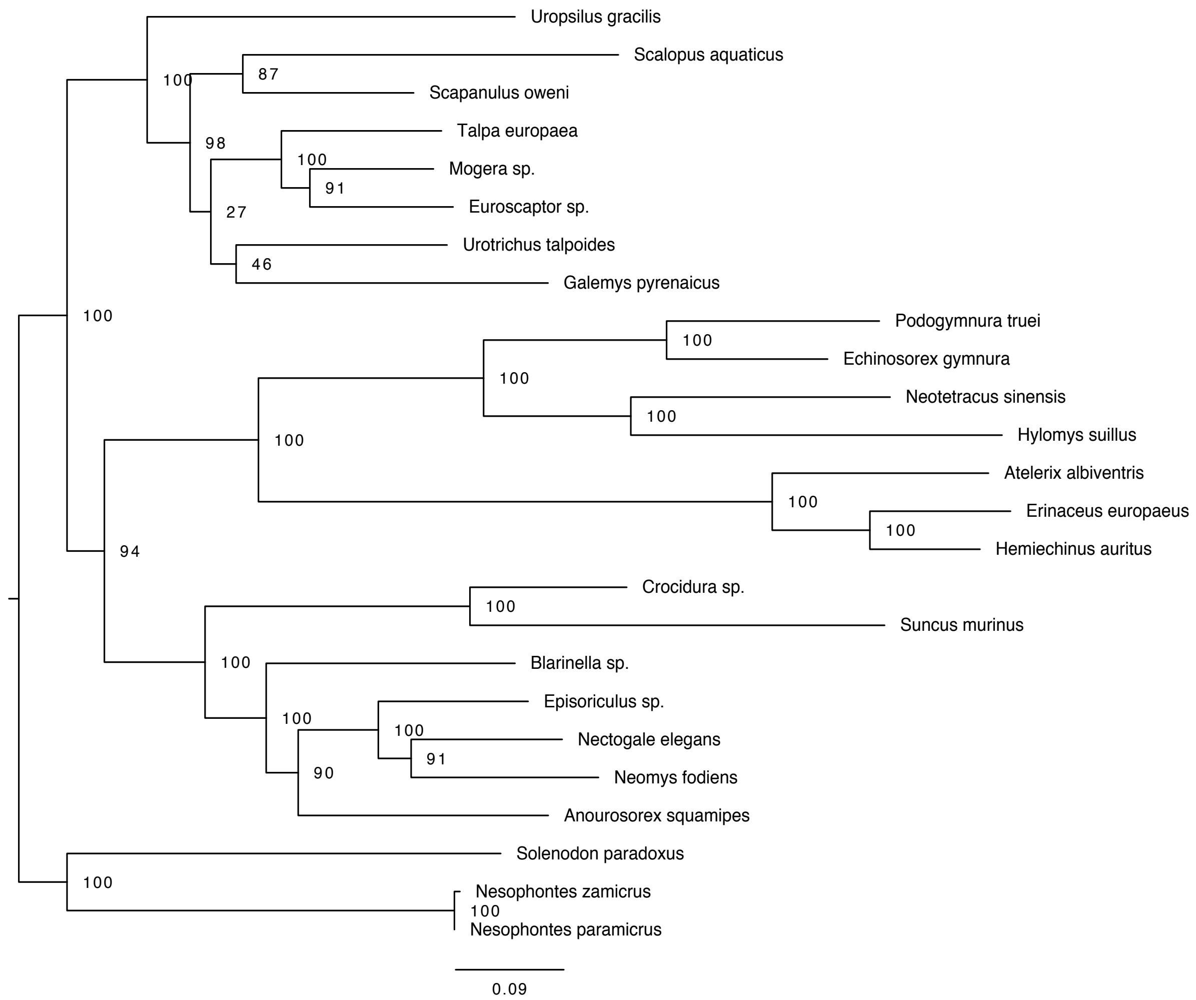
We were successfully able to extract and sequence aDNA from a single specimen of the extinct endemic Caribbean insectivore *Nesophontes zamicrus,* although additional aDNA extraction from *N. hypomicrus* was unfortunately not successful. This study is therefore limited to the investigation of the evolutionary relationship between *N. paramicrus* and *N. zamicrus*, rather than between all three species in this endemic Hispaniolan species group. A whole mitochondrial genome (15,280 bp) and 8 partial nuclear genes (APP, BMI1, CREM, PLCB4, ADORA3, APOB, ADRA2B and ADRAB2) were recovered, from the second *Nesophontes* species to be subjected to molecular study. As a result of HiSeqX sequencing a total of 58,248 reads were mapped to the reference mitochondrial genome used in our analysis with an average depth of coverage of 65.87 (Table 2).

Table 2. Sequencing results for N. zamicrus sample used for phylogenetic analysis with total reads mapped and averaged depth of coverage over the whole mitochondrial genome reference sequence used. Endogenous content was estimated using mapping results for mitochondrial genome reference sequence used.

Species	*Sequencing platform*	*Total nucleotides*	*Number of paired reads*	*Total reads mapped*	*Average Depth of coverage*	*Estimated endogenous content*
Nesophontes zamicrus	NextSeq	692,115,456	3,189,108	15	0.04	0.000470351
	HiSeq X	53,008,411,080	919,533,272	58,248	65.87	0.006334518

The maximum likelihood (Figure 1) and Bayesian phylogenetic analyses (Figure 2 and Appendix: Figure 2a) give congruent topology to the previous study of lipotyphlan relationships, which also included *Nesophontes* ancient DNA data, but which included only one described *Nesophontes* species (Brace et al. 2016). Our analysis places the newly sequenced *N. zamicrus* as most closely related to the previously sequenced *N. paramicrus*, as predicted by previous morphological analysis (Miller 1929; Miller 1930). The dated phylogeny (Figure 3) shows that these two *Nesophontes* species diverged from one another relatively recently, around 0.6 Ma (95% HPD lower/upper = 0.232–1.037 Ma). The two mitochondrial genes analysed (cytB and 12S) give pairwise genetic distance values of 0.053 (5.3%) and 0.022 (2.2%) respectively. All four nuclear genes analysed gave values above zero (Appendix: Tables 4a-4.9a).

Figure 1. Maximum likelihood phylogeny generated using RAxML, including N. zamicrus and N. paramicrus and extant lipotyphlan insectivores for which whole mitochondrial genome data are available. Node values represent bootstrap support (100 replicates).



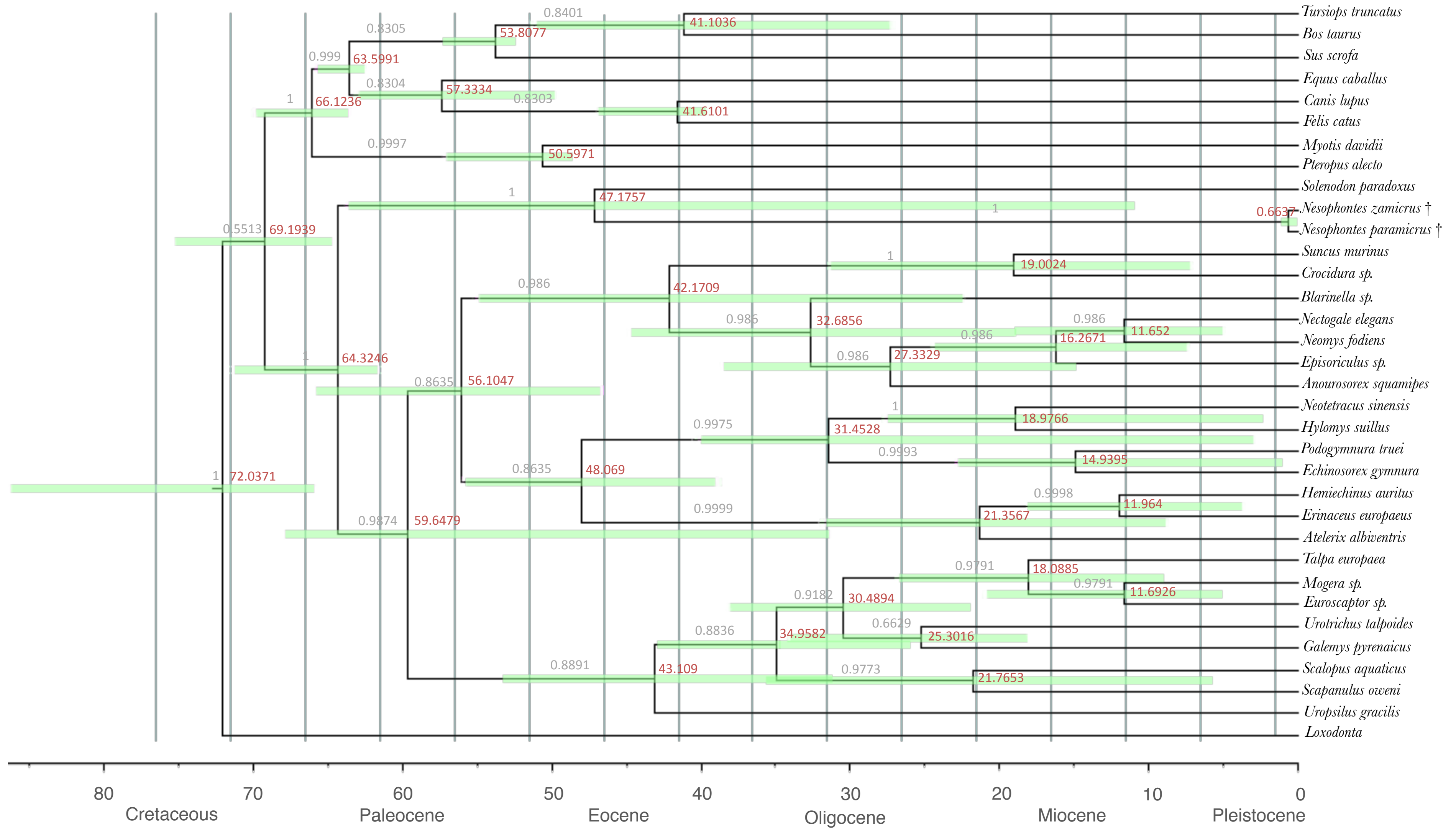


Figure 2. A time-calibrated phylogeny generated in BEAST, showing estimated divergence dates for N. zamicrus and N. paramicrus. Estimates of median divergence dates are shown (in red) above the nodes. Node bars indicate the 95% highest posterior density values. Posterior probability (pp) are shown (in grey) on branches.

3.6 Discussion and conclusions

Overview

This study has been able to successfully extract and sequence further novel ancient DNA data from the extinct Caribbean insectivore *Nesophontes*. Specifically, this is the first molecular work to be carried out on the smallest species in the genus, *N. zamicrus*. Because this is only the second molecular study to look at a species within the Nesophontidae, which represents an evolutionarily ancient lineage within the Lipotyphla, and because little is still known about the wider evolutionary history of lineages within this genus, the relationship between *N. zamicrus* and *N. paramicrus* may have represented a deep split, perhaps even at the genus level. However, we were instead able to show that the two Hispaniolan *Nesophontes* species for which we now have molecular data, *N. zamicrus* and *N. paramicrus*, diverged from one another during the late Pleistocene (95% HPD lower/upper = 0.2315–1.0372 Ma). Notably, Cuban and Hispaniolan solenodon species diverged from each other at least as early as the Early Pliocene (3.7–4.8 Ma [95% CI = 2.6–6.4 Ma]) (Sato et al. 2016), and previous estimates have suggested a much earlier date (25 Ma [95% CI = 16–38 Ma]) (Roca et al. 2004). The recent divergence between *N. zamicrus* and *N. paramicrus* therefore brings into question the validity of the species distinctiveness of *N. zamicrus* and other Caribbean *Nesophontes* species, all of which have been described on morphological grounds alone and often primarily on the basis of size differences.

Two possible hypotheses could explain the relatively recent divergence date between *N. zamicrus* and *N. paramicrus*: i) these two taxa may in fact represent within-species variation rather than two distinct species, for example due to either sexual dimorphism or other morphological variation, or ii) *N. zamicrus* and *N. paramicrus* could be two genetically distinct lineages that diverged recently due to either allopatric or peripatric speciation, caused by changes in their environment that arose during the fluctuating climate of the Pleistocene.

Interpretation of genetic distances and divergence dates in comparison with other lipotyphlan species

To determine whether phylogenetic analysis of mitochondrial and nuclear sequence variation supports a two-species hypothesis for *N. zamicrus* and *N. paramicrus,* a comparison with divergence dates of species that are part of other lipotyphlan radiations is necessary. *Nesophontes* species investigated in this study may be part of a single population where incomplete lineage sorting explains the genetic divergence between samples. Incomplete lineage sorting results in misleading estimations of phylogenetic relationships due to the retention and stochastic sorting of ancestral polymorphisms in a population. As the number of individuals sampled in this study was low, due to the restrictions on the extraction of aDNA from sub-fossil material from the tropics, the

possibility of incomplete lineage sorting being the causal factor for observed species diversity is higher. However increasing the number of loci used in a study has been shown to provide a more accurate phylogenetic tree, and this study was able to utilise multiple nuclear genes and the whole mitochondrial genome (Maddison & Knowles 2006). As the two *Nesophontes* species investigated in this study were recovered from the same location, *Nesphontes* species looked at as part of this study are likely two recently diverged populations that co-existed in the same landscape at the same time, because *N. zamicrus* and *N. paramicrus* appear to be sympatric, a comparison of phylogenetic divergences between lipotyphlan species which also exhibit this pattern of overlapping or shared distributions with closely related species is also necessary, to assess the relationship between Hispaniolan *Nesophontes*.

The genetic pairwise distance for cytB between these *Nesophontes* species is still relatively low (2.2%) in comparison to other closely related lipotyphlan species (Tables 3.3). Many studies use the genetic distance between mitochondrial genes such as cytB to delimit species and subspecies by comparing values found to closely related taxa of known species-level distinction. The relatively low level of divergence between *N. zamicrus* and *N. paramicrus* for cytB in comparison to other lipotyphlan species suggests that *N. zamicrus* and *N. paramicrus* are very closely related and may only represent sub-species level genetic divergence. However because *Nesophontes* species examined here were not geographically isolated, and therefore may have been subject to secondary contact introgression, we would expect levels of genetic distance to be far lower than that of allopatric lipotyphlan species.

If both *N. zamicrus* and *N. paramicrus* actually represent the same species, we might expect pairwise genetic distance between nuclear genes to be low or near to zero, as nuclear genes evolve more slowly than mitochondrial genes (Avise 1994). However, many of the pairwise genetic distances between incomplete nuclear genes analysed as part of this study are higher or equal to those of other lipotyphlan species included in our analysis (Appendix: Tables 4.1a-4.7a). Suggesting that the genetic distance between the two *Nesophontes* species included in this study represents species-level divergence. Considering the geographic distribution of our study species differences between pairwise distance analyses of mitochondrial and nuclear genetic markers may be the result of recent introgression of mitochondrial DNA between the two divergent species (Irwin et al., 2009). The inclusion of multiple nuclear molecular markers used in this study, were therefore highly important for corroboration of phylogenetic analysis based on mitochondrial genes.

The most recent available reported species divergences within the subfamilies Erinaceinae and Galericinae (gymnures and hedgehogs) are in the genus *Erinaceus* and date to the early Pleistocene (*E. europaeus* / *E. amurensis* = 1.06 Ma, 95% HPD = 0.6-1.6 Ma; *E. roumanicus* / *E. concolor* = 0.89 Ma, 95% HPD = 0.4-1.4 Ma) (He et al. 2012). A study of phylogenetic relationships and species relationships in Eurasian moles (*Talpa*) found that all studied sympatric species of

moles diverged before the Pleistocene, from 5.48-6.65 Ma to 3.09-4.65 Ma (95% HPD). Intraspecific divergence events were dated to the Early to Middle Pleistocene, 1.4 to 0.17 Ma. However, this pattern of early divergence may be due to the specific life history and evolution of moles, such as their dispersal ability and conserved morphological features that are restricted by their fossorial lifestyle (Bannikova et al. 2015). Few species relationships in the shrews (Soricidae) have late Pleistocene divergence dates. However, the most recent available divergence date between species in this family is mid-Pleistocene, between *Anourosorex squamipes / Anourosorex yamashinai* (0.66 Ma, 95% HPD = 0.18-1.35 Ma) (He et al. 2010), a comparable divergence date to the split between the two *Nesophontes* species included in this study. A study using cytB and nuclear region BRCA1 in the lesser white-toothed shrew (*Crocidura suaveolens* group) found that the group split into six distinct clades in the late Pliocene to early Pleistocene (1.9-09 Ma), and that these clades included populations exhibiting different morphotypes which should potentially be revaluated as representing distinct species (Dubey et al. 2006). The divergence dates estimated for *Nesophontes* species investigated as part of this study are relatively recent when compared with dates produced by other studies investigating lipotyphlan species. However Late Pleistocene divergence dates between Hispaniolan *Nesophontes* species do fall within the range of species divergence dates observed in lipotyphlan species, and because Hispaniolan *Nesophontes* species were sympatric we would expect the divergence date to be in the low end of the range for lipotyphlan species that continue to be geographically isolated.

Interpretation of genetic distances and divergence dates in comparison with other mammals

Comparison of genetic divergence data with extant lipotyphlan insectivores is limited by the general absence of well-resolved species-level molecular phylogenies for other lipotyphlan radiations. This limitation in part comes from the fact that lipotyphlans are amongst the most species-rich but understudied of all mammals. In shrews, for example, 376 species have been described so far from 26 genera (Willows-Munro & Matthee 2011). Mammals themselves, and insectivores in particular, are rarely found on oceanic-type islands, further limiting potential comparisons with the evolutionary history of other groups that may be expected to have experienced similar patterns of diversification.

The Malagasy tenrecs, although not lipotyphlan insectivores, may be comparable to *Nesophontes,* as they also form part of a rare mammalian oceanic-type island radiation and some species fill a similar ecological niche to the shrew-like *Nesophontes*. The existence of extant cryptic sympatric species of long-tailed shrew tenrecs (*Microgale*) has been revealed by combined mitochondrial DNA and morphological analysis (Olson et al. 2004). Studies using the mitochondrial NADH dehydrogenase subunit 2 gene found pairwise distance measures of 10.6–16.6% between sister species of oryzorictines (shrew and mole tenrecs) (Olson et al. 2009). Unfortunately, however, divergence dates between sister species are still unavailable for the majority of the Malagasy tenrecs.

Although a late Pleistocene divergence date seems recent in comparison to many other lipotyphlan species, uncontroversial species relationships in other mammal groups show similar divergence times. Indeed, many of these groups have much longer generation lengths than lipotyphlans, making these temporal divergence estimates reflect an even shorter number of generations. For example, polar bears (*Ursus maritimus*) and brown bears (*U. arctos*) diverged an estimated 600,000 years ago, with adaptation to a novel polar environment the likely driving force for rapid evolution of the major morphological differences between these species, although repeated hybridization is also thought to be an interacting factor in this specific case (Hailer et al. 2012). Recent and rapid speciation has also been demonstrated in *Rattus*, with multiple species arising in the mid-Pleistocene (Rowe et al. 2011). Other recognized Quaternary species-level divergences include many northern hemisphere species such as Eurasian and American lynx (*Lynx lynx* and *L. canadensis*) and three species of lemmings (*Lemmus*) (Lister 2004). However, these species are all found in temperate climates, where evidence suggests that species divergences were more recent and rapid due to the effects of climatic change between glacial and interglacial cycles. In the tropics, population-level divergences instead tend to extend further back to the early Pleistocene or Pliocene (Barnosky 2005).

The pairwise genetic distance value for the mitochondrial gene cytB in *Nesophontes* species is low in comparison with values for other mammal sister species, but is higher than average values for intra-population and sub-specific divergence (Bradley & Baker 2001). Sequence divergences in mtDNA between over 90 sister species of mammals have a median value that equates to a divergence date of 3.2 Ma under a standard molecular clock (Avise et al. 1998). This estimated divergence date is earlier than the estimated separation of the two Hispaniolan *Nesophontes* species found in this study. However, after using median phylogroup separation times to correct for the fact that speciation occurs over time and is not a point event, Avise *et al*. (1998) found that at least 50% of mammalian sister species speciation events occurred during the Pleistocene, which is broadly consistent with the observed 95% CI of estimated divergence seen between Hispaniolan *Nesophontes*.

Hispaniolan climate and geology

Pleistocene climate change in the tropics, and the associated formation of geographical barriers on the geotectonically complex island of Hispaniola, could explain the recent mid-Pleistocene divergence date observed between the two species of *Nesophontes*. Pleistocene climate was marked by repeated and rapid glacial cycles that could have interacted with geological features found on Hispaniola known to promote speciation (e.g., large island size, high mountain ranges and frequently submerged valleys) to provide vicariant and/or dispersal events that separated populations and drove speciation. Specifically, there is evidence of Quaternary glaciation on Hispaniola that would have caused the climate to be much drier than it is today, with cycles of xeric

and mesic habitats potentially driving population fragmentation (Schubert & Medina 1982). Whilst Pleistocene climate change would have affected other Greater Antillean islands, Hispaniola is perhaps more likely to generate speciation events due to the large size of the island (Losos & Schluter 2000). Hispaniola has some unique geological features that may also impact insular species diversity, including the hightest mountain ranges in the Greater Antilles and low valleys that were periodically flooded. Hispaniola also has a complex geological history and is made up of northern and southern palaeo-islands (Figure 3). Sea level rise and fall would have caused southern and northern parts of the island to be cut off by transient seaways in the Neiba Valley, a region that was periodically flooded during the Pleistocene. Bond's line formed a further flooded boundary subdividing the southern palaeo-island into eastern and western regions (Maurrasse et al. 1980) at the estimated time of the divergence between *N. zamicrus* and *N. paramicrus*. As *N. zamicrus* and *N. paramicrus* were sympatric, for allopatric speciation to have generated Hispaniolan *Nesophontes* species diversity, these potential barriers to gene flow must have been intermittent, allowing genetically diverged populations to move back into the same landscape after speciation had occurred.

Recent divergence of lineages followed by secondary sympatry has been demonstrated in insular birds (Andersen et al. 2015) and insects (Nattier et al. 2012) and more generally as a mechanism for producing observed mammalian diversity (Fitzpatrick & Turelli 2006). Currently only a single species is recognised on each of the other Greater Antillean islands where *Nesophontes* is found. This could be due to actual differences in the processes that generate species, for example some of the islands where *Nesophontes* are found are much smaller than Hispaniola (e.g.: the Cayman islands) and may have been unable to support multiple species. Or artificial differences in species delimitation, because genetic analysis has not yet been conducted on *Nesophontes* species from other Greater Antillean islands, where previous studies have suggested morphological differences (e.g. Cuba).

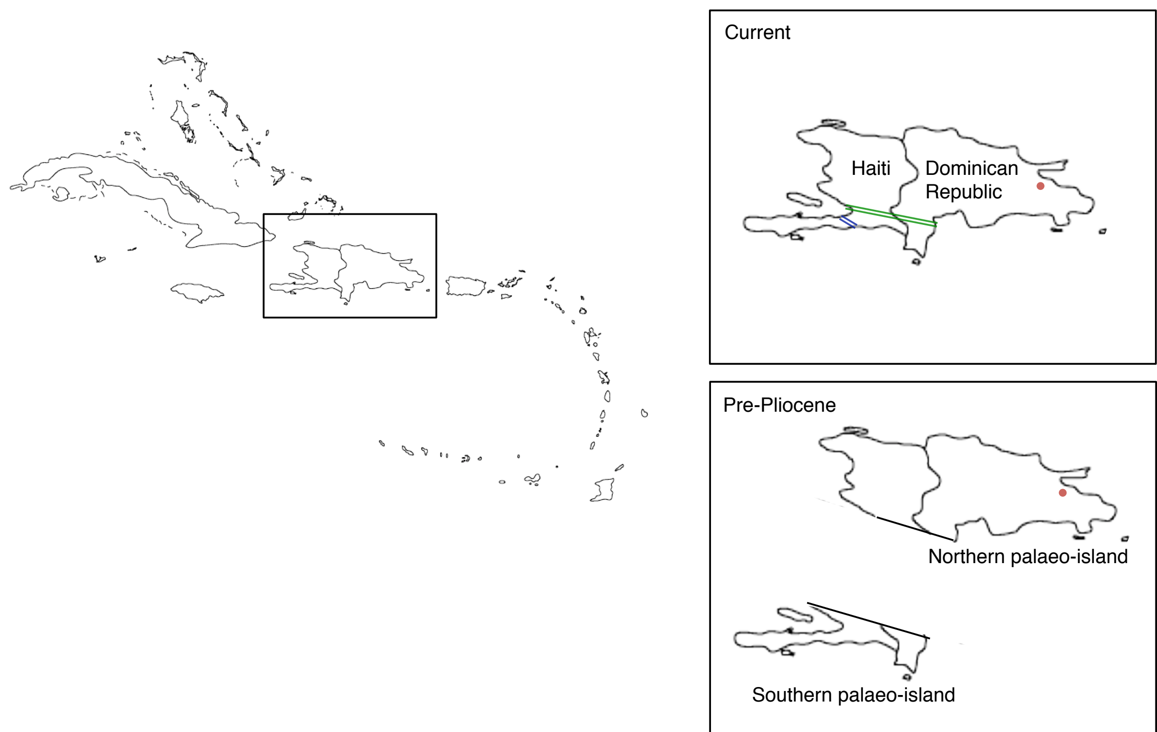


Figure 3. Map of the Caribbean and the island of Hispaniola, showing geological features of interest including paleo-island boundaries. Red dots indicate specimen collection location. Green line indicates the Neiba Valley. Blue line indicates the Jacmel-Fauché depression or Bond's Line.

A comparison with other Hispaniolan vertebrate radiations

Data available for other Hispaniolan mammals could also shed light on whether biological barriers that arose in the Pleistocene could have shaped the diversity of the island's endemic species, including *Nesophontes*. The Hispaniolan hutia (*Plagiodontia aedium*) is the only extant endemic Hispaniolan rodent, and part of the Caribbean's capromyid radiation. Divergences between lineages in the Hispaniolan hutia indicate that lineage separation between northern and southern palaeo-island populations occurred during the Pleistocene around 0.594 Ma (95% HPD), with a similar divergence date of 0.436 Ma (95% HPD) for east-west lineage separation in the southern palaeo-island. However, genetic differentiation between allopatric populations of the Hispaniolan hutia has been interpreted as representing subspecies-level rather than species-level differentiation, as associated morphological differentiation is relatively limited. Levels of sequence divergence between *Plagiodontia* subspecies ranged from 1.03-2.86% (Brace et al. 2012). A recent combined morphological and molecular analysis looking at the evolutionary relationships of extant Hispaniolan *Solenodon* populations also found three evolutionarily distinct allopatric populations representing distinct lineages (northern, southeastern and southwestern), and suggested subspecies designation for these genetically and morphologically distinct populations. Sequence divergence between

regionally sampled *Solenodon* populations for mitochondrial gene cyt b ranged from 1.10 – 2.20% (Turvey et al., 2016). A mitogenomic study of the extant Hispaniolan *Solenodon* also found evidence for distinct northern and southeastern populations and showed that the southeastern population had very limited genetic diversity (Brandt et al. 2016). The genetic divergence between Hispaniolan *Nesophontes* species investigated as part of this study is within the range found for both Hispaniolan *Plagiodontia* and *Solenodon* allopatric populations, suggesting a relatively high level of divergence in the sympatric *Nesophontes* species (2.2%). Although the genetic distance between *Plagiodontia* populations studied has been interpreted as sub-species level divergence, this is due to the limited morphological differentiation between prospective species. In *Nesophontes* not only have previous studies already established morphological differences between species, but species included in this analysis were sympatric; due to the lack of geographical isolation we would expect a lower level of genetic divergence between *Nesophontes* species than that found in allopatric *Plagiodontia* populations. In *Solenodon*, where genetic divergence levels between studied geographically isolated populations are comparable to sympatric *Nesophontes* species, recent studies suggest sub-species level distinction.

Radiations of other Hispaniolan vertebrates may also give some insight into the mechanisms that generated Pleistocene speciation events and current diversity on the island. Although there is some evidence of recent isolation in allopatric populations of Hispaniolan bird species, within-island speciation appears not to be the main causal factor in generating much of the current avian diversity; instead, geological processes predating the island's current formation are the likely cause (Sly et al. 2011). Allopatric genetic differentiation and population fragmentation caused by the flooding of Pliocene/Pleistocene seaways has been demonstrated in the Hispaniolan teiid lizard *Ameiva chrysolaema* (Gifford et al. 2004)m and patterns of geographic genetic differentiation have also been found in Hispaniolan *Anolis* lizards (Glor et al. 2003).

Allopatric speciation generated by Plestocene barriers to gene flow has been demonstrated in multiple Hispaniola vertebrate lineages including mammals, birds and reptiles. These same barriers to gene flow are likely to have also affected Hispaniolan *Nesophontes* species, which occupied the same insular environment. If recent Pleistocene events are accepted as the causal factors behind species diversity in other Hispaniolan vertebrate groups, a recent Pleistocene species level divergence in *Nesophontes* resulting in secondarily sympatric species, is a plausible explanation for the observed morphological and genetic diversity in sub-fossil specimens.

This study suggests that the relationship between two sympatric *Nesophontes* species may be more close than originally interpreted from morphological study of sub-fossil specimens. Molecular data suggest that *N. zamicrus* and *N. paramicrus* diverged recently, possibly as a result of mid-Pleistocene climate change. Although the divergence between *N. zamicrus* and *N. paramicrus* is recent in comparison with the few divergence data available for other lipotyphlan species, as these

two morphospecies represent two sympatric, morphologically and genetically distinct populations, a late Pleistocene (95% HPD lower/upper = 0.2315–1.0372 Ma) divergence date implies that *N. zamicrus* and *N. paramicrus* are distinct, albeit recently diverged species.

The disparity between morphological and molecular findings here highlights the difficulty in i) accurately describing and defining sub-fossil species from fragmented and scattered remains, and ii) disentangling questions over the status and validity of putative extinct species by extraction, sequencing and analysis of damaged and contaminated ancient DNA. Morphological interpretation of species status may have been further hindered by the fact that despite their recent divergence date, *N. zamicrus* and *N. paramicrus* display significant differences in relative body size. Islands are known to promote body size evolution in animal lineages, which are also often characterised by accelerated evolution (Millien 2006), unique or unusual morphological adaptations, and convergence (Losos & Ricklefs 2009b). Interspecific competition has also been shown to have generated multiple sympatrically occurring size classes or species in mainland insectivore groups (Kirkland, Jr. 1991). On islands, lack of competition from more distantly related mammals and other fauna, as well as reduced predation, may increase interspecific competition and drive morphological differentiation (Pinto et al. 2008). The high level of relative size differentiation between *Nesophontes* species may therefore have allowed the species to exploit different niches.

This study has demonstrated that the Late Quaternary Caribbean land mammal fauna consisted of not only deep divergences, such as the split between Nesophontide and Solenodontidae, but also recent radiations that were likely the result of changes in Quaternary environments. Studies of kiwi species (*Apteryx* spp.) from New Zealand provide another example of how island environments can protect and preserve ancient species but also generate recent radiations, made evident in kiwis by high population structure and cryptic species (Baker et al. 1995; Shepherd & Lambert 2008). The recent intra-island evolutionary radiation in extinct *Nesophontes* revealed here for the first time shows that Hispaniola not only represents an isolated island haven for the continued survival of ancient species, but has also persisted as a hotspot for the generation of species, until the recent loss of much of this diversity following human arrival.

References

Allen, G.M., 1918. Fossil mammals from Cuba. *Bulletin of the Museum of comparitive Zoology*, 62, pp.131–148.

Allen, G.M., 1917. Fossil mammals from Cuba. *Bulletin of the Museum of comparitive Zoology*, 61, pp.1–12.

Andersen, M.J. T Shult, H., & Cibois, A., 2015. Rapid diversification and secondary sympatry in Australo-Pacific kingfishers (Aves: Alcedinidae: Todiramphus). *Royal Society Open Science*, 2(140375), pp.1–13. Available at: http://rsos.royalsocietypublishing.org/ on.

Andersson, M. & Norberg, R., 1981. Evolution of reversed sexual size dimorphism and role partitioning among predatory birds, with a size scaling of flight performance. *Biological Journal of the Linnean Society*, 15.2, pp.105–130.

Anthony, H.E., 1925. Mammals of Porto Rico, living and extinct. Science Survey Puerto Rico and Virgin Islands. *New York Academy of Science*, 9, pp.1–96.

Anthony, H.E., 1916. Preliminary disagnosis of an apparently new family of insectivores. *Bulletin of the American Museum of Natural History*, 35, pp.726–728.

Avise, J.C., 1994. *Molecular Markers, Natural History and Evolution*, New York: Chapman & Hall.

Avise, J.C., Walker, D. & Johns, G.C., 1998. Speciation durations and Pleistocene effects on vertebrate phylogeography. *Proceedings of the Royal Society B: Biological Sciences*, 265(1407), pp.1707–1712.

Baker, A.J. et al., 1995. Flightless brown kiwis of New Zealand possess extremely subdivided population structure and cryptic species like small mammals. *Proceedings of the National Academy of Sciences of the United States of America*, 92(18), pp.8254–8258.

Bannikova, A.A. et al., 2015. An underground burst of diversity - a new look at the phylogeny and taxonomy of the genus *Talpa* Linnaeus, 1758 (Mammalia: Talpidae) as revealed by nuclear and mitochondrial genes. *Zoological Journal of the Linnean Society*, 175(4), pp.930–948.

Barnosky, A.D., 2005. Effects of quaternary climatic change on speciation in mammals. *Journal of Mammalian Evolution*, 12(1–2 SPEC. ISS.), pp.247–264.

Benson, D.A. et al., 2005. GenBank. *Nucleic Acids Research*, 33, pp.34–38.

Benton, M.J., Donoghue, P.C.J., Asher, R.J., 2009. Calibrating and constraining molecular clocks. In S. Hedges, S. B., Kumar, ed. *The Timetree of Life*. Oxford: Oxford University Pess, pp. 35–86.

Biknevicius, A.R., McFarlane, D. a. & Macphee, R.D.E., 1993. Body Size in Amblyrhiza inundata (Rodentia: Caviomorpha), an Extinct Megafaunal Rodent from the Anguilla Bank, West Indies: Estimates and Implications. *American Museum Novitates*, 3079(3079), pp.1–25.

Brace, S. et al., 2016. Evolutionary history of the Nesophontidae, the last unplaced Recent mammal family. *Molecular biology and evolution*, 33(12), p.msw186.

Brace, S. et al., 2012. Population history of the Hispaniolan hutia *Plagiodontia aedium* (Rodentia: Capromyidae): Testing the model of ancient differentiation on a geotectonically complex

Caribbean island. *Molecular Ecology*, 21(9), pp.2239–2253.

Bradley, R.D. & Baker, R.J., 2001. A Test of the Genetic Species Concept: Cytochrome-b Sequences and Mammals. *Journal of Mammalogy*, 82(4), pp.960–973. t.

Brandt, A.L. et al., 2016. Mitogenomic sequences support a north–south subspecies subdivision within *Solenodon paradoxus*. *Mitochondrial DNA Part A*, 1394(April), pp.1–9..

Brandt, J.F., 1833. De Solenodonte, novo mammalium insectivorum genere. *Memoires de l'Académie Impériale des Sciences de St. Pétersbourg, sér*, 2, pp.459–478.

Butler, M.A., Sawyer, S.A. & Losos, J.B., 2007. Sexual dimorphism and adaptive radiation in *Anolis* lizards. *Nature letters*, 447, pp.202–205.

Carstensen, D.W. et al., 2013. The functional biogeography of species: Biogeographical species roles of birds in Wallacea and the West Indies. *Ecography*, 36(10), pp.1097–1105.

Condis Fernández, M. M., Jiménez Vásquez, O., & Arredondo, C., 2005. Revisión taxonómica del género Nesophontes (Insectivora: Nesophontidae) en Cuba: análisis de los caracteres diagnóstico. *Monogra es de la Societat d'Història Natural de les Balears*, *12*, 95-100.

Dabney, J. et al., 2013. Complete mitochondrail genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences*, 110(39), pp.15758–15763.

Darwin, C., 1859. *On the origin of species by natural selection*, New York: P. F. Collier & Son.

Davalos, L.M. & Turvey, S.T., 2012. West Indian Mammals: The old the new and the recently extinct. In *Bones, Clones, and Biomes: The History and Geography of Recent Neotropical Mammals*. pp. 157–202.

Donoghue, M.J., 1985. A Critique of the Biological Species Concept and Recommendations for a Phylogenetic Alternative. *American Bryological and Lichenological Society*, 88(3), pp.172–181.

Douady, C.J. & Douzery, E.J.P., 2003. Molecular estimation of eulipotyphlan divergence times and the evolution of "Insectivora." *Molecular Phylogenetics and Evolution*, 28(2), pp.285–296.

Drummond, A.J. Suchard, M. A., Xie, D., & Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST. *Molecular Biology and Evolution*, 29, pp.1969–1973.

Dubey, S. et al., 2006. Pliocene and Pleistocene diversification and multiple refugia in a Eurasian shrew (*Crocidura suaveolens* group). *Molecular Phylogenetics and Evolution*, 38(3), pp.635–647.

Fitzpatrick, B.M. & Turelli, M., 2006. The Geography of Mammalian Speciation: Mixed Signals From Phylogenies and Range Maps. *Evolution*, 60(3), pp.601–615.

Gerrit S. Miller, J., 1929. Three Small Collections of Mammals from Hispaniola. *Smithsonian Miscellaneous Collections*, 82(15).

Gifford, M.E., Powell, R., Larson, A., & Gutberlet, R. L., 2004. Population structure and history of a phenotypically variable teiid lizard (*Ameiva chrysolaema*) from Hispaniola: The influence of a geologically complex island. *Molecular Phylogenetics and Evolution*, 32(3), pp.735–748.

Glor, R.E. et al., 2004. Partial island submergence and speciation in an adaptive radiation: a multilocus analysis of the Cuban green anoles. *Proceedings. Biological sciences / The Royal*

Society, 271(1554), pp.2257–2265.

Glor, R.E. et al., 2003. Phylogenetic Analysis of Ecological and Morphological Diversification in Hispaniolan Trunk-Ground Anoles (*Anolis Cybotes* Group). *Evolution*, 57(10), pp.2383–2397.

Gould, S.J. & Johnston, R.F., 1972. Geographic Variation. *Annual Review of Ecology and Systematics*, 3, pp.457–498.

Gutiérrez-garcía, T. a et al., 2014. Ancient DNA and the tropics : a rodent's tale Ancient DNA and the tropics : a rodent's tale. *Biology letters*, (June).

Hailer, F. et al., 2012. Nuclear Genomic Sequences Reveal that Polar Bears Are an Old and Distinct Bear Lineage. *Science*, 336(6079), pp.344–347.

Hallgrimsson, B. & Hall, B.K., 2005. *Variation and variability: central concepts in biology*, New York.

He, K. et al., 2010. A multi-locus phylogeny of Nectogalini shrews and influences of the paleoclimate on speciation and evolution. *Molecular Phylogenetics and Evolution*, 56(2), pp.734–746.

He, K. et al., 2012. An estimation of erinaceidae phylogeny: A combined analysis approach. *PLoS ONE*, 7(6).

Kawada, S.I. et al., 2007. Revision of the mole genus *Mogera* (Mammalia: Lipotyphla: Talpidae) from Taiwan. *Systematics and Biodiversity*, 5(2), pp.223–240.

Kearse, M. et al., 2012. Geneious Basic: an integrated and extendable desktop software platform for thr organization and analysis of sequence data. *Bioinformatics*, 28(12), pp.1647–1649.

Kirkland, Jr., G.L., 1991. Competition and Coexistence in Shrews (Insectivora: Soricidae). *Special publication The Museum of Southwestern Biology*, 1, pp.15–22.

Knell, R.J. Naish, D., Tomkins, J. L., & Hone, D. W., 2013. Sexual selection in prehistoric animals: detection and implications. *Trends in Ecology & Evolution*, 28(1), pp.38–47.

Lack, D., 1971. *Ecological isolation in birds*,

Lanfear, R. Calcott, B., Ho, S. Y., & Guindon, S., 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, 29, pp.1695–1701.

Larkin, M.A. et al., 2007. Clustal W and Clustal X. *Bioinformatics*, 23, pp.2947–2948.

Lindenfors, P., Gittleman, J. & Jones, K., 2007. Sexual size dimorphism in mammals. In *Sex, size and gender roles: evolutionary studies of sexual size dimorphism*. pp. 19–26.

Lister, A.M., 1996. Dwarfing in island elephants and deer: Processes in relation to time of isolation. *Symposium of the Zoological Society of London*, 69, pp.277–292.

Lister, A.M., 2004. The impact of Quaternary Ice Ages on mammalian evolution. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 359(1442), pp.221–41.

Losos, J.B. & Ricklefs, R.E., 2009. Adaptation and diversification on islands. *Nature*, 457(7231), pp.830–6.

Losos, J.B. & Schluter, D., 2000. Analysis of an evolutionary species-area relationship. *Nature*, 408(6814), pp.847–850.

MacPhee, R.D.E., Flemming, C. & Lunde, D., 1999. “Last Occurence” of the Antillean Insectivoran Nesophontes: New Radiometric Dates and Their Interpretation. *American Museum Novitates*, (3261), pp.1–20.

Maddison, W.P. & Knowles, L.L., 2006. Inferring Phylogeny Despite Incomplete Lineage Sorting. *Systematic Biology*, 55(1), pp.21–30.

Malmquist, M.G., 1985. Character Displacement and Biogeography of the Pygmy Shrew in Northern Europe. *Ecology*, 66(2), pp.372–377.

Maurrasse, F., Pierre-Louis, R. & Rigaud, J., 1980. Cenozoic facies distribution in the southern peninsula of Haiti and the Barahona Peninsula, Dominican Republic, and it’s relations concerning tectonic evolution of the La Selle-Baoruco block. *Caribbean Geology, Collected Contributions*, (9), pp.1–24.

Mayr, E., 2000. The biological species concept. In *Species concepts and phylogenetic theory: a debate*. New York: Columbia University Press, pp. 17–29.

McFarlane, D. a, 1999. A note on dimorphism in *Nesophontes edithae* (Mammalia: Insectivora), an extinct island-shrew from Puerto Rico. *Caribbean Journal of Science*, 35(1–2), pp.142–143.

Meyer, M. & Kircher, M., 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*, 5(6).

Miller, G.S., 1930. Three small collections of mammals from Hispaniola. *Smithsonian Miscellaneous Collections*, 82(15), pp.1–10.

Miller, M.A., Pfeiffer, W. & Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *Gateway Computing Enviroments Workshop (GCE)*, pp.1–8.

Millien, V., 2006. Morphological evolution is accelerated among island mammals. *PLoS Biology*, 4(10), pp.1863–1868.

Morgan, G.S., 1994. Late Quaternary fossil vertebrate from the Cayman Islands. In *The Cayman Islands: Natural History and Biogeography*. pp. 465–508.

Nadachowski, A., 1993. The species concept and quaternary mammals. *Quaternary International*, 19(C), pp.9–11.

Nattier, R. et al., 2012. Secondary Sympatry Caused by Range Expansion Informs on the Dynamics of Microendemism in a Biodiversity Hotspot. *PLoS ONE*, 7(11), p.e48047.

Olson, L.E. et al., 2009. Phylogeography of *Microgale Brevicaudata* (Tenrecidae) and Description of a New Species From Western Madagascar. *Journal of Mammalogy*, 90(5), pp.1095–1110.

Olson, L.E., Goodman, S.M. & Yoder, A.D., 2004. Illumination of cryptic species boundaries in long-tailed shrew tenrecs (Mammalia: Tenrecidae; *Microgale*), with new insights into geographic variation and distributional constraints. *Biological Journal of the Linnean Society*, 83(1), pp.1–22.

Orihuela, J., 2014. Endocranial morphology of the extinct Antillean shrew *Nesophontes* (Lipotyphla : Nesophontidae) from natural and digital endocasts of Cuban taxa. *Palaeontologia Electronica*, 17(2;22A), p.12.

Pinto, G. Mahler, D. L., Harmon, L. J., & Losos, J. B., 2008. Testing the island effect in adaptive

radiation: rates and patterns of morphological diversification in Caribbean and mainland *Anolis* lizards. *Proceedings. Biological sciences / The Royal Society*, 275(1652), pp.2749–2757.

De Queiroz, K. De, 2007. Species concepts and species delimitation. *Systematic Botany*, 56(6), pp.879–886.

Rambaut, A. et al., 2014. Tracer v1.6. *University of Edinburgh, Edinburgh, UK*.

Rambaut, A. & Drummond, A.J., TreeAnnotator v1.6.1. *University of Edinburgh, Edinburgh, UK*.

dos Reis, M. et al., 2012. Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. *Proceedings of the Royal Society B: Biological Sciences*, 279(1742), pp.3491–3500.

Ricklefs, R. & Bermingham, E., 2008. The West Indies as a laboratory of biogeography and evolution. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 363(1502), pp.2393–2413.

Roca, A.L. et al., 2004. Mesozoic origin for West Indian insectivores. *Nature*, 429(6992), pp.649–651.

Rofes, J. & Cuenca-bescós, G., 2012. The earliest shrew (Mammalia Soricidae) from the Balearic Islands : A new species of *Nesiotites*. *Palaeontologia Electronica*, (March 2017).

Ronquist, F. & Huelsenbeck, J.P., 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19, pp.1572–1574.

Rowe, K.C. Aplin, K. P., Baverstock, P. R., & Moritz, C., 2011. Recent and rapid speciation with limited morphological disparity in the genus *Rattus*. *Systematic Biology*, 60(2), pp.188–203.

Sato, J.J. et al., 2016. Molecular phylogenetic analysis of nuclear genes suggests a Cenozoic over-water dispersal origin for the Cuban solenodon. *Scientific Reports*, 6(April), p.31173.

Schluter, D., 2000. *The Ecology of Adaptive radiation*, OUP Oxford.

Schmidt, N.M. & Jensen, P.M., 2003. Changes in Mammalian Body Length over 175 Years—Adaptations to a Fragmented Landscape? *Conservation Ecology*, 7(2), p.6.

Schubert, C. & Medina, E., 1982. Evidence of Quaternary Glaciation in the Dominican Republic: Some implications for Caribbean Paleoclimateology. *Centro De Ecologia*, 39, pp.281–294.

Shepherd, L.D. & Lambert, D.M., 2008. Ancient DNA and conservation: Lessons from the endangered kiwi of New Zealand. *Molecular Ecology*, 17(9), pp.2174–2184.

Silva Taboada, G., Suarez Duque, W. & Diaz-Franco, S., 2007. Compendio de los mamiferos terrestres autoctonos de Cuba vivientes y extinguidos, La Habana: Museo Nacional de Historia Natural.

Sly, N.D. et al., 2011. Ancient islands and modern invasions: Disparate phylogeographic histories among Hispaniola's endemic birds. *Molecular Ecology*, 20(23), pp.5012–5024.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30, pp.1312–1313.

Stamatakis, A., Hoover, P. & Rougemont, J., 2008. A rapid bootstrap algorithm for the RAxML Web servers. *Systematic biology*, 57(5), pp.758–71.

Stewart, J.R., 1999. *The evolution of Quarternary birds in the western Palaeartic:aspects of*

taxonomy and ecomorphology. University College London.

Tamura K Dudley, J., Nei, M., & Kumar, S., 2007. Molecular Rvolutionary Genetic Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, 24, pp.1597–1599.

Turvey, S.T. et al., 2007. First human-caused extinction of a cetacean species? *Biology letters*, 3(5), pp.537–40.

Turvey, S.T. et al., 2016. Independent evolutionary histories in allopatric populations of a threatened Caribbean land mammal. *Diversity and Distributions*, 22(5), pp.589–602.

Turvey, S.T. Oliver, J. R., Storde, Y. N., & Rye, P., 2007. Late Holocene extinction of Puerto Rican native land mammals. *Biology letters*, 3(2), pp.193–196.

Van Valen, L., 1965. Morphological Variation and Width of Ecological Niche. *The Amercian Naturalist*, 99(908), pp.377–390.

VanValen, L., 1973. Pattern and the balance of nature. *Evolutionary Theory*, pp.31–49.

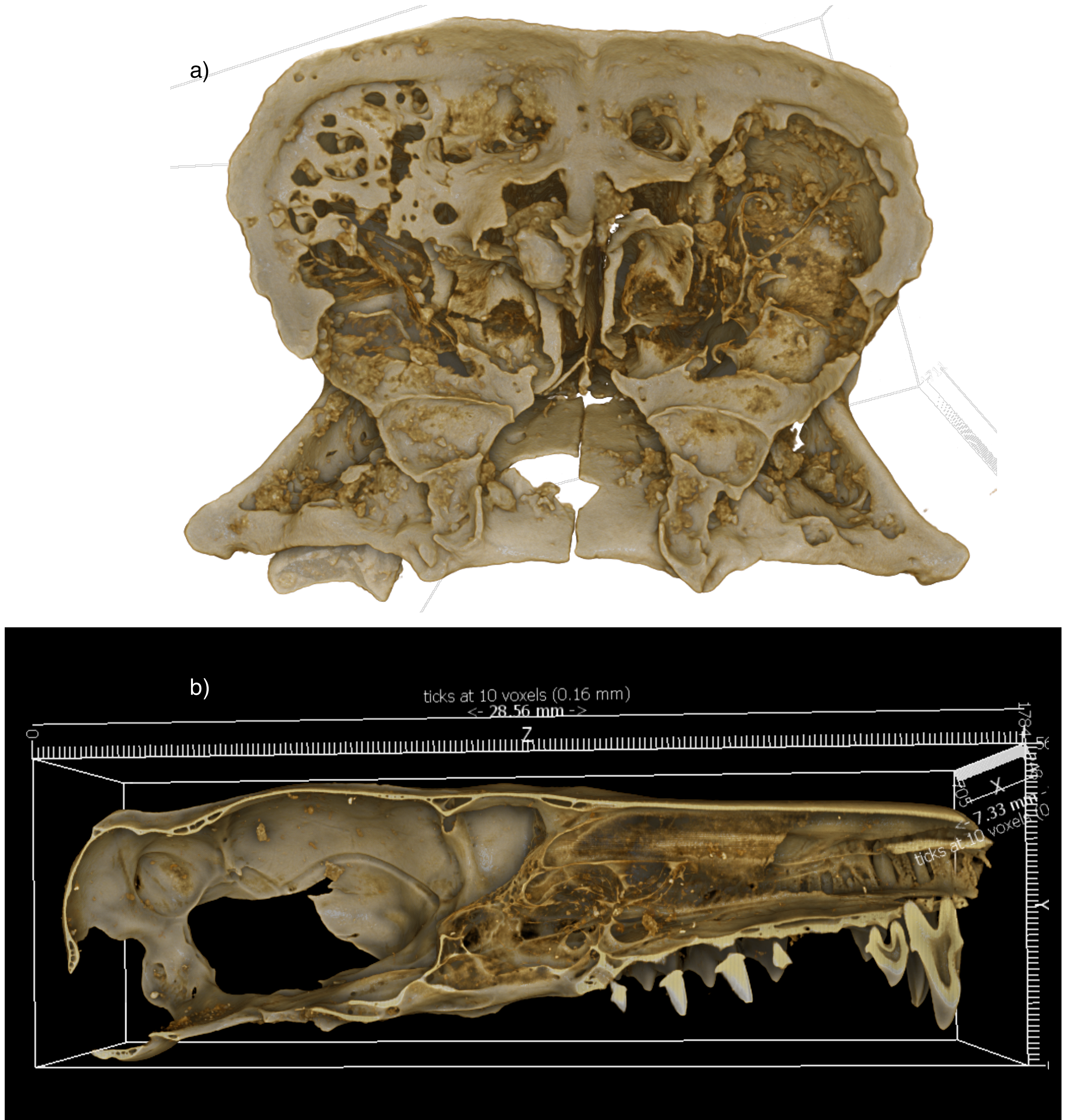
Wallace, A.R., 1869. *The Malay Archipelago: the land of the orang-utan and the bird of paradise; a narrative of travel, with studies of man and nature*, London, UK: Macmillan.

Wiens, J.J. and Penkrot, T.A., 2002. Delimiting species using DNA and morphological variation and discordant species limits in spiny lizards (Sceloporus). *Systematic biology*, *51*(1), pp.69-91.

Willows-Munro, S. & Matthee, C.A., 2011. Exploring the diversity and molecular evolution of shrews (family Soricidae) using mtDNA cytochrome b data. *African Zoology*, 46(2), pp.246–462.

Appendix Chapter 3. Palaeogenomic analyses reveal recent intra-island evolutionary radiation in the extinct Caribbean "island shrew" (Nesophontes)

Figure 1a. Nesophontes micro-CT scans: light areas indicate highly dense material such as bone or compacted earth, dark areas indicate lower density material such as possible soft tissue. a) a slice through the cranium and b) a lateral slice through the entire skull, with scale.



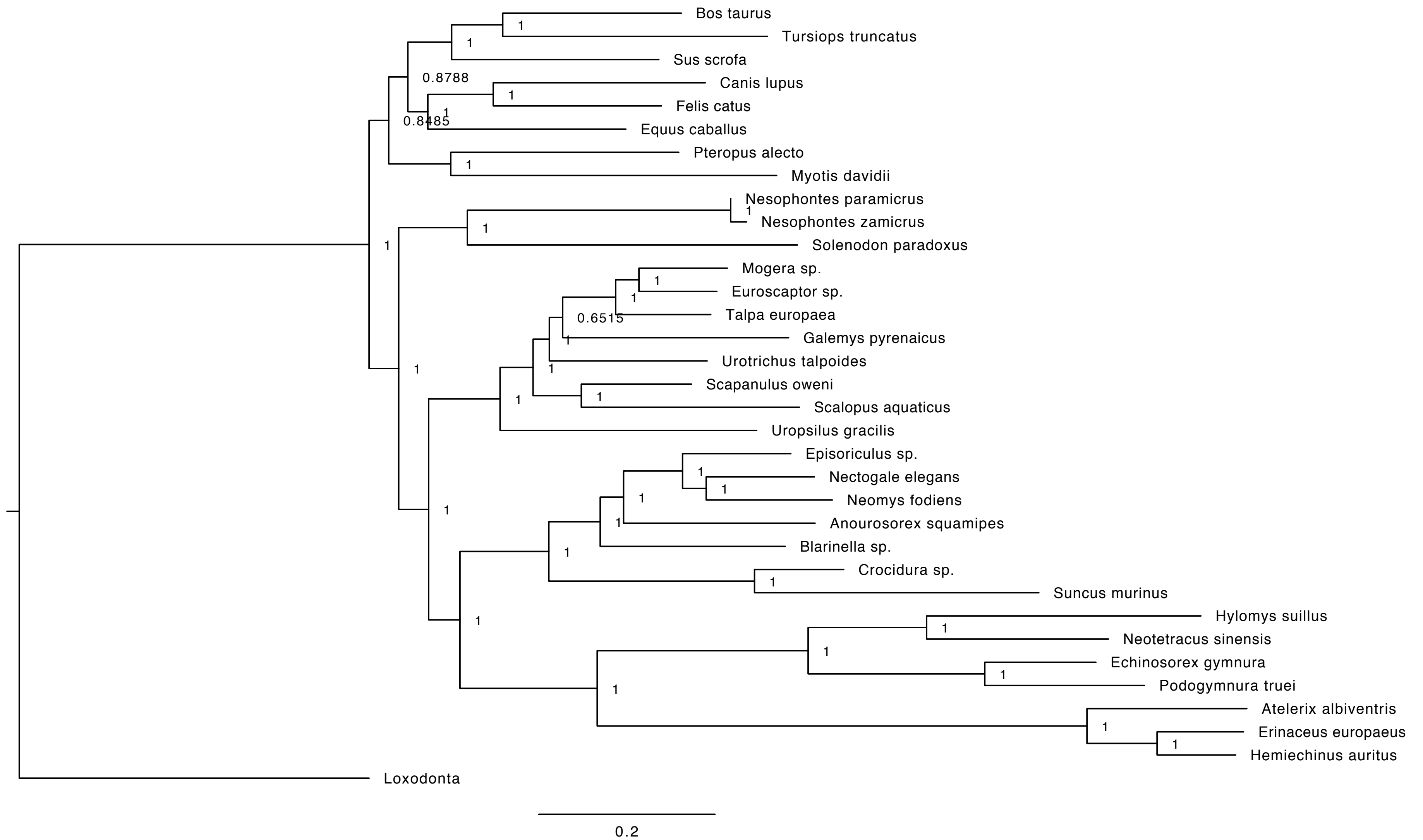


Figure 2a. Bayesian inference phylogeny inclusive of all out-group taxa. Node values indicate posterior probability. Scale indicates nucleotide substitutions per site

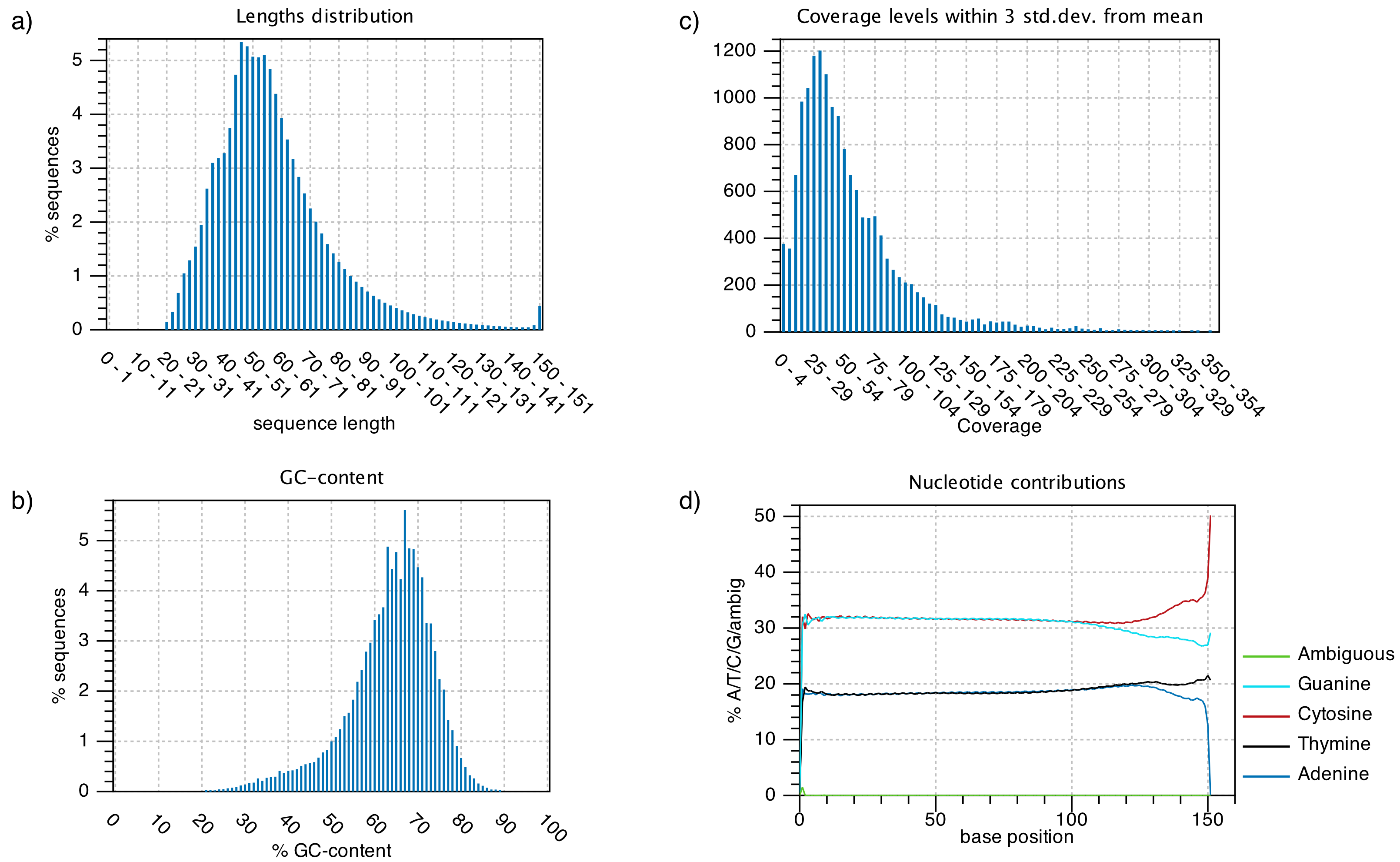


Figure 3a. Graphs output from CLC genomics workbench (v.8) ***a)*** *showing distribution of read lengths in paired reads dataset,* ***b)*** *showing GC content in paired reads dataset,* ***c)*** *showing contribution of different nucleotide bases (Guanine, Cytosine, Thymine and Adenine) for paired reads dataset,* ***d)*** *showing quality distribution of paired reads dataset using average PHRED score as measurement of quality and* ***e)*** *level of coverage across the reference mitochondrial genome in mapped reads dataset.*

Table 1a. Evolutionary models chosen for each gene in the alignment using the program Partitionfinder (Lanfear et al. 2012)

Gene partition	Model
mitochondrial genome	
rrnS, trnC, trnF, trnL2, trnM, trnN, trnQ, trnR, trnS2, trnY	GTR + G
rrnL, trnD, trnH, trnI, trnK, trnL1, trnT, trnV, trnW	GTR + G
atp6, cob, nad1, nad3, nad4, trnA	GTR + G + I
atp8, nad2, nad4l, nad5, trnP	GTR + G
cox1, cox2, cox3, trnG	GTR + G + I
nuclear genes	
ADORA3, ATP7A, GHR, VWF, BMI1, EDG1	GTR + G
CREM, APP, BRCA1, TTN	K80 + G
ADRB2, BDNF, RAG1, RAG2, BCHE	GTR + I

Table 2a. Species used in pairwise genetic distance analysis downloaded from Genbank NCBI database (Benson et al. 2005)

Species	Genetic distance ID	Species	Genetic distance ID
Mogera imaizumii	M1	*Neomys anomalus*	Neo1
Mogera wogura	M2	*Neomys teres*	Neo2
Mogera insularis	M3	*Neomys fodiens*	Neo3
Mogera tokudae	M4	*Suncus etruscus*	Su1
Euroscaptor klossi	Eu1	*Suncus murinus*	Su2
Euroscaptor malayana	Eu2	*Suncus montanus*	Su3
Euroscaptor longirotris	Eu3	*Suncus fellowesgordoni*	Su4
Euroscaptor mizura	Eu4	*Suncus varilla*	Su5
Euroscaptor parvidens	Eu5	*Suncus remyi*	Su6
Euroscaptor subanura	Eu6	*Suncus dayi*	Su7
Talpa stankovici	T1	*Suncus stoliczkanus*	Su8
Talpa europaea	T2	*Nesophontes paramicrus*	Nes1
Talpa altaica	T3	*Nesophontes zamicrus*	Nes2
Talpa caucasica	T4	*Solenodon paradoxus*	So1
Talpa davidiana	T5	*Atelerix albiventris*	A1
Uropsilus gracilis	U1	*Atelerix algirus*	A2
Uropsilus nivatus	U2	*Atelerix frontalis*	A3
Uropsilus soricipes	U3	*Erinaceus concolor*	Er1
Uropsilus atronates	U4	*Erinaceus amurensis*	Er2
Uropsilus investigator	U5	*Erinaceus europaeus*	Er3
Uropsilus aequodonenia	U6	*Erinaceus roumanicus*	Er4
Episoriculus caudatus	Ep1	*Anoutosorex squamipes*	An1
Episoriculus leucops	Ep2	*Anoutosorex yamashinai*	An2
Episoriculus macrurus	Ep3	*Blarinella griselda*	B1
Episoriculus fumidus	Ep4	*Blarinella quadraticauda*	B2
Episoriculus fumidus	Ep4	*Blarinella wardi*	B3

Table 3a. Genetic pairwise distance matrix values for N. paramicrus and N. zamicrus for four nuclear and two mitochondrial genes

Gene	Genetic distance (%)
CREM	2.2
BDNF	0.6
RAG1	0.1
BRCA1	0.4
12s	5.3
cytB	2.2

Table 4.1a. Pairwise distances for lipotyphlan insectivore sister species pairs in the mitochondrial gene 12s (mitochondrially encoded 12S ribosomal RNA) The number of base differences per site from between sequences are shown. The analysis involved 29 nucleotide sequences. All ambiguous positions were removed for each sequence pair. There were a total of 1009 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Tamura K et al. 2007).

	M1	M2	M3	
M1	-			
M2	0.025	-		
M3	0.032	0.034	-	
M4	0.027	0.034	0.029	
	Eu1	Eu2	Eu3	
Eu1	-			
Eu2	0.022	-		
Eu3	0.027	0.022	-	
Eu4	0.059	0.053	0.055	
	T1			
T1	-			
T2	0.095			
	U1	U2	U3	U4
U1	-			
U2	0.000	-		
U3	0.033	0.033	-	
U4	0.039	0.039	0.043	-
U5	0.032	0.032	0.037	0.042
	Ep1	Ep2	Ep4	
Ep1	-			
Ep2	0.039	-		
Ep3	0.063	0.057	-	
Ep4	0.080	0.072	0.071	
	Neo1	Neo2		
Neo1	-			
Neo2	0.035	-		
Neo3	0.053	0.045		
	Su1			
Su1	-			
Su2	0.071			
	Nes1			
Nes1	-			
Nes2	0.053			
	A1			
A1	-			
A2	0.023			

Table 4.2a. Pairwise distances for lipotyphlan insectivore sister species pairs the protein coding nuclear gene CREM (cAMP responsive element modulator) The number of base differences per site from between sequences are shown. The analysis involved 29 nucleotide sequences. All ambiguous positions were removed for each sequence pair. There were a total of 1009 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Tamura et al. 2007).

	T2		
T2	-		
T3	0.000		
	M2	M1	
M2	-		
M1	0.003	-	
M3	0.000	0.003	
	Eu2	Eu3	Eu5
Eu2	-		
Eu3	0.000	-	
Eu5	0.003	0.003	-
Eu4	0.005	0.005	0.003
	Nes1		
Nes1	-		
Nes2	0.022		

Table 4.3a. Pairwise distances for lipotyphlan insectivore sister species pairs the protein coding nuclear gene BRCA1 (Breast Cancer 1 gene). The number of base differences per site from between sequences are shown. The analysis involved 29 nucleotide sequences. All ambiguous positions were removed for each sequence pair. There were a total of 1009 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Tamura K et al. 2007).

	T2			
T2	-			
T4	0.004			
	M2			
M2	-			
M3	0.013			
	A3	A1		
A3	-			
A1	0.002	-		
A2	0.008	0.008		
	Er3	Er2		
Er3	-			
Er2	0.005	-		
Er4	0.005	0.005		
	Su5	Su6	Su2	Su7
Su5	-			
Su6	0.038	-		
Su2	0.072	0.088	-	
Su7	0.074	0.079	0.031	-
Su1	0.075	0.082	0.056	0.058
	Nes2			
Nes2	-			
Nes1	0.004			

Table 4.4a. Pairwise distances for lipotyphlan insectivore sister species pairs the protein coding nuclear gene RAG1 (recombination activating gene 1) The number of base differences per site from between sequences are shown. The analysis involved 29 nucleotide sequences. All ambiguous positions were removed for each sequence pair. There were a total of 1009 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Tamura K et al. 2007).

	T4	T3	T2
T4	-		
T3	0.018	-	
T2	0.022	0.021	-
T5	0.022	0.022	0.024
	M2	M1	M4
M2	-		
M1	0.003	-	
M4	0.008	0.005	-
M3	0.016	0.015	0.016
	Eu4	Eu2	
Eu4	-		
Eu2	0.012	-	
Eu6	0.019	0.019	
	U1		
U1	-		
U4	0.007		
	Su2	Su3	Su4
Su2	-		
Su3	0.005	-	
Su4	0.035	0.036	-
Su1	0.032	0.032	0.016
	Nes2		
Nes2	-		
Nes1	0.001		
	Er3	Er2	
Er3	-		
Er2	0.003	-	
Er4	0.004	0.001	
	A2	A1	
A2	-		
A1	0.002	-	
A3	0.006	0.007	

Table 4.5a. Pairwise distances for lipotyphlan insectivore sister species pairs the protein coding nuclear gene BDNF (brain derived neurotrophic factor) The number of base differences per site from between sequences are shown. The analysis involved 29 nucleotide sequences. All ambiguous positions were removed for each sequence pair. There were a total of 1009 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Tamura K et al. 2007).

	Su3	Su4	Su1
Su3	-		
Su4	0.000	-	
Su1	0.000	0.000	-
Su2	0.000	0.000	0.000
	T2	T3	
T2	-		
T3	0.006		
	Eu2	Eu3	Eu4
Eu2	-		
Eu3	0.006	-	
Eu5	0.010	0.008	-
Eu4	0.015	0.014	0.010
	M2	M1	
M2	-		
M1	0.000	-	
M3	0.006	0.006	
	Nes2		
Nes2	-		
Nes1	0.006		
	Er1		
Er1	-		
Er2	0.002		

Table 4.6a. Pairwise distances for lipotyphlan insectivore sister species pairs the protein coding nuclear gene CREM (cAMP responsive element modulator) **)** *The number of base differences per site from between sequences are shown. The analysis involved 29 nucleotide sequences. All ambiguous positions were removed for each sequence pair. There were a total of 1009 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Tamura K et al. 2007).*

	Su2	Su5	Su7	Su1
Su2	-			
Su5	0.059	-		
Su7	0.069	0.040	-	
Su1	0.080	0.076	0.087	
	T3			
T3	-			
T2	0.010			
	N2			
N2	-			
N1	0.000			

Table 4.7a. Pairwise distances for lipotyphlan insectivore sister species pairs in the mitochondrial gene cytB (cytochome B). The number of base differences per site from between sequences are shown. The analysis involved 29 nucleotide sequences. All ambiguous positions were removed for each sequence pair. There were a total of 1009 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Tamura et al. 2007).

	An1			
An1	-			
An2	0.061			
	B1	B2		
B1	-			
B2	0.048	-		
B3	0.117	0.119		
	Ep1	Ep2		
Ep1	-			
Ep2	0.133	-		
Ep3	0.156	0.156		
	Neo1			
Neo1	-			
Neo3	0.146			
	Su2			
Su2	-			
Su8	0.086			
	Eu1	Eu2		
Eu1	-			
Eu2	0.084	-		
Eu4	0.134	0.125		
	M1	M2	M4	
M1	-			
M2	0.073	-		
M4	0.098	0.096	-	
M3	0.096	0.111	0.111	
	T3			
T3	-			
T2	0.137			
	Nes1			
Nes1	-			
Nes2	0.022			
	U1	U2	U6	U4
U1	-			
U2	0.025	-		
U6	0.102	0.103	-	
U4	0.138	0.132	0.142	-
U3	0.126	0.124	0.121	0.096
	A1			
A1	-			
A3	0.125			
	Er3	Er4	Er2	
Er3	-			
Er4	0.003	-		
Er2	0.109	0.110	-	
Er1	0.128	0.129	0.131	

Table 5a. Genes used in phylogenetic analysis including GenBank NCBI database accession numbers for downloaded sequence data. indicates chimeric taxon made up of two species.*

Family	Subfamily	Genus	Species	Mitogenome	APP	BMI1	CREM	PLCB4	ADORA3	APOB	ADRA2B	ADRB2
Nesophontidae		*Nesophontes*	Sp		X		X			X	X	X
Solenodontidae		*Solenodon*	*Paradoxus*	X	AY530068	AY530071	AY451973, AY530072	AY530074	AY530066	JN414026	AY530081	AY530067
Erinaceidae	Erinaceinae	*Erinaceus*	*Europaeus*	NC.002080	XM.007519783 AY011319	XM.0075263 64	XM.007523995	XM.007539576	XM.007526934	JN414024	XM.007529903	XM.007516378
		Hemiechinus	*Auritus*	NC.005033								
		Atelerix	*Albiventris*	X					X	X	X	X
	Galericinae	*Echinosorex*	*Gymnura*	NC.002808		AY986735	AY451978	AY451986				
		Hylomys	*Suillus*	NC.010298				AY451987		DQ630204	AJ505819	
		Neotetracus	*Sinensis*	NC.019626								
		Podogymnura	*Truei*	AF434823, AF434829, AF298578-9	JN633548	AY986742	AY451980	JN633106	JN633426	JN414025	JN413881	JN633682
Soricidae	Crocidurinae	*Crocidura*	*russula/*fuliginosa /**gueldenstaedtii*	NC.006893		AY986736**	AY451977**	AY451985**		GU981117*	AY121766	
		Suncus	*murinus/*remyi*	AB175074 AB032842 EF507245 AB032845 AB033720 F444469 FJ814528 J813943 FJ813963 AB011390		AY986741	AY451981	AY451992		DQ630181		
	Soricinae	*Episoriculus*	*Fumidus*	NC.003040						DQ630193		
		Sorex	*araneus/*unguiculatus*	NC.005435*	XM.004621155	AY011509	AY011631	AY011754	AY011200	JN414027	XM.004609236	XM.004620597

Table 5a. (continued) Genes used in phylogenetic analysis including GenBank NCBI database accession numbers for downloaded sequence data. indicates chimeric taxon made up of two species.*

Family	Subfamily	Genus	Species	Mitogenome	APP	BMI1	CREM	PLCB4	ADORA3	APOB	ADRA2B	ADRB2
		Nectogale	*Elegans*	NC.023351						GU981129		
		Blarinella	Sp	NC.023950						DQ630187		
		Anourosorex	*Squamipes*	NC.024563								
		Neomys	*Fodiens*	NC.025559						DQ630177		
Talpidae	Talpinae	*Mogera*	*Wogura*	NC.005035	HG738058	HG738000	HG738020	HG738041	HG737959		AB638542	
		Talpa	*europaea /*altaica*	NC.002391	HG738065*	HG738009	AY011629*	AY011752*	HG737968	JN414028*	JN413891*	AY011259*
		Euroscaptor	*mizura/*parvidens*	AB106233 DQ630336, AB076828 DQ630413	HG738051	HG737991	HG738012	HG738032	HG737949	DQ630168		
		Scalopus	*Aquaticus*	AF069539 AY170059 AB076809							AF496636	
		Scapanulus	*Oweni*	NC.025777								
	(Scalopinae?)	*Condylura*	*Cristata*	X	XM.004675316	AY011508	AY011630	AY011753	AY011199	XM.004686059	XM.004686125	XM.004686868
		Galemys	*Pyrenaicus*	NC.008156	AY833412		AY833411	AY833416	AY833418		AY121767	AY833413
		Urotrichus	*Talpoides*	AB099483								
	Uropsilinae	*Uropsilus*	*gracilis*	NC.018598		KF778096	KF778144	KF778255	KF777913	DQ630199	AY121768	
Outgroups:												
Manidae		*Manis*	*pentadactyla*	NC.016008	AY011374	EF104992	AY011684	AY011807	AY011251	JN413999	JN413856	AY011311
Felidae		*Felis*	*Catus*	NC.001700	AY011369	AY011553	AY011679	AY011802	AY011246	XM.003984468	XM.006930331	AY011306
Canidae		*Canis*	*Lupus*	KF857179	GU167668	NM.001287063	GU167419	AY011805	NM.001003178	XM.005630596	XM.005642442	NM.001003234
Delphinidae		*Tursiops*	*Truncates*	NC.012059	AY011360	JQ002729	AY011670	AY011793	AY011237	JN413972	XM.004311473	AY011297

Table 5a. (continued) Genes used in phylogenetic analysis including GenBank NCBI database accession numbers for downloaded sequence data. indicates chimeric taxon made up of two species.*

Family	Subfamily	Genus	Species	Mitogenome	APP	BMI1	CREM	PLCB4	ADORA3	APOB	ADRA2B	ADRB2
Bovidae		*Bos*	*Taurus*	GU947019	NM.001076796		XM.005214245	NM.001166510	NM.001104611	XM.002691511	NM.001206628	NM.174231
Suidae		*Sus*	*Scrofa*	NC.012095	AY011364	NM.001285971	AY011674	AY011797	AY011241	XM.001501679	NM.001037148	AY011301
Equidae		*Equus*	*Caballus*	NC.001640	AY011366	XM.005606891	AY011676	AY011799	AY011243	JN414029	NM.001164012	AY011303
Vespertilionidae		*Myotis*	*Davidii*	KF111724	XM.006758757	XM.006759928	XM.006774616	XM.006758082		XM.006767541	XM.006761143	XM.006766218
Pteropodidae		*Pteropus*	*alecto/*giganteus*	NC.023122	XM.006908243	XM.006907001	XM.006911376	XM.006921597	XM.006919638	XM.006910423	XM.006909438	XM.006906373
Elephantidae		*Loxodonta*	*Africana*	NC.000934	AY011326	AY011513	AY011635	AY011758	AY011204	JN413956	JN413853	AY011264

Table 6a. Fossil constraints and priors used in divergence dating-analysis

Root	Node	Fossil and Deposit Locality	Age (Ma)	Prior distribution
	Placentalia	Minimum: Earliest carnivoramorphan: *Protictis* from the Fort Union/Polecat Bench Formation Soft maximum: Liaoning fossils of *Eomaia* and *Sinodelphys* (Theria)	Basal Torrejonian (To1) - Barremian 62.5 – 131.5	logNormalPrior mean="22.55" stdev="1.0" offset="62.5"
1	Laurasiatheria	*As above*	*As above*	logNormalPrior mean="16.02" stdev="1.0" offset="62.5"
2	Laurasiatheria minus Lipotyphla	*As above*	*As above*	*As above*
3	Ferungulata	*As above*	*As above*	*As above*
4	Lipotyphla	Minimum: Oldest erinaceomorph: *Adunator,* from Torrejonian. Soft maximum: *As above*	Early Danian – Barremian 61.5 – 131.5	logNormalPrior mean="16.25" stdev="1.0" offset="61.5"
5	Carnivora	Minimum: *Hesperocyon*, from earliest Duchesnean Soft maximum: Oldest stem carnivorans in Torrejonian	Earliest Duchesnean – early Palaeocene, 39.68 – 65.8	logNormalPrior mean="6.067" stdev="1.0" offset="39.68"
6	Cetartiodactyla	Minimum: *Himalayacetus* from the base of the Subathu Formation in Pakistan. Soft maximum: absence of crown Cetartiodactyla in Palaeocene	Early Eocene – early Palaeocene, 52.4 – 65.8	logNormalPrior mean="3.111" stdev="1.0" offset="52.4"
7	Chiroptera	Minimum: *Icaronycteris,* Soft maximum: as for other members of Laurasiatheria	48.6 – 65.8	logNormalPrior mean="3.995" stdev="1.0" offset="48.6"

Chapter 4. Ancient DNA of the extinct Jamaican monkey *Xenothrix* reveals extreme insular change within a morphologically conservative primate radiation

4.1 Abstract

The Caribbean region has often been used as a study system to investigate insular evolutionary dynamics. However, previous studies have rarely focused on mammals, as much of the archipelago's land mammal fauna has been lost in the world's largest post-glacial mammalian extinction event. Among the Caribbean mammal taxa that became extinct during the Holocene were four genera of endemic primates, formerly found on Jamaica, Hispaniola and Cuba. Morphological studies of the limited sub-fossil material of Caribbean primate taxa have tried to reconstruct their phylogenetic affinities with mainland Neotropical platyrrhines, and to determine whether they represent geologically ancient or recent Caribbean colonists and whether they constitute a monophyletic group or colonised the Caribbean multiple times. The extinct Caribbean primates, notably the Jamaican monkey *Xenothrix mcgregori*, are morphologically extremely unusual, and previous studies that have tried to determine their evolutionary history using morphological data from available sub-fossil specimens have generated multiple conflicting hypotheses. This study reports the first ancient DNA sequence data for an extinct Caribbean primate, and demonstrates that instead of being an ancient "relict" platyrrhine lineage outside the extant Neotropical radiation, *Xenothrix* falls within the species-rich but otherwise morphologically conservative titi monkey radiation (Callicebinae) and is sister to the newly recognized genus *Cheracebus*. *Xenothrix* and *Cheracebus* diverged c. 11 Ma (95% HPD = 5.2-14.9 Ma). As endemic primate species are known to have been present on other Caribbean islands by 16.5 Ma, our results therefore suggest that late Quaternary Caribbean primate diversity was generated by multiple over-water colonization events.

4.2 Introduction

Island evolution and extinctions: the insular Caribbean mammal fauna

Islands are the home of spectacular evolutionary radiations, preserve ancient relict species, and as 'natural laboratories' have generated theories that form the backbone of our evolutionary thinking (Darwin 1859; Wallace 1869; MacArthur & Wilson 1967; Lack 1947; Losos et al. 1998; Schluter 2000). However, oceanic-type (non-continental) islands are rarely colonised by terrestrial mammals, which limits investigation of this branch of evolutionary theory in one of the best-studied animal groups.

The insular Caribbean is unusual in this context, as it held an extensive late Quaternary mammal fauna that included insectivores, rodents, sloths and primates. Unfortunately many of the species that make up this remarkable group of colonists have since become extinct, as part of the largest regional-scale mammalian extinction of recent times (MacPhee & Flemming 1999; Macphee 2009; Turvey 2009). This event left a remnant, species-poor and highly threatened extant fauna, and investigations of the evolutionary history of the late Quaternary mammal fauna of the Caribbean have been based mainly on morphological study of incomplete and degraded sub-fossil material.

Four genera and five species were part of the Caribbean's extinct endemic primate assemblage, but the relationships and origins of this morphologically unusual group remain unresolved. Caribbean insular primates were all formally distributed in on the larger islands in the western Caribbean that make up the Greater Antilles. The assemblage includes: *Antillothrix bernensis* and *Insulacebus toussaintiana* excavated from Hispaniola (the single island that makes up Haiti and the Domican Republic), two species from the same genus, *Paralouatta varonai* and *P. marianae*, which were found on Cuba, and finally *Xenothrix mcgregori* formally found on Jamaica, is the focus of this study, and was the first fossil primate to be described from the Caribbean (Williams & Koopman 1952).

Radiocarbon dating of sub-fossil remains has been utilised to access whether the Greater Antillean primate fauna survived in to the mid-Holocene, postdating the colonization of the islands by humans. In *Antillothrix bernensis*, a species found on Hispaniola, an associated ^{14}C date gives an age of 3850 ± 150 calibrated radiocarbon years before present (cal. BP) for the type specimen (Rímoli 1977; MacPhee et al. 1995; Rosenberger et al. 2011). A study that looked at submerged remains of *Antillothrix* gave an associated uranium-series age of 1.32 ± 0.11 Ma (Rosenberger et al. 2015). For *Insulacebus toussaintiana*, also found on Hispaniola, two Holocene ^{14}C dates of 3850 ± 135 and 9,550 ± 150 cal. BP suggest this species may postdate the mid-Holocene arrival of humans in the Caribbean (Cooke et al. 2011b) The Cuban *Paralouatta varonai* was excavated with associated late Quaternary fauna in 1991 (Rivero & Arrendondo 1991), however remains of the

second species from this genus *P. marianae* are considerably older, dated to approximately 16.5 million years, and are the oldest Caribbean primate remains, providing an earliest constrained age for the presence of endemic primates in the Greater Antilles (MacPhee 1984). *Xenothrix* is thought to have survived until much more recently. A recent study dated remains from the late Holocene (direct AMS date of 1,477 ± 34 cal BP) and estimated a last occurrence date of ~ 900 BP (Cooke et al. 2017).

Like other Caribbean island mammal taxa, such as the large lipotyphlan insectivore *Solenodon* (Brandt 1833) and extinct giant rodent *Amblyrhiza inundata* (MacPhee & Flemming 2003), the Greater Antillean primates exhibit extremely unusual morphological features that suggest they have diverged from their mainland relatives to fill novel ecological niches that often exist in insular environments. These unusual morphological features are the most pronounced in *Xenothrix*. Study of the holotype, a lower jaw described by Williams and Koopman, initially suggested that *Xenothrix* shared morphological features with multiple living Neotropical platyrrhine taxa. Williams and Koopman suggested that a jaw studied was comparable with *Cebus* (capuchins), however they found that the dental formula was the same as callitrichines (marmosets) and also suggested affinities with callicebines (titi monkeys) and with *Callimico* (Goeldi's marmoset) (Williams & Koopman 1952). The collection of additional specimens in the 1990s allowed more thorough morphological investigation (MacPhee & Horovitz 2004; MacPhee & Meldrum 2006). These studies confirmed that *Xenothrix* was particularly unusual in terms of the occlusal surface of its molars and the shape of its jaws, which differed radically from any living platyrrhine. *Xenothrix* has a callitrichine-like dental formula where the third molar has been lost. This condition has been hypothesized to be the result of reduced body size in callitrichines (Ford 1980). However, the size of the jaw and especially the teeth in *Xenothrix* is unlike callitrichines and more comparable to *Alouatta* (howler monkeys) or *Brachyteles* (woolly spider monkeys) in the family Atelidae, despite both of these taxa retaining a third molar. The angle of the jaw and gradient of small front molars to larger back molars in *Xenothrix* is similar to the condition in the callicebines, currently recognized as part of the family Pitheciidae (Williams & Koopman 1952; MacPhee & Horovitz 2004). The postcranial morphology of *Xenothrix* is if anything even more unusual, as its short, stout femur suggests that it was an arboreal quadruped, a locomotory adaptation not seen in any recent platyrrhine (MacPhee & Meldrum 2006).

The exceptional morphology of the *Xenothrix* is not just at odds with that of mainland platyrrhines: the different extinct Greater Antillean species are also radically different to each other. *Paralouatta*, in contrast to *Xenothrix*, is morphologically similar to a single platyrrhine group, *Alouatta*. *Paralouatta* also has a appendicular skeleton associated with semiterrestriality, which is found in no other platyrrhine species (MacPhee & Meldrum 2006). Studies have suggested affinities between robust post cranial elements in *Antillothrix* and cebids, whilst dental features ally this species with *Saimiri* (squirrel monkeys) (Rosenberger et al. 2011). A recent study looking at *Antillothrix* cranial

remains suggested primitive morphology: a mosaic of morphological features shared with multiple extant platyrrhine groups including *Lagothrix* (woolly monkeys), *Alouatta* and *Brachyteles* (Halenar et al. 2017). *Insulacebus* shares several dental features with *Xenothrix* and both have been allied with callicebines and *Aotus* (Cooke et al. 2011b).

The evolutionary history of the extinct Greater Antillean primates

There is little consensus about the origin and relationships of the Caribbean primates, both to each other and to mainland Neotropical primates, and phylogenetic hypotheses have so far been based solely on morphological data. However, the unusual morphology of the Caribbean primates has made it difficult to reconstruct the phylogenetic affinities of each species based on morphological characters from fragmented sub-fossil remains alone. This difficulty has led to a large number of different hypotheses for the origin, colonisation and diversification of Caribbean primates, and in particular for *Xenothrix* (Figure 1). Williams and Koopman (1952) only went so far as to classify *Xenothrix* as a non-callitrichid platyrrhine. Hershkovitz (Hershkovitz n.d.) suggested that *Xenothrix* was not closely related to any living South American primate and placed it in its own family, Xenotricidae. Rosenberger (Rosenberger 2002; Rosenberger 1977) made two phylogenetic proposals: first, that *Xenothrix* is closely related to callicebines; second, after examining new specimens, that the enlarged orbit and comparably broad upper central incisors indicate that its closest living relatives are the nocturnal *Aotus* (owl monkeys) within the family Aotidae and the extinct Atelid *Tremacebus* (Kay et al. 2004). MacPhee and Horovitz evaluated the latter suggestion in their subsequent description of the new *Xenothrix* material (MacPhee & Horovitz 2004), where they looked at the relative placement and size of the zygomatic and maxillary components of the inferior orbital fissure, but concluded that *Xenothrix* was sister to *Callicebus*. In their analysis of the new *Xenothrix* material, MacPhee and Horovitz also concluded that morphological and molecular data (Schneider et al. 1993) excluded *Aotus* from the Pitheciidae, whereas there was a clear apparent relationship between *Xenothrix* and pitheciids, particularly *Callicebus*. All callicebines were at the time considered to be referable to the single genus *Callicebus*, which represented the most species-rich primate genus (Thomas 1903). However, a recent molecular review of the Callicebinae has recognised three separate clades within *Callicebus sensu lato* which diverged during the Miocene, and which have now been recognised as distinct genera (*Callicebus*, *Cheracebus*, *Plecturocebus*) (Byrne et al. 2016).

Recent studies using applied morphological and phylogenetic techniques have proposed further evolutionary hypotheses for *Xenothrix* and other Caribbean primates. Geometric morphometrical analysis of fossil and extant platyrrhines has suggested that *Xenothrix* could represent an ancient lineage that diverged before the radiation of all crown platyrrhines (Delgado et al. 2016). Another recent study used a molecular scaffold incorporating data from extant species in combination with morphological data to investigate extant and extinct platyrrhine relationships, and also suggested

that the extinct Caribbean taxa are stem platyrrhines (Kay 2015). However, there were limited support values for fossil taxa in this analysis, including the placement of *Xenothrix*, and differences between molecular and morphological analysis were also highlighted by this study, for example in terms of placement of *Callicebus*, which falls within Cebidae along with *Aotus* based on morphological data, and within Pitheciidae based on molecular data.

Colonization, biogeographic origin and phylogenetic relationships of the Greater Antillean primates

The biogeography of the Caribbean is a well-studied topic with a long history. Previous studies have attempted to analyze current and past species diversity seen in the insular Caribbean in the context of the route and timing of colonization events, with land bridges (Iturralde-Vinent & MacPhee 1999) and over-water dispersal (Hedges et al. 1992) as competing possible hypotheses to explain patterns of colonization.

Several alternative hypotheses have also been proposed by previous studies to explain both the relationships between different extinct Greater Antillean primates, and whether these taxa ultimately represent a monophyletic group or alternatively a series of distantly related taxa, which each have closer phylogenetic affinities to completely different mainland platyrrhine lineages. The first hypothesis is that all extinct Greater Antillean primates form a monophyletic group, which has been variously proposed as the sister group either to *Callicebus* (MacPhee & Horovitz 2004) or to all crown platyrrhines (Kay 2015). MacPhee and Horovitz concluded that *Xenothrix* and all other Caribbean primate taxa were monophyletic, and could be assigned to the endemic Caribbean tribe Xenotrichini. MacPhee and Horovitz also suggested a close relationship between *Xenothrix* and the Cuban *Paralouatta* due to the shared characteristic of enlarged nasal fossa. The description of *I. toussaintiana* also suggested a relationship between this Hispaniolan species and *Xenothrix,* citing tooth morphology unique to both taxa (Cooke et al. 2011a). The second hypothesis suggests that the diversity of extinct Caribbean primates has instead arisen from multiple mainland lineages, proposing either multiple colonisation events at different times, or a single multi-lineage colonisation event (Rosenberger, 2002), evidence for this hypothesis comes in the form of the variation in morphological features exhibited by the different Greater Antillean taxa.

Recovery of molecular data from extinct Caribbean fauna

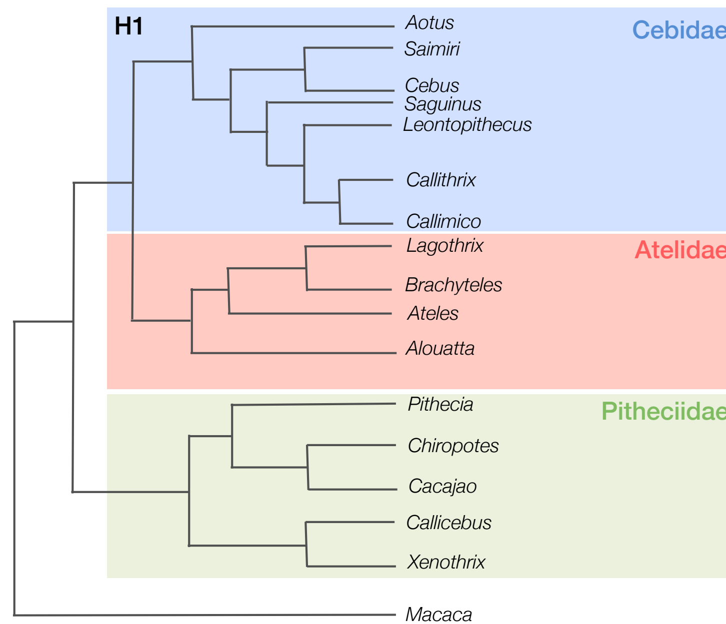
The study of ancient DNA (aDNA) – the extraction and sequencing of molecular data from poorly preserved specimens – has helped to shed light on many extinct faunal assemblages (Hofreiter et al. 2001), including some Caribbean land mammal groups (Brace et al. 2012; Fabre, Vilstrup, Raghavan, Sarkissian, Willerslev, Douzery & Orlando 2014; Brace et al. 2015; Brace et al. 2016; Kehlmaier et al. 2016). Ancient DNA and protein studies have the potential to reveal previously

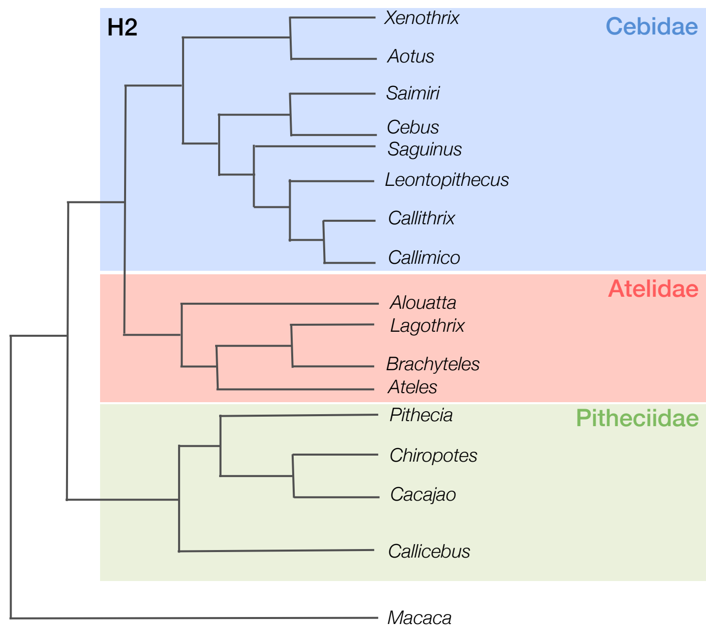
unsuspected phylogenetic relationships for extinct enigmatic taxa, such cases include: New Zealand ratite moa (Bunce et al. 2009), *Toxodon* and *Macrauchenia* (Welker et al. 2015) and Caribbean lipotyphlan insectivore *Nesophontes* (Brace et al. 2016) The Greater Antillean endemic primates have not yet been investigated using ancient DNA techniques. Molecular data could provide important new information to identify mainland sister taxa for the endemic Greater Antillean primates, and when the Greater Antillean primates diverged from related mainland lineages. Molecular study of extinct Caribbean species is limited by preservation of DNA, which is greatly inhibited by the high thermal age represented by humid tropical conditions (Reed et al. 2003) such as those found in the Caribbean. DNA that has been extracted from Greater Antillean sub-fossil specimens in previous studies is highly fragmented, with the chance of retrieving useable molecular markers for phylogenetic analysis greatly diminished, however the use of Next Generation Sequencing techniques combined with target capture enrichment have greatly increased the yield of endogenous DNA in recent ancient DNA studies (Horn 2012).

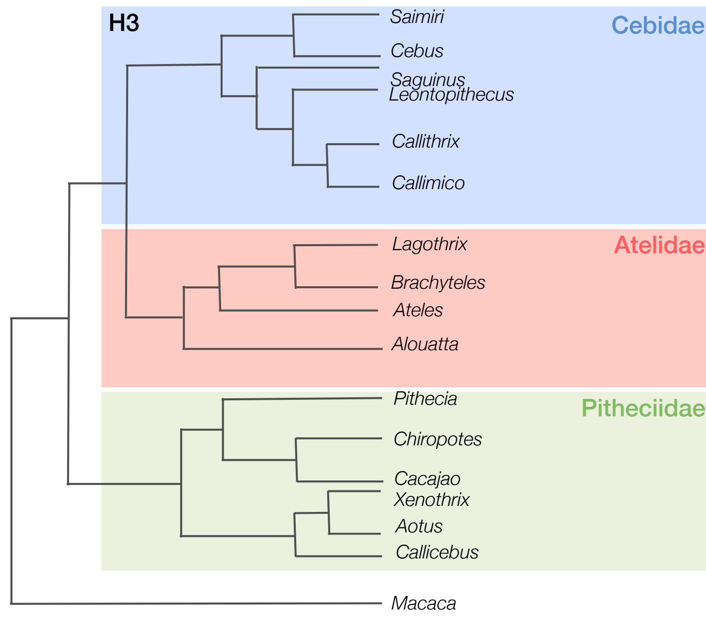
4.3 Study overview and aims

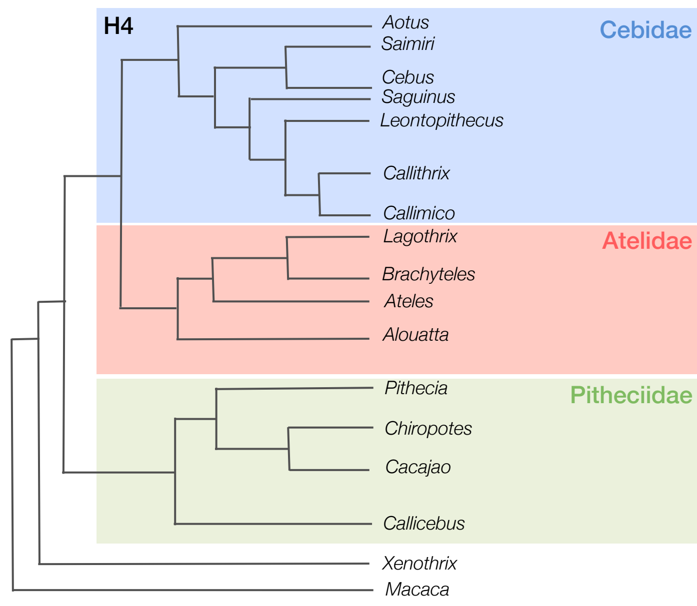
In this study we employ aDNA techniques and phylogenetic methods to examine the evolutionary relationships among the extinct Caribbean primates and their possible mainland Neotropical platyrrhine relatives, in order to test between the wide range of morphology-based phylogenetic hypotheses for the origination of these enigmatic primates that have been proposed by previous authors. Our primary objectives are to establish the phylogenetic association between *Xenothrix mcgregori* and mainland taxa, in order to reconstruct the phylogenetic history and dynamics of morphological evolution in this unusual extinct primate. This study also aims to date this divergence in relation to the known Caribbean primate fossil record, in order to determine whether all Caribbean primates belong to the same clade or whether there were multiple primate colonisations of the insular Caribbean.

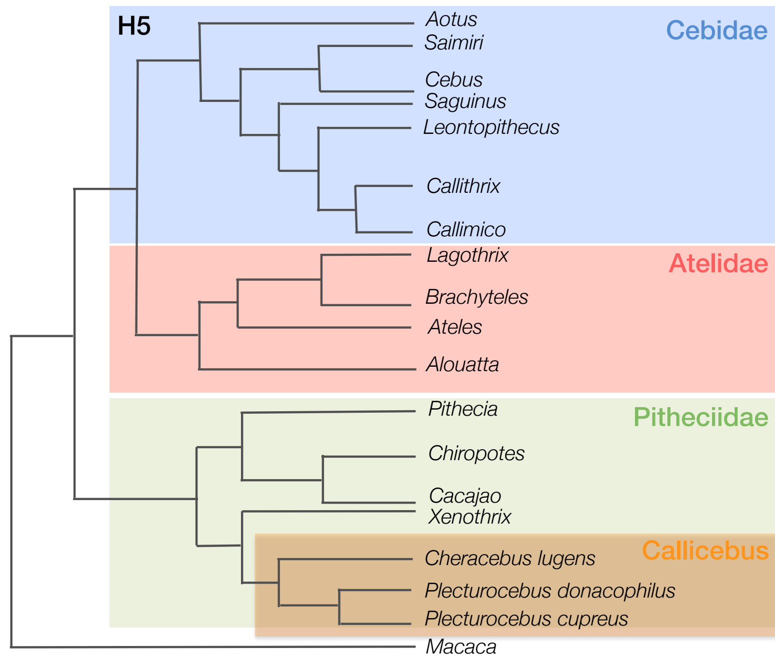
Figure 1. Five alternative tree topologies derived from previous proposals about the evolutionary origins of Xenothrix, including related South American extant genera. ***H1*** *(MacPhee 2004): Genus-level tree showing Xenothrix within Pitheciidae as sister to Callicebus,* ***H2*** *(Rosenberger 2002): Genus-level tree with Xenothrix placed within Cebidae as sister to Aotus,* ***H3*** *(Rosenberger 2002): Genus-level tree showing Xenothrix as sister to Aotus but with both genera placed in Pitheciidae,* ***H4*** *(Kay 2015): Genus-level tree where Xenothrix falls outside of all platyrrhine families (Pitheciidae, Cebidae and Atelidae),* ***H5*** *(MacPhee 2004, Byrne et al 2016): Species-level tree where Xenothrix is basal to all newly recognized callicebid genera.*




H2
Cebidae
Xenothrix
Aotus
Saimiri
Cebus
Saguinus
Leontopithecus
Callithrix
Callimico
Atelidae
Alouatta
Lagothrix
Brachyteles
Ateles
Pitheciidae
Pithecia
Chiropotes
Cacajao
Callicebus
Macaca


H3
Cebidae
Saimiri
Cebus
Saguinus
Leontopithecus
Callithrix
Callimico
Atelidae
Lagothrix
Brachyteles
Ateles
Alouatta
Pitheciidae
Pithecia
Chiropotes
Cacajao
Xenothrix
Aotus
Callicebus
Macaca


H4
Aotus
Saimiri
Cebus
Saguinus
Leontopithecus
Callithrix
Callimico
Cebidae
Lagothrix
Brachyteles
Ateles
Alouatta
Atelidae
Pithecia
Chiropotes
Cacajao
Callicebus
Pitheciidae
Xenothrix
Macaca

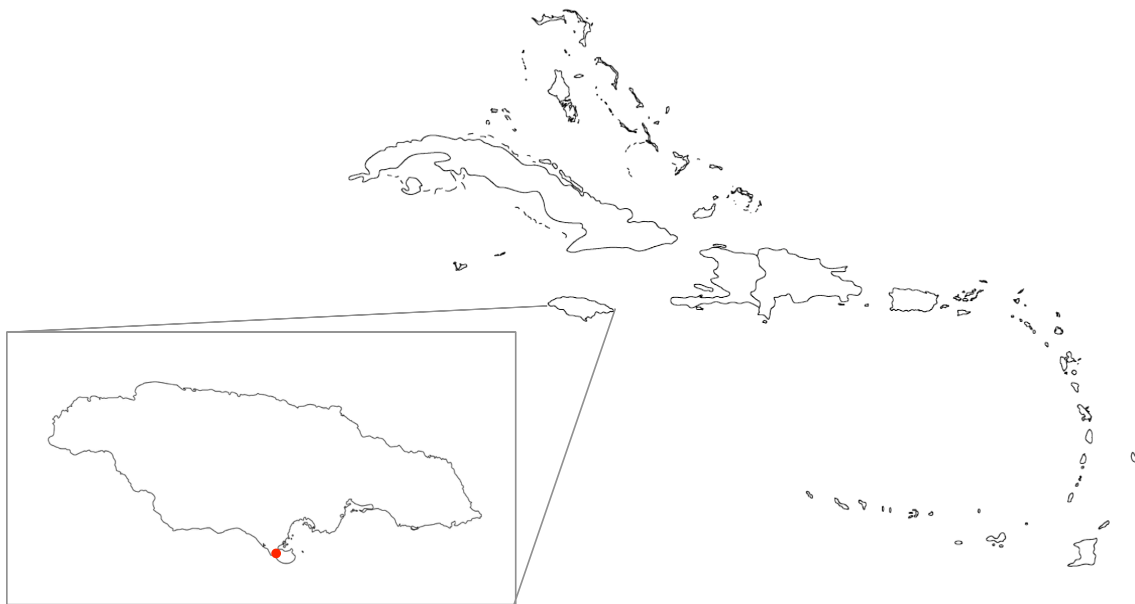

H5
Aotus
Saimiri
Cebus
Saguinus
Leontopithecus
Callithrix
Callimico
Cebidae
Lagothrix
Brachyteles
Ateles
Alouatta
Atelidae
Pithecia
Chiropotes
Cacajao
Xenothrix
Pitheciidae
Cheracebus lugens
Plecturocebus donacophilus
Plecturocebus cupreus
Callicebus
Macaca

4.4 Materials and Methods

Data Collection

Two subfossil primate specimens identified as *Xenothrix* (MacPhee & Fleagle 1991) were subjected to sampling for aDNA extraction. The first specimen, a femur (AMNH 268003), has previously given a direct AMS date of 1,477 ± 34 cal BP. The second specimen, a proximal ulna (AMNH 268010), has not been dated directly but is suspected to be similar in age. Both specimens were found in Somerville Cave, Clavendon, Jackson's Bay, Jamaica (Appendix: Figure 6a).

Figure 2. Map showing Greater and Lesser Antillean islands including Jamaica. Somerville Cave, where specimens analysed as part of this study were excavated, is indicated in red.



Extractions and Next Generation Sequencing library builds took place in a dedicated aDNA laboratory at the Natural History Museum, London, located in a separate building from any post-PCR analysis. A Dremel hand-held drill with a 2-3 mm drill bit was used to sample bone powder. The surface of the bone was first cleaned using the drill to minimize contamination. Drill bits were changed between specimens and all equipment used was sterilized with bleach and UV treated before and after use. A negative control was included in the set of samples during the entire process.

Single-index double-stranded DNA libraries were built following protocol from Meyer & Kircher, 2010 (Meyer & Kircher 2010). Libraries were screened for endogenous DNA using the Illumina MiSeq 500 Next Generation Sequencing platform. In-solution, hybridisation-capture enrichment was

carried out prior to sequencing in order to increase depth of sequencing of targeted areas (Enk et al. 2014). Capture enrichment kits (MYcroarray, Ann Arbor) were applied, using baits designed from the whole mitochondrial genome and five nuclear genes: epsilon-globin, oxytocin receptor (promotor region), vWF (von Willebrand factor), IRBP (interphotoreceptor retinoid binding protein), GHR (growth hormone receptor), TAS2R38 (Camopa bitter taste receptor) and ABCA1 (ATP-binding cassette sub-family A). These genes were selected due to their use in pervious studies including phyloegentic analysis of South American primates and their availablity on the NCBI database Genbank (Benson et al. 2005) (Appendix: Table 1a). As *Xenothrix* is an extinct taxon with no previous molecular data available, these reference sequences were interpreted as the best estimate of the closest living relative to *Xenothrix*, and were chosen on the basis of previous suggestions that *Xenothrix* may be most closely related to callicebines (MacPhee & Horovitz 2004). Specific species were chosen on the basis of whole mitochondrial genome and nuclear gene data availability (Appendix: Table 1a).

Sequence analysis

Raw data were analysed in CLC Workbench software v.8 (CLC Bio-Qiagen, Aarhus, Denmark), where reads were paired, merged and then trimmed of adapters using default settings. To reduce the potential for ascertainment bias during sequence assembly, reads were mapped to a range of 20 reference sequences for the whole mitochondrial genome and for each nuclear gene targeted. The set of reference sequences included available Neotropical platyrrhines and three outgroup taxa: *Homo sapiens, Macaca fuscata* and *Pan troglodytes* (Appendix: Table 2a). Mapping parameters were as follows: Length fraction: 0.8, Similarity fraction: 0.8. More reads mapped to callicebine reference sequences than to any other reference sequences, with the highest amount of reads mapping to *Cheracebus lugens* (Appendix: Figure 1a (parts 1 and 2)).

The *Xenothrix* sequence data were then aligned to callicebine species and with *Saimiri sciureus*, *Cebus albifrons*, *Pithecia pithecia*, *Chiropotes israelita* and *Cacajao calvus* as outgroup taxa, using the ClustalW (Larkin et al. 2007) alignment tool as implemented in Geneious v. 8.0.5 (Kearse et al. 2012) Alignments of each gene were concatenated using Seaview v.4 (Gouy et al. 2010). Phylogenetic relationships were estimated using maximum likelihood and Bayesian methods, with DNA substitution models chosen for the partitioned dataset using PartitionFinder (Lanfear et al. 2012) (Appendix: Table 3a). A maximum likelihood tree (Figure 3) with bootstrap support values was generated using RAxML v.8 (Stamatakis 2014) as implemented in CIPRES Science Gateway v.3.3 (Miller et al. 2010). Bayesian trees were constructed using MrBayes (Ronquist & Huelsenbeck 2003) with four chains (three heated, one cold) that were run for 1×10^6 generations, sampling every 1×10^3 generations with a burn-in period of 250 trees (Appendix: Figure 5a). Tests of alternative topologies suggested by previous studies (Figure 1) were made by submitting the sitewise log-likelihood values obtained from RAxML v.8 (Stamatakis 2014) to CONSEL (Shimodaira

& Hasegawa 2001), to calculate the P-values for each tree topology by using the AU test (Appendix: Table 4a).

Phylogeny and diversification times were simultaneously assessed under an uncorrelated relaxed lognormal molecular clock in BEAST v. 1.8.3 (Drummond & Rambaut 2007). Best-fit evolutionary models were chosen in Partitionfinder as in the previous phylogenetic analyses. A Yule model of speciation was used; the birth-death model was run for comparison and generated identical topology. Prior distributions on two nodes were set using two fossil calibration points: Cebinae (12.5 mya) and Pitheciidae (15.7 mya). (Appendix: Table 5a). To provide an ingroup calibration point a further prior distribution was set for the divergence between Callicebinae and Pitheciinae following the estimate given in Byrne et al (32) (95% HPD (highest posterior density) = 15.79-22.6 Ma), using tmrca for soft upper and lower bounds. All other priors were left as default values in BEAUti v. 1.8.3 (Drummond et al. 2012). The analysis was run for 25 million generations, sampling every 1000 generations. Tracer v. 1.6.0 (Rambaut et al. 2014) was used to access convergence and effective sample size for all parameters after a burn in of 10%. Finally, a maximum credibility tree was generated in TreeAnnotator v. 1.8.3 (58), using trees sampled in the prior distribution.

4.5 Results

Screening results suggested very poor survival of endogenous DNA, with estimated endogenous contents of 0.0064 and 0.0091 from samples taken from the *Xenothrix* specimens AMNH 268010 and AMNH 268003 respectively (Table 1). This high level of DNA degredation is expected from material preserved in a tropical locality with high humidity and temperature (Smith et al. 2001). The sample for which we recovered high levels of endogenous DNA from initial screening (AMNH 268010) was then used for further sequencing inclusive of target capture enrichment. This technique greatly increased endogenous DNA recovery, with an enrichment factor of over 180% (Table 1). We could therefore recover an incomplete mitochondrial genome, along with seven regions of the nuclear genome. Post-sequencing data processing generated a total of 8,567 total reads mapped to the mitochondrial reference genome used and an average depth of coverage of 13.81 (Table 1). The incomplete mitochondrial genome generated was used in preliminary genus-level analysis, and two mitochondrial genes and one nuclear gene were used in the final species-level analysis due to missing data in some extant species. In our tests of alternative tree topologies, AU P-valves were below 0.5 for all species level and almost all genus topologies suggested by previous morphological studies. Support was shown for genus level topology from MacPhee 2004 and species level topologies recovered by our analysis (Appendix: Table 4a). We recovered convergent maximum likelihood and Bayesian phylogenies (Figure 3, Figure 4 and Appendix: Figure 2a).

Table 1. Sequencing results including pre-target capture enrichment screening results and post-capture results. Endogenous content and enrichment factor were estimated using mapping results from the whole mitochonrial genome reference seqeunce used.

<table>
<tr><th rowspan="2">Species</th><th rowspan="2">Sample size used in the final analysis</th><th rowspan="2">Number of samples sequenced</th><th rowspan="2">Sample ID</th><th colspan="6">Pre-capture enrichment</th><th rowspan="2">Enrichment factor (mtgenome)</th><th rowspan="2">Percentage increase of total reads mapped</th></tr>
<tr><th>Total number of reads</th><th>Total number after trim</th><th>Reads mapped to reference sequence</th><th>Average depth of coverage</th><th>Reads mapped to mtgenome</th><th>Estimated endogenous content (mtgenome)</th></tr>
<tr><td rowspan="4">Xenothrix mcgreogori</td><td rowspan="4">1</td><td rowspan="4">2</td><td>RW55</td><td>175,560</td><td>175,557</td><td>16 (mtgenome), 7 (JX856254), 8 (KT182969)</td><td>0.02 (mtgenome), 0.10 (JX856254), 0.12 (KT182969)</td><td>16</td><td>0.009113849</td><td>na</td><td>na</td></tr>
<tr><td rowspan="3">RW56</td><td>6,833,646</td><td>6,822,002</td><td>438 (mtgenome), 253 (JX856254), 600 (KT182969)</td><td>0.56 (mtgenome), 3.66 (JX856254), 9.75 (KT182969)</td><td>438</td><td>0.006409463</td><td rowspan="3">184.15%</td><td rowspan="3">95%</td></tr>
<tr><td colspan="6">Post-capture enrichment</td></tr>
<tr><td>47,038,996</td><td>36,632,332</td><td>8,567 (mtgenome), 5,708 (JX856254), 3,236 (KT182969)</td><td>13.81 (mtgenome), 84.02 (JX856254), 48.51 (KT182969)</td><td>8,567</td><td>0.018212549</td></tr>
</table>

Due to the degraded nature of ancient DNA a number of associated errors may have affected our analysis. Degradation, the result of both time and the humidity and high temperatures associated with the tropics, can result in modifications that allow the incorporation of incorrect nucleotide bases in the sequence data. For example hydrolytic reactions can cause the deamination, depurination and depyrimidination of bases (Hoss et al., 1996). Deamination of cytosine to uracil, 5-methyl-cytosine to thymine and adenine to hypoxathine cause transitions (C -> T and G -> A) to be incorporated in to DNA strands. These aDNA specific changes could artfically inflate the rate of substitution causing erroneous estimates of branch lengths and the divergence dates associated with a phylogeny. The program MapDamage (Ginolhac et al 2010) that can be used to assess DNA damage patterns and is now often used by ancient DNA studies (Marx 2017) to authenticate ancient samples. Unfortunately as this program is most often utilised for whole genome mapping data the total number of reads mapped to the mitochondrial genome and depth of coverage generated from *Xenothrix* subfossil remains during our study were not sufficiently high to estimate DNA damage patterns. In our species level analysis (Appendix: Figure 2a) we recovered long branch lengths for *Xenothrix*, however we did not recover the same long branch lengths in our preliminary genus level analyses (Figure 3 and Figure 4). In order to include a wider range of species in our final analysis we used a more restricted set of molecular markers that were avaliable for the additional taxa included in the phylogeny. This reduction of molecular markers utilised in the second analysis may have increase the influence of errors in the sequencing data that are expected due to ancient DNA damage.

Our dated phylogeny (Figure 4) shows that *Xenothrix* falls within the group of taxa formerly assigned to *Callicebus sensu lato*, and is sister to the callicebine genus *Cheracebus*, with a mean estimated divergence date from *Cheracebus* of c. 11 Ma (95% HPD = 5.2-14.9 Ma).

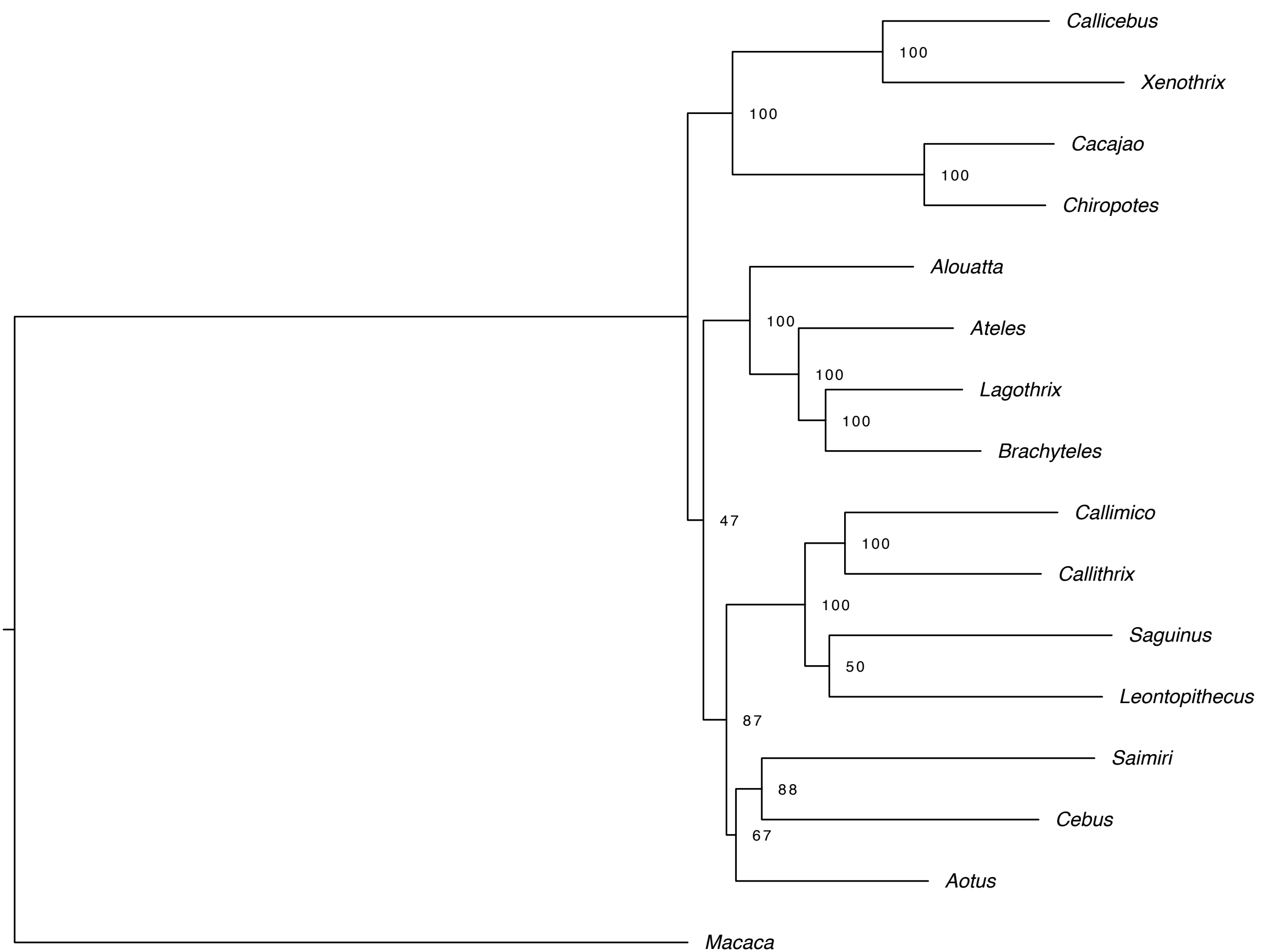


Figure 3. Genus-level maximum likelihood phylogeny generated using whole mitochondrial genomes and produced in RAxML, using data sequenced in this study for Xenothrix and data for 15 other primate genera from Genbank, and with Macaca fuscata selected as outgroup. Node values represent bootstrap support (100 replicates).

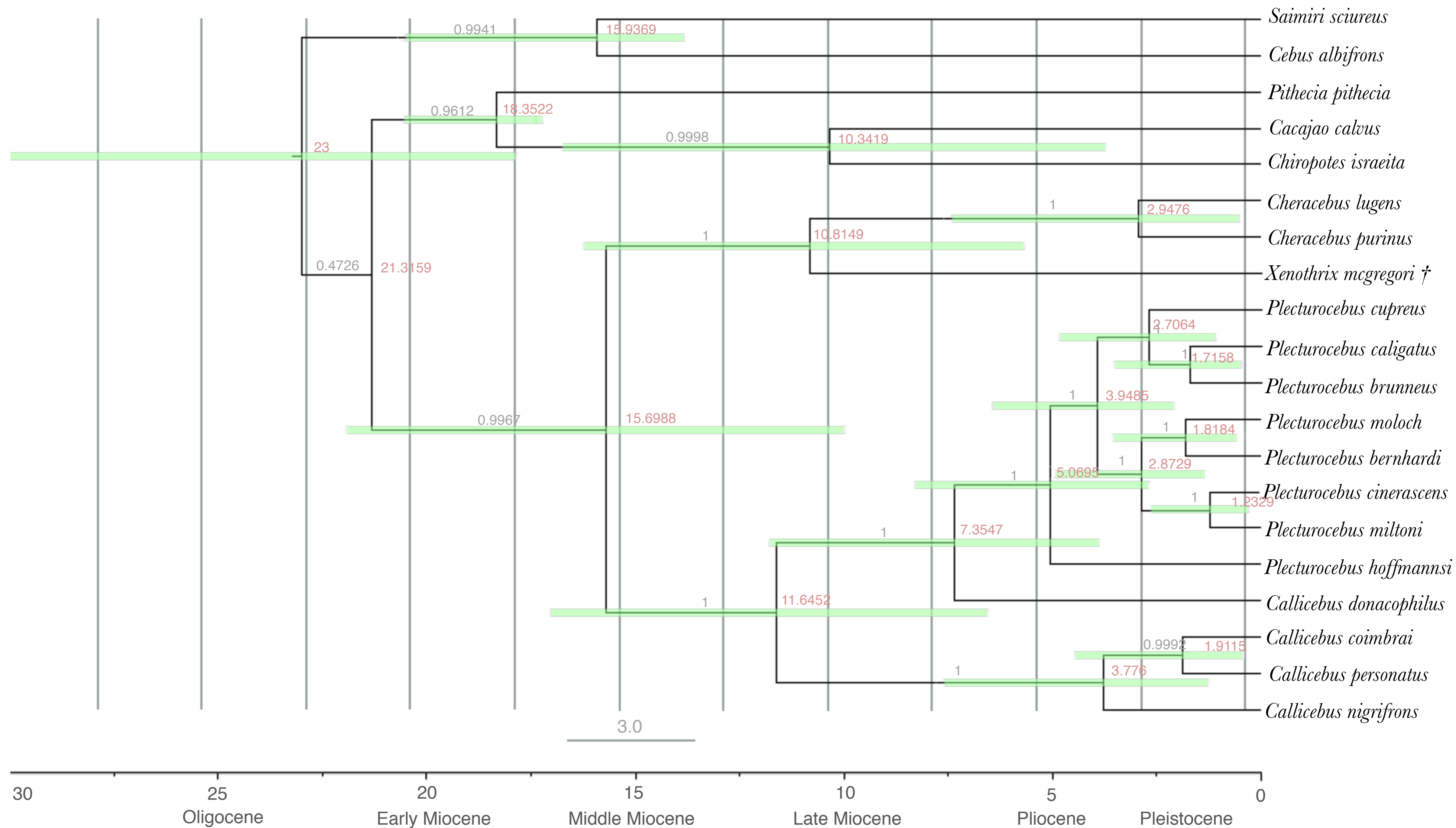


Figure 4. Time-calibrated phylogeny showing estimated divergence dates for Xenothrix, callicebine species, and other platyrrhine genera. Estimates of median divergence dates are shown in red above nodes. Node bars indicate 95% highest posterior density values. Branch values represent posterior probabilities.

4.6 Discussion and Conclusions

We were able to extract and sequence the first aDNA sequence data from an extinct Caribbean primate, despite the adverse preservational conditions that are prevalent in the Caribbean and reduce the likelihood of DNA preservation in sub-fossil samples. The results of our phylogenetic analysis are not congruent with any previous phylogenetic hypotheses proposed on the basis of morphological data. Instead of being a stem-group platyrrhine or having a close phylogenetic relationship to *Aotus*, *Alouatta*, *Brachyteles* or callitrichines, or being sister to the entire callicebine radiation, we find that *Xenothrix* is actually nested within the extant Callicebinae and is sister to the recently described genus *Cheracebus*.

Morphological versus molecular phylogenies for Caribbean primates

Disparities between morphological and molecular estimates of phylogeny are not unusual in the field of New World primate taxonomy. Whilst morphology-based analyses have previously produced inconsistent results on the phylogenetic affinities of both *Aotus* and callicebines, often suggesting a close relationship between the two genera (Rosenberger & Tejedor 2013; Schneider & Rosenberger 1996), molecular studies have consistently placed callicebines within the pitheciines and *Aotus* in the separate callitrichine/cebine clade (Schneider et al. 1993; Opazo et al. 2006). Partition homogeneity analysis has demonstrated that previous phylogenetic analyses for platyrrhines, and specifically including Caribbean primates, may recover different results when using craniodental versus postcranial data (Kay 2015), suggesting that phylogenetic hypotheses based on the restricted morphological character datasets available for extinct species must be interpreted with care. These considerations have implications for the phylogenetic explanatory value of morphology-only data for Caribbean primates, and for previous hypotheses based upon such data for the phylogenetic placement of *Xenothrix*. Most previous proposed hypotheses, including Rosenberger's (2002) suggestion of a close relationship between *Xenothrix* and *Aotus*, and MacPhee and Horovitz's proposal that *Xenothrix* is sister to *Callicebus* (MacPhee & Horovitz 2004), also rely upon taxonomy that is inconsistent with more recent platyrrhine molecular phylogenies. It is therefore not surprising that molecular and morphological analyses of *Xenothrix* draw different phylogenetic conclusions, when previous morphology-based and molecular-based studies produce such disparate hypotheses on the phylogenetic placement of its purported closest relatives.

Implications for primate insular evolution and morphological conservatism

Until recently, all extant callicebines were thought to be referable to the single genus *Callicebus*, to which between 13 and 28 species have been recognised following different species concepts (Wilson & Reeder 2005; Hershkovitz 1990; Groves 1992). Although this group is now interpreted as comprising three genera, the main differences between these taxa are in pelage characteristics and

body size, with craniodental and other skeletal characters exhibiting little variation across the subfamily (Hershkovitz 1990). Extant callicebines are therefore remarkably morphologically conserved in comparison with other platyrrhine lineages (Delgado et al. 2016; Aristide et al. 2015; Rosenberger 2011), so that the peculiar morphology of *Xenothrix* is evolutionarily unexpected.

The dramatic difference in morphology between *Xenothrix* and its closest extant relatives suggests that two contrasting modes of speciation are likely to have driven evolution in *Xenothrix* and mainland Callicebinae. Barriers to gene flow created by river systems (Gascon et al. 2000) and Pleistocene climate refugia (Brown & Ab'Saber 1979) are considered to be among the primary factors responsible for generating the high primate species diversity found in the Amazon Basin today, including the diversity observed in *Callicebus*, *Cheracebus* and *Plecturocebus*. Mainland callicebine populations are separated geographically but still inhabit relatively similar environments and occupy comparable ecological niches, which is associated with little morphological divergence over time. Conversely, the colonization of Jamaica by over-water dispersal of a callicebine lineage seemingly led to ecological release into a novel environment containing vacant niches, which was associated with equivalent divergence in primate morphospace. Caribbean islands apparently lacked medium-sized frugivores before the arrival of primates in this system (Fleagle 1999), and the unique and unexpected morphological traits exhibited by *Xenothrix* are likely to be associated with adaptation to this new ecological niche. Geographical isolation of other species is similarly known to result in peculiar morphologies (Sondaar 1991; Renaud & Millien 2001), drastic size changes (VanValen 1973; Lister 1996; Biknevicius et al. 1993) and even accelerated evolution (Millien 2006), such that a lineage's potential for phenetic plasticity when exposed to novel environmental conditions cannot be predicted on the basis of past morphological conservatism within more homogeneous systems.

Primates, like other non-volant mammals, are rarely found on non-continental island systems. However, evolutionary patterns seen across other insular species can also be seen in primates. Famously, the extinct insular hominin *Homo floresiensis* exhibits morphological divergence from mainland Asian and African hominins consistent with the general "island rule" (whereby larger-bodied lineages decrease in body size, and smaller-bodied lineages increase in body size, following insular colonization and geographical isolation) (Bromham, Lindell & Cardillo 2007). Macaques have also colonised multiple oceanic-type insular environments, with a distribution second only to humans among extant primates, and this group displays a series of characteristic morphological differences between island and mainland populations including divergence in body size, pelage coloration and tail length (Abegg & Thierry 2002; Schillaci et al. 2007; Schillaci et al. 2009). Our study provides further evidence of island evolution causing radical morphological changes over relatively short geological time frames in an insular primate. However, there are relatively few examples of this kind of extreme insular adaptation in extant island primate faunas, perhaps making

it easier to understand how morphological and molecular analysis can arrive at two starkly different phylogenetic conclusions for the evolutionary history of an unusual extinct island primate.

Colonisation and evolutionary history of the Caribbean primate fauna

Our data on the divergence of *Xenothrix* from *Cheracebus* suggests a colonisation of Jamaica by this lineage during the middle-late Miocene around 11 Ma, with an upper 95% HPD of 14.9 Ma. *Cheracebus* is distributed across northern South America, extending into the Orinoco region of Venezuela, suggesting that over-water dispersal of the ancestral *Xenothrix* lineage is most likely to have occurred northwards into the Caribbean from this region. The late Miocene was an interval of low eustatic sea level during which a northward riverine connection between Amazonia and the Caribbean may have existed, suggesting that over-water dispersal of *Xenothrix* may have been associated with natural rafts of floating vegetation from this palaeo-river system (Wilkinson et al. 2010).

Importantly, the oldest Caribbean primate fossils, referred to the Cuban species *Paralouatta marianae*, date to around 16.5 Ma, and therefore pre-date the oldest possible estimates for colonisation of the Caribbean by *Xenothrix* by more than 1.5 million years. Our estimates for the timing of the divergence between *Xenothrix* and *Cheracebus*, which constrain the earliest possible date for the arrival of the ancestral *Xenothrix* lineage in Jamaica, therefore reveal that at least two independent colonization events of the insular Caribbean by primates must have occurred at different times during the Neogene. As a result, the extinct Caribbean primate assemblage cannot have been a monophyletic group, contrary to the earlier morphology-based phylogenetic hypotheses (MacPhee & Horovitz 2004). This discovery matches the evolutionary history of several other Quaternary Caribbean vertebrate groups (e.g., leptodactylid frogs, mabuyid skinks, megalonychid sloths, Lesser Antillean oryzomyine rice rats), which have been shown to comprise multiple distantly related lineages representing separate colonization events (Woods & Sergile 2001; Hedges & Conn 2012; Brace et al. 2015), although other Caribbean mammal groups that exhibited considerable diversity in the Quaternary (e.g. solenodonotan insectivores) instead comprise a single monophyletic clade that experienced a within-Caribbean evolutionary radiation (Brace et al. 2016).

Our findings are also consistent with previous hypotheses about the origins and evolution of other components of the Jamaican vertebrate fauna. The Quaternary-modern fauna of Jamaica is biogeographically distinct from that of the other major Greater Antillean islands, lacking several groups that characterize these other islands (e.g., megalonychid sloths, solenodonotan insectivores), and showing the greatest level of species-level endemism in its avifauna for any Caribbean island (Vázquez-Miranda et al. 2007) Other vertebrate groups that are known from both Jamaica and elsewhere in the insular Caribbean have different colonization histories.

Molecular evidence supports the inclusion of all Jamaican *Anolis* species in a single monophyletic clade generated by a single colonization event, whereas *Anolis* species diversity found across the rest of the Caribbean was generated by two separate colonization events (Hedges & Burnell 1990). Oryzomyine rice rats were formerly present on Jamaica as well as in the Lesser Antilles, but whereas the rice rats of this latter island group comprise two distantly related clades that colonized from northern South America (Brace et al. 2015) the now-extinct Jamaican rice rat (*Oryzomys antillarum*) represents a separate colonization event that probably occurred over-water from Central America, as this species is extremely similar to the mainland Central American species *O. couesi* and has even been interpreted by several authors as being conspecific (Thomas 1898; Allen 1942; Hershkovitz 1966; Musser & D. 2005). The distinct evolutionary history of Jamaica's vertebrate fauna probably reflects a combination of geographic distance from other islands and the major marine barrier represented by the deep Cayman Trough, which may have hindered dispersal from other Greater Antillean islands to Jamaica even during periods of low sea level (Donnelly 1994).

Ancient DNA analysis suggests that far from being an ancient platyrrhine lineage outside the extant Neotropical radiation, the morphologically aberrant extinct Caribbean primate *Xenothrix* in fact falls within the otherwise morphologically conserved callicebine radiation, in which extant species were until recently all assigned to the same genus on the basis of their general similarity: a previously unsuspected phylogenetic relationship for this enigmatic and poorly understood animal. While we cannot yet identify the sister taxa of other extinct Caribbean primate species from Cuba and Hispaniola, the findings of our study further demonstrate that the Caribbean primate assemblage cannot represent a within-Caribbean evolutionary radiation resulting from a single over-water dispersal event. These findings are therefore able to reveal crucial insights into the evolutionary history and phylogenetic affinities of the wider extinct Caribbean primate assemblage even in the absence of aDNA sequence data for other species, and have important implications for reconstructing the evolution of both Neotropical primates and Caribbean mammal faunas across space and time.

References

Abegg, C. & Thierry, B., 2002. Macaque evolution and dispersal in insular south-east Asia. *Biological Journal of the Linnean Society*, 75(4), pp.555–576.

Allen, G.M., 1942. Extinct and Vanishing Mammals of the Western Hemisphere with the Marine Species of all the Oceans. *American Committee for InternationalWild Life Protection, Special Publication*, 11.

Aristide, L. Rosenberger, A. L., Tejedor, M. F., & Perez, S. I., 2015. Modeling lineage and phenotypic diversification in the New World monkey (Platyrrhini, Primates) radiation. *Molecular Phylogenetics and Evolution*, 82(PB), pp.375–385.

Benson, D.A. et al., 2005. GenBank. *Nucleic Acids Research*, 33, pp.34–38.

Biknevicius, A.R., McFarlane, D. a. & Macphee, R.D.E., 1993. Body Size in Amblyrhiza inundata (Rodentia: Caviomorpha), an Extinct Megafaunal Rodent from the Anguilla Bank, West Indies: Estimates and Implications. *American Museum Novitates*, 3079(3079), pp.1–25.

Brace, S. et al., 2016. Evolutionary history of the Nesophontidae, the last unplaced Recent mammal family. *Molecular biology and evolution*, 33(12), p.msw186.

Brace, S. et al., 2012. Population history of the Hispaniolan hutia *Plagiodontia aedium* (Rodentia: Capromyidae): Testing the model of ancient differentiation on a geotectonically complex Caribbean island. *Molecular Ecology*, 21(9), pp.2239–2253.

Brace, S. et al., 2015. Unexpected evolutionary diversity in a recently extinct Caribbean mammal radiation. *Proceedings of the Royal Society B: Biological Sciences*, 282(1807), pp.20142371–20142371.

Brandt, J.F., 1833. De Solenodonte, novo mammalium insectivorum genere. *Memoires de l'Académie Impériale des Sciences de St. Pétersbourg, sér*, 2, pp.459–478.

Brown, K.S.J. & Ab'Saber, A.N., 1979. Ice-age forest refugia and evolution in the neotropics; correlation of paleo-climatological, geomorphological and paleoecological data with modem bio- logical endemism. *Paleoclimas*, 5, pp.1–30.

Bunce, M. et al., 2009. The evolutionary history of the extinct ratite moa and New Zealand Neogene paleogeography. *Proceedings of the National Academy of Sciences*, 106(49), pp.20646–20651.

Byrne, H. et al., 2016. Phylogenetic relationships of the New World titi monkeys (Callicebus): first appraisal of taxonomy based on molecular evidence. *Frontiers in Zoology*, 13(10).

Cooke, S.B. Mychajliw, A. M., Southon, J., & MacPhee, R. D., 2017. Jamaica's last monkey: the extinction of *Xenothrix mcgregori*. *Journal of Mammalogy*. *98*(4), 937-949.

Cooke, S.B., Rosenberger, A.L. & Turvey, S., 2011b. An extinct monkey from Haiti and the origins of the Greater Antillean primates. *Proceedings of the National Academy of Sciences of the United States of America*, 108(7), pp.2699–2704.

Darwin, C., 1859. *On the origin of species by natural selection*, New York: P. F. Collier & Son.

Delgado, M.N., Galbany, J. & Pérez-pérez, A., 2016. Morphometric variation of extant platyrrhine

molars : taxonomic implications for fossil platyrrhines. *PeerJ*, pp.1–26.

Donnelly, T.W., 1994. The Caribbean Sea Floor. In *Caribbean Geology: An Introduction*. Kingston: U.W.I Publishers Association, pp. 41–60.

Drummond, A.J. M. A. Suchard, D. Xie, and A. Rambaut., 2012. Bayesian phylogenetics with BEAUti and the BEAST. *Molecular Biology and Evolution*, 29, pp.1969–1973.

Drummond, A.J. & Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC evolutionary biology*, 7, p.1.

Ginolhac, A., Rasmussen, M., Gilbert, M.T.P., Willerslev, E. and Orlando, L., 2011. mapDamage: testing for damage patterns in ancient DNA sequences. *Bioinformatics*, *27*(15), pp.2153-2155.

Enk, J.M. et al., 2014. Ancient whole genome enrichment using baits built from modern DNA. *Molecular Biology and Evolution*, 31(5), pp.1292–1294.

Fabre, P.-H. et al., 2014. Rodents of the Caribbean: origin and diversification of hutias unravelled by next-generation museomics. *Biology letters*, 10(7), p.20140266-.

Fleagle, J.G., 1999. *Primate adaptation and evolution* 2nd ed., New York: New York: Academic Press.

Ford, S.M., 1980. Callitrichids as phyletic dwarfs, and the place of the callitrichidae in platyrrhini. *Primates*, 21(1), pp.31–43.

Ford, S.M., 1986. Subfossil platyrrhine tibia (primates: Callitrichidae) from Hispaniola: a possible further example of island gigantism. *American journal of physical anthropology*, 70(1), pp.47–62.

Gascon, C. et al., 2000. Riverine barriers and the geographic distribution of Amazonian species. *Proceedings of the National Academy of Sciences*, 97(25), pp.13672–13677.

Gouy, M., Guindon, S. & Gascuel, O., 2010. Seaview version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution*, 27, pp.221–224.

Groves, C.P., 1992. Titis, New Yorld Monkeys of the Genus *Callicebus* (Cebidae, Platyrrhini): A Preliminary Taxonomic Review. *International Journal of Primatology*, 13(1), pp.111–112.

Halenar, L.B., Cooke, S. B., Rosenberger, A. L., & Rímoli, R., 2017. New cranium of the endemic Caribbean platyrrhine, *Antillothrix bernensis*, from La Altagracia Province, Dominican Republic. *Journal of Human Evolution*, 106, pp.133–153.

Hedges, S.B. & Burnell, K.L., 1990. The Jamaican radiation of *Anolis* (Sauria: Iguanidae): an analysis of relationships and biogeography using sequential electrophoresis. *Caribbean Journal of Science*, 26(1–2), pp.31–44.

Hedges, S.B. & Conn, C.E., 2012. A new skink fauna from Caribbean islands (Squamata, Mabuyidae, Mabuyinae). *Zootaxa*, 244(3288), pp.1–244.

Hedges, S.B., Hass, C. a & Maxson, L.R., 1992. Caribbean biogeography: molecular evidence for dispersal in West Indian terrestrial vertebrates. *Proceedings of the National Academy of Sciences of the United States of America*, 89(5), pp.1909–1913.

Hershkovitz, P., 1977. *Living new world monkeys (Platyrrhini)*, Chicago: Univ. Chicago Press.

Hershkovitz, P., 1966. South American Swamp and Fossorial Rats of the Scapteromyine Group (Cricetinae, Muridae) with comments on the Glans Penis in murid taxonomy. *Z. Saugetierk*, 31(2), pp.81–149.

Hershkovitz, P., 1990. Titis New World monkeys of the genus *Callicebus* (Cebidae, Platyrrhini): a preliminary taxonomic review. *Fieldiana Zoology*, 55, pp.1–109.

Hofreiter, M. et al., 2001. Ancient DNA. *Nature reviews. Genetics*, 2(5), pp.353–359.

Horn, S., 2012. Target enrichment via DNA hybridization capture. *Methods in Molecular Biology*, 840, pp.177–188.

Iturralde-Vinent, M.A. & MacPhee, R.D.E., 1999. Paleogeography of the Caribbean region: implications for Cenozoic biogeography. *American Museum of Natural history*.

Kay, R.F., 2015. Biogeography in deep time - What do phylogenetics, geology, and paleoclimate tell us about early platyrrhine evolution? *Molecular Phylogenetics and Evolution*, 82(PB), pp.358–374.

Kay, R.F. et al., 2004. Olfactory fossa of *Tremacebus harringtoni* (Platyrrhini, early Miocene, Sacanana, Argentina): Implications for activity pattern. *Anatomical Record - Part A Discoveries in Molecular, Cellular, and Evolutionary Biology*, 281(1), pp.1157–1172.

Kearse, M. et al., 2012. Geneious Basic: an integrated and extendable desktop software platform for thr organization and analysis of sequence data. *Bioinformatics*, 28(12), pp.1647–1649.

Kehlmaier, C. et al., 2016. Tropical ancient DNA reveals relationships of the extinct Bahamian giant tortoise *Chelonoidis alburyorum*. *Proceedings of the Royal Society B: Biological Sciences*, 21(16), pp.1–7.

Lack, D., 1947. *Darwin's finches*, Cambridge, UK: Cambridge Univ. Press.

Lanfear, R., Calcott, B., Ho, S. Y., & Guindon, S., 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, 29, pp.1695–1701.

Larkin, M.A. et al., 2007. Clustal W and Clustal X. *Bioinformatics*, 23, pp.2947–2948.

Lister, A.M., 1996. Dwarfing in island elephants and deer: Processes in relation to time of isolation. *Symposium of the Zoological Society of London*, 69, pp.277–292.

Losos, J.B. et al., 1998. Contingency and determinism in replicated adaptive radiations of island lizards. *Science*, 279(5359), pp.2115–2118.

MacArthur, R.H. & Wilson, D.E., 1967. *The theory of biogeography*, New Jersey: Princeton Univ. Press.

Macphee, R.D.E., 2009. Insulae infortunatae: establishing a chronology for Late Quaternary mammal extinctions in the West Indies. In G. Haynes, ed. *American Megafaunal Extinctions at the End of the Pleistocene*. New York: Springer, pp. 169–193.

MacPhee, R.D.E., Horovitz, I., Arredondo, O., & Vasquez, O. J., 1995. A new genus for the extinct Hispaniolan monky *Saimiri bernensis* Rimoli, 1977, with notes on its systematic position. *AmMusNov*, 3134, pp.1–21.

MacPhee, R.D.E., 1984. Quaternary mammal localities and heptaxodontid rodents of Jamaica.

AmMusNov, (2803), p.

MacPhee, R.D.E. & Fleagle, J.G., 1991. Postcrainial remains of *Xenothrix mcgregori* (Primates, Xenothrihidae) and other Late Quaternary mammals from Long Mile Cave, Jamaica. *Bulletin of the American Museum of Natural History*, (206), pp.287–321.

MacPhee, R.D.E. & Flemming, C., 2003. A Possible Heptaxodontine and Other Caviidan Rodents from the Quaternary of Jamaica. *American Museum Novitates*, 3422(1), p.1.

MacPhee, R.D.E. & Flemming, C., 1999. Requiem æternam: the last five hundred years of mammalian species extinctions. In *Extinctions in Near Time: Causes, Contexts, and Consequences*. New York: Kluwer Academic, pp. 333–371.

MacPhee, R.D.E. & Horovitz, I., 2004. New Craniodental Remains of the Quaternary Jamaican Monkey Xenothrix mcgregori (Xenotrichini, Callicebinae, Pitheciidae), with a Reconsideration of the Aotus Hypothesis. *American Museum Novitates*, 3434, pp.1–51.

MacPhee, R.D.E. & Meldrum, J., 2006. Postcranial remains of the extinct monkeys of the Greater Antilles, with evidence for semiterrestriality in Paralouatta. *American Museum Novitates*, 3516(1), p.1.

Marx, V., 2017. Genetics: new tales from ancient DNA. *Nature methods*, *14*(8), p.771.

Meyer, M. & Kircher, M., 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*, 5(6).

Miller, M.A., Pfeiffer, W. & Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *Gateway Computing Enviroments Workshop (GCE)*, pp.1–8.

Millien, V., 2006. Morphological evolution is accelerated among island mammals. *PLoS Biology*, 4(10), pp.1863–1868.

Musser, G.G. & D., C.M., 2005. Family Cricetidae. In D. E. Wilson & D. M. Reeder, eds. *Mammal Species of the World: A Taxonomic and Geographic Reference*. Baltimore: Johns Hopkins University Press, pp. 955–1189.

Opazo, J.C. et al., 2006. Phylogenetic relationships and divergence times among New World monkeys (Platyrrhini, Primates). *Molecular Phylogenetics and Evolution*, 40(1), pp.274–280.

Rambaut, A. et al., 2014. Tracer v1.6. *University of Edinburgh, Edinburgh, UK*.

Rambaut, A. & Drummond, A.J., TreeAnnotator v1.6.1. *University of Edinburgh, Edinburgh, UK*.

Reed, F.A. Kontanis, E. J., Kennedy, K. A., & Aquadro, C. F., 2003. Brief communication: Ancient DNA prospects from Sri Lankan highland dry caves support an emerging global pattern. *American Journal of Physical Anthropology*, 121(2), pp.112–116.

Renaud, S. & Millien, V., 2001. Intra- and interspecific morphological variation in the field mouse species Apodemus argenteus and A. speciosus in the Japanese archipelago: the role of insular isolation and biogeographic gradients. *Biological Journal of the Linnean Society*, 74, pp.557–569.

Rímoli, R., 1977. *A new species of monkey (Cebidae: Saimirinae:* Saimiri*) from Hispaniola (original title in Spanish)*, CENDIA, Cuadernos Cendia Univ Autónoma Santa Domingo, Dominican Republic.

Rivero, M. & Arrendondo, O., 1991. *Paralouatta varonai*, a new Quaternary platyrrhine from Cuba. *Journal of Human Evolution*, 21, pp.1–11.

Ronquist, F. & Huelsenbeck, J.P., 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19, pp.1572–1574.

Rosenberger, A.L. et al., 2015. 1.32 ± 0.11 Ma age for underwater remains constrain antiquity and longevity of the Dominican primate *Antillothrix bernensis*. *Journal of Human Evolution*, 88, pp.85–96.

Rosenberger, A.L., 2011. Evolutionary morphology, platyrrhine evolution, and systematics. *Anatomical Record*, 294(12), pp.1955–1974.

Rosenberger, A.L. et al., 2011. First skull of *Antillothrix bernensis*, an extinct relict monkey from the Dominican Republic. *Proceedings. Biological sciences / The Royal Society*, 278(1702), pp.67–74.

Rosenberger, A.L., 2002. Platyrrhine paleontology and systematics: the paradigm shifts. In W. C. Hartwig, ed. *The primate fossil record*. New York: New York: Cambridge Univ. Press, pp. 151–160.

Rosenberger, A.L., 1977. *Xenothrix* and ceboid phylogeny. *Journal of Human Evolution*, 6(5), pp.461–481.

Rosenberger, A.L. & Tejedor, M.F., 2013. The misbegotten: long lineages, long branches and the interrelationships of *Aotus*, *Callicebus* and the *saki-uacaris**. *Evolutionary Biology and Conservation of Titis, Sakis and Uacaris*, pp.13–22.

Schillaci, M.A. et al., 2007. Morphology and somatometric growth of long-tailed macaques (*Macaca fascicularis fascicularis*) in Singapore. *Biological Journal of the Linnean Society*, 92(4), pp.675–694.

Schillaci, M.A., Meijaard, E. & Clark, T., 2009. The effect of island area on body size in a primate species from the Sunda Shelf Islands. *Journal of Biogeography*, 36(2), pp.362–371.

Schluter, D., 2000. *The Ecology of Adaptive radiation*, OUP Oxford.

Schneider, H. et al., 1993. Molecular phylogeny of the New World Monkeys (Platyrrhini. Primates). *Molecular phylogenetics and evolution*, 2(3), pp.225–242.

Schneider, H. & Rosenberger, A.L., 1996. Molecules, morphology, and platyrrhine systematics. *Adaptive Radiations of Neotropical Primates*, pp.3–19.

Shimodaira, H. & Hasegawa, M., 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics*, 17(12), pp.1246–1247.

Smith, C.I. et al., 2001. Not just old but old and cold? *Nature*, 410(April), pp.771–772.

Sondaar, P.Y., 1991. Island mammals of the past. *Science Progress*, 75(298 Pt 3-4), pp.249–264.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30, pp.1312–1313.

Thomas, O., 1903. Notes on South American monkeys, bats, carnivores, and rodents, with descriptions of new species. *Ann. Mag. Nat. Hist.*, 7(12), pp.455–464.

Thomas, O., 1898. On indigenous Muridæ in the West Indies; with the description of a new Mexican

Oryzomys. *Journal of Natural History*, 1(2), pp.176–180.

Turvey, S.T., 2009. *Holocene extinctions*, OUP Oxford.

VanValen, L., 1973. Pattern and the balance of nature. *Evolutionary Theory*, pp.31–49.

Vázquez-Miranda, H., Navarro S, A.G. & Morrone, J.J., 2007. Biogeographical patterns of the avifaunas of the Caribean bassin islands: a parsimony perspective. *Cladistics*, 22, pp.1–21.

Wallace, A.R., 1869. *The Malay Archipelago: the land of the orang-utan and the bird of paradise; a narrative of travel, with studies of man and nature*, London, UK: Macmillan.

Welker, F. et al., 2015. Ancient proteins resolve the evolutionary history of Darwin's South American ungulates. *Nature*, 522(7554), pp.81–84.

Wilkinson, M.J. et al., 2010. Megafan environments in northern South America and their impact on Amazon Neogene aquatic ecosystems. In C. Hoorn & F. P. Wesselingh, eds. *Amazonia landscape and species evolution: a look into the past*. Oxford: Wiley-Blackwell, pp. 162–184.

Williams, E. & Koopman, K., 1952. West Indian fossil monkeys. *American Museum Novitates*, 1546(1546), pp.1–16.

Wilson, D.E. & Reeder, D.M., 2005. *Mammal Species of the World: A Taxonomic and Geographic Reference, Volume 1*,

Woods, C.A. & Sergile, F.E., 2001. *Biogeography of the West Indies: Patterns and Perspective, Second Edition: Past Present and Future*,

Appendix Chapter 4. Ancient DNA of the extinct Jamaican monkey Xenothrix reveals extreme insular change within a morphologically conservative primate radiation

Table 1a. Genes used to generate baits for capture enrichment

Gene	Length (bp)	NCBI accession	Species
epsilon globin (HBE)	1706	AAA35404.1	*Cheracebus torquatus*
oxytocin receptor (promotor region) (OXTR)	1872	KF701386.1	*Plecturocebus cupreus*
Von Willebrand factor (vWF)	1238	JX856254	*Plecturocebus moloch*
interphotoreceptor retinoid binding (IRBP)	1275	JX856220.1	*Plecturocebus moloch*
growth hormone receptor (GHR)	904	JX856185.1	*Plecturocebus moloch*
taste receptor 2 member 38 (TAS2R38)	1002	JQ272231.1	*Plecturocebus moloch*
mitochondrial genome	16689	NC024630.1	*Cheracebus lugens*

Table 2a. Genes used as reference sequences for alignment of raw reads during analysis

Gene	Length (bp)	Species	NCBI accession
epsilon-globin	1706	*Callicebus toruatus*	AAA35404.1
oxytocin receptor (promotor region)	1872	*Callicebus cupreus*	KF701386.1
vWF (von Willebrand factor)	1238	*Callicebus moloch*	JX856254.1
IRBP (interphotoreceptor retinoid binding protein)	1275	*Callicebus moloch*	JX856220.1
GHR (growth hormone receptor)	904	*Callicebus moloch*	JX856185.1
TAS2R38 (Camopa bitter taste receptor)	1002	*Callicebus moloch*	JQ272231.1
mitochondrion whole genome	16689	*Callicebus lugens*	NC_024630.1
ABCA1 (ATP-binding cassette sub-family A)	804	*Callicebus lugens*	KU694299.1

Figure 1a. (part 1) Coverage of the reference (whole mitochondrial genome) by mapped reads. Reads mapped using CLC genomics work bench v.8 (CLC Bio-Qiagen, Aarhus, Denmark)

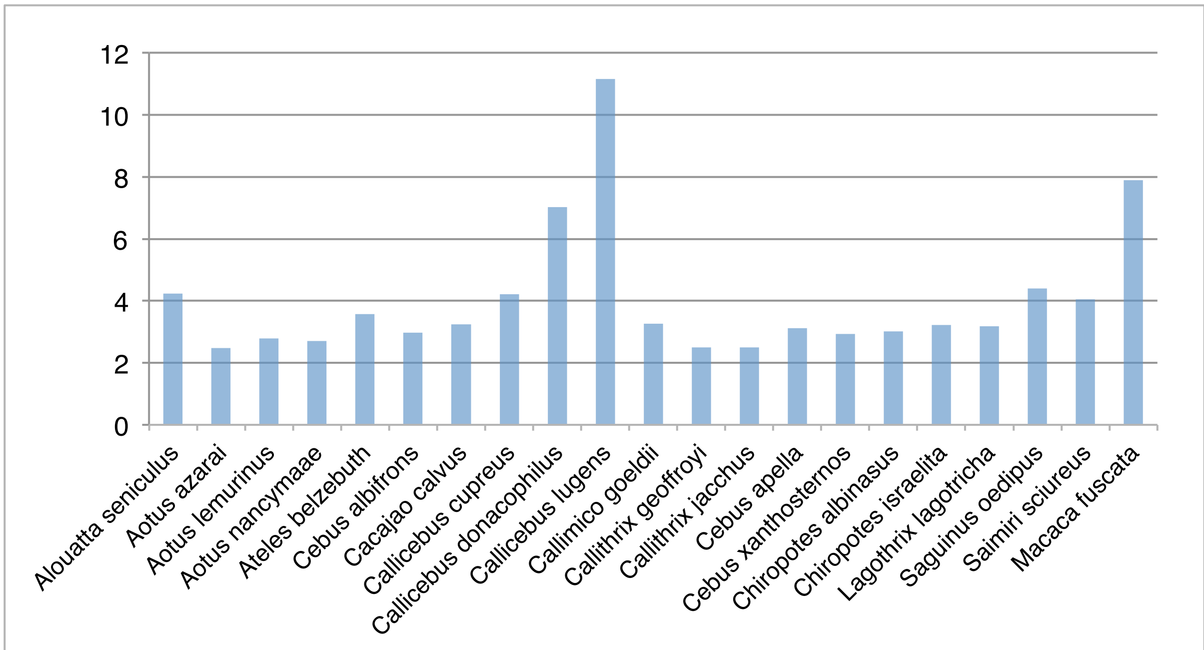


Figure 1a. (part 2) Total number of reads mapped to the reference (whole mitochondrial genome). Reads mapped using CLC genomics work bench v.8 (CLC Bio-Qiagen, Aarhus, Denmark)

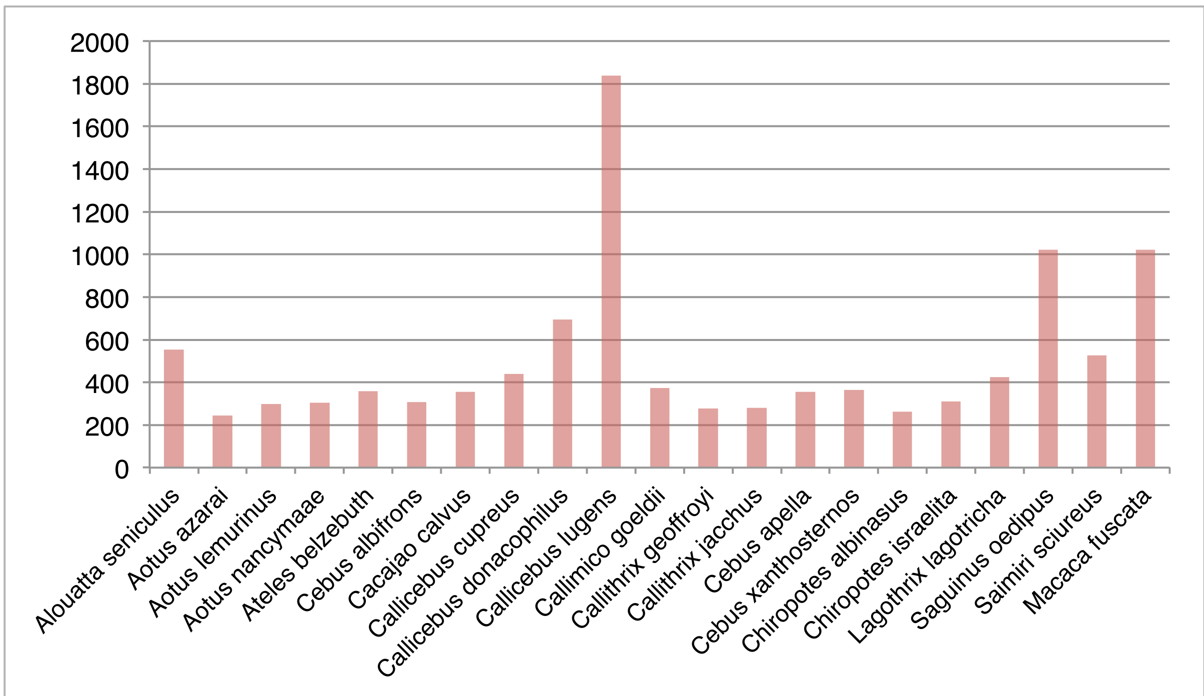


Table 3a. substitution models chosen for the partitioned dataset

Subset	Best model	Subset Partitions
		Mitochondrial genome
1	GTR+G	rrnL, rrnS, trnA, trnC, trnD, trnF, trnH, trnI, trnL1, trnL2, trnM, trnN, trnQ, trnR, trnS2, trnW, trnY
2	GTR+G	atp6, atp8, cob, cox3, nad1, nad2, nad3, nad4a, nad4b, nad4l, nad5, nad6, trnP, trnS1, trnV
3	GTR+G	cox1, cox2, trnE, trnG, trnK, trnT
		ADORA3 (adenosine A3 receptor)
1	K80+G	Gene
		APP (Amyloid beta A4 protein)
1	HKY	Gene
		CREM (CAMP Responsive Element Modulator)
1	K80	Gene
		DENND5A (DENN Domain Containing 5A)
1	K80	Gene
		DMRT1 (Doublesex And Mab-3 Related Transcription Factor 1)
1	HKY	Gene
		ERC2 (ELKS/RAB6-Interacting/CAST Family Member 2)
1	HKY	Gene
		FAM123B (APC Membrane Recruitment Protein 1)
1	K80	Gene
		FES (FES Proto-Oncogene)
1	HKY+G	Gene
		FOX G1 (Forkhead box protein G1)
1	HKY	Gene
		MAPKAP1 (mitogen-activated protein kinase associated protein 1)
1	HKY	Gene

Table 3a. (continued) substitution models chosen for the partitioned dataset

Subset	Best model	Subset Partitions
1	HKY	Gene
	NEGR1 (Neuronal Growth Regulator 1)	
1	F81	Gene
	NPAS3 (Neuronal PAS domain protein 3)	
1	HKY	Gene
	RAG1 (Recombination Activating 1)	
1	K80+G	Gene
	RAG2 (Recombination Activating 2)	
1	K80+G	Gene
	RPGRIP1 (Retinitis Pigmentosa GTPase Regulator Interacting Protein 1)	
1	K80	Gene
	SGMS1 (Sphingomyelin Synthase 1)	
1	F81	Gene
	SIM1 (Single-minded homolog 1)	
1	F81	Gene
	ZFX (Zinc Finger Protein)	
1	HKY	Gene

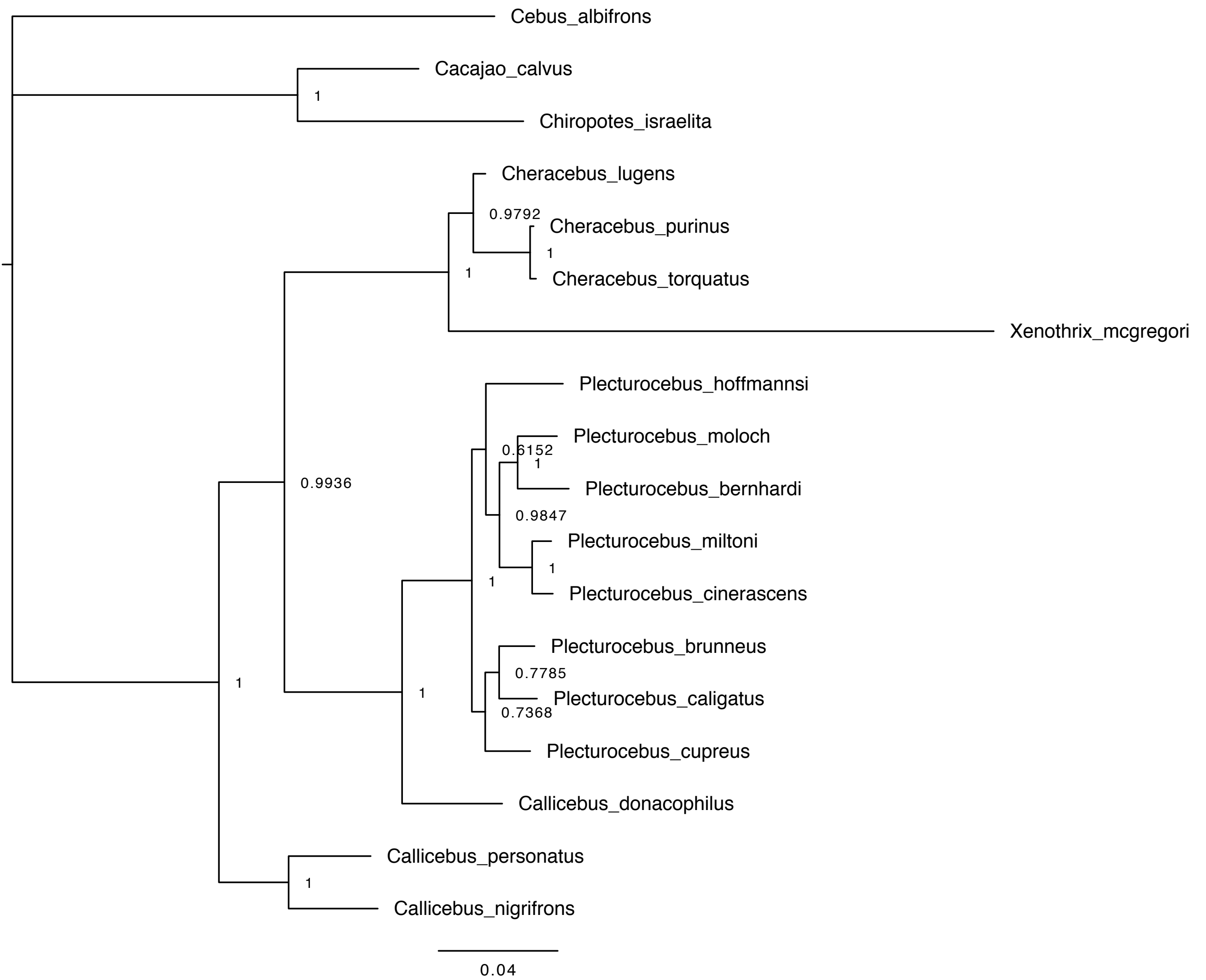


Figure 2a. Bayesian phylogeny constructed using MrBayes (Ronquist & Huelsenbeck 2003) node values represent posterior probabilities. Scale indicates nucleotide substitutions per site.

Table 4a. Tests of alternative topologies suggested by previous studies made by submitting the sitewise log-likelihood values obtained from RAxML v.8 (Stamatakis 2014) to CONSEL (Shimodaira & Hasegawa 2001), to calculate the p-values for each tree topology by using the AU test.

Hypothesis number	obs log liklihood score	au p-value	np p-value	bp p-value	pp p-value	kh p-value	sh p-value	wsh p-value
H1	-430.9	1	1	1	1	1	1	1
H2	430.9	1E-50	4E-16	0	7E-188	0	0	0
H3	722.6	0.003	0.000 00005	0	1e-314	0	0	0
H4	760.7	0.0002	0.000 1	0	0	0	0	0
H5	-523.9	1	1	1	1	1	1	1

Table 5a. Fossil constraints and priors used in divergence date analysis

Root	Node	Fossil	Age	Prior distribution
1	Cebinae *Cebus albifrons – Saimiri sciureus*	*Neosaimiri* (Stirton, 1951)	Mid Miocene (12.5 mya)	logNormalPrior mean="1.287" stdev="0.8" offset="12.5"
2	Pitheciidae *Cacajao calvus – Chiropotes israelita – Pithecia pithecia*	*Proteropithecia* (Kay et al., 1998)	Mid Miocene (15.7 mya)	logNormalPrior mean="1.016” stdev="0.8" offset="15.7"

Figure 3a. Specimens used for the extraction of aDNA. (left) A) femur: AMNH 268003 and (right) B) proximal ulna: AMNH 268010. Both specimens were excavated in Somerville Cave, Clavendon, Jackson's Bay, Jamaica.

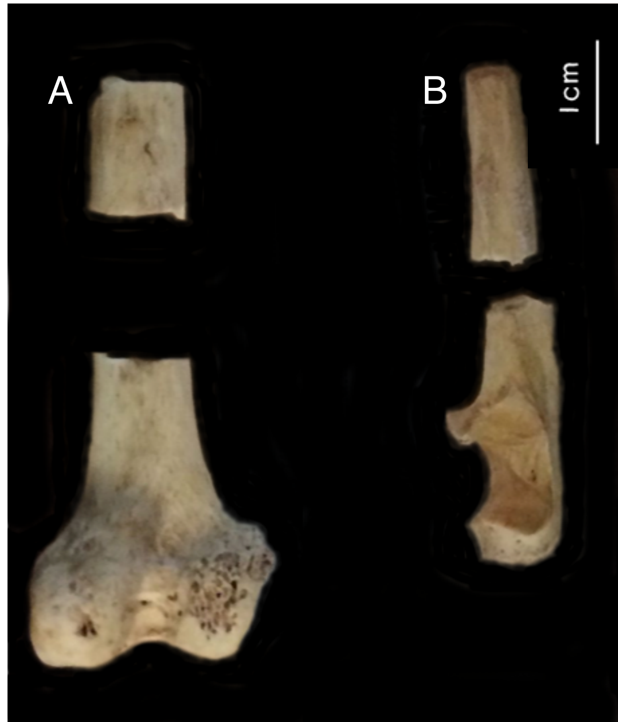


Figure 4a. Consensus tree adapted from MacPhee and Horovitz (2004) including Caribbean taxa. Phylogeny was constucted from nine most parsimonious trees based on 87 morphological charaters.

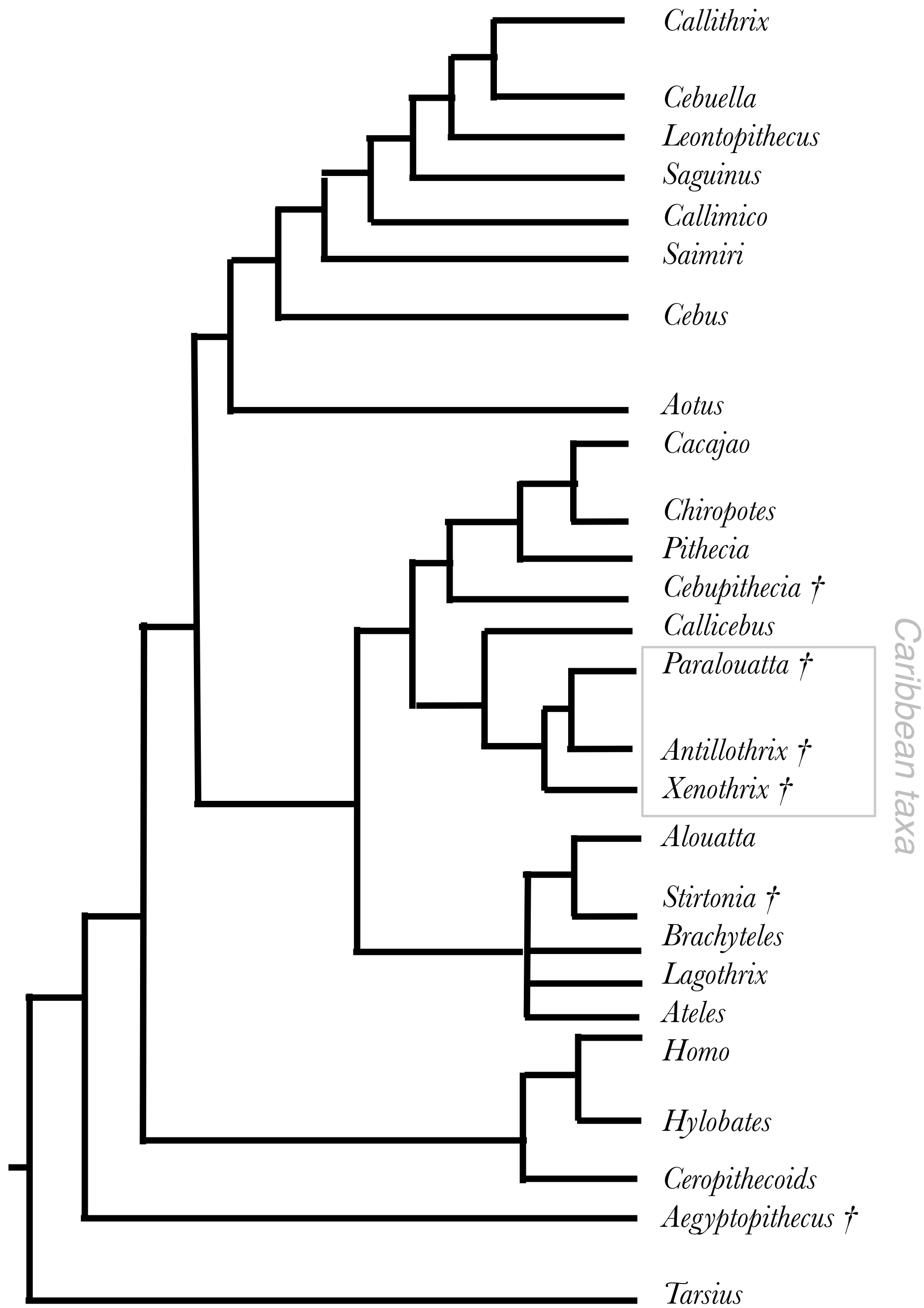


Table 6a. BLAST results from the NCBI database for reads mapped to reference sequences

Species mapped to:	Genbank ID	
Callicebus lugens	NC_024630.1	
1st BLAST hit		
Description	Query cover	Identity
Callicebus lugens mt genome	77%	85%
2nd BLAST hit		
Description	Query cover	Identity
Callicebus cupreus mt genome	28%	78%
3rd BLAST hit		
Description	Query cover	Identity
Callicebus donacophilus mt genome	36%	77%
Other hits:		
Callicebus 36-77%, Alouatta 19%, Aotus 13-18%, Saguinus 11%, Callithrix 14%, Leotopithecus 5%, Cebus 5% and Homo not above 13%		

Species mapped to:	Genbank ID	
Tarsius bancanus	NC_002811.1	
1st BLAST hit		
Description	Query cover	Identity
Tarsius bancanus mt genome	82%	84%
2nd BLAST hit		
Description	Query cover	Identity
Canis adustus mt genome	9%	75%
3rd BLAST hit		
Description	Query cover	Identity
Mustela	19%	74%
Other hits:		
Tarsius 7-82%, Canis 9-10%, Mustela 19%, Cercocebus 4% and Pan 12%		

Species mapped to:	Genbank ID	
Leontopithecus rosalia	NC_021952.1	
1st BLAST hit		
Description	Query cover	Identity
Leontopithecus whole mt genome	69%	89%
2nd BLAST hit		
Description	Query cover	Identity
Homo secernin (SCRN1)	1%	95%
3rd BLAST hit		
Description	Query cover	Identity
Pan troglodytes BAC CH251-499L3	1%	94%
Other hits:		
Leontopithecus rosalia 69%, Homo and Pan not above 1%		

Species mapped to:	Genbank ID	
Callicebus donacophilus	NC_019801.1	
1st BLAST hit		
Description	Query cover	Identity
Callicebus donacophilus mt genome	69%	86%
2nd BLAST hit		
Description	Query cover	Identity
Callicebus moloch 16s	5%	86%
3rd BLAST hit		
Description	Query cover	Identity
Homo nmt	9%	77%
Other hits:		
Callicebus 2-65% and Homo not above 9%		

Species mapped to:	Genbank ID	
Saguinus Oedipus	NC_021960.1	
1st BLAST hit		
Description	Query cover	Identity
Saguinus mt genome	76%	85%
2nd BLAST hit		
Description	Query cover	Identity
Callithrix mt genome	9%	79%
3rd BLAST hit		
Description	Query cover	Identity
Leontopithecus	3%	81%
Other hits:		
Saguinus 1-76%, Callithrix 9%, Leontopithecus 3%, Alouatta 3%, Aotus 7%, Lagothrix 13%, Cebus 55, Ateles 1%		

Species mapped to:	Genbank ID	
Macaca fuscata	NC_025513.1	
1st BLAST hit		
Description	Query cover	Identity
Macaca fuscata mt genome	90%	85%
2nd BLAST hit		
Description	Query cover	Identity
Macaca mulatta mt genome	85%	82%
3rd BLAST hit		
Description	Query cover	Identity
Macaca arctoides mt genome	79%	80%
Other hits:		
All Macaca 6-90%		

Species mapped to:	Genbank ID		
Alouatta seniculus	gi|927371820|ref|NC_027825.1		
1st BLAST hit			
Description	Query cover	Identity	
Alouatta seniculus mt genome	68%		89%
2nd BLAST hit			
Description	Query cover	Identity	
Ateles belzebuth mt genome	14%		79%
3rd BLAST hit			
Description	Query cover	Identity	
Lagothrix mt genome	13%		79%
Other hits:			
Aloutta 11-68%, Ateles 8-14% and Pan and Homo not above 12%			

Chapter 5. Phylogeographic analysis of sympatric Cuban hutia reveal parallel patterns evolution in a mammalian insular radiation

6.1 Abstract

The insular radiation of Cuban hutias (Capromyidae) has been characterised by high rates of Holocene extinction. Surviving species in three genera, *Capromys*, *Mesocapromys*, and *Mysateles*, are critically understudied at the molecular level, despite the fact that many of these taxa are threatened by extinction, and conservation priorities often rely upon the underlying taxonomy of species relationships. Morphological analyses of Cuban capromyids have proposed different levels of species diversity between these genera, despite parallel evolutionary pressures likely to have acted on all taxa, and recent molecular analysis using limited mitochondrial DNA data has suggested that *Capromys* is split into distinct eastern and western clades, which does not match any morphology-based hypotheses of hutia relationships. The recent extinction of many local Cuban hutia populations, and the threatened nature of most surviving populations and species, is a limiting factor in the ability of studies to accurately delimit species for the purposes of further evolutionary study and conservation. Here, species limits and phylogeographic structure in Cuban hutias has been investigated using molecular data (whole mitochondrial genomes and five nuclear genes) extracted from museum specimens using specialist ancient DNA techniques. Combined data analysis recovered two distinct clades within *Capromys* with a divergence date of ~1.7 Ma (95% HPD = 1.245–3.572 Ma) between the two lineages, which correspond to eastern and western clades proposed by previous studies. This study also finds a primary divergence between eastern and western clades within *Mesocapromys* with a divergence date of ~1.2 Ma (95% HPD = 0.726–2.293 Ma), suggesting congruent phylogeographic patterns and similar evolutionary histories between *Capromys* and *Mesocapromys* across Cuba. Levels of species diversity within Cuban hutia should therefore be revised, critically some of the highly threatened species of *Mesocapromys* that are currently recognised as conservation priorities may no longer constitute valid taxa.

6.2 Introduction

Island evolution and the Caribbean

Islands are often described as “natural laboratories” that can be used to study the intricate circumstances that generate both speciation events and extinctions (Losos & Ricklefs, 2009). The Caribbean islands are unusual in the fact that they were colonised by several lineages of mammals as well as by other vertebrate lineages. As few terrestrial mammals colonise oceanic island systems, the Caribbean offers a rare opportunity to study the colonisation histories, species relationships and patterns of evolutionary diversification of endemic insular mammals. However, a recent Holocene extinction event that eliminated over 100 Caribbean mammal species or separate island populations (Davalos & Turvey, 2012; Carstensen *et al*., 2013) left only a few scattered and threatened extant mammal populations, and limited remains of the Caribbean’s extinct endemic mammal fauna, inhibiting the ability of researchers to study their evolutionary history and dynamics. In order for the continued study of evolutionary processes in the Caribbean, palaeogenetic techniques need to be applied to the extinct or threatened mammal fauna to test hypotheses of evolutionary relationships proposed on morphological grounds. Of particular interest is phylogeographic investigation of surviving or recently extinct species, as this has the potential to uncover the mechanisms acting on diverging lineages (e.g. intra-island geological or ecological complexity, or inter-island or intra-island vicariance) and aid prioritisation of conservation attempts to protect surviving species.

Cuba

Cuba is the largest island in the Greater Antilles, and incorporates a wide variety of ecological habitat types across a range of elevations, and has a complex coastline with mangrove forests and small offshore islands. The largest of Cuba’s offshore islands is Isla de la Juventud (Isle of Youth), previously known as Isla de Pinos (Isle of Pines). Cuba’s offshore islands were previously connected to each other and the Cuban mainland when sea levels were low, and disconnected when sea levels were high, over periods in the island’s past (Milne et al. 2005). Around 17,000 years ago sea levels were 120 m below current levels, possibly leading to an increase in island size and removal of past and current biogeographic barriers (Gascoyne et al. 1979). Cuba, with its large size and varied habitat types, provides ample opportunities for the geographic and ecological isolation of populations and the generation of species. Past biogeographic barriers on the Cuban mainland provide the potential for allopatric speciation, the geographic isolation of populations, and additionally peripatric speciation may have also occurred between mainland Cuba and the offshore islands that surround it (Coyne & Orr, 2004; Losos & Glor, 2003).

Cuban hutias

The Greater Antillean hutias (Capromyidae) are part of a once more extensive endemic island radiation, that now includes only a few surviving members after an unprecedented Holocene extinction event. Over 26 proposed species of hutia were once living across the Greater Antilles that make up the larger islands in the west of the Caribbean, prior to the Holocene extinction event; however, ~57% (Borroto-Páez & Mancina, 2017) disappeared after the European historical arrival of humans on the islands (Siobhán B. Cooke et al. 2017). This study focuses on hutia from the island of Cuba, which is home to the majority of the Caribbean's surviving capromyid species (eight of the eleven currently recognized extant species) (Upham & Borroto-Páez, 2017). Cuban capromyids are split into three genera: *Capromys, Mesocapromys* and *Mysateles*. Representatives of *Capromys* and *Mesocapromys* are thought to have originally had similar distributions across the mainland of Cuba before some populations were driven to extinction.

As part of an insular radiation, there is evidence of niche-driven ecological adaptations between Cuban hutia genera. *Capromys* contains large bodied semi-arboreal forms, *Mesocapromys* species are smaller bodied and characterised by adaptations to mangrove habitats, and *Mysateles prehensilis* and *M. melanurus* are obligate arboreal specialists with prehensile tails.

Capromys, Mesocapromys and *Mysateles* have a complex taxonomic history. Historical species designations were sometimes based on morphological descriptions of populations that are no longer extant, and many of these descriptions were also associated with little or no geographic locality data (Arredondo 2012; Diaz-Franco 2001). Many populations have been categorised and re-categorised as synonyms, species, subspecies, or individuals with only population-level distinction (Borroto-Páez, Mancina, Woods, & Kilpatrick, 2012; Silva Taboada, Suarez Duque, & Diaz-Franco, 2007; Woods & Sergile, 2001). Current species designations also continue to change as the generation of molecular data is used to reassess taxonomic designations based solely on morphology-based methods.

Table 1. List of capromyid species and subspecies found on Cuba with IUCN threat status if accessed

Genus		Species	Location	IUCN threat status
Capromys	*Capromys p. pilorides*	Say, 1822	Mainland Cuba	Least concern
	Capromys p. relictus	Allen, 1911	North Isla de la Juventud	n/a
	Capromys p. ciprianoi	Páez, Pérez, & Garcia, 1992	South Isla de la Juventud	n/a
	Capromys p. gundlachianus	Varona, 1983	Sabana-Camaguey archipelago	n/a
	Capromys p. doceleguas	Varona, 1980	Jardines de la Reina archipelago	n/a
	C. p. ssp. 1	Borroto-Páez, Mancina, Woods, & Kilpatrick, 2012	Los Canarreos archipelago	n/a
	C. p. ssp. 2	Borroto-Páez et al., 2012; Rafael Borroto-Páez, 2005	Sabana-Camaguey archipelago	n/a
	C. p. ssp. 3	Borroto- Páez et al., 2012; Borroto-Páez, Pons, Mancina, & Oviedo, 2007	Zapata Swamp (South)	n/a
	C. p. ssp. 4	Borroto- Páez et al., 2012; Borroto-Páez, Pons, Mancina, & Oviedo, 2007	Zapata Swamp (South)	n/a
Mysateles	*Mysateles p. prehensilis*	Poeppig, 1824	Mainland Cuba	Near threatened
	Mysateles gundlachi	Chapman, 1901	North Isla de la Juventud	n/a
	Mysateles meridionalis	Varona, 1986	Cayos de Ana Maria	n/a
	Mysateles melanurus	Poey, 1865	Mainland Cuba (East)	Vulnerable A2cd
Mesocapromys	*Mesocapromys angelcabrerai*	Varona, 1979	Cayos de Ana Maria	Endangered C2a
	Mesocapromys auritus	Varona, 1970	Sabana-Camaguey archipelago	Endangered C2a(ii)
	Mesocapromys nanus	Allen, 1917	Zapata Swamp	Critically Endangered (Possibly Extinct) C2a(i)
	Mesocapromys sanfelipensis	Varona & Garrido, 1970	Cayos de San Felipe	Critically Endangered (Possibly Extinct) D

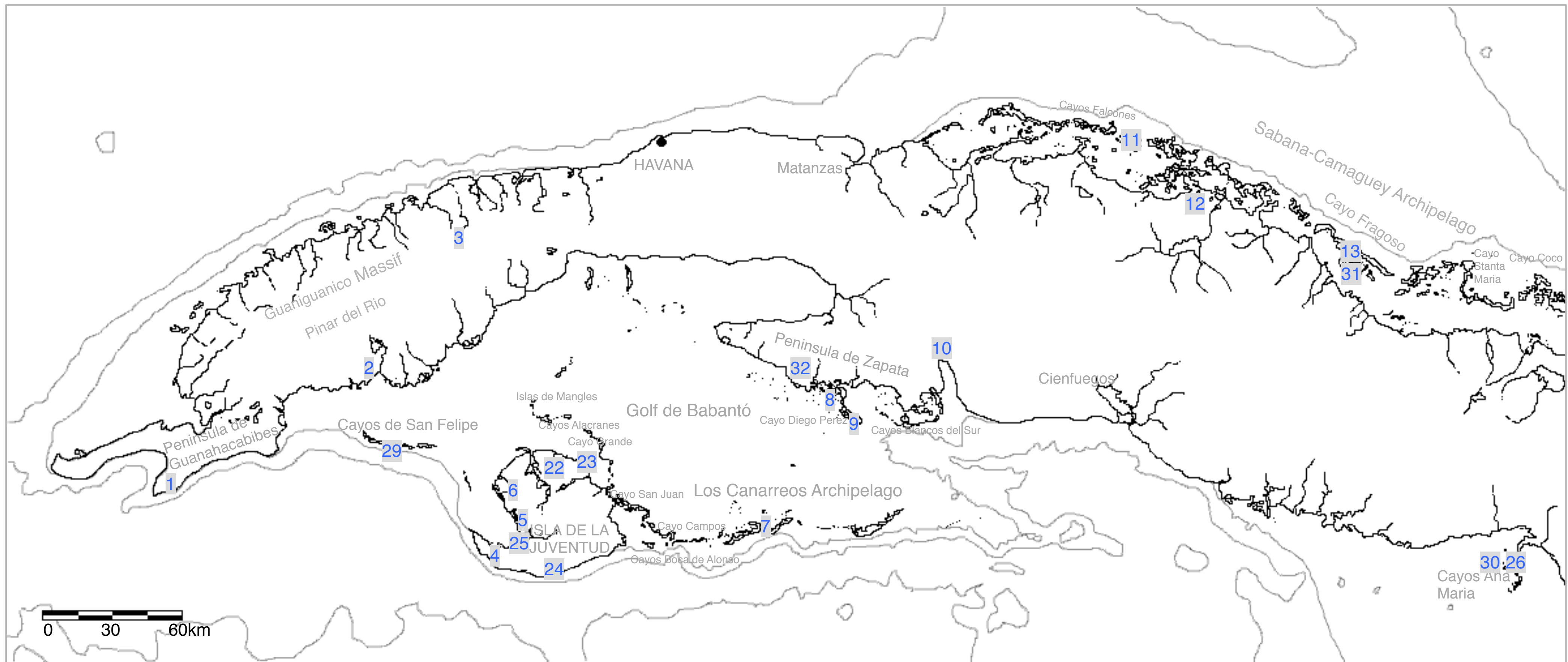


Figure 1, Part 1 Map of Cuba (west) and offshore islands including Isla de la Juventud, with scale bar and labels for localities. Numbered localities indicate the locations of different species, subspecies and populations of Cuban Capromys and Mesocapromys (Appendix: Table 1-3a)

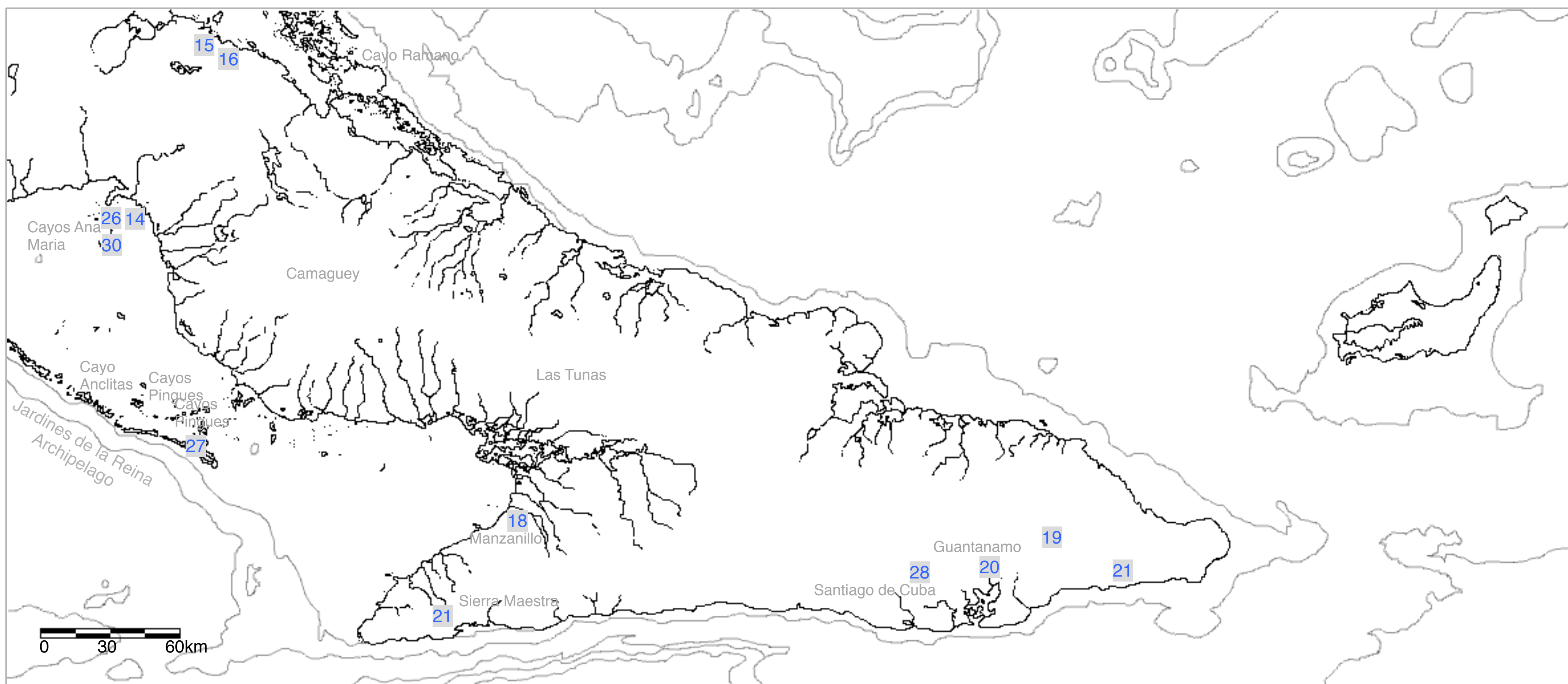


Figure 1, Part 2 Map of Cuba (east) and offshore islands including Isla de la Juventud, with scale bar and labels for localities. Numbered localities indicate the locations of different species, subspecies and populations of Cuban Capromys and Mesocapromys (Appendix: Table 1-3a)

Capromys

Previous studies have suggested that mainland *Capromys* populations represent a single subspecies *Capromys pilorides pilorides* (Desmarest's hutia), with geographically isolated offshore populations representing separate species or subspecies (Borroto-Páez, Mancina, Woods, & Kilpatrick, 2012; Woods & Sergile, 2001). Four subspecies have been described from populations that inhabit Cuba's many offshore islands. *Capromys pilorides gundlachianus* (Varona, 1983) was described from the Sabana cays and islands off of the northern cost of Cuba; the main basis for this description was the darkened, almost black coat colour of the population. *Capromys pilorides doceleguas* (Varona, 1980) was described from the Jardines de la Reina cays off the south-eastern coast of Cuba; the diagnostic features of this subspecies description included tail length and cranial dimensions that were larger then mainland specimens. A third subspecies, *Capromys pilorides relictus* is described from northern Isla de la Juventud (Hernandez et al 1987), whilst *Capromys pilorides ciprianoi* was later described from a population in the south of the same offshore island (Paez, Paez & Garcia 1992). Although *Capromys* populations are also found on many other offshore islands, these populations have not been assigned to any of the previously mentioned subspecies or formally described.

Sequence divergence in mitochondrial gene cytochrome *b* between these putative *Capromys* taxa was found to be very low (<0.4%) (Woods & Sergile 2001; Milishnikov et al. 2010). Conversely, sequence divergence estimates between species in the South American Echimyidae, the capromyids' closest relatives (Fabre *et al.,* 2015), range from 8.8% to 19.7% (Lara and Patton 1996), suggesting that sequence divergence between *Capromys* populations may therefore be below species level. Due to the low level of sequence divergence between *Capromys p. ciprianoi* and *Capromys p. relictus* the former is considered as a junior synonym of *Capromys p. relictus* by some authors (Woods & Sergile 2001) but is retained by others (Silva Taboada et al. 2007).

Additional putative *Capromys* subspecies and species have been discussed in other more recent studies. These also include a described species from Cayo Ballenato del Medio (*C. p. ssp*. 2), which was originally assigned to *C. p. gundlachianus*, but which has been shown to be divergent from other *Capromys* populations in cytB analysis (Borroto-Páez et al., 2012; Rafael Borroto-Páez, 2005). Morphology-only species delimitation and limited sampling in molecular studies have prevented a comprehensive analysis of Cuban *Capromys* species diversity, and assessment of the validity of these supposedly distinct taxa.

East-West split hypothesis

More recent studies have questioned earlier interpretations of within species diversity in Cuban huitas, particularly within the genus *Capromys*. Upham & Borroto-Páez (2017) suggested that rather

than genetic differentiation being restricted to off-shore island populations, with mainland hutias representing the same genetic population, Cuban huitas on the mainland and off-shore islands combined may exhibit a phylogeographic split between the east and west of the island. The authors tentatively suggest a line between Matanzas, Cienfuegos, and Villa Clara as a boundary between eastern and western *Capromys* lineages. The ~1.1 Ma divergence between eastern and western clades recovered by Upham & Borroto-Páez suggests that eastern and western clades should be regarded as discrete species. The authors call for a new or elevated name to be applied to eastern *Capromys* sp. nov., in order to distriguish eastern and western divergent taxa, but cite the confusion regarding type localities and specimens as preventing them from applying their practical taxonomic suggestions immediately.

In terms of the mechanisms responsible for the pattern of diversity seen in Cuban *Capromys*, allopatric differentiation could have been caused by partial submergence of palaeo-islands, or historical fragmentation of islands by over-water barriers (Stenson et al 2002). Reconstructions of historically emergent areas of Cuba suggest that this island consisted of three palaeo-islands separated by deep seaways throughout the Cenozoic before connecting to form the current mainland of Cuba in the Pliocene (Iturralde-Vinent & MacPhee 1999; MacPhee et al 2003; Graham 2003). Two historical seaways, the Cardenas-Cochinos and Guacanayayabo-Nipe fault lines, marked the borders between western, central, and eastern palaeo-archipelagos. The discovery of fossil marine vertebrates in central Cuba provides evidence for the continued fragmentation of Cuba during the Miocene (MacPhee et al 2003). Allopatric speciation generated by historical submergence and fragmentation of Cuba is suggested to have driven current diversity of green anole lizards (Glor et al. 2004). The arrival of anoles in the Caribbean has been dated to ~16-37 Ma ago, and capromyids are thought to have arrived in the Caribbean at around the same time (~15 Ma) (Fabre et al., 2014; Fabre et al., 2017), suggesting that similar causal factors for speciation may have affected both groups. However, unlike the Miocene divergence date estimated for green anoles, dated phylogenetic analysis of *Capromys* (Upham & Borroto-Páez 2017) suggests that divergence of eastern and western lineages (~1.1 Ma) post-dates periods of partial submergence of Cuba. A Pleistocene divergence date for the eastern and western *Capromys* clades could be related the periods of climate instability and sea level changes that were occurring globally including in the Caribbean during this period (Pregill & Olson 1981).

However, this proposed pattern of genetic structure seen in Cuban *Capromys* may also be artefactual, produced by geographic isolation due to the distance between populations (isolation by distance or IBD). In this phenomenon, any limitations to dispersal can generate genetic structure in population, as opposed to vicariance, which requires physical barriers to gene flow such as mountain ranges or seaways (Wright 1943). IBD can be identified by the observation of a relationship between genetic distance and geographic distance, due to increased migration between neighbouring populations. IBD can produce artificial phylogeographic groups when there are gaps in

the sampling of representative populations. Although there is a wide geographic spread in previous molecular analyses many populations are only represented by a single indervidual. Additionally, Upham and Borroto-Páez use a limited set of genetic markers, which were typically limited to only one or two mitochondrial genes, which could lead to erroneous evolutionary inferences (Rubinoff & Holland 2005). Although spatially high-resolution sampling may be beyond the scope of this study, the continued inclusion of additional individuals from *Capromys* populations with increased genetic sampling is necessary to test these alternative hypotheses.

An additional factor effecting Cuban *Capromys* populations or speculative species, are introgression events. Hybridization in secondary sympatric contact areas has been demonstarted in Cuban green *Anolis* (Glor et al. 2004), and as Cuban *Capromys* are found across the mainland of Cuba introgression between sympatric populations could have an impact phylogenetic analysis and species delimitation. Asymmetric introgression of mitochondrial DNA may be caused by a number of factors including the demographics of the hybridization zone (Petit and Excoffier 2009) and selection (Wielstra and Arntzen 2014). As introgression may only be visible when both mitochondrial and nuclear genetic markers are used for phylogenetic analyses, pevious studies using few and primarily mitochondrial genetic markers for analysis may result in incorrect phylogenetic assessments. The use of a more diverse range of nuclear markers is necessary to dectect fine-scale population processes obscured when only few mitochondrial markers are utilised (McGuire et al 2007).

Mesocapromys

The genus *Mesocapromys* has not been subjected to the same level of molecular study as *Capromys*, and presently contains several separate species based on morphological differences. *Mesocapromys* contains four currently recognised species, *M. nanus* (Allen 1917), *M. angelcabrerai* (Varona 1979), *M. auritus* (Varona 1970) and *M. sanfelipensis* (Varona & Garrido 1970). *M. nanus* and *M. sanfelipensis* are considered critically endangered although are possibly already extinct, *M. auritus* is considered endangered and *M. angelcabrerai* was recently elevated to the status of critically endangered. (Turvey et al. 2016).

Although morphological analysis of specimens is often the best or only way to describe species that are threatened or extinct, this method can result in misidentification, or assignment of species status to individuals that may only represent subspecies or are part of within-species variation (Knell et al. 2013). Almost the entire mainland population of *Mesocapromys* became extinct in the 20th century (Soy & Silva 2008) and because no mainland *Mesocapromys* specimens were available for comparison, previous studies may have misinterpreted the level of morphological diversity between offshore island populations that were discovered and described by Cuban researchers in the 1970s.

Because of the similarities in geographical distributions between both *Capromys* and *Mesocapromys* populations, and because causal factors for speciation in these groups (e.g. sea level changes, mountain ranges and varied ecology) would be expected to affect both genera in the same way, *Capromys* and *Mesocapromys* should be affected equally by causal factors for species generation, and so we would expect molecular studies to reflect this in similar tree topology and levels of divergence between allopatric populations within the two genera. However, recent molecular studies of Cuban hutia have focused on the genus *Capromys,* and suggest an east-west geographical split in this genus, but do not find the same pattern in *Mesocapromys*, perhaps due to incomplete sampling in this genus. Upham & Borroto-Páez included *M. nanus, M. auritus* and *M. angelcabrerai* and found very recent divergence dates between these species, ~0.7 Ma between *M. angelcabrerai* and the *M. nanus* + *M. auritus* clade and ~0.2 Ma between *M. nanus* and *M. auritus*, suggesting that previous morphology-based studies may have over estimated the diversity within this genus.

Mysateles

Previous studies have suggested that *Mysateles* contained up to five species (Taboada et al. 2007). However, more recent studies have suggested that only a single species *M. prehensilis* should be included in this genus (Turvey et al. 2016). The subspecies *Mysateles p. gundlanchi* was included in recent analyses by Upham & Borroto-Páez (2017) but recent (~0.3 Ma) divergence between this putative subspecies and *M. prehensilis* suggest that the subspecies designation may be invalid.

M. melanurus has historically been considered a member of *Mysateles* due to shared morphological features associated with arboreal lifestyles, but is now considered part of the genus *Mesocapromys* in several recent studies that included molecular data for this species (Woods & Sergile 2001; Borroto-Páez 2005; Upham & Borroto-Páez 2017). Upham & Borroto-Páez included four *M. melanurus* individuals that their phylogenetic reconstruction placed as sister to *Mesocapromys* species. However the inclusion of *M. melanurus* in previous molecular analysis has relied on few mitochondrial and nuclear genes sampled from one or two populations, and further sampling of individuals from known locations, and using additional molecular markers is necessary before morphological and ecological evidence can be fully dismissed.

Extraction and sequencing of ancient DNA from the tropics

A major hurdle for the molecular analysis of Cuban hutias has been recovering samples from populations or described species and subspecies from known geographic locations. Despite the fact that many of the surviving capromyid species are found on Cuba, many taxa have low population sizes, are rare, or inhabit areas that are remote and thus difficult to procure samples from. Museum specimens provide an excellent resource for the study of species, as many were collected before

populations declined or were lost due to extinction. However the extraction and sequencing of endogenous DNA from tropical environments is inhibited by high temperature and humidity, even from recent specimens (<200 years). Although sampling of modern DNA overcomes this problem, sampling may be limited by the current threatened nature of many capromyid species. Here we utilise specialist Next Generation Sequencing techniques and target capture enrichment in order to increase the yield of endogenous DNA from degraded specimens.

Conservation

Much of the surviving Cuban capromyid fauna is in danger of extinction (Table 1). Cuban hutias face multiple threats, such as habitat loss, competition and predation by introduced species, and hunting (Borroto-Páez & Mancina 2017; Siobhán B. Cooke et al. 2017), all of which are often compounded by small population size in island taxa. This study hopes to increase our current understanding of the genetic diversity within the group, allowing a re-assessment of species limits. For many capromyid species, little or no molecular work has been carried out to test species designations based solely on morphological analysis. Genetic differentiation between populations, subspecies or species is often cited as an important measure for the assessment of conservation priorities (Petit et al. 1998; Moritz 1995).

In *Mesocapromys*, populations found in different geographic locations have been assigned species status on morphological grounds. However, *Capromys* populations across a similar distribution are not considered to be genetically distinct to species level. Of the four currently recognised *Mesocapromys* species one is endangered (*M. auritus*), one is critically endangered (*M. angelcabrerai*) and two are likely extinct (*M. nanus* and *M. sanfelipensis*). Molecular analysis of some of these species by Upham & Borroto-Páez (2017) suggest very recent divergence dates for currently recognised species, calling in to question previous species level designations. These findings are important for the continued conservation of Cuban hutias because the re-assessment of *Mesocapromys* species as invalid, decreases the level of concern for threatened species, populations or subspecies. However if molecular analysis provides evidence to suggest that all four *Mesocapromys* species are valid, currently recognised species will remain of high conservation concern. Molecular evaluation of *Mesocapromys*, particularly *M. sanfelipensis* which is unstudied at the molecular level, and further study of *Capromys,* are needed to ensure that correct taxonomic designations are in place, as conservation priorities rely on the accuracy of species delimitations.

6.3 Study aims and objectives

This study aims to use specialist ancient DNA techniques in order to achieve three key aims: i) To extract and sequence ancient DNA from museum and sub-fossil specimens from Cuban *Capromys* and *Mesocapromys* species, subspecies and specific geographic populations, ii) to use this data

combined with data produced in previous molecular studies to reconstruct the phylogenetic relationships between taxa and interpret the relationships between allopatric populations of *Capromys* and *Mesocapromys,* in terms of whether molecular phylogenies produced are biogeographically congruent with previous studies and hypotheses, ii) to re-evaluate species definitions for *Capromys*, *Mesocapromys* and *Mysateles* using molecular data.

6.4 Methods

Specimen information

Mesocapromys, *Capromys* and *Mysateles* specimens identified as distinct species or subspecies, or from populations of known geographic location, were sampled for study. Specimens were obtained from museum collections, and included bone, hair and tissue samples taken from skins and taxidermy specimens of varying ages and preservation.

Extractions and sequencing

Extractions from bone, tooth, skin and tissue samples (Dabney et al. 2013) and Next Generation Sequencing library builds (Meyer & Kircher 2010) took place in a specialised aDNA laboratory at the Natural History Museum, London, that was separated from the post PCR laboratory. All equipment used was sterilized with bleach and UV treated before and after use. A negative control was included in the set of samples during the entire process.

Initially samples were screened using the Next Generation Sequencing platform Illumina NextSeq. To increase the amount of endogenous DNA sequenced, in-solution, hybridisation-capture enrichment was carried out on samples with low levels of endogenous DNA during mapping, prior sequencing on the Illumina NextSeq 500 (Enk et al. 2014). Capture enrichment kits (MYcroarray, Ann Arbor) were applied, using baits designed from whole mitochondrial genome and 5 nuclear genes: apolipoprotein B (apoB), recombination activating gene 1 (RAG1), von Willebrand factor (vWF), interphotoreceptor retinoid binding (IRBP) and growth hormone receptor (GHR) choosen on the basis of their availability on the NCBI database Genbank (Benson et al. 2005) for the following echimyid species: *Proechimys quadruplicatus, Proechimys simonsi, Echimys didelphoides, Phyllomys sp. Trinomys setosus* and *Proechimys longicaudatus* ((Appendix: Table 7a).

Post sequencing data processing and phylogenetic analysis

Data from this sequencing run were analysed using CLC Workbench software v.8 (CLC Bio-Qiagen, Aarhus, Denmark), where reads were paired, merged and then trimmed of adapters using default settings. Reads were then mapped to reference sequences from Genbank for whole mitochondrial genomes and nuclear genes using the mapping parameters and protocol described in the Methods

chapter of this thesis. Sequence data were then aligned using the ClustalW alignment tool (Larkin et al. 2007) as implemented in the programme Geneious v. 8.0.5 (Kearse et al. 2012).

Data generated as part of this study were combined with data generated in previous studies in order to produce a complete phylogeny including all available data for *Capromys*, *Mesocapromys* and *Mysateles* species, subspecies and individuals from known localities. Whole mitochondrial genome data were therefore combined with available mitochondrial genes and nuclear genes (Appendix: Table 2a-3a). Phylogenetic relationships were then estimated using maximum likelihood and Bayesian methods, with DNA substitution models chosen for the partitioned dataset using the software PartitionFinder (Lanfear et al. 2012) (Appendix: Table 6a). A maximum likelihood tree (Appendix: Figure 2a) with bootstrap support values was generated using RAxML v.8 (Stamatakis 2014) as implemented in CIPRES Science Gateway v.3.3 (Miller et al. 2010).

Divergence dating analysis and topology were simultaneously conducted in MrBayes using a relaxed clock model and birth death rate prior for branch lengths. In order to utilise *Proechimys,* a mainland echimyid genus*,* as the outgroup in all trees, a single topology constraint was enforced at the root, and branches were temporally scaled using the node-age prior. The sampling strategy was set to random with fraction 0.385 (= 10/26), which assumes random sampling of extant and recently extinct species. Runs included four chains for 33,330,000 generations, and trees were sampled every 10,000 generations. A single age constraint at the root divergence between Capromyidae and *Proechimys* was used, corresponding to the 95% highest posterior density interval of 17.1-19.3 Ma (mean = 18.2 Ma). This molecular prior is based on an analysis utilising 22 fossil calibrations across the Ctenohystrica parent clade to Capromyidae (Upham & Patterson, 2015). This analysis was also implemented in CIPRES Science Gateway v.3.3.

The mitochondrial gene cytB was used for a genetic pairwise distance analysis for comparison between individuals sampled by this and previous studies, with pairwise distances calculated as Kimura two-parameter distances in the programme MEGA v.5 (Tamura, Dudley, Nei, & Kumar, 2007) (Appendix: Tables 4a-5a).

6.5 Results

This study was able to extract degraded DNA successfully from 10 *Capromys* specimens collected from known geographical locations. Single specimens of *Mesocapromys* from all 4 previously described extant or possibly extant species and a single *Mysateles* specimen from the mainland of Cuba were also successfully sampled for aDNA (for specimen details and additional sampling information see Methods Chapter 2. Appendix Table 1a). As expected from DNA extracted from specimens found in sub-tropical regions endogenous content was low – an average estimates endogenous content of 0.020 in screening results, of samples extracted for this study. Target

capture enrichment techniques were utilsed for a total of 14 samples. Although enrichment did not always increase the endogenous content of samples (average enrichment factor was calculated at a percentage decrease of -%50), the total number of reads mapped to the mitochondrial genome in samples subjected to target capture enrichment represented an increase at an average of 66% (Table 2). These data were then combined with published data in order to generate a complete phylogeny for Cuban hutias (Appendix: Tables 1a-3a).

Table 2: Sequencing results including pre-target capture enrichment screening results and post-capture results. Endogenous content and enrichment factor were estimated using mapping results from the whole mitochonrial genome reference seqeunce used.

Species	*Number of samples screened*	*Sample size used in the final analysis*	*Sample ID*	*Pre-Capture/shotgun*					
				Sequencing platform	*Number of paired reads*	*Average coverage*	*Total reads mapped*	*Total reads mapped (mtgenome)*	*Endogenous content (mtgenome)*
Mesocapromys nanus	4	1	RW16	MiSeq	3,125,784	1.14 (whole mtgenome), 0.5 (vwF), 0.52 (RAG1), 0.83 (irbp), 0.13 (apoB), 0.34 (GHR)	252 (whole mtgenome), 25 (vwF), 18 (RAG1), 49 (irbp), 7 (apoB), 11 (GHR)	252	0.008061977
Capromys pilorides ciprianoi	4	3	RW41	NextSeq	5,407,328	1.96 (whole mtgenome), 4.16 (vwF), 3.49 (RAG1), 5.1 (irbp), 3.54 (apoB), 4.627 (GHR)	1,885 (whole mtgenome), 280 (vwF), 219 (RAG1), 374 (irbp), 234 (apoB), 215 (GHR)	1,885	0.034860101
			RW37	NextSeq	3,125,784	3.64 (whole mtgenome), 3.15 (vwF), 7.9 (RAG1), 4.40 (irbp), 8.79 (apoB), 4.44 (GHR)	1,759 (whole mtgenome), 102 (vwF), 494 (RAG1), 299 (irbp), 246 (apoB), 145 (GHR)	1,759	0.056273882
			RW38	NextSeq	7,678,551	7.21 (whole mtgenome), 1.98 (vwF), 2.07 (RAG1), 1.69 (irbp), 3.68 (apoB), 2.73 (GHR)	3,007 (whole mtgenome), 123 (vwF), 120 (RAG1), 114 (irbp), 218 (apoB), 124 (GHR)	3,007	0.039161034

Table 2. (continued) Sequencing results including pre-target capture enrichment screening results and post-capture results. Endogenous content and enrichment factor were estimated using mapping results from the whole mitochonrial genome reference seqeunce used.

Species	*Number of samples screened*	*Sample size used in the final analysis*	*Sample ID*	*Pre-Capture/shotgun*					
				Sequencing platform	*Number of paired reads*	*Average coverage*	*Total reads mapped*	*Total reads mapped (mtgenome)*	*Endogenous content (mtgenome)*
Capromys pilorides relictus	3	2	RW36	NextSeq	6,901,225	1.33 (whole mtgenome), 3.05 (vwF), 9.5 (RAG1), 1.8 (irbp), 2.9 (apoB), 4.6 (GHR)	1,146 (whole mtgenome), 107 (vwF), 600 (RAG1), 130 (irbp), 146 (apoB), 193 (GHR)	1,146	0.016605748
			RW33	NextSeq	6,416,105	0.9 (whole mtgenome), 0.9 (vwF), 0.49 (RAG1), 0.9 (irbp), 0.36 (apoB), 0.93 (GHR)	766 (whole mtgenome), 60 (vwF), 29 (RAG1), 46 (irbp), 35 (apoB), 41 (GHR)	785	0.012234837
Mesocapromys angelcabrerai	4	1	RW04	NextSeq	2,959,694	0.68 (whole mtgenome), 0.50 (vwF), 0.59 (RAG1), 0.56 (irbp),0.25 (apoB), 0.33 (GHR)	206 (whole mtgenome), 28 (vwF), 26 (RAG1), 30 (irbp), 12 (apoB), 10 (GHR)	206	0.006960179
Capromys pilorides gundlachanius	4	2	RW25	MiSeq	2,896,931	0.78 (whole mtgenome), 0.71 (vwF), 0.47 (RAG1), 0.69 (irbp), 0.35 (apoB), 0.57 (GHR)	295 (whole mtgenome), 47 (vwF), 29 (RAG1), 47 (irbp), 21 (apoB), 23 (GHR)	295	0.01018319
			RW26	MiSeq	3,372,096	0.02 (whole mtgenome), 0.67 (vwF), 0.22 (RAG1), 0.57 (irbp), 0.03 (apoB), 0.34 (GHR)	21 (whole mtgenome), 16 (vwF), 9 (RAG1), 35 (irbp), 2 (apoB), 12 (GHR)	21	0.000622758

Table 2. (continued) Sequencing results including pre-target capture enrichment screening results and post-capture results. Endogenous content and enrichment factor were estimated using mapping results from the whole mitochonrial genome reference seqeunce used.

Species	*Number of samples screened*	*Sample size used in the final analysis*	*Sample ID*	*Pre-Capture/shotgun*					
				Sequencing platform	*Number of paired reads*	*Average coverage*	*Total reads mapped*	*Total reads mapped (mtgenome)*	*Endogenous content (mtgenome)*
Capromys pilorides pilorides	2	2	RW32	NextSeq	6,386,952	1.62 (whole mtgenome), 0.79 (vwF), 0.68 (RAG1), 0.74 (irbp), 1.26 (apoB), 0.8 (GHR)	1,080 (whole mtgenome), 51 (vwF), 39 (RAG1), 48 (irbp), 80 (apoB), 38 (GHR)	1,080	0.016909474
			RW49	NextSeq	4,561,838	0.65 (whole mtgenome), 0.52 (vwF), 0.67 (RAG1), 0.4 (irbp), 0.63 (apoB), 0.55 (GHR)	413 (whole mtgenome), 34 (vwF), 41 (RAG1), 28 (irbp), 40 (apoB), 25 (GHR)	413	0.009053368
Capromys pilorides doceleguas	3	1	RW29	NextSeq	8,841,881	2.51 (whole mtgenome), 2.17 (vwF), 2.26 (RAG1), 1.92 (irbp), 2.8 (apoB), 2.5 (GHR)	2,130 (whole mtgenome), 135 (vwF), 141 (RAG1), 134 (irbp), 172 (apoB), 116 (GHR)	2,130	0.024089897
Mesocapromys/Mysateles melanurus	3	1	RW46	NextSeq	7,962,782	3.85 (whole mtgenome), 6.64 (vwF), 4.84 (RAG1), 3.63 (irbp), 8.35 (apoB), 5.19 (GHR)	3,248 (whole mtgenome), 420 (vwF), 303 (RAG1), 245 (irbp), 519 (apoB), 222 (GHR)	3,248	0.040789764
			RW10	NextSeq	3,537,618	1.14 (whole mtgenome), 0.65 (vwF), 6.88 (RAG1), 0.32 (irbp), 5.72 (apoB), 0.69 (GHR)	900 (whole mtgenome), 39 (vwF), 548 (RAG1), 20 (irbp), 320 (apoB), 30 (GHR)	900	0.025440847

Table 2. (continued) Sequencing results including pre-target capture enrichment screening results and post-capture results. Endogenous content and enrichment factor were estimated using mapping results from the whole mitochonrial genome reference seqeunce used.

Species	*Number of samples screened*	*Sample size used in the final analysis*	*Sample ID*	*Pre-Capture/shotgun*					
				Sequencing platform	*Number of paired reads*	*Average coverage*	*Total reads mapped*	*Total reads mapped (mtgenome)*	*Endogenous content (mtgenome)*
Mesocapromys sanfelipensis	1	1	RW52	NextSeq	4,328,791	0.20 (whole mtgenome), 0.48 (vwF), 0.47 (RAG1), 0.84 (irbp), 0.36 (apoB), 0.41 (GHR)	98 (whole mtgenome), 28 (vwF), 28 (RAG1), 54 (irbp), 21 (apoB), 17 (GHR)	98	0.002263912
Capromys pilorides pilorides (Cayman islands species)	3	1	RW58	NextSeq	3,903,375	1.48 (whole mtgenome), 0.23 (vwF), 0.66 (RAG1), 0.24 (irbp), 0.19 (apoB), 0.25 (GHR)	372 (whole mtgenome), 13 (vwF), 18 (RAG1), 15 (irbp), 8 (apoB), 5 (GHR)	372	0.009530214
Mesocapromys auritus	4	1	RW06	NextSeq	3,894,432	4.41 (whole mtgenome), 2.09 (vwF), 1.47 (RAG1), 2.35 (irbp), 1.22 (apoB), 2.33 (GHR)	1,385 (whole mtgenome), 17 (vwF), 14 (RAG1), 45 (irbp), 39 (apoB), 76 (GHR)	1,385	0.035563594
Mesocapromys prehensilis	2	1	RW20	NextSeq	4,117,670	3.97 (whole mtgenome), 0.42 (vwF), 0.18 (RAG1), 0.41 (irbp), 0.54 (apoB), 0.12 (GHR)	943 (whole mtgenome), 13 (vwF), 7 (RAG1), 26 (irbp), 14 (apoB), 7 (GHR)	943	0.022901301

Table 2. (continued) Sequencing results including pre-target capture enrichment screening results and post-capture results. Endogenous content and enrichment factor were estimated using mapping results from the whole mitochonrial genome reference seqeunce used.

Species	*Number of samples screened*	*Sample size used in the final analysis*	*Sample ID*	*Post-Capture*						*Enrichment factor*	*Percentage increase/decrease in total reads mapped*
				Sequencing platform	*Number of paired reads*	*Average coverage*	*Reads mapped*	*Total reads mapped (mtgenome)*	*Endogenous content (mtgenome)*		
Mesocapromys nanus	4	1	RW16	NextSeq	14,989,504	11.38 (whole mtgenome), 20.59 (vwF), 24.45 (RAG1), 14.85 (irbp), 19.72 (apoB), 19.19 (GHR)	4,202 (whole mtgenome), 782 (vwF), 796 (RAG1), 601 (irbp), 624 (apoB), 455 (GHR)	4,202	0.028032949	248%	94%
Capromys pilorides ciprianoi	4	3	RW41	NextSeq	21,201,540	2.64 (whole mtgenome), 8.47 (vwF), 7.15 (RAG1), 10.79 (irbp), 7.2 (apoB), 9.57 (GHR)	2,369 (whole mtgenome), 546 (vwF), 439 (RAG1), 766 (irbp), 477 (apoB), 441 (GHR)	2,369	0.011173717	-212%	20%
			RW37	NextSeq	18,000,223	8.37 (whole mtgenome), 10.27 (vwF), 11.59 (RAG1), 9.8 (irbp), 16.17 (apoB), 13.58 (GHR)	6,223 (whole mtgenome), 645 (vwF), 699 (RAG1), 680 (irbp), 1,015 (apoB), 616 (GHR)	6,223	0.034571794	-63%	72%
			RW38	Na	Na	Na	na	na	Na	na	na

Table 2. (continued) Sequencing results including pre-target capture enrichment screening results and post-capture results. Endogenous content and enrichment factor were estimated using mapping results from the whole mitochonrial genome reference seqeunce used.

Species	Number of samples screened	Sample size used in the final analysis	Sample ID	Post-Capture						Enrichment factor	Percentage increase/decrease in total reads mapped
				Sequencing platform	Number of paired reads	Average coverage	Reads mapped	Total reads mapped (mtgenome)	Endogenous content (mtgenome)		
Capromys pilorides relictus	3	2	RW36	NextSeq	23,126,372	4.59 (whole mtgenome), 4.77 (vwF), 16.87 (RAG1), 2.88 (irbp), 17.35 (apoB), 5.08 (GHR)	4,356 (whole mtgenome), 315 (vwF), 1,019 (RAG1), 209 (irbp), 1,079 (apoB), 236 (GHR)	4,356	0.018835639	12%	74%
			RW33	NextSeq	14,593,714	1.45 (whole mtgenome), 1.21 (vwF), 7.6 (RAG1), 1.43 (irbp), 5.16 (apoB), 1.36 (GHR)	825 (whole mtgenome), 81 (vwF), 466 (RAG1), 103 (irbp), 332 (apoB), 63 (GHR)	825	0.005653119	-116%	5%
Mesocapromys angelcabrerai	4	1	RW04	NextSeq	24,839,391	2.18 (whole mtgenome), 1.98 (vwF), 2.06 (RAG1), 1.69 (irbp), 3.67 (apoB), 2.72 (GHR)	2,338 (whole mtgenome),123 (vwF), 120 (RAG1), 114 (irbp), 218 (apoB), 124 (GHR)	2,338	0.009412469	26%	91%
Capromys pilorides gundlachanius	4	2	RW25	NextSeq	19,487,474	0.96 (whole mtgenome), 3.05 (vwF), 2.38 (RAG1), 2.37 (irbp), 3.12 (apoB), 2.9 (GHR)	695 (whole mtgenome), 185 (vwF), 134 (RAG1), 162 (irbp), 182 (apoB), 118 (GHR)	605	0.003104558	-228%	51%
			RW26	NextSeq	14,739,301	12.51 (whole mtgenome), 29.01 (vwF), 34.61 (RAG1), 32.21 (irbp), 13.2 (apoB), 24.9 (GHR)	11,408 (whole mtgenome), 1,715 (vwF), 1,880 (RAG1), 1,984 (irbp), 1,002 (apoB), 1,028 (GHR)	11,408	0.077398514	99%	100%

Table 2. (continued) Sequencing results including pre-target capture enrichment screening results and post-capture results. Endogenous content and enrichment factor were estimated using mapping results from the whole mitochonrial genome reference seqeunce used.

Species	*Number of samples screened*	*Sample size used in the final analysis*	*Sample ID*	*Post-Capture*						*Enrichment factor*	*Percentage increase/decrease in total reads mapped*
				Sequencing platform	*Number of paired reads*	*Average coverage*	*Reads mapped*	*Total reads mapped (mtgenome)*	*Endogenous content (mtgenome)*		
Capromys pilorides pilorides	2	2	RW32	NextSeq	35,108,115	4.03 (whole mtgenome), 3.61 (vwF), 3.74 (RAG1), 2.96 (irbp), 7.64 (apoB), 4.16 (GHR)	3,267 (whole mtgenome), 227 (vwF), 215 (RAG1), 192 (irbp), 445 (apoB), 172 (GHR)	3,267	0.009305541	-82%	67%
			RW49	NextSeq	46,970,522	4.7 (whole mtgenome), 21.22 (vwF), 14.75 (RAG1), 18.19 (irbp), 6.5 (apoB), 11.14 (GHR)	3,256 (whole mtgenome), 1,354 (vwF), 883 (RAG1), 1,252 (irbp), 414 (apoB), 510 (GHR)	3,256	0.006932007	-31%	87%
Capromys pilorides doceleguas	3	1	RW29	NextSeq	44,100,116	8.94 (whole mtgenome), 9.14 (vwF), 12.15 (RAG1), 8.68 (irbp), 19.85 (apoB), 14.51 (GHR)	7,320 (whole mtgenome), 598 (vwF), 725 (RAG1), 609 (irbp), 1,225 (apoB), 655 (GHR)	7,320	0.016598596	-45%	71%
Mesocapromys/ Mysateles melanurus	3	1	RW46	Na	Na	Na	Na	na	na	na	na
			RW10	Na	Na	Na	Na	na	na	na	na

Table 2. (continued) Sequencing results including pre-target capture enrichment screening results and post-capture results. Endogenous content and enrichment factor were estimated using mapping results from the whole mitochonrial genome reference seqeunce used.

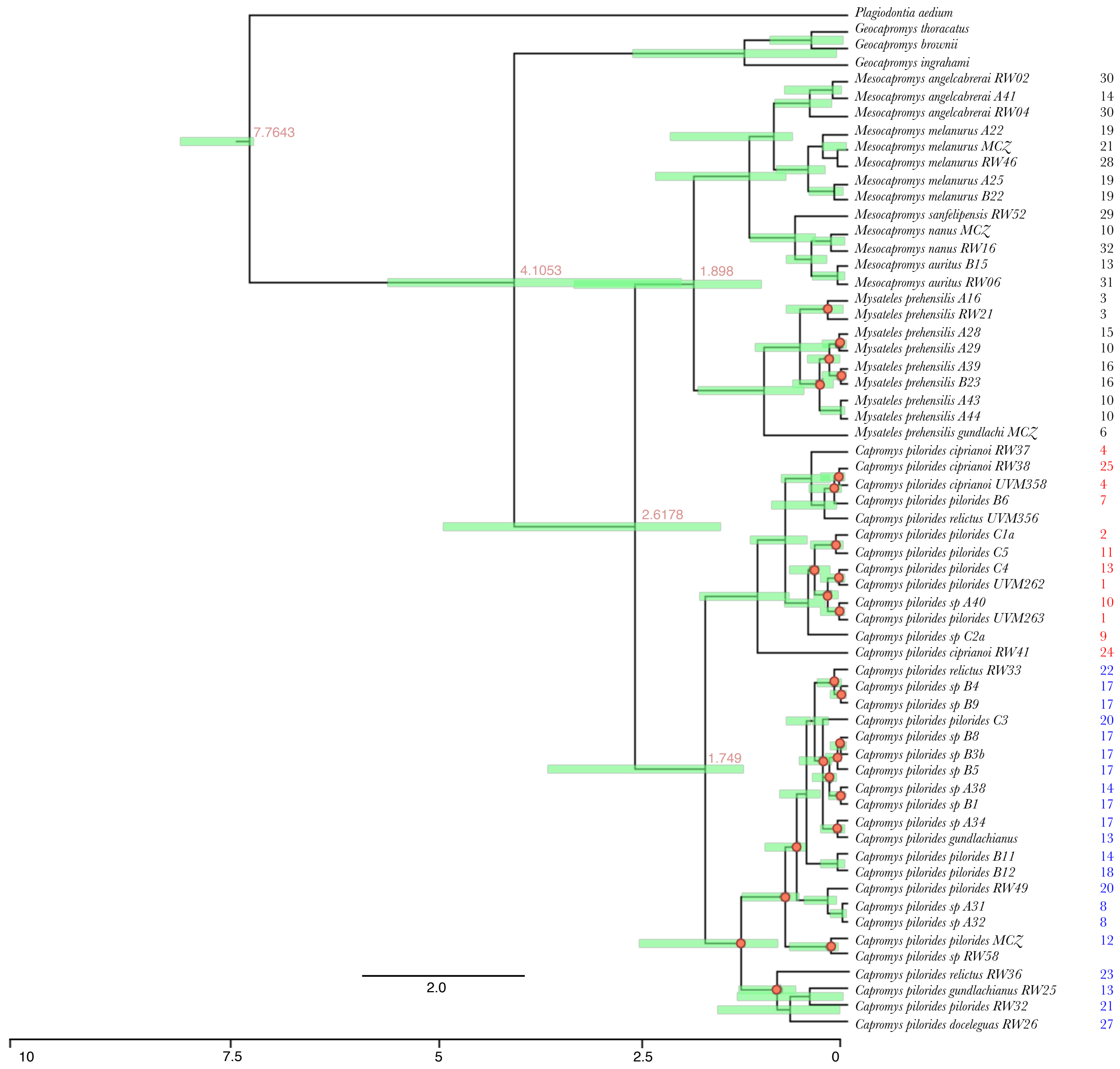
Species	*Number of samples screened*	*Sample size used in the final analysis*	*Sample ID*	*Post-Capture*						*Enrichment factor*	*Percentage increase/decrease in total reads mapped*
				Sequencing platform	*Number of paired reads*	*Average coverage*	*Reads mapped*	*Total reads mapped (mtgenome)*	*Endogenous content (mtgenome)*		
Mesocapromys sanfelipensis	1	1	RW52	NextSeq	44,153,937	0.87 (whole mtgenome), 3.89 (vwF), 3.31 (RAG1), 4.77 (irbp), 2.99 (apoB), 3.7 (GHR)	618 (whole mtgenome), 239 (vwF), 196 (RAG1), 306 (irbp), 166 (apoB), 164 (GHR)	618	0.001399649	-62%	84%
Capromys pilorides pilorides (Cayman islands species)	3	1	RW58	NextSeq	4,305,221	10.25 (whole mtgenome), 1.28 (vwF), 1.31 (RAG1), 0.75 (irbp), 2.37 (apoB), 0.9 (GHR)	1,980 (whole mtgenome), 15 (vwF), 13 (RAG1), 10 (irbp), 28 (apoB), 7 (GHR)	1,980	0.04599067	79%	81%
Mesocapromys auritus	4	1	RW06	Na	Na	Na	Na	na	na	na	na
Mesocapromys prehensilis	2	1	RW20	NextSeq	24,564,496	4.41 (whole mtgenome), 2.09 (vwF), 1.47 (RAG1), 2.35 (irbp), 1.22 (apoB), 2.33 (GHR)	1,302 (whole mtgenome), 120 (vwF), 776 (RAG1), 153 (irbp), 69 (apoB), 87 (GHR)	1,302	0.005300333	-332%	28%

Bayesian and maximum likelihood phylogenetic analysis produced congruent topology (Figure 2 and Appendix: Figure 2a). A visualization of sample locations in relation to phylogenetic placement is included here in the form of a map (Figure 3). The combined dataset mitochondrial and nuclear gene tree support a monophyletic Capromyidae clade. Both maximum likelihood and Bayesian analysis support reciprocal monophyly in five capromyid genera (*Capromys, Geocapromys, Mesocapromys, Mysateles* and *Plagiodontia*). In both analyses the genus *Capromys* is subdivided into two clades.

M. melanurus falls within *Mesocapromys*, confirming the findings of previous molecular studies (Upham & Borroto-Páez, 2017; Woods & Sergile, 2001). Within *Mesocapromys*, whole mitochondrial genome analysis recovers a different topology to previous study (Upham & Borroto-Páez, 2017): the genus is split into two clades, and with *Mysateles* as the sister group to all *Mesocapromys* species. *M. nanus* is sister to *M. auritus* as recovered by Upham & Borroto-Páez, (2017) but *M. angelcabrerai* and *M. melanurus* are recovered within a clade sister to a clade containing *M. auritus, M. nanus* and *M. sanfelipensis*.

In our analysis the crown radiation of Cuban hutias (*Capromys*, *Mesocapromys* and *Mysateles*) was dated to ~2.6 Ma (HPD 95% 1.511, 4.921), close to the estimated ~3.1 Ma divergence date recovered by Upham & Borroto-Páez (2017). We recovered the *Mysateles–Mesocapromys* split at ~1.8 Ma (1.038, 3.313). This analysis produced two clades within *Mesocapromys* with a divergence date of ~1.2 Ma (HPD 95% 0.726, 2.293). Two clades were also recovered within *Capromys* dated to ~1.7 Ma (HPD 95% 1.245, 3.572).

Figure 2. Divergence-dated Bayesian phylogeny constructed using MrBayes (Ronquist & Huelsenbeck 2003). Node values represent mean node age (HPD 95%) and bars indicate 95% intervals HPD. The scale indicates substitutions per site. Red nodes indicated low support (posterior probability < 0.9). This phylogeny includes data for Capromys, Mesocapromys and Mysateles species, subspecies and samples collected from individuals from known locations. For data produced as part of this study the whole mitochondrial genome and five nuclear genes are included (Table 2). The mainland echimyids Proechimys longcaudatus, Euryzygomatomys spinosus, Clyomys laticeps, Trinomys inheringi and Carterodon sulcidens were used as the outgroup.



Plagiodontia aedium
Geocapromys thoracatus
Geocapromys brownii
Geocapromys ingrahami
Mesocapromys angelcabrerai RW02 30
Mesocapromys angelcabrerai A41 14
Mesocapromys angelcabrerai RW04 30
Mesocapromys melanurus A22 19
Mesocapromys melanurus MCZ 21
Mesocapromys melanurus RW46 28
Mesocapromys melanurus A25 19
Mesocapromys melanurus B22 19
Mesocapromys sanfelipensis RW52 29
Mesocapromys nanus MCZ 10
Mesocapromys nanus RW16 32
Mesocapromys auritus B15 13
Mesocapromys auritus RW06 31
Mysateles prehensilis A16 3
Mysateles prehensilis RW21 3
Mysateles prehensilis A28 15
Mysateles prehensilis A29 10
Mysateles prehensilis A39 16
Mysateles prehensilis B23 16
Mysateles prehensilis A43 10
Mysateles prehensilis A44 10
Mysateles prehensilis gundlachi MCZ 6
Capromys pilorides ciprianoi RW37 4
Capromys pilorides ciprianoi RW38 25
Capromys pilorides ciprianoi UVM358 4
Capromys pilorides pilorides B6 7
Capromys pilorides relictus UVM356
Capromys pilorides pilorides C1a 2
Capromys pilorides pilorides C5 11
Capromys pilorides pilorides C4 13
Capromys pilorides pilorides UVM262 1
Capromys pilorides sp A40 10
Capromys pilorides pilorides UVM263 1
Capromys pilorides sp C2a 9
Capromys pilorides ciprianoi RW41 24
Capromys pilorides relictus RW33 22
Capromys pilorides sp B4 17
Capromys pilorides sp B9 17
Capromys pilorides pilorides C3 20
Capromys pilorides sp B8 17
Capromys pilorides sp B3b 17
Capromys pilorides sp B5 17
Capromys pilorides sp A38 14
Capromys pilorides sp B1 17
Capromys pilorides sp A34 17
Capromys pilorides gundlachianus 13
Capromys pilorides pilorides B11 14
Capromys pilorides pilorides B12 18
Capromys pilorides pilorides RW49 20
Capromys pilorides sp A31 8
Capromys pilorides sp A32 8
Capromys pilorides pilorides MCZ 12
Capromys pilorides sp RW58
Capromys pilorides relictus RW36 23
Capromys pilorides gundlachianus RW25 13
Capromys pilorides pilorides RW32 21
Capromys pilorides doceleguas RW26 27
7.7643
4.1053
1.898
2.6178
1.749
2.0
10
7.5
5
2.5
0

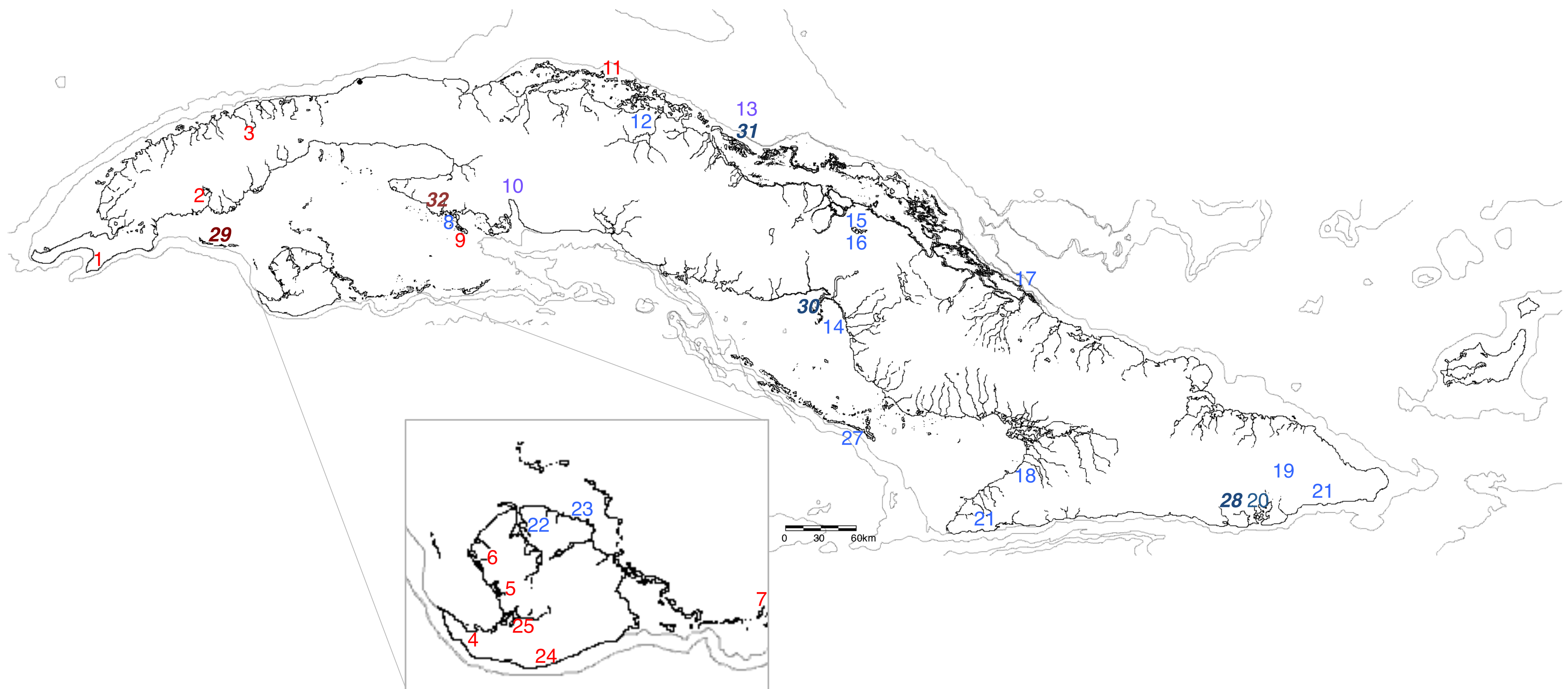


Figure 3. Map of Cuba showing Capromys and Mesocapromys sample locations included in Figure 1. (Appendix: Tables 6a-7a). Red indicates western clade and blue indicates eastern clades. Mesocapromys sample locations are indicated by bold italicised numbers.

6.5 Discussion and conclusions

Overview

To further assess the phylogeographic structure in Cuban hutias, we applied aDNA techniques to museum specimens in order to obtain whole mitochondrial genome data and five nuclear genes for *Capromys*, *Mesocapromys* and Mysateles individuals from known locations across Cuba. This analysis recovered two separate sister clades that include all extant and recently extinct Cuban hutias. Both clades show similar overall levels of genetic differentiation, with primary divergence at about 1.7-1.9 Ma. One clade is composed of large bodied hutias in the genus *Capromys*, the other is made up of Cuba's primarily small-bodied and obligate arboreal taxa.

Our analysis recovered two clades within *Capromys*, further confirming that genetic diversity with *Capromys* is better described by two discrete clades that may represent eastern and western populations, rather than multiple genetically isolated groups associated with offshore island locations. Our analysis differs to that of previous studies with respect to *Mesocapromys*, where we find two clades with *M. sanfelipensis, M. nanus* and *M. auritus* recovered in the first clade, and *M. melanurus* nested within the genus as sister to *M. angelcabrerai* placed in the second.

is the east-west split in Capromys supported?

Data collected as part of this study combined with data from previous studies provides a wider range of sampling locations for Cuban *Capromys* populations. In the phylogeny generated by this study a geographic east-west pattern can be detected, although some geographic mismatches are recovered. Whilst the two allopatric clades within *Capromys* are well-supported, however, many of the nodes for the topology recovered by this study and by Upham & Borroto-Páez (2017) have low support (posterior probability <0.9), including the nodes with the deepest divergence times within each of the *Capromys* clades. The low support for some nodes suggests that although there may be species or subspecies level divergence within prospective eastern and western clades, further studies including an increased number of samples and molecular markers are needed to assess species-level and population-level relationships within *Capromys*. This investigation would also benefit from statistical analysis such as the coalescent-based modeling approach (Wakeley 2009), which was utilised in testing demographic hypotheses in the Hispaniolan hutia *Plagiodontia* (Brace et al. 2012), in order to rule out other factors that may artificially produce this geographic pattern such as isolation by distance.

Do Mesocapromys and Capromys populations show the same phylogeographic structure?

One of the initial aims of this study was to determine whether different Cuban capromyid genera share parallel evolutionary relationships and phylogeographic structure. Our combined data phylogeny suggests that there may be a phylogeographic split between eastern and western *Capromys* populations, and that a similar trend can also be seen within *Mesocapromys*, for which two clades are also recovered. *M. nanus, M. auritus* and *M. sanfelipensis* are recovered in the same clade and are found on the western side of the putative line between eastern and western *Capromys* lineages (Matanzas, Cienfuegos, and Villa Clara). *M. auritus* is placed with *Mesocapromys* from western Cuba; in comparison, *Capromys* individuals found in the same locality (Cayo Fragoso) are recovered within both the proposed western and eastern clades. Upham & Borroto-Páez (2017) suggested that some *Capromys* individuals from Cayo Fragoso and Villa Clara may represent a relic population isolated from western populations.

The second *Mesocapromys* clade contains *M. angelcabrerai* and *M. melanurus*, both species found in the east of Cuba. *M. melanurus* was previously placed in *Mysateles,* and was associated with *Mysateles prehensilis* by morphology-based studies because of the taxon's adaptations to arboreal habitats. Not only does this study provide further support for the inclusion of *M. melanurus* within *Mesocapromys*, but our combined analysis recovers this species as sister to *M. angelcabrerai*, a species traditionally allied with *M. nanus* and *M. auritus,* because all three are small-bodied mangrove-associated taxa. This finding is further evidence that phylogenetic relationships among Cuban hutia populations and species reflect historical geographical relationships as opposed to morphological or ecological adaptations alone.

Insular species are known for convergent evolution and adaptive radiations (Losos & Ricklefs, 2009b). Our results have implications in this context, as this phylogenetic analysis demonstrates that individuals within one Cuban hutia clade taxa have adapted to multiple ecological niches, with a possible occurrence of convergence among arboreal prehensile tailed forms. Conversely in *Capromys*, where the divergence between taxonomic units is similar to *Mesocapromys*, taxa are morphologically conservative. This pattern in Cuban hutias is mirrored in other insular fauna where islands both preserve the ancestral species (e.g. Hispaniolan and Cuban *Solenodon*), but are also home to diverse and species rich radiations (e.g. Caribbean *Anoles* lizards), characterised by unusual adaptations and convergent evolution, corresponding to unique or empty ecological niches available.

Divergence dates

Capromys and *Mesocapromys + Mysateles* have similar estimated divergence dates for their respective radiations suggesting that the same mechanisms for the generation of divergence

affected both groups. Whilst eastern and western special spatial divergence patterns may be congruent with other Cuban vertebrate faunal divergence and distributions (e.g. Green *Anoles* (Glor et al. 2004), the Late Pleistocene divergence date observed in both clades is not comparable with the temporal divergence of these groups. However there is evidence of Late Pleistocene divergence in other Caribbean terrestrial vertebrate fauna including the Hispaniolan hutia (Brace et al. 2012) and *Nesophontes* (see Chapter 3). The Pleistocene was a period of turbulent climate and sea level changes, which within the insular environment may have provided cyclical partial barriers to gene flow and peripatric speciation.

If the wider divergence patterns across all Greater Antillean hutias, are considered in the context of the species-level differentiation observed across Cuban hutia, re-assessment of additional capromyid taxa may be warranted. Within *Geocapromys*, *G. thoracatus* and *G. brownii* (and even *G. columbinanus*, see Chapter 5) may not be valid species. Additionally, this mirrors the recent divergence between Cuban and Cayman Island *Capromys sp*. These patterns have a biogeographic implication and suggests that overwater dispersals southwards from the Greater Antilles may be more recent than previously suggested.

Taxonomic implications

The results of this study support recent molecular analyses that find two monophyletic mitochondrial lineages within Cuban *Capromys*. Combined analysis of all available individuals also supports the interpretation that these two lineages correspond with eastern and western geographic localities. Not only is this finding unexpected in light of previous morphology-based studies that suggested species relationships were connected to differences between mainland Cuban and offshore island populations, but it also has implications for the taxonomy and ultimately the conservation of Cuban *Capromys* populations, which are currently treated as a single species. The well supported ~1.7 Ma divergence estimated by our study suggests that, even with a conservative interpretation, Cuban *Capromys* should be split into two species corresponding to the two clades recovered in phylogenetic analyses. Upham and Borroto-Paez suggest that the western clade should retain *pilorides*, suggesting easterly distributed *doceleguas* and *gundlachianus* for the eastern clade. Individuals assigned to *C. p. relictus* are recovered in both clades, with the type specimen for this subspecies not included in our analysis the placement of this subspecies within the two clades remains unknown. Whilst *pilorides* is available for one of the two distinct clades the designation of a species epithet to the other clade will involve further sampling of museum type specimens, which necessary to clarify the taxonomy of this genus.

This study also examined the phylogenetic relationships between Cuban *Mesocapromys* and *Mysateles*. As the divergence between discrete clades within these groups share temporal distribution with the within genera divergence between eastern and western *Capromys* clades,

which has traditionally been considered a single genus, re-classification of *Mesocapromys* and *Mysateles* within the same genus is necessary. We suggest the senior synonym *Mysateles* (Poeppig 1824) should be applied to taxa within the *Mesocapromys* + *Mysateles* clade. This suggestion is supported by the placement of *M. melanurus*, a taxon that exhibits *Mysateles*-like morphology, as nested within a clade containing taxa previously referred to as *Mesocapromys*. The divergence between *Mysateles p. gundlachi* and other taxa in the *Mysateles* clade is high (~1 Ma) suggesting that this subspecies may represent a distinct species, however this species is represented by a single individual and only by three mitochondrial genetic markers and no nuclear DNA, therefore we conservatively retain the subspecies level distinction for this taxon subject to further sampling.

Traditionally taxa within *Mesocapromys* have been treated as discrete species. The morphological and ecological variation within this group supports the traditional view of the clade containing more than one species. As this study recovers the arboral, prehensilie tailed *M. melanurus* nested within the *Mesocapromys* clade sister to the morphologically dis-similar *M. angelcabrerai*, the divergence between these species can be used to calibrate the level of diveregnce between other taxa, for the purpose of clarifying the taxonomy of the group. Within the *M. sanfelipensis* and *M. auritus* + *M. nanus* clade the ~0.4 Ma divergence date recovered in this study between *M. auritus* and *M. nanus* is low, and is comparable to within-species divergences in *M. melanurus*, *M. prehensilis* and between *Capromys* populations. Pairwise distance analysis also suggest that the genetic distance between *M. auritus* and *M. nanus* is less than species level (<2%). The divergences between *M. sanfelipensis* and *M. auritus* + *M. nanus* (~0.6 Ma) is higher than the genetic distance between *M. auritus* and *M. nanus*, but still falls within the varation of within species diveregnce for other taxa. Due to the low level of divergence between *M. auritus* and *M. nanus* we suggest the senior synonym *nanus* (Allen 1917) be aplied to both taxa and possibly including *M. sanfelipensis*. Due to the conservation implications of this finding, and the fact that *M. sanfelipensis* is represented by a single individual in this analysis, further sampling is needed before the taxonomic status of *M. sanfelipensis* can be fully resolved.

Conservation implications

The results of this study suggest current taxonomic designations for Cuban hutia species need to be re-assessed, which has important implications for conservation priorities. Previously, *M. nanus* has been considered critically endangered and is most likely already extinct, whilst *M. auritus* is considered endangered. This study suggests species level designation of these taxa is invalid and calls for the amalgamation of these taxa to a single species (*Mysateles nanus*). In light of this taxonomic emendation and recent proposals for IUCN red list status for these taxa, a revision of the IUCN red list status to endangered for this taxa inclusive of *M. auritus*, seems appropriate. However *M. nanus* should still remain a priority for conservation, due to the fragmentation of existing

populations and the continued decline of this species preferred mangrove habitat. Practical efforts should now be focused on the protection surviving populations on Cayo Fragoso.

M. sanfelipensis is also considered critically endangered and possibly extinct. This study suggests that the level of genetic diveregence between *M. sanfelipensis* and *M. nanus* + *M. auritus* may also be below species level. We conservatively retain the species level distinction for this taxon until further samples can be included in this analysis, as such *M. sanfelipensis* retains the status of critically endangered and possibly extinct.

Final conclusions

Inter-island evolutionary relationships within Cuban hutias were analysed as part of this study, and phylogenetic results were complimentary to previous recent genetic analysis. The combined use of museum collections and aDNA techniques has allowed this study to include a more comprehensive dataset than previous studies including critically endangered and possibly extinct taxa, for which collection of specimens from living populations is not possible. The use of whole mitochondrial genomes and nuclear genes has allowed for better resolution and support for our phylogeny. The phylogenetic results of this study have important implications for taxonomy and conservation of Cuban hutias. Our analysis provides some evidence of parallel phylogeographic patterns found in the two Cuban capromyid genera *Capromys* and *Mesocapromys*, as opposed to traditional interpretation of species diversity and relationships. We also were able to demonstrate, for the first time using molecular data, examples of insular convergent evolution and adaptive radiations within Cuban hutias.

References

Allen, G.M., 1917. Fossil mammals from Cuba. *Bulletin of the Museum of comparitive Zoology*, 61, pp.1–12.

Arntzen, J.W., Wielstra, B. and Wallis, G.P., 2014. The modality of nine Triturus newt hybrid zones assessed with nuclear, mitochondrial and morphological data. *Biological Journal of the Linnean Society*, *113*(2), pp.604-622.

Arredondo, D.L.M., 2012. A traditional Cuban perspective on the taxonomy of Capromyid rodent fossils. In R. Borroto-páez, C. A. Woods, & F. E. Sergile, eds. *Terrestrial mammals of the West Indies*. Gainesville: Florida Museum of Natural History and Wacahoota Press, pp. 258–285.

Benson, D.A. et al., 2005. GenBank. *Nucleic Acids Research*, 33, pp.34–38.

Borroto-Páez, R., 2005. Sistematica de las jutias de las Antillas (Rodentia, Capromyidae). In P. Bover & J. A. Alcover, eds. *Proceedings of the International Symposium "Insular Vertebrate Evolution: the Palaeontological Approach."* Monografies de la Societat d'Historia Natural de les Balear, pp. 33-50

Borroto-Páez, R. Mancina, C. A., Woods, C. A., & Kilpatrick, C. W., 2012. Updated checklist of endemic terrestrial mammals of the West Indies. In R. Borroto-Páez, C. A. Woods, & F. E. Sergile, eds. *Terrestial mammals of the West Indies*. Florida Museum of Natural History, University of Florida, pp. 389–415.

Borroto-Páez, R. & Mancina, C.A., 2017. Biodiversity and conservation of Cuban mammals: past, present, and invasive species. *Journal of Mammalogy*, 98(4), pp.964–985.

Brace, S. et al., 2012. Population history of the Hispaniolan hutia *Plagiodontia aedium* (Rodentia: Capromyidae): Testing the model of ancient differentiation on a geotectonically complex Caribbean island. *Molecular Ecology*, 21(9), pp.2239–2253.

Carstensen, D.W. et al., 2013. The functional biogeography of species: Biogeographical species roles of birds in Wallacea and the West Indies. *Ecography*, 36(10), pp.1097–1105.

Cooke, S.B. et al., 2017. Anthropogenic Extinction Dominates Holocene Declines of West Indian Mammals. *Annual Review of Ecology, Evolution, and Systematics*, 48(1).

Coyne, J.A. & Orr, H.A., 2004. *Speciation*, Sunderland, MA, USA: Sinauer Associated.

Dabney, J. et al., 2013. Complete mitochondrail genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences*, 110(39), pp.15758–15763.

Davalos, L.M. & Turvey, S.T., 2012. West Indian Mammals: The old the new and the recently extinct. In *Bones, Clones, and Biomes: The History and Geography of Recent Neotropical Mammals*. pp. 157–202.

Diaz-Franco, S., 2001. Taxonomic situation of *Geocapromys megas* (Rodentia : Capromyidae). *Caribbean Journal of Science*, 37(1–2), pp.72–80.

Enk, J.M. et al., 2014. Ancient whole genome enrichment using baits built from modern DNA. *Molecular Biology and Evolution*, 31(5), pp.1292–1294.

Excoffier, L., Foll, M. and Petit, R.J., 2009. Genetic consequences of range expansions. *Annual Review of Ecology, Evolution, and Systematics*, *40*, pp.481-501.

Fabre, P.-H. et al., 2014. Rodents of the Caribbean: origin and diversification of hutias unravelled by next-generation museomics. *Biology Letters*, 10(7).

Fabre, P.H. et al., 2017. Mitogenomic phylogeny, diversification, and biogeography of South American spiny rats. *Molecular Biology and Evolution*, 34(3), pp.613–633.

Gascoyne, M., Benjamin, G.J. & Schwarz, H.P., 1979. Sea-level lowering during the Illinoian glaciation: evidence from a Bahama “blue hole.” *Science*, 205(806), p.8.

Glor, R.E. et al., 2004. Partial island submergence and speciation in an adaptive radiation: a multilocus analysis of the Cuban green anoles. *Proceedings. Biological sciences / The Royal Society*, 271(1554), pp.2257–2265.

Kearse, M. et al., 2012. Geneious Basic: an integrated and extendable desktop software platform for thr organization and analysis of sequence data. *Bioinformatics*, 28(12), pp.1647–1649.

Knell, R.J., Naish, D., Tomkins, J. L., & Hone, D. W., 2013. Sexual selection in prehistoric animals: detection and implications. *Trends in Ecology & Evolution*, 28(1), pp.38–47.

Lanfear, R. et al., 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, 29, pp.1695–1701.

Larkin, M.A. et al., 2007. Clustal W and Clustal X. *Bioinformatics*, 23, pp.2947–2948.

Losos, J.B. & Glor, R.E., 2003. Phylogenetic comparative methods and geography of speciation. *Trends in Ecology & Evolution*, 18, pp.220–227.

Losos, J.B. & Ricklefs, R.E., 2009a. Adaptation and diversification on islands. *Nature*, 457(7231), pp.830–836.

Losos, J.B. & Ricklefs, R.E., 2009b. Adaptation and diversification on islands. *Nature*, 457(7231), pp.830–6.

McGuire, J. A., Linkem, C. W., Koo, M. S., Hutchison, D. W., Lappin, A. K., Orange, D. I., Lemos-Espinol, J., Riddle, B. R., Jaeger, J. R. (2007). Mitochondrial introgression and incomplete lineage sorting through space and time: phylogenetics of crotaphytid lizards. *Evolution* 61: 2879-2897

Meyer, M. & Kircher, M., 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*, 5(6).

Milishnikov, A.N., Borroto, R. & Orlov, V.N., 2010. Correlation between morphological and genetic differentiation in Cuban hutia (Rodentia, Capromyinae): species fragmentation on the area and the problem of excessive taxonomic status problem. *Genetika*, 46(2), pp.218–225.

Miller, M.A., Pfeiffer, W. & Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *Gateway Computing Enviroments Workshop (GCE)*, pp.1–8.

Milne, G., Long, A. & Bassett, S., 2005. Modelling Holocene relative sea-level observations from the Caribbean and South America. *Quaternary Science Reviews*, 24(10–11), pp.1183–1202.

Moritz, C., 1995. Uses of molecular phylogenies for conservation. *Phil. Trans. R. Soc. Lond. B*, 349(1327), pp.113–118.

Petit, R.J., Mousadik, A.E.L. & Pons, O., 1998. Identifying populations for conservation on the basis of genetic markers. *Conservation Biology*, 12(4), pp.844–855.

Poeppig, E., 1824. Nova generis Capromys, Desm. species. *Journal of the Academy of Natural Sciences of Philadelphia*, 4, pp.11–15.

Pregill, G.K. & Olson, S.L., 1981. Zoogeography of West Indian Vertebrates in Relation to Pleistocene Climatic Cycles. *Annual Review of Ecology and Systematics*, 12(1981), pp.75–98.

Ronquist, F. & Huelsenbeck, J.P., 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19, pp.1572–1574.

Rubinoff, D. & Holland, B., 2005. Between Two Extremes: Mitochondrial DNA is neither the Panacea nor the Nemesis of Phylogenetic and Taxonomic Inference. *Systematic Biology*, 54(6), pp.952–961.

Silva Taboada, G., Suarez Duque, W. & Diaz-Franco, S., 2007. Compendio de los mamiferos terrestres autoctonos de Cuba vivientes y extinguidos La Habana: Museo Nacional de Historia Natural.

Soy, J. & Silva, G., 2008. Mesocapromys nanus. *The IUCN Red List of Threatened Species*, (e.T13217A3420679).

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30, pp.1312–1313.

Tamura K Dudley, J., Nei, M., & Kumar, S., 2007. Molecular Rvolutionary Genetic Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, 24, pp.1597–1599.

Turvey, S.T., Kennerley, R. J., Nuñez-Miño, J. M., & Young, R. P., 2016. The Last Survivors: current status and conservation of the non-volant land mammals of the insular Caribbean. , *Journal of Mammalogy*, *98*(4), 918-936.

Upham, N.S. & Borroto-Páez, R., 2017. Molecular phylogeography of endangered Cuban hutias within the Caribbean radiation of capromyid rodents. *Journal of Mammalogy*, 98(4), pp.950–963.

Upham, N.S. & Patterson, B.D., 2015. Evolution of caviomorph rodents: a complete phylogeny and timetree for living genera. In A. I. Vassallo & D. Antenucci, eds. *Biology of caviomorph rodents: diversity and evolution*. Buenos Aires, Argentine, pp. 63–120.

Varona, L.S., 1970. Descripción de una nueva especie de *Capromys* del sur de Cuba (Rodentia: Caviomorpha). *Poeyana Serie A*, 74, pp.1–16.

Varona, L.S., 1979. Subgénero y especie nuevos de *Capromys* (Rodentia: Caviomorpha) para Cuba. *Poeyana*, 194, pp.1–33.

Varona, L.S. & Garrido, O.H., 1970. Vertebrados de los cayos de San Felipe, Cuba, incluyendo una nueva especie de jutía. *Poeyana*, 75, pp.1–26.

Wakeley, J., 2009. *Coalescent Theory: An Introduction*, Greenwood Village, Colorado: Roberts and Company Publishers.

Woods, C.A. & Sergile, F.E., 2001. *Biogeography of the West Indies: Patterns and Perspective, Second Edition: Past Present and Future*,

Wright, S., 1943. Isolation by Distance. *Genetics*, 28(2), pp.114–138.

Appendix Chapter 5. Phyleogographic analysis of sympatric Cuban hutia reveal parallel patterns evolution in a mammalian insular radiation

Table 1a. Sample information from this study including sample identification number and location data as well as genes used in analysis

Species/subspecies	Museum/Sample ID	Location	Genes + mtgenome	Map reference number
Capromys p. pilorides	RW049	Guantanamo Province, Jatales de Manati, Los Comos Estate	mitogenome GHR apoB irBP	20
Capromys p. pilorides	RW032	Santiago de Cuba Province, Gran Piedra, East Sierra Maestra	mitogenome GHR RAG1 apoB irBP	21
Capromys p. relictus	RW033	North Isla Juventud (Isle of Pines)	mitogenome GHR RAG1 apoB irBP	22
Capromys p. relictus	RW036	North Isla Juventud (Isle of Pines), Cayo los Monos	mitogenome GHR vwF RAG1 apoB irBP	23
Capromys p. ciprianoi	RW041	South Isla Juventud (Isle of Pines), Hato de Milian	mitogenome GHR vwF RAG1 apoB irBP	24
Capromys p. ciprianoi	RW037	South Isla Juventud (Isle of Pines)	mitogenome GHR vwF RAG1 apoB irBP	4

Table 1a. (continued) Sample infromation from this study including sample identification number and location data as well as genes used in analysis

Species/subspecies	Museum/ Sample ID	Location	Genes + mtgenome	Map reference number
Capromys p. ciprianoi	RW038	South Isla Juventud (Isle of Pines), Estero las Piedras	mitogenome GHR vwF RAG1 apoB irBP	**25**
Capromys p. gundlachianus	RW025	Cayo Fragoso	mitogenome GHR RAG1 apoB irBP	**13**
Capromys p. doceleguas	RW029	Jucaro, Golfo de Ana Maria	mitogenome GHR vwF RAG1 apoB irBP	26
Capromys p. doceleguas	RW026	Cayo Caguana, Cayos y Laberinto de las Doce Leguas	mitogenome GHR vwF RAG1 apoB irBP	**27**

Table 1a. (continued) Sample information from this study including sample identification number and location data as well as genes used in analysis

Species/subspecies	Museum/ Sample ID	Location	Genes + mtgenome	Map reference number
Mysateles melanurus	RW046	Santiago de Cuba Province, Gran Piedra	mitogenome GHR vwF RAG1 apoB irBP	**28**
Mesocapromys sanfelipensis	RW052	Cayo Juan Garcia	mitogenome GHR vwF	**29**
Mesocapromys angelcabrerai	RW004	Cayos de Ana Maria	mitogenome GHR vwF RAG1 apoB irBP	**30**
Mesocapromys angelcabrerai	RW002	Cayos de Ana Maria	mitogenome GHR vwF RAG1 apoB irBP	**30**
Mesocapromys auritus	RW006	Cayo Fragoso	mitogenome GHR vwF RAG1 apoB irBP	**31**
Mesocapromys nanus	RW016	Zapata Peninsula	mitogenome GHR vwF RAG1 apoB irBP	**32**

Table 2a. Sample information from (Upham & Borroto-Páez 2017) including sample identification number, location data, and genes used in analysis

Species/subspecies	Museum/Sample ID	Location	Gene	Map reference number
Capromys p. pilorides	UVM262	Pinar del Rio, La Bajada	CytB	1
Capromys p. pilorides	UVM263	Pinar del Rio, La Bajada	CytB	1
Capromys p. ciprianoi	UVM358	South Isla Juventud (Isle of Pines), Hato de Milian	CytB	4
Capromys p. pilorides	RBP-A40	Matanzas, Cienaga de Zapata	CytB	10
Capromys p. ?ssp. 1	RBP-B6	Isla de la Juventud, Cayo Cantiles	CytB COI	7
Capromys p. pilorides	RBP-C1a	Pinar del Rio, La Coloma	CytB COI 12s	2
Capromys p. ?ssp. ¾	RBP-C2a	Matanzas, Cayo Diego Perez	CytB COI	9
Capromys p. ?gundlachianus	RBP-C4	Cayo Fragoso	CytB COI 12s	13
Capromys p. ?gundlachianus	RBP-C5	Cayo La Yana	CytB COI 12s	11
Capromys p. ?gundlachianus	RBP-C4/C5	Cayo Fragoso	CytB 12s GHR	13
Capromys p. pilorides	MVZ191417	Mogotes Jumagua	CytB 12s GHR vwF RAG1	12
Capromys p. ssp ¾	RBP-A31	Matanzas, Cayo Sinverguenza	CytB COI 12s	8
Capromys p. ssp ¾	RBP-A32	Matanzas, Cayo Sinverguenza	CytB 12s GHR vwF RAG1	8
Capromys p. ssp. 2	RBP-A34	Camaguey, Cayo Ballanato del Medio	CytB COI	17
Capromys p. ?ssp	RBP-A38	Ciego de Avila, Cayo Salinas	CytB	14

Table 2a. (continued) Sample information from (Upham & Borroto-Páez 2017) including sample identification number, location data, and genes used in analysis

Species/subspecies	Museum/Sample ID	Location	Gene	Map reference number
Capromys p. ssp. 2	RBP-B1	Camaguey, Cayo Ballanato del Medio	CytB	**17**
Capromys p. ssp. 2	RBP-B3b	Camaguey, Cayo Ballanato del Medio	CytB COI	**17**
Capromys p. ssp. 2	RBP-B4	Camaguey, Cayo Ballanato del Medio	CytB COI 12s	**17**
Capromys p. ssp. 2	RBP-B5	Camaguey, Cayo Ballanato del Medio	CytB COI	**17**
Capromys p. ssp. 2	RBP-B8	Camaguey, Cayo Ballanato del Medio	CytB COI 12s	**17**
Capromys p. ssp. 2	RBP-B9	Camaguey, Cayo Ballanato del Medio	CytB COI 12s	**17**
Capromys p. ?ssp	RBP-B11	Ciego de Avila, Cayo Salinas	CytB	**14**
Capromys p. pilorides	RBP-B12	Manzanillo	CytB	**18**
Capromys p. pilorides	RBP-C3	Guantanamo	CytB 12s	**20**

Table 3a. Sample information from (Upham & Borroto-Páez 2017) including sample identification number and location data as well as genes used in analysis

Species/subspecies	Museum/Sample ID	Location	Genes + mtgenome	Map reference number
Mesocapromys angelcabrerai	RBP-A41	Ciego de Avila, Cayo Salinas	cytB COI 12s GHR vwF RAG1	**14**
Mesocapromys auritus	RBP-B15	Cayo Fragoso	cytB COI 12s GHR vwF RAG1	**13**
Mesocapromys auritus	RBP-B18	Cayo Fragoso	cytB	**13**

Table 3a. (continued) Sample information from (Upham & Borroto-Páez 2017) including sample identification number and location data as well as genes used in analysis

Species/subspecies	Museum/Sample ID	Location	Genes + mtgenome	Map reference number
Mesocapromys melanurus	RBP-A22	Guantanamo, Monte Verde	cytB COI 12s	**19**
Mesocapromys melanurus	RBP-A25	Guantanamo, Monte Verde	cytB COI 12s	**19**
Mesocapromys melanurus	RBP-B22	Guantanamo, Monte Verde	cytB COI 12s	**19**
Mesocapromys melanurus	MCZ34406	Guantanamo, Imias	cytB COI 12s	**21**
Mesocapromys nanus	MCZ17730	Matanzas, Cienaga de Zapata	cytB COI 12s	**10**
Mysateles prehensilis	RBP-A16	Pinar del Rio, San Diego	cytB 12s	**3**
Mysateles prehensilis	RBP-A28	Ciego de Avila, Finca San Agustin	cytB	**15**
Mysateles prehensilis	RBP-A29	Matanzas, Cienaga de Zapata	cytB	**10**
Mysateles prehensilis	RBP-A39	Ciego de Avila, El Venero	cytB	**16**
Mysateles prehensilis	RBP-A43	Matanzas, Cienaga de Zapata	cytB	**10**
Mysateles prehensilis	RBP-A44	Matanzas, Cienaga de Zapata	cytB	**10**
Mysateles prehensilis	RBP-B23	Ciego de Avila, El Venero	cytB COI 12s GHR vwF RAG1	**16**
Mysateles prehensilis	MCZ17090	Isla de la Juventud, Estero de Soldado	cytB COI 12s	**6**

Table 4a. Estimates of evolutionary divergence between cytB sequence data from Mesocapromys and Mysateles individuals from different localities. The number of base differences per site from between sequences are shown. All ambiguous positions were removed for each sequence pair. There were a total of 798 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Tamura et al., 2007)

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	Mesocapromys melanurus RW46															
2	Mesocapromys melanurus A22	0.009														
3	Mesocapromys melanurus A25	0.015	0.006													
4	Mesocapromys melanurus B22	0.014	0.008	0.002												
5	Mesocapromys melanurus MCZ	0.000	0.009	0.015	0.014											
6	Mesocapromys angelcabrerai RW02	0.030	0.023	0.023	0.021	0.030										
7	Mesocapromys angelcabrerai RW04	0.020	0.023	0.029	0.026	0.020	0.014									
8	Mesocapromys angelcabrerai A41	0.031	0.026	0.026	0.023	0.031	0.000	0.016								
9	Mesocapromys sanfelipensis RW52	0.042	0.038	0.044	0.043	0.037	0.056	0.057	0.050							
10	Mesocapromys nanus MCZ	0.024	0.018	0.015	0.014	0.024	0.015	0.026	0.015	0.045						
11	Mesocapromys nanus RW16	0.023	0.018	0.018	0.017	0.022	0.021	0.024	0.026	0.049	0.011					
12	Mesocapromys auritus RW06	0.022	0.024	0.024	0.022	0.022	0.026	0.029	0.022	0.047	0.017	0.028				
13	Mesocapromys auritus B15	0.024	0.018	0.015	0.014	0.024	0.015	0.026	0.015	0.045	0.005	0.017	0.013			
14	Mysateles prehensilis B23	0.068	0.063	0.060	0.057	0.068	0.063	0.072	0.064	0.082	0.060	0.069	0.052	0.060		
15	Mysateles prehensilis gundlachi MCZ	0.070	0.065	0.062	0.060	0.070	0.063	0.072	0.064	0.078	0.060	0.069	0.052	0.060	0.005	
16	Mysateles prehensilis RW22	0.065	0.059	0.056	0.053	0.065	0.060	0.071	0.063	0.079	0.058	0.065	0.052	0.058	0.000	0.002

Table 5a Estimates of evolutionary divergence between cytB sequence data from Capromys inderviduals from different localities. The number of base differences per site from between sequences are shown. All ambiguous positions were removed for each sequence pair. There were a total of 798 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Tamura et al., 2007)

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	RW37																						
2	RW38	0.004																					
3	RW33	0.040	0.041																				
4	RW41	0.030	0.030	0.030																			
5	RW36	0.042	0.040	0.030	0.024																		
6	RW49	0.049	0.049	0.024	0.025	0.041																	
7	RW25	0.040	0.043	0.048	0.048	0.041	0.068																
8	RW32	0.105	0.105	0.085	0.105	0.082	0.094	0.054															
9	RW26	0.100	0.100	0.091	0.087	0.084	0.119	0.067	0.086														
10	RW58	0.025	0.026	0.016	0.019	0.027	0.018	0.048	0.090	0.092													
11	MCZ	0.027	0.028	0.017	0.017	0.027	0.016	0.048	0.088	0.090	0.001												
12	B4	0.030	0.031	0.014	0.024	0.027	0.019	0.049	0.097	0.098	0.003	0.004											
13	B9	0.030	0.031	0.014	0.024	0.027	0.019	0.049	0.097	0.098	0.003	0.004	0.000										
14	A31	0.023	0.023	0.019	0.022	0.029	0.020	0.047	0.103	0.096	0.003	0.003	0.003	0.003									
15	B8	0.030	0.031	0.014	0.024	0.027	0.019	0.049	0.097	0.098	0.003	0.004	0.000	0.000	0.003								
16	C1a	0.014	0.012	0.039	0.034	0.032	0.045	0.039	0.109	0.098	0.024	0.026	0.027	0.027	0.023	0.027							
17	C4	0.015	0.014	0.041	0.035	0.033	0.047	0.036	0.111	0.100	0.026	0.027	0.028	0.028	0.024	0.028	0.001						
18	C5	0.014	0.012	0.039	0.034	0.032	0.045	0.039	0.109	0.098	0.024	0.026	0.027	0.027	0.023	0.027	0.000	0.001					
19	A34	0.023	0.023	0.019	0.022	0.029	0.020	0.047	0.103	0.096	0.003	0.003	0.003	0.003	0.000	0.003	0.023	0.024	0.023				
20	B3b	0.030	0.031	0.014	0.024	0.027	0.019	0.049	0.097	0.098	0.003	0.004	0.000	0.000	0.003	0.000	0.027	0.028	0.027	0.003			
21	B5	0.030	0.031	0.014	0.024	0.027	0.019	0.049	0.097	0.098	0.003	0.004	0.000	0.000	0.003	0.000	0.027	0.028	0.027	0.003	0.000		
22	B6	0.004	0.000	0.044	0.032	0.041	0.051	0.044	0.106	0.102	0.028	0.030	0.031	0.031	0.023	0.031	0.012	0.014	0.012	0.023	0.031	0.031	
23	C2a	0.014	0.012	0.039	0.034	0.032	0.045	0.039	0.109	0.098	0.024	0.026	0.027	0.027	0.023	0.027	0.000	0.001	0.000	0.023	0.027	0.027	0.012

Table 6a. Substitution models chosen for the partitioned dataset

Subset	Best model	Subset Paritions
		Mitochonrial genome
1	GTR+G	rrnL, rrnS, trnC, trnD, trnF, trnH, trnI, trnK, trnL1, trnL2, trnM, trnN, trnP, trnQ, trnR, trnS2, trnV, trnW, trnY
2	GTR+G	atp6, atp8, cob, cox3, nad1, nad2, nad3, nad4a, nad4l, nad5, nad6
3	GTR+G	cox1, cox2, trnA, trnE, trnG, trnS1, trnT
		apoB (apolipoprotein B)
1	HKY+G	gene
		RAG1 (recombination activating gene 1)
1	K80+G	gene
		vWF (von Willebrand factor)
1	K80+G	gene
		IRBP (interphotoreceptor retinoid binding)
1	HKY+G	gene
		GHR (growth hormone receptor)
1	K80+G	gene

Table 7a Genes used to generate baits for capture enrichment

Gene	Length (bp)	NCBI accession	Species
apolipoprotein B (apoB)	1155	JX515310.1	*Proechimys quadruplicatus*
recombination activating gene 1 (RAG1)	1072	EU313333.1	*Proechimys simonsi*
von Willebrand factor (vWF)	1150	AJ849311.1	*Echimys didelphoides*
interphotoreceptor retinoid binding (IRBP)	1245	JF297775.1	*Phyllomys* sp.
growth hormone receptor (GHR)	798	JX515330.1	*Trinomys setosus*
mitochondrial genome	16816	HM544128.1	*Proechimys longicaudatus*

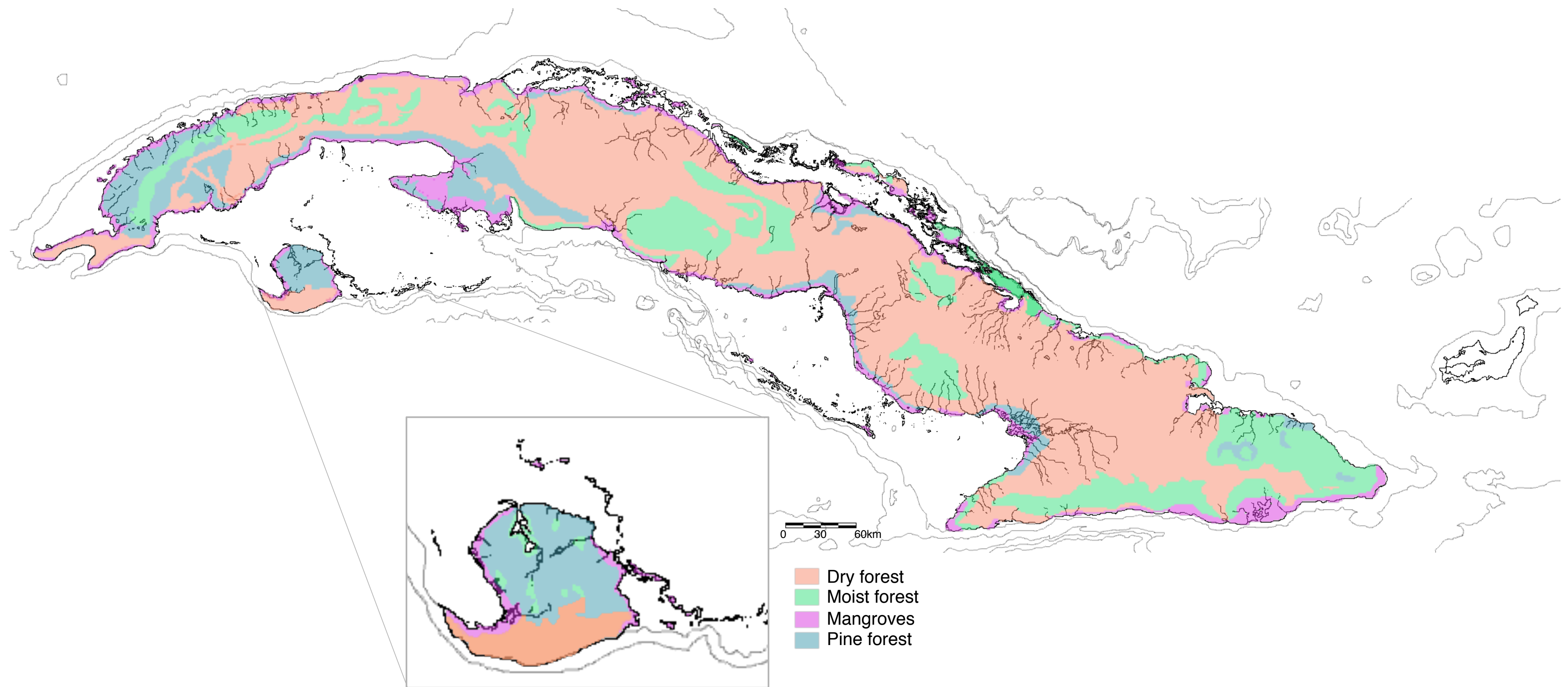


Figure 1a. Map of Cuba showing ecological regions adapted from WWF database http://assets.panda.org/img/original/hawksbill_caribbean_ecosystems.

Figure 2a. Maximum likelihood phylogeny produced using RAxML v.8 (Stamatakis 2014) which was implemented in CIPRES Science Gateway v.3.3 (Miller, Pfeiffer, & Schwartz, 2010).

Chapter 6. Evolutionary histories and biogeographic origins of the Greater Antillean rodent fauna

6.1 Abstract

The Caribbean islands are unique in their assemblage of diverse late Quaternary terrestrial mammal fauna. The most species rich and highly diverse group of non-volant Caribbean island mammals are the caviomorph rodents, traditionally interpreted as being represented by three families within the insular Caribbean: two endemic families, Capromyidae (hutias) and Heptaxodontidae ("giant hutias") and Heteropsomyinae (spiny rats)- an endemic subfamily within the South American family Echimyidae. Much of this insular diversity has been lost after an unprecedented archipelago-wide mammalian extinction event during the Holocene, leaving only two extant capromyid subfamilies. Extinct species hold the key for unraveling the evolutionary relationships and colonisation histories of Greater Antillean caviomorph rodents. Whilst previous studies have tried to investigate the Greater Antillean rodent taxa using morphology-based methods and molecular data from living taxa, investigators have been hampered by poorly preserved, fragmented remains of extinct taxa and few threatened surviving species. Despite recent molecular analyses of living capromyid taxa it is still unclear as to whether the Greater Antillean caviomorph rodents represent multiple colonisations of the Caribbean by evolutionarily distinct lineages, or alternately represent a single morphologically diverse monophyletic group. This study has used ancient DNA techniques to include extinct capromyids in the genera *Geocapromys* and *Capromys*, the heptaxodontid *Elasmodontomys*, and *Brotomys* and *Boromys*; two echmyid genera within the endemic subfamily Heteropsomyinae in a phylogenetic analysis carried out using partial mitochondrial genome and nuclear gene sequence data. We find that all of the extinct and living Caribbean caviomorph rodent species represent a single monophyletic clade within Echimyidae, which diverged from mainland sister taxa in the mid-Miocene (~12.1 Ma, 95% HPD 10.441–14.441 Ma).

6.2 Introduction

Caribbean island evolutions and extinctions

Islands have been highly important study systems for the investigation of questions surrounding evolution, colonisation histories and extinction dynamics of organisms (Losos & Ricklefs 2009b). The Caribbean in particular has played an important role in redefining theories on speciation (e.g. in *Anolis* lizards (Glor et al. 2003; Pinto et al. 2008; Losos & Queiroz 1997)). Whilst terrestrial mammals are rare on island systems, the Caribbean was successfully colonised by several groups including megalonychid sloths, platyrrhine monkeys, lipotyphlan insectivores and caviomorph rodents. This assemblage of unique but phylogenetically restricted Caribbean colonisers allows investigators the rare opportunity to study patterns of island evolution in insular mammals. However, due to archipelago-wide Holocene extinctions, the largest recent regional-scale mammalian extinction event (Siobhán B. Cooke et al. 2017), studies are restricted to scattered and fragmented sub-fossil materials and threatened extant species, hampering attempts to answer key questions about the living and extinct Caribbean taxa and providing a barrier to understanding mammalian evolution in insular systems. Molecular analysis, important in insular evolutionary studies for estimating divergence dates between taxa and establishing the mainland sister taxa of groups, is particularly impeded by the recent extinction of much of the Caribbean's Late Quaternary land mammal fauna, as DNA degradation is heightened in tropical localities (Hofreiter et al. 2015; Smith et al. 2001).

Greater Antillean caviomorph rodents

Rodents are well known for their high dispersal ability, and are some of the few terrestrial mammal colonisers of island systems (Gaines & McClenaghan Jr. 1980; Sax & Gaines 2008). The insular Caribbean is characterised by two geologically and biogeographically distinct sets of islands, the Lesser and Greater Antilles. Whilst the Lesser Antillean endemic rodent fauna is almost exclusively made up of oryzomyine rice rats, Greater Antillean rodent diversity is made up of several species groups of caviomorph rodents. Whilst the relationships between extant species of Greater Antillean caviomorph rodents have been investigated in recent molecular studies (Fabre et al., 2016; Fabre et al., 2014; Upham & Borroto-Páez, 2017), the relationships between recently extinct species, which includes entire families as well as informal species groups, remain unresolved. Greater Antillean caviomorph rodents are the most species rich group of non-volant mammals in the insular Caribbean, therefore investigation in to the evolutionary relationships and colonisation histories of this diverse

group is important for the clarification of evolutionary patterns across the Caribbean, and has wider implications for the study of other insular mammal faunas.

Wider colonisation histories and evolutionary relationships

One key area of Caribbean biogeographic research is the attempt to resolve colonisation histories of the major groups (Ali, 2012). Although colonisation events can be reconstructed using the fossil record to give approximate first occurrence dates, molecular data are often needed for a more accurate estimation of the colonisation histories of insular faunas (Donoghue & Benton 2007). This method has not yet been utilised for much of the Caribbean land mammal fauna, due to the poorly preserved nature of DNA taken from sub-fossil material or museum specimens from the tropics.

Two main theories have been suggested for the arrival of the Greater Antillean fauna: over-water dispersal via rafting (Hedges 2006; Alonso et al. 2012) and land-bridge hypotheses including the Greater Antilles Aves ridge hypothesis (GAAR-landia) (Iturralde-Vinent & MacPhee 1999). Recent studies have suggested overwater dispersal as the mechanism of colonisation for the living species within the endemic Caribbean rodent family Capromyidae, with evidence for a single colonisation event which gave rise to surviving taxa (Fabre et al., 2014). However there are two groups of recently extinct caviomorph rodents for which estimates of colonisation timings or wider evolutionary relationships have not been generated using molecular data. The echimyid subfamily Heteropsomyinae and family Heptaxodontidae both share similar Greater Antillean distributions with living capromyids and some species been linked to the Capromyidae by shared morphological features (Woods, 1982). Heteropsomyines share derived features with mainland echimyids (Carvalho & Salles 2004), but also share some cranial features with Caribbean capromyids (Allen 1918). Heptaxodontids are a possibly paraphyletic group and generally large in comparison to capromyids with distinctive tooth morphology, but have also been associated within Capromyidae by several authors (MacPhee, 2011; Ray, 1964, 1965; Woods & Kilpatrick, 2005). However previous morphology-based studies have not been able to satisfactorily clarify evolutionary relationships between all three groups, providing little evidence for shared colonisation histories.

Investigation of the well-studied Caribbean *Anolis* lizard radiation has shown that the diversity of this group in the Caribbean was formed by more than one colonisation of the islands, and also included back colonisation of the mainland (Nicholson et al. 2005). Lesser Antillean oryzomyine rodent diversity is also made up of two distinct colonisation events from South America (Brace, Turvey, Weksler, Hoogland, & Barnes, 2015). Molecular analysis of the extinct Jamaican monkey (*Xenothrix mcgregori*) provides evidence that the Greater Antillean

endemic primates also arrived in the Caribbean via more than one single colonisation event (see Chapter 4).

Woods & Sergile (2001) suggested that all Greater Antillean caviomorph rodents including living and extinct species formed a single monophyletic lineage, supporting a single ancestral colonisation of the Caribbean from South America. However other authors have suggested that some caviomorph rodent species and groups are more closely related to different mainland taxa suggesting multiple colonisation events (Kraglievich 1926; Landry 1957). The morphological distinction between Caribbean caviomorph rodent groups have led past studies to link various groups with different mainland rodent taxa. For example, endemic Caribbean heteropsomyines are currently classified as a subfamily within the mainland echimyid family because they share similar non-derived morphology to these taxa (Carvalho & Salles 2004). This morphological evidence suggests that endemic Caribbean heteropsomyines may represent a second, possibly more recent, colonisation event (McKenna & Bell 1997). Caribbean heptaxodontids have traditionally been considered as a discrete endemic family, and the distinctive cheek teeth and often larger body size of species in this group have led them to be associated with several mainland groups including two superfamilies Chinchilloidae (Anthony, 1917) and Octidontidae (Woods, 1993) as well as specifically within the family Echimyidae (MacPhee, 2011; Woods & Kilpatrick, 2005). The distinctive morphology in heptaxodontids, some of which is often described as primitive, suggests that this group may also represent a separate colonisation event from the capromyids (MacPhee 2011). Additionally the colonisation histories of other Caribbean vertebrate groups suggest that multiple colonisations of the Greater Antilles by caviomorph rodents may be a possible explanation for past species diversity and distributions.

Inter-island colonisation histories and evolutionary relationships

The production of molecular phylogenies is also key for understanding post-colonisation inter-island evolutionary histories in extinct and living caviomorph rodent taxa, where multiple species from different genera and even families are found on different islands. Living and recently extinct species within the family capromyidae are found across the Greater Antilles. Whilst recent studies have used molecular techniques to look at within-island relationships (Cuban species (Upham & Borroto-Páez 2017) and in the Hispaniolan hutia *Plagiodontia* (Brace et al., 2012)) little work has been done on the inter-island evolutionary relationships between living and extinct capromyid species. For entirely extinct groups such as the echimyid subfamily heteropsomyinaes, a similar molecular assessment of how species and genera on different islands are related is needed to establish insular patterns of evolution and colonisation histories. An understanding of inter-island evolutionary relationships is necessary for the investigation of how caviomorph rodents radiated across the Caribbean after initial

colonisation. The inclusion of extinct taxa with extant taxa in molecular phylogenies can improve resolution of divergence dates between species and genera, and help build a more complete picture of the evolution of insular Caribbean capromyid rodents. The current distribution and species diversity of living taxa could be the result of island hopping and intra-island speciation, or multiple re-colonisations of islands by different taxonomic groups (Fabre et al., 2014). The inclusion of extinct taxa may therefore allow this study to pinpoint the location of original colonisation and investigate the radiation of capromyids across the Greater Antilles.

The Greater Antilles consists of several large islands in the western Caribbean, including Cuba, Hispaniola (comprising the two countries of Haiti and the Dominican Republic), Jamaica, Puerto Rico, and smaller associated offshore islands. The Greater Antillean archipelago has a complex geological history that includes fragmentation of whole large islands (e.g. Cuban (Iturralde-Vinent & MacPhee 1999; Graham 2003) and Hispaniolan (Draper et al. 1994) palaeo-islands) and former sub-aerial connections between what are now distinct islands (e.g. between Hispaniola and Cuba, and Hispaniola and Puerto Rico (Iturralde-Vinent & Gahagan, 2002)). The interaction between the complex geology of the Greater Antillean islands and periods of Pleistocene sea level and climate change has been tied to the generation of Late Quaternary species diversity in several Caribbean vertebrate groups (Brace et al., 2015; Hedges & Conn, 2012; Hedges, Couloux, & Vidalj, 2009).

Whilst some molecular analyses have been conducted for living species (Fabre et al., 2014, 2016; Upham & Borroto-Páez, 2017), in extinct species the majority of studies have used morphology alone to establish evolutionary histories of taxa. The relationships between species within living and extinct Caribbean caviomorph groups currently recognised as different families remains unresolved, and for extinct species is unstudied at the molecular level. Molecular analysis of extinct species is therefore needed to test previous hypotheses of evolutionary relationships between these taxa, and for reconstruction of the caviomorph radiation across the Caribbean.

6.3 Capromyidae

Recent studies have reconstructed the evolutionary history of living species in the Greater Antillean rodent family Capromyidae (Fabre et al., 2014, 2016; Upham & Borroto-Páez, 2017). Whilst living capromyids can still be found on the Greater Antillean islands of Cuba, Jamaica, Hispaniola and the Bahamas, the archipelago-wide Holocene extinction event is thought to have removed 57.7% of the species diversity in this family, including considerable higher-order diversity, and also removed many capromyid genera and sometimes all species

from entire island groups (Borroto-Páez & Mancina 2017). This extinction event has hindered efforts to accurately investigate the evolutionary history of the group using morphology-based studies, and molecular data from extant species alone. This study includes four extinct capromyid species that represent considerable further diversity within the family: *Isolobodon portoricensis* and *Hexolobodon phenax* originally from Hispaniola, *Capromys* sp. from the Cayman Islands, *Geocapromys columbianus* from Cuba, and *Geocapromys thoracatus* from Little Swan Island.

Isolobodon and Hexolobodon

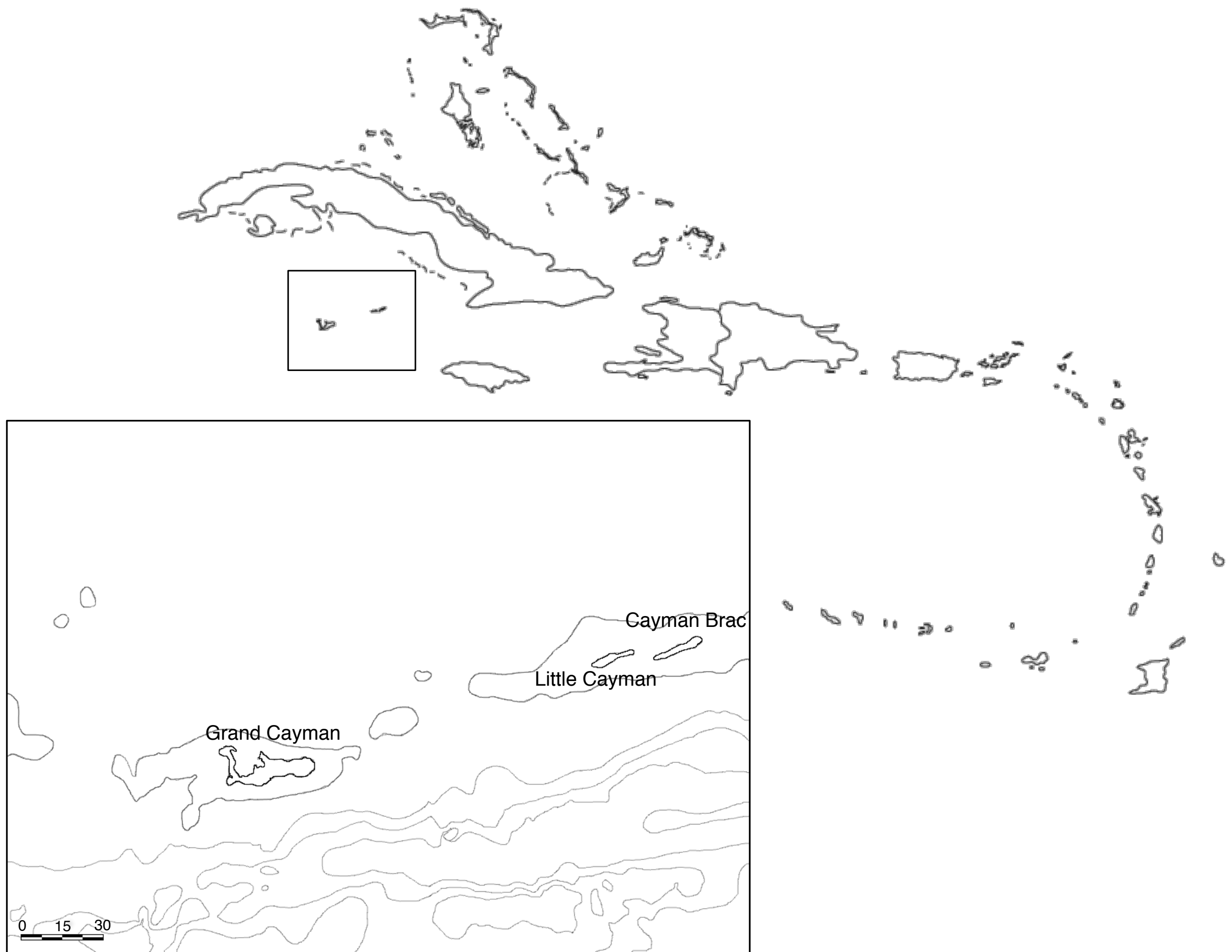
Isolobodon and *Hexolobodon* are extinct capromyid genera formerly endemic to Hispaniola, although individuals of *Isolobodon* were also transported to Puerto Rico and some of the Virgin Islands by early Amerindian colonizers of the Caribbean. The association of *Isolobodon* with Amerindian colonizers is made evident by the regular appearance of this taxon in archaeological midden sites (Woods, 1996). Authors have assigned two or alternatively three species to the genus: *I.* portoricensis, *I.* montanus, and *I.* levir; the last two species were originally placed in different monotypic genera by Miller (1922). *Hexolobodon* is a monotypic genus including the species *H. phenax* (Miller, 1930). Isolobodon can be distinguished from other capromyid genera by the derived features of cheek tooth cementum (Woods, 1989a; MacPhee and Iturralde-Vinent, 1995). As Fabre et al.'s (2014) study suggested a possible basal split for capromyids in Hispaniola, with the only surviving Hispaniolan hutia *Plagiodontia* as the outgroup for all other living hutia currently sampled, *Isolobodon* and *Hexalobodon* – two other independent genera from Hispaniola – could be vital in unraveling capromyid species relationships and colonization hypotheses. However, morphology-based studies have also linked *Isolobodon* to a possible crown capromyid, Cuban *Zazamys veronicae* (Macphee et al. 2003). *Zazamys* teeth were recovered from Cuban early Miocene deposits, making them the oldest caviomorph rodent fossils found in the Caribbean, and therefore suggesting that Cuba may be the origin of the capromyid radiation.

Capromys species from the Cayman Islands

The Cayman Islands are geographically situated to the south of the main island of Cuba on the opposite site of a deep-sea trench (the Cayman trench or trough). Morgan (1994) reported several undescribed extinct capromyid species from the Cayman Islands on the basis of subfossil remains. Morgan recovered remains that he assigned to two separate genera: *Capromys* and *Geocapromys*. These remains were found on two of the islands make up the Cayman Islands, Grand Cayman and Cayman Brac. This study will be the first to test these morphological taxonomic hypotheses using molecular data extracted from extinct hutia from Cayman Brac. Phylogenetic analysis including Cayman island *Capromys* species will allow

this study to estimate the date of divergence between Cayman island species and extant Cuban *Capromys* species, and thereby infer inter-island colonisation histories.

Figure 1. Map of the Caribbean region highlighting the location of the Cayman Islands, with a more detailed scaled map showing the three islands that make up the Cayman Islands group: Grand Cayman, Little Cayman and Cayman Brac.



Geocapromys columbianus and Geocapromys thoracatus

Two living species remain in the genus *Geocapromys*: *Geocapromys brownii*, found on Jamaica, and *Geocapromys ingrahami*, found on two islands in the Bahamas. Previous morphological studies and molecular data (Fabre *et al.,* 2014) suggest a sister relationship between *Geocapromys* and the genus *Capromys*. The remains of an extinct species, *Geocapromys columbianus*, have been excavated from Cuba (Allen 1918); this species is thought to have become extinct in the late 1500s when Europeans arrived in the Caribbean (Turvey & Helgen, 2008). *Geocapromys thoracatus* is another extinct species, formerly found on Little Swan Island. The island, which is very small in comparison with the Greater Antillean islands (2.4 km long), is geographically far removed from Jamaica, the Bahamas and Cuba where other living and extinct *Geocapromys* species are found. *G. thoracatus* became extinct recently, most likely the result of domestic cat introduction (Morgan 1989). Inclusion of *G. thoracatus* in divergence analysis will enable better understanding the colonisation history of Little Swan Island by *Geocapromys* and patterns of over water dispersal across the wider capromyid group. The generation of molecular data for both of these species can be used in phylogenetic analysis in order to give new insights into the region where the genus originated and radiated.

Figure 2. Map of the Caribbean region. The geographic position of Little Swan Island is indicated by the red dot.

This study will include extinct capromyid taxa in the genera *Capromys* and *Geocapromys* in order to investigate wider inter-island evolutionary relationships and colonisation histories within the endemic Caribbean capromyid radiation.

6.4 Heteropsomyinae

The Caribbean spiny rats within the family Echimyidae, subfamily Heteropsomyinae are known from Cuba (Allen 1918), Hispaniola (McFarlane, 2000) and Puerto Rico (Turvey, Grady, & Rye, 2006). Three distinct genera are recognised on the basis of morphological differences between forms found on different islands: *Boromys* (Miller, 1916) formerly found on Cuba and the off-shore Isla de Juventud, *Brotomys* (Miller, 1916) formerly found on Hispaniola and the off-shore Ile de la Gonave, and *Heteropsomys* (Anthony, 1916) formerly found on Puerto Rico. All of the Caribbean spiny rat species are extinct, most likely disappearing during the 16th century when Europeans first arrived in the Caribbean (Woods & Sergile, 2001). The family Echimyidae is one of the most species rich and diverse New World rodent lineages with ~40 extant and extinct genera (Carvalho & Salles 2004). Distinct morphological features found in Caribbean hutia led previous studies to establish Capromyidae as a distinct family (Woods & Kilpartrick, 2005), but recent molecular analysis showing that capromyids are part of the echimyid radiation calls this taxonomic distinction into question (Fabre et al., 2017). In contrast the Caribbean heteropsomyines have consistently been associated with South American echimyids on a morphological basis, and are recognised as one of five echimyid subfamilies (†Adelphomyinae, †Heterosomyinae, Eumysopinae, Dactylomtinae and Echimyinae) (Woods & Kilpatrick, 2005). The recent placement of the capromyids within the echimyid radiation raises questions surrounding the evolutionary relationships of heteropsomyines.

Several hypotheses could explain the geographic distribution of heteropsomyine species and their morphological similarity to echimyids but not capromyids: i) heteropsomyines are part of the echimyid radiation, but colonised the Caribbean separately from the Caribbean capromyids, ii) heteropsomyines are part of the Caribbean capromyid radiation and are basal to this group and retained the ancestral morphology of South American echimyids, iii) heteropsomyines are part of the Caribbean capromyid radiation but are not basal to the group, and the morphological distinction between heteropsomyines and capromyids are due to insular adaptations in order to occupy different niches. Molecular analysis of the extinct Caribbean echimyids is needed to test these alternative hypotheses.

Like many of the Caribbean's endemic rodents, different species within Heteropsomyinae are found on separate islands within the archipelago. Two species are included in this study:

Boromys offella formerly found on Cuba and *Brotomys voratus* found exclusively on Hispaniola. Here we further investigate intra-island evolutionary relationships within Heteropsomyinae by using ancient DNA techniques to include extinct Cuban and Hispaniolan species in this molecular phylogenetic analysis. The analysis of multiple species found on different Greater Antillean islands allows us to access the inter-island colonisation patterns of extinct Caribbean echimyids after their initial colonistion of the Caribbean.

6.5 Heptaxodontidae

All taxa placed within the family Heptaxodontidae (giant hutias) are extinct, and most or all species are thought to have become extinct well before the arrival of Europeans in the Caribbean (Biknevicius, McFarlane, & Macphee, 1993; McFarlane, 1999). Heptaxodontidae may be a paraphyletic group (MacPhee 2011), whilst most members share multilamellar teeth and large size, because convergence and gigantism are common amongst insular species (Van der Geer et al. 2010), shared evolutionary relationships between taxa identified as "giant hutia" cannot be assumed. Morphology-based analyses have previously included the following taxa in the group: *Amblyrhiza inundata* (Cope 1868), *Elasmodontomys obliquus* (Anthony, 1916), *Quemisia gravis* (Miller, 1929), *Clidomys osborni* (Anthony, 1920), *Xaymaca fulvopulvis* (MacPhee & Flemming 2003) and *Tainotherium valei* (Turvey et al., 2006). Like much of the rest of the Caribbean Late Quaternary land mammal fauna, the heptaxodontids have a complex taxonomic history. The genus name bearer *Heptaxodon* (Anthony, 1917) was shown to be a synonym of *Elasmodontomys* by Ray (1964). On discovering *Heptaxodon* Anthony placed it, along with *Amblyrhiza* and *Elasmodontomys,* in three monotypic subfamilies within the mainland caviomorph family Chinchillidae. Kraglievich (1926) and Landry (1957) concluded that Caribbean giant hutias should be placed with mid-Cenozoic South American mainland taxa that also shared the large body size and multilamellar dentition exhibited by heptaxodontids. Woods (1989a, 1990; see also Woods and Kilpatrick, 2005) however split the giant hutias into two subfamilies, with Heptaxodontinae including *Quemisia*, *Elasmodontomys*, and *Amblyrhiza*, and *Clidomys* in the separate subfamily Clidomyinae. Woods (1982) also suggested a close relationship between heptaxodontids and capromyids, proposing a shared Miocene ancestor (Woods and Kilpatrick 2005). MacPhee (2011) looked at the basicranial morphology in *Elasmodontomys* and other heptaxodontids and found evidence to suggest that the group may not be monophyletic. MacPhee found similarities between *Elasmodontomys* and mainland caviomorphs, but found no significant similarities between *Amblyrhiza* and *Elasmodontomys*. *Amblyrhiza*, MacPhee suggested resembles chinchilloids in its basicranial features. MacPhee allied *Elasmodontomys* with *Myocastor*, capromyids and echimyids, but did not go as far as to suggest sister taxa for the genera.

This study will look specifically at the phylogenetic placement of *Elasmodontomys* in relation to other Greater Antillean and South American rodents in order to assess evolutionary relationships and colonization histories in the group. Morphological evidence suggests that heptaxodontids either: i) represent a separate colonization of the Caribbean by South American rodents, most likely from the family Echimyidae, or ii) are a morphologically divergent part of the Greater Antillean capromyid radiation, representing a case of intra-island evolution.

6.6 Study aims and objectives

The key question of this study is to establish whether the insular Caribbean was colonised once by a South American echimyid ancestor followed by an adaptive radiation, with some lineages such as capromyids and heptaxodontids becoming morphologically divergent, whilst other lineages remained morphologically similar to mainland echimyids; or if different caviomorph rodent groups represent temporally separated colorizations with different possible mainland sister taxa. This study aims to uncover the evolutionary histories, date the divergence of groups from mainland taxa and investigate subsequent insular colonisation patterns in the Greater Antillean rodent fauna using aDNA techniques, and has three primary aims: i) to extract and sequence aDNA from degraded museum specimens and sub-fossil remains, ii) to use molecular phylogenetic techniques to reconstruct wider evolutionary histories and colonisation patterns in capromyids, heteropsomyines and heptaxodontids, and finally, iii) to look at inter-island patterns of evolution in living and extinct capromyids and Cuban and Hispaniolan heteropsomyines.

6.7 Methods

Specimen information

This study sampled 28 caviomorph rodent specimens sourced from several collections that included sub-fossil remains and museum specimens (tissue, skin and hair samples) for the extraction and sequencing of ancient DNA (Table 1).

Extractions and sequencing

A dedicated aDNA laboratory (Natural History Museum, London) located in a separate building from any post-PCR analysis was used for extractions (Dabney et al. 2013) and Next Generation Sequencing library builds. Samples of bone powder were taken from the

specimens using a Dremel hand held drill with a 2-3mm drill bit. The surface of the bone was cleaned prior to drilling in order to minimize contamination. Drill bits were changed between specimens and all equipment used was sterilized with bleach and UV treated before and after use. A negative control was included in the set of samples during the entire extraction protocol.

Single-index double-stranded DNA libraries were built following the protocol from Meyer & Kircher (2010). Initially libraries were screened for endogenous DNA (Table 2) using the Next Generation Sequencing platform the Illumina MiSeq. In order to increase depth of sequencing of targeted areas in-solution, hybridisation-capture enrichment was then used for samples with low endogenous content from reads mappings, prior to sequencing on the Illumina NextSeq 500 (Enk et al. 2014). Capture enrichment kits (MYcroarray, Ann Arbor) were applied, and baits were designed from whole mitochondrial genomes and 5 nuclear genes: apolipoprotein B (apoB), recombination activating gene 1 (RAG1), von Willebrand factor (vWF), interphotoreceptor retinoid binding (IRBP) and growth hormone receptor (GHR) choosen due to their availablity on the NCBI database Genbank (Benson et al. 2005) for the following echimyid species: *Proechimys quadruplicatus, Proechimys simonsi, Echimys didelphoides, Phyllomys sp. Trinomys setosus* and *Proechimys longicaudatus* (Appendix: Table 2a).

Table 1. Specimens sampled for aDNA as part of this study. Including: species, sample ID, material, locality of collection and Museum or collection

Species	Sample ID	Material	Locality	Museum or collection
Geocapromys thoracatus	RW053	Tissue	Little Swan Island	RAMM (Exeter)
Geocapromys thoracatus	RW054	Tissue	Little Swan Island	RAMM (Exeter)
Brotomys voratus	SB180	Bone	Hispaniola	ST
Boromys offella	SB293	Bone	Cuba	Peabody Museum, Yale University
Boromys offella	SB295	Bone	Cuba	Peabody Museum, Yale University
Boromys offella	SB294	Bone	Cuba	Peabody Museum, Yale University
Geocapromys columbianus	SB298	Bone	Cuba	Peabody Museum, Yale University
Geocapromys columbianus	SB299	Bone	Cuba	Peabody Museum, Yale University
Geocapromys columbianus	SB300	Bone	Cuba	Peabody, Museum, Yale University
Capromys sp.	RW057	Bone	Cayman Brac	FLMNH
Capromys sp.	RW058	Bone	Cayman Brac	FLMNH
Capromys sp.	RW059	Bone	Cayman Brac	FLMNH
Capromys sp.	RW066	Bone	Cayman Brac	FLMNH
Capromys sp.	RW067	Bone	Cayman Brac	FLMNH
Elasmodontomys obliquus	RW060	Bone (rib)	Cueva de la Vaca, Puerto Rico	ST
Elasmodontomys obliquus	RW061	Bone (rib)	Cueva de la Vaca, Puerto Rico	ST
Elasmodontomys obliquus	RW062	Bone	Cueva de la Vaca, Puerto Rico	ST
Elasmodontomys obliquus	RW063	Tooth	Cueva de la Vaca, Puerto Rico	ST
Elasmodontomys obliquus	RW064	Bone	Cueva de la Vaca, Puerto Rico	ST
Elasmodontomys obliquus	RW065	Tooth	Cueva de la Vaca, Puerto Rico	ST
Isolobodon portoricensis	SB165	Bone	Puerto Rico	ST
Isolobodon portoricensis	RWI78	Tooth (incisor)	Guana Island (Virgin Islands)	ST
Isolobodon portoricensis	RWI79	Tooth (molar)	Guana Island (Virgin Islands)	ST
Isolobodon portoricensis	RWI80	Tooth (incisor)	"Two caves", Parque del Este, Dom. Rep.	ST
Isolobodon portoricensis	RWI81	Tooth (molar)	Cueva de La Vaca, Puerto Rico	ST
Isolobodon portoricensis	RWI82	Bone	Cueva de La Vaca, Puerto Rico	ST
Isolobodon portoricensis	RWI83	Tooth (incisor)	Cueva de La Vaca, Puerto Rico	ST
Isolobodon montananus	RWI84	Bone (mandible)	"Two caves", Parque del Este, Dom. Rep.	ST
Hexolobodon phenax	RWH77	Tooth (molar)	San Gabriel Cave, Samana Bay, Dom. Rep.	ST

Museum collections abbreviations: Florida Museum of Natural History (FLMNH), Dr Samuel Turvey collection (ST) and Royal Albert Memorial Museum (RAMM)

Post sequencing data processing and phylogenetic analysis

Post sequencing data processing was achieved using CLC Workbench software v.8 (CLC Bio-Qiagen, Aarhus, Denmark), where reads were paired, merged and then trimmed of adapters using default settings. In order to account for potential mapping ascertainment bias during sequence assembly, an iterative mapping process based on Westbury et al. (2017) was applied to extinct species for which there was no close reference sequence available. Reads were mapped to four different reference sequences (Appendix: Table 2a) for the whole mitochondrial genome and for each nuclear gene targeted. All four reference sequences were used in six separate read mappings that were given increasingly strict parameters (Table 2 and Appendix: Table 3a). Consensus sequences for each reference sequence and each parameter were extracted from the read mappings. Consensus sequences for different reference sequences but the same mapping parameter were then aligned, variations between sequences were removed and a single consensus sequence was extracted. These mapping parameter consensus sequences were further aligned with each other and a final consensus was extracted.

Sequence data generated using the above method were aligned to other sequence data generated for this study and outgroup taxa (Appendix: Table: 5a), using the alignment tool ClustalW (Larkin et al. 2007) implemented in the programme Geneious v. 8.0.5 (Kearse et al. 2012) (http://www.geneious.com). The resulting alignments for five nuclear genes and the whole mitogenome were concatenated using the program Seaview v.4 (Gouy et al. 2010). Maximum likelihood and Bayesian methods were implemented to estimate phylogenetic relationships. DNA substitution models were chosen for the partitioned dataset using the software PartitionFinder (Lanfear et al. 2012) (Appendix: Table 5a). A maximum likelihood phylogeny (Appendix Figure. 3a) and corresponding bootstrap support values were generated using RAxML v.8 (Stamatakis 2014) which was implemented in CIPRES Science Gateway v.3.3 (Miller, Pfeiffer, & Schwartz, 2010). Bayesian phylogenies were constructed using MrBayes (Ronquist & Huelsenbeck 2003) with four chains (three heated, one cold) that were run for 1×10^6 generations, sampling every 1×10^3 generations with a burn-in period of 250 trees (Figure 3).

A divergence-dated phylogeny was estimated under an uncorrelated relaxed lognormal molecular clock in the programme BEAST v. 1.8.3 (Drummond & Rambaut 2007). As in the previous analyses the best-fit evolutionary models were chosen in the programme Partitionfinder. Two models of speciation were run for comparison, Yule and birth-death models, both of which generated identical topology. The prior distributions of five nodes were set using fossil calibration points (Appendix: Table: 6a). All other priors were left as default values in BEAUti v. 1.8.3 (Drummond et al. 2012). This analysis was run for 25 million

generations, sampling every 1000 generations. Convergence and effective sample size for all parameters were accessed in the programme Tracer v. 1.6.0 (Rambaut et al. 2014) after a burn in of 10% (Figure 4).

6.8 Results

Despite the adverse preservational conditions associated with tropical sub-fossil remains, this study was able to extract and sequence aDNA from a series of extinct endemic Caribbean rodents, including single samples from each of the extinct *Capromys* sp. from the Cayman Islands, *Geocapromys columbianus* and *Geocapromys thoracatus*, the heteropsomyines *Boromys offella* and *Brotomys voratus*, and the heptaxodontid *Elasmodontomys obliquus*, although additional aDNA extraction from *Isolobodon* and *Hexolobodon* was unfortunately not successful. Partial or whole mitochondrial genome data (15,259 bp) and 5 nuclear genes: apolipoprotein B (apoB), recombination activating gene 1 (RAG1), von Willebrand factor (vWF), interphotoreceptor retinoid binding (IRBP) and growth hormone receptor (GHR) were recovered for each species included in phylogenetic analysis. 10 samples were initially screened and showed low levels of endogenous content (average estimated endogenous content for the mitochondrial genome of 0.07). Samples with the highest levels of endogenous content for subject species were then subjected to additional sequencing including target capture enrichment. Although target capture enrichment did not always result in an increase of endogenous content (the average estimated enrichment factor; comparison between pre and post-target capture enrichment of libraries, was -212%) enrichment did result in an increase of reads mapped to the mitochondrial genome reference sequence with an average percentage increase in reads mapped of 68% (Table 2).

Table 2: Sequencing results including pre-target capture enrichment screening results and post-capture results. Endogenous content and enrichment factor were estimated using mapping results from the whole mitochonrial genome reference seqeunce used.

Species	*N° of samples sequenced*	*Sample size used in the final analysis*	*Sample ID*	*Pre-capture*					*Post capture*					*Enrichment factor*	*% increase / decrease reads mapped*
				N° of paired merged reads	*Av. coverage*	*Reads mapped*	*Reads mapped mtgenome*	*Est. endogenous content*	*N° of paired merged reads*	*Av. coverage*	*Reads mapped*	*Reads mapped mtgenome*	*Est endogenous content*		
Geocapromys thoracatus	1	1	RW54	2,709,500	72.3 (whole mtgenome), 0.17 (vwF), 0.34 (RAG1), 0.2 (irbp), 0.39 (apoB), 0.38 (GHR)	16,051 (whole mtgenome), 5 (vwF), 7 (RAG1), 4 (irbp), 9 (apoB), 8 (GHR)	12,927	0.477099096	5,405,344	63.14 (whole mtgenome), 1.85 (vwF), 2.92 (RAG1), 1.31 (irbp), 3.76 (apoB), 1.92 (GHR)	12,927 (whole mtgenome), 29 (vwF),46 (RAG1), 23 (irbp), 54 (apoB), 23 (GHR)	16,051	0.296946873	-61%	19%
Elasmodontomys obliquus	6	1	RW65	2,317,525	0.4 (whole mtgenome), 0 (vwF), 0 (RAG1), 0 (irbp), 0 (apoB), 0 (GHR)	6.18E-03 (whole mtgenome), 0 (vwF), 0 (RAG1), 0 (irbp), 0 (apoB), 0 (GHR)	6.18E-03	2.67E-07	4,426,535	0.45 (whole mtgenome), 2.88 (vwF), 1.47 (RAG1), 2.9 (irbp), 0.3 (apoB), 0.86 (GHR)	261 (whole mtgenome), 146 (vwF), 71 (RAG1), 157 (irbp), 16 (apoB), 31 (GHR)	261	0.00589626	100%	100%
			RW63	13,172,455	4.13 (whole mtgenome), 2.07 (vwF), 1.41 (RAG1), 1.79 (irbp), 0.59 (apoB), 1.07 (GHR)	3,764 (whole mtgenome), 132 (vwF), 845 (RAG1), 230 (irbp), 378 (apoB), 490 (GHR)	3,764	0.02857478	63,079,541	4.89 (whole mtgenome), 20.71 (vwF), 14.15 (RAG1), 17.9 (irbp), 5.91 (apoB), 10.72 (GHR)	4,135 (whole mtgenome), 1,324 (vwF), 845 (RAG1), 1,230 (irbp), 378 (apoB), 490 (GHR)	4,135	0.006555216	-336%	9%

Table 2 (continued): Sequencing results including pre-target capture enrichment screening results and post-capture results. Endogenous content and enrichment factor were estimated using mapping results from the whole mitochonrial genome reference seqeunce used.

Species	*N° of samples sequenced*	*Sample size used in the final analysis*	*Sample ID*	*Pre-capture*					*Post capture*					*Enrichment factor*	*% increase/ decrease reads mapped*
				N° of paired merged reads	*Av. coverage*	*Reads mapped*	*Reads mapped mtgenome*	*Est. endogenous content*	*N° of paired merged reads*	*Av. coverage*	*Reads mapped*	*Reads mapped mtgenome*	*Est endogenous content*		
Brotomys voratus	1	1	SB180	1,517,978	7.14 (whole mtgenome), 0.20 (vwF), 0.18 (RAG1), 0.34 (irbp), 0.24 (apoB), 0.13 (GHR)	1,934 (whole mtgenome), 6 (vwF), 6 (RAG1), 10 (irbp), 4 (apoB), 2 (GHR)	1,934	0.127406326	46,254,515	16.69 (whole mtgenome), 10.5 (vwF), 5.43 (RAG1), 8.64 (irbp), 3.23 (apoB), 2.16 (GHR)	47,251 (whole mtgenome), 204 (vwF), 59 (RAG1), 74 (irbp), 100 (apoB), 25 (GHR)	47,251	0.102154352	-25%	96%
Boromys offella	3	1	SB294	1,252,406	1.36 (whole mtgenome), 0.37 (vwF), 0.25 (RAG1), 0.25 (irbp), 0.13 (apoB), 0.11 (GHR)	569 (whole mtgenome), 21 (vwF), 14 (RAG1), 11 (irbp), 8 (apoB), 4 (GHR)	569	0.045432551	44,000,746	1.69 (whole mtgenome), 1.92 (vwF), 1.72 (RAG1), 2.09 (irbp), 0.36 (apoB), 0.81 (GHR)	2,519 (whole mtgenome), 102 (vwF), 94 (RAG1), 129 (irbp), 22 (apoB), 31 (GHR)	2,519	0.005724903	-694%	77%
			SB295	2,140,257	0.33 (whole mtgenome), 0.05 (vwF), 0.09 (RAG1), 0.13 (irbp), 0.06 (apoB), 0.02 (GHR)	327 (whole mtgenome), 2 (vwF), 3 (RAG1), 6 (irbp), 2 (apoB), 1 (GHR)	326	0.015231816	52,400,401	4.84 (whole mtgenome), 1.64 (vwF), 1.04 (RAG1), 1.19 (irbp), 0.43 (apoB), 0.50 (GHR)	1,620 (whole mtgenome), 104 (vwF), 64 (RAG1), 86 (irbp), 28 (apoB), 23 (GHR)	16,131	0.030784116	51%	98%
			SB293	1,102,436	0.0008896 (whole mtgenome), 0 (vwF), 0 (RAG1), 0 (irbp), 0 (apoB), 0 (GHR)	1 (whole mtgenome), 0 (vwF), 0 (RAG1), 0 (irbp), 0 (apoB), 0 (GHR)	1	9.07082E-05							

Table 2 (continued): Sequencing results including pre-target capture enrichment screening results and post-capture results. Endogenous content and enrichment factor were estimated using mapping results from the whole mitochonrial genome reference seqeunce used.

Species	*N° of samples sequenced*	*Sample size used in the final analysis*	*Sample ID*	*Pre-capture*					*Post capture*					*Enrichment factor*	*% increase / decrease reads mapped*
				N° of paired merged reads	*Av. coverage*	*Reads mapped*	*Reads mapped mtgenome*	*Est. endogenous content*	*N° of paired merged reads*	*Av. coverage*	*Reads mapped*	*Reads mapped mtgenome*	*Est endogenous content*		
Geocapromys columbinanus	3	1	SB299	3,762,236	1.56 (whole mtgenome), 6.67 (vwF), 3.27 (RAG1), 6.49 (irbp), 0.86 (apoB), 0 (GHR)	1,112 (whole mtgenome), 447 (vwF), 186 (RAG1), 410 (irbp), 57 (apoB), 0 (GHR)	1,112	0.029556891	78,429,431	2.33 (whole mtgenome), 6.95 (vwF), 5.03 (RAG1), 8.02 (irbp), 0.82 (apoB), 3.87 (GHR)	3,363 (whole mtgenome), 396 (vwF), 281 (RAG1), 505 (irbp), 51 (apoB), 160 (GHR)	3,363	0.004287931	-589%	67%
			SB300	3,451,270	0.04 (whole mtgenome), 0.69 (vwF), 0.10 (RAG1), 0.44 (irbp), 0.16 (apoB), 0 (GHR)	23 (whole mtgenome), 21 (vwF), 6 (RAG1), 26 (irbp), 4 (apoB), 0 (GHR)	23	0.000666421	40,120,287	0.07 (whole mtgenome), 0.37 (vwF), 0.11 (RAG1), 0.02 (irbp), 0.39 (apoB), 0.08 (GHR)	109 (whole mtgenome), 39 (vwF), 23 (RAG1), 2 (irbp), 25 (apoB), 3 (GHR)	109	0.000271683	-145%	79%
Isolobodon portoricensis	6	0	RW165	137,018	0.04 (whole mtgenome), 0 (vwF), 0 (RAG1), 0 (irbp), 0 (apoB), 0 (GHR)	14 (whole mtgenome), 0 (vwF), 0 (RAG1), 0 (irbp), 0 (apoB), 0 (GHR)	14	0.010217636	na	Na	na	na	na	na	na

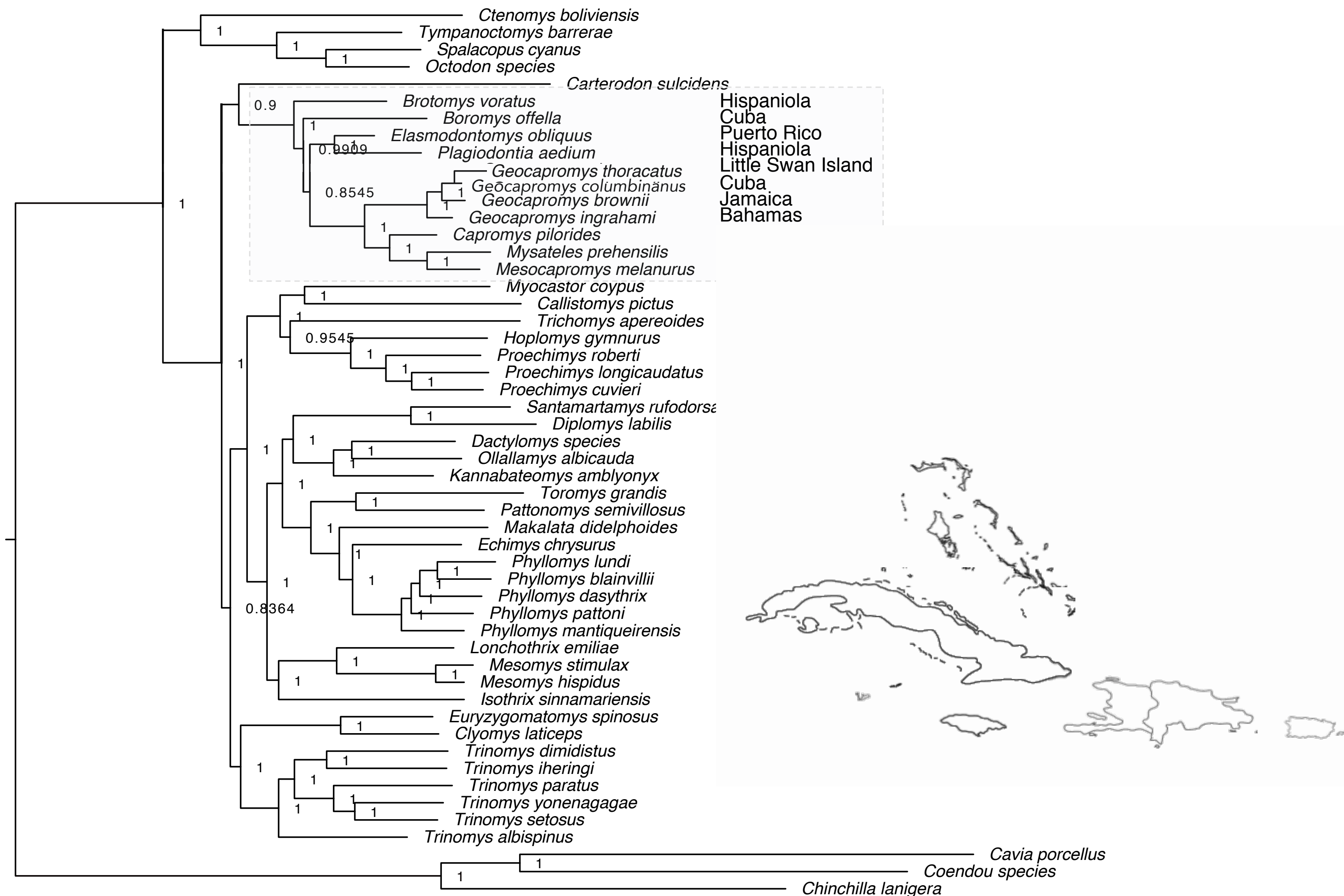


Figure 3. Bayesian phylogeny constructed using MrBayes (Ronquist & Huelsenbeck 2003). Node values represent posterior probabilities, and scale indicates nucleotide substitutions per site).

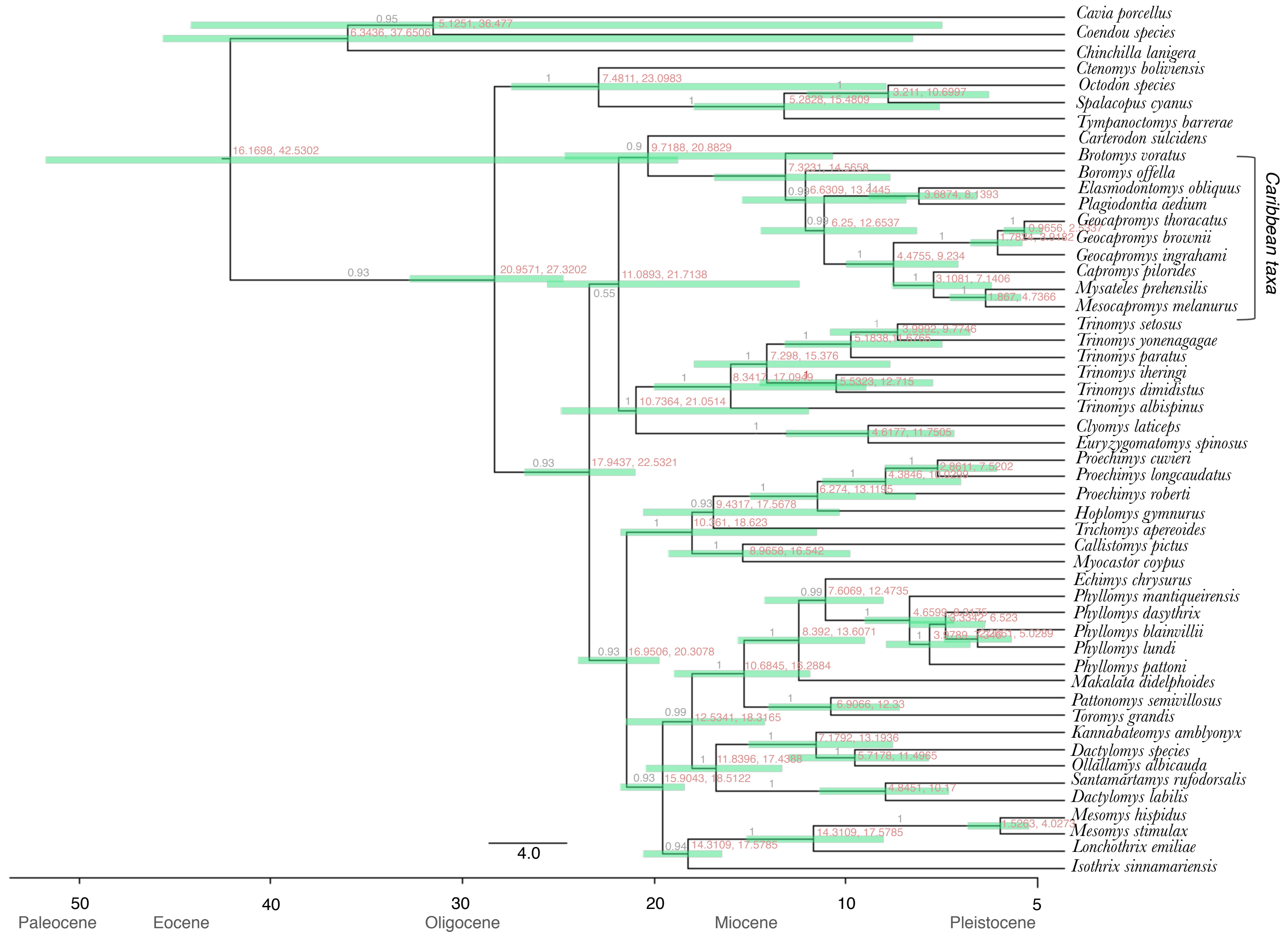


Figure 4. Time-calibrated phylogeny showing estimated divergence dates for Greater Antillean caviomorph species and extant Neotropical echimyid genera. Node bars indicate 95% highest posterior density values. Red branch values represent posterior probabilities.

Divergence-dated (Figure 4) and Bayesian phylogenies (Figure 3) were congruent apart from the placement of the *Carterodon* + Capromyidae clade, which in the Bayesian and Maximum likelihood phylogeny is basal to all other echimyid clades. The divergence-dated analysis places the *Carterodon* + Capromyidae clade as sister to Euryzygomatomyinae. Support for the latter arrangement is low (posterior probability = 0.567), whilst the Bayesian (Figure 3) and maximum likelihood (Appendix: Figure 1a) placement has a high posterior probability support value of 1. Fabre et al. (2017) were unable to resolve deeper relationships within the extant Echimyinae, with a trifurcation involving Capromyidae and Euryzygomatomyinae. This study found that mitogenomic and nuclear exon trees were congruent, apart from the phylogenetic position of *Carterodon*. One contributing factor to these phylogenetic inconsistences may be that, because the lineage leading to the genus *Carterodon* is characterised by long branch lengths, saturation in this taxon could inhibit accurate phylogenetic placement and affect the branching pattern between *Carterodon* and its closest relatives.

Despite the differences in the placement of the *Carterodon* + Capromyidae clade, divergence-dated and Bayesian and maximum likelihood phylogenetic analyses suggest that all Greater Antillean caviomorph rodents included in this study form a monophyletic group, with extinct heteropsomyines (*Boromys offella* and *Brotomys voractus*) and the extinct heptaxodontid *Elasmodontomys obliquus* all included within the capromyid radiation. This study dates the capromyid initial colonisation to the mid Miocene (~12.1 Ma 95% HPD 10.4406, 14.4407) consistent with previous molecular analyses. Heteropsomyines (*Boromys offella* and *Brotomys voractus*) are found sequentially nested within the capromyid clade with a diveregence date from the rest of the capromyids of ~11 Ma (HPD 95% 7.3231, 14.5658). The heptaxodontid *Elasmodontomys obliquus* is recovered as sister to *Plagiodontia aedium* with an estimated divergence date for this lineage of ~7.2 Ma (HPD 95% 3.6874, 8.1393).

6.9 Discussion and conclusions

This study has been able to successfully extract and sequence endogenous DNA from extinct capromyids, heteropsomyines and a single heptaxodontid. This is the first molecular work to be carried out on the majority of these species, and provides important new insights into the colonisation histories and evolutionary relationships between Greater Antillean caviomorph rodents. Divergence-dated phylogenetic analysis suggests that all Greater Antillean caviomorph rodents investigated as part of this study form a monophyletic group which diverged from their South American echimyid ancestors in the mid Miocene (~12.1 Ma). This colonisation was followed by a within-Caribbean radiation, with some taxa (heteropsomyines) retaining the ancestral spiny rat morphology whilst others (capromyids and heptaxodontids)

grew morphologically dissimilar to their closest mainland relatives, most likely as a result of adaptation to insular environments.

Taxonomic implications

Recent molecular analysis has confirmed the phylogenetic relationship between living Greater Antillean capromyid rodents and South American echymids (Fabre et al., 2014, 2016). However the relationships between this group of caviomorph rodents and two other morphologically distinct taxonomic groups, the heteropsomyines and heptaxodontids, was previously unclear, due in part to the recent extinction of both heteropsomyines and heptaxodontids.

Previous analysis had placed heteropsomyines as a subfamily within Echimyidae based on morphological similarities (Carvalho & Salles, 2004; Woods & Kilpartrick, 2005). Here we have been able to show that heteropsomyines instead fall within the capromyid radiation rather than being more closely related to mainland echimyid groups. Further more the two heteropsomyines included in this study were not recovered as a monophyletic clade suggesting that Heteropsomyinae can definitely no longer be recognised as a valid taxonomic grouping. The results of this study suggest that the extinct endemic Caribbean echimyids should be placed as a stem taxa within Capromyidae.

The heptaxodontids have also previously been allied to capromyids and echimyids, but morphological evidence did not rule out the possibility that this family may be more closely related to another echimyid group (MacPhee 2011; MacPhee & Flemming 2003). This study was able to include *Elasmodontomys* in our molecular analysis, showing a sister relationship between this species and the Hispaniolan hutia *Plagiodontia aedium*. Previous morphology-based studies have been unable to suggest specific sister taxa for *Elasmodontomys,* both Woods & Hermanson (1985) and MacPhee (2011) suggest that *Elasmodontomys* may be closely related to capromyids, neither suggest a relationship with *Plagiodontia,* although *Plagiodontia* was not included in the basicranial analysis carried out by MacPhee. Tate describes *Elasmodontomys* and the Hispaniolan heptaxodontid *Quemisia* as “offshoots” of an Isolobodontine subfamily, and goes on to link *Plagiodontia* with Isolobodon (Tate 1948) but this is the only reference to a connection between the two genera, aside from the suggestion by several authors that *Elasmodontomys* is associated with capromyids in general (MacPhee, 2011; Woods & Kilpartrick, 2005). The inclusion of other extinct heptaxodontids from Hispaniola (*Quemisia*) and Puerto Rico (*Tainotherium*) may help shed further light on the relationship between giant hutia and *Plagiodontia*.

The results of this study suggest that species currently placed in Heptaxodontidae may also be part of the capromyid radiation. However because this study was only able to include a single heptaxodontid species and previous studies have suggested that the group may be paraphyletic, and have also allied heptaxodontid species with a wider range of mainland sister taxa than heteropsomyines (MacPhee 2011), this study is unable to rule out the possibility that other taxa in this diverse group may be more closely related to another mainland group and thus form part of a separate colonisation of the insular Caribbean. Without the inclusion of further heptaxodontid taxa this study can only suggest the inclusion of *Elasmodontomys* as a part of the crown group capromyid clade.

This study recovers the capromyid clade (including *Elasmodontomys,* heteropsomyines and *Carterodon*) as sister to all other echimyids, suggesting that the familial level status should still be given to the endemic Caribbean caviomorph rodents now inclusive of heteropsomyines and at least one heptaxodontid taxaon.

Colonisation of the insular Caribbean by Greater Antillean caviomorph rodents

This monophyletic grouping of extant and extinct Greater Antillean caviomorphs supports the hypothesis that the diversity and distribution of Greater Antillean caviomorph rodents is the result of a single mid Miocene colonisation event. Whilst previous studies had suggested a Miocene colonisation date of the Caribbean for living capromyids, this study now includes extinct heteropsomyines and heptaxodontids within this group. Other Caribbean vertebrate taxa share mid-late Miocene colonisation dates (Brace et al., 2015; Hedges & Conn, 2012; Hedges et al., 2009). The late Miocene was a period defined by low global sea level, during which time a riverine association may have existed between South America and the Caribbean (Wilkinson et al. 2010) promoting overwater dispersal to the insular Caribbean by multiple vertebrate lineages.

Wider inter-island evolutionary relationships

Previous studies have suggested multiple alternative hypotheses to explain archipelago-wide inter-island evolutionary relationships between capromyids. These hypotheses are based on either vicariance or stepping stone colonisation between eastern and western Greater Antillean islands (Fabre et al., 2014). This study places extinct the giant hutia *Elasmodontomys* as sister to *Plagiodontia.* The distributions of both species are on eastern Greater Antillean islands: *Elasmodontomys* formerly on Puerto Rico and *Plagiodontia* on Hispaniola. These two islands may still have been connected until the early-mid Miocene (Iturralde-Vinent & Gahagan, 2002). However, the estimated divergence between

Elasmodontomys and *Plagiodontia* at ~6.4 Ma (HDP 5.007, 7.8201 Ma) post dates the separation of these islands.

Whilst *Plagiodontia* is identified as the most basal capromyid species in previous molecular phylogenies, the inclusion of extinct heteropsomyines and heptaxodontids within the monophyletic capromyid clade removes *Plagiodontia* from its basal position. The inclusion of extinct species suggests that heteropsomyines, with their ancestral morphology, are at the base of the capromyid radiation. Hispaniola may still be the origin of the capromyid radiation as *Brotomys voratus*, the Hispaniolan heteropsomyine included in this study, is recovered at the base of the capromyid radiation. However further analysis of other extinct Greater Antillean caviomorph species, particularly the Puerto Rican heteropsomyine *Heteropsomys* and Hispaniolan and Cuban isolobodontines, is necessary before the true biogeographic origns of this group can be fully resolved. This analysis also suggests that although only a single colonisation event from South America by Greater Antillean caviomorphs can be recognised, multiple inter-island colonisations or variance events occurred after the initial colonisation, generating past and current species diversity.

Inter-island evolutionary histories

Heteropsomyinae

Two species of heteropsomyines were included in this study: Cuban *Boromys offella* and Hispaniolan *Brotomys voractus*. *B. voractus* was recovered as a basal lineage to all other Greater Antillean caviomorph taxa included in this study, with *B. offella* as a second sequentially nested lineage, rather than as sister taxa to *B. voractus*. The dated phylogeny reveals a mid-late Miocene (~11.3 Ma, 95% HPD 9.569–13.370 Ma) divergence between *B. offella* and *B. voractus* and all other endemic Caribbean taxa. This mid-late Miocene divergence date between Cuban and Hispaniolan lineages does not correspond to geological reconstructions of a possible connection between south-eastern Cuba and northern Hispaniola, as this connection is thought to have been severed during the mid-Eocene (Iturralde-Vinent & MacPhee 1999; Lewis & Draper 1990), precluding this separation as an explanation for the biogeographic distribution and divergence dates of Cuban and Hispaniolan heteropsomyines. Several present day trenches in the Caribbean sea also formed during this time, including the Cayman Trench, the Anagada Trench, the Puerto Rico Trench and the Tortuga basin, possibly forming biogeographic barriers (Iturralde-Vinent 2006), isolating populations and facilitating the generation of species, if these species formed part of a single continuous populations across the region during this period.

Capromyidae

Three extinct capromyid species were included in this study: *Capromys sp.* from the Cayman Islands, *Geocapromys columbianus* from Cuba and *Geocapromys thoracatus* from Little Swan Island. As predicted, both extinct *Geocapromys* were placed in the same clade as living *Geocapromys* species. The Bahamian hutia *G. ingrahami* is recovered as basal to this clade (HPD 95% 1.782, 3.915), with *G. thoracatus* and *G. columbianus* most closely related to the Jamaican hutia *G. brownii*, suggesting that *Geocapromys* may have originated and radiated from the Bahamas. The *Capromys* individual from Cayman Brac appears to be conspecific to Cuban mainland species *Capromys pilorides* with a divergence date of ~0.7 (HPD 95% 0.086, 0707) (pairwise distance: 0.014 (see Chapter 5)) suggesting recent colonisation of the Cayman Islands by Cuban *C. pilorides* before their extinction. The recent divergence dates found in both extinct sub-fossil *Capromys* and *Geocapromys* taxa suggest that the southward radiations of both genera may have been more recent than previously thought.

Insular evolution

The morphology of both heptaxodontids and heteropsomyines is significantly different from living and extinct capromyids, so much so that previous studies have been unable to confirm evolutionary relationships between these three taxonomic groups despite shared distributions in the insular Caribbean, and the additional fact that a single rare, mammalian colonisation event is more parsimonious than multiple colonisation events. Islands are well known for driving the evolution of unusual morphology in endemic mammals (Sondaar 1977; Renaud & Millien 2001). Insular syndrome where the optimum size for a species is altered due to the different environmental conditions that islands often provide (e.g. the removal of predators, competition or change in diet or habitat), has been reported in many island mammal species (Lister 1996; Van Valen 1965). The Caribbean in particular lacks medium and large herbivorous mammals, potentially leaving this niche open for endemic caviomorph rodent species to exploit. Additionally, the rate of evolution may be increased on islands (Millien 2006). These factors may have affected the evolution of Greater Antillean caviomorph rodents, generating the morphological differences seen between taxa with close shared evolutionary histories. Other insular mammals demonstrate remarkable morphological variation in comparison with their mainland relatives, for example the Caribbean primate *Xenothrix*, which is most closely related to callicebines characterised by previously underestimated species diversity due to the morphological similarity between groups. However the South American echimyid family far from being morphologically restricted is a species rich taxonomic group containing several species with unusual or unique morphology such as the large coypu (*Myocastor coypus*) and as well as species with modified spine like hairs (Leite & Patton 2002).

This ancient DNA analysis suggests that Caribbean caviomorph rodents constitute a monophyletic radiation, the result of a single mid-Miocene colonisation event. This monophyletic clade contains both insular species with novel morphology in comparison with mainland echimyids and also species that retained the morphology of mainland sister taxa. This study has shown that a single colonisation event from South America to the insular Caribbean generated past and current species diversity within Greater Antillean caviomorphs, a group which is made up of multiple discrete, taxonomic species groups that are found on multiple islands within the archipelago, and which were previously interpreted as being distinct at the family or subfamily level. This finding suggests that inter-island divergence occurred after the initial colonisation of the Caribbean, making the current and past species diversity in Greater Antillean caviomorph rodents the result of a true insular radiation. Further molecular analysis of other extinct Greater Antillean caviomorphs, including the extinct capromyids *Isolobodon* and *Hexolobodon,* further species of giant hutia such as *Amblyrhiza inundata* and remaining heteropsomyine species, is necessary to fully resolve the evolutionary relationships in this group.

References

Ali, J. R. (2012). Colonizing the Caribbean: Is the GAARlandia land-bridge hypothesis gaining a foothold? *Journal of Biogeography*, *39*(3), 431–433.

Allen, G. M. (1918). Fossil mammals from Cuba. *Bulletin of the Museum of Comparitive Zoology*, *62*, 131–148.

Alonso, R., Crawford, A. J., & Bermingham, E. (2012). Molecular phylogeny of an endemic radiation of Cuban toads (Bufonidae: *Peltophryne*) based on mitochondrial and nuclear genes. *Journal of Biogeography*, *39*(3), 434–451.

Anthony, B. (1920). New mammals from Jamaica. *American Museum of Natural History*, *59*(9), 469–487.

Anthony, H. E. (1916). Preliminary Report of Fossil Mammals from Porto Rico, with Descriptions of a New Genus of Hystricomorph Rodents. *Annals of the New York Academy of Sciences*, *27*, 193–203.

Anthony, H. E. (1917). New fossil rodents from porto Rico; with additional notes on *Elasmodontomys obliquus* and *Heteropomys insulans*. *Bulletin of the American Museum of Natural History*, *37*(4).

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Wheeler, D. L. (2005). GenBank. *Nucleic Acids Research*, *33*, 34–38.

Biknevicius, A. R., McFarlane, D. a., & Macphee, R. D. E. (1993). Body Size in Amblyrhiza inundata (Rodentia: Caviomorpha), an Extinct Megafaunal Rodent from the Anguilla Bank, West Indies: Estimates and Implications. *American Museum Novitates*, *3079*(3079), 1–25.

Borroto-Páez, R., & Mancina, C. A. (2017). Biodiversity and conservation of Cuban mammals: past, present, and invasive species. *Journal of Mammalogy*, *98*(4), 964–985.

Brace, S., Barnes, I., Powell, A., Pearson, R., Woolaver, L. G., Thomas, M. G., & Turvey, S. T. (2012). Population history of the Hispaniolan hutia *Plagiodontia aedium* (Rodentia: Capromyidae): Testing the model of ancient differentiation on a geotectonically complex Caribbean island. *Molecular Ecology*, *21*(9),

Brace, S., Turvey, S. T., Weksler, M., Hoogland, M. L. P., & Barnes, I. (2015). Unexpected evolutionary diversity in a recently extinct Caribbean mammal radiation. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1807), 20142371–20142371.

Carvalho, G. A. S., & Salles, L. O. (2004). Relationships among extant and fossil echimyids (Rodentia: Hystricognathi). *Zoological Journal of the Linnean Society*, *142*(4), 445–477.

Cooke, S. B., Davalos, L., Mychajliw, A., Turvey, S., & Upham, N. S. (2017). Anthropogenic Extinction Dominates Holocene Declines of West Indian Mammals. *Annual Review of Ecology, Evolution, and Systematics*, *48*(1).

Cope, E. D. (1868). Exhibition of bones and teeth of a large rodent from the cave deposits of Anguilla, one of the Virgin West India Islands. *Proceedings of the Academy of Natural*

Sciences of Philadelphia, *20*, 313.

Dabney, J., Knapp, M., Glocke, I., Gansauge, M. T., Weihmann, A., Nickel, B., … Meyer, M. (2013). Complete mitochondrail genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences*, *110*(39), 15758–15763.

Davalos, L. M., & Turvey, S. T. (2012). West Indian Mammals: The old the new and the recently extinct. In *Bones, Clones, and Biomes: The History and Geography of Recent Neotropical Mammals* (pp. 157–202).

Donoghue, P. C. J., & Benton, M. J. (2007). Rocks and clocks: calibrating the Tree of Life using fossils and molecules. *Trends in Ecology and Evolution*, *22*(8), 424–431.

Draper, G., Mann, P., & Lewis, J. (1994). Hispaniola. In *Caribbean Geology: An introduction* (pp. 129–145). Kingston: U.W.I Publishers Association.

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

Drummond, A. J., Suchard, M. A., Xie, D., & Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST. *Molecular Biology and Evolution*, *29*, 1969–1973.

Enk, J. M., Devault, A. M., Kuch, M., Murgha, Y. E., Rouillard, J., & Poinar, H. N. (2014). Ancient whole genome enrichment using baits built from modern DNA. *Molecular Biology and Evolution*, *31*(5), 1292–1294.

Fabre, P.-H., et al. (2016). Mitogenomic phylogeny, diversification, and biogeography of South American spiny rats. *Molecular Biology and Evolution*, *34*(3), 613–633.

Fabre, P.-H., et al. (2014). Rodents of the Caribbean: origin and diversification of hutias unravelled by next-generation museomics. *Biology Letters*, *10*(7).

Gaines, M. S., & McClenaghan Jr., L. R. (1980). Dispersal in Small Mammals. *Annual Review of Ecology and Systematics*, *11*(1), 163–196.

Gerrit S. Miller, J. (1929). Three Small Collections of Mammals from Hispaniola. *Smithsonian Miscellaneous Collections*, *82*(15).

Glor, R. E., Kolbe, J. J., Powell, R., Larson, A., & Losos, J. B. (2003). Phylogenetic Analysis of Ecological and Morphological Diversification in Hispaniolan Trunk-Ground Anoles (*Anolis Cybotes* Group). *Evolution*, *57*(10), 2383–2397.

Gouy, M., Guindon, S., & Gascuel, O. (2010). Seaview version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution*, *27*, 221–224.

Graham, A. (2003). Geohistory models and Cenozoic paleoenvironments of the Caribbean region. *Systematic Botany*, *28*(2), 378–386.

Hedges, S. B. (2006). Paleogeography of the Antilles and Origin of West Indian Terrestrial Vertebrates. *Annals of the Missouri Botanical Garden*, *93*(2), 231–244.

Hedges, S. B., & Conn, C. E. (2012). A new skink fauna from Caribbean islands (Squamata, Mabuyidae, Mabuyinae). *Zootaxa*, *244*(3288), 1–244.

Hedges, S. B., Couloux, A., & Vidalj, N. (2009). Molecular phylogeny, classification, and biogeography of West Indian racer snakes of the Tribe Alsophiini (Squamata, Dipsadidae, Xenodontinae). *Zootaxa*, *28*(2067), 1–28.

Hofreiter, M., et al. (2015). The future of ancient DNA: Technical advances and conceptual shifts. *BioEssays*, *37*(3), 284–293.

Iturralde-Vinent, M. A. Gahagan, L. (2002). Latest Eocene to Middle Miocene Tectonic Evolution of the Caribbean: Some Principles and their Implications for Plate Tectonic Modelling. In *Caribbean Geology into the third Millennium: Transactions of the Fifteenth Caribbean Geological Conference* (pp. 47–62).

Iturralde-Vinent, M. A. (2006). Meso-Caribbean paleogeography:implications for the historical biogeography of the region. *International Geology Review*, *48*(9), 791–827.

Iturralde-Vinent, M. A., & MacPhee, R. D. E. (1999). Paleogeography of the Caribbean region: implications for Cenozoic biogeography. *American Museum of Natural History*.

Kearse, M., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for thr organization and analysis of sequence data. *Bioinformatics*, *28*(12), 1647–1649.

Kraglievich, L. (1926). Los grandes roedores terciarios de la Argentina y sus relaciones con ciertos generos pleistocenos de las Antillas. *Anales Del Museo Nacional de Historia Natural*, *34*, 121–135.

Landry, S. (1957). The interrelationships of the New and Old World hystricomorph rodents. *University of Caliornia Publications in Zoology*, *56*, 1–118.

Lanfear, R., Calcott, B., Ho, S. Y., & Guindon, S. (2012). PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, *29*, 1695–1701.

Larkin, M. A., et al. (2007). Clustal W and Clustal X. *Bioinformatics*, *23*, 2947–2948.

Leite, Y. L. R., & Patton, J. L. (2002). Evolution of South American spiny rats (Rodentia, Echimyidae): The star-phylogeny hypothesis revisited. *Molecular Phylogenetics and Evolution*, *25*(3), 455–464.

Lewis, J. F., & Draper, G. (1990). Resume of the geology of Hispaniola. In *Field Guide, Ninth Caribbean Geological conference* (pp. 5–31). Santo Domingo, Dominican Republic.

Lister, A. M. (1996). Dwarfing in island elephants and deer: Processes in relation to time of isolation. *Symposium of the Zoological Society of London*, *69*, 277–292.

Losos, J. B., & Queiroz, K. De. (1997). Evolutionary consequences of ecological release in Caribbean *Anolis* lizards. *Biological Journal of the Linnean Society*, *61*(4), 459–483.

Losos, J. B., & Ricklefs, R. E. (2009). Adaptation and diversification on islands. *Nature*, *457*(7231), 830–6.

MacPhee, R. D. E. (2011). Basicranial Morphology and Relationships of Antillean Heptaxodontidae (Rodentia, Ctenohystrica, Caviomorpha). *Bulletin of the American Museum of Natural History*, *363*, 3–70.

MacPhee, R. D. E., & Flemming, C. (2003). A Possible Heptaxodontine and Other Caviidan Rodents from the Quaternary of Jamaica. *American Museum Novitates*, *3422*(1), 1.

Macphee, R. D. E., Iturralde-Vinent, M., & Gaffney, E. S. (2003). Domo de Zaza, an Early Miocene Vertebrate Locality in South-Central Cuba , with Notes on the Tectonic Evolution of Puerto Rico and the Mona Passage. *American Museum Novitates*, *3394*(3394), 1–43.

McFarlane, D. A. (1999). Late Quaternary Fossil Mammals and Last Occurrence date from Cave at Barahona, Puerto Rico. *Caribbean Journal of Science*, *35.3*, 238–248.

McFarlane, D. A. (2000). New Specimens of Late Quaternary Extinct Mammals from Caves in Sanchez Ramirez Province , Dominican Republic., *36*(1), 163–166.

McKenna, M. C., & Bell, S. K. (1997). *Classification of mammals above the species level*. New York: Columbia University Press.

Meyer, M., & Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*, *5*(6).

Miller, G. S. (1916). Bones of mammals from indian sites in Cuba and Santo Domingo. *Smithsonian Miscellaneous Collections*, *66*(12), 1–10.

Miller, G. S. (1922). Remains of mammals from caves in the Republic of Haiti. *Smithsonian Institution*, *2707*.

Miller, G. S. (1930). Three small collections of mammals from Hispaniola. *Smithsonian Miscellaneous Collections*, *82*(15), 1–10.

Miller, M. A., Pfeiffer, W., & Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *Gateway Computing Enviroments Workshop (GCE)*, 1–8.

Millien, V. (2006). Morphological evolution is accelerated among island mammals. *PLoS Biology*, *4*(10), 1863–1868.

Morgan, G. S. (1989). *Geocapromys thoracatus*, *American Society of Mammalologists*, *52*(2), 1–5.

Nicholson, K. E., Glor, R. E., Kolbe, J. J., Larson, A., Hedges, S. B., & Losos, J. B. (2005). Mainland colonization by island lizards. *Journal of Biogeography*, *32*(6), 929–938.

Pinto, G., Mahler, D. L., Harmon, L. J., & Losos, J. B. (2008). Testing the island effect in adaptive radiation: rates and patterns of morphological diversification in Caribbean and mainland *Anolis* lizards. *Proceedings. Biological Sciences / The Royal Society*, *275*(1652), 2749–2757.

Rambaut, A., Suchard, M. A., Xie, D., & Drummond, A. J. (2014). Tracer v1.6. *University of Edinburgh, Edinburgh, UK*.

Ray, C. E. (1964). The Taxonomic Status of *Heptaxodon* and Dental Ontogeny in *Elasmodontomys* and *Ablyrhiza* (Rodentia:Caviomorpha). The Museum, 1964. *The Museum*.

Ray, C. E. (1965). The relationships of Quemisia gravis (Rodentia: Heptaxodontidae).

Proceedings of the Royal Society B: Biological Sciences, *275*, 923–928.

Renaud, S., & Millien, V. (2001). Intra- and interspecific morphological variation in the field mouse species Apodemus argenteus and A. speciosus in the Japanese archipelago: the role of insular isolation and biogeographic gradients. *Biological Journal of the Linnean Society*, *74*, 557–569.

Ronquist, F., & Huelsenbeck, J. P. (2003). MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, *19*, 1572–1574.

Sax, D. F., & Gaines, S. D. (2008). Species invasions and extinction: The future of native biodiversity on islands. *Proceedings of the National Academy of Sciences*, *105*(Supplement 1), 11490–11497.

Smith, C. I., Chamberlain, A. T., Riley, M. S., Cooper, A., Stringer, C. B., & Collins, M. J. (2001). Not just old but old and cold? *Nature*, *410*(April), 771–772.

Sondaar, P. Y. (1977). Insularity and its effect on mammal evolution. In *Major patterns in vertebrate evolution* (pp. 671–7007). Springer US.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*, 1312–1313.

Tate, G. H. H. (1948). Notes on the Haitian Hutia, *Plagiodontia*, and Extinct Related Genera. *Journal of Mammalogy*, *29*(2), 176–178.

Turvey, S. T., Grady, F. V., & Rye, P. (2006). A new genus and species of “giant hutia” (*Tainotherium valei*) from the Quaternary of Puerto Rico: An extinct arboreal quadruped? *Journal of Zoology*, *270*(4), 585–594.

Turvey, S. T., & Helgen, K. M. (2008). *Geocapromys columbianus*. *The IUCN Red List of Threatened Species*, (e. T9004A12949470).

Upham, N. S., & Borroto-Páez, R. (2017). Molecular phylogeography of endangered Cuban hutias within the Caribbean radiation of capromyid rodents. *Journal of Mammalogy*, *98*(4), 950–963.

Van der Geer, A., Lyras, G., de Vos, J., & Dermitzakis, M. (2010). *Evolution of Island Mammals: Adaptation and Extinction of Placental Mammals on Islands*. *Evolution of Island Mammals: Adaptation and Extinction of Placental Mammals on Islands*. John Wiley & Sons.

Van Valen, L. (1965). Morphological Variation and Width of Ecological Niche. *The Amercian Naturalist*, *99*(908), 377–390.

Westbury, M., et al. (2017). A mitogenomic timetree for Darwin’s enigmatic South American mammal *Macrauchenia patachonica*. *Nature Communications*, *8*(May), 15951.

Wilkinson, M. J., Marshall, L. G., Lundberg, J. G., & Kreslavsky, M. H. (2010). Megafan environments in northern South America and their impact on Amazon Neogene aquatic ecosystems. In C. Hoorn & F. P. Wesselingh (Eds.), *Amazonia landscape and species evolution: a look into the past* (pp. 162–184). Oxford: Wiley-Blackwell.

Woods, C. A. (1982). The history and classification of South American Hystricognath rodents:

reflections on the far away and long ago. In M. A. Mares & H. H. Genoways (Eds.), *Mammalian Biology in South America* (pp. 377–392). Pittsburgh: University of Pittsburgh.

Woods, C. A. (1993). Suborder Hystricognathi. In D. E. Wilson & D. M. Reeder (Eds.), *Mammal Species of the World: A Taxonomic and Geographic Reference* (pp. 771–806). Washington, DC: Smithsonian Institution Press.

Woods, C. A. (1996). The land mammals of Puerto Rico and the Virgin Islands. *Annals of the New York Academy of Sciences*, *776.1*, 131–148.

Woods, C. A., & Hermanson, J. W. (1985). Myology of hystricognath rodents: an analysis of form, function and phylogeny. In W. P. Luckett & J. L. Hertenbrger (Eds.), *Evolutionary relationships among rodents: a multidisiplinary analysis* (pp. 515–548). New York: Plenum.

Woods, C. A., & Kilpartrick, C. W. (2005). Infraorder Hystricognathi. In *Mammal Species of the World: A Taxonomic and Geographic Reference* (pp. 1538–1600).

Woods, C. A., & Kilpatrick, C. W. (2005). Family Echimyidae. In D. E. Wilson & D. M. Reeder (Eds.), *Mammal Species of the World: A Taxonomic and Geographic Reference* (3rd ed., pp. 1575–1592). Washington, DC: Johns Hopkins University Press.

Woods, C. A., & Sergile, F. E. (2001). *Biogeography of the West Indies: Patterns and Perspective, Second Edition: Past Present and Future*.

Appendix Chapter 6. Evolutionary histories and biogeographic origins of the Greater Antillean rodent fauna

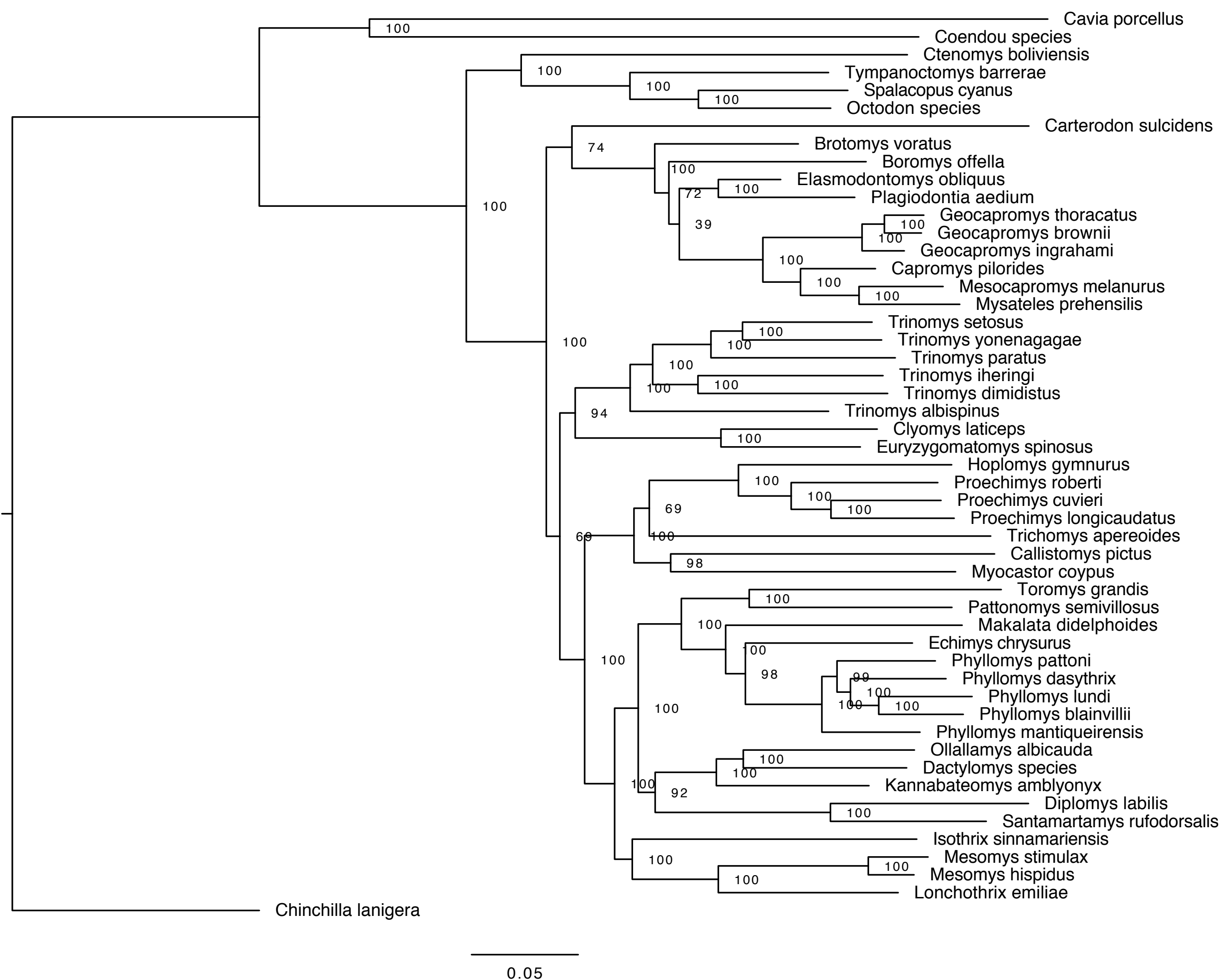


Figure 1a. Maximum likelihood phylogeny produced using RAxML v.8 (Stamatakis 2014) which was implemented in CIPRES Science Gateway v.3.3 (Miller, Pfeiffer, & Schwartz, 2010).

Table 1a. Genes used to generate baits for capture enrichment

Gene	Length (bp)	NCBI accession	Species
apolipoprotein B (apoB)	1155	JX515310.1	*Proechimys quadruplicatus*
recombination activating gene 1 (RAG1)	1072	EU313333.1	*Proechimys simonsi*
von Willebrand factor (vWF)	1150	AJ849311.1	*Echimys didelphoides*
interphotoreceptor retinoid binding (IRBP)	1245	JF297775.1	*Phyllomys* sp.
growth hormone receptor (GHR)	798	JX515330.1	*Trinomys setosus*
mitochondrial genome	16816	HM544128.1	*Proechimys longicaudatus*

Table 2a. Sequence data used in post-sequencing data processing as reference sequences in read mappings

Gene	Species	NCBI accession number
mitochondrion, complete genome	*Plagiodontia aedium*	KU892771.1
	Mysateles prehensilis	KU892770.1
	Chinchilla lanigera	NC021386.1
	Carterodon sulcidens	KU892752.1
	Capromys pilorides	KU892766.1
	Proechimys longicaudatus	NC020657.1
	Geocapromys brownii	KU892767.1
	Mesocapromys melanurus	KU892769.1
GHR (growth hormone receptor)	*Trinomys paratus*	JX515330.1
	Capromys pilorides	AF433950.1
	Geocapromys brownii	KM013984.1
	Geocapromys ingrahami	KM013985.1
	Mesocapromys auritus	KJ742633.1
	Mesocapromys angelcabrerai	KJ742632.1
	Mysateles prehensilis	KJ742634.1
RAG1 (recombination activating 1)	*Proechimys simonsi*	EU313333.1
	Capromys pilorides	JN633628.1
	Geocapromys ingrahami	KM013990.1
	Geocapromys brownii	KM013989.1
	Mesocapromys angelcabrerai	KJ742694.1
	Mesocapromys auritus	KJ742693.1
	Mesocapromys melanurus	KJ742691.1
	Mysateles prehensilis	KJ742696.1
irBP (interphotoreceptor retinoid-binding protein)	*Phyllomys* sp.	JF297775.1
	Capromys pilorides	KM013993.1
	Geocapromys ingrahami	KM013995.1
apoB (apolipoprotein B)	*Proechimys quadruplicatus*	JX515310.1
vWF (von Willebrand factor)	*Echimys didelphoides*	AJ849311.1
	Capromys pilorides	AJ251142.1
	Geocapromys ingrahami	KM014000.1
	Geocapromys brownii	KM013999.1
	Mesocapromys auritus	KJ742623.1
	Mesocapromys angelcabrerai	KJ742622.1

Table 3a. Mapping parameters used in CLC genomics work bench

Mapping parameter	Length fraction	Similarity fraction
default	0.5	0.8
L	0.7	0.7
M	0.8	0.8
M1	0.85	0.85
H	0.9	0.9
H1	0.95	0.95
XH	1	1

Table 3a. Echimyid outgroup taxa for whole mitochondrial genome and nuclear genes

Species	Genbank accession number					
	Mitogenome	apoB	RBP3	GHR	RAG1	vWF
Trinomys paratus	KU892763	JX515306	JX515320	JX515330	X	AJ849316
Trinomys albispinus	KU892761	KM013982	KM013997	KM013987	KM013992	KM014003
Plagiodontia aedium	KU892771	KM013981	KM013996	KM013986	KM013991	KM014002
Phyllomys blainvillii	KU892756	JX515302	JF297781	JX515331	JX515323	AJ849314
Myocastor coypus	KU892780	JX515307	JN414814	X	JN414955	AJ251140
Mesomys hispidus	KU892787	JX515305	X	JX515335	EU313322	KF590671
Makalata didelphoides	KU892782	JX515303	JF297756	JX515332	EU313314	AJ849311
Isothrix sinnamariensis	KU892785	JX515309	X	KF590685	EU313312	AJ849309
Hoplomys gymnurus	KU892779	X	X	JN414758	JN633632	JN415080
Echimys chrysurus	KU892781	JX515301	AJ427247	JX515533	EU313303	AJ251141
Clyomys laticeps	KU892753	JX515304	JX515316	JX515326	X	AJ849306
Cavia porcellus	NC_000884.1	AJ399524	AJ427248	AF433930	JN414957	AJ224663
Capromys pilorides	KU892766	KM013977	KM013993	KM013983	JX515322	KM013998
Trinomys yonenagagae	KU892765	KM013982	X	JX515328	X	AJ849318
Trinomys setosus	KU892764	X	X	JX515329	X	AJ849317
Euryzygomatomys spinosus	KU892755	X	JX515315	JX515327	KJ742680	AJ849319
Chinchilla lanigera	NC_021386.1	X	AF297280	AF332036	KF590658.1	AJ238385.1
Octodon lunatus/degus	HM544134.1	X	AM050863	X	KJ742676	AJ238386
Trinomys iheringi	KU892762	X	X	KF590695	EU313337	KF590677
Carterodon sulcidens	KU892752	KY303653	KY303657	KY303655	KJ742678	
Thrichomys apereoides	KU892773	X	JX515319	JX515325	EU313334	KY303660
Toromys grandis	KU892784	X	X	KF590694.1	EU313336.1	KF590676.1
Trinomys dimidiatus	JX312694.1	X	X	X	KJ742682.1	KJ742620.1
Tympanoctomys barrerae	HM544132.1	X	X	KX646530.1	X	X
Coendou species	NC_021387.1	X	X	X	X	X
Santamartamys rufodorsalis	KU892777	X	X	X	X	X
Spalacopus cyanus	HM544133.1	X	X	X	X	X
Pattonomys semivillosus	KU892783	X	X	X	KJ742687.1	KJ742616.1
Proechimys roberti	KU892772	X	X	X	X	X
Phyllomys pattoni	KU892760	X	JF297803	KJ742642	KJ742690	JF297754
Phyllomys mantiqueirensis	KU892759	X	JF297769.1	X	X	JF297720.1
Phyllomys lundi	KU892758	X	JF297770.1	X	X	JF297721.1
Proechimys longicaudatus	HM544128.1	X	X	FJ855217	KJ742681	KJ742619
Phyllomys dasythrix	KU892757	X	JF297757.1	KJ742641.1	KJ742689.1	JF297709.1

Table 3a. (continued) Echimyid outgroup taxa for whole mitochondrial genome and nuclear genes

Species	Genbank accession number					
	mitogenome	apoB	RBP3	GHR	RAG1	vWF
Proechimys cuvieri	KU892778	X	JX515318.1	KF590693.1	KF590665.1	KF590675.1
Olallamys albicauda	KU892774	X	X	KF590691.1	X	KF590673.1
Mesomys stimulax	KU892788	X	X	KJ742630.1	KJ742674.1	KJ742618.1
Lonchothrix emiliae	KU892786	X	X	X	X	X
Kannabateomys amblyonyx	KU892775	X	X	X	X	AJ849310.1
Dactylomys species	KU762015	X	X	KF590681.1	EU313300.1	KF590667.1
Ctenomys species	HM544130.1	JN414055.1	JN414816.1	FJ855214.1	JN414961.1	JN415078.1
Callistomys pictus	KU892754	KY303652.1	KY303656.1	KY303654.1	KY303658.1	KJ742614.1
Mesocapromys melanurus	KU892769	X	X	X	KJ742691	X
Mysateles prehensilis	KU892770	KM013980	X	X	X	X
Geocapromys brownii	KU892767	KM013978	X	KJ742644	KJ742692	KJ742621
Geocapromys ingrahami	KU892768	KM013979	KM013995	KM013985	KM013990	KM014000

Table 4a. Substitution models chosen for the partitioned dataset

Subset	Best model	Subset Paritions
		Mitochonrial genome
1	GTR+G	rrnL, rrnS, trnC, trnD, trnF, trnH, trnI, trnK, trnL1, trnL2, trnM, trnN, trnP, trnQ, trnR, trnS2, trnV, trnW, trnY
2	GTR+G	atp6, atp8, cob, cox3, nad1, nad2, nad3, nad4a, nad4l, nad5, nad6
3	GTR+G	cox1, cox2, trnA, trnE, trnG, trnS1, trnT
		apoB (apolipoprotein B)
1	HKY+G	Gene
		RAG1 (recombination activating gene 1)
1	K80+G	Gene
		vWF (von Willebrand factor)
1	K80+G	Gene
		IRBP (interphotoreceptor retinoid binding)
1	HKY+G	Gene
		GHR (growth hormone receptor)
1	K80+G	Gene

Table 5a. Fossil constraints and priors used in divergence dating analysis

Node	Fossil and Deposit Locality	Age (Ma)	Prior distribution
Cavioidea-Erethizontoidea	Minimum: *Andemys termasi* Tinguiririrican Maximum: *Canaanimys* and *Cachiyacuy* Barrancan	31.3 - 45.9	logNormalPrior mean="5.4" stdev="0.8" offset="31.3"
Octodontidae	Minimum: *Pseudoplateomys innominatus* Huayquerian Maximum: *Acarechimys* Laventan	6.8 - 11.8	logNormalPrior mean="1.846" stdev="0.8" offset="6.0"
Trinomys and Clyomys + Euryzygomatomys	*Theridomysops parvulus* Huayquerian	6.0 - 11.8	logNormalPrior mean="2.141" stdev="0.8" offset="6.0"
Echimyini	*Maruchito trilofodonte* Colloncuran	15.7 - 24.2	logNormalPrior mean="4.15" stdev="0.8" offset="15.7"
Myocastorini	*Pampamys emmonsae* Huayquerian	6.0 - 11.8	logNormalPrior mean="2.141" stdev="0.8" offset="6.0"

Chapter 7. Discussion and conclusions

7.1 Overview

The following chapter will summarize and critically discuss the four data chapters included in this thesis. The aims set out in the introductory chapter are also discussed and critically assessed. The final section of this chapter provides an overall summary of this thesis and details suggestions for future research on the subjects covered by this thesis.

7.2 Palaeogenomic analyses reveal recent intra-island evolutionary radiation in the extinct Caribbean “island shrew” (Nesophontes)

Nesophontes is an entirely extinct family, endemic to the insular Caribbean. Recent studies have uncovered the evolutionary relationships between this genus and the Caribbean’s other lipotyphian insectivore Solenodon (Brace et al. 2016). As part of this thesis intra-island evolutionary relationships were explored in Hispaniolan species *N. zamicrus* and *N. paramicrus*. This study was fortunate to have access to sequence data from this extinct genus obtained by earlier studies. The resulting data were used to generate a reference sequence with which to map the raw reads produced from NGS. The use of this closely related reference sequence was necessary for the prevention of mapping ascertainment bias and generation of high coverage whole mitochondrial genomic sequence data. This study found that morphologically discrete species fall within the expected sequence divergence for lipotyphian insectivore sister species, and are the result of recent Pleistocene divergence. This chapter focuses on the island of Hispaniola, and suggests that this Greater Antillean island not only acts as a shelter preserving ancient lineages such as the Hispaniolan *Solenodon*, but is also a hotspot for biodiversity, where recent speciation events were still occurring in mammalian groups such as *Nesophontes*, until their extinction in the Holocene.

7.3 Ancient DNA of the extinct Jamaican monkey Xenothrix reveals extreme insular change within a morphologically conservative primate radiation

This chapter utilised ancient DNA techniques to extract and sequence endogenous DNA from the extinct Greater Antillean primate *Xenothrix mcgreogori*. This project faced issues due to the degraded nature of genetic material extracted from sub-fossils found in the tropics and the rare nature of *Xenothrix* remains. DNA was extracted and sequenced using NGS techniques, enhanced by the use of target capture enrichment baits. The resulting data was used in phylogenetic analysis to reveal the mainland ancestor of this species, and through divergence date analysis uncover details of the colonisation histories of the Greater Antillean primate fauna. This study found that the unusual *Xenothrix mcgreogori* is most closely related to the species rich, but morphologically conservative South American platyrrhine group callicebines. Further demonstrating the results of

insular evolution in a rare island primate. Additionally divergence dated analysis suggests that the Greater Antillean primate taxa do not represent a monophyletic group, as has previously been proposed, but in fact past species diversity is the result of more than one colonisation event.

7.4 Phylogeographic analysis of sympatric Cuban hutia reveal parallel patterns evolution in a mammalian insular radiation

Cuban hutia are one of the few examples of a mammalian island radiation, however the extinction of many species and the fragmented and threatened nature of small populations of remaining survivors, mean that until recently very little molecular work had been carried out on the group. In order to combat these issues this study utilised museum collections that allowed us to extract DNA form individuals of known localities from specimens collected before local extinctions and population declines. Several new studies were published during data collection and analysis for this chapter (Fabre, Vilstrup, Raghavan, Sarkissian, Willerslev, Douzery & Orlando 2014; Fabre et al. 2017; Upham & Borroto-Páez 2017). The generation of molecular data by other studies increased the scope of this study allowing more complete phylogenies to be incorporated and providing testable hypotheses relating to evolutionary relationships and colonisation histories. One such study suggested a phylogeographic relationship between eastern and western Cuban *Capromys* populations (Upham & Borroto-Páez 2017). Our analysis confirmed that Cuban Capromys are separated in to two clades associated with eastern and western Cuba. In the genus *Mesocapromys* molecular analysis suggests that previous morphology-based studies may have over estimated diversity. A comparison between the evolutionary histories in both genera sheds light on shared Late Pleistocene divergences, and insular adaptive radiations and convergence. The results of this study have implications for both taxonomy and conservation of threatened species. The conservation of *Mesocapromys* species is particularly affected by this result, as populations previously accepted as discrete species and currently assessed as critically endangered or possibly extinct may represent a single genetic species.

7.5 Evolutionary histories and biogeographic origins of the Greater Antillean rodent fauna

The aim of this chapter was to include extinct sub-fossil Greater Antillean caviomorph rodents, previously categorised within several family or sub-family level groups, in a single phylogenetic analysis, in order to better understand the biogeography and evolutionary histories of this diverse group. Previous studies have used morphology to estimate the closest mainland ancestor of discrete morphologically similar groups, but so far this has been done without the use of molecular data (e.g.: MacPhee 2011; Woods & Sergile 2001). We utilised NGS and target capture enrichment to extract DNA from extinct sub-fossil species. Due to the fact that most the mainland and insular sister taxa of these species is unknown, mapping ascertainment bias was an issue faced by this project during the mapping process. We used an iterative mapping technique utilising multiple

reference sequences and mapping parameters to account for this bias. Our analysis reveals that the Greater Antillean caviomorphs constitute a single monophyletic group as opposed to multiple independent colonisers with different mainland ancestors. The taxonomically elusive extinct giant hutia *Elasmodontomys* was placed as sister to extant Hispaniolan hutia *Plagiodontia* by this analysis. Extinct heteropsomyines with their ancestral morphology were recovered as stem taxa. The results of this study have taxonomic implications and suggest that Greater Antillean caviomorphs previously separated in to distinct sub-family and family level groups should be amalgamated within the family Capromyidae.

7.6 Thesis aims

The four data chapters that make up this thesis have the combined objective to resolve unanswered questions in Caribbean insular mammalian evolutionary biology, by using specialist ancient DNA techniques. Four key aims were set out at the start of this project the section that follows outlines these aims and details how each one has been approached during this project:

i) To utilize recently-developed aDNA techniques, including NGS and target capture enrichment, to extract and sequence highly degraded DNA from museum and zooacheological specimens from the tropical Caribbean.

One of the major hurdles for this project was the poor preservation of genetic material in specimens used for aDNA extraction. Study species included in this thesis were all from the tropical location of the insular Caribbean, as discussed in the introduction these conditions are not ideal for the extraction and sequencing of endogenous DNA. In order to combat the issues associated with poor DNA preservation, this study utilised NGS and target capture enrichment to increase the depth of coverage for read mappings to specific genetic markers. A particular success of this study was the extraction and sequencing of DNA from Caribbean sub-fossil specimens (*Xenothix mcgregori*, *Nesophontes zamicrus*, *Elasmodontomys obliquus*, *Boromys offella*, *Brotomys voratus* and *Geocapromys columbinanus*).

Although the use of NGS techniques for the sequencing of aDNA from taxa included in this study negated many of the issues associated with PCR and shotgun Sanger sequencing techniques, other complications are involved with these methods, such as mapping ascertainment bias. We tested alternative reference sequences and applied a new iterative mapping method using multiple reference sequences and a range of mapping parameters to accommodate this issue (Westbury et al. 2017).

ii) To generate a phylogenetic placement for some recently extinct Caribbean species, where the evolutionary history is currently unresolved or highly debated. The use of aDNA is the only means for generating the first genetic phylogeny including these taxa.

For living taxa included in this study, molecular data produced will contribute to redefining species relationships to better inform conservation efforts.

Molecular phylogenetic analysis was conducted using maximum likelihood and Bayesian methods in order to resolve evolutionary relationships between species where previous analysis had been based on morphology alone. This study was able to uncover the mainland relatives of *Xenothrix mcgreogori* the first endemic Caribbean primate that has been included in molecular analysis. The surprising result of this analysis found that *Xenothrix* falls within the recently reassessed callicebine radiation, and is not a stem platyrrhine. This study also investigated the intra-island evolutionary relationships in *Nesophontes* species and found evidence to confirm previous morphological assessments. In extinct caviomorph rodents this study was able to include six taxa, five of which had not been subjected to molecular analysis before this study. This study provided evidence that endemic Greater Antillean caviomorph rodents form a single monophyletic group, despite previous morphology-based studies separating these taxa into family and sub-family level groups.

Cuban hutia include some of the Caribbean's few surviving mammal species, many of which are still threatened with extinction. Phylogenetic analysis produced as part of this study provides evidence for a suggested re-assessment of Cuban hutia taxonomy. An accurate taxonomy for this group critical for the continued conservation of threatened species. In particular the amalgamation of taxa previously accepted as discrete species within the genus *Mesocapromys* has important implications for conservation priorities.

iii) To use aDNA data to test long standing biogeographic hypotheses relating to the colonisation of the Caribbean by various land mammal fauna. Molecular data is imperative here, as it not only provides a statistically robust framework to identify extant sister taxa, and therefore the mainland region of origin, it also allows the use of molecular clocks to date the divergence between mainland and island species.

This thesis investigated the colonisation histories of taxa by generating Bayesian divergence-dated phylogenies for two separate groups Caribbean primates and caviomorph rodents. For both groups previous hypotheses based on morphology were over-turned by molecular data. The Greater Antillean primates had been previously put forward as a monophyletic group the result of a single colonisation event, however molecular dating analysis of *Xenothrix mcgreogori* suggests that multiple colonisation events generated the species diversity, and biogeographic pattern found in Greater Antillean primates. Almost the opposite result was found in the analysis of the Greater Antillean caviomorph rodents, which despite morphological differences leading previous studies to place species in three family or sub-family level groups, are recovered in a single monophyletic group by this study. This result suggests a single colonisation of the Caribbean by caviomorph rodents in the mid-Miocene gave rise to past and current species diversity.

iv) This study will investigate intra-island and wider inter-island relationships among species to explore evolutionary histories and the mechanisms of speciation acting on the Caribbean's endemic land mammal fauna.

This study looked an within island species relationships in *Nesophontes* and Cuban hutia. In *Nesophontes* this study was able to link recent late Pleistocene divergence to the turbulent climate and possible biogeographic barriers during this time period. Intra-island relationships were also investigated in Cuban hutia, where complex phylogeographic structure in three genera was analysed. Results suggested shared Mid-Pleistocene divergences in both clades. The alternative hypothesis of Miocene partial island submergence driven divergence was discounted.

This thesis also looked at the between island relationships among caviomorph rodent species. In the endemic Caribbean heteropsomyines, two genera one from Cuba and one from Hispaniola were analysed, the resulting phylogeny suggested that these taxa are sequentially nested rather than monophyletic with *B. voratus* as the basal lineage. Divergence-dated analysis of these taxa does not correspond to geological evidence for the mid-Eocene connection between Cuban and Hispaniolan palaeo-islands. In *Geocapromys* our analysis shows that *G. ingrahami* is basal to the Geocapromys clade with extinct species from Cuba and Little Swan Island as most closely related to the Jamaican species *G. brownii*, suggesting that the Bahamas is the origin for the radiation of this group. Cayman Island *Capromys* species were interpreted as con-specific to Cuban *Capromys* by this study due to the recent divergence date recovered in phylogenetic analysis. The recent divergence in both sub-fossil and recently extinct *Geocapromys* and *Capromys* species investigated here, suggest that southward radiation in both groups was more recent than previously thought.

7.7 Overall thesis summary, critical evaluation and future work

The chapters of this thesis set out to investigate a group of highly diverse insular mammal fauna, a group which despite its potential for contributing to investigation of key evolutionary and biogeographic questions, has been critically understudied at the molecular level due to unprecedentedly high Holocene extinct rates combined with the highly degraded nature of DNA in the sub-tropics. The section below discusses the central themes covered by this thesis, critically evaluates the outcomes of this work, and suggests how future research can continue to discover more about the evolution and biogeography of the Caribbean's remarkable insular mammals.

Ancient DNA

The species looked at in this study are united in their importance for the continued investigation of evolutionary concepts in insular mammals, but the inclusion of molecular data from these taxa in

studies such as this, is only made possible by aDNA techniques. Specialised techniques in the extraction and sequencing of highly degraded DNA have been instrumental in allowing this study to use novel molecular data, in order to answer questions about the evolution and biogeography of the Caribbean land mammal fauna.

New in-solution target capture enrichment techniques were utilised in three of the four data chapters that make up this thesis, and were vital for the inclusion of samples, which originally had very low endogenous content. Quantifying the success of target capture enrichment used by this study is made difficult by a lack of species specific reference sequences for subject species looked at by this project, and sub-genome level data used. Endogenous content was estimated on the basis of mapping results using non-species specific mitochondrial genome reference sequences, as a proportion of the total reads sequenced. This and variables associated with sequencing (such as the change in sequencing platform used) may explain why enrichment factors – a comparison between pre and post-target capture enrichment sequencing results, which takes in to account endogenous content – often represented a percentage decrease (an average of -97% across all data included in this thesis), rather than increase in this study. However target capture enrichment did result in an increase in the total number of reads mapped to the mitochondrial genome with a average of 1,412 reads mapped pre-capture and an average of 6,459 mapped post-capture.

The use of aDNA techniques can continue to increase our knowledge of the extinct and endangered Caribbean mammal fauna. Collection of whole mitochondrial genomic data as well as nuclear markers is necessary in order to generate accurate phylogenies. Further studies attempting to include additional extinct or threatened endemic Caribbean mammal taxa will also rely upon these techniques (particularly target capture enrichment) and the extensive museum collections that provide a repository for information for the study of evolution, biogeography and the conservation of threatened species.

Colonisation histories

Previous studies have suggested various routes, methods and divergence times for the colonisation of the Caribbean by endemic terrestrial mammal fauna. Two chapters in this thesis provide novel results in this topic. In the Greater Antillean extinct endemic primates this study has shown that multiple colonisation events generated Late Quaternary diversity. In caviomorph rodents evidence produced as part of this study suggests that a single colonisation may be more likely in this group.

Both colonisation events investigated here share estimated dates of divergence form mainland taxa during the Mid-Late Miocene. These results agree with other recent molecular analyses in discounting land-bridge dispersal via the Aves Ridge during the purposed 1-2 Ma interval ~35 Ma ago (GAARlandia hypothesis) (Iturralde-Vinent & MacPhee 1999). The Mid-Late Miocene was a

period of low global sea level and sea-surface currents may have contributed to over water dispersal of taxa from South American to the insular Caribbean during that time period (Wilkinson et al. 2010).

Our understanding of the inter-island colonisation histories and mainland origin of the Caribbean terrestrial mammal fauna is still far from complete, further studies inclusive other extinct taxa are now necessary to further build on work included in this thesis and other recent molecular analyses. Of particular importance is the inclusion of further extinct Caribbean primate taxa, as this may shed light on whether the Jamaican primate *Xenothix* is a monotypic lineage or if this species is part of an insular radiation. In the Greater Antillean caviomorphs, further sampling from giant hutia species is necessary to fully resolve the placement of this group within the diversity of the capromyid radiation, and to rule out the possibility of alternative coloniation histories for other members of Heptaxodontidae. For complete understanding of the capromyid radiation to be achieved further sampling of heteropsomyines, especially Puerto Rican species is needed.

Patterns of insular evolution

Due to high rates of Holocene extinction amongst the endemic mammal fauna of the insular Caribbean and the highly degraded nature of aDNA found in the tropics, little has been done to access patterns of evolution in these taxa using molecular data. This study has been able to demonstrate that Pleistocene climate change contributed to species generation in the insular Caribbean. Recent divergence dates found between taxa included in our analysis suggest that islands such as Hispaniola were not only sanctuaries for ancient relic species such as *Solenodon*, but were also hotspots for biodiversity and evolution until the recent extinction of the vast majority of newly diverging endemic mammal fauna.

This study has also been able to uncover and confirm some of the unusual results of insular evolution. The morphologically peculiar primate *Xenothrix* and caviomorph rodent *Elasmodontomys* were first analysed using molecular data by this study. The phylogenetic placement of these species was not predicted by some previous morphology-based studies. This study was able to reveal the mainland relatives of two groups of Caribbean taxa. In *Xenothrix* a taxon characterised by morphological obscurity, this species was most closely related to morphologically conservative South American callicebines (Byrne et al. 2016). This result demonstrates how insular evolution can dramatically affect colonising species. The highly diverse Greater Antillean caviomorph rodents were previously placed in multiple family or sub-family level groups with different mainland sister taxa by morphology-based studies, however molecular analyses suggest that they represent a single monophyletic group. The Greater Antillean caviomorph rodents are another example of dramatic island evolution, containing a diverse array or morphologies including island giants (McFarlane et al. 2014; Biknevicius et al. 1993). However this group also includes species that have

retained the morphology of their mainland sister taxa, providing evidence that, whilst islands often generate unusual morphology, in some circumstances they also preserve species with ancestral morphology. Like the callicebines the South American spiny rats (Echimyidea) are a species rich group. However unlike the callicebines the echimyids exhibit high levels of morphological diversity between some species and fill multiple ecotypioc niches (Galewski et al. 2005). This group can be compared to the mainland South American *Anoles*, which like their Caribbean relatives from a species rich and diverse radiation (Pinto et al. 2008).

To further assess insular evolution in Caribbean land mammals, genetic data from other fauna need to be analysed, particularly in extinct taxa where morphology-based analyses have been unable to confirm evolutionary relationships. Within the Greater Antillean endemic primates insular evolutionary relationships are still uncertain as the relatively recent divergence and colonisation data of *Xenothrix* suggests that past hypotheses made using morphological-methods may need re-assessment. Within island evolutionary relationships, such as those found in the Cuban hutia looked at as part of this study, can also be further examined further using statistical techniques to test alternative hypotheses of mechanisms of divergence.

References

Biknevicius, A.R., McFarlane, D. a. & Macphee, R.D.E., 1993. Body Size in Amblyrhiza inundata (Rodentia: Caviomorpha), an Extinct Megafaunal Rodent from the Anguilla Bank, West Indies: Estimates and Implications. *American Museum Novitates*, 3079(3079), pp.1–25.

Brace, S. et al., 2016. Evolutionary history of the Nesophontidae, the last unplaced Recent mammal family. *Molecular biology and evolution*, 33(12), p.msw186.

Byrne, H. et al., 2016. Phylogenetic relationships of the New World titi monkeys (Callicebus): first appraisal of taxonomy based on molecular evidence. *Frontiers in Zoology*, 13(10).

Enk, J.M. et al., 2014. Ancient whole genome enrichment using baits built from modern DNA. *Molecular Biology and Evolution*, 31(5), pp.1292–1294.

Fabre, P.-H. et al., 2014. Rodents of the Caribbean: origin and diversification of hutias unravelled by next-generation museomics. *Biology letters*, 10(7), p.20140266-.

Fabre, P.H. et al., 2017. Mitogenomic phylogeny, diversification, and biogeography of South American spiny rats. *Molecular Biology and Evolution*, 34(3), pp.613–633.

Galewski, T. et al., 2005. Ecomorphological diversification among South American spiny rats (Rodentia; Echimyidae): A phylogenetic and chronological approach. *Molecular Phylogenetics and Evolution*, 34(3), pp.601–615.

Iturralde-Vinent, M.A. & MacPhee, R.D.E., 1999. Paleogeography of the Caribbean region: implications for Cenozoic biogeography. *American Museum of Natural history*.

MacPhee, R.D.E., 2011. Basicranial Morphology and Relationships of Antillean Heptaxodontidae (Rodentia, Ctenohystrica, Caviomorpha). *Bulletin of the American Museum of Natural History*, 363, pp.3–70.

McFarlane, D.A., Lundberg, J. & Maincent, G., 2014. New specimens of Amblyrhiza inundata (Rodentia : Caviomorpha) from the Middle Pleistocene of Saint Barthélemy , French West Indies. *Caribbean Journal of Earth Science*, 47(December), pp.15–19.

Pinto, G. et al., 2008. Testing the island effect in adaptive radiation: rates and patterns of morphological diversification in Caribbean and mainland *Anolis* lizards. *Proceedings. Biological sciences / The Royal Society*, 275(1652), pp.2749–2757.

Upham, N.S. & Borroto-Páez, R., 2017. Molecular phylogeography of endangered Cuban hutias within the Caribbean radiation of capromyid rodents. *Journal of Mammalogy*, 98(4), pp.950–963.

Westbury, M. et al., 2017. A mitogenomic timetree for Darwin's enigmatic South American mammal *Macrauchenia patachonica*. *Nature Communications*, 8(May), p.15951.

Wilkinson, M.J. Marshall, L. G., Lundberg, J. G., & Kreslavsky, M. H., 2010. Megafan environments in northern South America and their impact on Amazon Neogene aquatic ecosystems. In C. Hoorn & F. P. Wesselingh, eds. *Amazonia landscape and species evolution: a look into the past*. Oxford: Wiley-Blackwell, pp. 162–184.

Woods, C.A. & Sergile, F.E., 2001. *Biogeography of the West Indies: Patterns and Perspective,*

Second Edition: Past Present and Future.

Appendix: Chapter 8. (a) The small and the dead: ancient DNA studies using micromammal species

Roseina Woods[1], Melissa Marr[1], Selina Brace[1,*] and Ian Barnes[1]

[1] Earth Sciences Department, Natural History Museum, Cromwell Road, London SW7 5BD, UK

* Corresponding author: s.brace@nhm.ac.uk

8a.1 Abstract

The field of ancient DNA has recently been in a state of exponential growth, largely driven by the uptake of Next Generation Sequencing (NGS) techniques. Much of this work has focused on the mammalian megafauna and ancient humans, with comparatively less studies looking at micromammal fauna, despite the potential of these species in testing evolutionary, environmental and taxonomic theories. Several factors make micromammal fauna ideally studied for aDNA extraction and sequencing. Micromammal subfossil assemblages often include the large number of individuals appropriate for population level analyses, and further, the assemblages are frequently found in cave sites where the constant temperature and sheltered environment provides favourable conditions for DNA preservation. This review looks at studies that include the use of aDNA in molecular analysis of micromammal fauna, in order to examine the wide array of questions that can be answered in the study of small mammals using new palaeogenetic techniques. This study highlights the bias in current aDNA studies and assesses the future use of aDNA as a tool for the study of micromammal fauna.

8a.2 Introduction

Techniques to enable extraction and sequencing of DNA from museum specimens, zooarchaeological and sub fossil material (Hagelberg et al. 2015) have continued to develop since the initial successful recovery of ancient DNA (aDNA) in 1984 (Higuchi et al. 1984). Whilst many of the earliest aDNA studies focused on taxonomy and resolving the phylogenetic placement of extinct species, the development of new techniques, and particularly the use of NGS, have allowed the field to expand into ancient pathogens (Bos et al. 2015; Roberts & Ingham 2008), population demographic studies (Hofreiter & Barnes 2010), conservation issues (Leonard 2008), domestication events (MacHugh et al. 2016) and palaeoecological studies (Anderson-Carpenter et al. 2011). The field of aDNA has however largely focused on mammal species (Hofreiter et al. 2001; Pääbo et al. 2004), with a strong bias towards large, charismatic species such as mammoths (Palkopoulou et al. 2015; Chang et al. 2017), and humans and their closest relatives (Dalton 2010; Fu et al. 2016; Llamas et al. 2017)(Dalton 2010; Llamas et al. 2017). This bias, however, does little to reflect real world species diversity with its plethora of small-bodied mammals. For example, half of all placental

diversity is contained within the order Rodentia, a highly speciose order principally composed of small bodied taxa (Blanga-Kanfi et al. 2009).

Here, we define small-bodied or micromammals as any mammal species under 500g. Definitions of micromammals have typically included members of several orders including Rodentia, Lipotyphla and occasionally Lagomorpha, Mustelidae and Chiroptera (Reed 2005). There are a wealth of research questions that can be addressed via the study of micromammal fauna. These species show rapid generational turnover (Martin & Palumbi 1993) which is associated with a high mutation rate (Li et al. 1987; Bromham et al. 1996; Nabholz et al. 2008) and an associated ability to swiftly adapt to new and/or changing environments (Orr 2005). Mitochondrial substitution rates within the Rodentia, for example, have been estimated to be around 20 times faster than in the Cetacea (Nabholz et al. 2008). This allows the identification of population level, biogeographic and adaptive responses to past changes in both biotic and abiotic environmental variables. Small bodied mammals typically have high dispersal abilities (Gaines & McClenaghan Jr. 1980) - a factor that facilitates frequent island colonisation events and thus further facilitates studies relating to biogeography and speciation. Micromammals are also considered key biostratigraphic indicators, providing a source of valuable proxy data with which to infer palaeoecological, palaeoclimatic and palaeogeographical change (Lister 1992; Schreve 2001; Currant & Jacobi 2001). This is due to their widespread abundance in faunal assemblages, their excellent fossil record, and their rapid rate of morphological and genetic change (Chaline et al. 1999; Abramson et al. 2009).

Whilst micromammals can make excellent model study species covering both diverse areas of interest and important concepts in evolutionary biology, palaeontology and conservation; they have been vastly under-represented in ancient DNA studies. Consequently, a review to highlight the potential of working with aDNA from small mammal is long overdue. This review catalogues past efforts to use aDNA to study micromammals and discusses prospects for the future of this field.

This review includes a meta analysis literature review of studies which utilse ancient DNA techniques to examine micromammal fauna, encompassing a range of materials from degraded museum specimens to sub-fossils (Table 1). Articles compiled for this review were extracted from ISI Web of Science, PubMed and Google Scholar in 2017 using several search terms including: “ancient DNA”, “aDNA”, “historic DNA” “palaeogenomic”, “paleogenomic”, “palaeogenetic”, “paleogenetic”, “museum” and “micromammal”, “micro mammal”, “small mammal”. We then analysed this data in terms of number of articles over time (Figure 1) and in potential taxanomic bias (Figure 2). The following sections of this article cover key trends in ancient DNA studies including micromammal fauna, as dictated by the articles assessed in this study.

8a.3 Next Generation Sequencing

The study of ancient DNA has been revolutionised by the emergence of new technologies; namely the shift from PCR and traditional Sanger sequencing to NGS (Knapp & Hofreiter 2010). This shift has impacted the field in numerous ways including the challenges faced by researchers hoping to extract endogenous genetic material from degraded specimens. Previous PCR based studies relied on designing primers, which was often a difficult task when studying extinct species with no previously recorded molecular data. The benefit of species-specific primers however is that they target both the region and species of interest, thus reducing the likelihood of sequencing contaminant DNA. In contrast, NGS shotgun approaches have no specificity, with all DNA in a sample, endogenous and contaminant, being sequenced. As an alternative to shotgun sequencing, target capture enrichment techniques can be applied prior to sequencing to increase the specificity of the sequencing process, but as with PCR, the process requires reference sequences from which to design the "baits" to capture the DNA. Even shotgun sequencing significantly benefits from a good reference genome, as *de novo* assembly of raw sequenced reads is inefficient, and particularly where the amount of data available is reduced and contigs are short. Different micromammal species have different levels of potential for these approaches. For species that are widely used in genetic studies (such as rats or mice), reference sequences are available, for other less studied, rare or extinct species, reference sequences may be unavailable, restricting analyses to either *de novo* assembly or aligning to the evolutionarily closest (presently often quite distant) sequenced relative.

8a.4 Cave sites

Micromammal remains are often found in cave environments where they form a large proportion of recovered faunal elements (Avery 1990). This may be the result of several factors including ecology (the cave was the natural habitat of the species), taphonomy (the small mammals were prey species for some cave-living avian or mammalian predator), or demography (the larger population size of small mammals simply increases the likelihood of their preservation). DNA recovery also seems to be higher in material from cave sites, possibly due to the amelioration of temperature variation which may help to slow down hydrolytic processes involved in DNA degradation, thus preserving DNA that may, outside of the cave environment, have become further fragmented (Willerslev et al. 2005; Kehlmaier et al. 2016). Many aDNA studies have taken advantage of cave site remains including Gutiérrez-garcía et al., (2014) where aDNA was successfully analysed from rodents (*Ototylomys phyllotis*) excavated from a cave site in the tropics. This further highlights that micromammals are an excellent study species for aDNA studies, particularly in places such as the tropics where DNA preservation outside of cave environments is poor.

8a.5 Resolving phylogenetic placement of extinct species

aDNA studies have been particularly useful in resolving the phylogenetic placement of rare or extinct groups or individual species. In the case of specious groups, such as insectivores and rodents, aDNA has helped resolve long-standing evolutionary questions. An example is the extinct Caribbean lipotiphyan insectivore *Nesophontes*. Brace et al., (2016) applied ancient DNA techniques to resolve the phylogenetic placement of *Nesophontes* by recovering a near-complete mitochondrial genome and 17 nuclear genes from a 750-year-old specimen. This study showed that despite their morphological differences, *Nesophontes* formed a monophylectic clade with the extant Caribbean insectivore *Solenodon*. It further highlighted that this was an ancient divergence, as the two genera shared a common ancestor more than 40 million years ago. Other aDNA studies of an extinct island endemic include the Lava mouse *Malpaisomys insularis* (Pagès et al. 2012). Here a more recent divergence was identified as this study revealed that *M. insularis* diverged from its morphologically incongruous closest mainland relative only 6.9 Ma.

8a.6 Species identification

Until the advent of ancient DNA analyses, the principle methods of identifying zooarchaeological and subfossil micromammals were via gross morphological examination and/or measurements taken from skeletal or dental elements (e.g., 36, 37)). The development of geometric morphometric methods greatly improved the sensitivity of statistical techniques for species identification but these usually incorporate some degree of error (Matthews & Stynder 2011; Adams & Otárola-Castillo 2013; Cornette et al. 2015). In addition, all morphological techniques are vulnerable to the confounding factors of physical damage and weak taxonomic signal. Many micromammal groups are highly species rich (e.g. Rodentia (Churakov et al. 2010) and rapid rates of diversification can result in genetic speciation without a concordant proliferation of morphological differences (Markova et al. 2012). In addition, some species groups show high levels of phenotypic plasticity (e.g. shrews (*Sorex*) (Fumagalli et al. 1999) and Arvicoline rodents (Chaline & Graf 1988; Chaline et al. 1999)). When high levels of *intra*-specific variation are coupled with low levels of *inter*-specific distinction, true evolutionary relationships can be obscured due to morphological overlap.

Molecular analysis of sub-fossil or museum specimens can be used to test hypotheses proposed by morphology. These studies often focus on extinct species or populations, and can provide additional data such as estimates of genetic distinction. Haring, Voyta, Däubl, & Tiunov, (2015) found that species delimitation based on genetic data from cave deposited specimens agreed with the previous morphometric analysis when examining far-eastern grey voles (Genus: *Alexandromys*). A further aDNA analysis of zooarchaeological rabbit remains in south-west America (Yang et al. 2005) also found agreement between genetic identification and morphological specifications. Interestingly, though, the genetic data further revealed the presence of the snowshoe hare (*Lepus americanus)*, a

species previously unknown from the region. Rodrigues et al., (2016) used both modern DNA and DNA extracted from museum specimens in order to test previous taxonomic hypotheses for the Egyptian weasel (*Mustela subplamata*). This study found no evidence for genetic distinction between *M. subplamata* and the least weasel (*M. nivalis*) contrary to previous morphology based studies. Rodrigues et al., (2016) suggested that the large size difference between these geographic populations and sexual dimorphism characteristic of *M. subplamata* are due to ecotypic rather than genetic variation. As the cost of sequencing continues to decrease, aDNA analyses become increasingly viable as a collaborative technique to identify species from archaeological sites.

8a.7 DNA barcoding

Micromammal remains, particularly rodents and insectivores, are often found in large quantities that provide limited morphological value. These large assemblages are perhaps optimally analysed using metabarcoding to retrieve viable data for ecological, taxonomic or biogeographical studies. Metabarcoding permits mass sequencing of genetic data using NGS. This technique uses universal PCR primers to mass-amplify DNA Barcodes from large collections of specimens or from environmental DNA (eDNA). The resulting sequence data can then be verified and identified to species or genus level using a database of existing DNA data. A study by Guimaraes et al., (2016) successfully used DNA barcoding to identify rodent species from a variety of differently preserved specimens and found that the method correctly identified 75.9% of museum samples and 81.8% of species from owl pellets.

8a.8 Owl pellets

Recent and palaeontological pellets from raptors, particularly owls, can provide a valuable source of micromammal skeletal material. Recent pellets can provide a useful, non-invasive method to sample current populations (Avenant 2005), whilst palaeontological pellets can provide a temporal snapshot of the micromammal fauna from a specific place and time (Reed 2005). Unfortunately much of the morphological data from pellets is lost leaving molecular methods as the most feasible means to address ecological, biogeographical and taxonomic questions. Pellets are commonly used to find and collect genetic material for extant micromammal populations as they can provide large numbers of individuals and a variety of species. They can additionally be employed to estimate the geographic distribution of micromammals if the range of the predator species is known. An initial study by Taberlet & Fumagalli (1996) successfully sequenced micromammal DNA from owl pellets, but found that the DNA quality was poor. A second study that applied specific aDNA techniques further suggested that extracting DNA from even the most recent specimens recovered from owl pellets warranted an aDNA approach (Poulakakis et al. 2005). It therefore seems likely that the continued development of aDNA techniques will only serve to increase the use of owl pellets as a valuable source of genetic data for the study of micromammal fauna.

8a.9 Biogeographic hypotheses and island micromammals

Micromammals are often adept dispersers and can therefore be found on many island systems (Gaines & McClenaghan Jr. 1980; Amori et al. 2008). As extinction rates on islands are higher than the mainland (Ricklefs & Bermingham 2008), aDNA studies play a pivotal role in our understanding of extinct or rare island micromammal evolution. Examples of studies that have used ancient DNA to examine island micromammals include: Caribbean rodents (Brace et al. 2015; Fabre, Vilstrup, Raghavan, Sarkissian, Willerslev, Douzery, Orlando, et al. 2014) and insectivores (Brace et al. 2016; Turvey et al. 2016), the Easter island Pacific rat (*Rattus exulans*) (Barnes et al. 2006), and the Christmas Island rat (*Rattus macleari*) (Wyatt et al. 2008). A further aDNA study that examined the Orkney vole (*Microtus arvalis orcadensis*) explored colonisation dynamics, founder effects and genetic drift, all of which are associated with island populations (Martinkova et al. 2013).

8a.10 Interaction with Humans

Using aDNA to explore our own human past history, typically generates significant interest (Pickrell & Reich 2014). However, reconstructing historic human migrations can be a difficult task. Early PCR based studies were plagued by contamination from modern human samples (Hofreiter et al. 2001) and whilst NGS techniques largely circumvent the problem, it remains relevant and validation of authenticity is required. When sample availability and permissions are also taken into account, an alternative approach to examining past human movements can be appealing. Due to their often-close interactions with humans, micromammals can make a useful proxy. Some species are domesticates (e.g.: rabbits (Carneiro et al. 2014)), others have adapted to live alongside us (e.g.: house mice (Jones et al. 2012)) to exploit the benefits we inadvertently provide, such as food or shelter, whilst humans have intentionally or accidentally transported and introduced other micromammals to new habitats (Clout & Russell 2008). Barnes et al., (2006) analysed aDNA from the Pacific rat (*Rattus exulans*), a rodent intentionally transported by ancient Polynesians as a food item, in order to study the colonisation route of ancient humans in the Pacific.

8a.11 Population genetics

The use of ancient DNA in population level genetic and genomic studies can be somewhat problematic. Large numbers of individuals and/or markers are often required to fully resolve population level processes and this is often difficult or impossible to achieve. Limiting factors in aDNA population level genetic analyses include availability/access to specimens and the variable quality of the DNA across samples. In contrast to many larger bodied species micromammals are often found at high population densities (Withers et al. 2016), increasing the chances of access to large, well preserved, sample sizes. An aDNA population level study that utilised the availability of

small mammal remains looked at ~80 samples of Gotland hares (Ahlgren et al. 2016). The large sample size permitted a temporal population study that compared Mesolithic and Neolithic hares from the island with mainland and other source populations, revealing extinction and colonisation events and the possibility of human transportation.

Chan et al (2006) introduced a novel approach, Approximate Bayesian Computation (ABC) and serial coalescent analyses to examine past population structure, specifically a historic population bottleneck in the tuco-tucos (*Ctenomys sociabillis*). The species, a social subterranean rodent, had previously been identified (Hadly et al., 2003) as having almost no genetic variation (mitochondrial gene cytB) over the last 1,000 years. A further study (Chan et al. 2005) looked at samples from 10,000-3,000 years ago and found far higher levels of variation, suggesting a severe population decline and a bottleneck. It was to this bottleneck that Chan et al (2006) applied the ABC coalescent method. Studies such as these were amongst the first to show that aDNA could improve our estimation of the timing of events, such as population declines, allowing greater understanding of past populations and even the ability to inform practices for current populations

8a.12 Conservation

The use of ancient DNA data to inform conservation management and planning (termed conservation palaeogenetics; (Leonard 2008) or, alternately, conservation archaeogenetics; (Hofman et al. 2015) is an emerging field which holds much promise for wildlife conservation. Heterochronous aDNA datasets allow comparison of contemporary levels of genetic diversity, populations size, population structure and geographical distributions with those that existed in the past (Leonard 2008; Hofman et al. 2015). This potentially allows conservation biologists to design and implement appropriate management programs based on data that is not biased by observations based only on extant populations that may be severely affected by human activity. While micromammals comprise some of the most globally abundant and widespread species (*e.g Microtus spp;* Jaarola et al. 2004), they also contain a number of rare niche specialists (*e.g* star nosed moles *Condylura cristata* (Catania 2005)). To date, few conservation palaeogenetics studies have been undertaken which focus on these.

Studies of Caribbean hutias (Brace et al. 2012; Fabre, Vilstrup, Raghavan, Sarkissian, Willerslev, Douzery, Orlando, et al. 2014) and solenodons (Samuel T Turvey et al. 2016) have shown that aDNA techniques can be used to analyse degraded museum and zooacheological specimens to examine living, highly endangered or rare species when sampling from wild populations is impractical. Endangered micromammals due to their small size can be particularly elusive in the wild. Museum samples can offer an alternative genetic resource to assess population structure, identify potential source populations and species boundaries to inform conservation efforts for endangered micromammal taxa (Leonard 2008; Shepherd & Lambert 2008). For example, DNA

extracted from museum specimens of the common hamster (*Cricetus cricetus)* allowed past levels of gene flow and population structure to be determined and this data subsequently formed the basis of population augmentation schemes in areas of the species range where it has suffered major declines (Smulders et al. 2003).

8a.13 Climate change

Due to biological factors, propensity to preserve well and their abundant presence in many archaeological and palaeontological sites, micromammals can act as excellent proxy indicators for past climate change. While early studies were constrained to use only presence-absence data (e.g (Lister 1992; Currant & Jacobi 2001; Schreve 2001)), more recent research has been able to take advantage of ancient DNA evidence (e.g (Prost et al. 2010)). A further key factor in relation to climate change studies is that prior to the large-scale, Neolithic, transition to agriculture (Ammerman & Luigi 2014), micromammals are free from the influence of anthropogenic activity. This allows the confounding factor of human-mediated effects to be removed. For example, Brace et al (2012) analysed collard lemming (*Dicrostonyx torquatus*) from Belgium, over a 40,000 year time period (10-50 Kyr ago). They observed dramatic loss in genetic diversity during this period and identified a series of extinction and recolonisation events which could be exclusively related to abrupt climatic change.

8a.14 Conclusion, future directions

Molecular analysis of sub fossil material, zooacheological skeletal remains and degraded museum specimens has transformed how we explore past histories, and has the potential to continue to do so as new techniques develop. Whilst many studies continue to apply the latest techniques to charismatic species such as extinct mega fauna and humans, groups such as micromammals conversely continue to remain an understudied and untapped information resource. This is despite the large numbers of specimens available, that due to cave environments, often exhibit superior preservation. We identify micromammals as an important future direction for the field of ancient DNA. Our study has further highlighted that even as an understudied group, there exists a bias towards rodent species in aDNA studies of micro-mammals. This bias could be a result of the widespread and speciose nature of the order Rodentia, or availability of specimens and close reference sequences. Many other micromammal species, such as lipotiphyan insectivores, remain drastically understudied in comparison, despite their potential to resolve taxonomy and provide information on past climates and environments. A further interesting area of focus would be the utilisation of micromammal assemblages to look at the response of small mammals to past climate change, that could also be applied to better inform our inferences from studies of present day threats to individual species and biodiversity. As ancient DNA studies continue to move away from

solely species identification and taxonomy and towards answering wider evolutionary questions, the inclusion of micromammals will become increasingly important.

References

Abramson, N.I., Lebedev, V. S., Tesakov, A. S., & Bannikova, A. A., 2009. Supraspecies relationships in the subfamily Arvicolinae (Rodentia, Cricetidae): An unexpected result of nuclear gene analysis. *Molecular Biology*, 43(5), pp.834–846.

Adams, D.C. & Otárola-Castillo, E., 2013. Geomorph: An r package for the collection and analysis of geometric morphometric shape data. *Methods in Ecology and Evolution*, 4(4), pp.393–399.

Ahlgren, H. et al., 2016. Multiple prehistoric introductions of the mountain hare (*Lepus timidus*) on a remote island, as revealed by ancient DNA. *Journal of Biogeography*, 43(9), pp.1786–1796.

Ammerman, A.J. & Luigi, L.C., 2014. *The Neolithic transition and the genetics of populations in Europe*, Princeton Univ. Press.

Amori, G., Gippoliti, S. & Helgen, K.M., 2008. Diversity, distribution, and conservation of endemic island rodents. *Quaternary International*, 182(1), pp.6–15.

Anderson-Carpenter, L.L. et al., 2011. Ancient DNA from lake sediments: Bridging the gap between paleoecology and genetics. *BMC Evolutionary Biology*, 11(1), p.30. Available at: http://bmcevolbiol.biomedcentral.com/articles/10.1186/1471-2148-11-30.

Avenant, N.L., 2005. Barn owl pellets : A useful tool for monitoring small mammal communities ? *Belgian Journal of Zoology*, 135, pp.39–43.

Avery, D., 1990. Holocene climatic change in southern Africa: the contribution of micromammals to its study. *South African Journal of Science*, 86(January 1990), pp.407–412. Available at: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Holocene+climatic+change+in+southern+Africa:+the+contribution+of+micromammals+to+its+study#0.

Avery, D.M., 1998. An assessment of the Lower Pleistocene micromammalian fauna from Swartkrans members 1–3, Gauteng, South Africa. *Geobios*, 31(3), pp.393–414.

Barnes, S.S., Matisoo-Smith, E. & Hunt, T.L., 2006. Ancient DNA of the Pacific rat (*Rattus exulans*) from Rapa Nui (Easter Island). *Journal of Archaeological Science*, 33(11), pp.1536–1540.

Blanga-Kanfi, S. et al., 2009. Rodent phylogeny revised: analysis of six nuclear genes from all major rodent clades. *BMC evolutionary biology*, 9, p.71.

Bos, K.I. et al., 2015. Parallel detection of ancient pathogens via array-based DNA capture. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 370(1660), p.20130375. Available at: http://rstb.royalsocietypublishing.org/content/370/1660/20130375.long.

Brace, S., Thomas, J.A., et al., 2016. Evolutionary history of the Nesophontidae, the last unplaced Recent mammal family. *Molecular biology and evolution*, 33(12), p.msw186. Available at: http://www.ncbi.nlm.nih.gov/pubmed/27624716.

Brace, S. et al., 2012. Population history of the Hispaniolan hutia *Plagiodontia aedium* (Rodentia: Capromyidae): Testing the model of ancient differentiation on a geotectonically complex Caribbean island. *Molecular Ecology*, 21(9), pp.2239–2253.

Brace, S. et al., 2012. Serial population extinctions in a small mammal indicate Late Pleistocene

ecosystem instability. *Proceedings of the National Academy of Sciences*, 109(50), pp.20532–20536. Available at: http://www.pnas.org/cgi/doi/10.1073/pnas.1213322109.

Brace, S., Ruddy, M., et al., 2016. The colonization history of British water vole (*Arvicola amphibius* (Linnaeus, 1758)): origins and development of the Celtic fringe. *Proceedings of the Royal Society B: Biological Sciences*, 283(1829), p.20160130. Available at: http://rspb.royalsocietypublishing.org/lookup/doi/10.1098/rspb.2016.0130.

Brace, S. et al., 2015. Unexpected evolutionary diversity in a recently extinct Caribbean mammal radiation. *Proceedings of the Royal Society B: Biological Sciences*, 282(1807), pp.20142371–20142371. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4424637&tool=pmcentrez&rendertype=abstract.

Bromham, L., Rambaut, A. & Harvey, P.H., 1996. Determinants of rate variation in mammalian DNA sequence evolution. *Journal of Molecular Evolution*, 43(6), pp.610–621.

Carneiro, M. et al., 2014. Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. *Science*, 345(6200), pp.1074–1079. Available at: http://www.sciencemag.org/cgi/doi/10.1126/science.1253714.

Catania, K.C., 2005. Evolution of sensory specializations in insectivores. *Anatomical Record - Part A Discoveries in Molecular, Cellular, and Evolutionary Biology*, 287(1), pp.1038–1050.

Chaline, J. et al., 1999. Anatomy of the arvicoline radiation (Rodentia): palaeogeographical, palaeoecological history and evolutionary data. *Annales Zoologici Fennici*, 36(December), pp.239–267.

Chaline, J. & Graf, J.D., 1988. Phylogeny of the Arvicolidae (Rodentia): biochemical and paleontological evidence. *Journal of Mammalogy*, 69(1), pp.22–33.

Chan, Y.L. et al., 2005. Ancient DNA reveals Holocene loss of genetic diversity in a South American rodent. *Biology Letters*, 1(4), pp.423–426. Available at: http://rsbl.royalsocietypublishing.org/cgi/doi/10.1098/rsbl.2005.0354.

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), pp.451–460.

Chan, Y.L. & Hadly, E.A., 2011. Genetic variation over 10000 years in Ctenomys: Comparative phylochronology provides a temporal perspective on rarity, environmental change and demography. *Molecular Ecology*, pp.4592–4605.

Chang, D. et al., 2017. The evolutionary and phylogeographic history of woolly mammoths: a comprehensive mitogenomic analysis. *Scientific Reports*, 7(September 2016), p.44585. Available at: http://www.nature.com/articles/srep44585.

Churakov, G. et al., 2010. Rodent evolution: Back to the root. *Molecular Biology and Evolution*, 27(6), pp.1315–1326.

Clout, M.N. & Russell, J.C., 2008. The invasion ecology of mammals: A global perspective. *Wildlife Research*, 35(3), pp.180–184.

Cornette, R. et al., 2015. Specific information levels in relation to fragmentation patterns of shrew

mandibles: Do fragments tell the same story? *Journal of Archaeological Science*, 53, pp.323–330.

Currant, A. & Jacobi, R., 2001. A formal mammalian biostratigraphy for the Late Pleistocene of Britain. *Quaternary Science Reviews*, 20(16–17), pp.1707–1716.

Dalton, R., 2010. Ancient DNA set to rewrite human history. *Nature*, 465(7295), pp.148–149.

Demastes, J.W. et al., 2003. Systematics of a rare species of pocket gopher, *Pappogeomys alcorni*. *Journal of Mammalogy*, 84(2), pp.753–761.

Fabre, P.-H. et al., 2014. Rodents of the Caribbean: origin and diversification of hutias unravelled by next-generation museomics. *Biology Letters*, 10(7). Available at: http://rsbl.royalsocietypublishing.org/cgi/doi/10.1098/rsbl.2014.0266.

Fraser, M., Sten, S. & Götherström, A., 2012. Neolithic Hedgehogs (*Erinaceus europaeus*) from the Island of Gotland show early contacts with the Swedish mainland. *Journal of Archaeological Science*, 39(2), pp.229–233. Available at: http://dx.doi.org/10.1016/j.jas.2011.08.006.

Fu, Q. et al., 2016. The genetic history of Ice Age Europe. *Nature*, advance on(7606), pp.200–205. Available at: http://dx.doi.org/10.1038/nature17993.

Fulton, T.L. et al., 2013. Ancient DNA supports southern survival of Richardson's collared lemming (*Dicrostonyx richardsoni*) during the last glacial maximum. *Molecular Ecology*, 22(9), pp.2540–2548.

Fumagalli, L. et al., 1999. Molecular phylogeny and evolution of Sorex shrews (Soricidae: insectivora) inferred from mitochondrial DNA sequence data. *Molecular phylogenetics and evolution*, 11(2), pp.222–235.

Gaines, M.S. & McClenaghan Jr., L.R., 1980. Dispersal in Small Mammals. *Annual Review of Ecology and Systematics*, 11(1), pp.163–196.

Guimaraes, S. et al., 2016. A cost-effective high-throughput metabarcoding approach powerful enough to genotype ~44 000 year-old rodent remains from Northern Africa. *Molecular Ecology Resources*.

Gutiérrez-garcía, T. a et al., 2014. Ancient DNA and the tropics : a rodent ' s tale Ancient DNA and the tropics : a rodent ' s tale. *Biology letters*, (June).

Hadly, E.A. et al., 2003. Ancient Dna Evidence of Prolonged Population Persistence With Negligible Genetic Diversity in an Endemic Tuco-Tuco (Ctenomys Sociabilis). *Journal of Mammalogy*, 84(2), pp.403–417. Available at: http://www.bioone.org/doi/abs/10.1644/1545-1542%282003%29084%3C0403%3AADEOPP%3E2.0.CO%3B2%5Cn(null).

Hagelberg, E., Hofreiter, M. & Keyser, C., 2015. Introduction. Ancient DNA: the first three decades. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 370(1660), p.20130371. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4275880&tool=pmcentrez&rendertype=abstract.

Haring, E. et al., 2015. Comparison of genetic and morphological characters in fossil teeth of grey voles from the Russian Far East (Rodentia: Cricetidae: Alexandromys). *Mammalian Biology*,

80(6), pp.496–504. Available at: http://dx.doi.org/10.1016/j.mambio.2015.08.001.

Higuchi, R. et al., 1984. DNA sequences from the quagga, an extinct member of the horse family. *Nature*, pp.282–284.

Hofman, C.A. et al., 2015. Conservation archaeogenomics: Ancient DNA and biodiversity in the Anthropocene. *Trends in Ecology and Evolution*, 30(9), pp.540–549.

Hofreiter, M. et al., 2001. Ancient DNA. *Nature reviews. Genetics*, 2(5), pp.353–359.

Hofreiter, M. & Barnes, I., 2010. Diversity Lost: Are All Holarctic Large Mammal Species Just Relict Populations? *BMC Biology*, 8(1), pp.1–3. Available at: http://dx.doi.org/10.1186/1741-7007-8-46.

Jones, E. et al., 2012. Fellow travellers: a concordance of colonization patterns between mice and men in the North Atlantic region. *BMC Evolutionary Biology*, 12(1), p.35. Available at: http://www.biomedcentral.com/1471-2148/12/35.

Kalkvik, H.M. et al., 2012. Investigating niche and lineage diversification in widely distributed taxa: phylogeography and ecological niche modeling of the Peromyscus maniculatus species group. *Ecography*, 35(1), pp.54–64.

Kehlmaier, C. et al., 2016. Tropical ancient DNA reveals relationships of the extinct Bahamian giant tortoise *Chelonoidis alburyorum*. *Proceedings of the Royal Society B: Biological Sciences*, 21(16), pp.1–7.

Kerhoulas, NJ and Arbogast, BS, 2010. Molecular systematics and Pleistocene biogeography of Mesoamerican flying squirrels. *Journal of Mammalogy*, 91(3), pp.654–667.

Knapp, M. & Hofreiter, M., 2010. Next generation sequencing of ancient DNA: Requirements, strategies and perspectives. *Genes*, 1(2), pp.227–243.

Krystufek, B. et al., 2007. Phylogeography of the rare Balkan endemic Martino's vole, *Dinaromys bogdanovi*, reveals strong differentiation within the western Balkan Peninsula. *Molecular Ecology*, 16(6), pp.1221–1232.

Kuch, M. et al., 2002. Molecular analysis of an 11,700-year-old rodent midden from the Atacama Desert, Chile. *Molecular ecology*, 11(5), pp.913–924.

Leonard, J.A., 2008. Ancient DNA applications for wildlife conservation. *Molecular Ecology*, 17(19), pp.4186–4196.

Li, W.H., Tanimura, M. & Sharp, P.M., 1987. An evaluation of the molecular clock hypothesis using mammalian DNA sequences. *Journal of Molecular Evolution*, 25(4), pp.330–342.

Lister, A.M., 1992. Mammalian fossils and quaternary biostratigraphy. *Quaternary Science Reviews*, 11(3), pp.329–344.

Llamas, B., Willerslev, E. & Orlando, L., 2017. Human evolution: a tale from ancient genomes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1713), p.20150484. Available at: http://rstb.royalsocietypublishing.org/lookup/doi/10.1098/rstb.2015.0484.

MacHugh, D.E., Larson, G. & Orlando, L., 2016. Taming the Past: Ancient DNA and the Study of Animal Domestication. *Annual Review of Animal Biosciences*.

Markova, E. et al., 2012. Differentiating sibling species in the Quaternary fossil record: a comparison of morphological and molecular methods to identify *Microtus arvalis* and *M. rossiaemeridionalis* (Arvicolinae, Rodentia). *Journal of Systematic Palaeontology*, 10(3), pp.585–597.

Martin, A.P. & Palumbi, S.R., 1993. Body size, metabolic rate, generation time, and the molecular clock. *Proceedings of the National Academy of Sciences of the United States of America*, 90(9), pp.4087–4091. Available at: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=8483925&retmode=ref&cmd=prlinks.

Martinkova, N. et al., 2013. Divergent evolutionary processes associated with colonization of offshore islands. *Molecular Ecology*, 22(20), pp.5205–5220.

Matisoo-Smith, E. et al., 1997. Ancient DNA from Polynesian rats: extraction, amplification and sequence from single small bones. *Electrophoresis*, 18(9), pp.1534–1537.

Matisoo-Smith, E. & Robins, J., 2009. Mitochondrial DNA evidence for the spread of Pacific rats through Oceania. *Biological Invasions*, 11(7), pp.1521–1527.

Matthews, T. & Stynder, D.D., 2011. An analysis of two Myosorex species (Soricidae) from the Early Pliocene site of Langebaanweg (West coast, South Africa) using geometric morphometrics, linear measurements, and non-metric characters. *Geobios*, 44(1), pp.87–99. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0016699510001129.

Nabholz, B., Glémin, S. & Galtier, N., 2008. Strong variations of mitochondrial mutation rate across mammals - The longevity hypothesis. *Molecular Biology and Evolution*, 25(1), pp.120–130.

Orlando, L. et al., 2003. Napoleon Bonaparte and the fate of an amazonian rat: New data on the taxonomy of *Mesomys hispidus* (Rodentia: Echimyidae). *Molecular Phylogenetics and Evolution*, 27(1), pp.113–120.

Orr, H.A., 2005. The genetic theory of adaptation: a brief history. *Nature Reviews Genetics*, 6(2), pp.119–127. Available at: http://www.nature.com/doifinder/10.1038/nrg1523.

Pääbo, S. et al., 2004. Genetic analyses from ancient DNA. *Annual review of genetics*, 38, pp.645–679.

Pacioni, C. et al., 2014. Genetic diversity loss in a biodiversity hotspot: ancient DNA quantifies genetic decline and former connectivity in a critically endangered marsupial. *Molecular ecology*.

Pagès, M. et al., 2012a. Paleogenetic analyses reveal unsuspected phylogenetic affinities between mice and the extinct *Malpaisomys insularis*, an endemic rodent of the Canaries. *PLoS ONE*, 7(2).

Pagès, M. et al., 2012b. Paleogenetic analyses reveal unsuspected phylogenetic affinities between mice and the extinct *Malpaisomys insularis*, an endemic rodent of the Canaries. *PLoS ONE*, 7(2), pp.10–13.

Palkopoulou, E. et al., 2015. Complete genomes reveal signatures of demographic and genetic declines in the woolly mammoth. *Current Biology*, 25(10), pp.1395–1400.

Pickrell, J.K. & Reich, D., 2014. Toward a new history and geography of human genes informed by ancient DNA. *Trends in Genetics*, 30(9), pp.377–389. Available at: http://dx.doi.org/10.1016/j.tig.2014.07.007.

Poulakakis, N. et al., 2005. Isolation and amplification of shrew DNA from barn owl pellets. *Biological Journal of the Linnean Society*, 85(3), pp.331–340.

Prost, S. et al., 2013. Effects of late quaternary climate change on Palearctic shrews. *Global Change Biology*, 19(6), pp.1865–1874.

Prost, S. et al., 2010. Influence of climate warming on arctic mammals? new insights from ancient DNA studies of the collared lemming *Dicrostonyx torquatus*. *PLoS ONE*, 5(5).

Reed, D.N., 2005. Taphonomic implications of roosting behavior and trophic habits in two species of African owl. *Journal of Archaeological Science*, 32(11), pp.1669–1676.

Ricklefs, R. & Bermingham, E., 2008. The West Indies as a laboratory of biogeography and evolution. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 363(1502), pp.2393–2413.

Roberts, C. & Ingham, S., 2008. Using ancient DNA analysis in palaeopathology: a critical analysis of publshed papers, with recommendations for future work. *International Journal of Osteoarchaeology*, 18(6), pp.600–613.

Robins, J.H. et al., 2014. Phylogenetic species identification in *Rattus* highlights rapid radiation and morphological similarity of New Guinean species. *PLoS ONE*, 9(5).

Rodrigues, M. et al., 2016. Taxonomic status and origin of the Egyptian weasel (*Mustela subpalmata*) inferred from mitochondrial DNA. *Genetica*, 144(2), pp.191–202.

Schreve, D.C., 2001. Mammalian evidence from Middle Pleistocene fluvial sequences for complex environmental change at the oxygen isotope substage level. *Quaternary International*, 79(1), pp.65–74.

Shepherd, L.D. & Lambert, D.M., 2008. Ancient DNA and conservation: Lessons from the endangered kiwi of New Zealand. *Molecular Ecology*, 17(9), pp.2174–2184.

Smith, S. et al., 2016. *Non-receding hare lines: genetic continuity since the Late Pleistocene in European mountain hares (*Lepus timidus*).*

Smulders, M.J.M. et al., 2003. Complete loss of MHC genetic diversity in the Common Hamster (Cricetus cricetus) population in The Netherlands. Consequences for conservation strategies. *Conservation Genetics*, 4(4), pp.441–451.

Spaeth, P.A. et al., 2009. Phylogeography of *Microtus Longicaudus* in the Tectonically and Glacially Dynamic Central Rocky Mountains. *Journal of Mammalogy*, 90(3), pp.571–584.

Taberlet, P. & Fumagalli, L., 1996. Owl pellets as a source of DNA for genetic studies of small mammals. *Molecular ecology*, 5(2), pp.301–305.

Thomas, W.K. et al., 1990. Spatial and temporal continuity of kangaroo rat populations shown by sequencing mitochondrial DNA from museum specimens. *Journal of Molecular Evolution*, 31(2), pp.101–112.

Turvey, S.T. et al., 2016. Independent evolutionary histories in allopatric populations of a threatened

Caribbean land mammal. *Diversity and Distributions*, 22(5), pp.589–602.

Van, M.A.J., 1973. *Middle Pleistocene smaller mammals from the Monte Peglia (Orvieto, Italy) with special reference to the phylogeny of Microtus*,

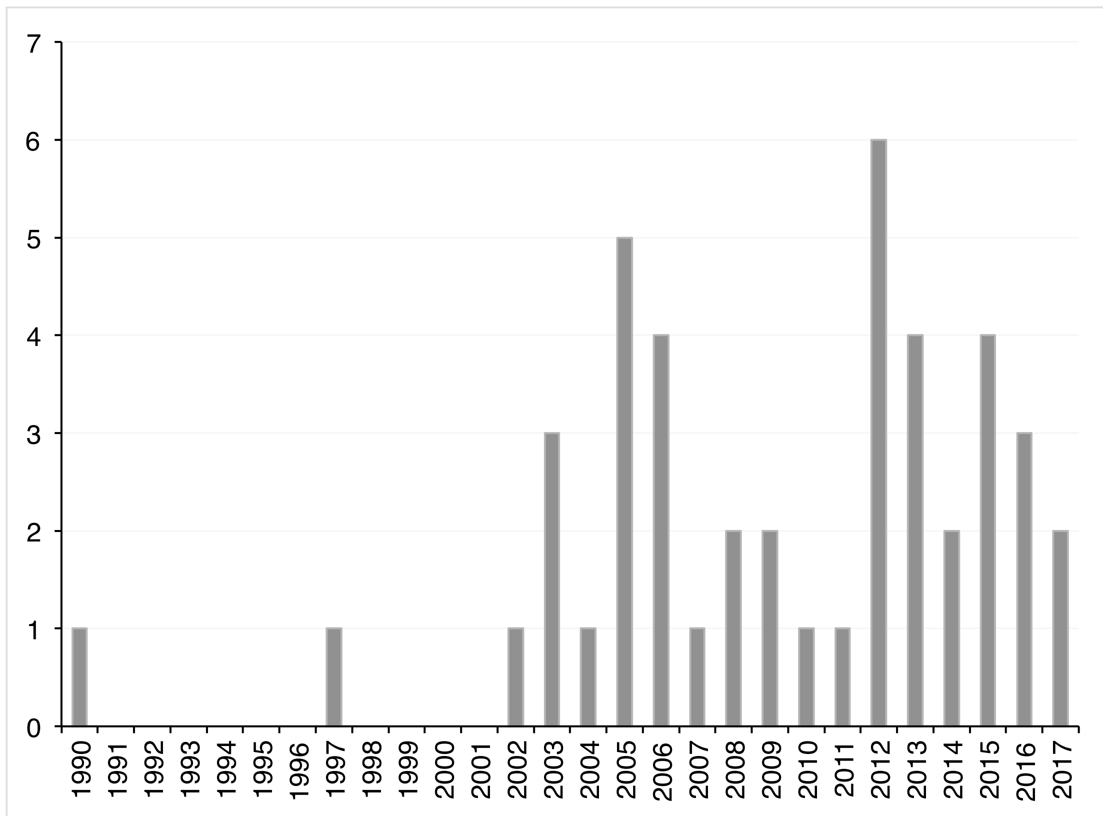
Willerslev, E. et al., 2005. Ancient DNA. *Proceedings. Biological sciences / The Royal Society*, 272(1558), pp.3–16.

Withers, P. et al., 2016. Introduction to mammals. In *Ecological and Enviromental Physiology of Mammals*. Oxford University Pess.

Wyatt, K.B. et al., 2008. Historical mammal extinction on Christmas Island (Indian Ocean) correlates with introduced infectious disease. *PLoS ONE*, 3(11).

Yang, D.Y., Woiderski, J.R. & Driver, J.C., 2005. DNA analysis of archaeological rabbit remains from the American Southwest. *Journal of Archaeological Science*, 32(4), pp.567–578.

Figure 1a. Ancient DNA studies including micromammals from 1990-2017



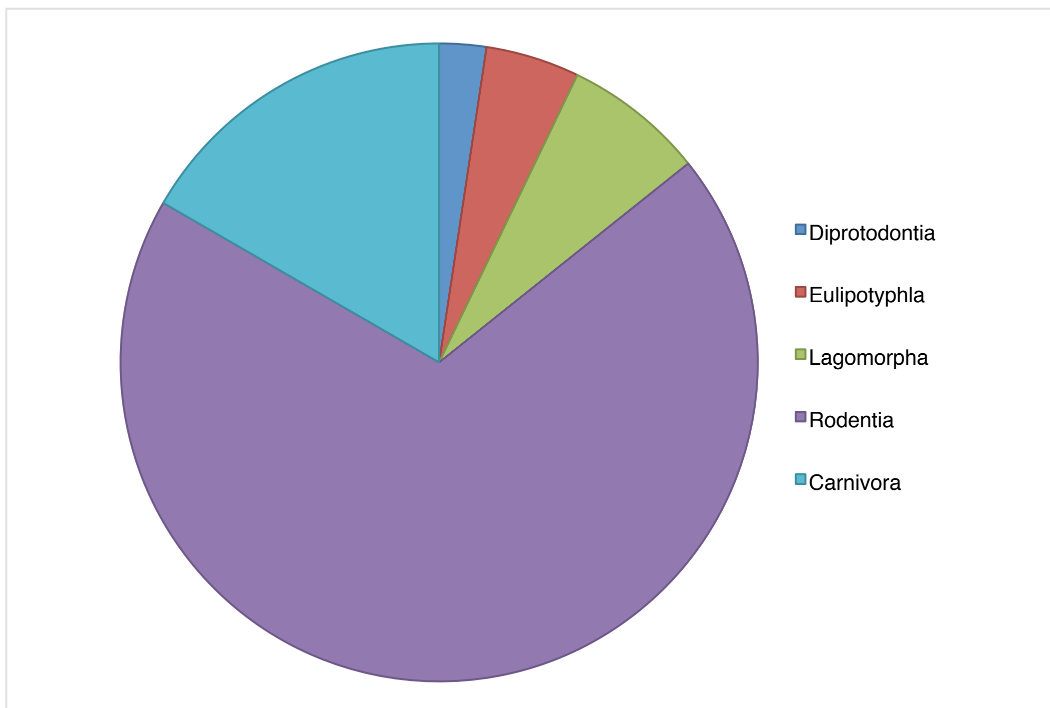


Figure 2a. Ancient DNA studies including different micromammal orders (1990-2015)

Appendix Table 1a. All published studies from 1990 to 2017 in which ancient or degraded DNA was utilised to study micromammal fauna

Order	Genus	Species	Date	Author	Genetic Marker
Rodentia	*Dipodomys*	*Panamintinus*	1990	Thomas et al.	Control Region
Rodentia	*Rattus*	*Exulans*	1997	Matisoo-Smith et al.	Control Region
Rodentia	-	-	2002	Kuch et al.	12S & 16S rDNA, Cytochrome *b*
Rodentia	*Mesomys*	*Hispidus*	2003	Orlando et al.	Cytochrome *b*
Rodentia	*Pappogeomys*	*Alcorni*	2003	Demastes et al.	Cytochrome *b*
Rodentia	*Microtus*	*Montanus*	2004	Hadly, et al. (Hadly et al. 2003)	Cytochrome *b*
Lagomorpha	*Lepus, Sylvilagus*	-	2005	Yang et al.	Cytochrome *b*
Rodentia	*Ctenomys*	*sociabilis*	2005	Chan et al.	Cytochrome *b*
-	-	-	2005	Avenant et al.	Cytochrome *b*
-	-	-	2005	Poulakakis et al.	Cytochrome *b*
Rodentia	*Ctenomys*	*Sociabilis*	2006	Chan et al.	Cytochrome *b*
Rodentia	*Rattus*	*Exulans*	2006	Barnes et al.	Control Region
Rodentia	*Dinaromys*	*Bogdanovi*	2007	Krystufek et al.	Cytochrome *b*
Rodentia	*Microtus*	*longicaudus*	2009	Spaeth et al.	Cytochrome *b*
Rodentia	*Rattus*	*exulans*	2009	Matisoo-Smith & Robins	Control region, Cytochrome *b*, Cytochrome Oxidase Subunit I

Appendix Table 1a. (continued). All published studies from 1990 to 2017 in which ancient or degraded DNA was utilised to study micromammal fauna

Order	Genus	Species	Date	Author	Genetic Marker
Rodentia	*Glaucomys*	*Volans*	2010	Kerhoulas & Arbogast	Cytochrome *b*
Rodentia	*Dicrostonyx*	*torquatus*	2010	Prost et al.	Control Region
Rodentia	*Ctenomys*	*haigi, sociabilis*	2011	Chan & Hadley	Cytochrome *b*
Eulipotyphla	*Erinaceus*	*europaeus*	2012	Fraser et al.	Cytochrome *b*
Rodentia	*Dicrostonyx*	*torquatus*	2012	Brace et al.	Cytochrome *b*
Rodentia	*Malpaisomys, Mus, Murinae*	-	2012	Pages et al.	Cytochrome b, Interphotoreceptor Retinoid Binding Protein (IRBP)
Rodentia	*Microtus*	*arvalis, rossiaemeridionalis*	2012	Markova et al.	Cytochrome *b*
Rodentia	*Peromyscus*	*polionotus*	2012	Kalkvik et al.	Cytochrome *b*
Rodentia	*Plagiodontia*	*Aedium*	2012	Brace et al.	Cytochrome *b*
Eulipotyphla	*Sorex*	*araneus, tundrensis*	2013	Prost et al.	Cytochrome *b*
Rodentia	*Dicrostonyx*	*richardsoni*	2013	Fulton et al.	Cytochrome *b*
Rodentia	*Dicrostonyx*	*torquatus*	2013	Prost et al.	Cytochrome *b*
Rodentia	*Microtus*	*gregalis*	2013	Prost et al.	Cytochrome *b*

Appendix Table 1a. (continued). All published studies from 1990 to 2017 in which ancient or degraded DNA was utilised to study micromammal fauna.

Order	Genus	Species	Date	Author	Genetic Marker
Rodentia	*Ototylomus*	*phyllotis*	2014	Gutierrez-Garcia et al.	Cytochrome *b*
Rodentia	*Rattus*	-	2014	Robins et al.	Control region, Cytochrome Oxidase Subunit I
Diprotodontia	*Bettongia*	Penicillata	2015	Pacioni et al.	Control Region, Microsatellites
Rodentia	*Alexandromys*	*fortis, indet, maximowiczii, oeconomus*	2015	Haring et al.	Control Region
Rodentia	*Megalomys*	*georginae, desmarestii,*	2015	Brace et al.	Cytochrome *b*
Rodentia	*Pennatomys*	*nivalis, luciae*	2015	Brace et al.	Cytochrome *b*
Lagomorpha	*Lepus*	*timidus*	2016	Ahlgren et al.	Control Region
Rodentia	*Arvicola*	*Amphibious*	2016	Brace et al.	Control Region
Eulipotyphla	*Nesophontidae*	*paramicrus*	2016	Brace et al.	Mitochondrial genome, ADORA3, ADRA2B, ADRB2, APOB, APP, ATP7A, BCHE, BDNF, BMI1, BMP4, BRCA1, CREM, EDG1, GHR, PLCB4, RAG1, RAG2, RHO, TTN, TYR, VWF
Eulipotyphla	Solenodontidae	*paradoxus*	2016	Turvey et al.	

Appendix Table 1a. (continued) All published studies from 1990 to 2017 in which ancient or degraded DNA was utilised to study micromammal fauna.

Order	Genus	Species	Date	Author	Genetic Marker
Rodentia	*Dicrostonyx*	*groenlandicus, richardsoni, torquatus*	2016	Palkopoulou et al.	Cytochrome *b*
Lagomorpha	*Lepus*	*timidus*	2017	Smith et al.	Control Region, Cytochrome *b*
Rodentia	*Mus, Meriones, Apodemus, Lemniscomys, Rattus, Gerbillus, Cricetulus, Microtus, Eliomys*	-	2017	Guimaraes et al.	Single Nucleotide Polymorphisms (SNPs)