Speciation and diversification of the beetle tribe Cratopini on the islands of the South West Indian Ocean area.

James J.N. Kitson

A thesis submitted for the degree of Doctor of Philosophy.

School of Biology
University of East Anglia, UK

September 2012

© This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that use of any information derived there-from must be in accordance with current UK Copyright Law. In addition, any quotation or extract must include full attribution.

Thesis abstract

The aim of this thesis is to examine a spectacular radiation of phytophagous beetles (Curculionidae: Entiminae: Cratopini) that inhabit the islands of the South-Western Indian Ocean. As there are no previous molecular studies of these beetles, it was necessary to take a varied approach that employed both population genetic methods and phylogenetic methods to gain a broad understanding of the evolution of the group. Questions addressed range from aspects of the ecology of individual species to the relationships within islands amongst morphospecies and finally to the biogeography of the group as a whole. Throughout, where appropriate, reference is made to the current understanding of Indian Ocean geology and how this effects the interpretation of the patterns seen.

The relationships amongst morphospecies are complex with many found to be paraphyletic within much larger monophyletic clades of multiple morphospecies. Coalescent model based species delimitation techniques are used to define the boundaries of putative species for future examination. The overall biogeography of the group is found to be highly complex with multiple colonisation events across islands and archipelagos. Flight is found to have been lost on multiple occasions and the taxonomy of the group is discussed. Finally many Cratopine lineages are found to display strong population structure despite being flighted, possible reasons for this are discussed.

Supervisor details

Main supervisor:

Dr Brent Emerson

Island Ecology and Evolution Research Group,

Instituto de Productos Naturales y Agrobiología (IPNA-CSIC),

C/Astrofísico Francisco Sánchez 3,

La Laguna,

Tenerife (Canary Islands, 38206),

Spain

Secondary supervisor:

Dr David S Richardson

School of Biological Sciences,

University of East Anglia,

Norwich Research Park,

Norwich,

Norfolk

NR4 7TJ

United Kingdom

List of Contents

Thesis abstract i

Supervisor details ii

List of Contents iii

List of Tables and Figures vii

Acknowledgements xiv

Chapter 1: Introduction 1

1.1 Evolutionary Biology and Islands 2

1.2 The study system and thesis structure 3

1.3 The geology of the South-West Indian Ocean 5

1.4 References 10

Chapter 2: Molecular characterisation of a trophic shift within an island radiation of insect herbivores (Curculionidae: Entiminae: *Cratopus*). 14

2.1 Abstract 15

2.2 Introduction 15

2.3 Methods 18

2.3.1 Beetle sampling 18

2.3.2 Plant sampling 19

2.3.3 DNA extraction, PCR amplification and sequencing 19

2.3.4 Phylogenetic, network and population genetic analyses 21

2.3.5 Divergence time estimation 22

2.3.6 Species assignment of trnL sequences 23

2.3.7 Statistical analyses 24

2.4 Results 25

2.4.1 MtDNA and nuclear gene sequencing 25

2.4.2 Phylogenetic analyses and haplotype networks 26

2.4.3 Population genetic analyses and cryptic species assessment 30

2.4.4 Divergence times 32

2.4.5 Plant analyses ... *32*

2.5 Discussion ... 36

2.5.1 Geographic structuring of genetic variation within Cratopus murinus and Cratopus ovalis ... *36*

2.5.2 Trophic ecology of Cratopus murinus and Cratopus ovalis from dry forest sites ... *38*

2.5.3 Variation in Cratopus murinus diet across dry forest sites ... *39*

2.5.4 Trophic ecology of coralline islet populations of Cratopus murinus ... *39*

2.6 Conclusions ... 40

2.7 References ... 41

Chapter 3: Baseline molecular data for poorly understood diversity: a case study of Mauritian Cratopine weevils. ... 45

3.1 Abstract ... 46

3.2 Introduction ... 46

3.3 Methods ... 49

3.3.1 Beetle sampling ... *49*

3.3.2 DNA extraction, PCR amplification and sequencing ... *51*

3.3.3 Phylogenetic analyses ... *53*

3.3.4 Divergence time estimation ... *53*

3.3.5 GMYC analyses ... *53*

3.3.6 Population structure ... *54*

3.3.7 Phylogenetic diversity ... *54*

3.4 Results ... 55

3.4.1 MtDNA sequencing ... *55*

3.4.2 Phylogenetic analyses ... *56*

3.4.3 Divergence times and GMYC analyses ... *59*

3.4.4 Population structure ... *67*

3.4.5 Phylogenetic diversity ... *68*

3.5 Discussion ... **73**

3.5.1 The Phylogenetics of Mauritian Cratopine weevils ... *73*

3.5.2 GMYC analyses ... *74*

3.5.3 Population structure ... *75*

3.5.4 Phylogenetic diversity ... *76*

3.6 Conclusions ... **77**

3.7 References ... **78**

4.1 Abstract ... **83**

4.2 Introduction ... **83**

4.3 Methods ... **84**

4.3.1 Beetle sampling ... *84*

4.3.2 DNA extraction, PCR amplification and sequencing ... *89*

4.3.3 Phylogenetic analyses ... *90*

4.3.4 Divergence time estimation ... *90*

4.3.5 GMYC analyses ... *90*

4.3.6 Population structure ... *91*

4.4 Results ... **91**

4.4.1 MtDNA sequencing ... *91*

4.4.2 Phylogenetic analyses ... *93*

4.4.3 Divergence times and GMYC analyses ... *97*

4.4.4 Population structure ... *106*

4.5 Discussion ... **109**

4.5.1 Phylogenetics of Reunionaise Cratopine weevils ... *109*

4.5.2 Divergence times and colonisation of La Réunion ... *110*

4.5.3 GMYC analyses ... *110*

4.5.4 Population structure ... *112*

4.6 Conclusions ... **113**

4.7 References ... **114**

5.1 Abstract **120**

5.2 Introduction **120**

5.3 Methods **123**

5.3.1 Beetle sampling *123*

5.3.2 DNA extraction, PCR amplification and sequencing *125*

5.3.3 Phylogenetic analyses *126*

5.3.4 Divergence time estimation *127*

5.4 Results **127**

5.4.1 DNA sequencing *127*

5.4.2 Phylogenetic analyses *135*

5.5 Discussion **145**

5.5.1 Biogeography of South-West Indian Ocean Cratopine weevils *145*

5.5.2 Cratopopsis bistigma and Cratopopsis mauritianus *149*

5.5.3 Loss of flight capability in Cratopine weevils *149*

5.6 Conclusions **150**

5.7 References **151**

6.1 Overview **157**

6.2 Taxonomy **157**

6.3 Ecology **158**

6.4 Flight **159**

6.5 Future directions **160**

6.5.1 Rapid colonisation of coastal habitats by Cratopus *160*

6.5.2 Range mapping of Cratopine species *160*

6.5.3 Next generation approaches to diet analysis *161*

6.5.4 Comparative analyses *161*

6.6 References **161**

List of Tables and Figures

Fig 1.1: The current configuration of South-West Indian Ocean land masses. The first number associated with islands is the oldest age estimate available while the ages of the oldest radiometrically dated lavas are contained in brackets. References detailing these dates are in the text. 7

Fig 1.2: (A) Current island configuration in the Indian Ocean. (B) Estimated island extents during an 80 m BPSL lowstand. (C) Estimated island extents during a 135 m BPSL lowstand. Both lowstand maps assume constant ocean floor topography. Diagram taken from Warren *et al.* (2010). Abbreviated island labels: RE (La Réunion), MU (Mauritius) and RD (Rodrigues). 9

Table 2.1: Names and locations of sampled populations with description of habitat. 19

Table 2.2: PCR and sequencing primers used in this study. Primers suffixed with (sequencing) were only used during the sequencing reactions.......... 20

Figure 2.1. Mr Bayes consensus tree for *C. murinus* and *C. ovalis* with *C. melanocephalus* and *C. vulgaris* as outgroups. The analysis was performed using 4 MCMC chains for 10 million generations with the evolutionary model HKY+G. Values above the nodes are posterior probabilities for each node. Nodes used to define tmrcas in BEAST analyses and populations for pairwise F_{st} calculations are labelled. Insert is a map of Mauritius with collecting sites marked.......... 27

Figure 2.2. Haplotype networks derived from putatively non-recombined sections of ArgK as identified by the Dual Brothers analysis. The area of each haplotype is proportional to its frequency. 29

Figure 2.3. Haplotype network produced by TCS for the ITS2 alignment with gaps removed. The area of each haplotype is proportional to its frequency. 30

Table 2.3. Pairwise F_{st} values between sampling locations for *C. murinus.* All values are significant (significance level = 0.05). Values above the diagonal are calculated using ITS2 data, those below the diagonal are calculated using ArgK.......... 31

Table 2.4. Pairwise F_{st} values between sampling locations for *C. ovalis.* Non-significant results are formatted in bold (significance level = 0.05). Values above the diagonal are calculated using ITS2 data, those below the diagonal are calculated using ArgK.......... 31

Table 2.5. Plant species consumed by *C. murinus*, with the number of feeding records for each, by sampling locality including the *Betula sp.* and *Pinus sp.* sequences deemed to be contamination. 33

Table 2.6. The number of sequences retrieved for each food plant from *C. ovalis* samples split by collection locality including the *Betula sp.* sequences deemed to be contamination. 33

Table 2.7. Potential *Cratopus* foodplants identified during plant surveys by CB and BFVF on Corps de Garde and Trois Mamelles. 34

Table 2.8. Feeding and abundance records for food plants of *C. murinus* in Corps de Garde and Trois Mamelles with sequences deemed to be contamination removed.. 35

Figure 2.4. Scatter plot of Number of feeding records vs. foodplant availability split by sampling site. The solid black line is a line of best fit though the data. The dashed arrow highlights the points for *Erythroxylum sideroxyloides* in both sites for comparison. 35

Figure 2.5. Scatter plot of Number of feeding records vs. foodplant availability split by sampling site with values for *Erythroxylum sideroxyloides* removed. The solid black line is a line of best fit though the data..36

Table 3.1: The collection localities included in this study and their associated latitude and longitude..50

Fig 3.1: Collecting sites sampled on Mauritius. Inset is a magnification of the Brise Fer and Machabee regions. 51

Fig 3.2: A linear regression of uncorrected genetic distance against genetic distance corrected using the GTR model. The solid red line represents the fitted model and the solid black line represents no difference between uncorrected and corrected genetic distances..56

Fig 3.3: A simplified MrBayes tree for all Mauritian Cratopine weevils sampled. The depth of the black triangles represents the coalescent depth of the group, the height of the rectangles represents the number of tips and the width of the rectangles represents the proportion of tips in the group belonging to that morphospecies. 58

Fig 3.4: The multimodel GMYC analysis result for section A of Fig 3.3 (unique sequences only).....60

Fig 3.5: The multimodel GMYC analysis result for section B of Fig 3.3 (unique sequences only).....61

Fig 3.6: The multimodel GMYC analysis result for sections C and D of Fig 3.3 (unique sequences only).....62

Fig 3.7: The multimodel GMYC analysis result for sections F and G of Fig 3.3 (unique sequences only).....63

Fig 3.8: The multimodel GMYC analysis result for sections H and I of Fig 3.3 (unique sequences only).....63

Fig 3.9: The multimodel GMYC analysis result for section J of Fig 3.3 (unique sequences only).....64

Fig 3.10: The multimodel GMYC analysis result for sections L and M of Fig 3.3 (unique sequences only).....65

Fig 3.11: The multimodel GMYC analysis result for section N of Fig 3.3 (unique sequences only).....66

Fig 3.12: The multimodel GMYC analysis result for sections Q and R and I of Fig 3.3 (unique sequences only).....67

Fig 3.13: Collecting sites included in each AMOVA performed.....68

Fig 3.14: A linear regression of phylogenetic diversity against morphospecies richness. The continuous red line represents the straight fitted model while the dashed blue line represents the curved fitted model.....69

Fig 3.15: A linear regression of mean pairwise distance against morphospecies richness. The red line represents the fitted model.....70

Fig 3.16: A linear regression of mean nearest taxon distance against morphospecies richness. The red line represents the fitted model.....71

Fig 3.17: Rarefacted phylogenetic diversity for each site. Each section represents a different rarefaction level. Black dots are mean phylogenetic diversity values and red bars are 95% confidence intervals. 72

Table 4.1: Decimal latitudes and longitudes of samples sites on La Réunion.....86

Fig 4.1a: Relative locations of sample sites in the northern half of La Réunion. 87

Fig 4.1b: Relative locations of sample sites in the southern half of La Réunion. 88

Fig 4.2: A linear regression of uncorrected genetic distance against genetic distance corrected using the GTR model. The solid green line represents the fitted model and the solid black line represents no difference between uncorrected and corrected genetic distances. 93

Fig 4.3a: A simplified MrBayes tree for all Reunionaise Cratopine weevils sampled. The depth of the black triangles represents the coalescent depth of the group, the height of the rectangles represents the number of tips and the width of the rectangles represents the proportion of tips in the group belonging to that morphospecies. Numbers beside nodes are the Bayesian posterior probability for that node. 95

Fig 4.3b: Detail for lineage E. Inset is a map of collecting sites on La Réunion. Tip labels are coloured by morphological similarity and sites are coloured to match tip labels. When more than one lineage is present at a site, the marker is subdivided and coloured to represent this. Numbers beside nodes are the Bayesian posterior probability for that node......96

Fig 4.4: A simplified chronogram of Reunionaise Cratopine weevils generated in BEAST. Node bars represent 95% HPD intervals on the age estimate for that node. Stars represent nodes for which monophyly was enforced and numbers beside nodes are Bayesian posterior probabilities for that node. 98

Fig 4.5: The multimodel GMYC analysis result for lineage A of Fig 4.3a (unique sequences only). Stars represent GMYC clusters from the single threshold GMYC analysis. 100

Fig 4.6: The multimodel GMYC analysis result for lineages B, D, F, H and I of Fig 4.3a (unique sequences only). Stars represent GMYC clusters from the single threshold GMYC analysis. 101

Fig 4.7: The multimodel GMYC analysis result for lineage D of Fig 4.3a (unique sequences only). Stars represent GMYC clusters from the single threshold GMYC analysis. 102

Fig 4.8: The multimodel GMYC analysis result for lineage E of Fig 4.3a (unique sequences only). Stars represent GMYC clusters from the single threshold GMYC analysis. 103

Fig 4.9: The multimodel GMYC analysis result for lineage G of Fig 4.3a (unique sequences only). Stars represent GMYC clusters from the single threshold GMYC analysis. 104

Fig 4.10: The multimodel GMYC analysis result for lineage J of Fig 4.3a (unique sequences only). Stars represent GMYC clusters from the single threshold GMYC analysis. 105

Fig 4.11: The multimodel GMYC analysis result for lineages K, L, M and N of Fig 4.3a (unique sequences only). Stars represent GMYC clusters from the single threshold GMYC analysis. 106

Table 4.2: F_{st} and AMOVA results for the tested Reunionaise Cratopine lineages 107

Fig 4.12: Collecting sites included in the AMOVA performed on each lineage. 108

Table 5.1: Morphospecies included in this study with island and island group information. 124

Table 5.2: Sequencing success and summary statistics for each locus used in the phylogeny. 128

Table 5.3: Uncorrected p distances and GTR corrected distances for all loci used in the phylogeny. 128

Table 5.4: Sequencing success split by locus for each morphospecies. Green boxes indicate success and red boxes are failures. 129

Fig 5.2: uncorrected genetic distances plotted against GTR corrected genetic distances for all sequences in the COII alignment. The black line represents a perfect match between uncorrected and corrected distances and zero saturation whereas the green line represents the linear model fitted to the data. 131

Fig 5.3: uncorrected genetic distances plotted against GTR corrected genetic distances for all sequences in the ArgK alignment. The black line represents a perfect match between uncorrected and corrected distances and zero saturation whereas the green line represents the linear model fitted to the data. 132

Fig 5.4: uncorrected genetic distances plotted against GTR corrected genetic distances for all sequences in the Histone 3 alignment. The black line represents a perfect match between uncorrected and corrected distances and zero saturation whereas the green line represents the linear model fitted to the data. 133

Fig 5.5: uncorrected genetic distances plotted against GTR corrected genetic distances for all sequences in the 28S alignment. The black line represents a perfect match between uncorrected and corrected distances and zero saturation whereas the green line represents the linear model fitted to the data. 134

Table 5.5: Xia's test for saturation test statistics and critical values for each locus. The test statistic must be significantly less than the critical value to indicate no saturation. Protein coding loci are partitioned into fist and second codon positions vs third codon position.. 135

Table 5.6: Coefficients of the linear models fitted to the saturation plots of uncorrected p distances vs GTR corrected distances for each locus......................... 135

Fig 5.6: A Bayesian inference tree generated by Mr Bayes using an alignment of Cratopine COII sequences. Values placed on nodes represent Bayesian posterior probabilities for that node. *Cratopopsis bistigma* and *Cratopopsis mauritianus* are highlighted to show incongruence between nuclear loci and COII. 137

Fig 5.7: A Bayesian inference tree generated by Mr Bayes using an alignment of Cratopine ArgK sequences. Values placed on nodes represent Bayesian posterior probabilities for that node. *Cratopopsis bistigma* and *Cratopopsis mauritianus* are highlighted to show incongruence between nuclear loci and COII............................ 138

Fig 5.8: A Bayesian inference tree generated by Mr Bayes using an alignment of Cratopine Histone 3 sequences. Values placed on nodes represent Bayesian posterior probabilities for that node. *Cratopopsis bistigma* and *Cratopopsis mauritianus* are highlighted to show incongruence between nuclear loci and COII 139

Fig 5.9: A Bayesian inference tree generated by Mr Bayes using an alignment of Cratopine 28S sequences. Values placed on nodes represent Bayesian posterior probabilities for that node. *Cratopopsis bistigma* and *Cratopopsis mauritianus* are highlighted to show incongruence between nuclear loci and COII............................ 140

Fig 5.10: A Bayesian inference tree generated by Mr Bayes using an alignment of Cratopine sequences from all four loci. Values placed on nodes represent Bayesian posterior probabilities for that node. Node labels coloured red were constrained to be monophyletic in the BEAST analysis. Stars represent key nodes of interest for inferring colonisation times and red boxes delimit the extent of Fig 5.11a and Fig 5.11b. 142

Fig 5.11a: An enlargement of lineage 1 and part of lineage 2 from Fig 5.10. The top number in the box is the mean estimated age of the node indicated, the numbers in brackets represent the 95% HPD interval on the node age estimate. Values placed on nodes represent Bayesian posterior probabilities for that node generated in the combined Mr Bayes analysis. 143

Fig 5.11b: An enlargement of lineage 3 and part of lineage 2 from Fig 5.10. The top number in the box is the mean estimated age of the node indicated, the numbers in brackets represent the 95% HPD interval on the node age estimate. Values placed on nodes represent Bayesian posterior probabilities for that node generated in the combined Mr Bayes analysis. 144

Acknowledgements

The work presented here would have been impossible without the help of a large number of people. First and foremost, my supervisor Brent Emerson has been patient and knowledgeable throughout the whole project and without him this would not have been possible. It has also been a pleasure to work with my secondary supervisor David Richardson who has provided salient, timely advice throughout the PhD.

In the field I have been aided immeasurably by Ben Warren who provided advice and logistical support throughout my time in the Mascarenes. Additionally Juliane Casquet and Emilie Lesseguince who made fieldwork so enjoyable with their discussion of the finer points of cheese (especially the awful Mauritian cheese), Death Metal, French womens' boxing and the delights of Creole cooking. Antoine Franck deserves a special mention for providing all my samples from Europa and Juan de Nova as well as contributing samples from numerous other islands. Extensive sampling of Reunionaise *Cratopopsis* and *Cratopus bernei* would not have been possible without Jaques Rochat and much of the identification would have been impossible without Jaques Poussereau.

With regard to the Mauritian field work, I would specifically like to thank Vincent Florens and Claudia Baider for all their work surveying floral diversity for chapter two; all the staff at the Mauritian Wildlife Foundation (especially Sayed Jhumka who introduced me to flightless *C. murinus* populations that were previously undocumented); the staff at the National Parks and Conservation Service (especially Kevin Ruhomaun who arranged our permits); Owen Griffiths and his conservation manager Nada Padayatchy who allowed me to access their private reserve and provided me with dried specimens of the rarest and most beautiful *Cratopus* of all (*Cratopus chrysochlorus* – known only from Owen's swimming pool).

For samples from other localities (given in brackets) I would like to thank Karl Phillips (Cousine Island); Dave Wright (Cousin Island and Aride Island); Justin Gerlach (Silhouette Island); N. Borowiec (Mahe, Praslin and La Digue); Janske van de Crommenacker (Aldabra) and finally, for his stalwart efforts and unsurpassed generosity, Silvain Hugel (Mauritius, Rodrigues, Silhouette Island, La Digue, Praslin, Moheli, Anjouan and Grande Comore). Without the efforts of these people,

this thesis would be distinctly focussed on the Mascarenes and much of the biogeography would be unknown.

This work would not be possible without the overall project leader Dominique Strasbourg and generous funding from ANR, NERC and the Pamela Salter fund.

Finally I would like to thank all of Team Weevil past and present (Gerardo Hernandez-Vera, Conrad Gillett, Christiana Faria and Alicia Mastretta-Yanes) for being awesome in every way; Matt Ellis for accompanying me in my efforts to find the (seemingly mythical) *Cratopus* of Madagascar; Lewis Spurgin and Phil Leftwich for pub times and mountain times and Lucy Friend, James Norris and Lotty Packman for cycling japes and whisky. I would like to thank my parents for supporting me through the last 30 years (including nine years of university!). Finally I would like to thank my girlfriend Anna Millard for being there over the last few years and providing endless support and respite from the stress of writing.

Chapter 1: Introduction

Sometimes fieldwork happens in bad places but not always!
Anse Bouteille, Rodrigues 2009

1.1 Evolutionary Biology and Islands

"… it is necessary to state the relations and affinities [of taxa] to each other in order to get an idea of the peculiarities of the geographical distribution and to find out their cause." (Ortmann 1894)

The study of the biotic patterns and the processes that have led to their formation are fundamental to the understanding of life as a whole. Populations and species are shaped by a complex array of processes over long periods of time. This historical aspect makes direct quantification of evolutionary processes impractical for all but the fastest evolving bacteria or viruses and makes rigorous testing of hypotheses and theories extremely difficult.

Emerson (2002) states that oceanic islands provide ideal terrestrial environments for examining evolutionary hypotheses, since : (1) they are discrete geographic entities within a matrix of inhospitable ocean that is usually present over evolutionary timescales; (2) gene flow between islands is often greatly reduced by oceanic barriers; (3) their flora and fauna are often more completely known than those of continental systems; (4) they frequently contain a wide array of habitats despite their small size, and; (5) they are frequently geologically active and subject to ongoing volcanic, tectonic or erosional processes. These characteristics mean that islands typically have very high levels of endemicity, often in the form of spectacular species radiations. Perhaps the most famous of these radiations is the collection of 15 species of finch that form an adaptive radiation on the Galapagos islands (Darwin 1869). Despite the simplicity of island systems compared to their continental counterparts, biogeographic patterns and, crucially, the geographic origins of groups can be difficult to ascertain, with the evolutionary relationships and origins of Darwin's finches only being addressed relatively recently (Sato *et al.* 1999; Sato *et al.* 2001).

For most island groups, it is extremely rare to find continuous fossil records, especially for invertebrate groups. Additionally, convergent evolution can result in similar morphologies arising independently on different islands (Gillespie 2004), leading to incorrect conclusions regarding their origin when based on morphology alone. Using molecular markers can provide a wealth of characters that are often independent of morphology, and mutational rate variation across

markers allows researchers to tailor the level of variability in their datasets to those appropriate for the questions being asked. Perhaps most importantly, these characters are available for any extant organism that can be sampled and indeed for many that are no longer extant (e.g. Austin *et al.* 2004).

Molecular data have been used for the examination of patterns and processes on islands at varying evolutionary scales. At interspecific scales, phylogenetic approaches have revealed a wealth of information related to the timing of divergence events (Magnacca & Danforth 2006; Sequeira *et al.* 2000), patterns of diversification (Holland & Hadfield 2002; Jordal *et al.* 2004; Jordan *et al.* 2005) and the resolution of taxonomic uncertainty (e.g. Clement *et al.* 2004; Jordal *et al.* 2006). While at intraspecific scales, population genetic approaches and coalescent models have typically been used to address questions relating to genetic structure across populations within the same species (e.g. Emerson *et al.* 2000; Grobler *et al.* 2011; Illera *et al.* 2007; Polihronakis *et al.* 2010). Both approaches can yield information that advances evolutionary understanding (e.g. Fukami *et al.* 2007; Gillespie 2004; Spurgin *et al.* 2011) and conservation management (e.g. Cegelski *et al.* 2006; de Thoisy *et al.* 2010; Newton *et al.* 1999).

With approximately 62,000 described species (Oberprieler *et al.* 2007), weevils are the most speciose invertebrate superfamily and have a worldwide distribution that covers all major terrestrial habitats except mainland Antarctica. This makes weevils among the most speciose and geographically available pool of island colonisers. Within the weevils, one of the biggest groups is the subfamily Entiminae. As such, it is perhaps not surprising that on most major island archipelagos the Entiminine weevils are represented by at least one species rich genus. These genera can provide a microcosm in which to examine the evolutionary processes that have driven the radiation of weevils in general and allow us to draw conclusions regarding the formation of hyperdiverse lineages.

1.2 The study system and thesis structure

The overarching aim of this thesis is to use molecular data to infer ecological and evolutionary patterns in a spectacular phytophagous insect radiation on the Islands of the Indian Ocean. The focal taxa are three genera in the weevil tribe Cratopini (Curculionidae: Entiminae). The Cratopini contains nine genera

(*Cratopophilus*, *Cratopopsis,* *Cratopus*, *Hemicratopus*, *Lujaiella*, *Pseudophisus*, *Scaevinus*, *Staimus* and *Zyrcosoides*) (Alonso-Zarazaga & Lyal 1999) and is distributed across the Indian Ocean, Africa and the Middle East. Of the nine genera, only three are encountered frequently on the Islands of the Indian Ocean (*Cratopus*, *Cratopopsis* and *Scaevinus*). They are found across a wide range of islands from Europa in the South-West to the Andamans in the North-East, with representatives on each major archipelago. The Genus *Cratopus* is by far the most diverse, with approximately 80 species described (Williams & Cox 2003) on Mauritius, La Réunion and Rodrigues and approximately 100 species in total. *Cratopopsis* has far fewer representatives, with only 14 species, and is only known from Moheli, Mauritius, La Réunion and Rodrigues. Finally there are only two described species of *Scaevinus*, one on Mauritius and one on La Réunion.

Most of the described morphospecies of *Cratopus* and *Cratopopsis* are single island endemics; very few species have been historically recorded across islands (Williams & Cox 2003). Many of these are very rare and are associated with questionable collecting data, leading Williams and Cox (2003) to conclude that many of these records are likely to be erroneous. There is also much disagreement among authors over the status of many species, with recent revisions of many species (see the following for a complete history of Cratopus taxonomy: Champion 1914; Ferragu & Richard 1990, 1993; Galman *et al.* 2011; Hustache 1919; Hustache 1920; Poussereau & Voisin 2009; Richard 1957; Richard 1958, 1961, 1977; Richard 1995a, b; Voisin & Poussereau 2007a, b, 2009; Williams & Cox 2003).

The biology and ecology of individual Cratopine species are largely unknown. Adults are folivorous and seem to feed largely on native plant species, with a few species being minor pests of agricultural products (*C. punctum* attacks sugar cane on Mauritius (Nair 2007; Williams & Cox 2003), while *C. brunnipes* and *C. humeralis* attack fruit trees on La Réunion (Menzel *et al.* 2005)). Eggs are laid in leaf axils and larvae probably feed either on or around the roots of plants (Menzel *et al.* 2005; Williams & Cox 2003). The three genera exhibit substantial morphological and ecological variation both within and among species (Hustache 1920; Williams & Cox 2003), and species are present in all native terrestrial habitats available on the Mascarene islands between the foreshore and high alpine scrub up to at least 2500m. Within *Cratopus*, a large proportion of described

species are seemingly rare or have highly restricted ranges (Williams & Cox 2003), making them vulnerable to global extinction.

Throughout this thesis Bayesian tree building methods are used on both single and multilocus datasets to infer phylogenies and estimate divergence times (using previously published mutation rates) between populations and species. The trees generated are used as the basis for further analyses involving phylogenetic diversity and putative species delimitation. Four data chapters are included in this thesis and each is written in the style of a self contained manuscript. Chapter two examines population structure, ecology and a putative trophic shift in a widespread Mauritian *Cratopus* species. Chapters three and four use a single locus screen of a large number of individuals on Mauritius and La Réunion combined with a coalescent theory-based approach to estimate putative species boundaries for future investigation. Chapter three also examines comparable sites across Mauritius to look for areas of high phylogenetic diversity and identify sites of conservation interest. Chapter five presents a dated phylogeny of the group as a whole, with inferences about biogeography, flight loss and mitochondrial introgression. The final chapter draws overall conclusions and discusses possible future work using this group.

1.3 The geology of the South-West Indian Ocean

The evolution of island groups cannot be fully understood without first understanding the basic geological processes that have formed their habitats. The South-West Indian Ocean has a rich geological history that involves the formation of today's islands through tectonic, volcanic and biological activity. The current configuration of islands is shown in Fig 1.1. The breakup of the Gondwanan supercontinent occurred between 160 Ma (megaannum) and 64 Ma with the split of Antarctica/Seychelles/Madagascar/India from Africa occurring between 160 Ma and 115 Ma (Coffin & Rabinowitz 1987; Kingdon 1990; Plummer & Belle 1995; Rabinowitz *et al.* 1983). Around 130 Ma Seychelles/Madagascar/India split from Antarctica and around 88 Ma Seychelles/India split from Madagascar (Storey *et al.* 1995). The Indian subcontinent started moving northwards and around 64 Ma the granitic Seychelles split from India (Plummer & Belle 1995). The next oldest group of islands are the volcanic islands of the Comoros and Mascarene archipelagos. The

sequential ages of the islands in the Comoros suggest that they were formed by a classic volcanic hotspot over approximately the last 10 Ma (Emerick & Duncan 1982). Estimated maximum ages available for Comorian islands are 10-15 Ma (Mayotte), 11.5 Ma (Anjouan) and 0.5 Ma (Grande Comore) (Montaggioni & Nougier 1981; Nougier *et al.* 1986), while radiometric dating of the oldest lavas returns ages of 7.7 ± 1.0 Ma (Mayotte), 3.9 ± 0.3 Ma (Anjouan), 0.13 ± 0.02 Ma (Grande Comore) and 5.0 ± 0.4 Ma (Moheli) (Emerick & Duncan 1982; Nougier *et al.* 1986). The Mascarene Islands are also of volcanic origin, but rather than being formed in sequence by a volcanic hotspot, islands are separated from each other by fracture zones that have caused the islands to evolve independently (McDougal 1971; McDougal & Chamalaun 1969). This has resulted in islands that are separated by ocean around 4000m deep and have never been connected, unlike other groups of islands in the area (see below). The oldest lavas on each island are thought to be 8.9 ± 0.17 Ma (Mauritius (Moore *et al.* 2011)), 2.0 ± 0.05 Ma (La Réunion (McDougal 1971)) and 1.5 ± 0.05 Ma (Rodrigues (McDougall *et al.* 1965)). However it should be noted that many of the exposed lavas in the Mascarenes are the result of volcanic reactivation with multiple phases of island building and erosion having occurred on each island (Montaggioni & Nativel 1988). There is evidence that La Réunion may be up to 5 Ma (Bonneville *et al.* 1988) and despite the young age of the Rodriguan lavas, it is possible that the island is as old as the seamount it sits upon (up to 15 Ma B. Warren *pers. comm.*).

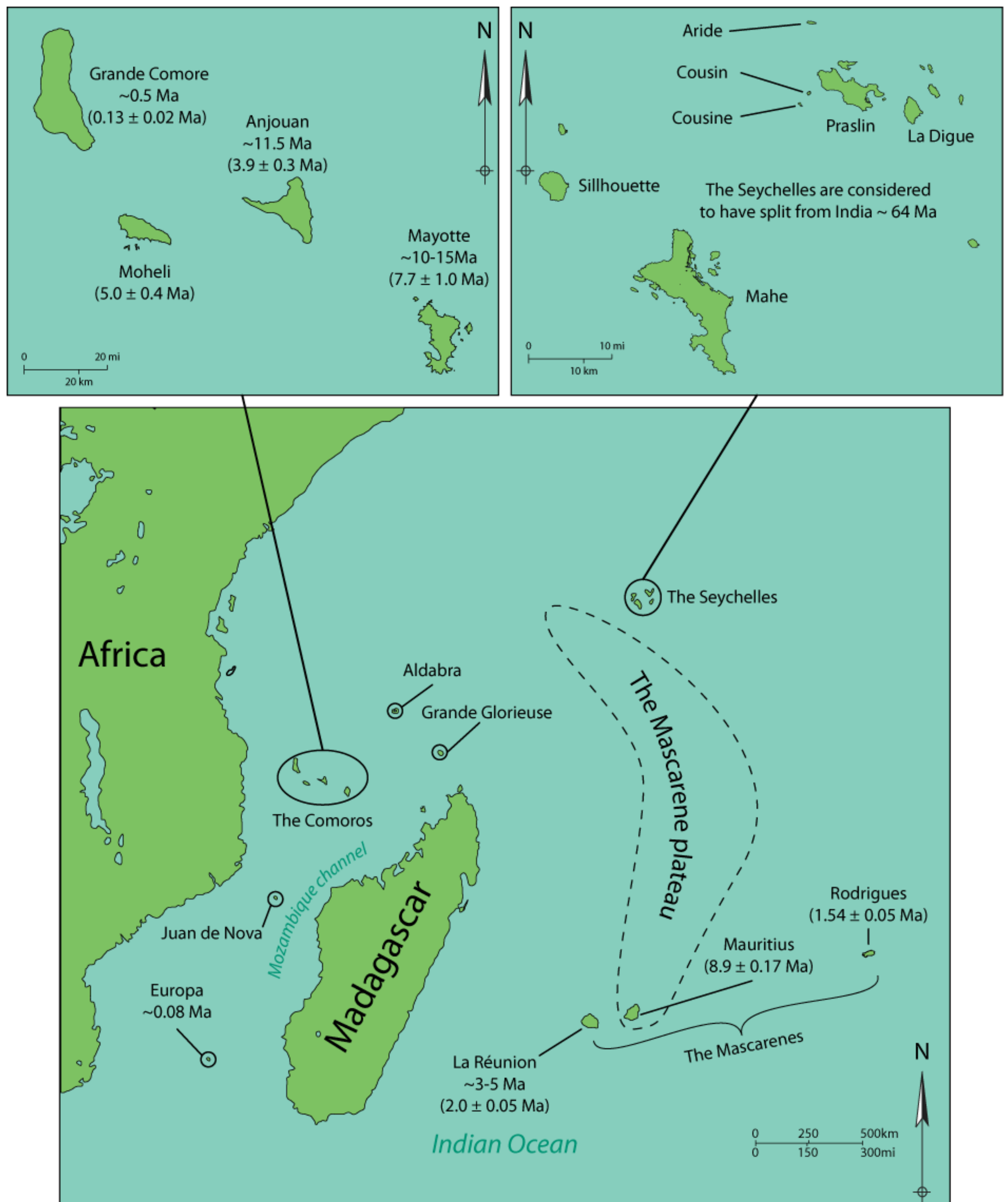


Fig 1.1: The current configuration of South-West Indian Ocean land masses. The first number associated with islands is the oldest age estimate available while the ages of the oldest radiometrically dated lavas are contained in brackets. References detailing these dates are in the text.

The youngest islands are the raised coralline islets of Europa, Juan de Nova, the Glorieuses, Farquars, Amerานties and Aldabra (including Assumption, Cosmoledo and Astove). The ages of these islands are generally unknown, but Europa is estimated to have emerged within the last 80,000 years (Battistini 1966)

and other islands in the Aldabra group may be less than 15,000 years old (Peake 1971). All of these islands are extremely low lying (e.g. Aldabra has a maximum elevation of approximately 8m above sea level) suggesting that they are extremely susceptible to sea level changes; at least one inundation of Grande Glorieuse and Aldabra is estimated to have occurred 125,000 years ago (Battistini & Cremer 1972; Thomson & Walton 1972). This inundation event is believed to have been sufficiently high to have eradicated the terrestrial fauna present at the time (Thomson & Walton 1972).

Sea levels are believed to have changed repeatedly throughout geological history (Haq *et al.* 1987; Miller *et al.* 2005), with the last 500,000 years being particularly variable due to repeated glacial cycles (Camoin *et al.* 2004; Colonna *et al.* 1996; Rohling *et al.* 1998). Several lowstands in the last 500,000 years are believed to have been as low as 145 m below present sea level (BPSL) in the Indian Ocean with lowstands of between 80 m and 100 m BPSL in the preceding 10 Ma (Haq *et al.* 1987). Lowstands such as these would have revealed a number of islands in the Mascarene plateau that are currently submerged. Figure 1.2 shows the effects of lowstands of 80 m and 135 m BPSL on island size compared to current sea levels. When islands are revealed by falling sea levels, the maximum distance that colonisers must traverse to reach a new island drops dramatically. As a consequence, the spread of flora and fauna around the islands of the Indian Ocean may have been facilitated by changing sea levels (Warren *et al.* 2010). Where applicable the following chapters address changes in sea levels and potential for colonisation.

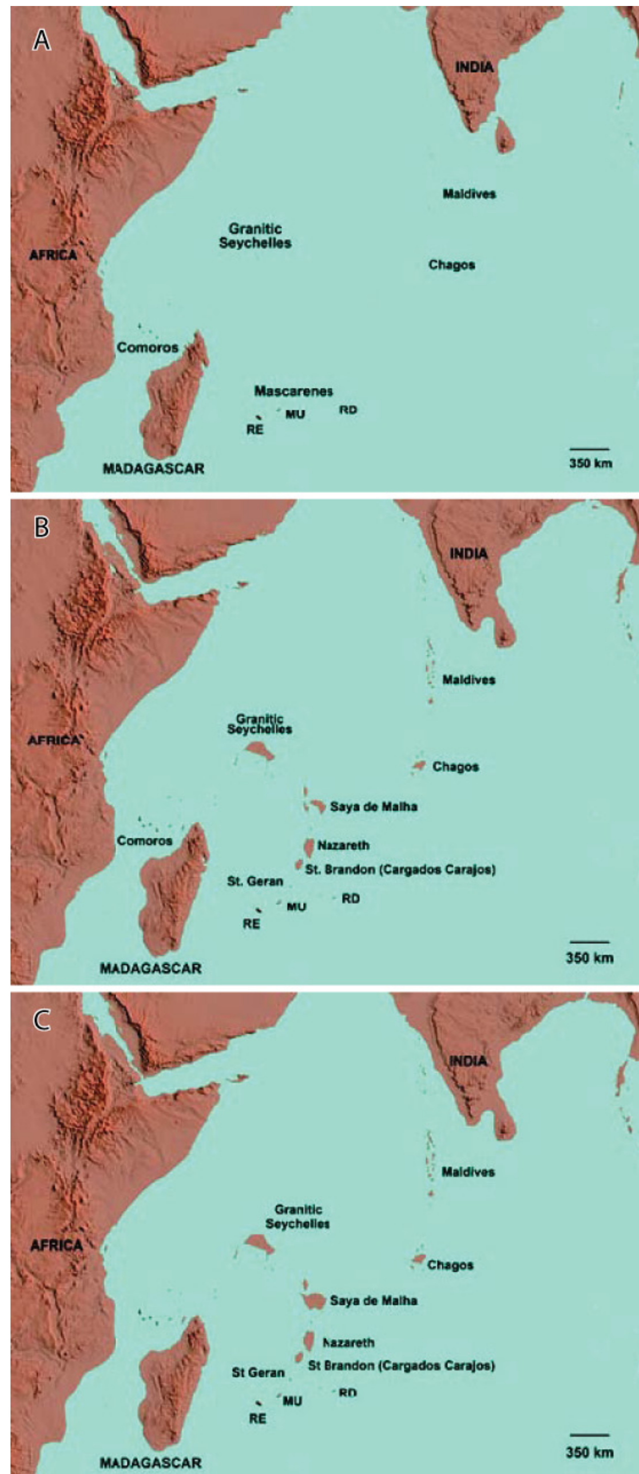


Fig 1.2: (A) Current island configuration in the Indian Ocean. (B) Estimated island extents during an 80 m BPSL lowstand. (C) Estimated island extents during a 135 m BPSL lowstand. Both lowstand maps assume constant ocean floor topography. Diagram taken from Warren *et al.* (2010). Abbreviated island labels: RE (La Réunion), MU (Mauritius) and RD (Rodrigues).

1.4 References

Alonso-Zarazaga MA, Lyal CHC (1999) A world catalogue of families and genera of Curculionoidea (Insecta: Coleoptera)(Excepting Scolytidae and Platypodidae), 315 pp. *Entomopraxis, Barcelona*.

Austin JJ, Arnold EN, Jones CG (2004) Reconstructing an island radiation using ancient and recent DNA: the extinct and living day geckos (Phelsuma) of the Mascarene islands. *Molecular Phylogenetics and Evolution* **31**, 109-122.

Battistini R (1966) La morphologie de l'Ile Europa. *Mem. mus. natn. Hist. Nat., Paris* **91**, 7-17.

Battistini R, Cremer G (1972) Geomorphology and vegetation of Iles Glorieues. *ATOLL RESEARCH BULLETIN* **159**, 153-169.

Bonneville A, Barriot JP, Bayer R (1988) Evidence From Geoid Data of a Hotspot Origin for the Southern Mascarene Plateau and Mascarene Islands (Indian Ocean). *J. Geophys. Res.* **93**, 4199-4212.

Camoin GF, Montaggioni LF, Braithwaite CJR (2004) Late glacial to post glacial sea levels in the Western Indian Ocean. *Marine Geology* **206**, 119-146.

Cegelski CC, Waits LP, Anderson NJ*, et al.* (2006) Genetic diversity and population structure of wolverine (*Gulo gulo*) populations at the southern edge of their current distribution in North America with implications for genetic viability. *Conservation Genetics* **7**, 197-211.

Champion H (1914) *The Percy Sladen Trust Expedition to the Indian Ocean in 1905.*

Clement WL, Tebbitt MC, Forrest LL*, et al.* (2004) Phylogenetic position and biogeography of Hillebrandia sandwicensis (Begoniaceae): A rare Hawaiian relict. *American Journal of Botany* **91**, 905-917.

Coffin MF, Rabinowitz PD (1987) Reconstruction of Madagascar and Africa: Evidence From the Davie Fracture Zone and Western Somali Basin. *J. Geophys. Res.* **92**, 9385-9406.

Colonna M, Casanova J, Dullo WC, Camoin G (1996) Sea-level changes and delta O-18 record for the past 34,000 yr from Mayotte reef, Indian Ocean. *Quaternary Research* **46**, 335-339.

Darwin C (1869) *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*, 5th edn. John Murray, London.

de Thoisy B, da Silva AG, Ruiz-Garcia M*, et al.* (2010) Population history, phylogeography, and conservation genetics of the last Neotropical mega-herbivore, the lowland tapir (Tapirus terrestris). *BMC Evolutionary Biology* **10**.

Emerick CM, Duncan RA (1982) Age progressive volcanism in the comores archipelago, western Indian Ocean and implications for Somali plate-tectonics. *Earth and Planetary Science Letters* **60**, 415-428.

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

Emerson BC, Oromi P, Hewitt GM (2000) Colonization and diversification of the species *Brachyderes rugatus* (Coleoptera) on the Canary Islands: Evidence from mitochondrial DNA COII gene sequences. *Evolution* **54**, 911-923.

Ferragu M, Richard R (1990) Trois nouveaux Curculionides de Madagascar et de l'Île de la Réunion [Coleoptera]. *Revue Française d'Entomologie (N.S.)* **12**, 105-110.

Ferragu M, Richard R (1993) Etude portant sur les Cratopus de l'Ile de la Réunion [Coleoptera, Curculionidae]. *Revue Française d'Entomologie (N.S.)* **14**, 111-116.

Fukami T, Beaumont HJE, Zhang XX, Rainey PB (2007) Immigration history controls diversification in experimental adaptive radiation. *Nature* **446**, 436-439.

Galman G, Matyot P, Voisin J-F (2011) Le genre Cratopus aux îles Seychelles (Coleoptera : Curculionidae). *Annales De La Societe Entomologique De France*.

Gillespie R (2004) Community assembly through adaptive radiation in Hawaiian spiders. *Science* **303**, 356-359.

Grobler GC, Bastos ADS, Chimimba CT, Chown SL (2011) Inter-island dispersal of flightless *Bothrometopus huntleyi* (Coleoptera: Curculionidae) from the sub-Antarctic Prince Edward Island archipelago. *Antarctic Science* **FirstView**, 1-10.

Haq BU, Hardenbol J, Vail PR (1987) Chronology of Fluctuating Sea Levels since the Triassic. *Science* **235**, 1156-1167.

Holland BS, Hadfield MG (2002) Islands within an island: phylogeography and conservation genetics of the endangered Hawaiian tree snail *Achatinella mustelina*. *Molecular Ecology* **11**, 365-375.

Hustache A (1919) Synopsis des Curculionides de la Faune Malgache. I. Brachydérides et Otiorrhynchides. *Annales De La Societe Entomologique De France* **87**, 441-520.

Hustache A (1920) Curculionides des Iles Mascareignes. *Extrait des Annales de la Societe Entomologique de France* **89**.

Illera JC, Emerson BC, Richardson DS (2007) Population history of Berthelot's pipit: colonization, gene flow and morphological divergence in Macaronesia. *Molecular Ecology* **16**, 4599-4612.

Jordal BH, Emerson BC, Hewitt GM (2006) Apparent 'sympatric' speciation in ecologically similar herbivorous beetles facilitated by multiple colonizations of an island. *Molecular Ecology* **15**, 2935-2947.

Jordal BH, Kirkendall LR, Harkestad K (2004) Phylogeny of a Macaronesian radiation: host-plant use and possible cryptic speciation in *Liparthrum* bark beetles. *Molecular Phylogenetics and Evolution* **31**, 554-571.

Jordan S, Simon C, Foote D, Englund RA (2005) Phylogeographic patterns of Hawaiian Megalagrion damselflies (Odonata : Coenagrionidae) correlate with Pleistocene island boundaries. *Molecular Ecology* **14**, 3457-3470.

Kingdon J (1990) *Island Africa: the evolution of Africa's rare animals and plants.* Collins, London.

Magnacca KN, Danforth BN (2006) Evolution and biogeography of native Hawaiian Hylaeus bees (Hymenoptera : Colletidae). *Cladistics* **22**, 393-411.

McDougal I (1971) Geochronology and evolution of the young volcanic island of Reunion, Indian Ocean. *Geochimica Et Cosmochimica Acta* **35**, 261-&.

McDougal I, Chamalaun F (1969) Isotopic Dating and Geomagnetic Polarity Studies on Volcanic Rocks from Mauritius, Indian Ocean. *Geological Society of America Bulletin* **80**, 1419-&.

McDougall I, Upton BGJ, Wadsworth WJ (1965) A Geological Reconnaissance of Rodriguez Island, Indian Ocean. *Nature* **206**, 26-27.

Menzel C, Mitra S, Waite G (2005) *Litchi and Longan: Botany, Cultivation and Uses* CABI Publishing.

Miller KG, Kominz MA, Browning JV, *et al.* (2005) The Phanerozoic Record of Global Sea-Level Change. *Science* **310**, 1293-1298.

Montaggioni L, Nativel P (1988) *La Reunion, Ile Maurice: Geologie et apercus biologiques, plantes et animaux* Masson, Paris.

Montaggioni L, Nougier J (1981) Les enclaves de roches detritiques dans les Volcans d'Anjouan (Archipel des Comores); Origine et interpretation dans le cadre de l'evolution du Canal de Mozambique. *Bulletin De La Societe Geologique De France* **23**, 595-601.

Moore J, White WM, Paul D, *et al.* (2011) Evolution of shield-building and rejuvenescent volcanism of Mauritius. *Journal of Volcanology and Geothermal Research* **207**, 47-66.

Nair K (2007) *Tropical Forest Insect Pests: Ecology, Impact, and Management* Cambridge University Press, Cambridge.

Newton AC, Allnutt TR, Gillies ACM, Lowe AJ, Ennos RA (1999) Molecular phylogeography, intraspecific variation and the conservation of tree species. *Trends in Ecology & Evolution* **14**, 140-145.

Nougier J, Cantagrel JM, Karche JP (1986) The Comores archipelago in the western Indian Ocean: volcanology, geochronology and geodynamic setting. *Journal of African Earth Sciences (1983)* **5**, 135-144.

Oberprieler RG, Marvaldi AE, Anderson RS (2007) Weevils, weevils, weevils everywhere. *Zootaxa*, 491-520.

Ortmann AE (1894) A Study of the Systematic and Geographical Distribution of the Decapod Family Atyidæ Kingsley. *Proceedings of the Academy of Natural Sciences of Philadelphia* **46**, 397-416.

Peake JF (1971) Evolution of Terrestrial Faunas in Western Indian-Ocean. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **260**, 581-&.

Plummer PS, Belle ER (1995) Mesozoic tectono-stratigraphic evolution of the Seychelles microcontinent. *Sedimentary Geology* **96**, 73-91.

Polihronakis M, Caterino MS, Chatzimanolis S (2010) Elucidating the phylogeographic structure among a mosaic of unisexual and bisexual populations of the weevil Geodercodes latipennis (Coleoptera: Curculionidae) in the Transverse Ranges of southern California. *Biological Journal of the Linnean Society* **101**, 935-948.

Poussereau J, Voisin J-F (2009) Cratopus francki, espèce nouvelle de la Réunion (Coleoptera, Curculionidae). *Bulletin de la Société entomologique de France* **114**, 361-364.

Rabinowitz PD, Coffin MF, Falvey D (1983) The Separation of Madagascar and Africa. *Science* **220**, 67-69.

Richard R (1957) Coleopteres Curculionidae de La Reunion. *Memoires de l'Institut Sceientifique de Madagascar* **8**.

Richard R (1958) Curculionides nouveaux ou mal connus des Comores, ile Glorieuse, ile Europa New or little known Curculionidae from the Comoro Islands, Glorioso Island and Europa Island. *Mem Inst Sci Madagascar Ser E Ent* **10**, 41-63.

Richard R (1961) Notes Sur Quelques Curculionides de l'ile Rodriguez. *The Mauritius Institute Bulletin* **5**.

Richard R (1977) Nouveaux Curculionides des iles Europa at de la Réunion. *Bulletin de la Société entomologique de France* **82**, 125-127.

Richard R (1995a) Description de deux Cratopini nouveaux de la région Malgache [Coleoptera: Curculionidae]. *Revue française d'entomologie* **17**, 63-65.

Richard R (1995b) Note sur l'apparition brusque d'une variation importante de la squamulation de Cratopus humeralis Boheman (Coleoptera, Curculionidae). *Bulletin de la Société entomologique de France* **100**, 303-306.

Rohling EJ, Fenton M, Jorissen FJ*, et al.* (1998) Magnitudes of sea-level lowstands of the past 500,000 years. *Nature* **394**, 162-165.

Sato A, O'hUigin C, Figueroa F*, et al.* (1999) Phylogeny of Darwin's finches as revealed by mtDNA sequences. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 5101-5106.

Sato A, Tichy H, O'hUigin C*, et al.* (2001) On the origin of Darwin's finches. *Molecular Biology and Evolution* **18**, 299-311.

Sequeira AS, Lanteri AA, Scataglini MA, Confalonieri VA, Farrell BD (2000) Are flightless Galapaganus weevils older than the Galapagos Islands they inhabit? *Heredity* **85**, 20-29.

Spurgin LG, van Oosterhout C, Illera JC*, et al.* (2011) Gene conversion rapidly generates major histocompatibility complex diversity in recently founded bird populations. *Molecular Ecology* **20**, 5213-5225.

Storey M, Mahoney JJ, Saunders AD*, et al.* (1995) Timing of hot spot related volcanism and the breakup of Madagascar and India. *Science* **267**, 852-855.

Thomson J, Walton A (1972) Redetermination of Chronology of Aldabra Atoll by 230Th/234U Dating. *Nature* **240**, 145-146.

Voisin J-F, Poussereau J (2007a) Charactérisation de Cratopus frappieri Deyrolle, 1862 (Coleoptera, Curculionidae). *Nouv. Revue Ent (N.S.)* **24**, 329-334.

Voisin J-F, Poussereau J (2007b) L'identité de Cratopus sumptuosus Boheman, 1834, de Cratopus exquisitus Boheman, 1840, et de Cratopus sandi Deyrolle, 1862, espèces endémiques de l'île de la Réunion (Coleoptera, Curculionidae). *Nouv. Revue Ent (N.S.)* **24**, 279-283.

Voisin J-F, Poussereau J (2009) A propos de Cratopus nanus et de Cratopus parvus de l'île de la Réunion (Coleoptera, Curculionidae). *Nouv. Revue Ent. (N.S.)* **26**, 59-65.

Warren BH, Strasberg D, Bruggemann JH, Robert PP-J, Christophe T (2010) Why does the biota of the Madagascar region have such a strong Asiatic flavour? *Cladistics* **26**, 526-538.

Williams JR, Cox ML (2003) A Contribution to the study of Mascarene Weevils of the genus *Cratopus* Schonherr (Coleoptera: Curculionidae: Entiminae: Cratopini): The species of Mauritius and Rodrigues. *The Mauritius Institute Bulletin* **12**.

Chapter 2: Molecular characterisation of a trophic shift within an island radiation of insect herbivores (Curculionidae: Entiminae: *Cratopus*).

Top left: Cratopus murinus, Ille aux Vacoas, Mauritius.
Top right: Cratopus ovalis, Snail Rock, Mauritius.
Bottom: Cratopus murinus feeding damage on Scaevola taccada, Ille de la Passe, Mauritius.

2.1 Abstract

The phytophagus beetle family Curculionidae is the most species rich insect family known. Much of this diversity has been attributed to both co-evolution with foodplants and host-shifts at key points in the group's early evolutionary history. Less well understood is the extent to which populations and individuals can host-shift over much more recent timescales. Here molecular techniques are used to (1) identify the most frequent dietary components for two Mauritian species of weevil, (2) examine how the most frequent dietary components vary across populations for one of the species and test for an adaptive shift vs. a neutral response to resource availability, (3) determine whether or not any host shift involves a phylogenetically novel host and (4) assess breadth of diet and whether or not trophic shifts are from specialist to specialist or generalist to specialist. The study species are polyphagous and most populations consume a much wider range of plants than would be suggested by the current literature with local diet being largely determined by food availability rather than by preferences indicative of an adaptive shift. Locally specialist populations are found to consume food plants that are not phylogenetically novel but which are also not consumed in other populations Molecular dating techniques are used to determine that these specialist populations are derived from more generalist ones. Finally a modified set of primers with improved performance for barcoding of intestinal contents are reported. This study demonstrates that using molecular methods, it is possible to unambiguously quantify variation in the most frequent dietary components across populations of polyphagous insect herbivores, demonstrate true polyphagy within individuals and begin to assess how local plant diversity can shape the dietary preferences of insect herbivores.

2.2 Introduction

Weevils are the most successful of all insect herbivore lineages, with approximately 62,000 described species and an estimated 220,000 species in total (Oberprieler *et al.* 2007). Interactions between weevils and angiosperms have been hypothesised to be a significant driver of the great diversity we see today (Farrell 1998; Oberprieler *et al.* 2007) under a scenario where much of the family level diversity evolved on gymnosperms and monocots before host shifting, to and

radiating across, angiosperms (McKenna *et al.* 2009). However, at more recent evolutionary timescales below the species level, or among closely related species, patterns of host use and changes in host use through time are less understood. Progress in this area has largely been made through the investigation of species that inhabit tissue of a host plant at some stage of their life cycle e.g. (Barat *et al.* 2008; Hernandez-Vera *et al.* 2010; Jordal *et al.* 2004; Podlussany *et al.* 2001), where a reliable host/herbivore relationship can be inferred. Trophic relationships are much more complicated to describe for weevil species where adult forms feed on plant foliage and larval stages do not develop inside host tissue. This is frequently the case for groups such as the Entiminine weevils (Curculionidae: Entiminae) where trophic ecology is particularly complicated to quantify for three reasons. Firstly, Entiminine weevils generally feed at night so direct feeding observations are rare (Morris 1997). Secondly, multiple Entiminine species may feed upon the same plant species (Williams 2000; Williams & Cox 2003) so directly attributing feeding damage to one species over another is difficult. Finally, weevils move through the environment (for example while searching for mates or looking for food plants) so there is no specific reason to suspect that the plant you collect any given weevil on is indicative of what it feeds upon.

Host choice tests under laboratory conditions provide one way to quantify the trophic ecology of insect herbivores (Barone 1998; Novotny *et al.* 2002; Novotny *et al.* 2006), but such experiments may either be costly, labour intensive or difficult to achieve in remote field conditions. In addition to these disadvantages, there is the obvious complication that feeding associations inferred under laboratory conditions may not be an accurate reflection of feeding associations under natural conditions. Studies such as that of Otte (1976) have attempted to circumvent these shortcomings by microscopically examining digestive system contents, but this approach is also extremely labour intensive. Recent advances in molecular ecology provide a new tool for the direct quantification of trophic interactions between plants and insect herbivores in the field (Jurado-Rivera *et al.* 2009), but have yet to be fully exploited for this purpose (but see Pinzón-Navarro *et al.* 2010).

Jurado-Rivera *et al.* (2009) have demonstrated that field caught chrysomelid beetles may contain sufficient plant tissue in their digestive system to permit the amplification of chloroplast DNA genes for the quantification of plant

species consumed across a range of diverse chrysomelid species. Here the approach of Jurado-Rivera *et al.* (2009) is extended for an analysis of trophic ecology within a radiation of Entiminine weevils across the Mascarene Islands, part of the southwest Indian Ocean area biodiversity hotspot (Myers *et al.* 2000). The genus *Cratopus* is represented by approximately 100 species distributed across the islands of the Indian Ocean, with the great majority of species described being restricted to the Mascarene Islands, with approximately 80 species across La Réunion, Mauritius and Rodrigues. The genus is particularly unusual for an oceanic island radiation of beetles, in that all but a few species are flighted. All other Entiminine radiations studied on oceanic islands to date involve flightless lineages. These include *Syzygops* on the Mascarene Islands (Williams 2000), *Laparocerus* within Macaronesia (Machado 2006, 2007a, b, 2008; Machado *et al.* 2008), *Rhyncogonus* within Polynesia (Gillespie *et al.* 2008), *Ectemnorhinus* of the Prince Edward Islands (Grobler *et al.* 2006) and *Galapaganus* from the Galapagos Islands (Sequeira *et al.* 2008). While geographic isolation within islands has apparently acted to facilitate speciation within island invertebrate radiations (e.g. Holland & Hadfield 2002), this would appear to hold less explanatory power for a flighted group such as *Cratopus*.

The available information on the trophic ecology of *Cratopus*, although limited, indicates that they feed on a broad range of plant species, with Mauritian species ranging from generalists to specialists (Williams & Cox 2003). This suggests the potential for *Cratopus* species to have diversified with respect to plant feeding relationships, but given the complications described above for quantifying trophic ecology within Entiminine species, details of specific plant feeding relationships are lacking. In this study attention is focussed on the Mauritian species *Cratopus murinus*, and its close relative *Cratopus ovalis*. Both species are restricted to isolated remnant dry forests that cover the mountains of the Mauritian central plateau (Williams & Cox 2003), with *C. murinus* also occurring on several small coralline islets off the east coast of Mauritius. Within these habitats, a range of plant species exhibit foliage damage typical of Entiminine feeding and both species can be collected from a range of plant species, including species not exhibiting feeding damage to foliage. In contrast, the islet populations of *C. murinus* are found only on one plant species, *Scaevola taccada*, a species that does not naturally occur within the dry forest habitat of the main island. *Cratopus murinus*

populations from the islets also have atrophied remnants of wings, suggesting these populations to be derived from mainland source populations.

Cratopus murinus and *C. ovalis* were sampled from across Mauritius to apply a molecular approach to quantify (1) geographic structure of genetic variation within species and (2) the most important components of diet. Multi-gene approaches are used to quantify population structure and phylogeographic history within both species, with a specific focus on identifying the evolutionary and temporal origin of the coralline islet populations of *C. murinus*. Chloroplast DNA barcoding efficiency was assessed for an alternative set of primers to those used previously (Jurado-Rivera et al., 2009: Pinzón-Navarro et al. 2010) to address the following aims: (1) quantify the plant species most frequently consumed by *C. ovalis* and *C. murinus*; (2) quantify trophic variation among dry forest populations of *C. murinus* and test between explanations of adaptive shift, or neutral response to resource availability; (3) determine whether *S. taccada* represents the acquisition of a phylogenetically novel food plant (i.e. is *S. taccada* closely or distantly related to one or more plant species consumed within mainland populations); (4) distinguish between a specialist to specialist transition, or a generalist to specialist transition, for the trophic shift to *S. taccada* with the colonisation of coralline islets.

2.3 Methods

2.3.1 Beetle sampling

Beetles were collected by foliage beating on Mauritius in April 2009. Samples were placed directly in 99% ethanol solution, with the exception of a few individuals that were kept alive until the same evening for photographing, before being placed in 99% ethanol. Sampling sites were recorded on a handheld GPS unit (Garmin GPS 60, Garmin Ltd). A total of 201 beetles were collected across 9 sites (136 *C. murinus* and 65 *C. ovalis*). See Table 2.1 for descriptions and locations of sites.

Table 2.1: Names and locations of sampled populations with description of habitat.

Site	GPS co-ordinates	Description
Corps de Garde	20°15'35.72"S 57°27'7.25"E	Mountainous dry forest/scrub
Ile aux Vacoas	20°23'51.84"S 57°46'12.78"E	Coralline islet
Ile de la Passe	20°23'56.26"S 57°46'2.29"E	Coralline islet
Ile Marianne	20°22'49.47"S 57°47'14.22"E	Coralline islet
Snail Rock	20°11'30.09"S 57°30'24.20"E	Dry forest/forestry commission land
Mondrain	20°19'30.65"S 57°27'19.41"E	Botanic Garden
Lion Mountain	20°21'42.96"S 57°43'31.48"E	Semi-dry forest surrounded by sugar cane fields
Trois Mamelles	20°18'46.51"S 57°26'56.74"E	Dry forest/private hunting reserve
Yemen	20°17'57.54"S 57°24'39.18"E	Remnant dry forest
Chamarel	20°25'49.50"S 57°22'27.78"E	Partially restored dry forest

2.3.2 Plant sampling

Plant sampling was carried out by Claudia Baider and Vincent Florens between 5th September 2010 and 1st October 2010. For every GPS co-ordinate where beetles were sampled in Corps de Garde and Snail Rock (the two sites with sufficient samples of *C. murinus* for more detailed analysis), a 5m-radius circle was marked and a complete survey of vascular plants was undertaken. Leaf samples of all species were taken and preserved in silica gel. Plant reference sequences were obtained by sequencing DNA extracted from identified plant tissue and amplifying with the same primers as used for the beetle dietary analysis (see below).

2.3.3 DNA extraction, PCR amplification and sequencing

DNA was extracted from the head and pronotum of beetles using the DNeasy 96 well Blood and Tissue Extraction kit (QIAGEN, West Sussex, UK) with the digestion buffer volumes amended for large specimens as recommended by the manufacturer. Plant leaf samples were ground by hand and extracted using the DNeasy Plant kit (QIAGEN, West Sussex, UK) as per the manufacturer's instructions. To undertake a genealogical analysis of population structure within beetle species the mitochondrial gene Cytochrome Oxidase II (COII) was sequenced along with two nuclear genes: arginine kinase (ArgK) and internal transcribed spacer 2 (ITS2). Beetle feeding records were obtained directly from beetle DNA extractions by amplifying the plant plasmid gene transfer RNA Leucine UAA (trnL). Primers for each gene are listed in Table 2.2. Primers COIICraF and COIICraR were designed by aligning Curcujiform COII sequences from GenBank and selecting conserved regions as close to the COII gene as possible. COII

amplification conditions were; 0.5 mM of each primer, 5 mM $MgCl_2$ and a thermal profile of: 95°C for 60s, 58°C for 60s and 72°C for 90s, 40 cycles. For ArgK, the primers ArgKforB2 and ArgKrevB1 (attributed to Danforth *et al.* (2005) in McKenna *et al.* (2009)) were used with a thermal profile of: 95°C for 60s, 50°C for 60s and 72°C for 120s, 37 cycles. ITS2 was amplified using the primers CAS28sB1d (Ji *et al.* 2003) and M13REV-CAS5 (CAS5p8sFt in *Ji et al.* (2003) but modified to have a M13REV tail (Cho *et al.* 1995; Regier & Shi 2005)) with a thermal profile of; 95°C for 40s, 50°C for 60s and 72°C for 40s, 35 cycles. All PCRs had an initial denaturing step of 94°C for 5 minutes and a final extension step of 72°C for 2 minutes.

Table 2.2: PCR and sequencing primers used in this study. Primers suffixed with (sequencing) were only used during the sequencing reactions.

Primer Name	Sequence (5'-3')	Target
COIICraF	TAATATGGCAGAWTAGTGCAATGGA	COII
COIICraR	TGCTTTCAGTCATCTAATGATCTRTTTACAGA	COII
ArgKforB2	GAYTCCGGWATYGGWATCTAYGCTCC	ArgK
ArgKrevB1	TCNGTRAGRCCCATWCGTCTC	ArgK
ArgMurf1	CCGGATTCACCAAGAGYGAYAARCA	ArgK (sequencing)
M13REV-CAS5	CAGGAAACAGCTATGACCTGAACATCGACATTTYGAACGCATAT	ITS2
CAS28sB1d	TTCTTTTCCTTCSCTTAYTRATATGCTTAA	ITS2
M13REV	CAGGAAACAGCTATGACC	ITS2 (sequencing)
A49325	CGAAATCGGTAGACGCTACG	trnL
M13(-21)B49863	TGTAAAACGACGGCCAGTGGGGATAGAGGGACTTGAAC	trnL
H-rev	GATAGGTGCAGAGACTCAATG	trnL (sequencing)

For the trnL amplifications a range of primer combinations were explored and amplification conditions with primers described in Taberlet *et al.* (2007) and tailed modifications of these. Previous studies (Jurado-Rivera *et al.* 2009; Pinzón-Navarro *et al.* 2010) have used primers A49325 and B49863 (Taberlet *et al.* 2007), but with limited success. Although Jurado-Rivera *et al.* (2009) did not report amplification success rates, Pinzón-Navarro *et al.* (2010) reported only a 36% amplification success rate for beetles sampled from the field. Preliminary investigations achieved improved amplification success with an alternative pair of primers, A49425 (Taberlet *et al.* 2007) and M13(-21)B49863, a modification of B49863 by the incorporation of an M13(-21) tail to improve PCR yield (Regier & Shi 2005). A touchdown PCR protocol (Don *et al.* 1991) was used with a thermal profile of: 94°C for 60s, 60°C for 60s (touched down by 1°C per cycle for 17 cycles) and 72°C for 60s, followed by 29 cycles with an annealing temperature of 42°C,

giving a range of annealing temperatures between 60°C and 42°C and 46 cycles in total. PCR products comprised of more than one band were separated on a 2% agarose gel and individual bands were sampled from the gel with a pipette tip and resuspended in 30ul of AE buffer (QIAGEN, West Sussex, UK). Resuspensions were used as template for subsequent PCR reactions using the same primers and conditions as above.

Sequencing reactions were performed with the Big Dye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems, California, USA). For COII, the primer COIICraF was used, and an internal sequencing primer, ArgMurF1, was designed for ArgK. CAS28sB1d and M13REV were used for ITS2 sequencing and for trnL the primer H-rev (a reverse form of primer H (Taberlet *et al.* 2007) was used. The thermal profile used for all sequencing primers was: 96°C for 10s, 50°C for 5s and 60°C for 240s, 25 cycles, with the exception of COIICraF, for which an annealing temperature of 58°C was used. Sequences were read on a 3730XL sequencer (Applied Biosystems).

All sequences were checked and ambiguous bases called in Geneious Pro version 5.4 (Drummond *et al.* 2011). Sequences were aligned using MAFFT v6.814b (Katoh *et al.* 2002) with the following parameter values: scoring matrix 200PAM/k=2, Gap open penalty = 1.53, Offset value = 0.123. The alignments produced were then checked by eye. For the nuclear loci, alleles were resolved for heterozygotes involving nucleotide polymorphism either by direct comparison to homozygous sequences or with PHASE 2.1 (Stephens & Scheet 2005; Stephens *et al.* 2001). Indel heterozygotes within ITS2 were resolved by sequencing individuals with both forward and reverse primers. The readable portions of each read where then aligned against all known unambiguous homozygote sequences allowing both haplotypes to be inferred for each individual as demonstrated in Flot *et al.* (2006) and Peters *et al.* (2007).

2.3.4 Phylogenetic, network and population genetic analyses

For the mtDNA COII sequence data phylogenetic trees were constructed using Bayesian Inference with MrBayes 3.2 (Huelsenbeck & Ronquist 2001) with *C. melanocephalus* and *C. vulgaris* as outgroups. Four analyses were run; each for 10 million generations using 4 MCMC chains discarding 25% of the samples as burnin with a model of sequence evolution determined using jModelTest (Posada 2008)

under the BIC and AIC criteria. All parameters permitted under this model were estimated. The output was assessed for stationarity and convergence in Tracer v1.5.0 (Rambaut & Drummond 2007) and the consensus tree was visualised in FigTree v1.3.1 (Rambaut 2011). To visualise relationships among the less divergent nuclear gene sequences, haplotype networks were constructed with TCS 1.21 (Clement *et al.* 2000) for non-recombinant sections within each gene, as in Jordal *et al.* (2006). TCS uses statistical parsimony to infer relationships amongst very closely related sequences. These are represented as a network. When TCS encounters recombination or alternative mutational pathways with equal parsimony scores, loops are added. For the quantification of non-recombinant sections both genes were analysed with the recombination detection program, Dual Brothers (Minin *et al.* 2005) with the default MCMC settings. Dual Brothers detects recombination using a Bayesian sliding window approach that detects putative recombination sites by comparing the probability of the observed portion of the alignment having arisen by mutation to it having arisen by recombination. Two possible methods for detecting the potential recombination points are employed; firstly by examining how inferred phylogenetic history changes across the alignment and secondly by examining how substitution parameters change across the alignment. Average and maximum pairwise distances within gene partitions were calculated in MEGA 5 (Tamura *et al.* 2011). Pairwise F_{st} statistics were also generated among sampling locations for nuclear sequence data using Arlequin 3.5 (Excoffier & Lischer 2010).

2.3.5 Divergence time estimation

Three approaches were taken to estimate the timing of divergence between the flightless coral islets populations of *C. murinus* and populations of the main island. Each of the three approaches uses Bayesian inference but the actual analyses are very different; IMa2 (Hey & Nielsen 2007 & Hey 2010) uses an isolation with migration model that allows for incomplete isolation between populations, BEAST (Drummond & Rambaut 2007) assumes complete isolation between populations but allows for differing tree models to be applied to different sections of the tree and *BEAST (Heled & Drummond 2010) allows for the similtaneous estimation of divergence times from possibly conflicting genealogies. The isolation with migration model implemented within a Bayesian framework is described in Hey &

Nielsen (2007) and Hey (2010). An IMa2 input file was generated containing COII, ArgK and ITS2 with all sites containing gaps removed. ArgK and ITS2 were tested for recombination using the program DualBrothers (Minin *et al.* 2005) and reduced to the single non-recombinant section containing the most segregating sites if necessary. The HKY model was implemented for COII, with the infinite sites model implemented for the ArgK and ITS2 partitions. A coleopteran COII mutation rate of 0.0154 substitutions/site/myr was taken from Cicconardi *at al.* (2010) and converted to a mutation rate in substitutions/locus/year for the input file as specified by the program's author. To allow for uncertainty surrounding the true mutation rate, the mutation rate was allowed to vary in a normal fashion with a standard deviation of 0.075 substitutions/site/myr. An 'M mode' analysis was performed using 100 geometrically heated Markov chains (heating values between 0.6 and 0.975) and stationarity was assessed every 1.5 million generations. Eight parallel analyses were performed to assess convergence. The parameters saved in 'M mode' were then used to manually calculate the splitting time. The BEAST analysis was performed with only the COII partition, employing a constant size coalescent model for all nodes corresponding to intra-population level variation, and a Yule model specified for all the remainder of the tree, following the approach of Ho *et al.* (2008). Three replicate analyses were performed using the best fit model characterised by jModelTest and an uncorrelated relaxed clock (Drummond *et al.* 2006). Analyses were run for 100 million generations, and convergence and stationarity were assessed with Tracer v1.5.0. Finaly *BEAST was used with the complete COII and ITS2 (gaps included) data sets and the largest non-recombinant section of ArgK. An uncorrelated relaxed clock was used for all three partitions and substitution models were determined using jModelTest. Three separate analyses, each of 500 million generations were run and convergence and stationarity were assessed with Tracer v1.5.0.

2.3.6 Species assignment of trnL sequences

TrnL sequences obtained from beetle DNA extracts were grouped according to sequence identity. A representative of each group was then compared to trnL sequences obtained from leaf samples taken during the floral surveys. The most similar match was considered to be the most likely identity of the foodplant. In cases where a sequence obtained from a beetle could not be matched to any of the

reference sequences, the program SAP (Munch *et al.* 2008) was used to determine the most likely closest match on Genbank. SAP achieves this by searching for each ambiguous query in the nucleotide database provided by NCBI (National Centre for Biotechnology Information) using BLAST (Basic Local Alignment Search Tool) (Altschul *et al.* 1990) and downloading the top 100 results. These are then aligned and a Bayesian tree is made using the alignment. Posterior probabilities in the SAP tree are then interpreted as a measure of how well the closest matches in the NCBI nucleotide database fit the query sequence. The result of the SAP analysis was then compared to the floral list of the Mascarenes to determine which Mauritian species were the most probable identifications.

To compare the breadth of the observed diets of both *C. murinus* and *C. ovalis* the current APG III classification of angiosperms (The Angiosperm Phylogeny Group 2009) was used. For this study dietary breadth is defined as the sum of the branch lengths of the APG III tree that that include all food plants. The phylogenetic placement of *Scaevola taccada,* with regard to the food plants consumed by mainland populations of *C. murinus*, is used as a measure of the novelty of *S. taccada* as a food plant.

2.3.7 Statistical analyses

To assess whether or not changes in diet have evolved among populations of *C. murinus*, tests for significant dietary differences among populations must be performed, then whether any such differences are functional, or a response to differing resource availability among collecting sites evaluated. Analyses were performed on the two sites sampled for the greatest number of *C. murinus* where vegetation surveys were subsequently undertaken, providing data on both the availability and frequency of each plant species. As feeding records are in the form of count data containing zeros, generalised linear models (GLMs) with a quasipoisson error structure were used to simultaneously examine the effects of site and food availability on feeding frequency. The number of sampling plots containing a given plant species was used as a proxy for the availability of each plant species within a site. All analyses were performed in R 2.15.0 (R Development Core Team 2010).

2.4 Results

2.4.1 MtDNA and nuclear gene sequencing

Of the 201 beetles sampled, 182 (123 *C. murinus* and 59 *C. ovalis*) were successfully amplified and sequenced for the mtDNA COII gene (GenBank accession numbers JN982748 - JN982931), yielding sequences of 607bp. One hundred and forty sites were variable across both species, with 107 sites variable within *C. murinus* and 53 within *C. ovalis.* There were 69 unique mitotypes (41 in *C. murinus* and 28 in *C. ovalis*) and the average pairwise p-distance was 7.23% across both species with a maximum of 10.38%. Average pairwise p-distance for *C. murinus* sequences was 5.21% with a maximum of 10.38% and the average pairwise distance for *C. ovalis* was 2.07% with a maximum of 3.62%. For ArgK 554bp of sequence was obtained for 118 *C. murinus* (236 haplotypes) and 57 *C. ovalis* (114 haplotypes) with 34 unique haplotypes found within *C. murinus* and 21 within *C. ovalis* gene (GenBank accession numbers JN987963 - JN988312). Across both species there are 41 variable sites with an average pairwise p-distance of 1.32% and a maximum of 2.53%. The average pairwise distance for *C. murinus* sequences was 1.01% with a maximum of 2.53% and the average pairwise distance for C. ovalis was 0.56% with a maximum of 1.08%. ITS2 sequences were obtained for 69 *C. murinus* (138 haplotypes) and 49 *C. ovalis* (98 haplotypes). Although 98% of specimens were successfully amplified for ITS2, sequencing success was much lower (58.7%) due to a high frequency of heterozygotes with multiple indels that could not be resolved. *Cratopus murinus* ITS2 sequences varied in length from 337bp to 348bp and *C. ovalis* sequences varied from 336bp to 337bp with 24 unique haplotypes within *C. murinus*, and 12 within *C. ovalis*. An alignment of both species was 366bp long with 88 variable sites, an average pairwise p-distance of 4.96% and a maximum pairwise distance of 10.94%. An alignment of *C. murinus* sequences was 366bp in length with 56 variable sites, an average pairwise p-distance of 1.61% and a maximum of 5.09%. An alignment of *C. ovalis* sequences was 337bp long with 13 variable sites, an average pairwise p-distance of 0.87% and a maximum pairwise p-distance of 2.08%.

2.4.2 Phylogenetic analyses and haplotype networks

The Bayesian analysis (Fig 2.1) reveals 8 well supported mtDNA lineages. Cratopus *murinus* is divided into 6 main lineages: (1) all Corps de Garde samples, plus some individuals from Mondrain and Yemen; (2) all samples from Snail Rock; (3) all coralline islet samples; (4 & 5) all samples from Trois Mamelles, with a single sample from Yemen within lineage 4 and (6) all samples from Lion Mountain. *Cratopus ovalis* is divided into two distinct lineages: (7) samples from Snail Rock, and (8) samples from all sites.

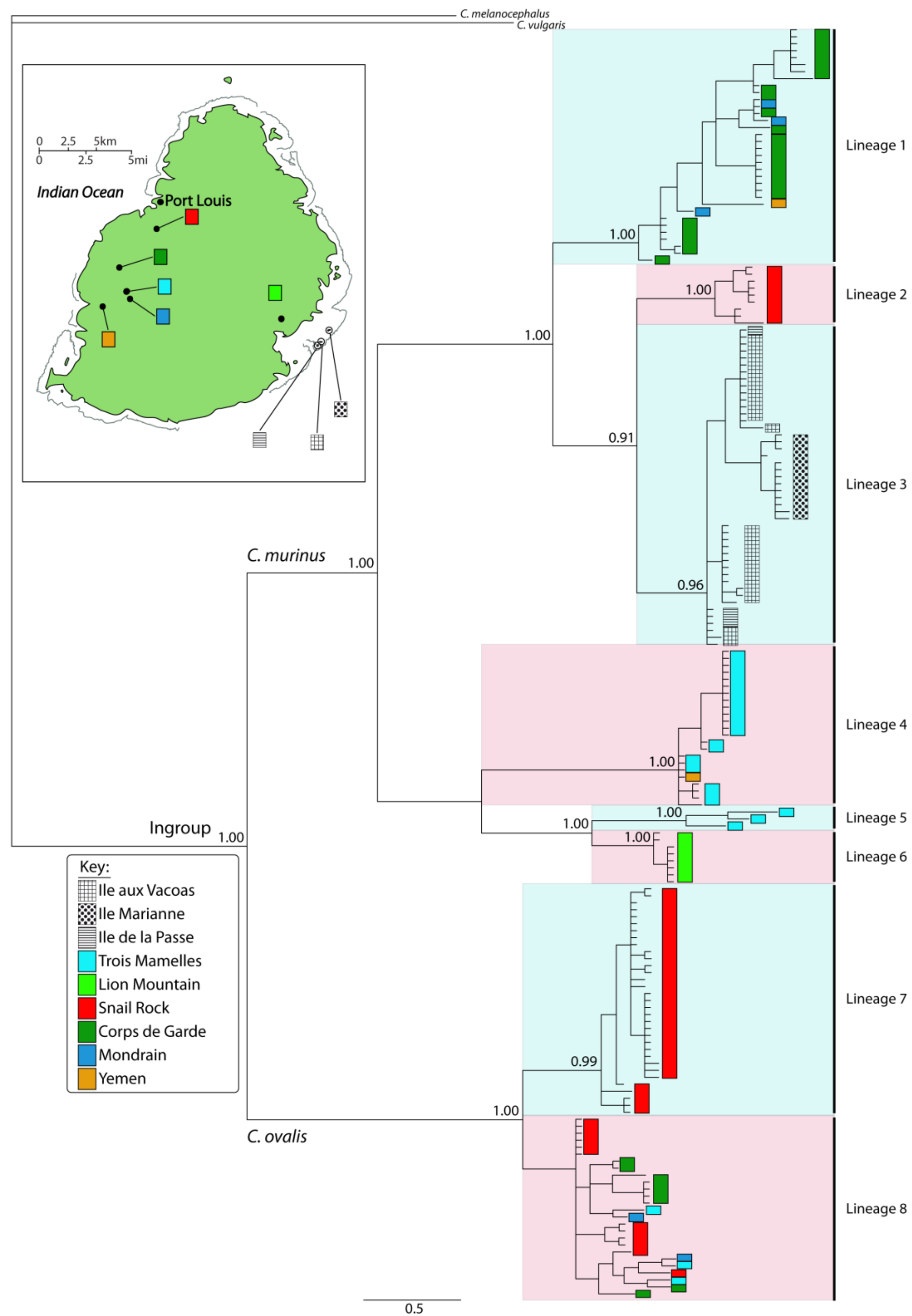


Figure 2.1. Mr Bayes consensus tree for *C. murinus* and *C. ovalis* with *C. melanocephalus* and *C. vulgaris* as outgroups. The analysis was performed using 4 MCMC chains for 10 million generations with the evolutionary model HKY+G. Values above the nodes are posterior probabilities for each node. Nodes used to define tmrcas in BEAST analyses and populations for pairwise F_{st} calculations are labelled. Insert is a map of Mauritius with collecting sites marked.

Recombination analyses performed on the ArgK alignment using the program DualBrothers revealed there to be at least 3 recombination breakpoints within this locus, so the alignment was divided into three partitions of length 30bp, 200bp and 208bp for network analyses. The remainder of the locus was unusable

due to limited information content within non-recombinant regions. No recombination was detected in the ITS2 alignment. For network analyses, an ITS2 alignment was produced with gaps removed, resulting in an alignment length of 339bp, with an average pairwise p-distance of 4.68% and a maximum of 7.67%.

Haplotype networks were generated for each of the three informative non-recombinant ArgK partitions (Fig. 2.2). The first ArgK partition recovers *C. ovalis* as distinct from *C. murinus*, with no haplotype sharing and monophyly for both species. The most ancestral haplotype within *C. murinus* is defined by the network placement of *C. ovalis* haplotypes and reveals that haplotypes from Trois Mamelles and Lion Mountain are derived from haplotypes shared across Snail Rock, Corps de Garde, and the coralline islets. The second ArgK partition mostly distinguishes *C. murinus* from *C. ovalis*, with one haplotype being shared across these two species. Haplotypes unique to the coral islets are derived from a haplotype distributed across the islets and several mainland sites. The probable ancestral *C. murinus* haplotype is found in Trois Mamelles with the single haplotype from Lion Mountain being derived from this. As with the first partition, haplotypes from the coralline islets are shared with, or derived from, a haplotype found in Snail Rock and Corps de Garde. Assuming that *C. murinus* and *C. ovalis* are good species, the shared haplotype may represent an ancient retained polymorphism or it could be evidence of historic hybridisation between the two species. Ancient hybridisation could possibly explain the differences between partitions one and two as each could potentially have had very different evolutionary histories. The third ArgK partition has few informative sites, so lacks resolution, but does resolve lineage 6 in a manner consistent with the COII tree.

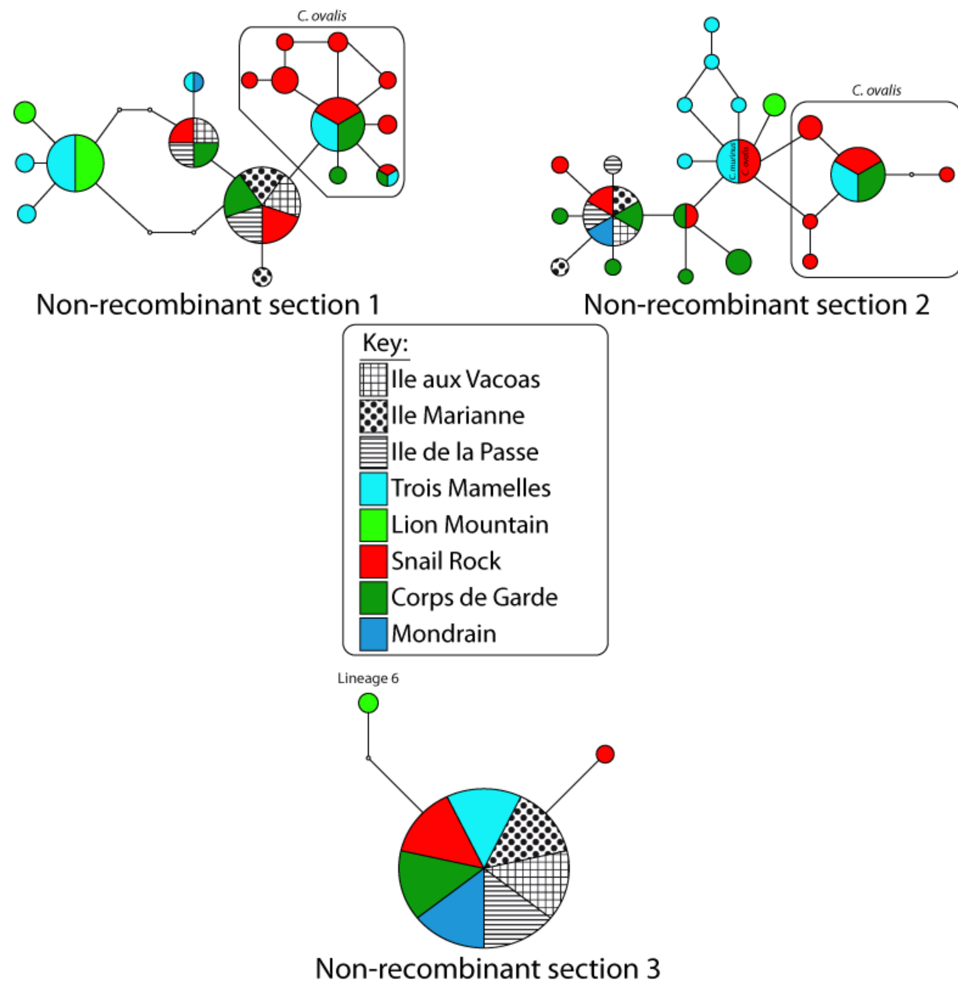


Figure 2.2. Haplotype networks derived from putatively non-recombined sections of ArgK as identified by the Dual Brothers analysis. The area of each haplotype is proportional to its frequency.

A single haplotype network was generated from the ITS2 haplotype data (Fig. 2.3), with some minor reticulation. *Cratopus murinus* and *C. ovalis* are clearly distinct, with the ancestral haplotype within *C. murinus* found in Snail Rock and the coralline islets. Corps de Garde and Snail Rock haplotypes form a closely related group, from which Lion Mountain and Trois Mamelles haplotypes are derived.

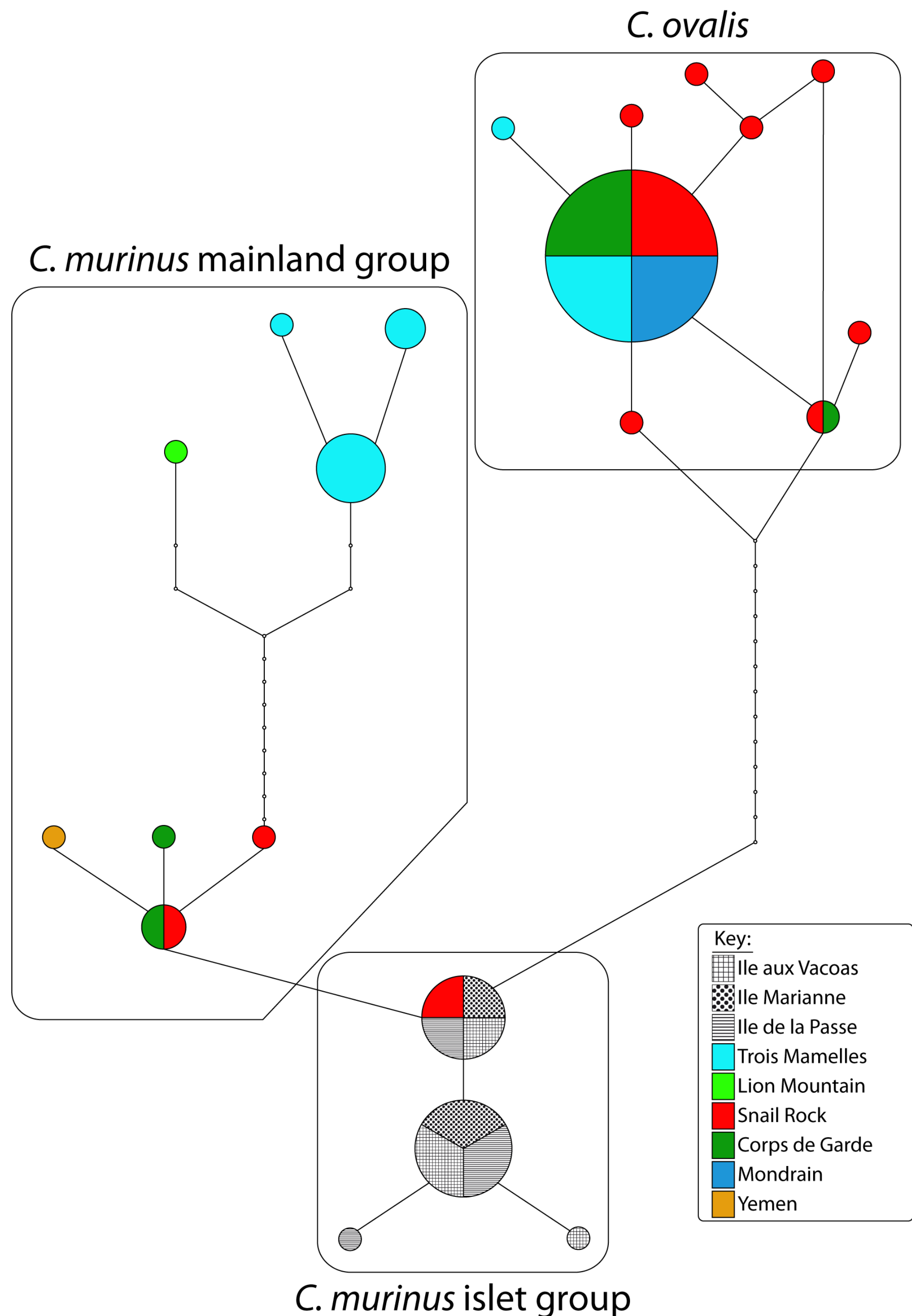


Figure 2.3. Haplotype network produced by TCS for the ITS2 alignment with gaps removed. The area of each haplotype is proportional to its frequency.

2.4.3 Population genetic analyses and cryptic species assessment

Pairwise F_{st} values were calculated for both nuclear loci between all sampling locations for *C. murinus* and *C. ovalis,* with the exception of Mondrain and Yemen (see Discussion). All sampling locations are significantly and often highly differentiated for both loci for *C. murinus* (Table 2.3). In comparison, populations

of *C. ovalis* are on average less differentiated, with some sites revealing no statistically significant differentiation (Table 2.4).

Table 2.3. Pairwise F_{st} values between sampling locations for *C. murinus.* All values are significant (significance level = 0.05). Values above the diagonal are calculated using ITS2 data, those below the diagonal are calculated using ArgK.

	Corps de Garde	Corraline islets	Snail Rock	Lion Mountain	Trois Mamelles
Corps de Garde	-	0.58891	0.18046	0.90129	0.95221
Corraline islets	0.27	-	0.68357	0.93589	0.94921
Snail Rock	0.15	0.13	-	0.98733	0.98245
Lion Mountain	0.80	0.83	0.85	-	0.96731
Trois Mamelles	0.77	0.82	0.86	0.86	-

Table 2.4. Pairwise F_{st} values between sampling locations for *C. ovalis.* Non-significant results are formatted in bold (significance level = 0.05). Values above the diagonal are calculated using ITS2 data, those below the diagonal are calculated using ArgK.

	Corps de Garde	Snail Rock	Trois Mamelles
Corps de Garde	-	0.09102	0.434
Snail Rock	0.05179	-	0.73695
Trois Mamelles	**-0.03067**	**0.02431**	-

The existence of divergent mtDNA lineages in sympatry provides an opportunity to test for the existence of cryptic species when combined with the analysis of nuclear genetic data by testing for linkage disequilibrium (e.g. Cicconardi *et al.* 2010). Divergent mtDNA lineages co-occurring in sympatry are found within both *C. murinus* (Trois Mamelles) and *C. ovalis* (Snail Rock), and the biological significance of these was evaluated using pairwise F_{st} statistics generated from nuclear sequence data. Within Trois Mamelles, two mtDNA lineages with an average p-distance of 7.21% were found within *C. murinus*. These two lineages are not significantly differentiated at either nuclear locus (ArgK: F_{st} = -0.02, p = 0.62; ITS2: F_{st} = -0.08, p > 0.99). Within Snail Rock *C. ovalis* is represented by two divergent mtDNA lineages with an average p-distance of 2.73% and again, the lineages were not significantly differentiated at either nuclear locus (ArgK: F_{st} = -0.01, p = 0.75; ITS2: F_{st} = 0.00, p = 0.50).

2.4.4 Divergence times

IMa2 analyses returned a t0 parameter of 6.29 which when divided by the mutation rate in substitutions/locus/year gives a splitting time of 0.67 mya (95% HPD interval: 0.01 mya - 1.6 mya) for the divergence between *C. murinus* populations of the main island and the corraline islets. The BEAST analysis places a similar age on the split with a tmrca for the Snail Rock and corraline islet populations of 0.89 mya (95% HPD interval: 0.59 mya – 1.23 mya, ESS = 1992.68) and a tmrca for the corraline islet populations of 0.68 mya (95% HPD interval: 0.42 mya – 0.95 mya, ESS = 3789.91). The *BEAST estimate for the divergence between the corraline islet and Snail Rock populations is slightly older at 1.20 mya (95% HPD interval: 0.22 mya – 2.39 mya, ESS = 2168.04).

2.4.5 Plant analyses

Across all sites, clearly readable trnL sequences were obtained from 160 of the 201 sampled beetles, comprising 104 *C. murinus* and 56 *C. ovalis*, an overall success rate of 80%. This represents a substantial improvement over the 36% success rate reported in the only other comparable study to date (Pinzón-Navarro *et al.* 2010). Of the 160 individuals, 16 yielded two trnL sequences from differently sized bands that were individually reamplified. An additional 9 individuals produced mixed trnL sequence traces that can be attributed to more than one trnL PCR product of similar length. Thus in real terms the success rate is 85%. Thirty two individuals either did not amplify at all or produced sequences of poor quality. TrnL sequence length varied between 255bp and 409bp and a total of 17 different plant species were amplified from beetle DNA extracts. Across all sites, fourteen plant species that could be matched to reference sequences were consumed by *C. murinus* (Table 2.5) and six by *C. ovalis* (Table 2.6), with five species in common. Species of *Eugenia* could not be distinguished by trnL sequences, precluding specific identification of *Eugenia* foodplants. Two sequences could not be matched to a reference sequence and were analysed using SAP. The first was identified as belonging to the genus *Betula* and the second was attributed to the genus *Pinus*. The *Betula* sequence was found in 21 individuals (15 *C. murinus* from Corps de Garde, Snail Rock and Trois Mamelles and 6 *C. ovalis* from Snail Rock and Corps de Garde). The *Pinus* sequence was identified from a single specimen on Ille aux

Vacoas. Neither of the plant genera identified were found during floral surveys and given that both genera are exclusively represented by large woody species, it is unlikely they could have been overlooked. As such both sequences were attributed to contamination and excluded from further analysis.

Table 2.5. Plant species consumed by *C. murinus*, with the number of feeding records for each, by sampling locality including the *Betula sp.* and *Pinus sp.* sequences deemed to be contamination.

Locality	Species consumed	Number of sequences retrieved
Corps de Garde	*Litsea glutinosa* (LAURALES)	1
	Hilsenbergia petiolaris (MALVALES)	4
	Pittosporum ferugineum (APIALES)	6
	Schinus terebinthifolius (SAPINDALES)	4
	Betula sp. (FAGALES)	5
Coralline islets	*Scaevola taccada* (ASTERALES)	46
	Ipomoea sp. (SOLANALES)	2
	Pinus sp.(PINALES)	1
Trois Mamelles	*Hilsenbergia petiolaris* (MALVALES)	1
	Erythroxylum sideroxyloides (MALPIGHIALES)	11
	Eugenia sp. (MYRTALES)	2
	Cassine orientalis (CELESTRALES)	1
	Flacourtia indica (MALPIGHIALES)	1
	Margaritaria anomala (MALPIGHIALES)	2
	Coffea myrtifolia (GENTIANALES)	1
	Betula sp. (FAGALES)	5
Snail Rock	*Hilsenbergia petiolaris* (MALVALES)	8
Lion Mountain	*Psiadia viscosa* (ASTERALES)	8

Table 2.6. The number of sequences retrieved for each food plant from *C. ovalis* samples split by collection locality including the *Betula sp.* sequences deemed to be contamination.

Locality	Species consumed	Number of sequences retrieved
Corps de Garde	*Schinus terebinthifolius* (SAPINDALES)	1
	Flacourtia indica (MALPIGHIALES)	1
	Betula sp.(FAGALES)	3
Trois Mamelles	*Hilsenbergia petiolaris* (MALVALES)	2
	Margarataria anomala (MALPIGHIALES)	1
	Ageratum conysoides (ASTERALES)	1
Snail Rock	*Hilsenbergia petiolaris* (MALVALES)	48
	Margarataria anomala (MALPIGHIALES)	2
	Betula sp. (FAGALES)	3

A total of twenty-five plots were surveyed for vascular plants, 13 in Corps de Garde and 12 in Trois Mamelles, and a total of 83 plant species were identified across

both sites (Table 2.7). DNA sequences successfully obtained from identified plant material can be found using GenBank accession numbers JF804883 - JF804940. Across both sites 11 species of plant were characterised from trnL sequences obtained from *C. murinus* samples (Table 2.8). Six of the 11 plant species consumed were found in both Trois Mamelles and Corps de Garde, but only one of these plant species was detected as being consumed in both sites (Table 2.8). The minimally significant GLM suggests there is an effect of foodplant availability on feeding frequency (F= 2.77, df = 19, p=0.013) removal of this variable significantly changes the fit of the GLM (F = 7.12, df = 18, p = 0.016). A scatter plot of feeding frequency vs. proportion of sites (Fig. 2.4) shows that there is one outlier, *Erythroxylum sideroxyloides* in Trois Mamelles, which could be driving this relationship. The GLM was therefore repeated with *Erythroxylum sideroxyloides* omitted (Fig. 2.5). There were no significant interactions or terms in the GLM model using this reduced dataset.

Table 2.7. Potential *Cratopus* foodplants identified during plant surveys by CB and BFVF on Corps de Garde and Trois Mamelles.

Unique to Corps de Garde	Unique to Trois Mamelles	Shared
Agarista salicifolia	Calophyllum tacamahaca	Dodonaea viscose
Albizia lebbeck	Cassine orientalis	Doratoxylon apetalum
Antirhea borbonica	Clerodendrum heterophyllum	Erythroxylum sideroxyloides
Aphloia theiformis	Coffea myrtifolia	Eugenia cf sieberi
Cordia currasavica	Cordia myxa	Eugenia tinifolia
Cossinia pinnata	Diospyros melanida	Eugenia uniflora
Erythrospermum monticolum	Diospyros tessellaria	Ficus reflexa
Eugenia cf orbiculata	Dracaena reflexa	Flacourtia indica
Ixora parviflora	Erythroxylum hypericifolium	Hilsenbergia petiolaris
Leucaena leucocephala	Eugenia fasciculata	Lantana camara
Molinaea alternifolia	Fernelia buxifolia	Ligustrum robustum
Olea lancea	Foetidia mauritiana	Litsea glutinosa
Pithecelobium cf dulce	Ixora borboniae	Ludia mauritiana
Psiadia viscosa	Mangifera indica	Maytenus pyria
Santalum album	Margaritaria anomala	Ochna mauritiana
Sideroxylon cinereum	Mimusops petiolaris	Pittosporum ferrugineum
Stillingia lineata	Molinaea laevis	Protium obtusifolium
Tambourissa cf peltata	Murraya paniculata	Psidium cattleianum
Turraea thouarsiana	Phyllanthus casticum	Psidium guajava
	Premna serratifolia	Schinus terebinthifolius
	Scolopia heterophylla	Syzygium cumini
	Stadmannia oppositifolia	
	Tabernaemontana persicariifolia	

Table 2.8. Feeding and abundance records for food plants of *C. murinus* in Corps de Garde and Trois Mamelles with sequences deemed to be contamination removed..

	Corps de Garde		Trois Mamelles	
	Feeding records	proportion of plots	Feeding records	proportion of plots
Litsea glutinosa	1	0.62	0	0.25
Hilsenbergia petiolaris	4	0.15	1	0.08
Pittosporum ferugineum	6	0.15	0	0.08
Erythroxylum sideroxyloides	0	0.15	11	1.00
Schinus terebinthifolius	4	0.77	0	0.08
Eugenia sp.	0	0.31	2	0.92
Cassine orientalis	0	0.00	1	0.33
Flacourtia indica	0	0.46	1	0.17
Margaritaria anomala	0	0.00	2	0.33
Coffea myrtifolia	0	0.00	1	0.25

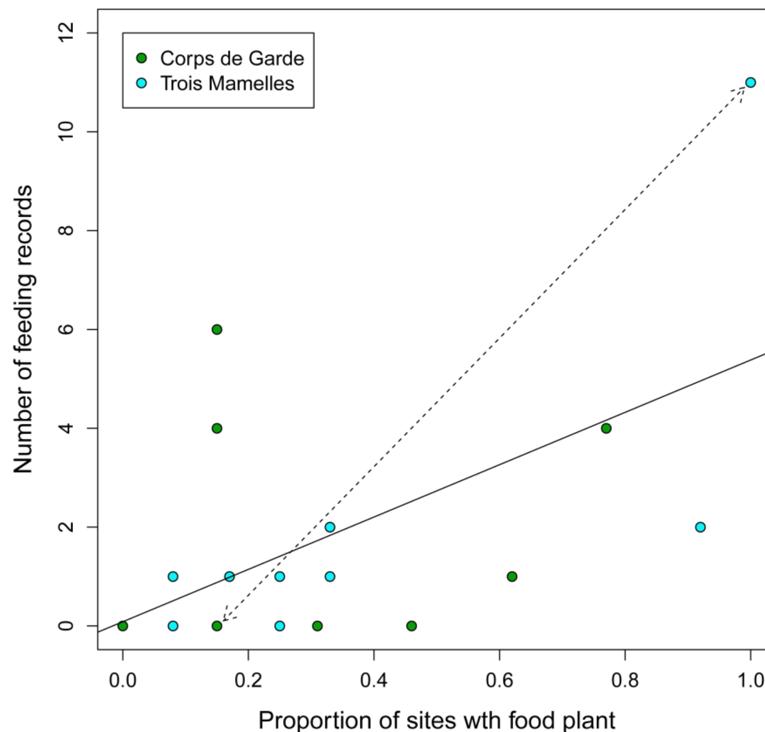


Figure 2.4. Scatter plot of Number of feeding records vs. foodplant availability split by sampling site. The solid black line is a line of best fit though the data. The dashed arrow highlights the points for *Erythroxylum sideroxyloides* in both sites for comparison.

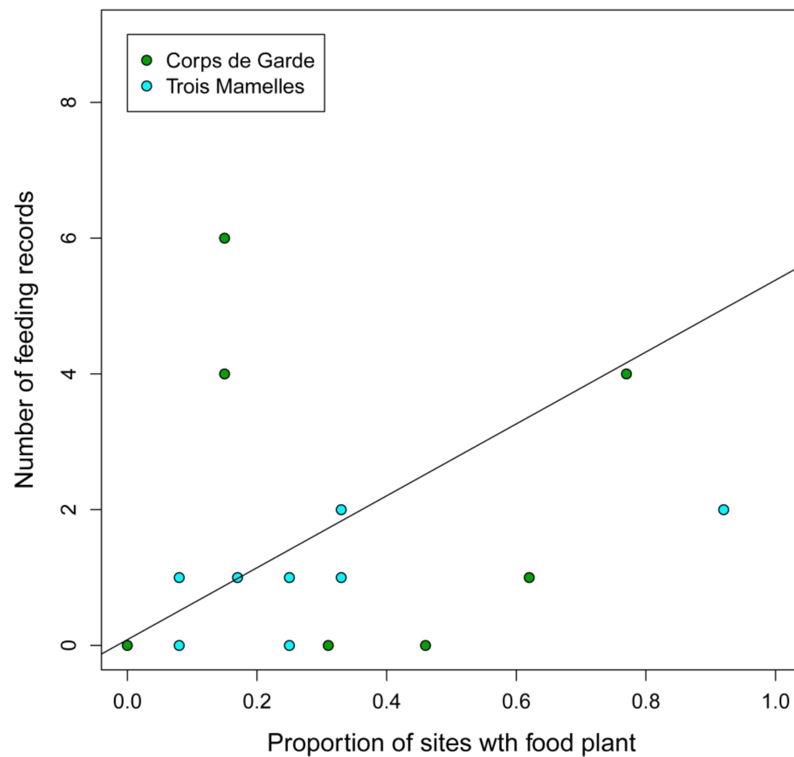


Figure 2.5. Scatter plot of Number of feeding records vs. foodplant availability split by sampling site with values for *Erythroxylum sideroxyloides* removed. The solid black line is a line of best fit though the data.

2.5 Discussion

2.5.1 Geographic structuring of genetic variation within Cratopus murinus and Cratopus ovalis

There is a striking difference between the geographic structuring of genetic variation of *C. murinus* and *C. ovalis*. *Cratopus murinus* populations are highly differentiated with mitotypes largely restricted to one collecting site. This level of structure is reflected in the pairwise Fst values between populations for the ArgK and ITS2 data sets. This suggests that both sexes of *C. murinus* have low dispersal ability or tendency, with no evidence for sex-biased dispersal, in contrast to other beetle genera in studies such as Lagisz (2010). Notable exceptions to this highly structured pattern can be seen in samples from Yemen, Mondrain, Ille de La Passe and Ille aux Vacoas.

Cratopus murinus samples from Yemen and Mondrain possess mtDNA sequences that are phylogenetically closely related, or in several cases identical, to *C. murinus* samples from Corps de Garde Trois Mamelles. These seemingly unusual

relationships, in the context of otherwise strongly geographically structured populations, are most probably explained by recent anthropogenic events. In contrast to all other sampling sites, neither Yemen nor Mondrain can be considered natural dry forests. Mondrain is a botanic garden, and as such plants from various locations across Mauritius are present here. It is therefore conceivable that *C. murinus* in Mondrain are the descendants of larvae or adults accidentally transported with dry forest plants sampled from Corps de Garde. Yemen is also heavily influenced by botanical introductions from across Mauritius, as it is a tourist attraction with a 'botanical trail'. Again, the DNA sequence affinities of *C. murinus* from Yemen with *C. murinus* from Corps de Garde and Trois Mamelles are suggestive of recent anthropogenic introductions.

Two mitotypes are shared across the islets of Ille aux Vacoas and Ille de la Passe, presumably a consequence of their close proximity (220m at their closest point) compared to the more distant Ille Marianne (2,440m from Ille aux Vacoas and 2,720m from Ille de la Passe). Fluctuating sea levels over the last 0.45 mya mean that all three islets were connected to each other and the mainland on at least several occasions (Warren *et al.* 2010). Due to their close proximity and the shallow water bordering the western shores of these islets, Ille aux Vacoas and Ille de la Passe would have had the most substantial connectivity, and as such they could be considered as one recently fragmented population.

C. ovalis populations exhibit greater homogeneity with regard to the structuring of genetic variation, suggesting a more recent origin for the distribution of this species, or higher levels of dispersal relative to *C. murinus*. Both explanations imply a dispersal difference between the two species, which is intriguing for two ecologically and evolutionarily closely related species. It has been suggested that the ability to fly and dispersal itself are not necessarily linked, and McCulloch *et al.* (2009) have shown that some insect populations can retain wings while not routinely dispersing by flight. Associations between reduced flight and energetic savings as well as life history benefits such as increased reproductive output are well documented. (Roff 1990). Zera and Denno (1997) suggests that reduced flight musculature provides most of the energetic savings resulting from flight loss, meaning that individuals retain some flight capability with the benefits that brings (e.g. escape from predators), while also achieving energetic savings. A detailed comparative analysis of flight musculature would

reveal if *C. murinus* is functionally incapable of larger dispersal distances. Brachyapterous corraline islet populations would then represent a further functional reduction of dispersal capability within *C. murinus* in an environment where it has been suggested that selection should favour non-flighted individuals due to reduced recruitment of flighted individuals dispersing out of a limited habitat patch (Harrison 1980).

2.5.2 Trophic ecology of Cratopus murinus and Cratopus ovalis from dry forest sites

Dry forest populations of both *C. murinus* and *C. ovalis* are dietary generalists with diets encompassing one member of the Laurales and much of the higher order diversity within the Core Eudicots as defined in APG III (2009), but with *C. ovalis* having a narrower diet than *C. murinus*. However, most foodplants belong to orders that were also represented within each site by other species, none of which were identified as foodplants. Diet within the two focal species is not restricted to a few closely related foodplants, as is suggested to be the norm for tropical coleopteran herbivores by Novotny *et al.* (2002) and Ødegaard *et al.* (2005), Within the study species diet is probably not constrained by phylogenetic conservatism, with a plausible explanation being that foodplants present a range structural and chemical defences that fall within a spectrum of palatable values for *C. murinus* and *C. ovalis*.

Both the physiological efficiency hypothesis (Dethier 1954) and the enemy free space hypothesis (Jeffries & Lawton 1984) have been suggest to promote host specialisation in small insect herbivores (Tilmon 2008). This would not seem to be the case for *C. murinus* and *C. ovalis*. However, selective advantages have been described for true polyphagy. where any individual can eat multiple foodplants, as opposed to composite generalism proposed by Fox and Morrow (1981), where individuals are monophagous but the whole species or population is polyphagous. True polyphagy carries with it the advantage of being able to gain the full range of nutrients required, whilst reducing intake of any single plant secondary metabolite (Singer *et al.* 2002). Additionally, if enemy free space is variable in space or time (e.g. changing exposure or predator activity patterns during the day), herbivores that can feed on different plants in response to this would be at a selective advantage over those that cannot (Tilmon 2008). Since individuals with multiple

foodplant sequences in their digestive systems were found, it seems likely that dry forest *C. murinus* and *C. ovalis* are at least capable of true food mixing, and their evolutionary history may be influenced by the aforementioned selective pressures.

2.5.3 Variation in Cratopus murinus diet across dry forest sites

By adopting a methodology that combines direct molecular quantification of diet at the level of the individual weevil with floral surveys of two densely sampled sites, it has been possible to attempt to examine how and why diet varies across populations of *C. murinus.* While observed feeding frequencies vary between the populations of Trois Mamelles and Corps de Garde, so does the availability of each foodplant, and a GLM reveals that feeding frequency co-varies with foodplant availability rather than site. This result suggests that dietary choices in dry forest populations of *C. murinus* are regulated by the availability of foodplants, and that diet variation is a neutral response to the foodplant community, and not indicative of any adaptive differences between populations. However, this relationship is probably influenced by one particular foodplant species and removal of this species from the analysis removes any significant effects. This in turn suggests that overall dietary choice may be random within the dietary range of *C. murinus* and that any link between foodplant availability and feeding frequency should be viewed with caution. It is also unclear whether the large number of feeding records for *Erythroxylum sideroxyloides* represents a particular feeding bias towards this plant or whether this just a random effect. Feeding records and abundance data for *E. sideroxyloides* across many sites would be needed to address this.

2.5.4 Trophic ecology of coralline islet populations of Cratopus murinus

In contrast to the dry forest populations of *C. murinus*, populations from the coralline islets appear to have a restricted diet that consists mostly of *Scaevola taccada* with two records of a species of *Ipomoea,* despite other Core Eudicot genera being available. Both *S. taccada* and *Ipomoea sp.* represent novel foodplants, in the context of foodplants consumed within dry forest populations of *C. murinus*. However, neither represents a phylogenetically novel foodplant as they belong to the orders Asterales and Solanales respectively. However, through the fortuitous anthropogenic introduction of *S. taccada* to another recently discovered site,

Ebony Forest, it does appear that *S. taccada* represents an evolutionary change in host plant preference and/or tolerance in coralline islet populations, as no feeding damage has been observed on *S. taccada* in Ebony Forest. While it would seem that there is a stark contrast between the diets, we cannot entirely rule out that the diet of coralline islet *C. murinus* is variable across time. However given that all *C. murinus* were collected in the same season of the same year and *S. taccada* was available to at least one inland population of *C. murinus* without any apparent feeding damage; it would appear likely that the variation in diet is real. Dating estimates suggest that coralline islet populations of *C. murinus* probably diverged from extant mainland dry forest populations somewhere between 0.68-1.2mya. However, the true age of coralline islet populations could be much younger if local population extinction events have featured within the last 1my of the island's history (Emerson 2002). Given that molecular data consistently supports the closest population to the coralline islets to be the geographically most distant population on the main island (Figs. 1-3), and that molecular data also consistently supports a very limited capacity for long distance dispersal within *C. murinus*, a recent history of population extinction seems likely. Thus although a change in the trophic ecology of coralline islet populations has occured, the timing of this is less clear.

2.6 Conclusions

Quantifying trophic interactions between insect herbivore and plant species is fundamental to the understanding of ecosystem functioning. However, it is complicated by the difficulty of obtaining plant-feeding records, particularly when an insect herbivore's life cycle does not involve direct association with plant tissue. Population-level molecular analysis of insect digestive system contents using DNA sequences can overcome this complication. The improved success rate in obtaining plant feeding records from DNA sequence data is at least in part due to the reduced fragment length amplified compared with previous studies. While this may limit the power to distinguish between closely related plant species, this is offset by increased sample sizes. A strength of the approach used here was to create a local plant sequence reference library, reducing uncertainty regarding species identify and potential contamination that can emerge with indirect molecular identification

techniques, due to limited species representation on public databases such as GenBank. Further studies combining direct DNA sequence-based measures of plant feeding, DNA sequence libraries for plant species identification, population genetic data and local ecological data, will help further develop our understanding of trophic interactions between insect herbivore and plant species.

2.7 References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic Local Alignment Search Tool. *Journal of Molecular Biology* **215**, 403-410.

Barat M, Tarayre M, Atlan A (2008) Genetic divergence and ecological specialisation of seed weevils (*Exapion* spp.) on gorses (*Ulex* spp.). *Ecological Entomology* **33**, 328-336.

Barone JA (1998) Host-specificity of folivorous insects in a moist tropical forest. *Journal of Animal Ecology* **67**, 400-409.

Cho SW, Mitchell A, Regier JC, *et al.* (1995) A highly conserved nuclear gene for low-level phylogenetics - Elongation Factor-1-Alpha recovers morphologically-based tree for Heliothine moths. *Molecular Biology and Evolution* **12**, 650-656.

Cicconardi F, Nardi F, Emerson BC, Frati F, Fanciulli PP (2010) Deep phylogeographic divisions and long-term persistence of forest invertebrates (Hexapoda: Collembola) in the North-Western Mediterranean basin. *Molecular Ecology* **19**, 386-400.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology* **9**, 1657-1659.

Danforth BN, Lin CP, Fang J (2005) How do insect nuclear ribosomal genes compare to protein-coding genes in phylogenetic utility and nucleotide substitution patterns? *Systematic Entomology* **30**, 549-562.

Dethier VG (1954) Evolution of feeding preferences in phytophagous insects. *Evolution* **8**, 33-54.

Don RH, Cox PT, Wainwright BJ, Baker K, Mattick JS (1991) 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Research* **19**, 4008-4008.

Drummond A, Ashton B, Buxton S*, et al.* (2011) Geneious v5.4, Available from http://www.geneious.com/.

Drummond AJ, Ho SYW, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biol* **4**, e88.

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

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

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**, 564-567.

Farrell BD (1998) "Inordinate fondness" explained: Why are there so many beetles? *Science* **281**, 555-559.

Flot J-F, Tillier A, Samadi S, Tillier S (2006) Phase determination from direct sequencing of length-variable DNA regions. *Molecular Ecology Notes* **6**, 627-630.

Fox LR, Morrow PA (1981) Specialization: species property or local phenomenon? *Science* **211**, 887-893.

Gillespie RG, Claridge EM, Goodacre SL (2008) Biogeography of the fauna of French Polynesia: diversification within and between a series of hot spot archipelagos. *Philosophical Transactions of the Royal Society B-Biological Sciences* **363**, 3335-3346.

Grobler GC, van Rensburg LJ, Bastos ADS, Chimimba CT, Chown SL (2006) Molecular and morphometric assessment of the taxonomic status of *Ectemnorhinus* weevil species (Coleoptera : Curculionidae, Entiminae) from the sub-Antarctic Prince Edward Islands. *Journal of Zoological Systematics and Evolutionary Research* **44**, 200-211.

Harrison RG (1980) Dispersal Polymorphisms in Insects. *Annual Review of Ecology and Systematics* **11**, 95-118.

Heled J, Drummond AJ (2010) Bayesian Inference of Species Trees from Multilocus Data. *Molecular Biology and Evolution* **27**, 570-580.

Hernandez-Vera G, Mitrovic M, Jovic J*, et al.* (2010) Host-associated genetic differentiation in a seed parasitic weevil *Rhinusa antirrhini* (Coleptera: Curculionidae) revealed by mitochondrial and nuclear sequence data. *Molecular Ecology* **19**, 2286-2300.

Hey J (2010) Isolation with migration models for more than two populations. *Molecular Biology and Evolution* **27**, 905-920.

Hey J, Nielsen R (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 2785-2790.

Ho SYW, Larson G, Edwards CJ*, et al.* (2008) Correlating Bayesian date estimates with climatic events and domestication using a bovine case study. *Biology Letters* **4**, 370-374.

Holland BS, Hadfield MG (2002) Islands within an island: phylogeography and conservation genetics of the endangered Hawaiian tree snail *Achatinella mustelina*. *Molecular Ecology* **11**, 365-375.

Huelsenbeck JP, Ronquist F (2001) MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754-755.

Jeffries MJ, Lawton JH (1984) Enemy free space and the structure of ecological communities. *Biological Journal of the Linnean Society* **23**, 269-286.

Ji YJ, Zhang DX, He LJ (2003) Evolutionary conservation and versatility of a new set of primers for amplifying the ribosomal internal transcribed spacer regions in insects and other invertebrates. *Molecular Ecology Notes* **3**, 581-585.

Jordal BH, Emerson BC, Hewitt GM (2006) Apparent 'sympatric' speciation in ecologically similar herbivorous beetles facilitated by multiple colonizations of an island. *Molecular Ecology* **15**, 2935-2947.

Jordal BH, Kirkendall LR, Harkestad K (2004) Phylogeny of a Macaronesian radiation: host-plant use and possible cryptic speciation in *Liparthrum* bark beetles. *Molecular Phylogenetics and Evolution* **31**, 554-571.

Jurado-Rivera JA, Vogler AP, Reid CAM, Petitpierre E, Gomez-Zurita J (2009) DNA barcoding insect-host plant associations. *Proceedings of the Royal Society B: Biological Sciences* **276**, 639-648.

Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* **30**, 3059-3066.

Lagisz M, Wolff K, Sanderson RA, Laskowski R (2010) Genetic population structure of the ground beetle, *Pterostichus oblongopunctatus*, inhabiting a fragmented and polluted landscape: Evidence for sex-biased dispersal. *Journal of Insect Science* **10:105**.

Machado A (2006) The type material of the species of *Laparocerus* Schonherr, 1834 (Coleoptera, Curculionidae, Entiminae). *Journal of Natural History* **40**, 2001-2055.

Machado A (2007a) New species of *Laparocerus* Schonherr, 1834 (Coleoptera, Curculionidae) from the island of Tenerife (Canary Islands). *Zootaxa* **1457**, 35-48.

Machado A (2007b) New species of *Laparocerus* Schonherr, 1834 from La Gomera, Canary Islands (Coleoptera, Curculionidae, Entiminae). *Zootaxa* **1643**, 1-38.

Machado A (2008) Three new endogean species of *Laparocerus* Schonherr, 1834 from the Canary Islands (Coleoptera, Curculionidae). *Journal of Natural History* **42**, 1277-1288.

Machado A, Lopez M, Almeida T, Hernandez M (2008) Mitochondrial DNA phylogenetic analysis of the genus *Laparocerus* (Coleoptera, Curculionidae, Entiminae). 1 The madeiran clade. *Zoologica Scripta* **37**, 415-427.

McCulloch GA, Wallis GP, Waters JM (2009) Do insects lose flight before they lose their wings? Population genetic structure in subalpine stoneflies. *Molecular Ecology* **18**, 4073-4087.

McKenna DD, Sequeira AS, Marvaldi AE, Farrell BD (2009) Temporal lags and overlap in the diversification of weevils and flowering plants. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 7083-7088.

Minin VN, Dorman KS, Fang F, Suchard MA (2005) Dual multiple change-point model leads to more accurate recombination detection. *Bioinformatics* **21**, 3034-3042.

Morris MG (1997) *Broad-Nosed Weevils*, 17a edn. Royal Entomological Society.

Munch K, Boomsma W, Huelsenbeck JP, Willerslev E, Nielsen R (2008) Statistical assignment of DNA sequences using Bayesian phylogenetics. *Systematic Biology* **57**, 750-757.

Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature* **403**, 853-858.

Novotny V, Basset Y, Miller SE, *et al.* (2002) Low host specificity of herbivorous insects in a tropical forest. *Nature* **416**, 841-844.

Novotny V, Drozd P, Miller SE, *et al.* (2006) Why are there so many species of herbivorous insects in tropical rainforests? *Science* **313**, 1115-1118.

Oberprieler RG, Marvaldi AE, Anderson RS (2007) Weevils, weevils, weevils everywhere. *Zootaxa*, 491-520.

Ødegaard F, Diserud OH, Østbye K (2005) The importance of plant relatedness for host utilization among phytophagous insects. *Ecology Letters* **8**, 612-617.

Otte D (1976) On feeding patterns in desert grasshoppers and the evolution of specialized diets. *Proceedings of the Academy of Natural Sciences of Philadelphia* **128**, 89-126.

Peters JL, Zhuravlev Y, Fefelov I, Logie A, Omland KE (2007) Nuclear loci and coalescent methods support ancient hybridization as cause of mitochondrial paraphyly between gadwall and falcated duck (*Anas* spp.). *Evolution* **61**, 1992-2006.

Pinzón-Navarro S, Jurado-Rivera JA, Gómez-Zurita J, Lyal CHC, Vogler AP (2010) DNA profiling of host–herbivore interactions in tropical forests. *Ecological Entomology* **35**, 18-32.

Podlussany A, Jermy T, Szentesi A (2001) On the leguminous host plants of seed predator weevils (Coleoptera : Apionidae, Curculionidae) in Hungary. *Acta Zoologica Academiae Scientiarum Hungaricae* **47**, 285-299.

Posada D (2008) jModelTest: Phylogenetic model averaging. *Molecular Biology and Evolution* **25**, 1253-1256.

R Development Core Team (2010) R: A Language and Environment for Statistical Computing, Vienna, Austria.

Rambaut A (2011) FigTree v1.3.1. In: *http://tree.bio.ed.ac.uk/software/figtree/*

Rambaut A, Drummond A (2007) Tracer v1.5.0. In: *http://beast.bio.ed.ac.uk/Tracer*

Regier JC, Shi D (2005) Increased yield of PCR product from degenerate primers with nondegenerate, nonhomologous 5 ' tails. *Biotechniques* **38**, 34.

Roff DA (1990) The evolution of flightlessness in insects. *Ecological Monographs* **60**, 389-421.

Sequeira AS, Lanteri AA, Albelo R, Bhattacharya S, Sijapati M (2008) Colonization history, ecological shifts and diversification in the evolution of endemic Galapagos weevils. *Molecular Ecology* **17**, 1089-1107.

Singer MS, Bernays EA, Carrière Y (2002) The interplay between nutrient balancing and toxin dilution in foraging by a generalist insect herbivore. *Animal Behaviour* **64**, 629-643.

Stephens M, Scheet P (2005) Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. *American Journal of Human Genetics* **76**, 449-462.

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics* **68**, 978-989.

Taberlet P, Coissac E, Pompanon F*, et al.* (2007) Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. *Nucleic Acids Research* **35**, Article No.: e14.

Tamura K, Peterson D, Peterson N*, et al.* (2011) MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* **28**, 2731–2739.

The Angiosperm Phylogeny Group (2009) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society* **161**, 105-121.

Tilmon KJ (2008) *Specialization, Speciation, and Radiation. The Evolutionary Biology of Herbivorous Insects* University of California Press, Los Angeles.

Warren BH, Strasberg D, Bruggemann JH, Robert PP-J, Christophe T (2010) Why does the biota of the Madagascar region have such a strong Asiatic flavour? *Cladistics* **26**, 526-538.

Williams JR (2000) A revision of the Mascarene weevil genus *Syzygops* Schonherr (Coleoptera : Curculionidae : Entiminae). *Invertebrate Taxonomy* **14**, 411-432.

Williams JR, Cox ML (2003) A Contribution to the study of Mascarene Weevils of the genus *Cratopus* Schonherr (Coleoptera: Curculionidae: Entiminae: Cratopini): The species of Mauritius and Rodrigues. *The Mauritius Institute Bulletin* **12**.

Zera AJ, Denno RF (1997) Physiology and ecology of dispersal polymorphism in insects. *Annual Review of Entomology* **42**, 207-230.

Chapter 3: Baseline molecular data for poorly understood diversity: a case study of Mauritian Cratopine weevils.

Cratopus nigrogranatus: Machabée, Mauritius 2009.

3.1 Abstract

A single locus screen of the Cratopine weevil species of Mauritius was undertaken to address taxonomic uncertainty surrounding the Mauritian members of the tribe. Phylogenetic analyses revealed 18 separate lineages, 15 of which contained a single morphospecies and three of which contained at least two morphospecies. Multimodel averaged GMYC analysis revealed strong variation in the degree to which lineages are subdivided using coalescent models with the most widespread and common species being subdivided the most. Population structure was assessed for selected lineages using AMOVA. Population structure was common and the possible reasons for this are discussed. Measures of phylogenetic diversity were calculated for selected sites and patterns of diversity are discussed.

3.2 Introduction

Oceanic island ecosystems often present impressive species radiations that allow for DNA sequence based molecular examination of evolutionary patterns and processes at both interspecific and intraspecific scales within relatively simple natural systems (e.g. Gillespie *et al.* 2008; Sequeira *et al.* 2008; Thacker & Hadfield 2000). Much attention has been focussed at the interspecific scale and has typically used phylogenetic techniques to address questions related to the timing of divergence events (Magnacca & Danforth 2006; Sequeira *et al.* 2000), patterns of diversification (Holland & Hadfield 2002; Jordal *et al.* 2004; Jordan *et al.* 2005) and the resolution of taxonomic uncertainty (e.g. Clement *et al.* 2004; Jordal *et al.* 2006). At the intraspecific scale, coalescent and population genetic approaches are used to characterise genetic structure across populations within the same species (e.g. Emerson *et al.* 2000; Grobler *et al.* 2011; Illera *et al.* 2007; Polihronakis *et al.* 2010). Both approaches can yield information that advances evolutionary understanding (e.g. Fukami *et al.* 2007; Gillespie 2004; Spurgin *et al.* 2011) and conservation management (e.g. Cegelski *et al.* 2006; de Thoisy *et al.* 2010; Newton *et al.* 1999), however knowing which approach to take can prove difficult when the taxonomy for a particular group is incomplete or poorly resolved, and the boundary between interspecific divergence and intraspecific diversity cannot be assumed *a priori*.

A number of attempts have been made to infer putative species boundaries using either DNA sequence or tree-based analyses such as 'barcode gaps' (Hebert *et al.* 2003), cladistic haplotype aggregation (Brower 1999), population aggregation analysis (Davis & Nixon 1992) and the Weins-Penkrot method (Sites & Marshall 2003; Wiens & Penkrot 2002). However, a limitation common to all of these approaches is that populations or species need to be defined *a priori* and then tested, rather than being derived directly from the data. A more recently described approach developed by Pons *et al.* (2006) and extended by Monaghan *et al.* (2009) uses an algorithmic approach based on changing branching patterns as population-level coalescent processes give way to species-level divergences to delineate putative species. This general mixed Yule coalescent (GMYC) approach has been advocated as a useful technique for accelerating species discovery and inventory in environments under threat from anthropogenic pressure (Monaghan *et al.* 2009), and in a geographic context it has proved valuable for characterising the temporal and spatial dimensions of diversification (e.g. Pons *et al.* 2006; Pons *et al.* 2011; Puillandre *et al.* 2011).

The substantial potential of the GMYC approach for the delimitation of species boundaries makes it attractive for the analysis of under studied taxa of biodiversity interest. In this study the utility of the GMYC is exploited in combination with a mtDNA sequence screen of biodiversity within such a taxon. The islands of the southwest Indian Ocean are recognised as a biodiversity hotspot (Myers *et al.* 2000), but most are under severe anthropogenic pressure due to deforestation, farming, invasive species and urban sprawl. An emblematic feature of the arthropod diversity of these islands is the weevil tribe Cratopini, specifically the genera *Cratopus*, *Cratopopsis* and *Scaevinus*. This group is comprised of approximately 100 species of folivorous weevils and has the majority of its diversity concentrated on the islands of Mauritius and La Réunion (Hustache 1920; Williams & Cox 2003). *Cratopopsis* and *Scaevinus* contain far fewer species than *Cratopus* and are all flightless, whereas *Cratopus* is unusual among island inhabiting weevils in that most species are flighted. The three genera exhibit substantial morphological and ecological variation both within and among species (Hustache 1920; Williams & Cox 2003), and species are present in all native terrestrial habitats available on the Mascarene islands between the foreshore and high alpine scrub up to at least 2500m. Within *Cratopus*, a large proportion of

described species are seemingly rare or have highly restricted ranges (Williams & Cox 2003), which makes them vulnerable to global extinction. Additionally, even relatively common species may present complex population structures, with multiple isolated populations vulnerable to local extinction (see chapter 2). Extinction risk coupled with taxonomic uncertainty surrounding a number of species means that a molecular assessment of diversity within this group is needed before conservation priorities can be suggested or evolutionary units can be defined for further phylogenetic analyses.

Our broad aim is to provide a measure of diversity both above and below the level of traditionally defined morphospecies within *Cratopus* and *Cratopopsis*, and to gain estimates of how evolutionary diversity is spatially structured at both the interspecific and intraspecific level. To achieve this aim, all three genera were sampled extensively across the island of Mauritius and individuals were sequenced for the mitochondrial Cytochrome Oxidase C subunit II (COII) gene to address a series of fundamental baseline questions concerning the evolution and conservation importance of morphologically defined species within this group. Mauritius has a minimum subaerial age of 8.9 million years (Ma) (Moore *et al.* 2011) suggesting that much of the endemic Cratopine diversity could be quite old. However, recent comparable studies of the origin of species diversity within oceanic islands frequently reveal diversity to be much younger that what might be expected given island age (e.g. Contreras-Diaz *et al.* 2007; Monaghan *et al.* 2006; Price & Clague 2002; Rees *et al.* 2001). We first ask how old is the Cratopine weevil diversity within Mauritius? By using phylogenetic methods and published evolutionary rates it is possible to generate an estimate of the age of the Most Recent Common Ancestor (mrca) for Mauritian Cratopine weevils. Morphologically defined species are then examined in the context of molecular phylogenetic data and putative model-based inferences of species boundaries. This is particularly useful in the context of rare species and species described on the basis of subtle morphological differences where uncertainty exists as to the validity of species assignment. Specimens collected from La Réunion that exhibit morphological similarities to the Mauritian species *C. punctum* and *C. murinus* were also included in this study to determine whether or not these individuals are likely to represent part of the same species; the null hypothesis is that they are part of the same species. The third specific objective is to assess evidence for geographic

structuring of genetic variation within more commonly occurring species across sampling sites. As they are flighted, one might expect limited population structure within species of *Cratopus*. However, strong population structure is a feature of at least one species of *Cratopus,* despite its potential for dispersal (see *C. murinus*, chapter 2). Understanding the extent to which other Mauritian *Cratopus* species are geographically structured can provide insight into both local extinction risk, and potential incipient speciation. The final objective is to generate estimates of phylogenetic diversity and related measures within collection sites to test if there are areas containing higher evolutionary diversity than might be expected by chance. Such areas would be of interest both for their conservation value for Cratopine evolutionary diversity and also as a focus of attention for further biodiversity analyses within Mauritius.

3.3 Methods

3.3.1 Beetle sampling

Samples were collected by foliage beating on Mauritius during the wet season between June 2007 and December 2011. Samples were placed directly in 99% ethanol solution, with the exception of a few individuals that were kept alive until the same evening for photographing, before being placed in 99% ethanol. Sampling sites were recorded on a handheld GPS unit (Garmin GPS 60, Garmin Ltd). A total of 1445 beetles were collected across 28 sites and identified using the morphological key in Williams & Cox (2003). See Table 3.1 for GPS coordinates of sites and Fig 3.1 for a map of the sites. Five individuals (unless fewer than five were obtained) for each morphospecies within a site were selected for DNA extraction and sequencing, giving a data set of 451 individuals. An additional thirty-five specimens from the neighbouring island of La Réunion were also included, as these samples conform to the Mauritian species *C. punctum* and *C. murinus* (*C. murinus* samples are from Piton de Grande Anse, Cap Jaune and Cap Méchant and *C. Punctum* samples are from Forêt domaniale de la Côte sous le Vent).

Table 3.1: The collection localities included in this study and their associated latitude and longitude.

Island	Site	Latitude	Longitude	Island	Site	Latitude	Longitude
Mauritius	Brise Fer 1	-20.378	57.442	Mauritius	Montagne Cocotte	-20.440	57.467
Mauritius	Brise Fer 2	-20.381	57.448	Mauritius	Montagne de Bel Ombre	-20.467	57.417
Mauritius	Brise Fer 3	-20.378	57.442	Mauritius	Petrin	-20.408	57.473
Mauritius	Camp Thorel	-20.209	57.639	Mauritius	Pigeon wood	-20.441	57.487
Mauritius	Chamarel	-20.430	57.374	Mauritius	Plaine Champagne	-20.419	57.445
Mauritius	Corps de Garde	-20.249	57.451	Mauritius	Roches 1	-20.121	57.698
Mauritius	Gorges	-20.404	57.432	Mauritius	Roches 2	-20.110	57.727
Mauritius	Île Marianne	-20.380	57.787	Mauritius	Roches 3	-20.112	57.749
Mauritius	Ile aux Vacoas	-20.398	57.770	Mauritius	Round Island	-19.852	57.788
Mauritius	Ile de la Passe	-20.399	57.767	Mauritius	Seama Beach	-20.355	57.361
Mauritius	Ilot Sancho	-20.504	57.449	Mauritius	Snail Rock	-20.192	57.507
Mauritius	Kanaka	-20.406	57.519	Mauritius	Souillac	-20.525	57.528
Mauritius	Le Pouce	-20.198	57.529	Mauritius	Trois Mamelles	-20.313	57.449
Mauritius	L'Etoile	-20.323	57.670	Mauritius	Vallee de l'est	-20.332	57.726
Mauritius	Lion Mountain	-20.362	57.725	Mauritius	Yemen	-20.299	57.411
Mauritius	Machabee 1	-20.396	57.456	La Réunion	Cap Jaune	-21.378	55.677
Mauritius	Machabee 2	-20.390	57.460	La Réunion	Cap Méchant	-21.376	55.709
Mauritius	Mondrian	-20.325	57.455	La Réunion	Forêt Domaniale de la Côte sous le Vent	-21.282	55.363
Mauritius	Montagne Camizard	-20.336	57.715	La Réunion	Piton de Grande Anse	-21.373	55.551

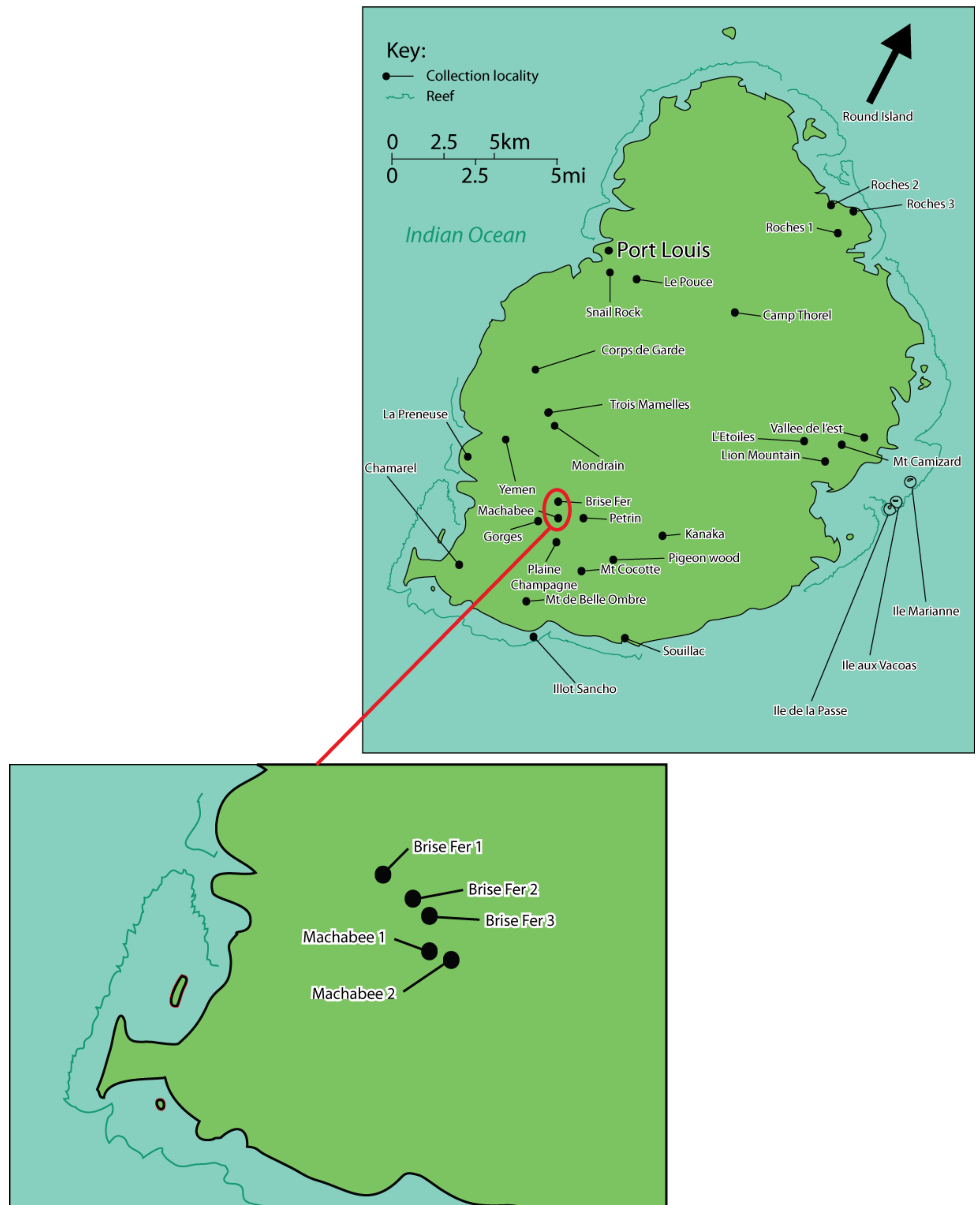


Fig 3.1: Collecting sites sampled on Mauritius. Inset is a magnification of the Brise Fer and Machabee regions.

3.3.2 DNA extraction, PCR amplification and sequencing

DNA was extracted from the head and pronotum using the DNeasy 96 well Blood and Tissue Extraction kit (QIAGEN, West Sussex, UK) with the digestion buffer volumes amended for large specimens as recommended by the manufacturer.

Primers COIICraF (5' TAATATGGCAGAWTAGTGCAATGGA 3') and COIICraR (5' TGCTTTCAGTCATCTAATGATCTRTTTACAGA 3') were designed by aligning Curcujiform mitochondrial sequences from GenBank and selecting conserved regions in tRNA-Leucine and tRNA-Lysine, which flank the COII gene. COII amplification conditions were: 0.5 mM of each primer, 5 mM $MgCl_2$, and a thermal profile of: 95°C for 60s, 58°C for 60s and 72°C for 90s, 40 cycles.

Sequencing reactions were performed with the Big Dye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems, California, USA). For COII, COIICraF was used for forward sequencing and COIICraR was used for reverse sequencing when forward sequences of fewer than 600bp were produced. When reverse sequencing was employed, the consensus of each pair of forward and reverse sequences was generated in Geneious Pro version 5.4 (Drummond *et al.* 2011). The thermal profile used for all sequencing reactions was: 96°C for 10s, 58°C for 5s and 60°C for 240s, 25 cycles. Sequences were read on a 3730XL sequencer (Applied Biosystems). All sequences were checked and ambiguous bases called in Geneious Pro version 5.4. Sequences were aligned using MAFFT v6.814b (Katoh *et al.* 2002) with the following parameter values: scoring matrix 200PAM/k=2, Gap open penalty = 1.53, Offset value = 0.123, and then checked by eye. The aligned sequences were tested for saturation using the entropy-based index of substitution saturation (Xia *et al.* 2003) as implemented in DAMBE v5.2.78 (Xia & Xie 2001). Xia's index is the ratio of the mean proportion of each nucleotide across all sites of your alignment to the proportion expected when the sequence is fully saturated. The significance of this ratio is tested by comparison to a set of critical values derived by simulation (Xia *et al.* 2003). Xia's index was calculated for two data sets, one comprised of the first and second codon positions and the second comprised of third codon positions. Substitution saturation was visualised by plotting uncorrected pairwise p-distances against corrected pairwise distances. A linear regression of the plotted values provides an alternative method to qualitatively assess saturation as an x coefficient much less than one in the fitted model would indicate a strong disagreement between modelled and observed genetic distances and that the sequences are likely heavily saturated.

3.3.3 Phylogenetic analyses

Trees were constructed from the COII alignment using MrBayes 3.2 (Huelsenbeck & Ronquist 2001) with *Naupactus xanthographus* as an outgroup (GenBank accession number: GU176345). For MrBayes, eight analyses were performed, each for 12 million generations using eight MCMC chains, discarding 25% of the samples as burnin with a model of sequence evolution determined using jModelTest (Posada 2008) under the BIC and AIC criteria. All parameters permitted under this model were estimated. The output was assessed for stationarity and convergence in Tracer v1.5.0 (Rambaut & Drummond 2007), with only ESS scores greater than 200 being accepted, and the consensus tree was visualised in FigTree v1.3.1 (Rambaut 2011).

3.3.4 Divergence time estimation

To estimate the root age of our sample, and generate an ultrametric tree for the GMYC analyses a BEAST analysis was performed on an alignment of unique sequences only. Ten replicate analyses were performed using the best fit model characterised by jModelTest and an uncorrelated relaxed clock (Drummond *et al.* 2006) with a constant population size as suggested in Monaghan *et al.* (2009). A coleopteran COII mutation rate of 0.0154 substitutions/site/Ma was taken from Cicconardi et al. (2010). As this estimate may not accurately reflect the mutation rate in Cratopine weevils, the mutation rate was allowed to have a normal distribution with a mean of 0.0154 substitutions/ site/Ma and a standard deviation of 0.0075 substitutions/site/Ma. Analyses were run for 120 million generations and convergence and stationarity were assessed with Tracer v1.5.0. As with the MrBayes analysis, only ESS scores greater than 200 were accepted. After discarding burnin, the tree files were combined in LogCombiner and the tree for the GMYC analysis was generated in TreeAnnotator and visualised in FigTree v1.3.1.

3.3.5 GMYC analyses

The maximum likelihood number of GMYC clusters and GMYC entities (GMYC clusters plus single branches that cannot be assigned to a GMYC cluster) was determined in R 2.13.1 (R Development Core Team 2010) using the single

threshold GMYC model (Pons *et al.* 2006), the multiple threshold GMYC model (Monaghan *et al.* 2009) and the multimodel inference approach detailed by Powell (2012). The difference between single and multiple threshold approaches was tested using a likelihood ratio test. GMYC clusters and probability values were then annotated on to the ultrametric tree.

3.3.6 Population structure

Population structure was tested using AMOVA (Excoffier *et al.* 1992) as implemented in Arlequin v 3.5 (Excoffier & Lischer 2010). Only morphospecies where there were at least three sites each with a minimum of three samples were included. Instances where strongly supported GMYC groups contained only a single collecting site were also excluded from the analyses as this in itself is evidence for population structure and this signal would confound any weaker signal contained in geographically mixed clades. Finally, if a lineage was subdivided by the GMYC analysis, then only samples from the largest GMYC cluster were used as this removes any bias caused by potential cryptic species.

3.3.7 Phylogenetic diversity

Phylogenetic diversity (PD) (Faith 1992) was calculated in R 2.13.1 using the package picante (Kembel *et al.* 2010). Only sites of comparable sampling effort were included in the community matrix, and a single individual from each morphospecies was chosen to represent that morphospecies across all sites. Thus, the community matrix represents presence/absence for each species within each site. PD was then calculated based on the MrBayes tree pruned to include only the sequences contained in the community matrix. Sites with only one morphospecies were excluded as picante uses an unrooted tree for its calculations. Standard effect size mean pairwise distance (SES_{MPD}) and standard effect size mean nearest taxon distance (SES_{MNTD}) (Webb *et al.* 2002) were both calculated using 10,000 tip label randomisations and a cophenetic distance matrix generated in picante. These values were then used to determine whether or not each site has more diversity than a null model generated by species randomisation across the entire tree. As the community matrix contains only presence/absence data, the chances of drawing any given species during randomisation are related to the number of sites in which

it occurs, as opposed to how many specimens were collected, thus preventing population size variation across sites confounding our analyses.

Direct comparison of sites was achieved by rarefaction of PD values using the R package phylocurve.perm (available from http://homepage.mac.com/davidnipperess), which implements a methodology similar to that of Gotelli & Colwell (2001), but where PD is the calculated measure instead of species richness. Sites were rarefied by species so that for each site the expected PD for n species (the number of species you are sub-sampling from each site) out of N species (the total number of species collected in each site) could be calculated. Comparisons of sites were made for varying values of n and 10,000 replicates of each rarefaction were performed. For each comparison, mean values and 95% confidence intervals were plotted for visual comparison.

Linear regressions of species richness vs mean pairwise distance (MPD) and mean nearest taxon distance (MNTD) were performed and their residuals examined to identify sites that had higher or lower MPD and MNTD than predicted by a linear model. Species compositions for these sites were then compared to see which species may be responsible for unusually high or low diversity values.

3.4 Results

3.4.1 MtDNA sequencing

Of the 486 beetles selected for sequencing, 438 were successfully amplified and sequenced for COII with 234 samples that required reverse sequencing (a 90.1% success rate) yielding sequences from 582bp to 673bp with a mean length of 659.2bp and a standard deviation of 23.4bp. Two hundred and eighty-two unique sequences were recovered from across 25 morphospecies with an average pairwise p-distance of 14.7%, a maximum of 18.4% and a standard deviation of 2.0%. When corrected using the general time reversible model (GTR), the average pairwise distance was 17.1%, the maximum was 22.2% and the standard deviation was 2.6%.

Xia's index for the first and second codon positions produced a value of 0.07, significantly less than the critical value for symmetric topologies (0.69, $P<0.01$), suggesting that sites have not reached mutational saturation. For the third codon position, Xia's index was 0.54, which is also significantly less than the critical

value for symmetric topologies (0.63, *P*=0.01). However, the index for this codon position is closer to the critical value, suggesting that third codon position data is approaching maximum divergences that can be reliably inferred. Plotting uncorrected p-distances against genetic distances corrected using the GTR model for all codon position combined (Fig. 3.2) reveals that the modelled distances do deviate from the observed distances, suggesting that there may be some tendency towards saturation. However, a linear regression of this relationship returns an x coefficient of 0.77 (*P*<0.01), suggesting that this effect is weak.

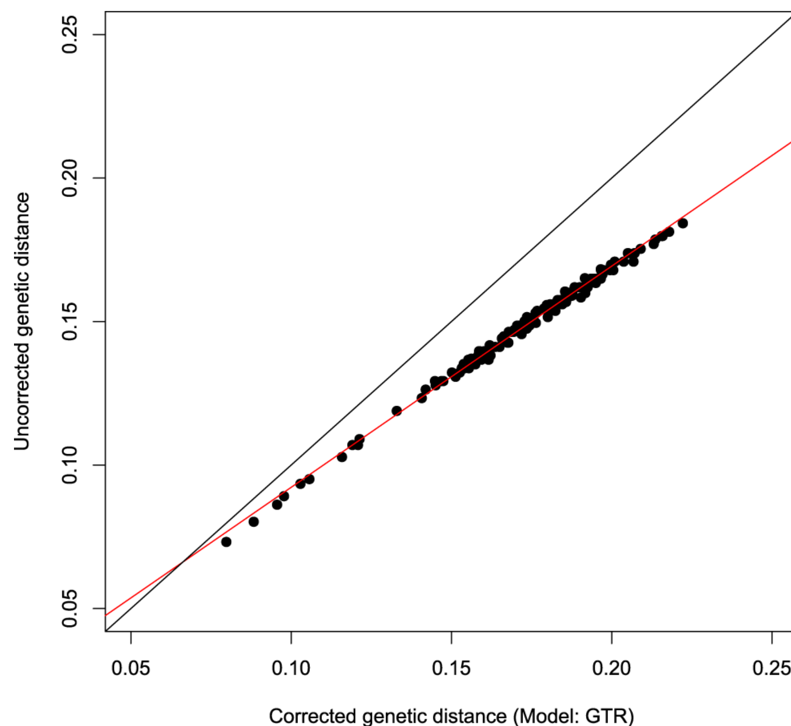


Fig 3.2: A linear regression of uncorrected genetic distance against genetic distance corrected using the GTR model. The solid red line represents the fitted model and the solid black line represents no difference between uncorrected and corrected genetic distances.

3.4.2 Phylogenetic analyses

MrBayes analyses were performed using the GTR+G substitution model returned as the best fit model by jModeltest, yielding a tree with 18 distinct lineages with a lineage being defined as either a monophyletic morphospecies or the smallest monophyletic group that contains paraphyletic morphospecies (Fig. 3.3). Eleven of these are monophyletic groups each containing a single morphospecies (lineages A, C, D, F, G, H, I, J, L, Q and R). A further four lineages are rare morphospecies for which only a single individual was amplified (lineages E, K, O and P). All of these

are found on long branches and high Bayesian posterior probabilities (greater than 0.9) suggest that these individuals do not fall within their sister lineages. The final three lineages (B, M and N) are each composed of more than one morphospecies. Lineage N is composed of individuals belonging to the *C. aeneoniger* species complex as defined by Williams & Cox (2003) (*C. aeneoniger*, *C. confusus*, *C. deceptus* and *C. vulgaris*) and two additional morphospecies (*C. emarginatus and C. stigmaeus)*. Lineage B contains *C. murinus*, *C. tigratus* and 5 individuals of *C. caliginosus* (all other *C. caliginosus* individuals are found in lineage A). The final lineage (M) comprises the two sampled *Cratopopsis* (hereafter abbreviated to *Cr.*) morphospecies, *Cr. mauritianus* and *Cr. impressus*. As expected, individuals from La Réunion are nested within the Mauritian samples of the same morphospecies suggesting that these species are indeed shared across islands. Each site for Réunionaise *C. murinus* and the single site for Réunionaise *C. punctum* were all recovered as monophyletic with posterior probabilities ≥ 0.9. The sole Mauritian representative of the genus *Scaevinus* is recovered outside the other genera whereas *Cratopopsis* is placed within *Cratopus*.

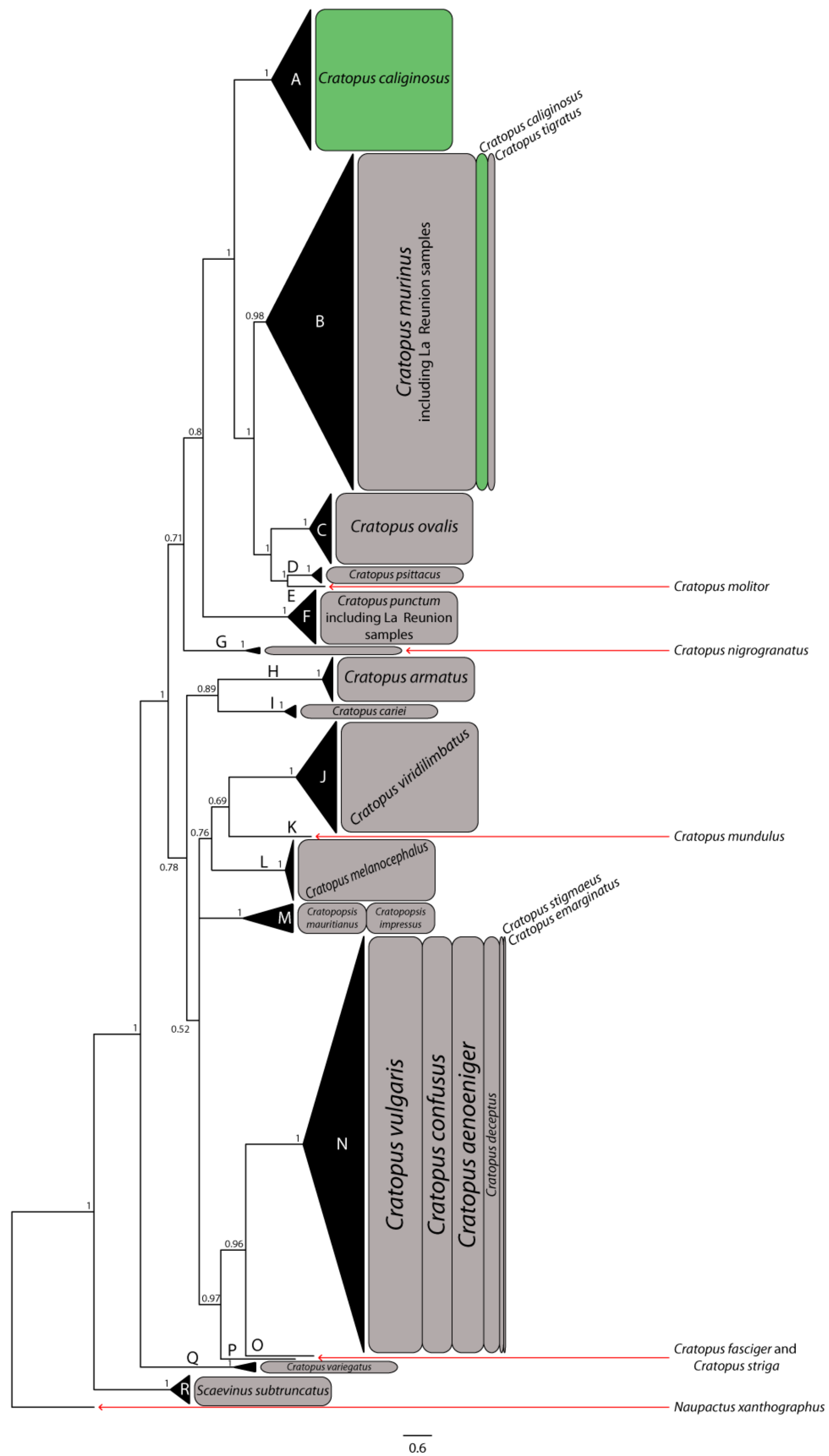


Fig 3.3: A simplified MrBayes tree for all Mauritian Cratopine weevils sampled. The depth of the black triangles represents the coalescent depth of the group, the height of the rectangles represents the number of tips and the width of the rectangles represents the proportion of tips in the group belonging to that morphospecies.

3.4.3 Divergence times and GMYC analyses

The BEAST analysis produced a root age of 9.40 Ma (3.55 Ma – 18.59 Ma HPD interval) 5.11 Ma (3.21 Ma- 7.25 Ma 95% HPD interval) and an ultrametric tree suitable for GMYC analyses. The single threshold GMYC analysis suggested 18 GMYC clusters (confidence interval 16 - 21 clusters) and 22 GMYC entities (confidence interval 19 - 26 entities), providing a significantly better fit to the data than a single GMYC cluster (P <0.01). The multiple threshold analysis produced 23 GMYC clusters (confidence interval 21 - 23 clusters) and 30 GMYC entities (confidence interval 25 - 30 entities), and was also a significantly better fit to the data than a single GMYC cluster (P <0.01). A likelihood ratio test for significant differences between single and multiple threshold models was not significant (X^2 = 3.88, P = 0.69) indicating that the multiple threshold model does not perform significantly better than the single threshold model. Multimodel averaging suggested that across all the best fitting GMYC models, there was an average of 21.50 GMYC clusters (sd = 1.22 clusters) and a total of 26.07 GMYC entities (sd = 1.40 entities). There are 35 separate GMYC entities if nodes with a GMYC threshold probability of less than 0.95 are used as the criterion for separating clusters. 28.5% of the GMYC entities defined using this cut-off contain more than one morphospecies. GMYC clusters of unique sequences defined by the multimodel averaging approach are presented in Figs. 3.4- 3.12.

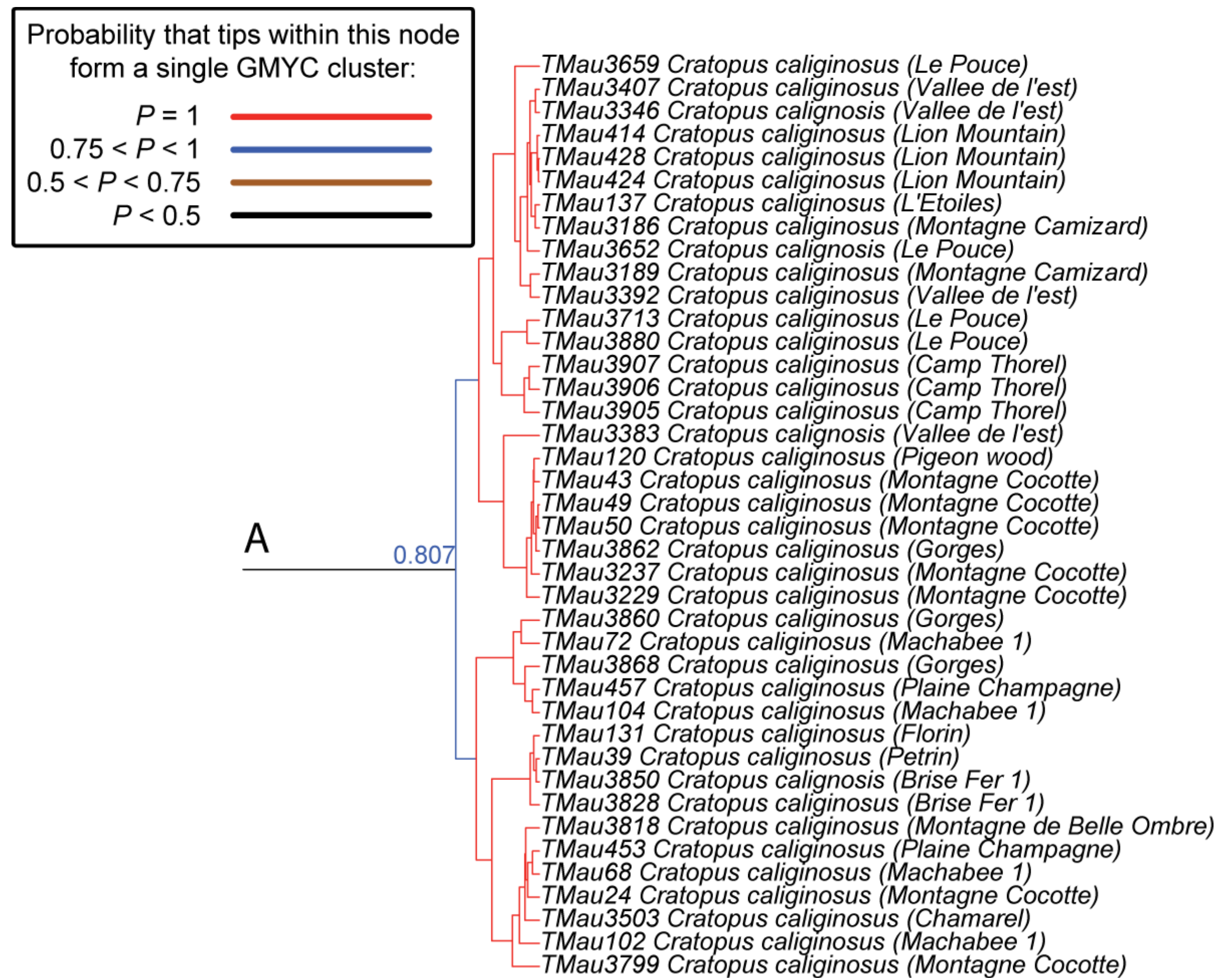


Fig 3.4: The multimodel GMYC analysis result for section A of Fig 3.3 (unique sequences only).

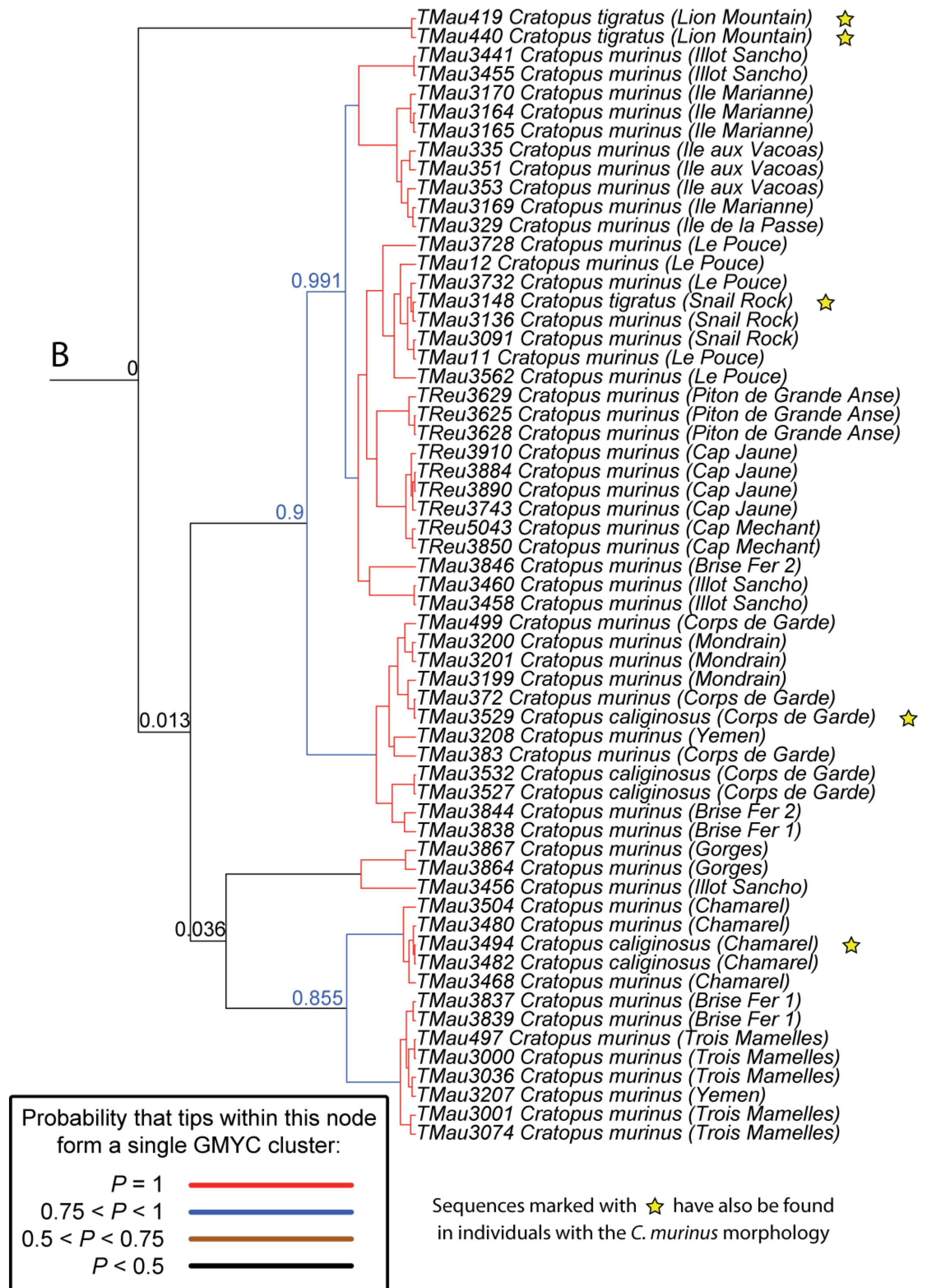


Fig 3.5: The multimodel GMYC analysis result for section B of Fig 3.3 (unique sequences only).

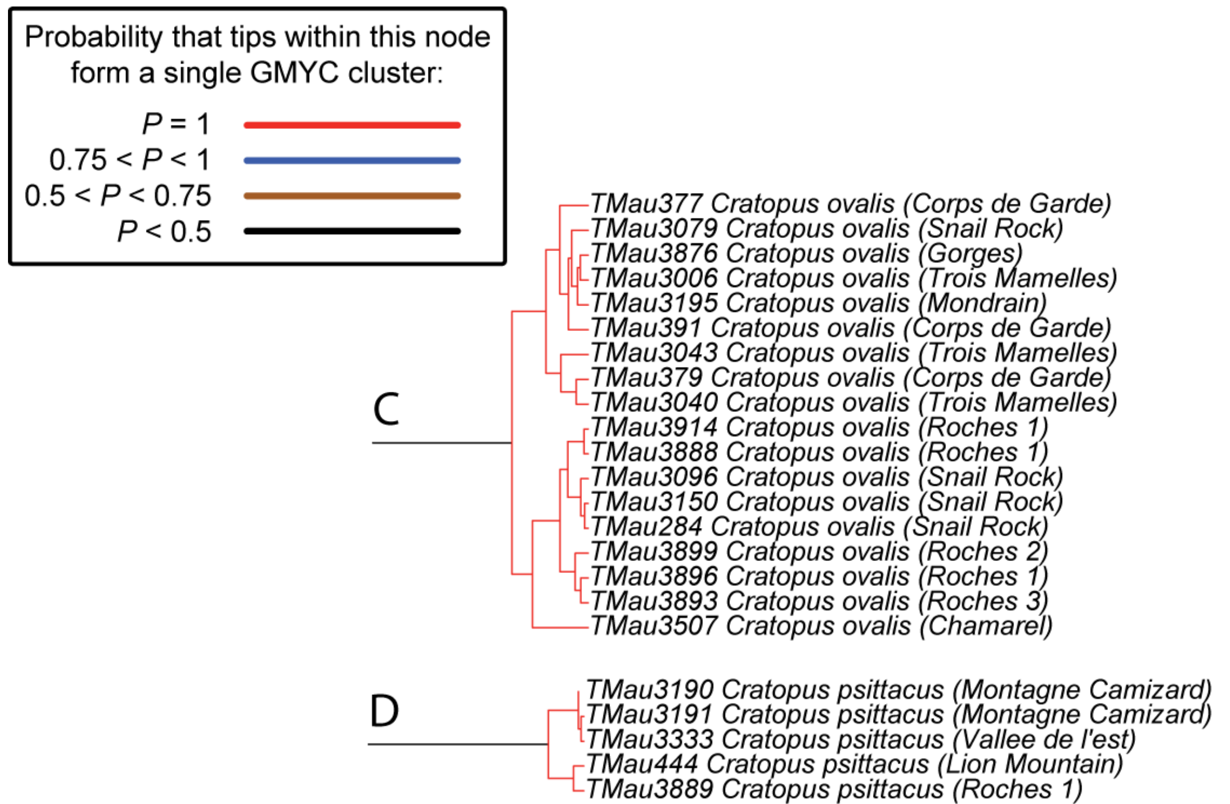


Fig 3.6: The multimodel GMYC analysis result for sections C and D of Fig 3.3 (unique sequences only).

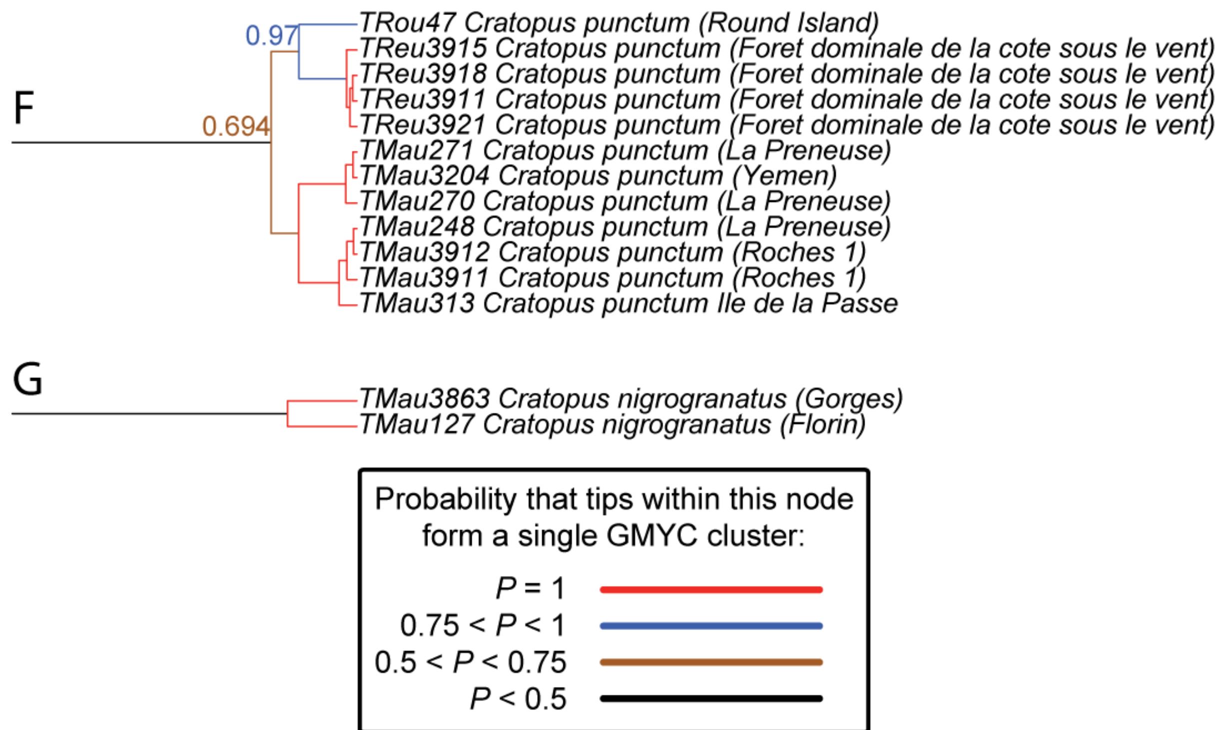


Fig 3.7: The multimodel GMYC analysis result for sections F and G of Fig 3.3 (unique sequences only).

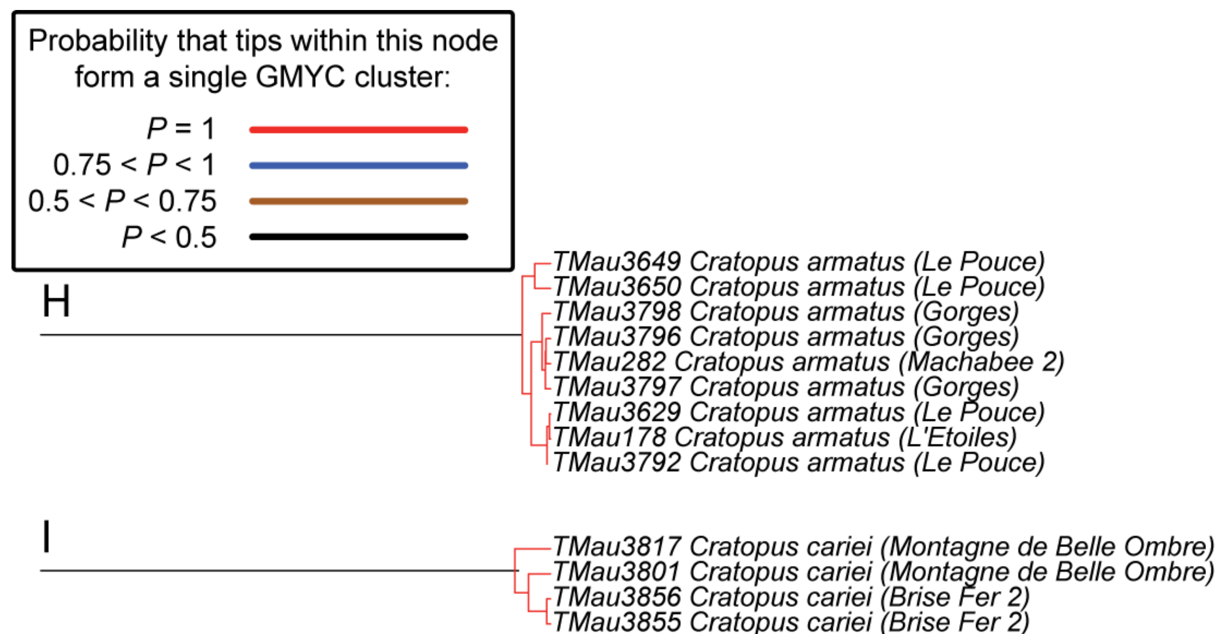


Fig 3.8: The multimodel GMYC analysis result for sections H and I of Fig 3.3 (unique sequences only).

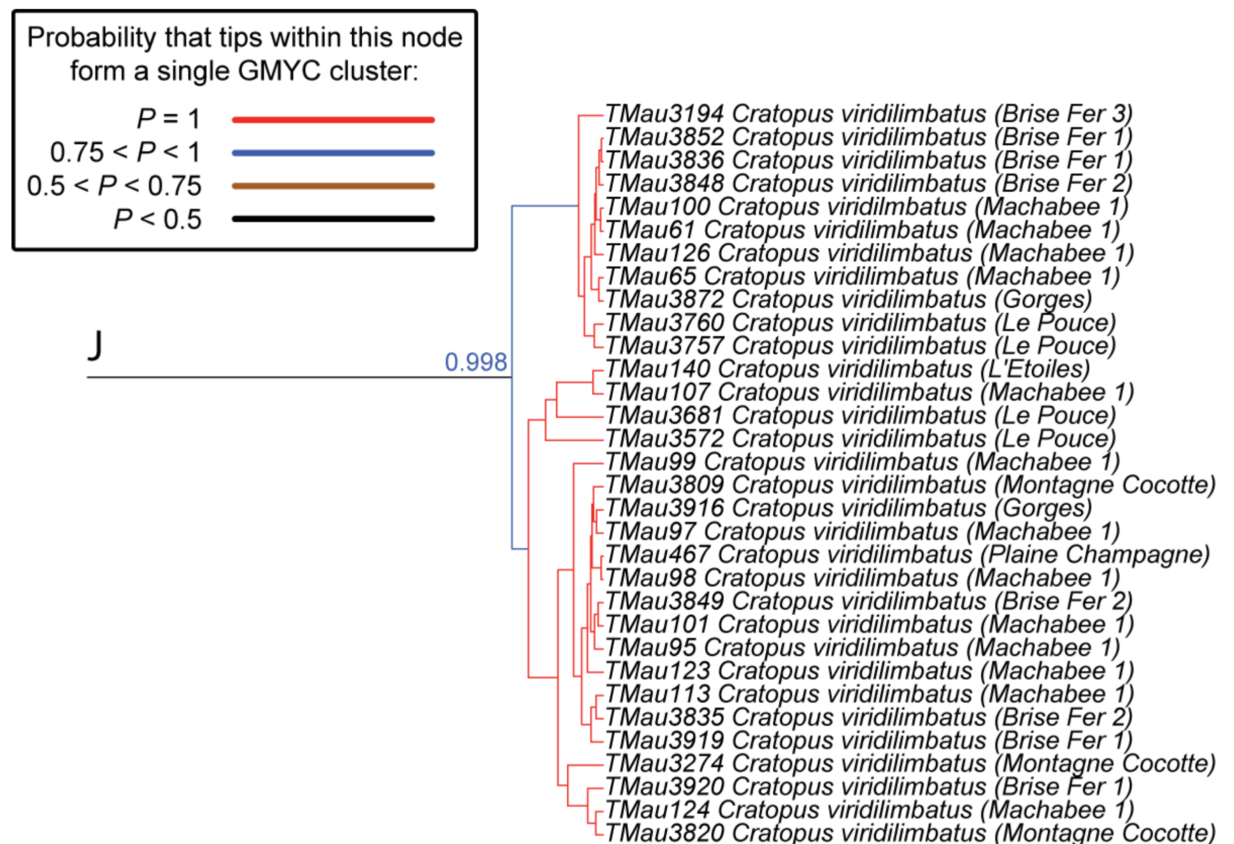


Fig 3.9: The multimodel GMYC analysis result for section J of Fig 3.3 (unique sequences only).

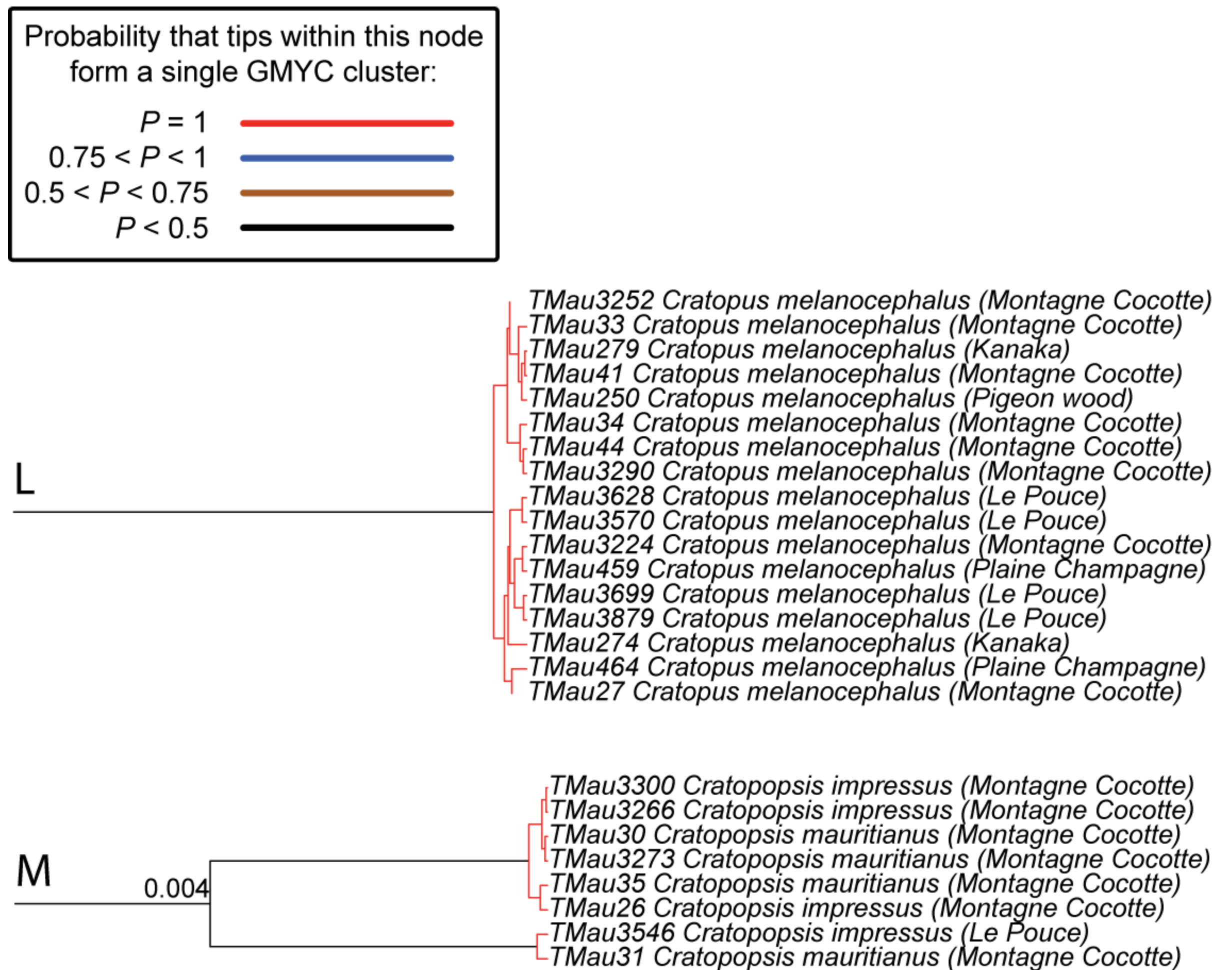


Fig 3.10: The multimodel GMYC analysis result for sections L and M of Fig 3.3 (unique sequences only).

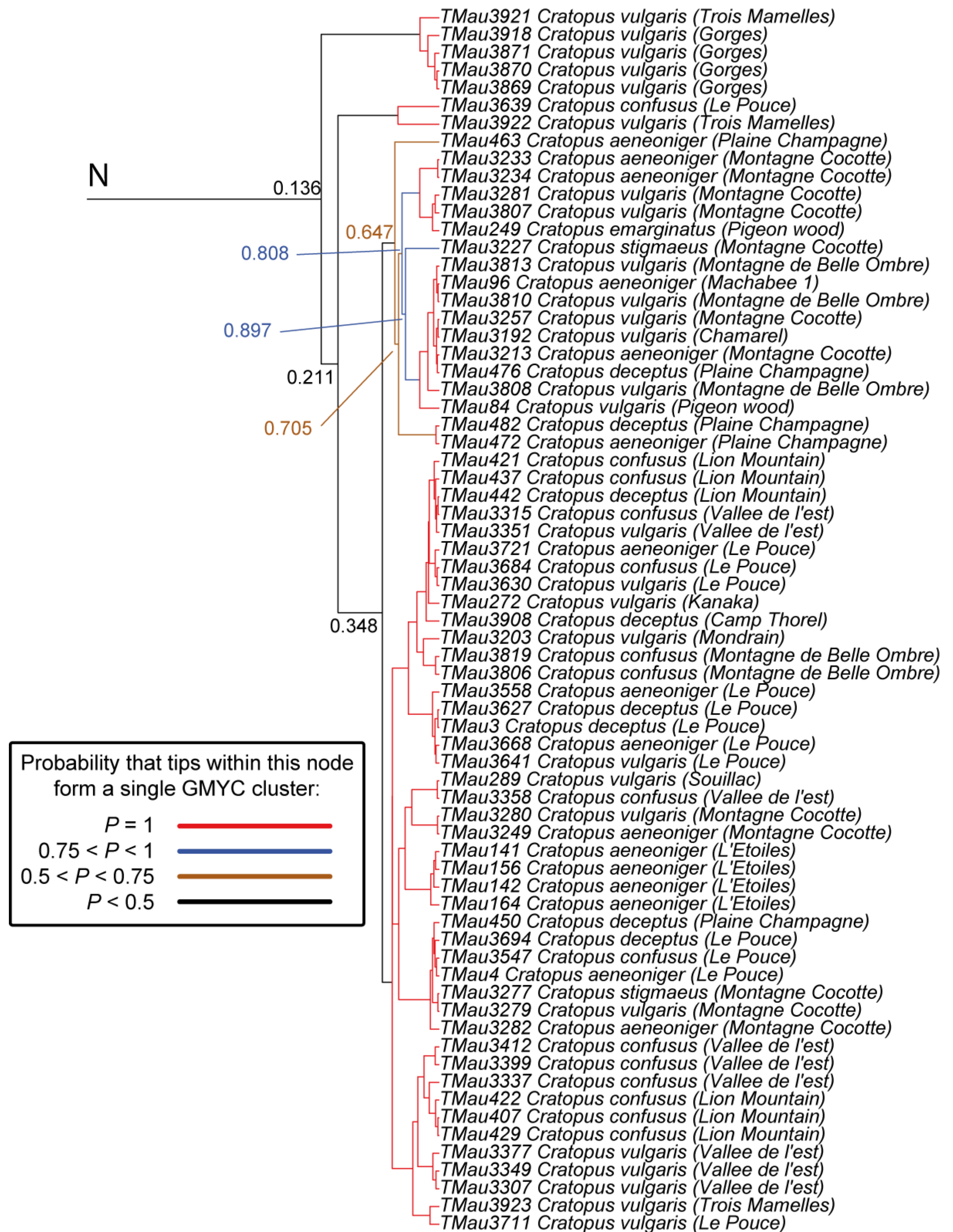


Fig 3.11: The multimodel GMYC analysis result for section N of Fig 3.3 (unique sequences only).

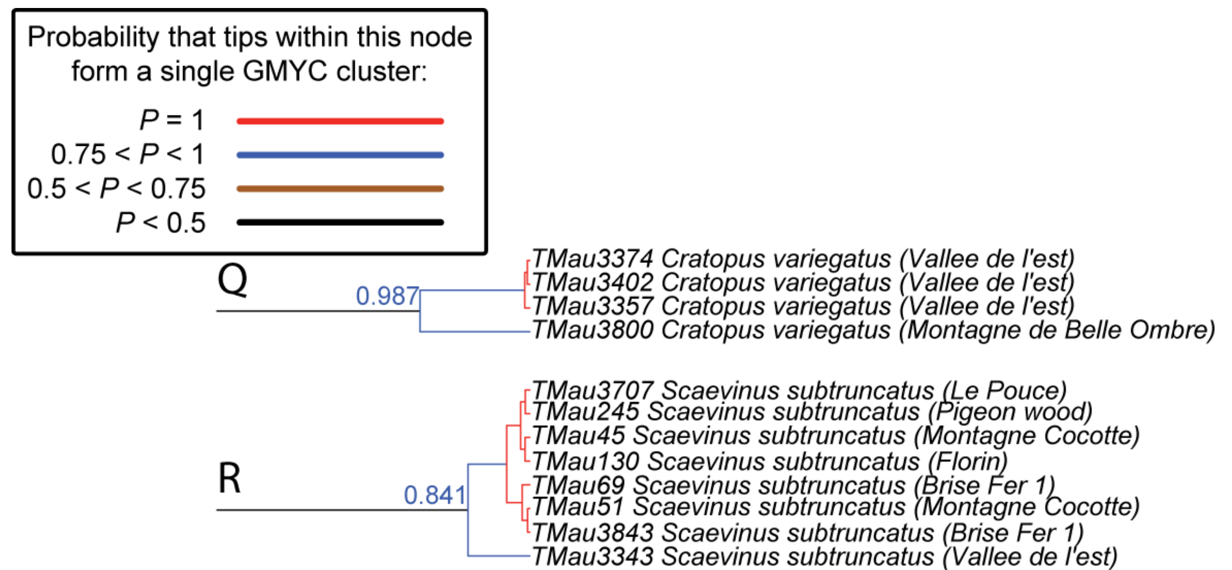


Fig 3.12: The multimodel GMYC analysis result for sections Q and R and I of Fig 3.3 (unique sequences only).

3.4.4 Population structure

Once sites with fewer than 3 samples and geographically monophyletic GMYC clusters had been excluded, four lineages were available for testing (the *C. aeneoniger* morphospecies complex, *C. murinus*, *C. viridilimbatus* and *C. caliginosus*). Due to the apparent lack of monophyly for morphospecies in the *C. aeneoniger* complex, all morphospecies in this complex were treated as a single entity for the purposes of the AMOVA. Additionally, for all AMOVA calculations, sites very close to each other were combined. This relates to Brise Fer 1, 2 and 3, Roches 1, 2 and 3 and Machabee 1 and 2. This results in data sets of six sites for the *C. aeneoniger* complex, eight sites for *C. murinus*, nine sites for *C. caliginosus* and five sites for *C. viridilimbatus*. Significant structure was detected in the *C. aeneoniger* complex, *C. murinus*, *C. caliginosus* (F_{ST} = 0.30, P <0.001, F_{ST} = 0.60, P < 0.001 and F_{ST} = 0.62, P < 0.001 respectively), while no population structure was detected for *C. viridilimbatus* (F_{ST} = 0.03, P = 0.29). The relative locations of the collecting sites included in each AMOVA are presented in Fig. 3.13.

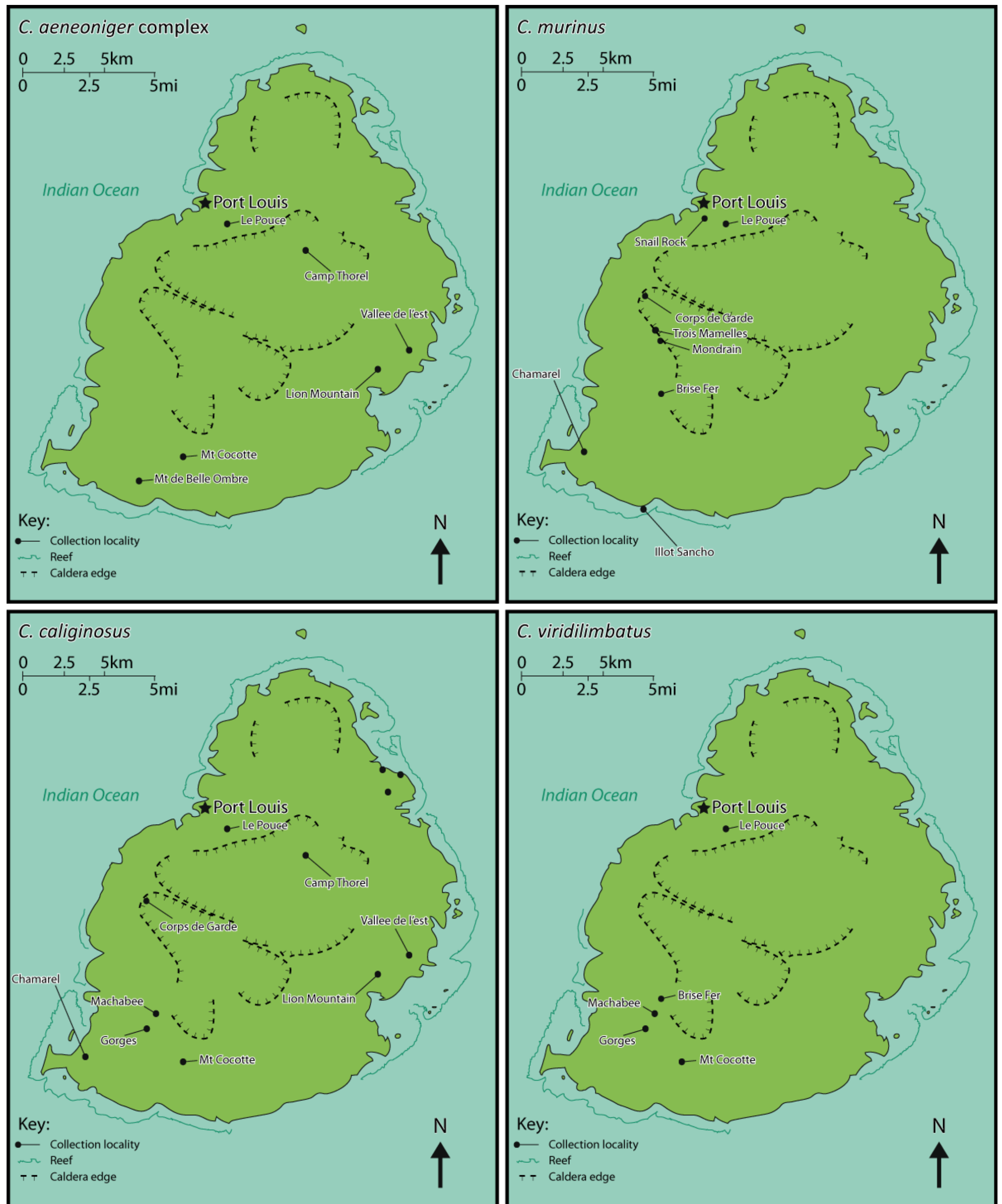


Fig 3.13: Collecting sites included in each AMOVA performed.

3.4.5 Phylogenetic diversity

Sites where a morphospecies was found, but was not successfully amplified, were included in the community matrix for all PD analyses, as they all involved morphospecies that could be unambiguously identified (*C. caliginosus* in Corps de Garde, *C. murinus* in Roches, *C. viridilimbatus* in Montagne de Bel Ombre, *C.*

melanocephalus in Brise Fer and Machabee, *C. fasciger* in Brise Fer and *Cr. impressus* in Brise Fer and Machabee). Additionally, for all PD calculations, sites very close to each other were combined. This relates to Brise Fer 1, 2 and 3, Roches 1, 2 and 3 and Machabee 1 and 2.

Perhaps not surprisingly, Fig 3.14 reveals that PD increases with increasing morphospecies richness. A linear regression of this relationship is significant (R^2_{adj} = 0.90, P < 0.001). Fitting a curved linear model to the data is also significant overall (R^2_{adj} = 0.90, P < 0.001), but the quadratic component of the model is not significant so overall the more complex model does not provide a better fit to the data.

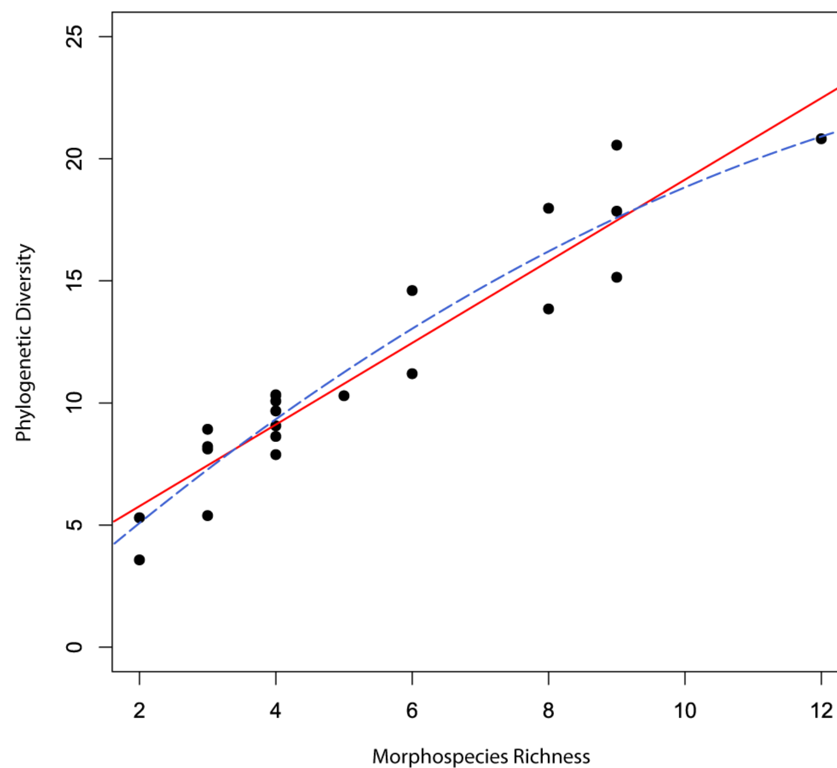


Fig 3.14: A linear regression of phylogenetic diversity against morphospecies richness. The continuous red line represents the straight fitted model while the dashed blue line represents the curved fitted model.

SES_{MPD} values suggest that on average no site has either higher or lower average genetic distance among morphospecies than would be expected by chance (all positive SES values have P values less than 0.95 and all negative SES values have P values greater than 0.05). Conversely, SES_{MNTD} values suggest that 4 sites (Brise Fer, Gorges, Machabee and Le Pouce) all have higher mean nearest taxon distances than would be expected by chance (all have positive SES values and P

values greater than 0.95). A linear regression of non-standardised MPD against morphospecies richness (Fig 3.15) was not significant (R^2_{adj} = 0.05, P = 0.99), whereas a regression of non-standardised MNTD against morphospecies richness (Fig 3.16) was significant (R^2_{adj} = 0.34, P = 0.001) with a slope of -0.21 suggesting that as morphospecies richness increases, mean nearest taxon distance decreases.

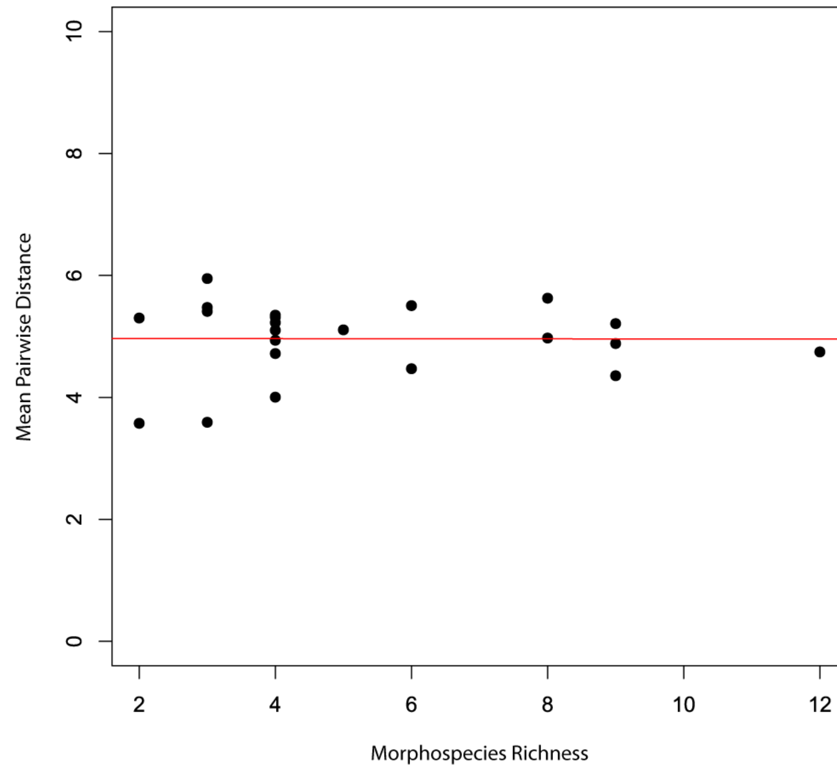


Fig 3.15: A linear regression of mean pairwise distance against morphospecies richness. The red line represents the fitted model.

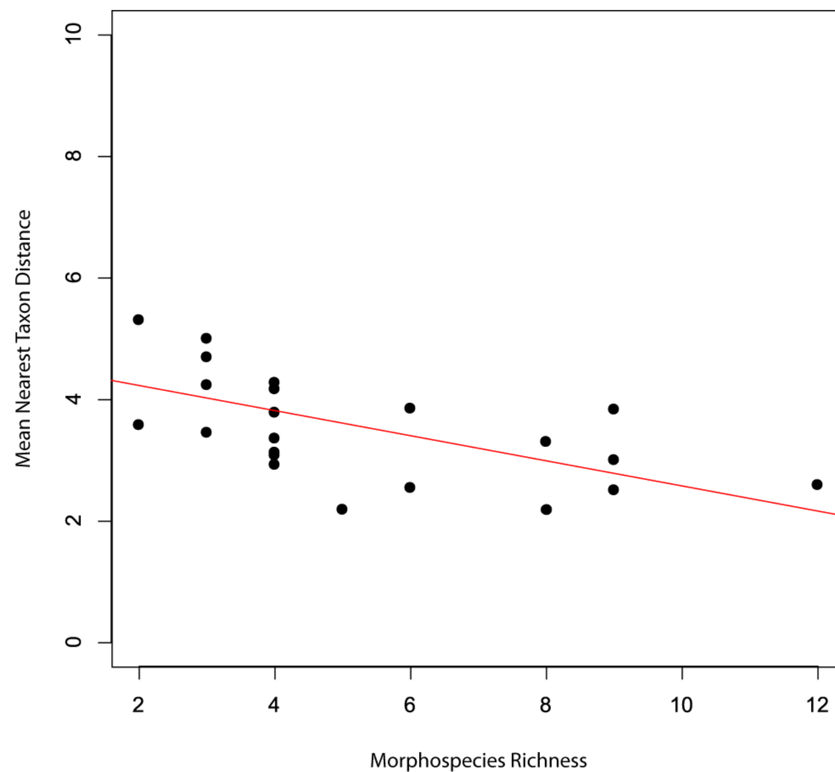


Fig 3.16: A linear regression of mean nearest taxon distance against morphospecies richness. The red line represents the fitted model.

Rarefaction of PD was performed across all possible sites for three, six and nine morphospecies. Three morphospecies was considered to be the lowest bound that it would be worthwhile rarefacting to, while six and nine morphospecies rarefactions permit more accurate comparison for morphospecies rich sites. Fig 3.17 shows the rarefacted PD values for each rarefaction level. For all three graphs, 95% confidence intervals overlap across nearly all sites. This suggests that on average any subsample of morphospecies from a morphospecies rich site will not have any more PD than that of at least one less morphospecies rich site.

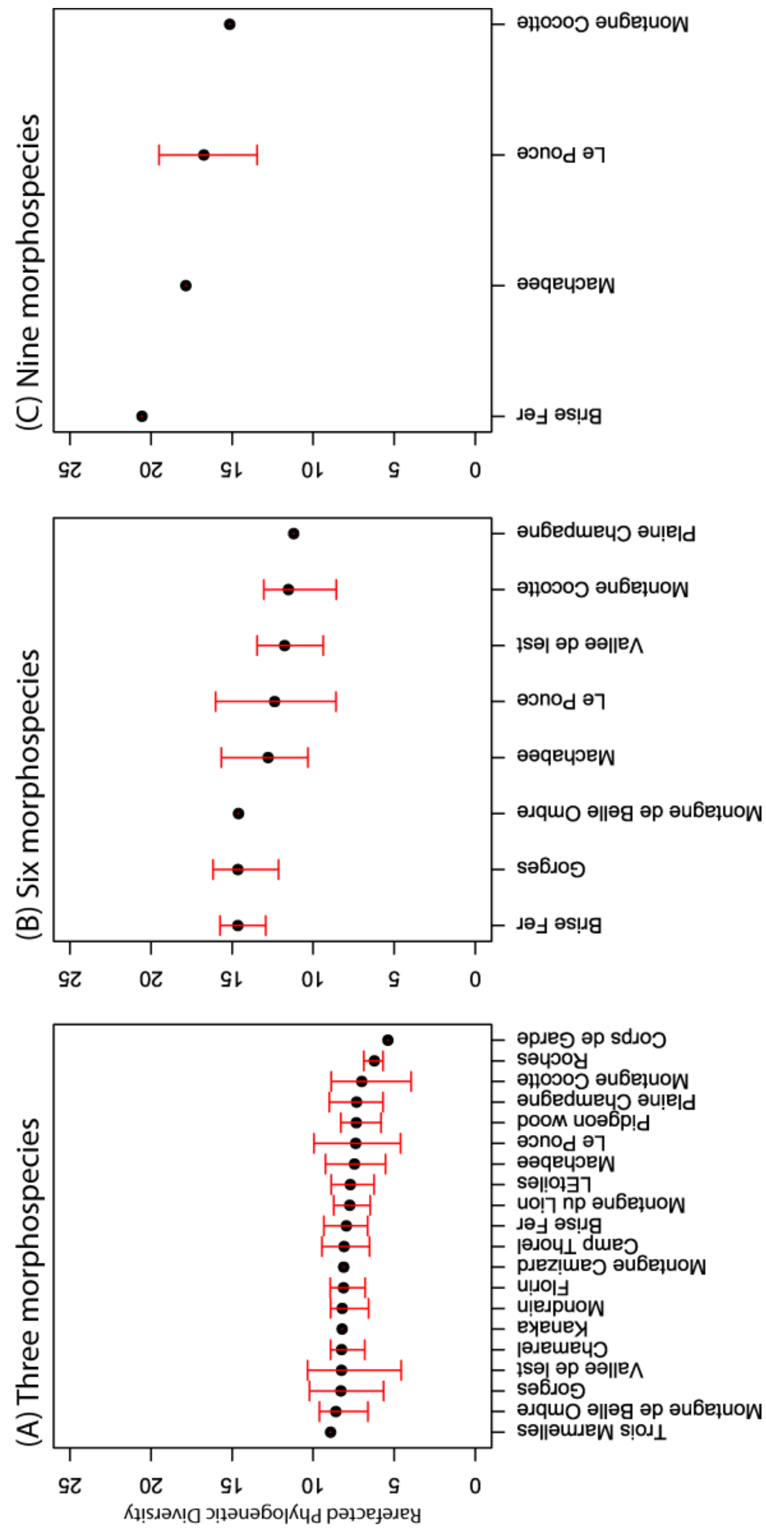


Fig 3.17: Rarefacted phylogenetic diversity for each site. Each section represents a different rarefaction level. Black dots are mean phylogenetic diversity values and red bars are 95% confidence intervals.

3.5 Discussion

3.5.1 The Phylogenetics of Mauritian Cratopine weevils

Cratopine weevil diversity on Mauritius is possibly much younger than Mauritius itself; this is not uncommon in island groups (e.g. Contreras-Diaz *et al.* 2007; Monaghan *et al.* 2006; Price & Clague 2002; Rees *et al.* 2001). However the error associated with our root age is not inconsiderable and could mean that the group started diversifying on Mauritius as little as 1.25 Ma after it emerged above sea level.

Many lineages present in the MrBayes tree (Fig 3.3) contain only one morphospecies, suggesting that the morphospecies contained in lineages C, D, F, G, H, I, J, L, Q and R are likely to represent distinct evolutionary units that equate to species. Four of the rarest morphospecies (specifically ones for which only single sequences were obtained: *C. molitor*, *C. mundulus*, *C. fasciger* and *C. striga*) are recovered as separate monophyletic lineages. This also suggests that these may represent distinct species and not unusual forms of a more common species. Specimens of *C. murinus* and *C. punctum* collected on La Réunion were placed in lineages B and F with their Mauritian counterparts (Fig 3.3). Additionally the monophyly of each Réunionaise site for *C. murinus* and *C. punctum* was well supported. This suggests that both morphospecies have been present on La Réunion long enough to evolve unique mutations *in situ*.

Not all lineages recovered in the Mr Bayes tree contain a single morphospecies. Lineage N (Fig 3.3) contains six morphospecies. Four of these (*C. aeneoniger*, *C. confusus*, *C. deceptus* and *C. vulgaris*) are highly variable and lack single characters that separate them (Williams & Cox 2003). While the remaining two (*C. stigmaeus* and *C. emarginatus*) do have specific characters that define the morphospecies, the characters are subtle so the inclusion of these two morphospecies in lineage N is perhaps unsurprising. Whether the presence of so many morphologies in one lineage is indicative of one highly variable species or incipient speciation is unclear from our data.

Multiple morphospecies are also present within lineage B (Fig 3.3). *C. murinus* and *C. tigratus* are very similar morphospecies; only the dermal scales and colour of the derm (the underlying exoskeleton) differentiate the two. As such, their placement in the same lineage is unsurprising and has already been

suggested as likely by Williams and Cox (2003). This is further emphasised by the fact that identical COII sequences are shared across both morphospecies in two localities (highlighted in Fig. 3.5). Lineage B also contains five individuals of the morphospecies *C. caliginosus*. This morphospecies is distinct from *C. murinus* and *C. tigratus* (there are multiple morphological characters that do not overlap facilitating easy separation of the morphospecies) and indeed all of the remaining *C. caliginosus* sequences obtained are restricted to lineage A. The *C. caliginosus* individuals in lineage B are sympatric with *C. murinus* individuals and the haplotype they possess was also found in *C. murinus* individuals (highlighted in Fig. 3.5). Perhaps the most plausible explanation is hybridisation between *C. murinus* and *C. caliginosus*. This could be further quantified with more detailed sampling of individuals and nuclear genes.

The third lineage containing multiple morphologies is lineage M (*Cr. mauritianus* and *Cr. impressus*). Once again, the characters separating these species are subtle so the inclusion of both morphospecies in the same lineage is unsurprising. As with lineage N, both intraspecific variation and incipient speciation could plausibly explain the pattern of mtDNA relationships between these two morphospecies. As lineage M contains all the described morphospecies of Mauritian *Cratopopsis* and it has been recovered within the monophyletic group containing all the *Cratopus* species sampled, it is likely that *Cratopopsis* is not a valid genus. Whether it is a sub-genus of *Cratopus* or whether it is entirely invalid will depend on the relative positions of the Réunionaise and Rodriguan *Cratopopsis* in a complete phylogeny.

3.5.2 GMYC analyses

Lineages vary in the extent to which they are partitioned by the GMYC analysis. Overall, the most geographically widespread or common lineages such as B and N (Figs 3.5 and 3.11) are subdivided the most strongly. These lineages contain multiple nodes with GMYC probabilities less than 0.5 suggesting that the individuals within these lineages probably do not form a single coalescent unit. For lineage B this appears to be strongly influenced by geography. This is in agreement with the conclusions from chapter two that showed strong population structure for this lineage. Lineage N does not form obvious geographically-linked GMYC clusters. This could either be that a complex demographic history for this lineage

has caused it to deviate from coalescent assumptions or that the morphological variation present is indicative of incipient speciation. Further investigation using more individuals and nuclear markers may be able to resolve this. More narrowly distributed lineages or lineages that contained rarer morphospecies were more likely to contain only a single GMYC cluster. This could be an effect of geography with narrowly distributed lineages having low population structure (e.g. lineage J Fig 3.9). The interaction between geographic range and dispersal ability is likely to be a major factor in determining the degree of population structure since dispersal generally acts to suppress the formation of population structure.

3.5.3 Population structure

Four lineages were available for an assessment of population structure using AMOVA, three of which (lineages A, B and N) revealed significant population structure. The fourth lineage (lineage J) was sampled mostly from sites that are comparatively close together in the southwest of the island, with the only outlying site being Le Pouce further north. As such, dispersal in this species may be enough to suppress population structure forming. One caveat to be aware of when interpreting these results is that the sample sizes used are smaller than would be ideal so the F_{st} values obtained are, at best, approximations. However, the high values seen for some of the lineages would be unlikely to disappear completely with the addition of more specimens.

Studies of flight-capable weevils in continental systems have also revealed population structure, but they have typically been investigated over much broader geographic scales (e.g. Aoki *et al.* 2009; Toju & Sota 2006). Some species, such as the boll weevil (*Anthonomus grandis*), have low F_{ST} values over distances of up to 300 miles (Kim & Sappington 2004) or, as in the case of *Rhinusa antirrhini*, virtually no structure at all (Hernandez-Vera *et al.* 2010). The population structure observed within some species of *Cratopus* suggests that flight, in at least some *Cratopus spp*, is not associated with dispersal over longer distances (in the context of the size of Mauritius), but has been retained for other purposes such as escape from predators or short distance movement between locally patchy food resources. This phenomenon has been noted before and insects that are fully winged yet apparently flightless have been the subject of previous investigations (e.g. Carroll *et al.* 2003; Jackson 1933; Jackson 1956a, b). Stoneflies capable of

flight but with reduced flight musculature have been studied by McCulloch (2009) who observed population structure over relatively small distances. McCulloch (2009) suggested that many of the benefits of flight loss postulated by Harrison (1980) and Roff (1990) can be obtained by simply reducing flight musculature (Zera & Denno 1997); the same may be true of fully winged *Cratopus*. Future work comparing wing morphology or flight musculature across populations with varying structure could provide further insight into how non-flighted species evolve on islands.

3.5.4 Phylogenetic diversity

Measures of MPD and SES_{MPD} suggest that there are no sites that have more or less PD that would be expected by chance. This leads to the conclusion that there are no sites which have a special set of morphospecies that confer a higher level of PD than we might expect for a given number of morphospecies and as such no site that would be considered of special conservation interest based solely on the species present. MNTD values reveal that the more morphospecies there are in a site, the more likely it is that two of them are closely related. This would be expected in a situation where multiple morphospecies are very similar or indistinct in terms of their COII sequences. In this case, some of the inter-morphospecies branch lengths probably represent intraspecific variation rather than all branch lengths representing interspecific variation. However, SES_{MNTD} reveals that four sites have a higher MNTD value than would be expected by chance (Brise Fer, Gorges, Machabee and Le Pouce), suggesting that these sites have taxa that are slightly less related than expected for a given number of morphospecies. This suggests that in these sites, a smaller proportion of the morphospecies present are very closely related or indistinct in terms of their COII sequences. The apparent contradiction between SES_{MPD} and SES_{MNTD} may be attributed to their method of calculation. Standard effect size MNTD and MNTD are calculated using the distance to the nearest tip in the tree and as such they are very sensitive to how closely related sister morphospecies are. This is in contrast to SES_{MPD} and MPD which use the pairwise distances between all morphospecies in the tree. A consequence of this is that SES_{MPD} and MPD are much less sensitive to patterns of diversity within tip groups (Webb *et al.* 2002) and could miss the subtle decrease in relatedness seen in Brise Fer, Gorges, Machabee and Le Pouce.

Rarefacted PD values for small numbers of morphospecies revealed no significant differences among sites once we controlled for morphospecies richness. Small but statistically significant differences were seen when only the most species rich sites were included (Brise Fer vs all other sites in Fig 3.17(C)). This is likely due to the increased numbers of morphospecies from lineage N in Machabee, Le Pouce and Montagne Cocotte adding disproportionately small amounts to the subsampled PD values compared to the more phylogenetically even assemblage of Brise Fer.

Overall, PD values are mostly influenced by morphospecies richness; the actual combinations of morphospecies in each site have only small effects on PD. As such, morphospecies richness is a good proxy for PD in Cratopine weevils on Mauritius. Selection of sites for conservation priorities or for biodiversity assessments of other groups should reflect this with sites such as Brise Fer and Le Pouce being prioritised over less species rich sites.

3.6 Conclusions

Cratopine weevil diversity on Mauritius may be young in comparison to the island the weevils inhabit. However the error associated with our root age is large and it cannot be rejected that the group started diversifying on the island soon after it emerged above sea level. Many morphospecies are likely to represent discrete evolutionary units, but there are some notable exceptions that would appear to represent either morphologically variable species or species complexes in the early stages of speciation. Population structure was present in most of the species sampled sufficiently for assessment, and although our measures provide only an approximation, F_{st} values were reasonably high. Further sampling within widespread species on La Réunion would permit an assessment of the generality of this pattern. Phylogenetic diversity within sampling sites appears to be largely driven by morphospecies richness, such that a simple relationship appears to exist where sites with more morphospecies are the most valuable in terms of conservation priorities. However, population structure within lineages suggests genetic diversity within species is not captured within a single site, and suggests that local populations extinctions may not necessarily be easily reseeded by dispersal.

3.7 References

Aoki K, Kato M, Murakami N (2009) Phylogeographical patterns of a generalist acorn weevil: insight into the biogeographical history of broadleaved deciduous and evergreen forests. *BMC Evolutionary Biology* **9**, 103.

Brower AVZ (1999) Delimitation of phylogenetic species with DNA sequences: A critique of Davis and Nixon's population aggregation analysis. *Systematic Biology* **48**, 199-213.

Carroll SP, Marler M, Winchell R, Dingle H (2003) Evolution of Cryptic Flight Morph and Life History Differences During Host Race Radiation in the Soapberry Bug, Jadera haematoloma Herrich-Schaeffer (Hemiptera: Rhopalidae). *Annals of the Entomological Society of America* **96**, 135-143.

Cegelski CC, Waits LP, Anderson NJ*, et al.* (2006) Genetic diversity and population structure of wolverine (*Gulo gulo*) populations at the southern edge of their current distribution in North America with implications for genetic viability. *Conservation Genetics* **7**, 197-211.

Cicconardi F, Nardi F, Emerson BC, Frati F, Fanciulli PP (2010) Deep phylogeographic divisions and long-term persistence of forest invertebrates (Hexapoda: Collembola) in the North-Western Mediterranean basin. *Molecular Ecology* **19**, 386-400.

Clement WL, Tebbitt MC, Forrest LL*, et al.* (2004) Phylogenetic position and biogeography of Hillebrandia sandwicensis (Begoniaceae): A rare Hawaiian relict. *American Journal of Botany* **91**, 905-917.

Contreras-Diaz HG, Moya O, Oromi P, Juan C (2007) Evolution and diversification of the forest and hypogean ground-beetle genus Trechus in the Canary Islands. *Molecular Phylogenetics and Evolution* **42**, 687-699.

Davis JI, Nixon KC (1992) Populations, Genetic-Variation, and the Delimitation of Phylogenetic Species. *Systematic Biology* **41**, 421-435.

de Thoisy B, da Silva AG, Ruiz-Garcia M*, et al.* (2010) Population history, phylogeography, and conservation genetics of the last Neotropical mega-herbivore, the lowland tapir (Tapirus terrestris). *BMC Evolutionary Biology* **10**.

Drummond A, Ashton B, Buxton S*, et al.* (2011) Geneious v5.4, Available from http://www.geneious.com/.

Drummond AJ, Ho SYW, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biol* **4**, e88.

Emerson BC, Oromi P, Hewitt GM (2000) Colonization and diversification of the species *Brachyderes rugatus* (Coleoptera) on the Canary Islands: Evidence from mitochondrial DNA COII gene sequences. *Evolution* **54**, 911-923.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**, 564-567.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of Molecular Variance inferred from metrix distances among haplotypes - application to human mitochondrial DNA restriction data. *Genetics* **131**, 479-491.

Faith DP (1992) Conservation evaluation and phylogenetic diversity. *Biological Conservation* **61**, 1-10.

Fukami T, Beaumont HJE, Zhang XX, Rainey PB (2007) Immigration history controls diversification in experimental adaptive radiation. *Nature* **446**, 436-439.

Gillespie R (2004) Community assembly through adaptive radiation in Hawaiian spiders. *Science* **303**, 356-359.

Gillespie RG, Claridge EM, Goodacre SL (2008) Biogeography of the fauna of French Polynesia: diversification within and between a series of hot spot archipelagos. *Philosophical Transactions of the Royal Society B-Biological Sciences* **363**, 3335-3346.

Gotelli NJ, Colwell RK (2001) Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters* **4**, 379-391.

Grobler GC, Bastos ADS, Chimimba CT, Chown SL (2011) Inter-island dispersal of flightless *Bothrometopus huntleyi* (Coleoptera: Curculionidae) from the sub-Antarctic Prince Edward Island archipelago. *Antarctic Science* **FirstView**, 1-10.

Harrison RG (1980) Dispersal Polymorphisms in Insects. *Annual Review of Ecology and Systematics* **11**, 95-118.

Hebert PDN, Cywinska A, Ball SL, DeWaard JR (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London Series B-Biological Sciences* **270**, 313-321.

Hernandez-Vera G, Mitrovic M, Jovic J, *et al.* (2010) Host-associated genetic differentiation in a seed parasitic weevil *Rhinusa antirrhini* (Coleptera: Curculionidae) revealed by mitochondrial and nuclear sequence data. *Molecular Ecology* **19**, 2286-2300.

Holland BS, Hadfield MG (2002) Islands within an island: phylogeography and conservation genetics of the endangered Hawaiian tree snail *Achatinella mustelina*. *Molecular Ecology* **11**, 365-375.

Huelsenbeck JP, Ronquist F (2001) MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754-755.

Hustache A (1920) Curculionides des Iles Mascareignes. *Extrait des Annales de la Societe Entomologique de France* **89**.

Illera JC, Emerson BC, Richardson DS (2007) Population history of Berthelot's pipit: colonization, gene flow and morphological divergence in Macaronesia. *Molecular Ecology* **16**, 4599-4612.

Jackson DJ (1933) Observations on the flight muscles of *Sitona* weevils. *Annals of Applied Biology* **20**, 731-770.

Jackson DJ (1956a) The capacity for flight of certain water beetles and its bearing on their origin in the western scottish isles. *Proceedings of the Linnean Society of London* **167**, 76-96.

Jackson DJ (1956b) Observations on flying and flightless water beetles. *Journal of the Linnean Society of London, Zoology* **43**, 18-42.

Jordal BH, Emerson BC, Hewitt GM (2006) Apparent ‘sympatric’ speciation in ecologically similar herbivorous beetles facilitated by multiple colonizations of an island. *Molecular Ecology* **15**, 2935-2947.

Jordal BH, Kirkendall LR, Harkestad K (2004) Phylogeny of a Macaronesian radiation: host-plant use and possible cryptic speciation in *Liparthrum* bark beetles. *Molecular Phylogenetics and Evolution* **31**, 554-571.

Jordan S, Simon C, Foote D, Englund RA (2005) Phylogeographic patterns of Hawaiian Megalagrion damselflies (Odonata : Coenagrionidae) correlate with Pleistocene island boundaries. *Molecular Ecology* **14**, 3457-3470.

Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* **30**, 3059-3066.

Kembel SW, Cowan PD, Helmus MR*, et al.* (2010) Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* **26**, 1463-1464.

Kim KS, Sappington TW (2004) Boll Weevil (Anthonomus grandis Boheman) (Coleoptera: Curculionidae) Dispersal in the Southern United States: Evidence from Mitochondrial DNA Variation. *Environmental Entomology* **33**, 457-470.

Magnacca KN, Danforth BN (2006) Evolution and biogeography of native Hawaiian Hylaeus bees (Hymenoptera : Colletidae). *Cladistics* **22**, 393-411.

McCulloch GA, Wallis GP, Waters JM (2009) Do insects lose flight before they lose their wings? Population genetic structure in subalpine stoneflies. *Molecular Ecology* **18**, 4073-4087.

Monaghan MT, Balke M, Pons J, Vogler AP (2006) Beyond barcodes: complex DNA taxonomy of a south pacific island radiation. *Proceedings of the Royal Society B-Biological Sciences* **273**, 887-893.

Monaghan MT, Wild R, Elliot M*, et al.* (2009) Accelerated Species Inventory on Madagascar Using Coalescent-Based Models of Species Delineation. *Systematic Biology* **58**, 298-311.

Moore J, White WM, Paul D*, et al.* (2011) Evolution of shield-building and rejuvenescent volcanism of Mauritius. *Journal of Volcanology and Geothermal Research* **207**, 47-66.

Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature* **403**, 853-858.

Newton AC, Allnutt TR, Gillies ACM, Lowe AJ, Ennos RA (1999) Molecular phylogeography, intraspecific variation and the conservation of tree species. *Trends in Ecology & Evolution* **14**, 140-145.

Polihronakis M, Caterino MS, Chatzimanolis S (2010) Elucidating the phylogeographic structure among a mosaic of unisexual and bisexual populations of the weevil Geodercodes latipennis (Coleoptera: Curculionidae) in the Transverse Ranges of southern California. *Biological Journal of the Linnean Society* **101**, 935-948.

Pons J, Barraclough TG, Gomez-Zurita J*, et al.* (2006) Sequence-Based Species Delimitation for the DNA Taxonomy of Undescribed Insects. *Systematic Biology* **55**, 595-609.

Pons J, Fujisawa T, Claridge EM*, et al.* (2011) Deep mtDNA subdivision within Linnean species in an endemic radiation of tiger beetles from New Zealand (genus Neocicindela). *Molecular Phylogenetics and Evolution* **59**, 251-262.

Posada D (2008) jModelTest: Phylogenetic model averaging. *Molecular Biology and Evolution* **25**, 1253-1256.

Powell JR (2012) Accounting for uncertainty in species delineation during the analysis of environmental DNA sequence data. *Methods in Ecology and Evolution* **3**, 1-11.

Price JP, Clague DA (2002) How old is the Hawaiian biota? Geology and phylogeny suggest recent divergence. *Proceedings of the Royal Society of London Series B-Biological Sciences* **269**, 2429-2435.

Puillandre N, Meyer CP, Bouchet P, Olivera BM (2011) Genetic divergence and geographical variation in the deep-water Conus orbignyi complex (Mollusca: Conoidea). *Zoologica Scripta* **40**, 350-363.

R Development Core Team (2010) R: A Language and Environment for Statistical Computing, Vienna, Austria.

Rambaut A (2011) FigTree v1.3.1. In: *http://tree.bio.ed.ac.uk/software/figtree/*

Rambaut A, Drummond A (2007) Tracer v1.5.0. In: *http://beast.bio.ed.ac.uk/Tracer*

Rees DJ, Emerson BC, Oromi P, Hewitt GM (2001) Mitochondrial DNA, ecology and morphology: interpreting the phylogeography of the Nesotes (Coleoptera : Tenebrionidae) of Gran Canaria (Canary Islands). *Molecular Ecology* **10**, 427-434.

Roff DA (1990) The evolution of flightlessness in insects. *Ecological Monographs* **60**, 389-421.

Sequeira AS, Lanteri AA, Albelo R, Bhattacharya S, Sijapati M (2008) Colonization history, ecological shifts and diversification in the evolution of endemic Galapagos weevils. *Molecular Ecology* **17**, 1089-1107.

Sequeira AS, Lanteri AA, Scataglini MA, Confalonieri VA, Farrell BD (2000) Are flightless Galapaganus weevils older than the Galapagos Islands they inhabit? *Heredity* **85**, 20-29.

Sites JW, Marshall JC (2003) Delimiting species: a Renaissance issue in systematic biology. *Trends in Ecology & Evolution* **18**, 462-470.

Spurgin LG, van Oosterhout C, Illera JC*, et al.* (2011) Gene conversion rapidly generates major histocompatibility complex diversity in recently founded bird populations. *Molecular Ecology* **20**, 5213-5225.

Thacker RW, Hadfield MG (2000) Mitochondrial phylogeny of extant Hawaiian tree snails (Achatinellinae). *Molecular Phylogenetics and Evolution* **16**, 263-270.

Toju H, Sota T (2006) Phylogeography and the geographic cline in the armament of a seed-predatory weevil: effects of historical events vs. natural selection from the host plant. *Molecular Ecology* **15**, 4161-4173.

Webb CO, Ackerly DD, McPeek MA, Donoghue MJ (2002) Phylogenies and community ecology. *Annual Review of Ecology and Systematics* **33**, 475-505.

Wiens JJ, Penkrot TA (2002) Delimiting species using DNA and morphological variation and discordant species limits in spiny lizards (Sceloporus). *Systematic Biology* **51**, 69-91.

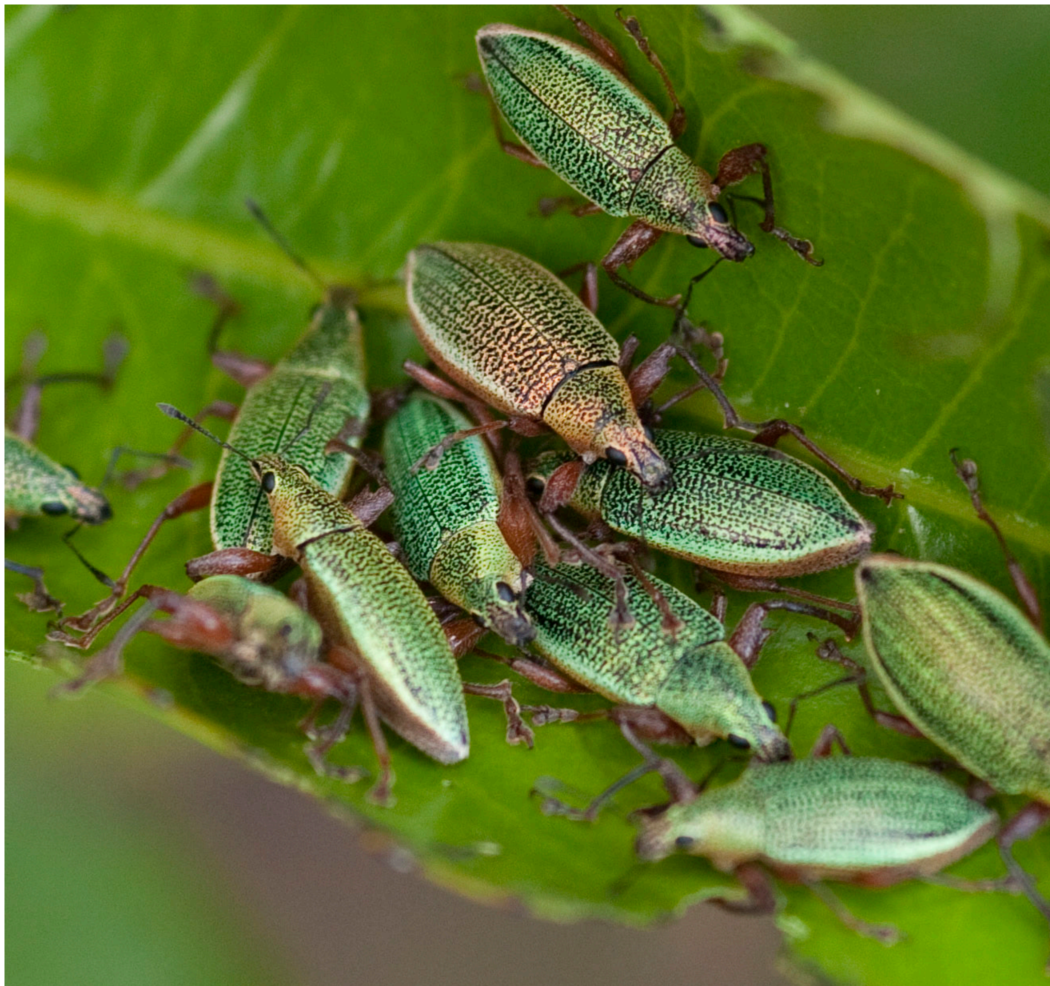
Williams JR, Cox ML (2003) A Contribution to the study of Mascarene Weevils of the genus *Cratopus* Schonherr (Coleoptera: Curculionidae: Entiminae: Cratopini): The species of Mauritius and Rodrigues. *The Mauritius Institute Bulletin* **12.**

Xia X, Xie Z (2001) DAMBE: Software package for data analysis in molecular biology and evolution. *Journal of Heredity* **92**, 371-373.

Xia XH, Xie Z, Salemi M, Chen L, Wang Y (2003) An index of substitution saturation and its application. *Molecular Phylogenetics and Evolution* **26**, 1-7.

Zera AJ, Denno RF (1997) Physiology and ecology of dispersal polymorphism in insects. *Annual Review of Entomology* **42**, 207-230.

Chapter 4: Baseline molecular data for poorly understood diversity: a case study of the Cratopine weevils of La Réunion.



Cratopus sandi resting communally: Les Makes, La Réunion 2010.

4.1 Abstract

A single locus genetic screen of the Cratopine weevil species of La Réunion was undertaken to address taxonomic uncertainty surrounding the Reunionaise members of the tribe. Phylogenetic analyses revealed fourteen separate lineages, nine of which contained a single morphospecies and five of which contained at least two morphospecies. Single threshold GMYC analysis suggests all but two of the lineages examined formed single coalescent groups. Population structure was assessed for selected lineages using AMOVA. As with the Mauritian Cratopine weevils, despite most lineages being flight capable, population structure was common and the possible reasons for this are discussed.

4.2 Introduction

Due to their isolation and ecological simplicity in comparison to continental systems, volcanic oceanic islands are often used to study the evolutionary patterns and processes that shape populations, species and genera (e.g. Gillespie *et al.* 2008; Sequeira *et al.* 2008; Thacker & Hadfield 2000). Much attention has been focussed on interspecific questions and has typically used phylogenetic techniques to address the timing of divergence events (Magnacca & Danforth 2006; Sequeira *et al.* 2000), patterns of diversification (Holland & Hadfield 2002; Jordal *et al.* 2004; Jordan *et al.* 2005) and the resolution of taxonomic uncertainty (e.g. Clement *et al.* 2004; Emerson & Oromi 2005; Jordal *et al.* 2006). At the intraspecific scale coalescent and population genetic approaches are used to characterise genetic structure across populations within the same species (e.g. Emerson *et al.* 2000; Grobler *et al.* 2011; Illera *et al.* 2007; Polihronakis *et al.* 2010). Both approaches can yield information that advances evolutionary understanding (e.g. Fukami *et al.* 2007; Gillespie 2004; Spurgin *et al.* 2011) and conservation management (e.g. Cegelski *et al.* 2006; de Thoisy *et al.* 2010; Newton *et al.* 1999), however knowing which approach to take can prove difficult when the taxonomy for a particular group is incomplete or poorly resolved, and the boundary between interspecific divergence and intraspecific diversity cannot be assumed *a priori*.

In chapter 3, the lack of knowledge regarding species boundaries was addressed for the Mauritian Cratopine weevils. The sampled individuals were partitioned into hypothetical species by applying the GMYC approaches of Pons *et*

al. (2006), Monaghan *et al.* (2009) and Powell (2012) to a single locus sequence dataset (see chapter 3 for a summary of other approaches to molecular species delimitation). Comparisons across all GMYC models using the method of Powel (2012) suggested that sampled individuals could be divided into approximately 35 separate GMYC entities (GMYC clusters and singletons with a GMYC probability >0.95). Ten (28.5%) contained more than one morphospecies suggesting that morphospecies are not reliable units for a phylogenetic analysis of this group without prior investigation. For this reason, the Cratopine morphospecies of La Réunion will be examined using the same methodology. The first objective is to estimate the ages of Cratopine lineages on La Réunion using phylogenetic methods and published evolutionary rates. These are then compared to both the age of the oldest subaerial lavas (2 Ma - McDougal 1971) and the oldest potential age for volcanic activity (5 Ma - Bonneville *et al.* 1988; Gillot *et al.* 1994) to assess the extent to which intra-island diversification has been important in the formation of the Reunionaise Cratopine fauna. Second, morphologically defined species are examined in the context of molecular phylogenetic data and putative model-based inferences of species boundaries. This is particularly useful in the context of rare species and those described on the basis of subtle morphological differences where uncertainty exists as to the validity of taxonomic assignment. Finally genetic structure is assessed for common species on La Réunion. Population structure is a feature of several species of *Cratopus,* despite their potential for dispersal (see chapters 2 and 3). Understanding the extent to which Reunionaise *Cratopus* species are geographically structured can provide insight into both local extinction risk, and potential for incipient speciation.

4.3 Methods

4.3.1 Beetle sampling

Samples were collected by foliage beating on La Réunion during the wet season between June 2007 and December 2011. Samples were placed directly in 99% ethanol solution, with the exception of a few individuals that were kept alive until the same evening for photographing, before being placed in 99% ethanol. Sampling sites were recorded on a handheld GPS unit (Garmin GPS 60, Garmin Ltd). A total of 3430 beetles were collected across 62 sites. See Table 4.1 for GPS co-ordinates of sites and Figs. 4.1a and 4.1b for maps of the sites. Samples were identified using

morphological keys to *Cratopus* and *Cratopopsis* provided by Jaques Poussereau (*pers. comm.*) along with the original descriptions of most species (see the following papers for species descriptions; Ferragu & Richard 1990, 1993; Hustache 1919; Hustache 1920; Richard 1957; Richard 1977; Richard 1995a, b; Vinson 1967; Voisin & Poussereau 2007a, b, 2009). Within each site five individuals (unless fewer than five were obtained) of each morphospecies were selected for DNA extraction and sequencing, giving a data set of 925 individuals.

Table 4.1: Decimal latitudes and longitudes of samples sites on La Réunion.

Site	Latitude °	Longitude °	Site	Latitude °	Longitude °
Basse Vallee	-21.351	55.702	Les Makes	-21.185	55.433
Bayonne	-21.221	55.463	Mare Longue	-21.350	55.743
Bebour	-21.120	55.568	Nez de Boeuf	-21.203	55.619
Bois Ozoux	-21.198	55.648	Petrel Noir	-21.208	55.496
Bras Patate	-21.187	55.436	Pic Adam	-20.946	55.465
Cap Anglais	-21.073	55.521	Pic de Bellevue	-21.289	55.604
Cap Blanc	-21.277	55.661	Piton de Grande Anse	-21.373	55.551
Cap Jaune	-21.378	55.677	Piton de l'eau	-21.175	55.685
Cap Mechant	-21.376	55.709	Piton Desforges	-21.196	55.575
Cascade du Chien	-21.033	55.610	Piton Enchaing	-21.044	55.500
Col de Bellevue	-21.166	55.591	Piton Hyacinthe	-21.218	55.537
Colimacon	-21.123	55.366	Piton Mahot	-21.225	55.586
Colorado	-20.905	55.428	Piton Textor	-21.185	55.631
Etang du Gol	-21.284	55.391	Plaine d'Affouches	-20.983	55.403
Etang St Paul	-20.996	55.295	Plaine des Chicots	-20.980	55.446
Foret de Bel Air	-21.307	55.602	Plaine des Fougiers	-20.980	55.505
Foret de Tevelave	-21.184	55.375	Plaine des Gregues	-21.326	55.610
Foret dominale de la cote sous le Vent	-21.282	55.363	Plaine des Lianes	-21.054	55.576
Foret dominale Notre Dame de la Paix	-21.264	55.603	Plaine des Tamarinds	-21.079	55.443
Foret du Grande Malarum	-21.113	55.490	Ravine des Trois Bassins	-21.099	55.290
Foret du Volcane	-21.206	55.605	Ravine Jean Payet	-21.297	55.527
Gite des Tamarinds	-21.108	55.363	Route de Plaine de les Fougiers	-20.980	55.522
Grand Etang	-21.096	55.649	Route des Tamarinds – start	-21.059	55.366
Grande Chaloupe	-20.928	55.388	Route forrestiere des Makes	-21.192	55.431
Grande Coude	-21.278	55.631	Sentier Botanique - Forest de Bebour	-21.145	55.588
Gros Piton Rond	-21.155	55.598	Sentier Trophee Mondial	-21.200	55.493
Hauts des Makes	-21.177	55.426	Site Dior	-20.999	55.567
La Fenetre	-21.186	55.434	Takamaka	-21.094	55.609
Le Block	-21.131	55.495	Trois Bassins	-21.118	55.361
Le Maido	-21.070	55.387	Vallee Heureuse	-21.329	55.697
Le Tampon	-21.281	55.517	Volcano	-21.218	55.687

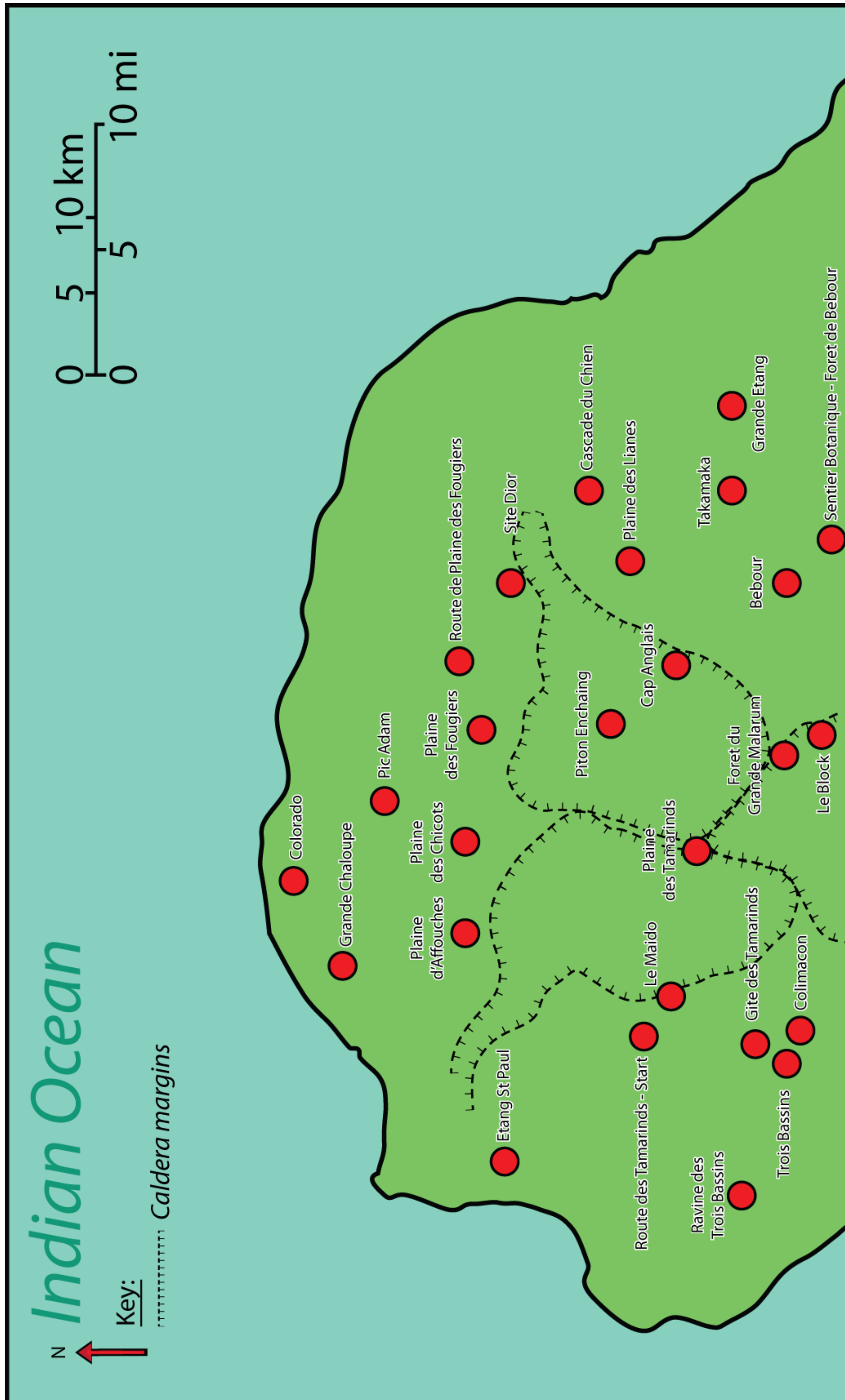


Fig 4.1a: Relative locations of sample sites in the northern half of La Réunion.

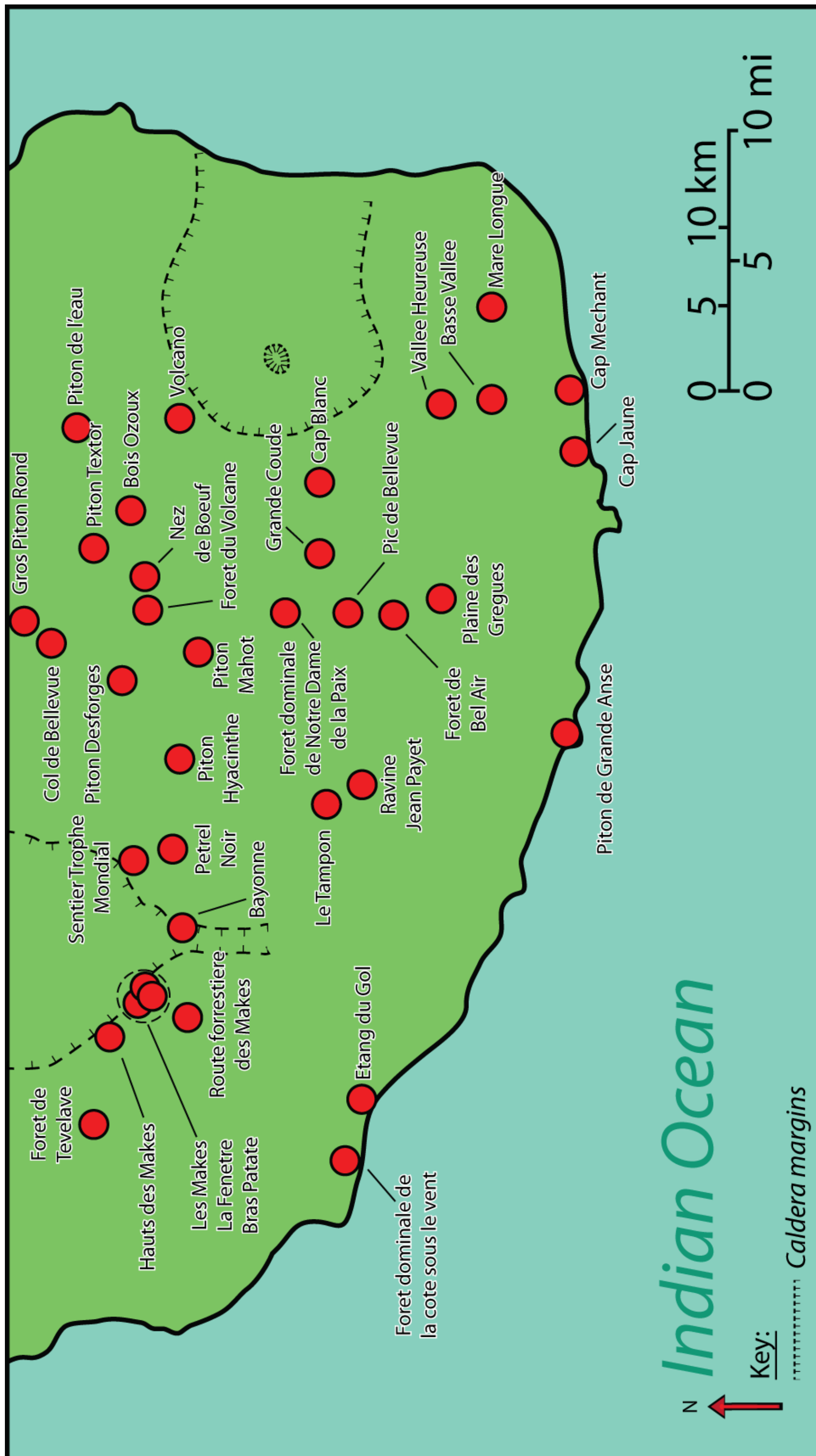


Fig 4.1b: Relative locations of sample sites in the southern half of La Réunion.

4.3.2 DNA extraction, PCR amplification and sequencing

DNA was extracted from the head and pronotum using the DNeasy 96 well Blood and Tissue Extraction kit (QIAGEN, West Sussex, UK) with the digestion buffer volumes amended for large specimens as recommended by the manufacturer. The primers COIICraF (5' TAATATGGCAGAWTAGTGCAATGGA 3') and COIICraR (5' TGCTTTCAGTCATCTAATGATCTRTTTACAGA 3') were used as in chapters two and three to amplify the mitochondrial gene Cytochrome Oxidase II (COII). Amplification conditions were; 0.5 mM of each primer, 5 mM $MgCl_2$ and a thermal profile of: 95°C for 60s, 58°C for 60s and 72°C for 90s, 40 cycles.

Sequencing reactions were performed with the Big Dye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems, California, USA). COIICraF was used for forward sequencing and COIICraR was used for reverse sequencing when forward sequences of less than 600bp were produced. When reverse sequencing was employed, the consensus of each pair of forward and reverse sequences was generated in Geneious Pro version 5.6 (Drummond *et al.* 2012). The thermal profile used for all sequencing reactions was: 96°C for 10s, 58°C for 5s and 60°C for 240s, 25 cycles. Sequences were read on a 3730XL sequencer (Applied Biosystems). All sequences were checked in Geneious Pro version 5.4. Sequences were aligned using MAFFT v6.814b (Katoh *et al.* 2002) with the following parameter values: scoring matrix 1PAM/k=2, Gap open penalty = 1.53, Offset value = 0.123, and then checked by eye. The aligned sequences were tested for saturation using the entropy-based index of substitution saturation (Xia *et al.* 2003) as implemented in DAMBE v5.2.78 (Xia & Xie 2001). This was performed on two data sets, one comprised of the first and second codon positions and the second comprised of the third codon positions. Substitution saturation was visualised by plotting uncorrected pairwise p-distances against corrected pairwise distances. A linear regression of the plotted values provides an alternative method to qualitatively assess saturation as an *x* and coefficient in the fitted model much less than one would indicate a strong disagreement between modelled and observed genetic distances and that the sequences are likely heavily saturated.

4.3.3 Phylogenetic analyses

Trees were constructed from the COII alignment using MrBayes 3.2 (Huelsenbeck & Ronquist 2001) with *Naupactus xanthographus* as an outgroup (GenBank accession number: GU176345). For MrBayes, eight analyses were performed, each for 16 million generations using eight MCMC chains, discarding 25% of the samples as burnin with a model of sequence evolution determined using jModelTest2 (Guindon & Gascuel 2003; Posada 2008) under the BIC and AIC criteria. All parameters permitted under this model were estimated. The output was assessed for stationarity and convergence in Tracer v1.5.0 (Rambaut & Drummond 2007), with only ESS scores greater than 200 being accepted, and the consensus tree was visualised in FigTree v1.3.1 (Rambaut 2011).

4.3.4 Divergence time estimation

To estimate the root age of our sample, and generate an ultrametric tree for the GMYC analyses, a BEAST analysis was performed on an alignment of unique sequences only. Ten replicate analyses were performed using the SRD06 model (Shapiro *et al.* 2006) and an uncorrelated relaxed clock (Drummond *et al.* 2006) with a constant population size as suggested in Monaghan *et al.* (2009). A coleopteran COII mutation rate of 0.0154 substitutions/site/Ma was taken from (Cicconardi *et al.* 2010). The complexity of the analysis was reduced by fixing nodes with posterior probabilities greater than 0.9 in the MrBayes analysis to be monophyletic in the BEAST analysis. Additionally a normal root height prior of 9.5Ma (sd = 1Ma) was applied to restrict the root age of the analysis to estimates that might be expected given the sequence divergences seen in our data set. Analyses were run for 120 million generations and convergence and stationarity were assessed with Tracer v1.5.0. As with the MrBayes analysis, only ESS scores greater than 200 were accepted. After discarding burnin, tree files were combined and subsampled in LogCombiner and the tree for the GMYC analysis was generated in TreeAnnotator and visualised in FigTree v1.3.1.

4.3.5 GMYC analyses

The maximum likelihood number of GMYC clusters and GMYC entities (GMYC clusters plus single branches that cannot be assigned to a GMYC cluster) was

determined in R 2.13.1 (R Development Core Team 2010) using the single threshold GMYC model (Pons *et al.* 2006), the multiple threshold GMYC model (Monaghan *et al.* 2009) and the multimodel inference approach detailed by Powell (2012). The significance of any difference between single and multiple threshold approaches was tested using a likelihood ratio test. GMYC clusters and probability values were then annotated on to the ultrametric BEAST tree.

4.3.6 Population structure

Population structure was tested for lineages using AMOVA (Excoffier *et al.* 1992) as implemented in Arlequin v 3.5 (Excoffier & Lischer 2010). Only lineages containing morphospecies that occurred within at least three sites, each with a minimum of three samples were tested. If a lineage contained more than one morphospecies, then these were treated as the same morphospecies. Instances where strongly supported GMYC clusters contained only a single collecting site were excluded from the analyses as this in itself is evidence for population structure, and this signal would confound any weaker signal contained in geographically mixed clades. Finally, if a lineage was subdivided by the GMYC analysis, then only samples from the largest GMYC cluster were used as testing for structure across putative species may confound the results.

4.4 Results

4.4.1 MtDNA sequencing

Of the 925 beetles sampled for sequencing, 752 were successfully amplified and sequenced for COII. Reverse sequencing was required for 164 individuals giving an overall success rate of 81.3%. Sequences varied from 538bp to 663bp with a mean length of 656.1bp and a standard deviation of 15.1bp. Five hundred and sixty-five unique sequences were recovered across 30 morphospecies with an average pairwise p-distance of 12.8% with a maximum of 19.6% and a standard deviation of 4.4%. When corrected using the general time reversible model (GTR), the average pairwise distance was 14.9%, the maximum was 24.6% and the standard deviation was 5.5%. *Cratopopsis bistigma* (*Cratopopsis* hereafter abbreviated to *Cr.*) sequences fell into two groups. The first group (n=4) aligned well with the

other Cratopine sequences while the second group (n=40) contained multiple frame shift indels and stop codons. These were considered to be NuMts (see Bensasson *et al.* 2001; du Buy & Riley 1967) and excluded from further analyses.

Xia's index for the first and second codon positions produced a value of 0.08, significantly less than the critical value for symmetric topologies (0.71, $P<0.01$) suggesting that sites have not reached mutational saturation. For the third codon position, Xia's index was 0.52, which is also significantly less than the critical value for symmetric topologies (0.67, $P=0.01$). However, the index for this codon position is closer to the critical value, suggesting that third codon position data is approaching maximum divergences that can be reliably inferred. Plotting uncorrected p-distances against genetic distances corrected using the GTR model for all codon position combined (Fig. 4.2) reveals that the modelled distances do deviate from the observed distances suggesting that there may be some tendency towards saturation. Fitting a linear model with a quadratic component to this relationship returns an x coefficient of 0.996 ($P<0.01$) and an x^2 coefficient of -0.787 ($P<0.01$) suggesting that the deviation from a straight line of x=y is small and saturation may not be a serious problem.

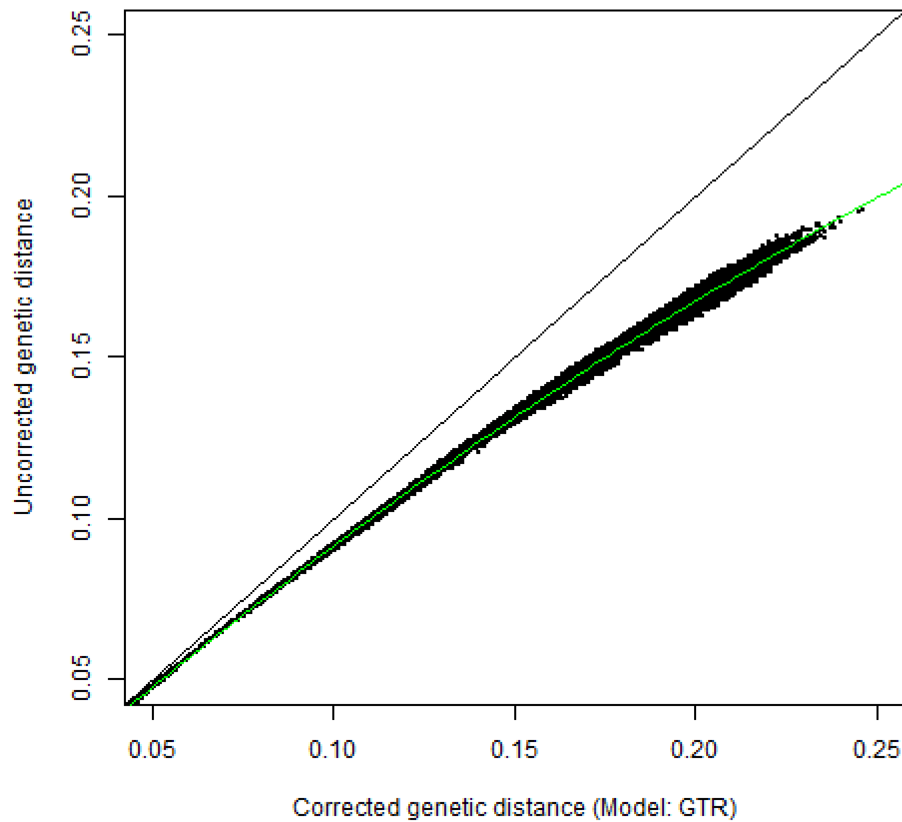


Fig 4.2: A linear regression of uncorrected genetic distance against genetic distance corrected using the GTR model. The solid green line represents the fitted model and the solid black line represents no difference between uncorrected and corrected genetic distances.

4.4.2 Phylogenetic analyses

MrBayes analyses were performed using the GTR+G substitution model returned as the best fit model by jModeltest, yielding a tree (Fig. 4.3a) with 14 distinct lineages, with a lineage being defined as either a monophyletic morphospecies or the smallest monophyletic group that contains paraphyletic morphospecies. Nine of these are monophyletic groups each containing a single morphospecies (C, F, G, H, I, K, L M and N). The remaining five lineages (A, B, D, E, J) are composed of multiple morphospecies. Lineage A comprises *C. sumptuosus* and *C. frappieri*, lineage B consists of *C. nigridorsis*, *C. circumcinctus* and *C. marmoreus* and the morphospecies in lineage D all belong to the *C. humeralis* species group as defined by J. Poussereau (*pers. comm.*). The majority of the samples conform to *C. humeralis* and just three samples conform to the much rarer morphospecies *C.*

alboscuttelatus, *C. fulvescens* and *C. leucophaetus*. Lineage J consists of *C. exquisitus* and *C. sandi*. Finally, lineage E contains all Reunionaise *Cratopopsis* morphospecies other than *Cr*. *bistigma*, (*Cr. alluaudi*, *Cr. antiquus*, *Cr. coquereli*, *Cr. cribatus*, *Cr. fulvicornis*, *Cr. nitidifrons*, *Cr. obscurus* and *Cr. villosulus*). This lineage has considerable substructure with some sublineages containing a single morphospecies while others contain a mixture of morphospecies (see Fig 4.3b for detail). When specimens are compared to the tree, it becomes apparent that there is strong intraspecific morphological variation that partitions by sublineage and locality rather than morphospecies identity (*pers. obs.*). As with chapter 3, the single Reunionaise species of *Scaevinus* (*S. dombaye*) was recovered outside the other genera and the two *Cratopopsis* lineages are placed within *Cratopus* with high support.

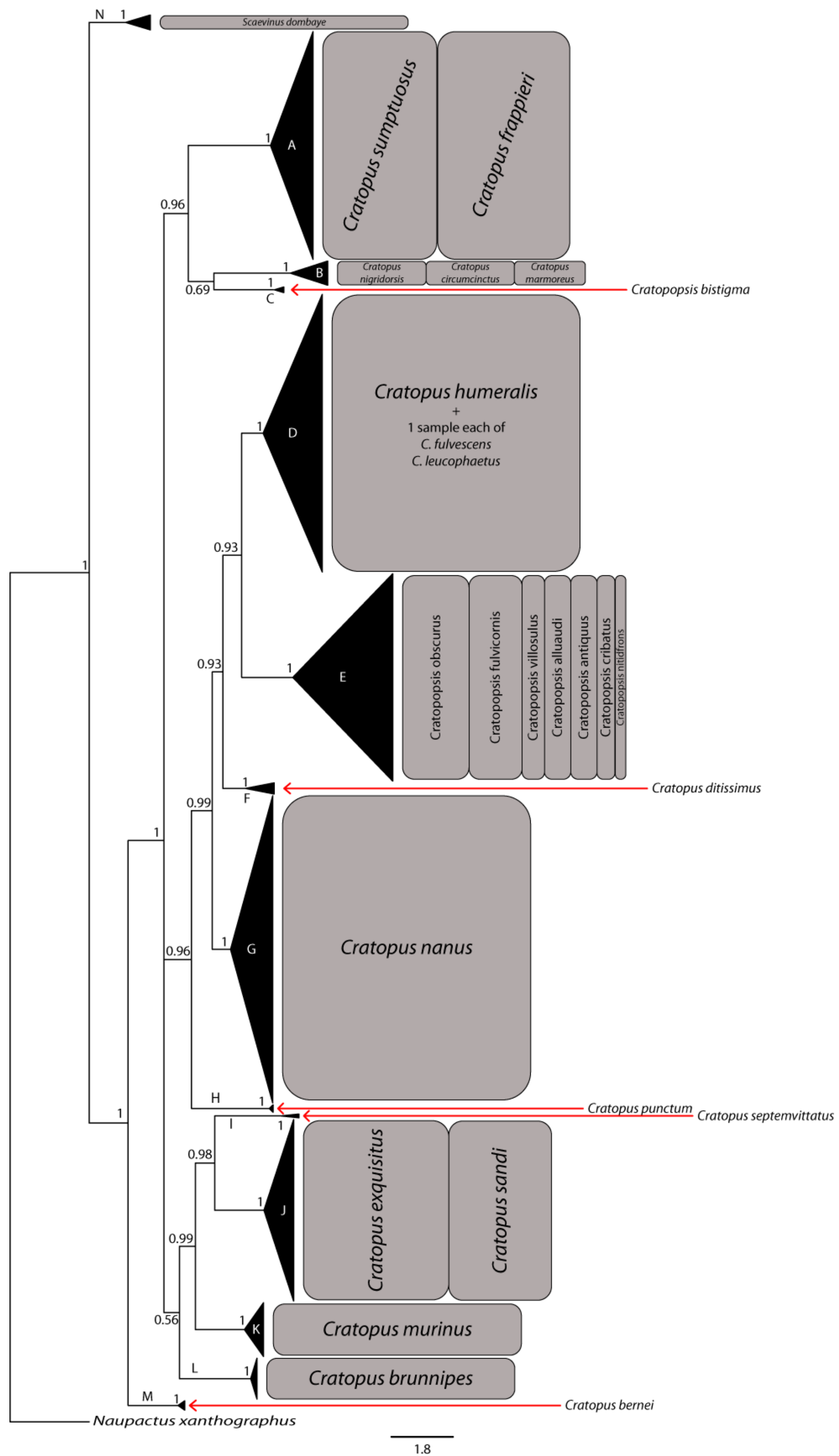


Fig 4.3a: A simplified MrBayes tree for all Reunionaise Cratopine weevils sampled. The depth of the black triangles represents the coalescent depth of the group, the height of the rectangles represents the number of tips and the width of the rectangles represents the proportion of tips in the group belonging to that morphospecies. Numbers beside nodes are the Bayesian posterior probability for that node.

Fig 4.3b: Detail for lineage E. Inset is a map of collecting sites on La Réunion. Tip labels are coloured by morphological similarity and sites are coloured to match tip labels. When more than one lineage is present at a site, the marker is subdivided and coloured to represent this. Numbers beside nodes are the Bayesian posterior probability for that node.

4.4.3 Divergence times and GMYC analyses

The age estimates for all nodes in the BEAST tree can be seen in Fig. 4.4. Two lineages (D and E) have mean age estimates older than the oldest subaearial lavas on La Réunion (approximately 2Ma (McDougal 1971)). Lineage D has an estimated mean age of 2.2Ma (1.52Ma – 2.95Ma 95% HPD interval) while lineage E is older with an estimated mean age of 2.65Ma (1.97Ma – 3.39Ma 95% HPD interval). These are both less than the estimated 5Ma age of La Réunion from Bonneville (1988).

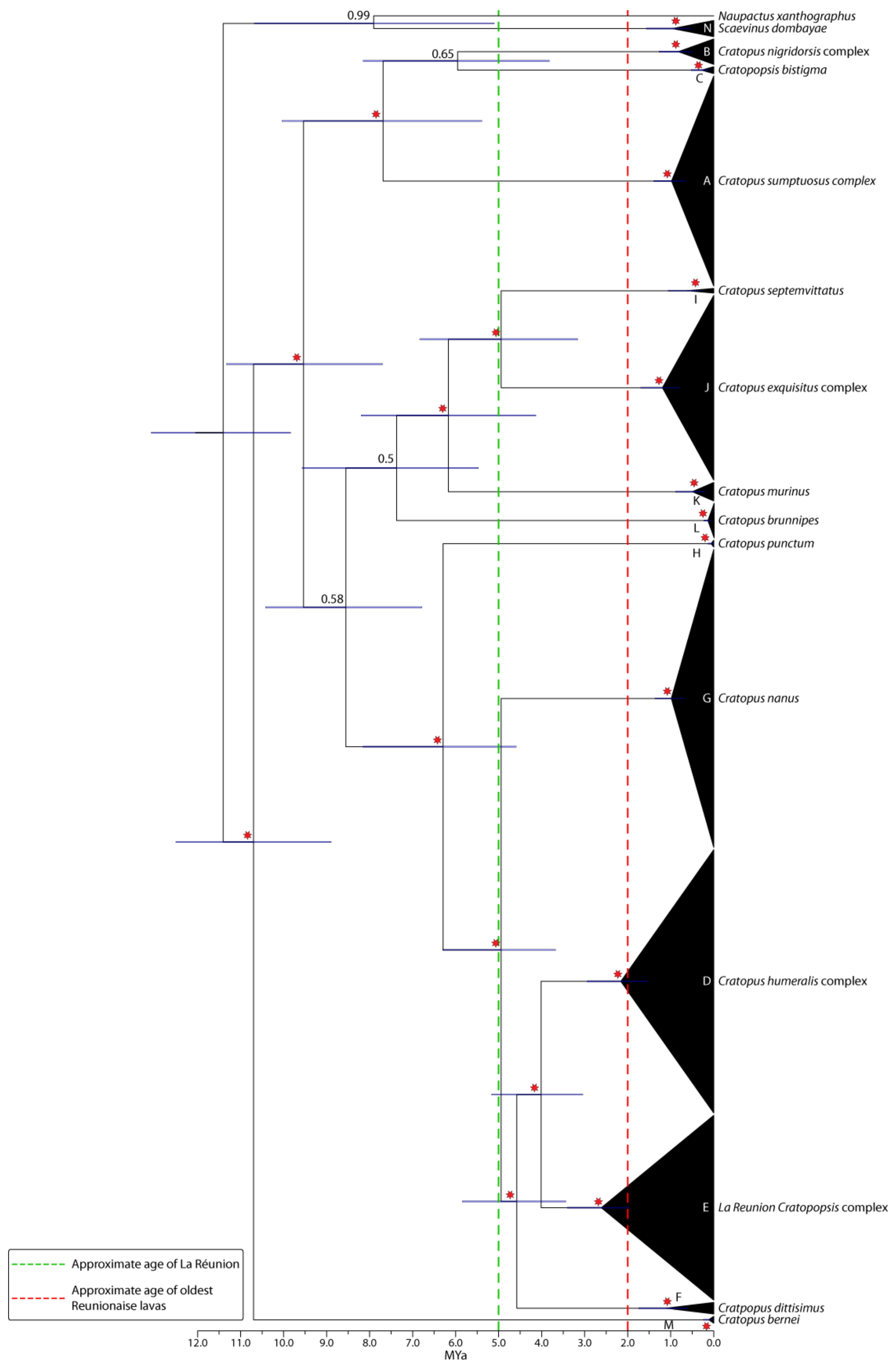


Fig 4.4: A simplified chronogram of Reunionaise Cratopine weevils generated in BEAST. Node bars represent 95% HPD intervals on the age estimate for that node. Stars represent nodes for which monophyly was enforced and numbers beside nodes are Bayesian posterior probabilities for that node.

The most likely single threshold GMYC analysis suggested 17 GMYC clusters (confidence interval 14-21 clusters) and 18 GMYC entities (confidence interval 15-23 entities), providing a significantly better fit to the data than a single GMYC cluster (X^2 = 35.28, P <0.01). The most likely multiple threshold analysis produced 105 GMYC clusters (confidence interval 101-113 clusters) and 191 GMYC entities (confidence interval 179-207 entities), and was also a significantly better fit to the data than a single GMYC cluster (X^2 = 43.39, P <0.01). A likelihood ratio test for significant differences between single and multiple threshold models was not significant (X^2 = 8.10, P = 0.99) indicating that the multiple threshold model does not perform significantly better than the single threshold model. Multimodel averaging suggested that across all the best fitting GMYC models, there was an average of 42.64 GMYC clusters (sd = 6.00 clusters) and a total of 66.78 GMYC entities (sd = 8.33 entities). There are 203 separate GMYC entities if nodes with a GMYC threshold probability of less than 0.95 are used as the criterion for separating clusters. 15.3% of the GMYC entities defined using this cut-off contain more than one morphospecies. GMYC clusters of unique sequences defined by the multimodel averaging approach are presented in Figs. 4.5 – 4.11. Nodes marked with red stars represent the GMYC cut-offs suggested by the single threshold analysis

Probability that tips within this node form a single GMYC cluster:
$P \geq 0.95$
$0.75 < P < 0.95$
$0.5 < P < 0.75$
$P < 0.5$

A

Reu2434 Cratopus septemvittatus
Reu2606 Cratopus frappieri Le Maido
Reu2617 Cratopus frappieri Le Maido
Reu5049 Cratopus sumptuosus Plaine des Tamarinds
Reu5062 Cratopus frappieri Plaine des Tamarinds
Reu5082 Cratopus sumptuosus Plaine des Tamarinds
Reu2745 Cratopus frappieri Route des Tamarinds start
Reu5008 Cratopus frappieri Le Block
Reu4900 Cratopus sumptuosus Le Block
Reu4024 Cratopus frappieri Les Makes
Reu3181 Cratopus sumptuosus Les Makes
Reu3927 Cratopus sumptuosus Les Makes
Reu3167 Cratopus frappieri Foret de Tevelave
Reu3926 Cratopus frappieri Les Makes
Reu3988 Cratopus sumptuosus Les Makes
Reu3220 Cratopus sumptuosus Route forrestiere des Makes
Reu3226 Cratopus frappieri Route forrestiere des Makes
Reu3191 Cratopus frappieri Les Makes
Reu4008 Cratopus frappieri Les Makes
Reu3223 Cratopus sumptuosus Route forrestiere des Makes
Reu3930 Cratopus frappieri Les Makes
Reu4029 Cratopus sumptuosus Les Makes
Reu3165 Cratopus frappieri Foret de Tevelave
Reu4757 Cratopus frappieri Trois Bassins
Reu3265 Cratopus frappieri Colimacon
Reu3274 Cratopus sumptuosus Colimacon
Reu4755 Cratopus sumptuosus Trois Bassins
Reu2667 Cratopus frappieri Colimacon
Reu3280 Cratopus frappieri Colimacon
Reu4758 Cratopus frappieri Trois Bassins
Reu4492 Cratopus frappieri Cap Blanc
Reu245 Cratopus tristis Volcano
Reu4725 Cratopus frappieri Grande Coude
Reu4708 Cratopus frappieri Grande Coude
Reu4726 Cratopus sumptuosus Grande Coude
Reu4203 Cratopus sumptuosus Foret de Bel Air
Reu4513 Cratopus sumptuosus Foret Dominale Notre Dame de la Paix
Reu4512 Cratopus sumptuosus Foret Dominale Notre Dame de la Paix
Reu4271 Cratopus frappieri Foret de Bel Air
Reu4272 Cratopus frappieri Foret de Bel Air
Reu3447 Cratopus sumptuosus Piton Mahot
Reu3668 Cratopus sumptuosus Foret du Volcane
Reu2264 Cratopus frappieri Piton Desforges
Reu3387 Cratopus sumptuosus Piton Mahot
Reu3704 Cratopus sumptuosus Foret du Volcane
Reu4508 Cratopus frappieri Foret Dominale Notre Dame de la Paix
Reu3464 Cratopus frappieri Piton Hyacinthe
Reu2191 Cratopus frappieri Piton Hyacinthe
Reu3667 Cratopus sumptuosus Bois Ozoux
Reu3641 Cratopus frappieri Nez de Boeuf
Reu2968 Cratopus frappieri Nez de Boeuf
Reu3642 Cratopus sumptuosus Nez de Boeuf
Reu3821 Cratopus frappieri Sentier Trophee Mondial
Reu3816 Cratopus sumptuosus Sentier Trophee Mondial
Reu3831 Cratopus sumptuosus Sentier Trophee Mondial
Reu2428 Cratopus frappieri Plaine des Fougiers
Reu3767 Cratopus frappieri Plaine des Fougiers
Reu3790 Cratopus frappieri Plaine des Fougiers
Reu3752 Cratopus sumptuosus Plaine des Fougiers
Reu3776 Cratopus sumptuosus Plaine des Fougiers
Reu2873 Cratopus frappieri Gros Piton Rond
Reu2878 Cratopus sumptuosus Gros Piton Rond
Reu3550 Cratopus frappieri Sentier Botanique Foret de Bebour
Reu2876 Cratopus frappieri Gros Piton Rond
Reu2121 Cratopus sumptuosus Cap Anglais
Reu4101 Cratopus frappieri Plaine d'Affouches
Reu4107 Cratopus frappieri Plaine d'Affouches
Reu4106 Cratopus sumptuosus Plaine d'Affouches
Reu4091 Cratopus sumptuosus Plaine d'Affouches
Reu4839 Cratopus frappieri Plaine des Chicots
Reu4838 Cratopus sumptuosus Plaine des Chicots
Reu4783 Cratopus frappieri Plaine des Chicots
Reu4904 Cratopus sumptuosus Plaine des Chicots
Reu3127 Cratopus frappieri Plaine d'Affouches
Reu2755 Cratopus sumptuosus Grand Etang
Reu55 Cratopus sumptuosus Basse Vallee
Reu54 Cratopus sumptuosus Basse Vallee
Reu2894 Cratopus sumptuosus Mare Longue
Reu2884 Cratopus sumptuosus Mare Longue
Reu2449 Cratopus frappieri Vallee Heureuse
Reu2911 Cratopus sumptuosus Mare Longue
Reu2199 Cratopus frappieri Vallee Heureuse
Reu2548 Cratopus sumptuosus Vallee Heureuse
Reu2284 Cratopus sumptuosus Vallee Heureuse
Reu2144 Cratopus frappieri Vallee Heureuse
Reu2027 Cratopus sumptuosus Vallee Heureuse
Reu3832 Cratopus frappieri Basse Vallee
Reu2021 Cratopus frappieri Basse Vallee
Reu2451 Cratopus sumptuosus Vallee Heureuse
Reu4315 Cratopus sumptuosus Plaine des Gregues
Reu3834 Cratopus sumptuosus Basse Vallee
Reu2447 Cratopus sumptuosus Vallee Heureuse
Reu2448 Cratopus sumptuosus Vallee Heureuse

Fig 4.5: The multimodel GMYC analysis result for lineage A of Fig 4.3a (unique sequences only). Stars represent GMYC clusters from the single threshold GMYC analysis.

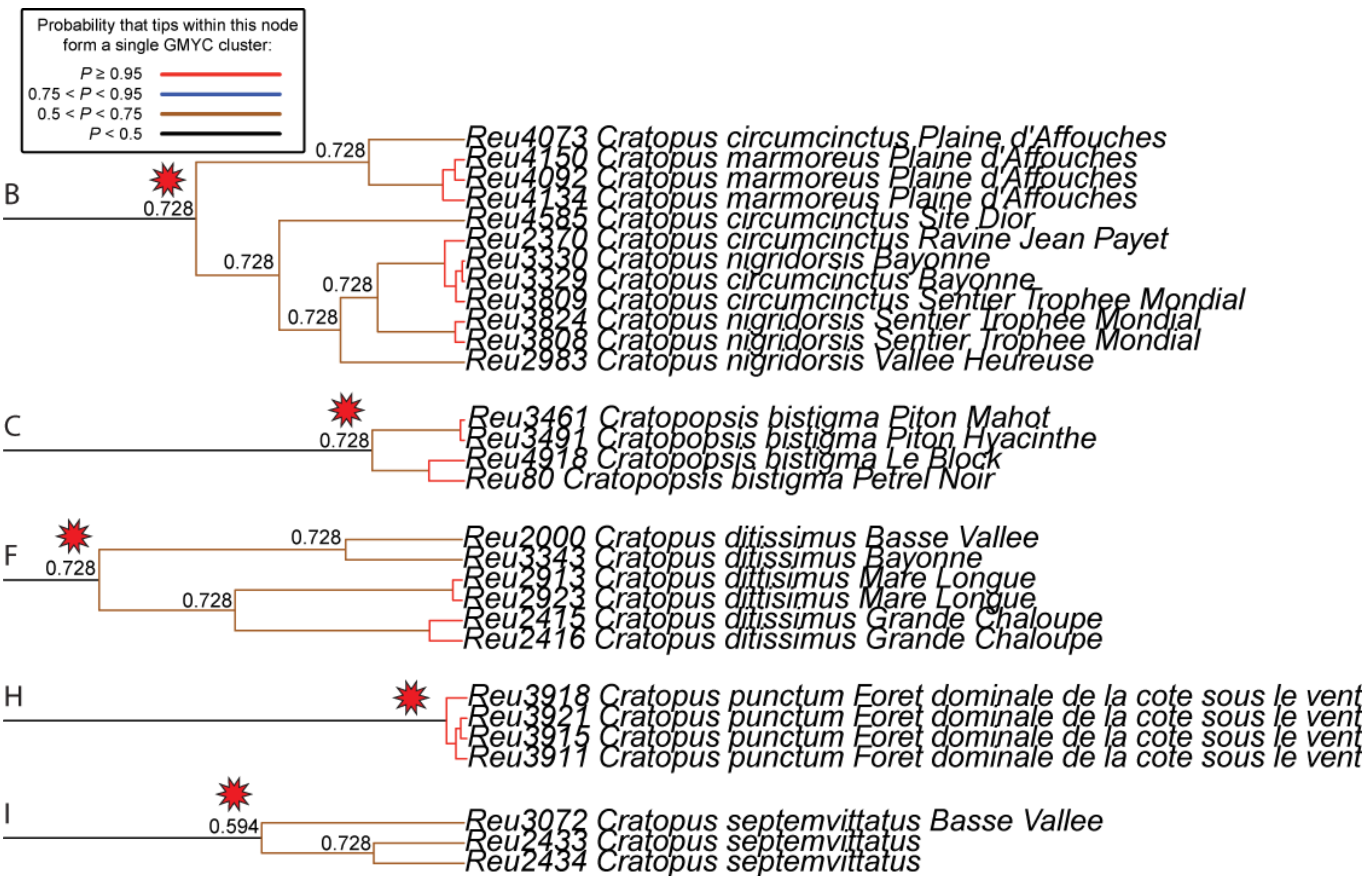


Fig 4.6: The multimodel GMYC analysis result for lineages B, D, F, H and I of Fig 4.3a (unique sequences only). Stars represent GMYC clusters from the single threshold GMYC analysis.

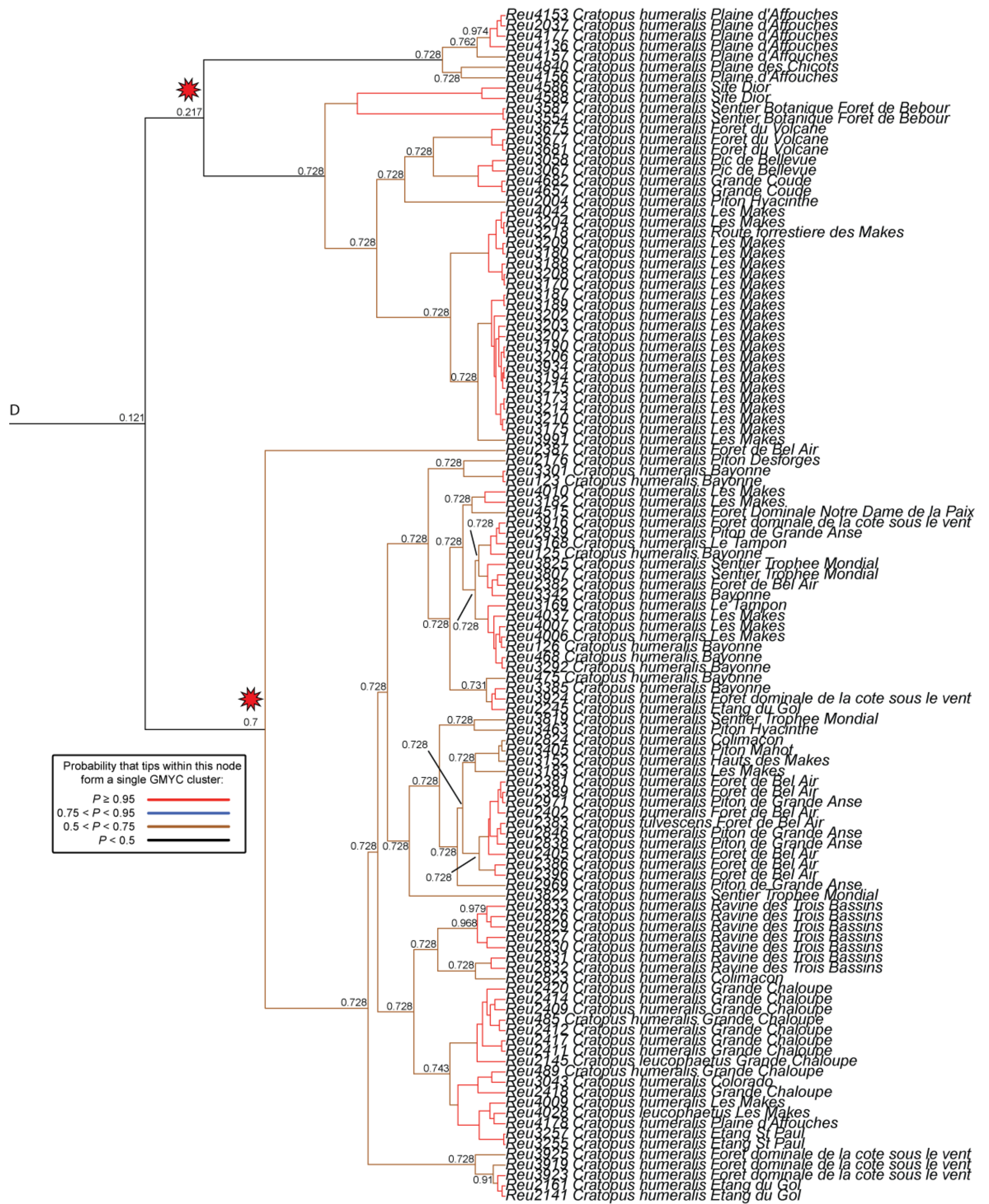


Fig 4.7: The multimodel GMYC analysis result for lineage D of Fig 4.3a (unique sequences only). Stars represent GMYC clusters from the single threshold GMYC analysis.

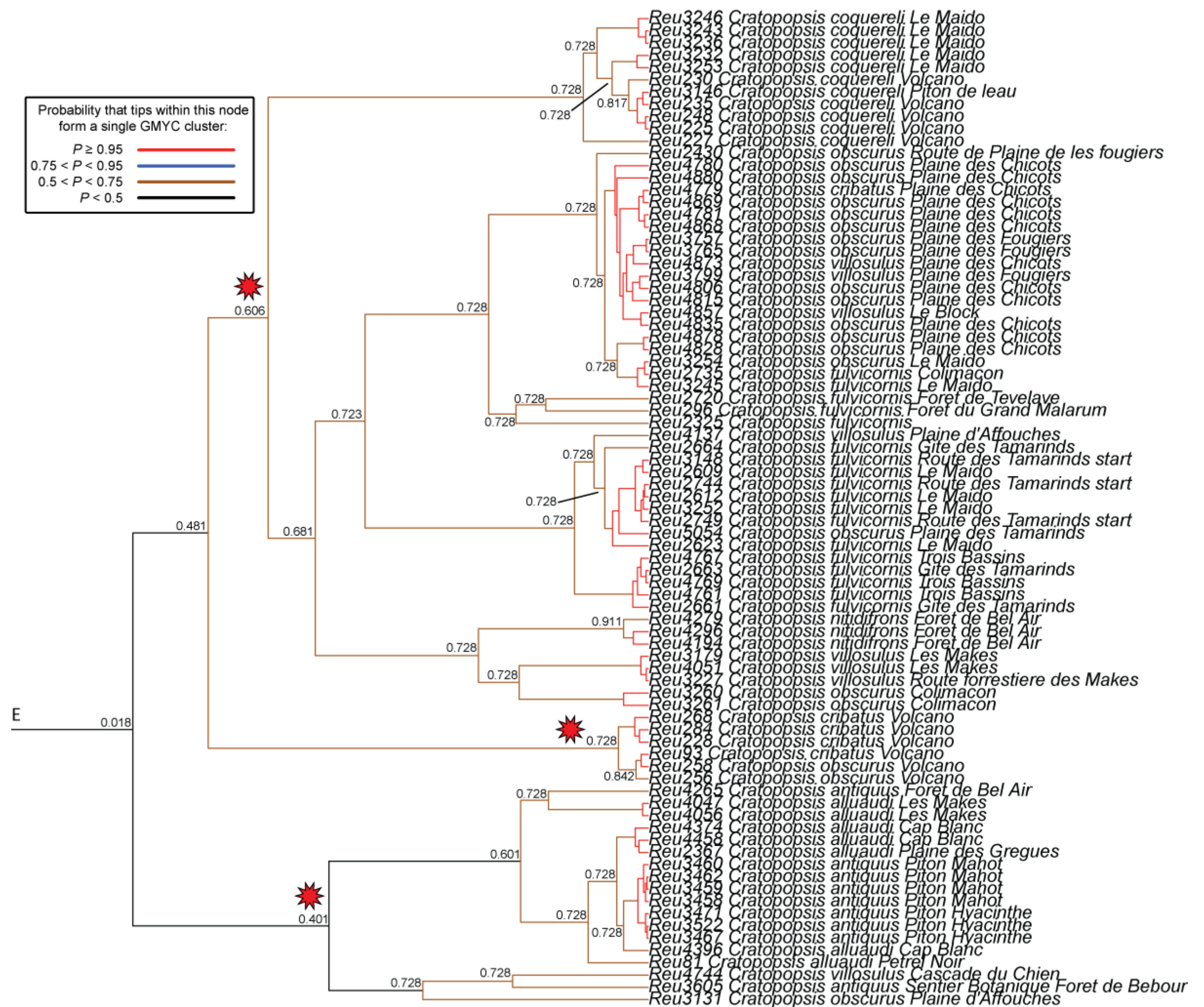


Fig 4.8: The multimodel GMYC analysis result for lineage E of Fig 4.3a (unique sequences only). Stars represent GMYC clusters from the single threshold GMYC analysis.

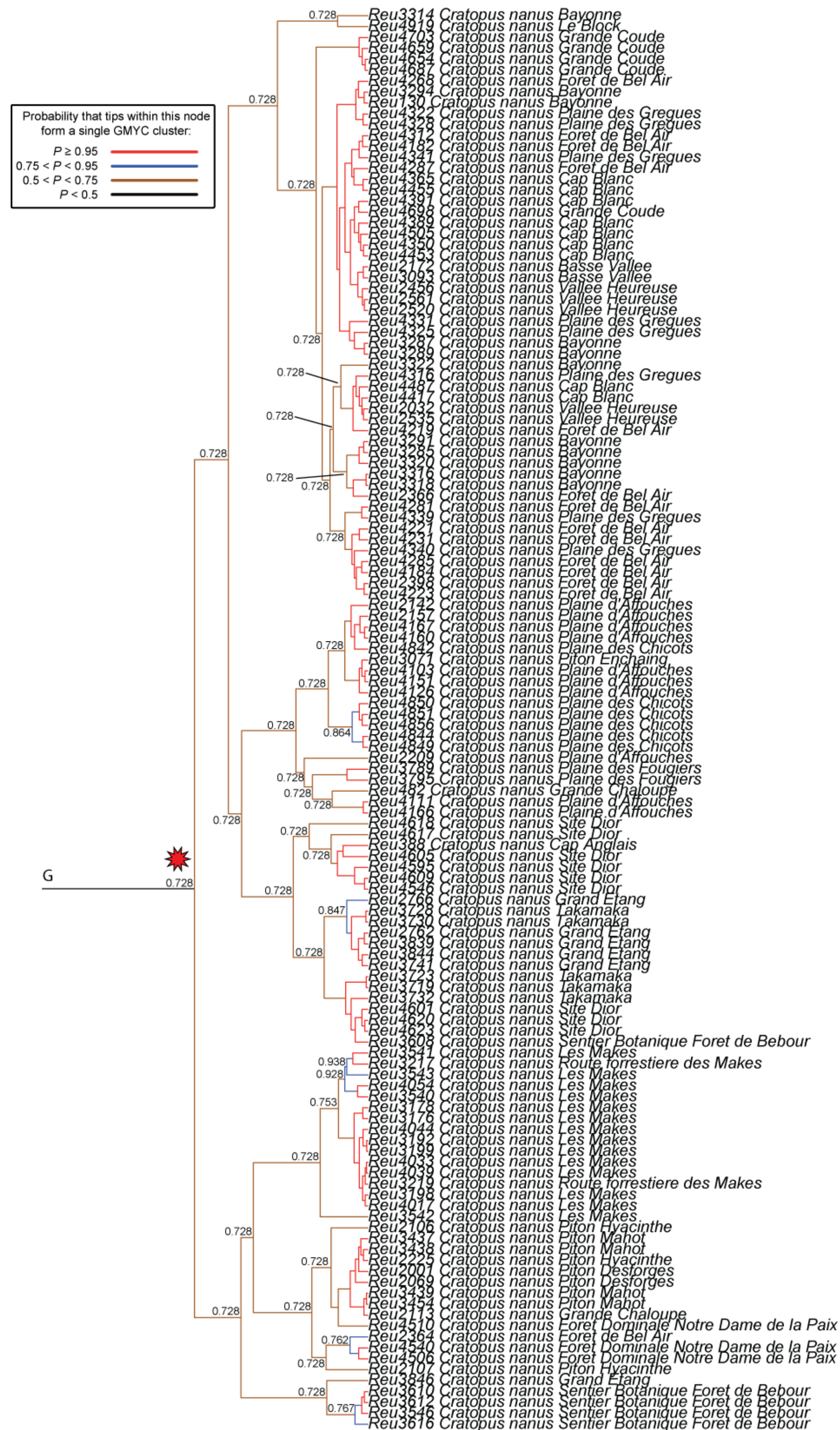


Fig 4.9: The multimodel GMYC analysis result for lineage G of Fig 4.3a (unique sequences only). Stars represent GMYC clusters from the single threshold GMYC analysis.

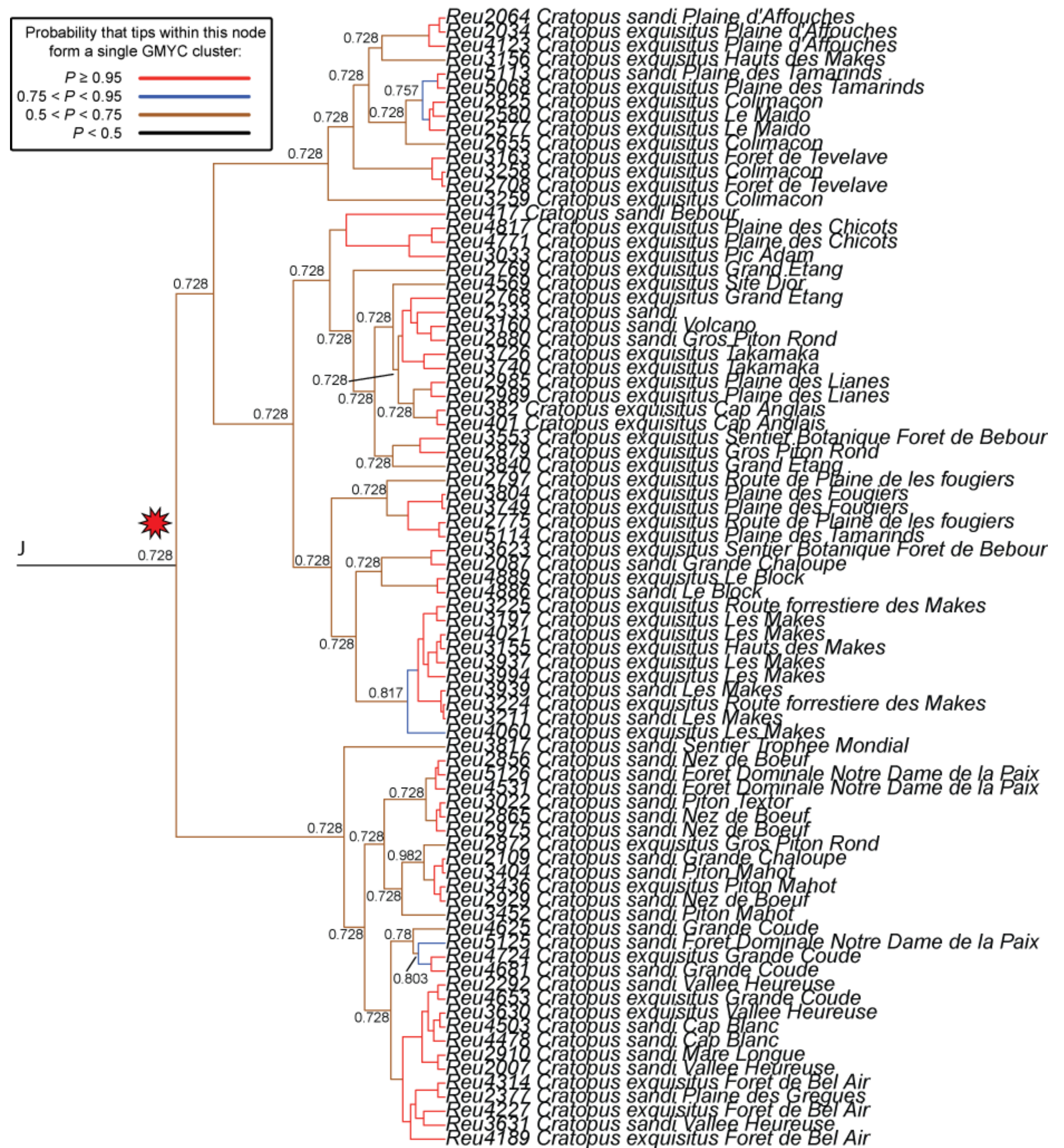


Fig 4.10: The multimodel GMYC analysis result for lineage J of Fig 4.3a (unique sequences only). Stars represent GMYC clusters from the single threshold GMYC analysis.

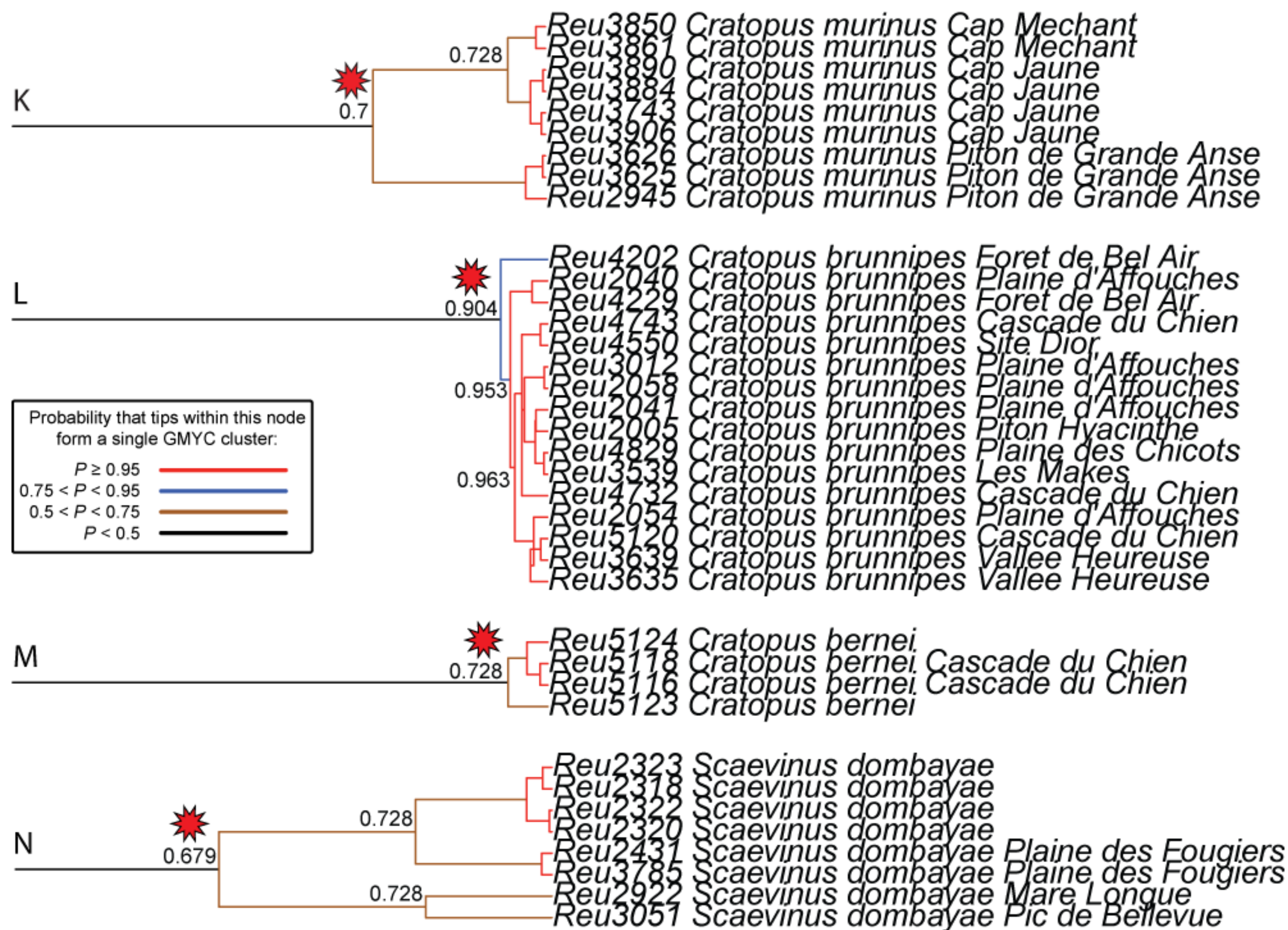


Fig 4.11: The multimodel GMYC analysis result for lineages K, L, M and N of Fig 4.3a (unique sequences only). Stars represent GMYC clusters from the single threshold GMYC analysis.

4.4.4 Population structure

If the multimodel or multiple threshold GMYC approaches are used then there are no groups that contain more than three samples in at least three sites so AMOVA and F_{ST} calculations were based on the GMYC groups estimated using the single threshold analysis. Once sites with fewer than 3 samples and geographically monophyletic GMYC clusters had been excluded, seven lineages were available for testing (A, D, E, G, J, K, and L). Due to the apparent lack of monophyly for morphospecies in lineages A, D, E and J, morphospecies within each lineage were treated as a single entity for the purposes of AMOVA. Additionally, for all AMOVA calculations, sites within 1km of each other were combined. This relates to La Fenetre, Bras Patate and Les Makes. The results are summarised in Table 4.2 and the relative locations of the collecting sites in each AMOVA are presented in Fig 4.12.

Table 4.2: F_{st} and AMOVA results for the tested Reunionaise Cratopine lineages

Lineage	Morphospecies in lineage	Number of sites	F_{ST}	P
A	C. *sumptuosus* *C. frappieri*	17	0.74	< 0.001
D	*C. humeralis* *C. fulvescens* *C. leucophaetus*	9	0.39	< 0.001
E	*Cr. coquereli* *Cr. obscurus* *Cr. villosulus* *Cr. cribatus* *Cr. fulvicornis* *Cr. nitidifrons*	9	0.7	< 0.001
G	*C. nanus*	20	0.69	< 0.001
J	*C. exquisitus* *C. sandi*	15	0.77	< 0.001
K	*C. murinus*	3	0.96	< 0.001
L	*C. brunnipes*	3	0.05	0.24

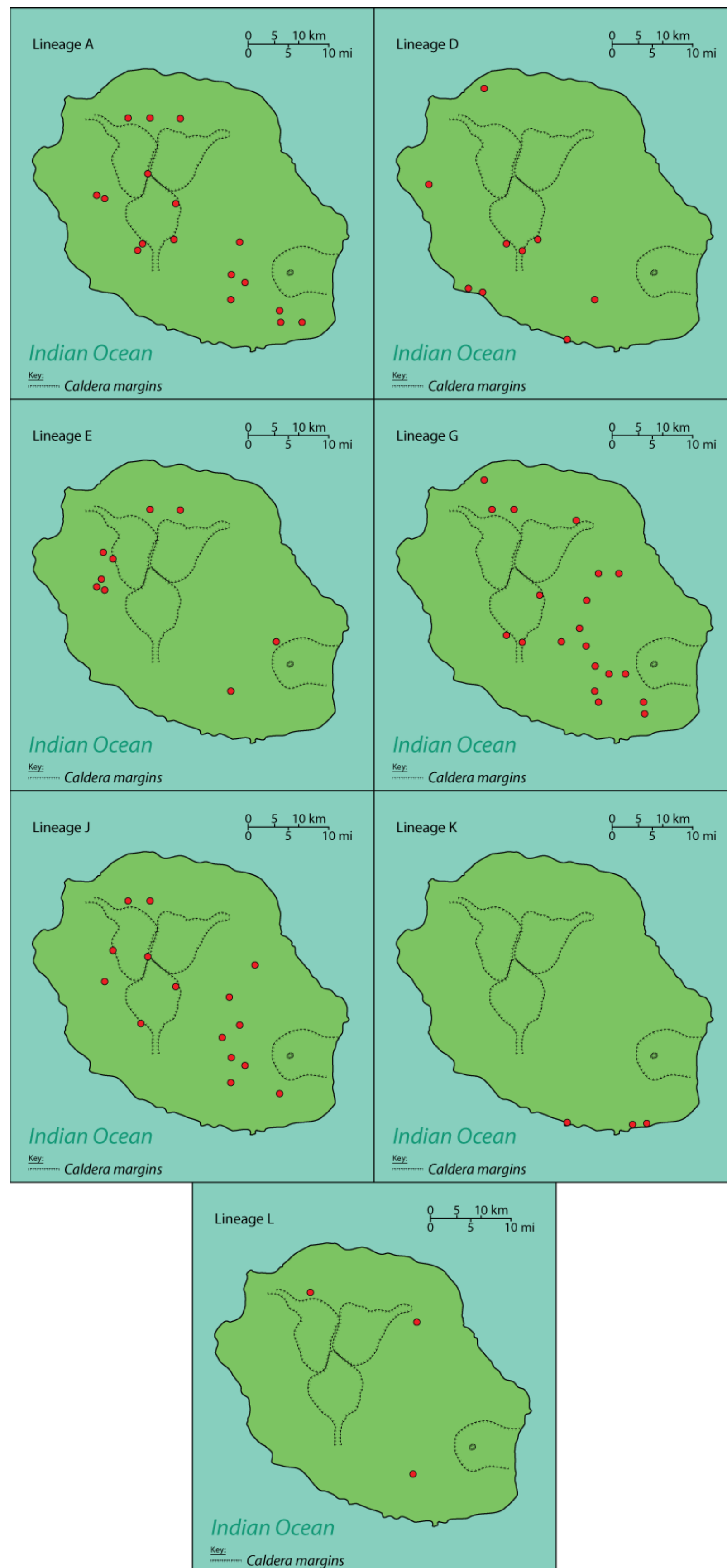


Fig 4.12: Collecting sites included in the AMOVA performed on each lineage.

4.5 Discussion

4.5.1 Phylogenetics of Reunionaise Cratopine weevils

The MrBayes analysis revealed a well-supported tree (Fig 4.3a) with fifteen strongly supported lineages that varied in the number of morphospecies they contain. Nine of the lineages (lineages C, F, G, H, I, K, L, M and N) contain a single morphospecies and as such these morphospecies are likely to represent distinct evolutionary units that equate to species. The remaining five lineages contain more than one morphospecies. Lineages A (*C. sumptuosus* and *C. frappieri*), B (*C. nigridorsis*, *C. circumcinctus* and *C. marmoreus*), D (*C. humeralis*, *C. fulvescens* and *C. leucophaetus*) and J (*C. exquisitus* and *C. sandi*) are all comparable in that each lineage contains morphospecies among which intermediate forms are identifiable. Molecular data suggest that while each of these four lineages are distinct, the morphospecies within each lineage represent forms within a morphological continuum and are consequently unlikely to represent to distinct evolutionary units.

Lineage E contains eight morphospecies that belong to the genus *Cratopopsis* (see Fig 4.3b). There is considerable substructure within this lineage and at least one of the morphospecies (*Cr. coquereli*) forms a monophyletic unit. The remaining sublineages comprise varying combinations of morphospecies suggesting that morphospecies for most *Cratopopsis* species on La Réunion are not a reliable way of selecting distinct evolutionary units. Future analyses involving Reunionaise *Cratopopsis* should take this into account and perhaps use the sublineages of lineage E as evolutionary units for analysis rather than relying on morphospecies. Future work to quantify morphological variation across sublineages and examine the relationships of sublineages using nuclear loci would be useful for defining distinct evolutionary units.

The original descriptions of many of the *Cratopus* and *Cratopopsis* morphospecies found on La Réunion (see Hustache 1919; Hustache 1920) are at best fragmentary and often contradictory, as such it is unsurprising that morphospecies diversity presents discrepancies with patterns of relationship inferred by mtDNA sequence data. The Reunionaise Cratopine morphospecies are currently under review by J. Possereau and J-F Voisin (see Voisin & Poussereau

2007a, b, 2009) and the data presented here will be useful for informing the taxonomy of this group.

4.5.2 Divergence times and colonisation of La Réunion

Previous studies suggest it is not uncommon for island radiations to be younger than the island they inhabit (e.g. Contreras-Diaz *et al.* 2007; Monaghan *et al.* 2006; Price & Clague 2002; Rees *et al.* 2001). However, groups such as the *Galapaganus* weevils of the Galapagos Islands (Sequeira 2000) suggest that this may not always be the case. Indeed, mtDNA sequence data obtained from the Cratopine diversity on La Réunion is consistent with the root age of Reunionaise Cratopine diversity being older than the island itself, see Fig 4.4. This is possible if the early diversification of Cratopine weevils took place on a different island (or set of islands) with multiple subsequent colonisations of La Réunion. Multiple colonisations such as this are also well documented (Gillespie 2004; Jordal *et al.* 2006; Page *et al.* 2007) and should be considered when conflict between the root age of a group and island age arises. In this study all of the defined lineages have estimated coalescent ages less than the 5Ma age estimate of Bonneville *et al.* (1988) and Gillot *et al.* (1994) while many of the divergences between lineages are estimated as older than this. This again argues for the early diversification of the group elsewhere followed by multiple colonisations of La Réunion and subsequent *in situ* radiation. At least six nodes in Fig 4.4 have HPD intervals that include 5Ma. This makes the precise number of colonisations difficult to determine and would be better addressed by adding taxa from potential source populations such as Mauritius into a multilocus analysis that covers the geographic range of the group.

4.5.3 GMYC analyses

The two GMYC approaches applied provide very different estimates of the number of GMYC clusters present in the dataset. The single threshold approach produces estimates of GMYC cluster boundaries that closely approximate the lineages defined in Fig 4.3a, only lineages D and E are partitioned into more than one cluster. Lineage D contains two GMYC clusters with no obvious geographic or morphological basis while lineage E contains three GMYC clusters that are congruent with the major groupings of the morphological variation discussed

previously and presented in Fig 4.3b. Further analyses with additional specimens and nuclear loci would be necessary to evaluate the validity of these clusters as genuine species and refute alternative hypotheses such as complex demographic histories that deviate from coalescent expectations.

The multiple threshold approach produces a very large number of GMYC entities such that GMYC clusters are never more than a few individuals. A likelihood ratio test revealed that the more complex model was not a better fit to the data given the extra parameters required to fit the model. The large discrepancy between the single and multiple threshold analyses means that the multimodel averaging approach of Powel (2012) is less appropriate for summarising this data set than it was for chapter 3, where the results were much more similar across the two GMYC approaches. The reasons for this are likely to stem from the models used during tree building. Coalescent models may not be appropriate for datasets with very divergent sequences as it might be expected that only branching events close to the tips of the tree conform to coalescent assumptions. Similarly, it has been noted in previous studies that Yule models are also inappropriate as the trees produced have branch lengths that deviate from the assumptions in the GMYC model causing an overestimation of species richness (Monaghan *et al.* 2009; Williams *et al.* 2011). Here, a coalescent model was applied and yet still produced a multiple threshold result that seems implausibly high. There may be a number of reasons for this. Firstly, topological constraints were enforced to simplify the analysis. This may result in an elongation of tip branches and an overestimation of GMYC entities. Secondly, it has been suggested that sampling error can result in an over estimation of the number of GMYC clusters present (Lohse 2009). Incomplete sampling cannot be discounted in this study and may be a factor in overestimation of GMYC entity number. It should be noted that Papadopoulou *et al.* (2009) disagree with Lohse *et al.* and argue that while their point is valid, it is overstated and does not diminish the utility of the GMYC approach to provide “an initial hypothesis of the number and extent of species-level groups, given a particular sampling regime and an assumed process of lineage branching” (Papadopoulou *et al.* 2009 page 3).

4.5.4 Population structure

Seven lineages were available for an assessment of population structure using AMOVA, six of which (lineages A, D, E, G, J, and K) revealed significant population structure. Lineage L did not have any significant structure despite the sites available for sampling being distributed widely. Lineages A, D, G, and J are all flight capable so it might be expected that their potential dispersal ability would erode population structure. While studies of flight capable weevils in continental systems have revealed population structure (e.g. Aoki *et al.* 2009; Toju & Sota 2006) this is typically over much larger geographical scales with some species such as the boll weevil (*Anthonomus grandis*) having low F_{ST} values over distances of up to 300 miles (Kim & Sappington 2004) or in the case of *Rhinusa antirrhini*, virtually no structure at all across its western European range (Hernandez-Vera *et al.* 2010). The population structure observed within some species of *Cratopus* suggests that flight is not associated with dispersal over distances such as those between sampling sites, but has been retained for other purposes, perhaps escape from predators or movement between locally patchy food resources. This phenomenon has been noted previously and insects that are fully winged yet apparently flightless have been the subject of previous investigations examining groups as diverse as Rhopalid bugs, water beetles and weevils (e.g. Carroll *et al.* 2003; Jackson 1933; Jackson 1956a, b). Population structure over relatively small distances has been observed in stoneflies capable of flight but with reduced flight musculature (McCulloch 2009). While discussing possible reasons for this, McCulloch (2009) suggested that many of the benefits of flight loss postulated by Harrison (1980) and Roff (1990) can be obtained without complete loss of flight by simply reducing the amount highly metabolically active flight musculature (Zera & Denno 1997), this may also be the case in the fully winged species of *Cratopus*. Examining population structure in the context either wing morphology or flight musculature across multiple populations and lineages could provide further insight into how non-flighted species evolve on islands.

Population structure in lineages E and K is to be expected as both these lineages are flightless. Lineage K (*C. murinus*) is of particular note as it is hypothesised to have a single foodplant with a restricted and patchy distribution on the south coast of La Réunion (see chapters 2 and 3). The population structure seen in lineage K suggests that the habitats for this lineage are stable over

evolutionary time as organisms that exploit patchy or ephemeral resources frequently have excellent dispersal ability that erodes population structure (e.g. carrion beetles Ikeda & Sota 2011).

4.6 Conclusions

Overall the evolution of Cratopine weevils on La Réunion appears to be complex. The root age of the Cratopine diversity of La Réunion is estimated to be older than the island itself. This suggests that the initial radiation of the group occurred elsewhere with multiple colonisation of La Réunion once it became subaerial. The single threshold GMYC analyses suggested that only the two oldest Reunionaise lineages contain more than one GMYC cluster, and phylogenetic analyses indicated that nine morphospecies were monophyletic and might be appropriate units for future analyses examining the biogeography of the group as a whole. The remaining morphospecies were paraphyletic and were recovered in monophyletic clades with multiple morphospecies. Within monophyletic lineages composed of more than one morphospecies, the morphospecies are variations of the same basic morphology and the morphospecies themselves are likely to be inappropriate units for future analyses but this would benefit from examination with nuclear markers. The exception to this is the Reunionaise *Cratopopsis* in which sublineages are visually distinct but share taxonomic characters suggesting that the characters are insufficient to describe the variation present.

As has been noted among Mauritian Cratopine weevils, the presence of population structure was common among the lineages tested. This further argues that the dispersal ability of *Cratopus* is limited and makes the geographic range of the genus and the potential number of colonisation events on La Réunion even more impressive. The precise reasons for a generally reduced tendency to disperse are unknown but the fact that there is variation across lineages both in the presence of wings and the extent of population structure, argues that lineages may be experiencing different selection pressures with respect to the trade-off between the costs and benefits of flight.

4.7 References

Aoki K, Kato M, Murakami N (2009) Phylogeographical patterns of a generalist acorn weevil: insight into the biogeographical history of broadleaved deciduous and evergreen forests. *BMC Evolutionary Biology* **9**, 103.

Bensasson D, Zhang D-X, Hartl DL, Hewitt GM (2001) Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends in Ecology & Evolution* **16**, 314-321.

Bonneville A, Barriot JP, Bayer R (1988) Evidence From Geoid Data of a Hotspot Origin for the Southern Mascarene Plateau and Mascarene Islands (Indian Ocean). *J. Geophys. Res.* **93**, 4199-4212.

Carroll SP, Marler M, Winchell R, Dingle H (2003) Evolution of Cryptic Flight Morph and Life History Differences During Host Race Radiation in the Soapberry Bug, Jadera haematoloma Herrich-Schaeffer (Hemiptera: Rhopalidae). *Annals of the Entomological Society of America* **96**, 135-143.

Cegelski CC, Waits LP, Anderson NJ*, et al.* (2006) Genetic diversity and population structure of wolverine (*Gulo gulo*) populations at the southern edge of their current distribution in North America with implications for genetic viability. *Conservation Genetics* **7**, 197-211.

Cicconardi F, Nardi F, Emerson BC, Frati F, Fanciulli PP (2010) Deep phylogeographic divisions and long-term persistence of forest invertebrates (Hexapoda: Collembola) in the North-Western Mediterranean basin. *Molecular Ecology* **19**, 386-400.

Clement WL, Tebbitt MC, Forrest LL*, et al.* (2004) Phylogenetic position and biogeography of Hillebrandia sandwicensis (Begoniaceae): A rare Hawaiian relict. *American Journal of Botany* **91**, 905-917.

Contreras-Diaz HG, Moya O, Oromi P, Juan C (2007) Evolution and diversification of the forest and hypogean ground-beetle genus Trechus in the Canary Islands. *Molecular Phylogenetics and Evolution* **42**, 687-699.

de Thoisy B, da Silva AG, Ruiz-Garcia M*, et al.* (2010) Population history, phylogeography, and conservation genetics of the last Neotropical mega-herbivore, the lowland tapir (Tapirus terrestris). *BMC Evolutionary Biology* **10**.

Drummond A, Ashton B, Buxton S*, et al.* (2012) Geneious v5.6, Available from http://www.geneious.com/.

Drummond AJ, Ho SYW, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biol* **4**, e88.

du Buy HG, Riley FL (1967) Hybridization between the nuclear and kinetoplast DNA's of Leishmania enriettii and between nuclear and mitochondrial DNA's of mouse liver. *Proc Natl Acad Sci USA* **57**, 790-797.

Emerson BC, Oromi P (2005) Diversification of the forest beetle genus Tarphius on the Canary Islands, and the evolutionary origins of island endemics. *Evolution* **59**, 586-598.

Emerson BC, Oromi P, Hewitt GM (2000) Colonization and diversification of the species *Brachyderes rugatus* (Coleoptera) on the Canary Islands: Evidence from mitochondrial DNA COII gene sequences. *Evolution* **54**, 911-923.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**, 564-567.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of Molecular Variance inferred from metrix distances among haplotypes - application to human mitochondrial DNA restriction data. *Genetics* **131**, 479-491.

Ferragu M, Richard R (1990) Trois nouveaux Curculionides de Madagascar et de l'Île de la Réunion [Coleoptera]. *Revue Française d'Entomologie (N.S.)* **12**, 105-110.

Ferragu M, Richard R (1993) Etude portant sur les Cratopus de l'Ile de la Réunion [Coleoptera, Curculionidae]. *Revue Française d'Entomologie (N.S.)* **14**, 111-116.

Fukami T, Beaumont HJE, Zhang XX, Rainey PB (2007) Immigration history controls diversification in experimental adaptive radiation. *Nature* **446**, 436-439.

Gillespie R (2004) Community assembly through adaptive radiation in Hawaiian spiders. *Science* **303**, 356-359.

Gillespie RG, Claridge EM, Goodacre SL (2008) Biogeography of the fauna of French Polynesia: diversification within and between a series of hot spot archipelagos. *Philosophical Transactions of the Royal Society B-Biological Sciences* **363**, 3335-3346.

Gillot P-Y, Lefèvre J-C, Nativel P-E (1994) Model for the structural evolution of the volcanoes of Réunion Island. *Earth and Planetary Science Letters* **122**, 291-302.

Grobler GC, Bastos ADS, Chimimba CT, Chown SL (2011) Inter-island dispersal of flightless *Bothrometopus huntleyi* (Coleoptera: Curculionidae) from the sub-Antarctic Prince Edward Island archipelago. *Antarctic Science* **FirstView**, 1-10.

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

Harrison RG (1980) Dispersal Polymorphisms in Insects. *Annual Review of Ecology and Systematics* **11**, 95-118.

Hernandez-Vera G, Mitrovic M, Jovic J, *et al.* (2010) Host-associated genetic differentiation in a seed parasitic weevil *Rhinusa antirrhini* (Coleptera: Curculionidae) revealed by mitochondrial and nuclear sequence data. *Molecular Ecology* **19**, 2286-2300.

Holland BS, Hadfield MG (2002) Islands within an island: phylogeography and conservation genetics of the endangered Hawaiian tree snail *Achatinella mustelina*. *Molecular Ecology* **11**, 365-375.

Huelsenbeck JP, Ronquist F (2001) MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754-755.

Hustache A (1919) Synopsis des Curculionides de la Faune Malgache. I. Brachydérides et Otiorrhynchides. *Annales De La Societe Entomologique De France* **87**, 441-520.

Hustache A (1920) Curculionides des Iles Mascareignes. *Extrait des Annales de la Societe Entomologique de France* **89**.

Ikeda H, Sota T (2011) Macroscale evolutionary patterns of flight muscle dimorphism in the carrion beetle Necrophila japonica. *Ecology and evolution* **1**, 97-105.

Illera JC, Emerson BC, Richardson DS (2007) Population history of Berthelot's pipit: colonization, gene flow and morphological divergence in Macaronesia. *Molecular Ecology* **16**, 4599-4612.

Jackson DJ (1933) Observations on the flight muscles of *Sitona* weevils. *Annals of Applied Biology* **20**, 731-770.

Jackson DJ (1956a) The capacity for flight of certain water beetles and its bearing on their origin in the western scottish isles. *Proceedings of the Linnean Society of London* **167**, 76-96.

Jackson DJ (1956b) Observations on flying and flightless water beetles. *Journal of the Linnean Society of London, Zoology* **43**, 18-42.

Jordal BH, Emerson BC, Hewitt GM (2006) Apparent 'sympatric' speciation in ecologically similar herbivorous beetles facilitated by multiple colonizations of an island. *Molecular Ecology* **15**, 2935-2947.

Jordal BH, Kirkendall LR, Harkestad K (2004) Phylogeny of a Macaronesian radiation: host-plant use and possible cryptic speciation in *Liparthrum* bark beetles. *Molecular Phylogenetics and Evolution* **31**, 554-571.

Jordan S, Simon C, Foote D, Englund RA (2005) Phylogeographic patterns of Hawaiian Megalagrion damselflies (Odonata : Coenagrionidae) correlate with Pleistocene island boundaries. *Molecular Ecology* **14**, 3457-3470.

Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* **30**, 3059-3066.

Kim KS, Sappington TW (2004) Boll Weevil (Anthonomus grandis Boheman) (Coleoptera: Curculionidae) Dispersal in the Southern United States: Evidence from Mitochondrial DNA Variation. *Environmental Entomology* **33**, 457-470.

Lohse K (2009) Can mtDNA Barcodes Be Used to Delimit Species? A Response to Pons et al. (2006). *Systematic Biology* **58**, 439-441.

Magnacca KN, Danforth BN (2006) Evolution and biogeography of native Hawaiian Hylaeus bees (Hymenoptera : Colletidae). *Cladistics* **22**, 393-411.

McCulloch GA, Wallis GP, Waters JM (2009) Do insects lose flight before they lose their wings? Population genetic structure in subalpine stoneflies. *Molecular Ecology* **18**, 4073-4087.

McDougal I (1971) Geochronology and evolution of the young volcanic island of Reunion, Indian Ocean. *Geochimica Et Cosmochimica Acta* **35**, 261-&.

Monaghan MT, Balke M, Pons J, Vogler AP (2006) Beyond barcodes: complex DNA taxonomy of a south pacific island radiation. *Proceedings of the Royal Society B-Biological Sciences* **273**, 887-893.

Monaghan MT, Wild R, Elliot M*, et al.* (2009) Accelerated Species Inventory on Madagascar Using Coalescent-Based Models of Species Delineation. *Systematic Biology* **58**, 298-311.

Newton AC, Allnutt TR, Gillies ACM, Lowe AJ, Ennos RA (1999) Molecular phylogeography, intraspecific variation and the conservation of tree species. *Trends in Ecology & Evolution* **14**, 140-145.

Page TJ, von Rintelen K, Hughes JM (2007) An island in the stream: Australia's place in the cosmopolitan world of Indo-West Pacific freshwater shrimp (Decapoda: Atyidae: *Caridina*). *Molecular Phylogenetics and Evolution* **43**, 645-659.

Papadopoulou A, Monaghan MT, Barraclough TG, Vogler AP (2009) Sampling Error Does Not Invalidate the Yule-Coalescent Model for Species Delimitation. A Response to Lohse (2009). *Systematic Biology* **58**, 442-444.

Polihronakis M, Caterino MS, Chatzimanolis S (2010) Elucidating the phylogeographic structure among a mosaic of unisexual and bisexual populations of the weevil Geodercodes latipennis (Coleoptera: Curculionidae) in the Transverse Ranges of southern California. *Biological Journal of the Linnean Society* **101**, 935-948.

Pons J, Barraclough TG, Gomez-Zurita J, *et al.* (2006) Sequence-Based Species Delimitation for the DNA Taxonomy of Undescribed Insects. *Systematic Biology* **55**, 595-609.

Posada D (2008) jModelTest: Phylogenetic model averaging. *Molecular Biology and Evolution* **25**, 1253-1256.

Powell JR (2012) Accounting for uncertainty in species delineation during the analysis of environmental DNA sequence data. *Methods in Ecology and Evolution* **3**, 1-11.

Price JP, Clague DA (2002) How old is the Hawaiian biota? Geology and phylogeny suggest recent divergence. *Proceedings of the Royal Society of London Series B-Biological Sciences* **269**, 2429-2435.

R Development Core Team (2010) R: A Language and Environment for Statistical Computing, Vienna, Austria.

Rambaut A (2011) FigTree v1.3.1. In: *http://tree.bio.ed.ac.uk/software/figtree/*

Rambaut A, Drummond A (2007) Tracer v1.5.0. In: *http://beast.bio.ed.ac.uk/Tracer*

Rees DJ, Emerson BC, Oromi P, Hewitt GM (2001) Mitochondrial DNA, ecology and morphology: interpreting the phylogeography of the Nesotes (Coleoptera : Tenebrionidae) of Gran Canaria (Canary Islands). *Molecular Ecology* **10**, 427-434.

Richard R (1957) Coleopteres Curculionidae de La Reunion. *Memoires de l'Institut Sceientifique de Madagascar* **8**.

Richard R (1977) Nouveaux Curculionides des iles Europa at de la Réunion. *Bulletin de la Société entomologique de France* **82**, 125-127.

Richard R (1995a) Description de deux Cratopini nouveaux de la région Malgache [Coleoptera: Curculionidae]. *Revue française d'entomologie* **17**, 63-65.

Richard R (1995b) Note sur l'apparition brusque d'une variation importante de la squamulation de Cratopus humeralis Boheman (Coleoptera, Curculionidae). *Bulletin de la Société entomologique de France* **100**, 303-306.

Roff DA (1990) The evolution of flightlessness in insects. *Ecological Monographs* **60**, 389-421.

Sequeira AS, Lanteri AA, Albelo R, Bhattacharya S, Sijapati M (2008) Colonization history, ecological shifts and diversification in the evolution of endemic Galapagos weevils. *Molecular Ecology* **17**, 1089-1107.

Sequeira AS, Lanteri AA, Scataglini MA, Confalonieri VA, Farrell BD (2000) Are flightless Galapaganus weevils older than the Galapagos Islands they inhabit? *Heredity* **85**, 20-29.

Shapiro B, Rambaut A, Drummond AJ (2006) Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. *Molecular Biology and Evolution* **23**, 7-9.

Spurgin LG, van Oosterhout C, Illera JC, *et al.* (2011) Gene conversion rapidly generates major histocompatibility complex diversity in recently founded bird populations. *Molecular Ecology* **20**, 5213-5225.

Thacker RW, Hadfield MG (2000) Mitochondrial phylogeny of extant Hawaiian tree snails (Achatinellinae). *Molecular Phylogenetics and Evolution* **16**, 263-270.

Toju H, Sota T (2006) Phylogeography and the geographic cline in the armament of a seed-predatory weevil: effects of historical events vs. natural selection from the host plant. *Molecular Ecology* **15**, 4161-4173.

Vinson J (1967) Liste Chorologique des Coleopteres des Mascareignes. *The Mauritius Institute Bulletin* **4**, 299-357.

Voisin J-F, Poussereau J (2007a) Charactérisation de Cratopus frappieri Deyrolle, 1862 (Coleoptera, Curculionidae). *Nouv. Revue Ent (N.S.)* **24**, 329-334.

Voisin J-F, Poussereau J (2007b) L'identité de Cratopus sumptuosus Boheman, 1834, de Cratopus exquisitus Boheman, 1840, et de Cratopus sandi Deyrolle, 1862, espèces endémiques de l'île de la Réunion (Coleoptera, Curculionidae). *Nouv. Revue Ent (N.S.)* **24**, 279-283.

Voisin J-F, Poussereau J (2009) A propos de Cratopus nanus et de Cratopus parvus de l'île de la Réunion (Coleoptera, Curculionidae). *Nouv. Revue Ent. (N.S.)* **26**, 59-65.

Williams S, Apte D, Ozawa T, Kaligis F, Nakano T (2011) Speciation and dispersal along continental coastlines and island arcs in the Indo-West Pacific turbinid gastropod genus Lunella. *Evolution* **65**, 1752-1771.

Xia X, Xie Z (2001) DAMBE: Software package for data analysis in molecular biology and evolution. *Journal of Heredity* **92**, 371-373.

Xia XH, Xie Z, Salemi M, Chen L, Wang Y (2003) An index of substitution saturation and its application. *Molecular Phylogenetics and Evolution* **26**, 1-7.

Zera AJ, Denno RF (1997) Physiology and ecology of dispersal polymorphism in insects. *Annual Review of Entomology* **42**, 207-230.

Chapter 5:
The phylogenetics and biogeography of the Cratopine weevils of the South-West Indian Ocean area (Curculionidae: Entiminae: Cratopini)

Scaevinus dombayae: Plaine des Fougères, La Réunion 2009.

5.1 Abstract

Invertebrate radiations within the South-West Indian Ocean islands are poorly studied in comparison to those of other archipelagos such as Hawaii and the Canary Islands. Here a multi-locus phylogenetic analysis was performed with estimated node ages for a group of phytophagous beetles that have radiated across the region. Results indicate a complex biogeographic history that can (with some exceptions) typically be reconciled with island ages. Much of the lineage diversity is the product of repeated colonisations out of Mauritius to La Réunion, Rodrigues and the coralline islets of the Seychelles, Aldabra and the Iles Eparses. The evolution of flightlessness is inferred to have occurred at least five times, with evidence for mitochondrial introgression within one lineage following colonisation of a new island.

5.2 Introduction

Oceanic island systems are an attractive system for the study of evolution for a number of reasons. Oceanic islands are discrete identities with easily defined boundaries. Barriers to gene flow are more easily characterised as these islands are separated by inhospitable expanses of ocean that are often continuously present over evolutionary timescales. Islands often contain a diverse range of habitats and in many cases are still geologically dynamic places with ongoing volcanic and erosional processes. Finally, island archipelagos frequently host impressive species radiations that are ideal for DNA sequence based molecular examination of patterns and processes at interspecific scales. To date, much attention has been focused on the invertebrate taxa of the Hawaiian Islands (e.g. Nitta & O'Grady 2008; O'Grady *et al.* 2011; Shaw 2002; Thacker & Hadfield 2000), Macaronesian Islands (e.g. Contreras-Diaz *et al.* 2007; Emerson *et al.* 2006; Emerson *et al.* 1999; Machado *et al.* 2008; Rees *et al.* 2001) and the Galapagos Islands (e.g. Sequeira *et al.* 2008; Sequeira *et al.* 2012) with phylogenetic techniques being used to address questions related to the timing of divergence events (e.g. Magnacca & Danforth 2006; Sequeira *et al.* 2000), patterns of diversification (e.g. Holland & Hadfield 2002; Jordal *et al.* 2004; Jordan *et al.* 2005)

and the resolution of taxonomic uncertainty (Clement *et al.* 2004; Jordal *et al.* 2006).

To date, only two published studies have examined the invertebrate radiations of the Indian Ocean islands using molecular methods. Both focus on highly dispersive spiders (Kuntner & Agnarsson 2011a; Kuntner & Agnarsson 2011b) which may not be representative of other invertebrate taxa. The remaining studies concentrate on vertebrate radiations such as *Phelsuma* day geckos (Austin *et al.* 2004), chameleons (Raxworthy *et al.* 2002) and birds (Warren *et al.* 2003; Warren *et al.* 2005, 2006). This is surprising as the number and variability of Indian Ocean islands provides an ideal setting in which to address questions concerning island biogeography.

The South-West Indian ocean islands can be broadly divided into groups based on age and geology. Madagascar and the granitic Seychelles are the remains of continental blocks left after the breakup of Gondwana 64-130mya (million years ago) (Coffin & Rabinowitz 1987; Kingdon 1990; Rabinowitz *et al.* 1983). Volcanic islands such as the Comoros and Mascarene archipelagos are the next oldest. The islands of the Comoros get progressively younger as you move east suggesting that they were formed by classic volcanic hotspot activity over the last 10 - 15 million years (Emerick & Duncan 1982). The maximum estimated ages available for Comorian islands are 10-15 mya (Mayotte), 11.5 mya (Anjouan) and 0.5 mya (Grande Comore) (Montaggioni & Nougier 1981; Nougier *et al.* 1986) with radiometric dating of the oldest lavas producing slightly younger ages of 7.7 ± 1.0 mya (Mayotte), 3.9 ± 0.3 mya (Anjouan), 0.13 ± 0.02 mya (Grande Comore) and 5.0 ± 0.4 mya (Moheli) (Emerick & Duncan 1982; Nougier *et al.* 1986). Similar to the Comores, the Mascarene Islands are of volcanic origin but rather than being formed in sequence by oceanic crust moving over a volcanic hotspot, each island is separated from the rest by fracture zones which redirect flowing magma causing it to emerge in places and at times we might not expect. This means the islands have evolved independently of one another (McDougal 1971; McDougal & Chamalaun 1969). Radiometric dating of the oldest lavas on each island produces ages of 8.9 ± 0.17 mya (Mauritius (Moore *et al.* 2011)), 2.0 ± 0.05 mya (La Réunion (McDougal 1971)) and 1.5 ± 0.05 mya (Rodrigues (McDougall *et al.* 1965)) respectively. However it should be noted that many of the exposed lavas in the Mascarenes are the result of volcanic reactivation with multiple phases of island building and

erosion having occurred on each island (Montaggioni & Nativel 1988), as such earlier, older lavas may be unsampled or even destroyed. The result of this is that even accurate radiometric dating may underestimate the ages of islands, for example there is indirect evidence that La Réunion may be up to 5 mya (Bonneville *et al.* 1988). The youngest archipelagos in the region are the raised coralline islets of the Aldabra group and many low lying islands including Europa, Juan de Nova and Iles Glorieuses. Europa is estimated to have emerged within the last 80,000 years (Battistini 1966). All of these islands are low lying and sea levels over the last 500,000 years are suggested to have repeatedly been higher than current levels (for a review of sealevel data see Warren *et al.* 2010) with direct evidence of the inundation of Aldabra and Iles Glorieuses 125,000 years ago (Battistini & Cremer 1972; Thomson & Walton 1972) that is likely to have eliminated the terrestrial fauna at that time. As a consequence, the fauna of these islands may be much younger than other islands in the area. Taken together, the islands of this region are an attractive setting for studies that relate divergence times and patterns of speciation to island ages. Figure 1.1 shows the relative positions and ages of the islands studied.

One of the most speciose invertebrate groups across the South-West Indian Ocean islands is the Cratopine weevils. The tribe Cratopini consists of nine genera (Alonso-Zarazaga & Lyal 1999) but only three are commonly found on the Islands of the South-West Indian Ocean (*Cratopus*, *Cratopopsis* and *Scaevinus*). With approximately 100 morphospecies described, the genus *Cratopus* is the largest of the three genera while *Cratopopsis* and *Scaevinus* have 14 and 2 morphospecies respectively. Previous work in this thesis (chapters two, three and four) has shown that populations of *Cratopus* are often highly structured despite being flight capable and has suggested that many described morphospecies may in fact be components of highly variable lineages which are better candidates for species status than each morphospecies separately. There is also evidence that species within the genus *Cratopopsis* may be phylogenetically nested within *Cratopus,* raising questions about its taxonomic validity. To date studies published on *Cratopus*, *Cratopopsis* and *Scaevinus* (*Cratopopsis* hereafter abbreviated to *Cr.*) have largely dealt with describing their morphology and defining species based on this (see Ferragu & Richard 1990, 1993; Galman *et al.* 2011; Hustache 1919; Hustache 1920; Poussereau & Voisin 2009; Richard 1957; Richard 1958, 1961, 1977;

Richard 1995a, b; Voisin & Poussereau 2007a, b, 2009; Williams & Cox 2003), but other than some occasional commentary within these publications regarding morphological similarities, no attempt has been made to examine the relationships across the three genera as a whole, particularly regarding biogeographic relationships. The aim of this study is to generate a multi-locus phylogeny of the three Cratopine genera that inhabit the islands of the South-West Indian Ocean. Additionally the ages of colonisation events will be estimated using phylogenetic methods and molecular clocks.

5.3 Methods

5.3.1 Beetle sampling

Beetles were collected from the islands listed in Table 5.1 by foliage beating during the wet season between June 2007 and December 2011. Samples were placed directly in 99% ethanol solution, with the exception of a few individuals that were kept alive until the same evening for photographing, before being placed in 99% ethanol. Sampling sites were recorded on handheld GPS units. A total of 5565 beetles were collected across 19 islands that contain the majority of the described Cratopine diversity. Samples were identified using morphological keys for *Cratopus* and *Cratopopsis* provided by Jaques Poussereau (*pers. comm.*) and the original descriptions of most species (Ferragu & Richard 1990, 1993; Galman *et al.* 2011; Hustache 1919; Hustache 1920; Poussereau & Voisin 2009; Richard 1957; Richard 1958, 1961, 1977; Richard 1995a, b; Voisin & Poussereau 2007a, b, 2009; Williams & Cox 2003). Within each island a single individual of each morphospecies was selected rather than one individual per lineage defined in chapters three and four. This allowed congruence between mitochondrial and nuclear loci to be tested across morphospecies. If morphospecies group identically to chapters three and four this would further suggest that morphospecies are not appropriate units of diversity for this lineage. Reunionaise *Cratopopsis* (excluding *Cr bistigma*) were selected based on sublineages of lineage E in chapter four rather than morphospecies as this is likely to better reflect the diversity present than morphospecies. Finally, specimens of *C. murinus* and *C. punctum* from La Réunion were not included as chapter three strongly suggested that these are nested within the Mauritian *C. murinus* and *C. punctum* lineages and neither displays any morphological differences across islands. This gives a final data set of 77 individuals.

Table 5.1: Morphospecies included in this study with island and island group information.

Island	**Morphospecies**	**Island group**
La Réunion	*Cratopopsis bistigma*	Mascarenes
La Réunion	*Cratopopsis E1*	Mascarenes
La Réunion	*Cratopopsis E2*	Mascarenes
La Réunion	*Cratopopsis E3*	Mascarenes
La Réunion	*Cratopopsis E4*	Mascarenes
La Réunion	*Cratopopsis E5*	Mascarenes
La Réunion	*Cratopopsis E6*	Mascarenes
La Réunion	*Cratopopsis E7*	Mascarenes
La Réunion	*Cratopus bernei*	Mascarenes
La Réunion	*Cratopus brunnipes*	Mascarenes
La Réunion	*Cratopus circumcinctus*	Mascarenes
La Réunion	*Cratopus ditissimus*	Mascarenes
La Réunion	*Cratopus exquisitus*	Mascarenes
La Réunion	*Cratopus frappieri*	Mascarenes
La Réunion	*Cratopus fulvescens*	Mascarenes
La Réunion	*Cratopus humeralis*	Mascarenes
La Réunion	*Cratopus leucophaetus*	Mascarenes
La Réunion	*Cratopus marmoreus*	Mascarenes
La Réunion	*Cratopus. murinus*	Mascarenes
La Réunion	*Cratopus nigridorsis*	Mascarenes
La Réunion	*Cratopus. sandi*	Mascarenes
La Réunion	*Cratopus septemvittatus*	Mascarenes
La Réunion	*Cratopus sumptuosus*	Mascarenes
La Réunion	*Cratopus tristis*	Mascarenes
La Réunion	*Scaevinus dombayae*	Mascarenes

Island	**Morphospecies**	**Island group**
Mauritius	*Cratopopsis mauritianus*	Mascarenes
Mauritius	*Cratopus aeneoniger*	Mascarenes
Mauritius	*Cratopus armatus*	Mascarenes
Mauritius	*Cratopus caliginosus*	Mascarenes
Mauritius	*Cratopus cariei*	Mascarenes
Mauritius	*Cratopus confusus*	Mascarenes
Mauritius	*Cratopus deceptus*	Mascarenes
Mauritius	*Cratopus emarginatus*	Mascarenes
Mauritius	*Cratopus fasciger*	Mascarenes
Mauritius	*Cratopus melanocephalus*	Mascarenes
Mauritius	*Cratopus molitor*	Mascarenes
Mauritius	*Cratopus mundulus*	Mascarenes
Mauritius	*Cratopus murinus*	Mascarenes
Mauritius	*Cratopus nigrogranatus*	Mascarenes
Mauritius	*Cratopus nothus*	Mascarenes
Mauritius	*Cratopus ovalis*	Mascarenes
Mauritius	*Cratopus psittacus*	Mascarenes
Mauritius	*Cratopus punctum*	Mascarenes
Mauritius	*Cratopus stigmaeus*	Mascarenes
Mauritius	*Cratopus striga*	Mascarenes
Mauritius	*Cratopus variegatus*	Mascarenes
Mauritius	*Cratopus viridilimbatus*	Mascarenes
Mauritius	*Cratopus viridulus*	Mascarenes
Mauritius	*Cratopus vulgaris*	Mascarenes
Mauritius	*Scaevinus subtruncatus*	Mascarenes

Island	Morphospecies	**Island group**
Rodrigues	*Cratopopsis pauliani*	Mascarenes
Rodrigues	*Cratopus inornatus*	Mascarenes
Rodrigues	*Cratopus rocki*	Mascarenes
Rodrigues	*Cratopus virescens*	Mascarenes
Rodrigues	*Cratopus viridipunctatus*	Mascarenes
Aldabra	*Cratopus griseovestitus*	Seychelles
Aldabra	*Cratopus viridisparsus*	Seychelles
Aride Island	*Cratopus griseovestitus*	Seychelles
Cousin Island	*Cratopus griseovestitus*	Seychelles
Cousine Island	*Cratopus griseovestitus*	Seychelles
La Digue	*Cratopus aurostriatus*	Seychelles
La Digue	*Cratopus Dig_sp1*	Seychelles
La Digue	*Cratopus griseovestitus*	Seychelles
Mahe	*Cratopus Mah sp1*	Seychelles
Praslin	*Cratopus segregatus*	Seychelles
Silhouette	*Cratopus griseovestitus*	Seychelles
Silhouette	*Cratopus segregatus*	Seychelles
Anjouan	*Cratopus AnjCom sp1*	Comoros
Anjouan	*Cratopus AnjCom sp2*	Comoros
Grande Comore	*Cratopus GCom sp1*	Comoros
Grande Comore	*Cratopus subdenudatus*	Comoros
Mayotte	*Cratopus ditissimus*	Comoros
Moheli	*Cratopus MohCom sp1*	Comoros
Europa	*Cratopus viridisparsus*	Iles Eparse
Grande Glorieuse	*Cratopus gloriosus*	Iles Eparse
Juan de Nova	*Cratopus adspersus*	Iles Eparse
Madagascar	*Polyclaeis equestris*	Madagascar

5.3.2 DNA extraction, PCR amplification and sequencing

DNA was extracted from the head and pronotum using the DNeasy 96 well Blood and Tissue Extraction kit (QIAGEN, West Sussex, UK) with the digestion buffer volumes amended for large specimens as recommended by the manufacturer. The primers COIICraF (5'-TAATATGGCAGAWTAGTGCAATGGA-3') and COIICraR (5'-TGCTTTCAGTCATCTAATGATCTRTTTACAGA-3') were used to amplify the mitochondrial gene Cytochrome Oxidase II (COII). Amplification conditions were; 0.5 mM of each primer, 5 mM $MgCl_2$ and a thermal profile of: 95°C for 60s, 58°C for 60s and 72°C for 90s, 40 cycles. . For Argenine Kinase (ArgK) a nested PCR approach was used. The first PCR used the primers ArgKforB2 (5' GAYTCCGGWATYGGWATCTAYGCTCC 3') and ArgKrevB1 (5' TCNGTRAGRCCCATWCGTCTC 3') (both attributed to Danforth *et al.* (2005) in McKenna *et al.* (2009)) with 0.5 mM of each primer, 5 mM $MgCl_2$ and a thermal profile of: 95°C for 60s, 50°C for 60s and 72°C for 120s, 37 cycles. The second PCR was performed using the PCR product from the first PCR as a template. This PCR used the primers ArgK_F1_semidg (5'-GATCCCATCATHGARGAYTARCA-3') and ArgK_R2_fulldg (5'-GTNCCYAARTTNGTNGGRCARAA-3') (both designed by C. Gillett, University of East Anglia) with 0.5 mM of each primer, 4.3 mM $MgCl_2$ and a thermal profile of: 95°C for 60s, 57°C for 60s and 72°C for 120s, 37 cycles. For Histone 3 a nested PCR approach was also used. The first PCR used the primers M13REVH3AF (5'-CAGGAAACAGCTATGACCATGGCTCGTACCAAGCAGACVGC-3') and M13(-21)H3AR (5'-TGTAAAACGACGGCCAGTATATCCTTRGGCATRATRGTGAC-3'), these are the primers listed in (Colgan *et al.* 1998) but modified with M13 tails to improve PCR yield (Regier & Shi 2005). The thermal profile for Histone 3 PCRs employed a touchdown protocol (Don *et al.* 1991) of 94°C for 60s, 60°C for 60s and 72°C for 120s (touched down by 1°C per cycle), 5 cycles, followed by 35 cycles with an annealing temperature of 55°C with 0.5 mM of each primer, 5 mM $MgCl_2$ and. The PCR product from the first PCR was separated and the band of approximately 330bp was sampled using a pipette tip. This was soaked in 10 µl of H_2O to release the PCR product. The resulting solution was used as the template for a second PCR with the primers H3AF (5'-ATGGCTCGTACCAAGCAGACVGC-3') and H3AR (5'-ATATCCTTRGGCATRATRGTGAC-3'). The thermal profile was identical to the first PCR. Finally 28S was amplified using the primers 28SDD (5'-GGGACCCGTCTTGAAACAC-3') and 28SFF (5'-TTACACACTCCTTAGCGGAT-3')

(Ahrens *et al.* 2007) with 0.5 mM of each primer, 2 mM $MgCl_2$ and a thermal profile of: 94°C for 60s, 50°C for 30s and 72°C for 30s, 31 cycles.

Sequencing reactions were performed with the Big Dye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems, California, USA). COIICraF was used for forward sequencing and COIICraR was used for reverse sequencing when forward sequences of less than 600bp were produced. Due to its short fragment size, Histone 3 was sequenced in both directions using H3AF and H3AR. ArgK was sequenced with ArgK_F1_semidg and 28S was sequenced with 28SDD. When reverse sequencing was employed, the consensus of each pair of forward and reverse sequences was generated in Geneious Pro version 5.6 (Drummond *et al.* 2012). The thermal profile used for COII sequencing reactions was: 96°C for 10s, 58°C for 5s and 60°C for 240s, 25 cycles while all other reactions had an annealing temperature of 50°C. Sequences were read on a 3730XL sequencer (Applied Biosystems). All sequences were checked in Geneious Pro version 5.4. Sequences were aligned using MAFFT v6.814b (Katoh *et al.* 2002) with the following parameter values: scoring matrix 1PAM/k=2, Gap open penalty = 1.53, Offset value = 0.123, and then checked by eye. The aligned sequences were tested for saturation using the entropy-based index of substitution saturation (Xia *et al.* 2003) as implemented in DAMBE v5.2.78 (Xia & Xie 2001). This was performed on two data sets, one comprised of the first and second codon positions and the second comprised of the third codon positions except for 28S which was not partitioned. Substitution saturation was visualised by plotting uncorrected pairwise p-distances against corrected pairwise distances. A linear regression of the plotted values provides an alternative method to qualitatively assess saturation as x or x^2 coefficients in the fitted model much less than one would indicate a disagreement between modelled and observed genetic distances and that the sequences are possibly saturated.

5.3.3 Phylogenetic analyses

Trees were constructed both from individual alignments of each locus and from a combined dataset using MrBayes 3.2 (Huelsenbeck & Ronquist 2001) with *Polyclaeis equestris* as an outgroup. For MrBayes, four analyses were performed, each for 20 million generations using eight MCMC chains, discarding 25% of the samples as burnin with a model of sequence evolution determined using

jModelTest2 (Guindon & Gascuel 2003; Posada 2008) under the BIC and AIC criteria. All parameters permitted under this model were estimated and in the combined analysis, each gene partition was permitted to have its own estimated values. The output was assessed for stationarity and convergence in Tracer v1.5.0 (Rambaut & Drummond 2007), with only ESS scores greater than 200 being accepted, and the consensus tree was visualised in FigTree v1.3.1 (Rambaut 2011).

5.3.4 Divergence time estimation

To estimate the ages of colonisation events, fifteen replicate BEAST analyses were performed on a combined dataset of all four loci. An uncorrelated relaxed clock (Drummond *et al.* 2006) was used with a Yule speciation tree model. A coleopteran COII mutation rate of 0.0154 substitutions/site/Ma was taken from (Cicconardi *et al.* 2010) and applied to the COII partition with all other partition estimated relative to this. To prevent over parameterisation of the analysis, the HKY substitution model (Hasegawa *et al.* 1985) was used and nodes with posterior probabilities greater than 0.9 in the MrBayes analysis were constrained to be monophyletic in the BEAST analysis. Additionally a normal root height prior of 10.0Ma (sd = 0.5Ma) was applied to restrict the root age of the analysis to estimates that might be expected given the sequence divergences seen in the COII data set. Analyses were run for 150 million generations and convergence, stationarity and node ages were assessed with Tracer v1.5.0. As with the MrBayes analysis, only ESS scores greater than 200 were accepted.

5.4 Results

5.4.1 DNA sequencing

A full summary of sequencing success for each locus can be found in Table 5.2 while Table 5.3 contains the corrected and uncorrected genetic distances for each locus. The sequencing success for each individual split by locus can be seen in Table 5.4.

Table 5.2: Sequencing success and summary statistics for each locus used in the phylogeny.

Locus	Number of samples	% success	Minimum length (bp)	Maximum length (bp)	Alignment length (bp)	Mean length (bp)	sd length (bp)
COII	72	93.5%	607	685	711	663.1	17.3
ArgK	73	94.8%	432	500	500	499.1	7.9
28S	68	88.3%	621	630	638	621.9	2.1
H3	59	76.6%	178	328	328	290	55.7
All samples	77						

Table 5.3: Uncorrected p distances and GTR corrected distances for all loci used in the phylogeny.

	p distance			**GTR corrected distance**		
Locus	**Maximum**	**Mean**	**sd**	**Maximum**	**Mean**	**sd**
COII	0.20	0.15	0.03	0.25	0.18	0.04
ArgK	0.09	0.04	0.02	0.09	0.04	0.02
28S	0.04	0.01	0.01	0.04	0.01	0.01
H3	0.22	0.08	0.04	0.26	0.09	0.04

Table 5.4: Sequencing success split by locus for each morphospecies. Green boxes indicate success and red boxes are failures.

Island	Morphospecies	COII/ArgK/28S/H3			
La Réunion	*Cratopopsis bistigma*	green	green	green	green
La Réunion	*Cratopopsis E1*	green	green	green	green
La Réunion	*Cratopopsis E2*	green	green	green	green
La Réunion	*Cratopopsis E3*	green	green	green	green
La Réunion	*Cratopopsis E4*	green	green	green	green
La Réunion	*Cratopopsis E5*	green	green	green	green
La Réunion	*Cratopopsis E6*	green	green	green	green
La Réunion	*Cratopopsis E7*	green	green	green	green
La Réunion	*Cratopus bernei*	green	green	green	green
La Réunion	*Cratopus brunnipes*	green	green	green	red
La Réunion	*Cratopus circumcinctus*	green	green	green	green
La Réunion	*Cratopus ditissimus*	green	green	green	green
La Réunion	*Cratopus exquisitus*	green	green	green	red
La Réunion	*Cratopus frappieri*	green	green	green	green
La Réunion	*Cratopus fulvescens*	green	green	green	green
La Réunion	*Cratopus humeralis*	green	green	green	green
La Réunion	*Cratopus leucophaetus*	green	green	green	green
La Réunion	*Cratopus marmoreus*	green	green	green	green
La Réunion	*Cratopus nanus*	green	green	green	green
La Réunion	*Cratopus nigridorsis*	green	green	green	green
La Réunion	*Cratopus sandi*	green	green	green	green
La Réunion	*Cratopus septemvittatus*	green	green	red	red
La Réunion	*Cratopus sumptuosus*	green	green	green	green
La Réunion	*Cratopus tristis*	green	green	green	red
La Réunion	*Scaevinus dombayae*	green	green	green	green
Mauritius	*Cratopopsis mauritianus*	green	green	green	green
Mauritius	*Cratopus aeneoniger*	green	green	green	green
Mauritius	*Cratopus armatus*	green	green	green	green
Mauritius	*Cratopus caliginosus*	green	green	green	green
Mauritius	*Cratopus cariei*	green	green	red	green
Mauritius	*Cratopus confusus*	green	green	green	green
Mauritius	*Cratopus deceptus*	green	green	green	green
Mauritius	*Cratopus emarginatus*	green	green	green	green
Mauritius	*Cratopus fasciger*	green	green	green	red
Mauritius	*Cratopus melanocephalus*	green	green	green	red
Mauritius	*Cratopus molitor*	green	red	red	red
Mauritius	*Cratopus mundulus*	green	green	green	green
Mauritius	*Cratopus murinus*	green	green	green	green
Mauritius	*Cratopus nigrogranatus*	green	green	green	red
Mauritius	*Cratopus nothus*	red	green	green	red
Mauritius	*Cratopus ovalis*	green	green	green	red
Mauritius	*Cratopus psittacus*	green	green	green	green
Mauritius	*Cratopus punctum*	green	green	green	green
Mauritius	*Cratopus stigmaeus*	green	green	green	green
Mauritius	*Cratopus striga*	green	green	green	red
Mauritius	*Cratopus variegatus*	green	green	green	red
Mauritius	*Cratopus viridilimbatus*	green	green	green	green
Mauritius	*Cratopus viridulus*	green	green	green	green
Mauritius	*Cratopus vulgaris*	green	green	green	green
Mauritius	*Scaevinus subtruncatus*	green	green	green	green
Rodrigues	*Cratopopsis pauliani*	green	green	green	green
Rodrigues	*Cratopus inornatus*	red	green	green	green
Rodrigues	*Cratopus rocki*	green	green	green	green
Rodrigues	*Cratopus virescens*	green	green	green	green
Rodrigues	*Cratopus viridipunctatus*	green	green	green	green
Aldabra	*Cratopus griseovestitus*	green	green	green	green
Aldabra	*Cratopus viridisparsus*	green	green	green	red
Aride Island	*Cratopus griseovestitus*	green	green	green	green
Cousin Island	*Cratopus griseovestitus*	green	green	green	green
Cousine Island	*Cratopus griseovestitus*	green	green	green	green
La Digue	*Cratopus aurostriatus*	green	green	green	green
La Digue	*Cratopus Dig sp1*	green	green	green	green
La Digue	*Cratopus griseovestitus*	green	green	green	green
Mahe	*Cratopus Mah sp1*	green	red	red	red
Praslin	*Cratopus segregatus*	green	green	green	red
Silhouette	*Cratopus griseovestitus*	green	green	green	green
Silhouette	*Cratopus segregatus*	red	red	red	red
Anjouan	*Cratopus AnjCom sp1*	green	green	green	green
Anjouan	*Cratopus AnjCom sp2*	green	green	red	green
Grande Comore	*Cratopus GCom sp1*	green	green	green	green
Grande Comore	*Cratopus subdenudatus*	green	green	green	green
Mayotte	*Cratopus ditissimus*	red	red	red	red
Moheli	*Cratopus MohCom sp1*	green	green	red	green
Europa	*Cratopus viridisparsus*	green	green	green	green
Grande Glorieuse	*Cratopus gloriosus*	green	green	green	green
Juan de Nova	*Cratopus adspersus*	green	green	red	green
Madagascar	*Polyclaeis equestris*	red	green	green	green

The results of Xia's test for saturation for each locus split into first and second codon position vs. third codon position (except 28S which was not partitioned) can be found in Table 5.5. Overall the results indicated that sequences were not saturated with the possible exception of the third codon position for COII which had a test statistic that was not significantly less than the critical value. This suggests that this codon position for COII may be saturated or nearly so. Plotting uncorrected p-distances against genetic distances corrected using the GTR model provided similar results but the deviation of the GTR corrected distances from the uncorrected p distances was not as high as might be expected for totally saturated sequences. The saturation plots for each locus can be seen in Figs 5.2-5.5. A summary of the linear model coefficients for each locus can be seen in Table 5.6. Overall x coefficients were not dissimilar to one but x^2 coefficients vary suggesting disagreement between fitted and corrected genetic distances. This is especially true for COII and Histone 3. For COII this has already been shown to likely be caused by saturation at the third codon position whereas for Histone 3 Xia's test did not detect saturation in either partition. The discrepancy may be a due to differences in the way that Xia's test and genetic distance matrices in R deal with missing data.

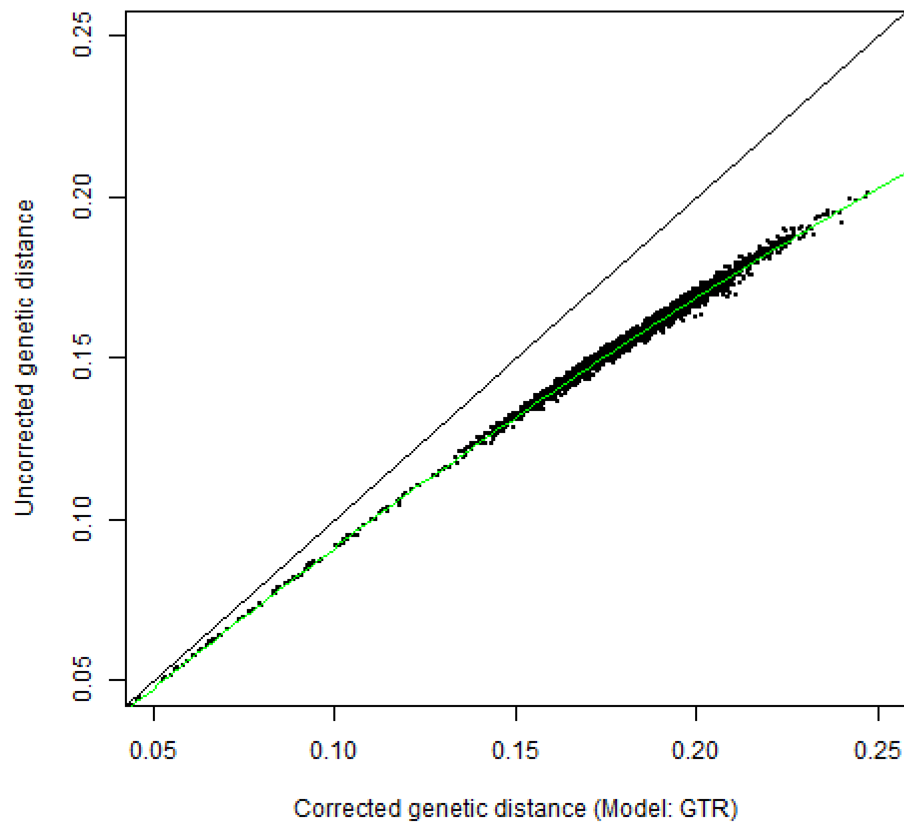


Fig 5.2: uncorrected genetic distances plotted against GTR corrected genetic distances for all sequences in the COII alignment. The black line represents a perfect match between uncorrected and corrected distances and zero saturation whereas the green line represents the linear model fitted to the data.

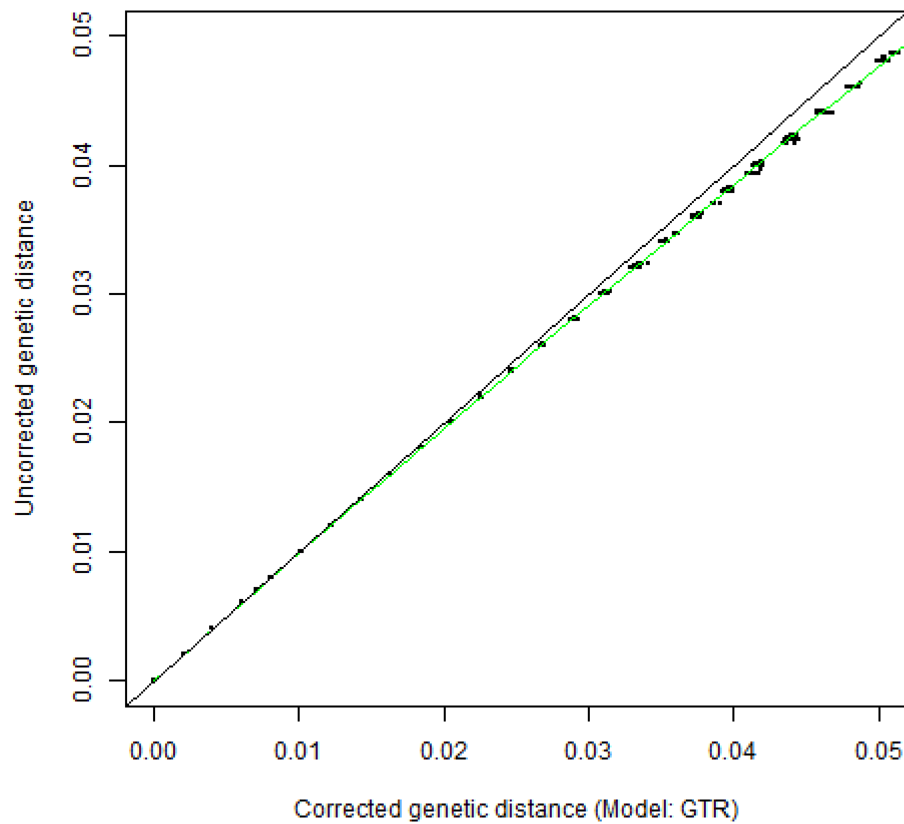


Fig 5.3: uncorrected genetic distances plotted against GTR corrected genetic distances for all sequences in the ArgK alignment. The black line represents a perfect match between uncorrected and corrected distances and zero saturation whereas the green line represents the linear model fitted to the data.

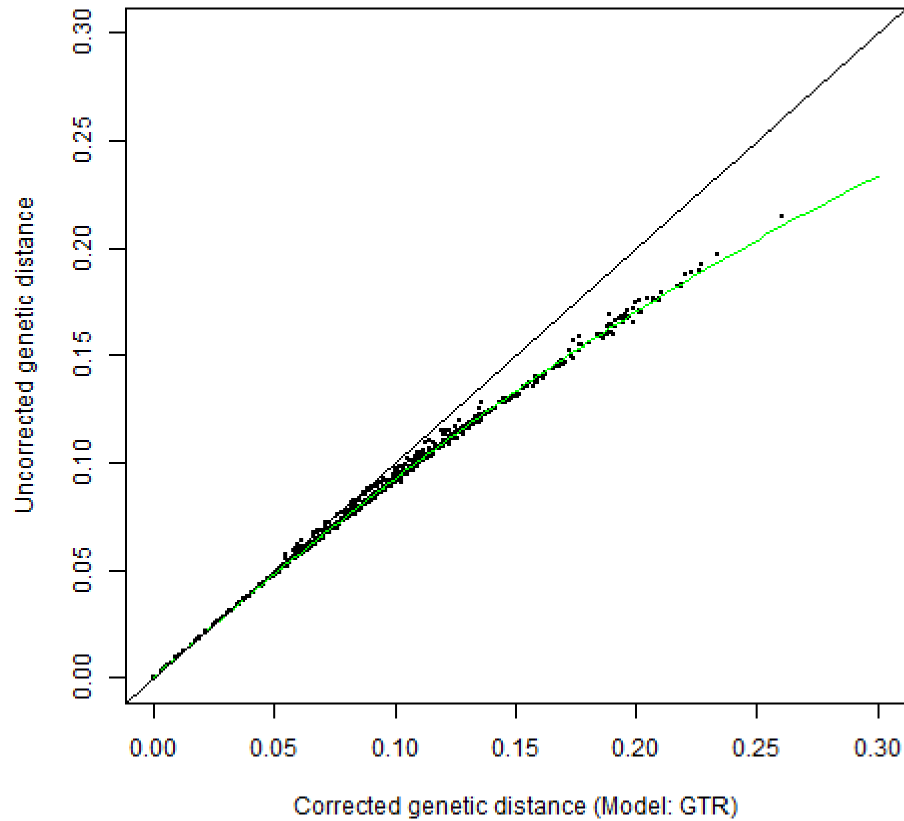


Fig 5.4: uncorrected genetic distances plotted against GTR corrected genetic distances for all sequences in the Histone 3 alignment. The black line represents a perfect match between uncorrected and corrected distances and zero saturation whereas the green line represents the linear model fitted to the data.

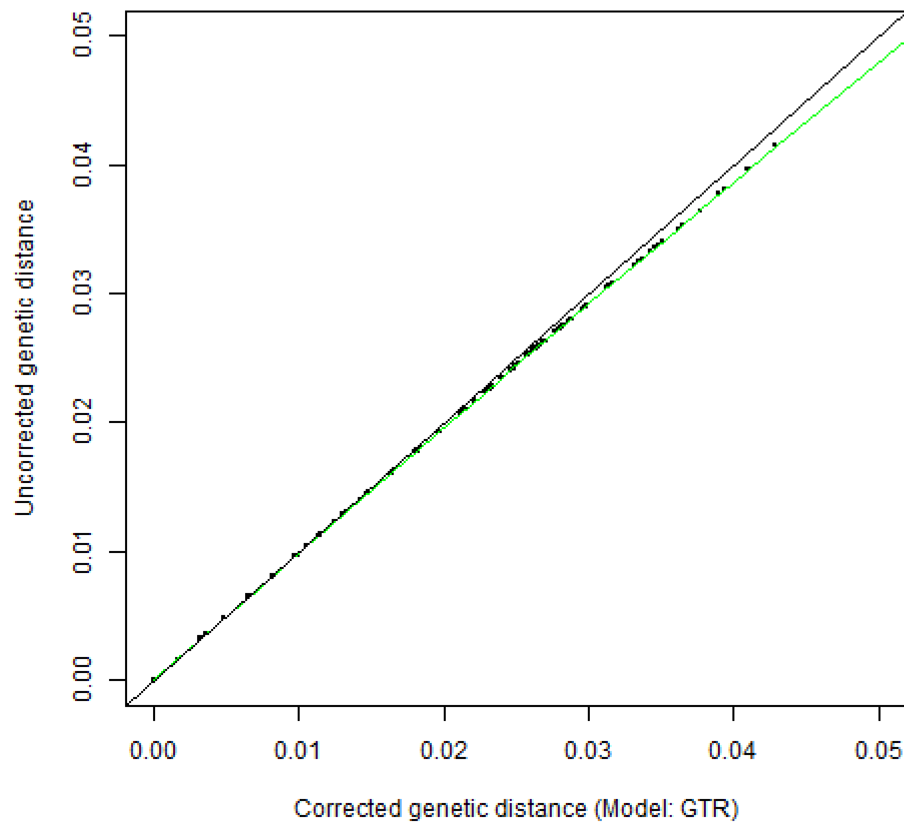


Fig 5.5: uncorrected genetic distances plotted against GTR corrected genetic distances for all sequences in the 28S alignment. The black line represents a perfect match between uncorrected and corrected distances and zero saturation whereas the green line represents the linear model fitted to the data.

Table 5.5: Xia's test for saturation test statistics and critical values for each locus. The test statistic must be significantly less than the critical value to indicate no saturation. Protein coding loci are partitioned into fist and second codon positions vs third codon position.

	Locus						
Locus	**COII**		**ArgK**		**H3**		**28s**
Codon positions	1 & 2	3	1 & 2	3	1 & 2	3	all
Obsverved index	0.075	0.581	0.024	0.147	0.007	0.304	0.033
Critical value	0.692	0.599	0.652	0.541	0.619	0.619	0.734
P	<0.001	0.614	<0.001	<0.001	<0.001	<0.001	<0.001

Table 5.6: Coefficients of the linear models fitted to the saturation plots of uncorrected p distances vs GTR corrected distances for each locus.

Locus	x	*P*	x2	*P*
COII	0.97	<0.001	-0.66	<0.001
ArgK	0.99	<0.001	-0.8	<0.001
28S	0.95	<0.001	-0.71	<0.001
H3	0.99	<0.001	-0.73	<0.001

5.4.2 Phylogenetic analyses

MrBayes analyses were performed using the GTR+G substitution model (COII, ArgK and H3) and HKY+I+G model (28S) returned as the best fit by jModeltest2. The COII (Fig.5.6) and ArgK (Fig.5.7) trees are well supported with ArgK having better resolution towards the base of the tree. The Histone 3 (Fig.5.8) and 28S (Fig.5.9) trees are less well supported. Histone 3 contains much missing data (11.8% of the alignment is missing or ambiguous compared to 10.5% for COII, 0.2% for ArgK and 0.3% for 28S). Weins and Moen (2008) have stated that missing data is not problematic in Bayesian analyses however they add the caveat that this refers to analyses where the number of characters is high and in small data sets this may not be the case. As Histone 3 is a much shorter fragment than COII with fewer overall characters available, we might expect that this locus may not perform as well as longer loci even with similar levels of missing data. Missing data is unlikely to be a problem for the analyses of the multi-locus dataset due to the number of characters available overall. The lack of support in the 28S tree is likely to be due to the low number of variable sites (91.5% of aligned sites are identical compared to 58.9% for COII, 74.4% for ArgK and 60.7% for Histone 3). The individual gene trees are in broad agreement with three notable exceptions. First, *Cr. bistigma* and *Cr. mauritianus* are recovered as sister species in all nuclear loci

but are recovered on separate branches with strong support in the COII tree. The sequences used in the tree for *Cr. mauritianus* and *Cr. bistigma* were previously used in chapters three and four and both grouped strongly with other sequences from the same morphospecies. This suggests that the incongruence between nuclear and mitochondrial loci is not simply the result of sequencing error, because of this; these individuals were excluded from the overall phylogeny. Second, the position of *C. variegatus* in the COII tree is inconsistent with the 28S, ArgK and the COII tree in Fig 3.3 which strongly supported an early divergence of this morphospecies from all other Mauritian species. Additionally this sequence is placed on a long tip branch suggesting that the analysis has inferred a very different mutation rate for this sequence compared to its sister sequence. This is unlikely to be the case and it is more likely that the differences between these two sequences are the result of *C. variegatus* being placed with the incorrect sister sequence. As a result this sequence was not excluded from the overall phylogeny. Finally the position of *C. sandi* in the Histone 3 tree appears inconsistent with all other loci. This sequence is not in a well supported position and contains missing data as such this sequence is not definitely incongruent with other loci and has not been excluded from the overall phylogeny.

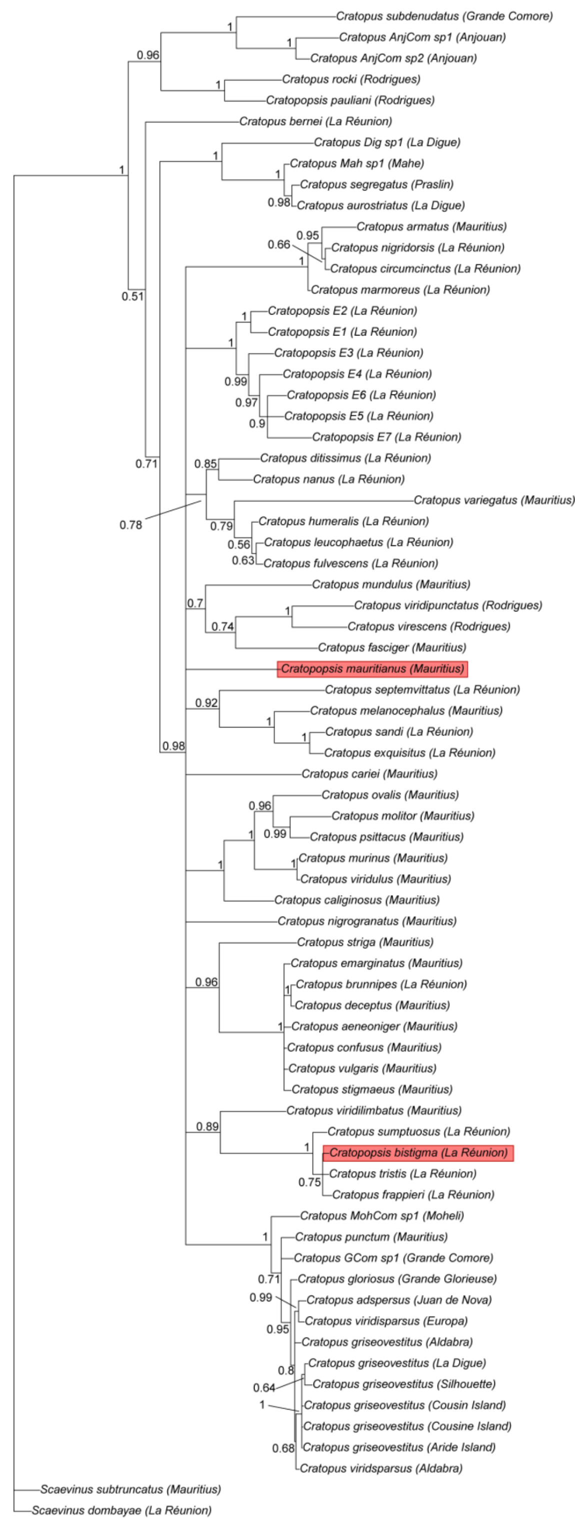


Fig 5.6: A Bayesian inference tree generated by Mr Bayes using an alignment of Cratopine COII sequences. Values placed on nodes represent Bayesian posterior probabilities for that node. *Cratopopsis bistigma* and *Cratopopsis mauritianus* are highlighted to show incongruence between nuclear loci and COII.

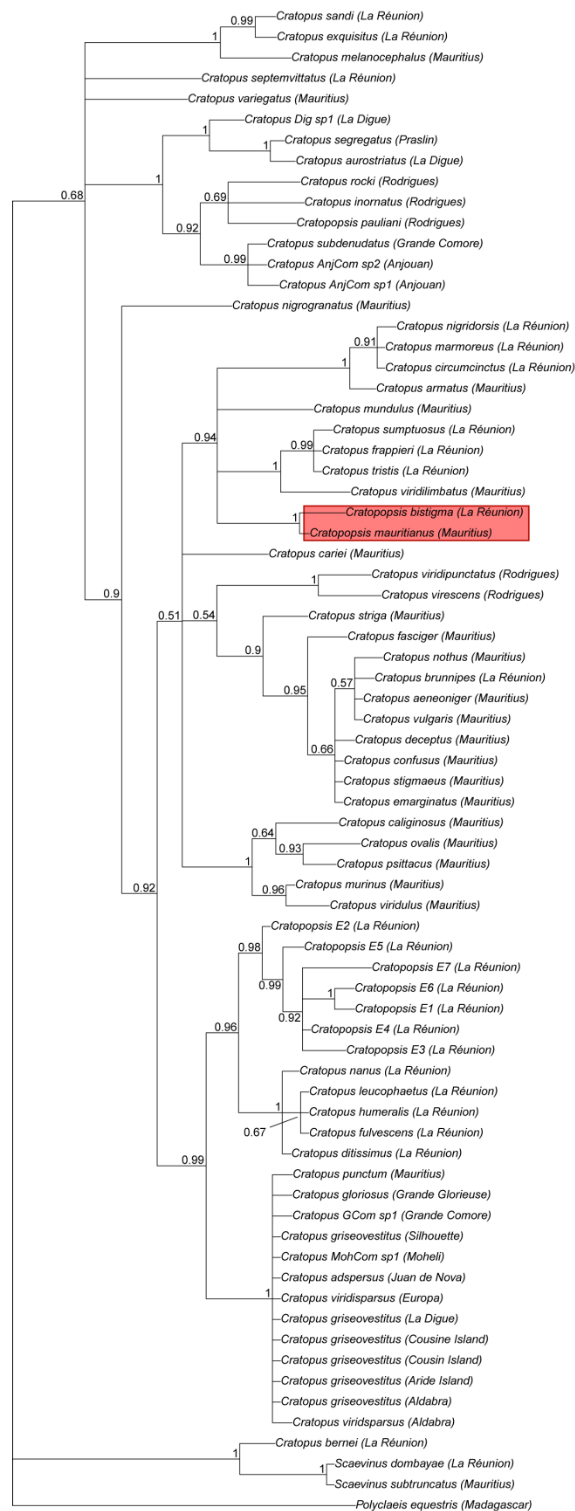


Fig 5.7: A Bayesian inference tree generated by Mr Bayes using an alignment of Cratopine ArgK sequences. Values placed on nodes represent Bayesian posterior probabilities for that node. *Cratopopsis bistigma* and *Cratopopsis mauritianus* are highlighted to show incongruence between nuclear loci and COII

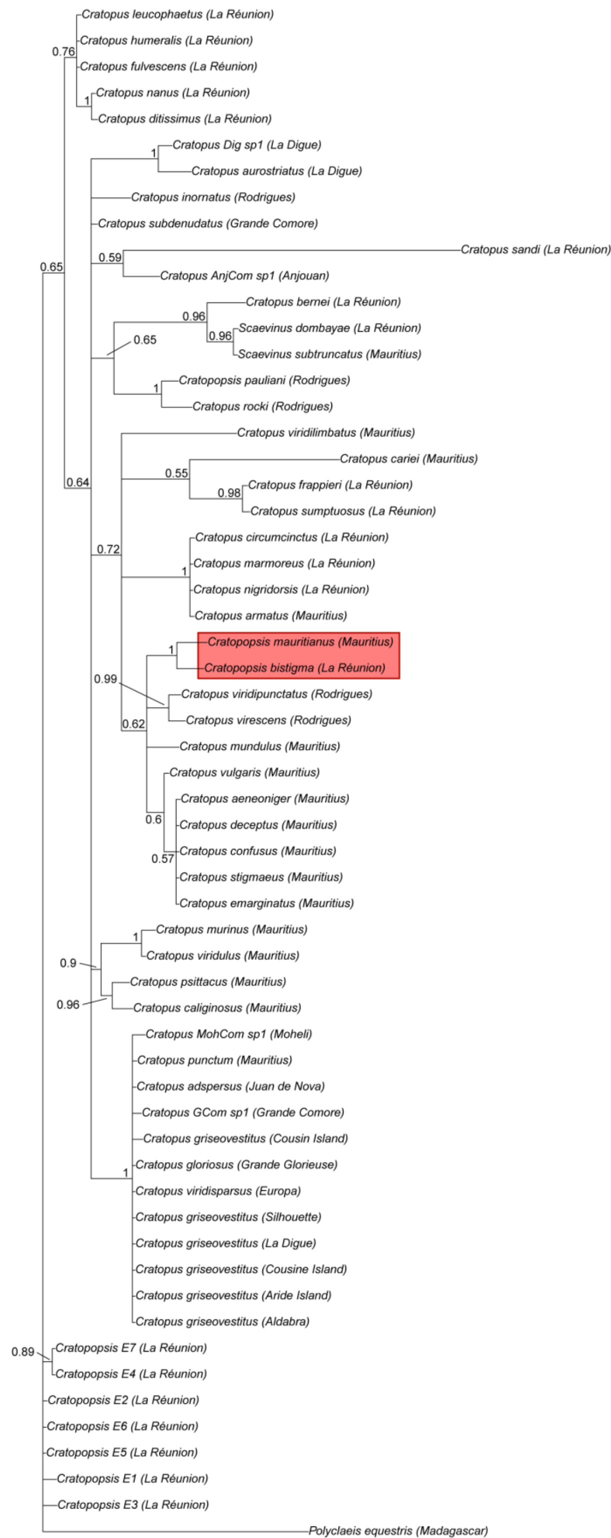


Fig 5.8: A Bayesian inference tree generated by Mr Bayes using an alignment of Cratopine Histone 3 sequences. Values placed on nodes represent Bayesian posterior probabilities for that node. *Cratopopsis bistigma* and *Cratopopsis mauritianus* are highlighted to show incongruence between nuclear loci and COII

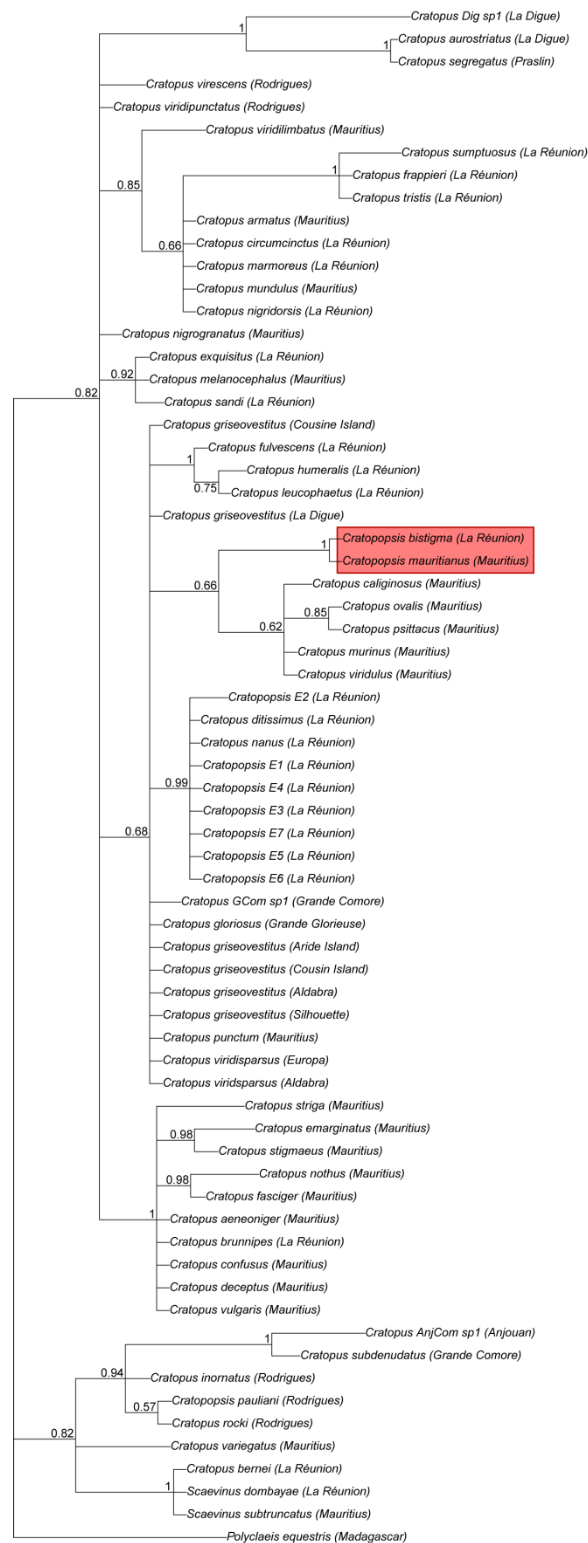


Fig 5.9: A Bayesian inference tree generated by Mr Bayes using an alignment of Cratopine 28S sequences. Values placed on nodes represent Bayesian posterior probabilities for that node. *Cratopopsis bistigma* and *Cratopopsis mauritianus* are highlighted to show incongruence between nuclear loci and COII

The multilocus phylogeny inferred from all four loci contained 73 samples and the same substitution model was used for each partition as in individual

analyses. The concatenated alignment was 2,173 bp long with 10% of sites missing data. The tree produced (Fig 5.10) was well supported with three main Cratopine lineages although the relationships among these lineages was unclear. Lineage 1 consists of morphospecies restricted to the granitic Seychelles (specifically La Digue, Mahe and Praslin), the Comoros (specifically Anjouan and Grande Comore) and Rodrigues. Lineage 2 is the largest and contains morphospecies from all three Mascarene Islands, the coralline Seychelles (specifically Cousin Island and Cousine Island), the granitic Seychelles (specifically La Digue, Silhouette and Aride Island), the Comoros (specifically Moheli and Grande Comore), Aldabra, Europa, Juan de Nova and Grande Glorieuse. The final lineage contains the two morphospecies of *Scaevinus* and *Cratopus bernei*. *Cratopus variegatus* is recovered separately from all other lineages with support values suggesting it may be closest to either lineage one or two. *Cratopus exquisitus* is recovered with strong support in the same lineage as *C. sandi*. For ease of presentation Fig 5.10 has been divided into sections with boundaries that contain all relevant nodes for the colonisation events discussed (see Fig 5.11a and Fig 5.11b). Divergence times of key nodes in the overall phylogeny are given in these figures.

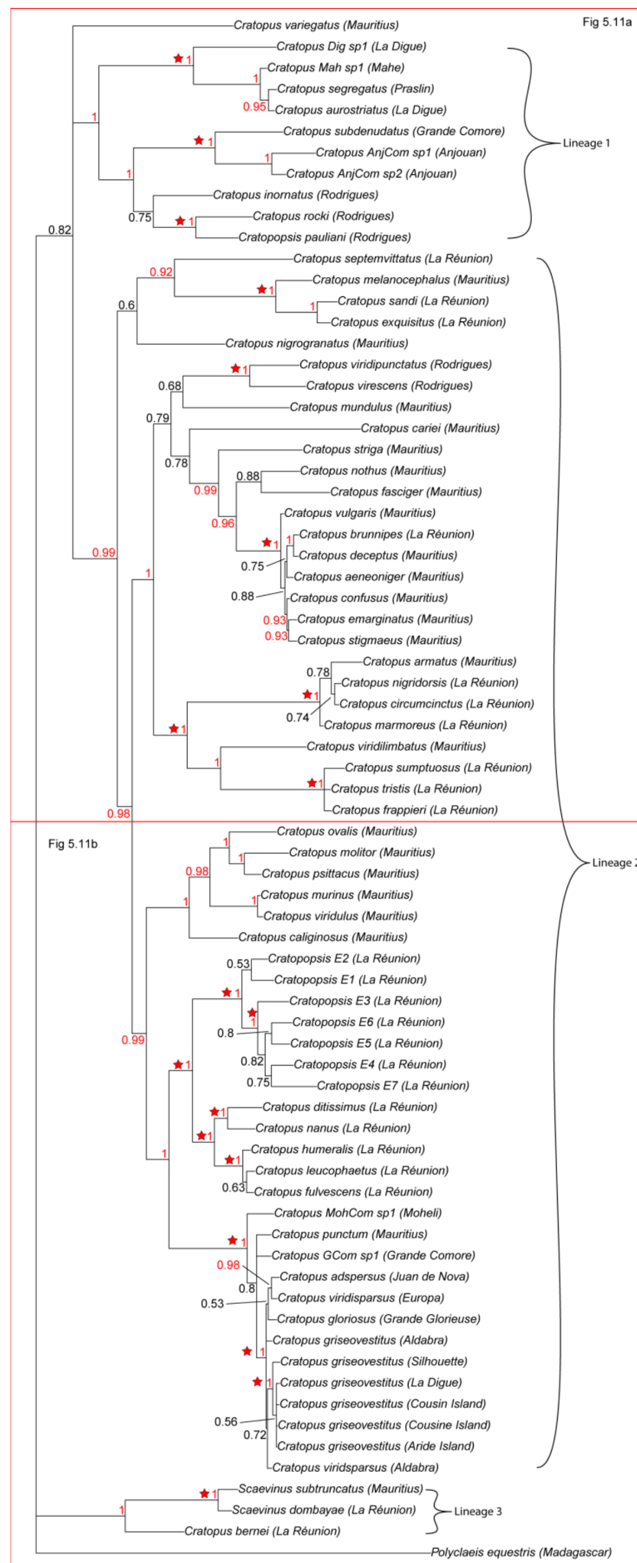


Fig 5.10: A Bayesian inference tree generated by Mr Bayes using an alignment of Cratopine sequences from all four loci. Values placed on nodes represent Bayesian posterior probabilities for that node. Node labels coloured red were constrained to be monophyletic in the BEAST analysis. Stars represent key nodes of interest for inferring colonisation times and red boxes delimit the extent of Fig 5.11a and Fig 5.11b.

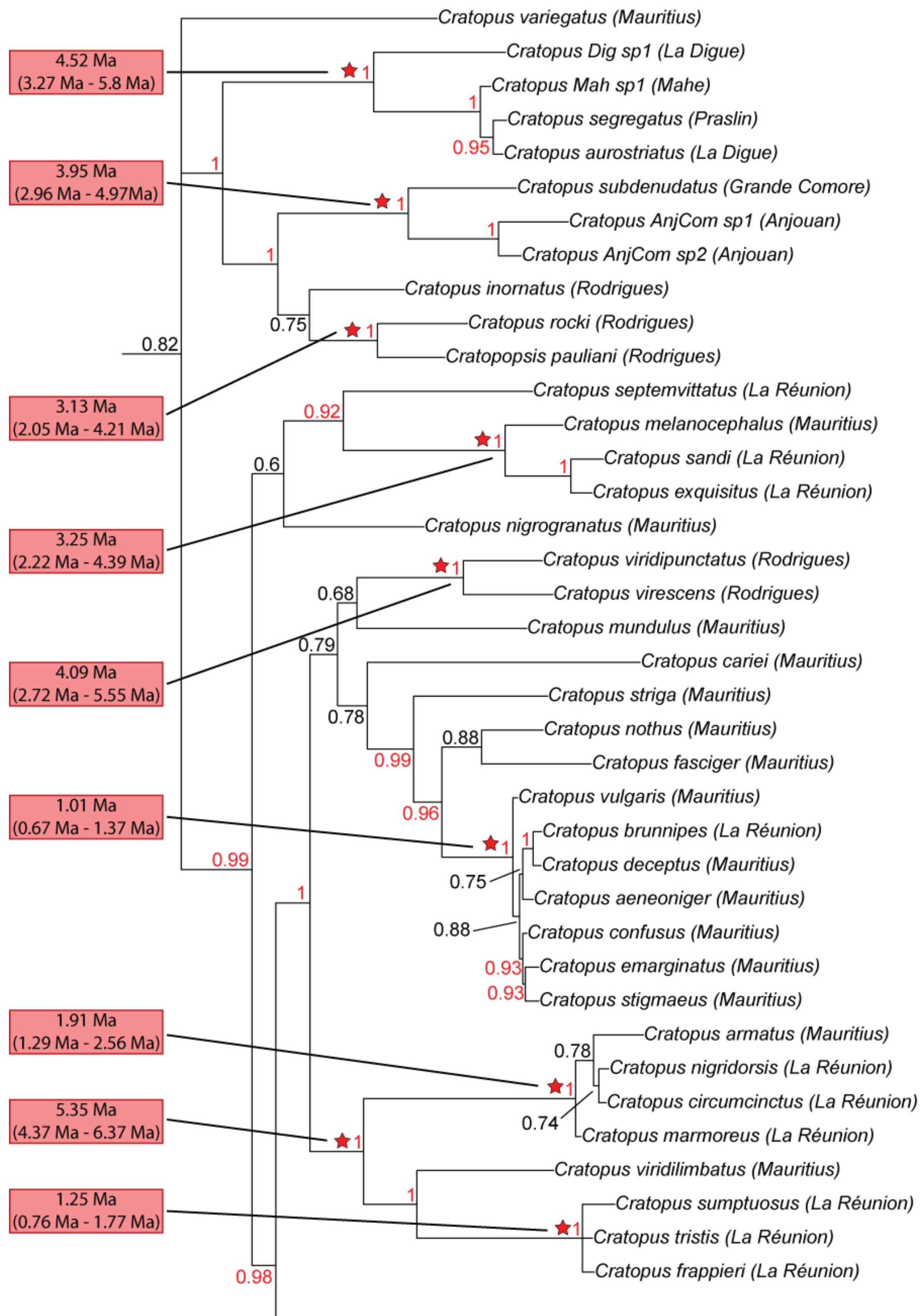


Fig 5.11a: An enlargement of lineage 1 and part of lineage 2 from Fig 5.10. The top number in the box is the mean estimated age of the node indicated, the numbers in brackets represent the 95% HPD interval on the node age estimate. Values placed on nodes represent Bayesian posterior probabilities for that node generated in the combined Mr Bayes analysis.

Fig 5.11b: An enlargement of lineage 3 and part of lineage 2 from Fig 5.10. The top number in the box is the mean estimated age of the node indicated, the numbers in brackets represent the 95% HPD interval on the node age estimate. Values placed on nodes represent Bayesian posterior probabilities for that node generated in the combined Mr Bayes analysis.

5.5 Discussion

5.5.1 Biogeography of South-West Indian Ocean Cratopine weevils

The multi-gene phylogeny partitioned all but one morphospecies into three well-supported monophyletic lineages (Fig 5.10). Across these three lineages there appears to be a complex biogeographic pattern with multiple colonisations of islands/archipelagos. While much of the Cratopine diversity of the South-West Indian Ocean appears to be in place by three to five million years ago, the geographic source of this diversity remains undetermined. Within the Mascarenes, it is possible to infer at least three colonisations of La Réunion from Mauritius, two colonisation of Mauritius from La Réunion, one major Reunionaise lineage with an undetermined source and two colonisations between Mauritius and La Réunion for which a direction cannot be determined. Additionally, there are two colonisations of Rodrigues (both with unknown origins), two of the Seychelles and two of the Comoros. One of the two colonisations in each of the Seychelles and Comoros involves a group of morphospecies that have colonised coastal habitats in either Mauritius or the Comoros and eventually spread northwards to the Seychelles via the coralline islets between the two archipelagos.

Of the three Reunionaise colonisations with direct evidence that they are from Mauritius, only one is estimated here; the ancestor of *C. brunnipes* is estimated to have colonised La Réunion from Mauritius no more than 1.0 Ma (Fig 5.11a). The remaining colonisations from Mauritius to La Réunion are seen in chapter three and involve the Reunionaise populations of *C. murinus* and *C. punctum*, both of which are nested within the lineages of their Mauritian counterparts. The dates for these colonisations are not estimated but divergences between islands are small (Figs 3.5 and 3.7) suggesting that these colonisations are also comparatively recent.

Two Mauritian morphospecies are inferred to be the result of colonisation from La Réunion. The first is *C. melanocephalus* (Fig 5.11a), the ancestor of which is estimated to have colonised Mauritius no more than 3.3 Ma, while *S. subtruncatus* (Fig 5.11b) is the result of a second colonisation to Mauritius from La Réunion no more than 1.67 Ma. The remaining Mascarene lineages have undetermined sources. Figure 5.11b shows a large monophyletic radiation of Reunionaise *Cratopopsis* morphospecies, *C. ditissimus*, *C. nanus, C. humeralis, C. leucophaetus* and

C. fulvescens, estimated to have originated at least 5.7 Ma. This age is much older than the 2.0 Ma age for the oldest lavas sampled on La Réunion (McDougal 1971) but is consistent with the maximum age of La Réunion inferred using non-radiometric methods by Bonneville (1988). The simplest explanation is that La Réunion is older than the oldest lavas would suggest and that a single colonisation generated this radiation. If La Réunion is no older than the oldest lavas, at least four colonisations would need to be inferred to account for the pattern seen leading to the conclusion that a 2.0 Ma age of La Réunion is less likely that it being older.

Directions cannot be determined for the colonisations between Mauritius and La Réunion in the clade that contains *C. armatus*, *C. nigridorsis*, *C. circumcinctus*, *C. marmoreus*, *C. viridilimbatus*, *C. sumptuosus*, *C. tristis* and *C. frappieri* (Fig 5.11a). The root is around 5.4 Ma and the topology of the clade is such that at least two colonisations have occurred. This could either be two colonisations from La Réunion to Mauritius or the opposite scenario of two colonisations from Mauritius to La Réunion with insufficient evidence to decide between scenarios.

The final major radiation in lineage two contains the morphospecies *C. punctum*, *C. griseovestitus*, *C. viridisparsus*, *C. adspersus*, *C. gloriosus*, *GCom sp1* and *MohCom sp1* (Fig 5.11b). These morphospecies are exclusively restricted to coastal habitats. BEAST estimates suggest the group formed at least 3.1 Ma and spread from the Comoros or Mauritius to the low lying islands of Juan de Nova, Europa, Grande Glorieuse and Aldabra. This occurred no less than 1.2 Ma and was followed more recently by a colonisation of Seychelles no less than 0.6 Ma. The precise route of colonisation cannot be ascertained due to low levels of support in this section of the tree and the rapid nature of the movements between islands may mean that relationships between the populations on different islands may be difficult to estimate due to incomplete lineage sorting. This may be best addressed in future using multiple individuals per island and population genetic analysis programs such as migrate-N that allow the most probable relationships amongst populations to be calculated.

Rodrigues appears to have been colonised twice (Fig 5.11a). One colonisation comes from within lineage one and consists of *C. inornatus*, *C. rocki* and *Cr. pauliani*. These are morphologically very different from *C. virescens* and *C. viridipunctatus* (Williams & Cox 2003) which are the result of the second

colonisation from within lineage two. The age of the colonisation from lineage one is difficult to estimate due to the low support for the position of *C. inornatus* but it is at least approximately 3.1 Ma while the colonisation from lineage two is no younger than approximately 4 Ma. A plausible source population for the *C. virescens*/*C. viridipunctatus* colonisation is Mauritius but given the isolation of Rodrigues from both the granitic Seychelles and the Comoros it seems unlikely that either of these archipelagos would be the direct source of colonists that led to *C. inornatus*, *C. rocki*, and *Cr. pauliani*. Warren *et al.* (2010) have hypothesised that sea level changes in the Indian ocean over the last 500,000 years would have exposed many islands now submerged (currently the shallowest parts of the Mascarene plateau, see Fig 1.1) and that these may have acted as stepping stones for organisms to disperse across longer distances. Miller *et al.* (2005) have hypothesised global sea level changes of a similar magnitude as Warren *et al.* over longer time scales suggesting that stepping stone islands may have been intermittently present over much of the evolutionary history of the Cratopine weevils and provided routes for dispersal between the granitic Seychelles, the Comoros and Rodrigues. The estimated ages of the Rodriguan lineages are also of note. The estimated divergence times which are likely to have occurred on Rodrigues (4.1 Ma between *C. virescens* and *C. viridipunctatus*; 3.1 Ma between *C. rocki* and *Cr. pauliani*) are both much older than the estimated age of the oldest lavas on Rodrigues (1.5 Ma (McDougall *et al.* 1965)). This could be accounted for if there were in reality more than two colonisations of Rodrigues with much more recent divergences from unsampled taxa on different islands. However, it should also be noted that many of the lavas on the Mascarene Islands are thought to be a result of recent volcanic reactivation and are potentially younger than the islands they cover (Montaggioni & Nativel 1988). While the oldest exposed lavas on Rodrigues are only 1.5 Ma (McDougall *et al.* 1965), it is possible that the island is as old as the seamount it sits on (approximately 15 Ma, B. Warren *pers comm.*). As the morphospecies within both Rodriguan Cratopine lineages are closely related, multiple origins within each clade would necessarily involve the same or a very closely related ancestor from outside the island. That both clades separately show similar patterns argues strongly that Rodrigues is older than the oldest lavas suggest.

The non-Rodriguan morphospecies in lineage 1 (Fig 5.11a) are recovered in two well supported clades. The first consists of morphospecies restricted to the granitic Seychelles while the second exclusively contains Comorian morphospecies. The granitic Seychelles appear to have been colonised by the ancestor of *Dig sp1*, *Mah sp1*, *C. segregatus* and *C aurostriatus* at least 4.5 Ma while the Comoros are estimated to have been colonised by the ancestor of *C. subdenudatus*, *AnjCom sp1* and *AnjCom sp2* at least 4.0 Ma, however the source of both of these clades cannot be resolved at this time. The divergence of *C. subdenudatus* from other Comorian Cratopines (approximately 4.0 Ma) is much older than the maximum age of Grande Comore (0.5 Ma (Nougier *et al.* 1986)). While this might argue for an older age of Grande Comore as was the case for Rodrigues, unlike Rodriguan Cratopines, Comorian species are poorly studied. No systematic collection of Cratopines was undertaken on the Comoros during this study and all samples included were kindly donated by researchers studying separate groups. Given the lack of knowledge regarding Comorian Cratopine diversity and the lack of intense sampling in the region, it is highly possible that closely related species exist or have recently existed on the much older nearby islands. These would have acted as a source for a recent colonisation of Grande Comore. Additionally, Grande Comore is geologically very active with island formation still in progress (the most recent major eruption of Mount Karthala was in 2005) arguing that the young inferred age of the island is probably correct.

Previous studies have suggested that much of the biota of the South-West Indian Ocean is the result of dispersal from Africa but with a significant component that has an Asian origin (reviewed in Yoder & Nowak 2006). However, this review focuses largely on Madagascar and the only studies published to date that use molecular methods to examine the biogeography of invertebrate groups in the surrounding islands are based on the spider genera Nephila and Nephilengys (Kuntner & Agnarsson 2011a; Kuntner & Agnarsson 2011b). These revealed a recent West to East colonisation of the Mascarenes from Africa via Madagascar. Africa may be the source for the Cratopine diversity in general as the other genera in the tribe are described from the Congo, Iran, Madagascar and Zaire (Alonso-Zarazaga & Lyal 1999). However these genera and indeed species of *Cratopus* from the Andamans, Maldives, Laccadives, Nicobar and Sri Lanka (and possibly Madagascar although reports are unconfirmed) remain unsampled preventing

identification of the overall geographical source. Given its age and position within the Mascarenes, its seems likely that Mauritius is the source of much of the diversity in the area and repeated colonisations of La Réunion would fit well with island biogeography theory's prediction that most colonists should come from the nearest source population (MacArthur & Wilson 1967). That some of the colonisations of La Réunion are comparatively recent suggests that this process is ongoing and that all potential niches for *Cratopus* may not yet be occupied. Comparing the ecology of recent colonisers to the established *in situ* radiations that formed much earlier in La Réunion's history may provide valuable insights into island community formation.

5.5.2 Cratopopsis bistigma and Cratopopsis mauritianus

The incongruence between COII and the nuclear genes for the morphospecies *Cr. bistigma* and *Cr. mauritianus* is highly suggestive of post colonisation mitochondrial introgression between the ancestor of *Cr. bistigma* and another Cratopine possibly related to the *C. frappieri* morphospecies group. This phenomenon appears to be common and has been documented in a number of groups including Hawaiian crickets (Shaw 2002), New Zealand grasshoppers (Trewick 2008), New Zealand cicadas (Marshall *et al.* 2011), Skipper butterflies (Zakharov *et al.* 2009) and leaf beetles (Mardulyn *et al.* 2011). Investigation of this event using further nuclear loci (e.g. using a RAD sequencing approach) would reveal the extent to which introgression has occurred. The morphologies of *Cr. bistigma* and *Cr. mauritianus* are nearly identical and both very different to the flighted *C. frappieri* morphospecies group arguing that while hybridisation has occurred, it's effects have been limited and may not be ongoing.

5.5.3 Loss of flight capability in Cratopine weevils

The phylogenies presented here and in chapter two suggest at least five instances where flight has been lost across the three Mascarene Islands (*Cr. pauliani*, the *Cr. bistigma*/*Cr. mauritianus* group, the remaining Reunionaise *Cratopopsis*, *C. murinus* and in *Scaevinus*). The number of events may actually be higher as two Cratopine weevils known to be flightless were not sampled during this study (*C. triangularis* from Mauritius and *Cr. matilei* from Anjouan). This aspect of Cratopine biology

would benefit from a more rigorous comparative analysis using Bayesian or maximum likelihood analyses, perhaps including wing development as a continuous character rather than simply flight capable vs flightless. Regardless of the details, it is clear that such a dramatic morphological change has occurred so frequently may argue that the benefits of being flightless such as energetic savings and increased reproductive output (Roff 1990) could be important in this group. Selection against dispersal capabilities has been hypothesised for organisms inhabiting limited habitat patches (Harrison 1980) and this may be applicable to the flightless *C. murinus* morphology which inhabits exposed coastal areas on Mauritius and La Réunion but the remaining flightless groups inhabit the relatively extensive and sheltered interiors of islands arguing that selection against dispersal may not be important for these beetles. An alternative explanation to selections might be that small isolated weevil populations might be unable to purge mutations that disable flight due to their small effective population sizes. Such exaptations (or spandrels by some authors) may then be adaptive by chance (Gould & Lewontin 1979). In this case, flight loss may be not be controlled by selection but may still provide the benefits postulated by Roff (1990).

5.6 Conclusions

The Cratopine weevils of the South-West Indian Ocean present three lineages with a complex biogeographic history. Mauritius is the probable source of the majority of morphospecies groups (lineage 2) with multiple colonisations of La Réunion and at least one colonisation of Rodrigues from Mauritius. There has been at least one diversification event that has lead to a radiation of coastal specialist morphospecies that have colonised the low-lying coralline islets that surround Madagascar. Flight has been independently lost in each of the three lineages and five times in total suggesting that there are distinct benefits to being flightless in this group though differing ecologies may mean that the selection pressures leading to a flightless state are different across flight loss events. Finally there is evidence of mitochondrial introgression following a colonisation of La Réunion. Further examination of this event may reveal more about the fate of new colonisers arriving on an island with closely related species already present.

5.7 References

Ahrens D, Monaghan MT, Vogler AP (2007) DNA-based taxonomy for associating adults and larvae in multi-species assemblages of chafers (Coleoptera: Scarabaeidae). *Molecular Phylogenetics and Evolution* **44**, 436-449.

Alonso-Zarazaga MA, Lyal CHC (1999) A world catalogue of families and genera of Curculionoidea (Insecta: Coleoptera)(Excepting Scolytidae and Platypodidae), 315 pp. *Entomopraxis, Barcelona*.

Austin JJ, Arnold EN, Jones CG (2004) Reconstructing an island radiation using ancient and recent DNA: the extinct and living day geckos (Phelsuma) of the Mascarene islands. *Molecular Phylogenetics and Evolution* **31**, 109-122.

Battistini R (1966) La morphologie de l'Ile Europa. *Mem. mus. natn. Hist. Nat., Paris* **91**, 7-17.

Battistini R, Cremer G (1972) Geomorphology and vegetation of Iles Glorieues. *ATOLL RESEARCH BULLETIN* **159**, 153-169.

Bonneville A, Barriot JP, Bayer R (1988) Evidence From Geoid Data of a Hotspot Origin for the Southern Mascarene Plateau and Mascarene Islands (Indian Ocean). *J. Geophys. Res.* **93**, 4199-4212.

Cicconardi F, Nardi F, Emerson BC, Frati F, Fanciulli PP (2010) Deep phylogeographic divisions and long-term persistence of forest invertebrates (Hexapoda: Collembola) in the North-Western Mediterranean basin. *Molecular Ecology* **19**, 386-400.

Clement WL, Tebbitt MC, Forrest LL*, et al.* (2004) Phylogenetic position and biogeography of Hillebrandia sandwicensis (Begoniaceae): A rare Hawaiian relict. *American Journal of Botany* **91**, 905-917.

Coffin MF, Rabinowitz PD (1987) Reconstruction of Madagascar and Africa: Evidence From the Davie Fracture Zone and Western Somali Basin. *J. Geophys. Res.* **92**, 9385-9406.

Colgan DJ, McLauchlan A, Wilson GDF*, et al.* (1998) Histone H3 and U2 snRNA DNA sequences and arthropod molecular evolution. *Australian Journal of Zoology* **46**, 419-437.

Contreras-Diaz HG, Moya O, Oromi P, Juan C (2007) Evolution and diversification of the forest and hypogean ground-beetle genus Trechus in the Canary Islands. *Molecular Phylogenetics and Evolution* **42**, 687-699.

Danforth BN, Lin CP, Fang J (2005) How do insect nuclear ribosomal genes compare to protein-coding genes in phylogenetic utility and nucleotide substitution patterns? *Systematic Entomology* **30**, 549-562.

Don RH, Cox PT, Wainwright BJ, Baker K, Mattick JS (1991) 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Research* **19**, 4008-4008.

Drummond A, Ashton B, Buxton S*, et al.* (2012) Geneious v5.6, Available from http://www.geneious.com/.

Drummond AJ, Ho SYW, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biol* **4**, e88.

Emerick CM, Duncan RA (1982) Age progressive volcanism in the comores archipelago, western Indian Ocean and implications for Somali plate-tectonics. *Earth and Planetary Science Letters* **60**, 415-428.

Emerson BC, Forgie S, Goodacre S, Oromi P (2006) Testing phylogeographic predictions on an active volcanic island: Brachyderes rugatus (Coleoptera :

Curculionidae) on La Palma (Canary Islands). *Molecular Ecology* **15**, 449-458.

Emerson BC, Oromı, amp*, et al.* (1999) MtDNA Phylogeography and Recent Intra-island Diversification among Canary Island Calathus Beetles. *Molecular Phylogenetics and Evolution* **13**, 149-158.

Ferragu M, Richard R (1990) Trois nouveaux Curculionides de Madagascar et de l'Île de la Réunion [Coleoptera]. *Revue Française d'Entomologie (N.S.)* **12**, 105-110.

Ferragu M, Richard R (1993) Etude portant sur les Cratopus de l'Ile de la Réunion [Coleoptera, Curculionidae]. *Revue Française d'Entomologie (N.S.)* **14**, 111-116.

Galman G, Matyot P, Voisin J-F (2011) Le genre Cratopus aux îles Seychelles (Coleoptera : Curculionidae). *Annales De La Societe Entomologique De France*.

Gould SJ, Lewontin RC (1979) The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme. *Proceedings of the Royal Society of London. Series B. Biological Sciences* **205**, 581-598.

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

Harrison RG (1980) Dispersal Polymorphisms in Insects. *Annual Review of Ecology and Systematics* **11**, 95-118.

Hasegawa M, Kishino H, Yano TA (1985) Dating of the Human-Ape Splitting by a Molecular Clock of Mitochondrial DNA. *Journal of Molecular Evolution* **22**, 160-174.

Holland BS, Hadfield MG (2002) Islands within an island: phylogeography and conservation genetics of the endangered Hawaiian tree snail *Achatinella mustelina*. *Molecular Ecology* **11**, 365-375.

Huelsenbeck JP, Ronquist F (2001) MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754-755.

Hustache A (1919) Synopsis des Curculionides de la Faune Malgache. I. Brachydérides et Otiorrhynchides. *Annales De La Societe Entomologique De France* **87**, 441-520.

Hustache A (1920) Curculionides des Iles Mascareignes. *Extrait des Annales de la Societe Entomologique de France* **89**.

Jordal BH, Emerson BC, Hewitt GM (2006) Apparent 'sympatric' speciation in ecologically similar herbivorous beetles facilitated by multiple colonizations of an island. *Molecular Ecology* **15**, 2935-2947.

Jordal BH, Kirkendall LR, Harkestad K (2004) Phylogeny of a Macaronesian radiation: host-plant use and possible cryptic speciation in *Liparthrum* bark beetles. *Molecular Phylogenetics and Evolution* **31**, 554-571.

Jordan S, Simon C, Foote D, Englund RA (2005) Phylogeographic patterns of Hawaiian Megalagrion damselflies (Odonata : Coenagrionidae) correlate with Pleistocene island boundaries. *Molecular Ecology* **14**, 3457-3470.

Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* **30**, 3059-3066.

Kingdon J (1990) *Island Africa: the evolution of Africa's rare animals and plants.* Collins, London.

Kuntner M, Agnarsson I (2011a) Biogeography and diversification of hermit spiders on Indian Ocean islands (Nephilidae: Nephilengys). *Mol Phylogenet Evol* **59**, 477-488.

Kuntner M, Agnarsson I (2011b) Phylogeography of a successful aerial disperser: the golden orb spider Nephila on Indian Ocean islands. *BMC Evolutionary Biology* **11**, 119.

MacArthur RH, Wilson EO (1967) *The Theory of Island Biogeography* Princeton University Press, Princeton.

Machado A, Lopez M, Almeida T, Hernandez M (2008) Mitochondrial DNA phylogenetic analysis of the genus Laparocerus (Coleoptera, Curculionidae, Entiminae). I. The madeiran clade. *Zoologica Scripta* **37**, 415-427.

Magnacca KN, Danforth BN (2006) Evolution and biogeography of native Hawaiian Hylaeus bees (Hymenoptera : Colletidae). *Cladistics* **22**, 393-411.

Mardulyn P, Othmezouri N, Mikhailov YE, Pasteels JM (2011) Conflicting mitochondrial and nuclear phylogeographic signals and evolution of host-plant shifts in the boreo-montane leaf beetle Chrysomela lapponica. *Molecular Phylogenetics and Evolution* **61**, 686-696.

Marshall DC, Hill KBR, Cooley JR, Simon C (2011) Hybridization, Mitochondrial DNA Phylogeography, and Prediction of the Early Stages of Reproductive Isolation: Lessons from New Zealand Cicadas (Genus Kikihia). *Systematic Biology* **60**, 482-502.

McDougal I (1971) Geochronology and evolution of the young volcanic island of Reunion, Indian Ocean. *Geochimica Et Cosmochimica Acta* **35**, 261-&.

McDougal I, Chamalaun F (1969) Isotopic Dating and Geomagnetic Polarity Studies on Volcanic Rocks from Mauritius, Indian Ocean. *Geological Society of America Bulletin* **80**, 1419-&.

McDougall I, Upton BGJ, Wadsworth WJ (1965) A Geological Reconnaissance of Rodriguez Island, Indian Ocean. *Nature* **206**, 26-27.

McKenna DD, Sequeira AS, Marvaldi AE, Farrell BD (2009) Temporal lags and overlap in the diversification of weevils and flowering plants. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 7083-7088.

Miller KG, Kominz MA, Browning JV*, et al.* (2005) The Phanerozoic Record of Global Sea-Level Change. *Science* **310**, 1293-1298.

Montaggioni L, Nativel P (1988) *La Reunion, Ile Maurice: Geologie et apercus biologiques, plantes et animaux* Masson, Paris.

Montaggioni L, Nougier J (1981) Les enclaves de roches detritiques dans les Volcans d'Anjouan (Archipel des Comores); Origine et interpretation dans le cadre de l'evolution du Canal de Mozambique. *Bulletin De La Societe Geologique De France* **23**, 595-601.

Moore J, White WM, Paul D*, et al.* (2011) Evolution of shield-building and rejuvenescent volcanism of Mauritius. *Journal of Volcanology and Geothermal Research* **207**, 47-66.

Nitta JH, O'Grady PM (2008) Mitochondrial phylogeny of the endemic Hawaiian craneflies (Diptera, Limoniidae, Dicranomyia): Implications for biogeography and species formation. *Molecular Phylogenetics and Evolution* **46**, 1182-1190.

Nougier J, Cantagrel JM, Karche JP (1986) The Comores archipelago in the western Indian Ocean: volcanology, geochronology and geodynamic setting. *Journal of African Earth Sciences (1983)* **5**, 135-144.

O'Grady PM, Lapoint RT, Bonacum J*, et al.* (2011) Phylogenetic and ecological relationships of the Hawaiian Drosophila inferred by mitochondrial DNA analysis. *Molecular Phylogenetics and Evolution* **58**, 244-256.

Posada D (2008) jModelTest: Phylogenetic model averaging. *Molecular Biology and Evolution* **25**, 1253-1256.

Poussereau J, Voisin J-F (2009) Cratopus francki, espèce nouvelle de la Réunion (Coleoptera, Curculionidae). *Bulletin de la Société entomologique de France* **114**, 361-364.

Rabinowitz PD, Coffin MF, Falvey D (1983) The Separation of Madagascar and Africa. *Science* **220**, 67-69.

Rambaut A (2011) FigTree v1.3.1. In: *http://tree.bio.ed.ac.uk/software/figtree/*

Rambaut A, Drummond A (2007) Tracer v1.5.0. In: *http://beast.bio.ed.ac.uk/Tracer*

Raxworthy CJ, Forstner MRJ, Nussbaum RA (2002) Chameleon radiation by oceanic dispersal. *Nature* **415**, 784-787.

Rees DJ, Emerson BC, Oromi P, Hewitt GM (2001) Mitochondrial DNA, ecology and morphology: interpreting the phylogeography of the Nesotes (Coleoptera : Tenebrionidae) of Gran Canaria (Canary Islands). *Molecular Ecology* **10**, 427-434.

Regier JC, Shi D (2005) Increased yield of PCR product from degenerate primers with nondegenerate, nonhomologous 5 ' tails. *Biotechniques* **38**, 34.

Richard R (1957) Coleopteres Curculionidae de La Reunion. *Memoires de l'Institut Sceientifique de Madagascar* **8**.

Richard R (1958) Curculionides nouveaux ou mal connus des Comores, ile Glorieuse, ile Europa New or little known Curculionidae from the Comoro Islands, Glorioso Island and Europa Island. *Mem Inst Sci Madagascar Ser E Ent* **10**, 41-63.

Richard R (1961) Notes Sur Quelques Curculionides de l'ile Rodriguez. *The Mauritius Institute Bulletin* **5**.

Richard R (1977) Nouveaux Curculionides des iles Europa at de la Réunion. *Bulletin de la Société entomologique de France* **82**, 125-127.

Richard R (1995a) Description de deux Cratopini nouveaux de la région Malgache [Coleoptera: Curculionidae]. *Revue française d'entomologie* **17**, 63-65.

Richard R (1995b) Note sur l'apparition brusque d'une variation importante de la squamulation de Cratopus humeralis Boheman (Coleoptera, Curculionidae). *Bulletin de la Société entomologique de France* **100**, 303-306.

Roff DA (1990) The evolution of flightlessness in insects. *Ecological Monographs* **60**, 389-421.

Sequeira AS, Lanteri AA, Scataglini MA, Confalonieri VA, Farrell BD (2000) Are flightless Galapaganus weevils older than the Galapagos Islands they inhabit? *Heredity* **85**, 20-29.

Sequeira AS, Sijapati M, Lanteri AA, Albelo LR (2008) Nuclear and mitochondrial sequences confirm complex colonization patterns and clear species boundaries for flightless weevils in the Galapagos archipelago. *Philosophical Transactions of the Royal Society B-Biological Sciences* **363**, 3439-3451.

Sequeira AS, Stepien CC, Sijapati M, Albelo LR (2012) Comparative Genetic Structure and Demographic History in Endemic Galapagos Weevils. *Journal of Heredity* **103**, 206-220.

Shaw KL (2002) Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: What mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 16122-16127.

Thacker RW, Hadfield MG (2000) Mitochondrial phylogeny of extant Hawaiian tree snails (Achatinellinae). *Molecular Phylogenetics and Evolution* **16**, 263-270.

Thomson J, Walton A (1972) Redetermination of Chronology of Aldabra Atoll by 230Th/234U Dating. *Nature* **240**, 145-146.

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

Voisin J-F, Poussereau J (2007a) Charactérisation de Cratopus frappieri Deyrolle, 1862 (Coleoptera, Curculionidae). *Nouv. Revue Ent (N.S.)* **24**, 329-334.

Voisin J-F, Poussereau J (2007b) L'identité de Cratopus sumptuosus Boheman, 1834, de Cratopus exquisitus Boheman, 1840, et de Cratopus sandi Deyrolle, 1862, espèces endémiques de l'île de la Réunion (Coleoptera, Curculionidae). *Nouv. Revue Ent (N.S.)* **24**, 279-283.

Voisin J-F, Poussereau J (2009) A propos de Cratopus nanus et de Cratopus parvus de l'île de la Réunion (Coleoptera, Curculionidae). *Nouv. Revue Ent. (N.S.)* **26**, 59-65.

Warren BH, Bermingham E, Bowie RC, Prys-Jones RP, Thebaud C (2003) Molecular phylogeography reveals island colonization history and diversification of western Indian Ocean sunbirds (Nectarinia: Nectariniidae). *Mol Phylogenet Evol* **29**, 67-85.

Warren BH, Bermingham E, Prys-Jones RP, Thebaud C (2005) Tracking island colonization history and phenotypic shifts in Indian Ocean bulbuls (Hypsipetes : Pycnonotidae). *Biological Journal of the Linnean Society* **85**, 271-287.

Warren BH, Bermingham E, Prys-Jones RP, Thebaud C (2006) Immigration, species radiation and extinction in a highly diverse songbird lineage: white-eyes on Indian Ocean islands. *Molecular Ecology* **15**, 3769-3786.

Warren BH, Strasberg D, Bruggemann JH, Robert PP-J, Christophe T (2010) Why does the biota of the Madagascar region have such a strong Asiatic flavour? *Cladistics* **26**, 526-538.

Wiens JJ, Moen DS (2008) Missing data and the accuracy of Bayesian phylogenetics. *Journal of Systematics and Evolution* **46**, 307-314.

Williams JR, Cox ML (2003) A Contribution to the study of Mascarene Weevils of the genus *Cratopus* Schonherr (Coleoptera: Curculionidae: Entiminae: Cratopini): The species of Mauritius and Rodrigues. *The Mauritius Institute Bulletin* **12**.

Xia X, Xie Z (2001) DAMBE: Software package for data analysis in molecular biology and evolution. *Journal of Heredity* **92**, 371-373.

Xia XH, Xie Z, Salemi M, Chen L, Wang Y (2003) An index of substitution saturation and its application. *Molecular Phylogenetics and Evolution* **26**, 1-7.

Yoder AD, Nowak MD (2006) Has vicariance or dispersal been the predominant biogeographic force in Madagascar? Only time will tell. *Annual Review of Ecology Evolution and Systematics* **37**, 405-431.

Zakharov EV, Lobo NF, Nowak C, Hellmann JJ (2009) Introgression as a likely cause of mtDNA paraphyly in two allopatric skippers (Lepidoptera: Hesperiidae). *Heredity* **102**, 590-599.

Chapter 6:
General Discussion

Typical ravine habitat, Plaine des Fougères, La Réunion 2009.

6.1 Overview

The objective of this thesis was to investigate the evolution of one of the most speciose groups of island invertebrates in the Indian Ocean. With virtually no previous information on Indian Ocean Cratopine weevils other than the conflicting opinions of taxonomists, it was necessary to investigate this group at multiple evolutionary scales from the interpopulation level through to the relationships among morphospecies on separate islands.

6.2 Taxonomy

Overall, the validity of many *Cratopus* morphospecies remains in doubt. Chapters three and four revealed that while many Mauritian and Reunionaise morphospecies form distinct well supported monophyletic units many are placed in mixed clades. Individual morphospecies within these mixed clades are paraphyletic. Often these morphospecies are variations on a theme with only subtle defining characters suggesting that the current taxonomy does not match the evolutionary units it is attempting to define. This is supported by the nuclear data in chapter five as nuclear gene trees place morphospecies in the same mixed clades and chapters three and four. However, it should be noted that the nuclear data was only examined for single individuals of each morphospecies and a more definitive answer would be possible by examining the same number of individuals for nuclear genes as were examined for the mitochondrial data.

Despite the molecular evidence presented, uncertainty still lingers over the validity of some morphospecies. Firstly the genus *Cratopopsis* is probably not valid. The *Cratopopsis* morphotype seems to have arisen independently on at least three occasions. This would benefit from formal analysis using maximum likelihood or Bayesian approaches. Should it prove to be the case that the *Cratopopsis* morphotypes are polyphyletic, it cannot even be retained as a subgenus without defining each derivation of the morphotype separately. Secondly the three lineages of Cratopine weevils defined in chapter five are well supported but the relationships among them are unknown. Whether these lineages are separate genera or simply subgenera of *Cratopus* is matter of opinion as no strict rules exist defining either taxonomic rank although resolving the relationships between

lineages would help inform this decision. Were the lineages redefined as subgenera, only lineage two could be called *Cratopus sensu stricto* as this lineage contains *C. melanocephalus*, the type species of the genus (Williams & Cox 2003). Lineage three (chapter five) also requires some revision. *Cratopus bernei* and the two *Scaevinus* species are at first inspection very different however they share a number of similarities such as acute humeral callai (the elytral 'shoulder') and an elongate deeply furrowed rostrum (Ferragu & Richard 1990; Hustache 1920) that are not possessed by other Cratopine weevils (*pers. obs.*). This may argue for a separate genus for these species.

Recent colonisations of islands also may have possible taxonomic implications. *Cratopus murinus* and *C. punctum* are apparently recent colonisations of La Réunion from Mauritius and are currently considered the same species as their Mauritian counterparts. Conversely *C. brunnipes* is currently considered a separate species to its most closely related Mauritian relatives and is probably no more divergent from them than *C. murinus* and *C. punctum* are from their Mauritian counterparts. Williams and Cox (2003) even note that *C. brunnipes* males are indistinguishable from *C. vulgaris* males from Mauritius. While the rate of migration between Mauritius and La Réunion was not estimated for these species, it is likely to be extremely low with the result that these populations are in the early stages of allopatric speciation. The taxonomy of these species should be amended to include geographically isolated subspecies. While the taxonomic information gained refers specifically to this group, understanding how genetic variation partitions with regard to morphology is fundamental to selecting taxa for the future work outlined in the final section.

6.3 Ecology

Despite the efforts of previous authors (Williams & Cox 2003) and the analyses presented in chapter two of this thesis, much of the ecology (and indeed lifecycle) of Cratopine weevils remains unknown. Chapter two suggested that *C. murinus* and *C. ovalis* at least are polyphagous with their diets comprising a subset of the plants available. Diet within these species is not restricted to a few closely related foodplants, as is suggested to be the norm for tropical coleopteran herbivores by Novotny *et al.* (2002) and Ødegaard *et al.* (2005) and it appears that phylogenetic

conservatism does not constrain the diet of *C. murinus* or *C. ovalis*. Instead it is possible that diet is limited to a set of plants that display similar structural or chemical defences to which *C. murinus* and *C. ovalis* are adapted.

Williams and Cox (2003) have hypothesised based on field observations that the dietary breadth of different species of *Cratopus* may vary and in contrast to the diets of *C. murinus* and *C. ovalis*, *C. nigridorsis* was only ever collected on Bois d'Olive Blanc (*Olea lancea*) (*pers. obs.*) during this study. Investigating the variation in diet across a wider variety species, especially ones that are sympatric could provide insight into phytophagous insect community assembly. It may generally be the case that sympatric species are able to exist through resource partitioning or alternatively, it may be the case that they have such broad trophic ranges that competition for any one foodplant does not prevent coexistence. Similar explanations have been invoked to explain the apparent lack of competition for foodplants in some butterflies (Turlure *et al.* 2009). Future work outlined in the final section would include a much broader analysis of Cratopine trophic ecology to further understand how trophic ecology varies across the group and how this interacts with species assemblages.

6.4 Flight

Cratopine weevils, particularly the species currently described as *Cratopus* are unusual among island beetle radiations in that they are flighted. Chapters two, three and four revealed that there is variation in the degree of population structure across lineages. Strong population structure argues against frequent dispersal between populations. As flight would be the most likely mode of transport between populations, structure suggests that flight may not be used as a dispersal mechanism in some Cratopine lineages. Flight capable with reduced flight musculature display population structure over relatively small distances (McCulloch 2009) and winged but apparently flightless individuals are known from a range of insect taxa (e.g. Carroll *et al.* 2003; Jackson 1933; Jackson 1956a, b). It has been suggested by McCulloch (2009) that many of the benefits of flight loss postulated by earlier authors (especially Harrison (1980) and Roff (1990)) can be obtained without complete loss of flight by simply reducing energetically expensive flight musculature (Zera & Denno 1997), this may also be applicable to

fully winged Cratopine weevils. The variation in structure across species may be indicative of a general trend in Cratopine weevils towards the flightless state seen in other island groups. If this is the case, coastal populations of C. murinus that have highly atrophied wings may represent an intermediate stage between fully flighted species and the fully flightless species of Cratopopsis. A study of wing morphology/flight musculature across species or lineages in the context of varying population structure would be useful for examining this hypothesis.

6.5 Future directions

6.5.1 Rapid colonisation of coastal habitats by Cratopus

Chapter five revealed a lineage of *Cratopus* morphospecies that have colonised coastal habitats across a large area of the Indian Ocean. There is the possibility of collaboration with Kenneth Rijsdijk (University of Amsterdam) to model island areas and sea level changes during the period in which these beetles colonised the area. In combination with population genetic techniques such as Migrate-N and phylogenetic techniques using fast evolving nuclear genes such as ITS2, it should be possible to estimate colonisation pathways and times for this group. Similar molecular approaches have been used to address recent colonisation of Hawaiian islands by happy face spiders (Croucher *et al.* 2012). The addition of the island area data would allow the hypotheses in Warren *et al.* (2010) to be evaluated and used to draw more general conclusions about sea level fluctuations and island colonisation.

6.5.2 Range mapping of Cratopine species

Specific GPS co-ordinates are available for most of the individual beetles collected. Given the dense sampling on La Réunion (and to a lesser extent Mauritius), it would be possible to use ArcGIS, possibly in conjuction with data layers (e.g. elevation, vegetation composition, rainfall or temperature) for La Réunion, to create species specific range maps. One use of these maps would be to compare the ranges of recently colonising lineages to older lineages with questions such as; (1) which species ranges overlap significantly? (2) Can the ranges of newly colonising lineages be related to other factors such as geology? The overall aim of this would be to examine how new lineages can colonise an island when older lineages should

in theory be occupying many of the niches available. This has much broader implications for island community assembly.

6.5.3 Next generation approaches to diet analysis

New sequencing approaches (particularly Illumina sequencing) could be used to rapidly assess population level diets in Cratopine weevils. Rather than look for variation across populations within the same species as in chapter two, the aim would be to ascertain the dietary overlap between sympatric species and examine whether resource partitioning permits sympatry. The benefits of using next generation sequencing include automatic separation of similarly sized DNA fragments and the ability to use multiple plant barcode loci simultaneously for more accurate identification of foodplants.

6.5.4 Comparative analyses

Weevils currently in the genus *Cratopus* vary in the degree of sexual dimorphism they display. In sexually dimorphic species, males have impressive femoral spines with an unknown function. Examining speciation rates in the context of sexual dimorphism may reveal areas in which sexual selection is driving the evolution of the group.

6.6 References

Carroll SP, Marler M, Winchell R, Dingle H (2003) Evolution of Cryptic Flight Morph and Life History Differences During Host Race Radiation in the Soapberry Bug, Jadera haematoloma Herrich-Schaeffer (Hemiptera: Rhopalidae). *Annals of the Entomological Society of America* **96**, 135-143.

Croucher PJP, Oxford GS, Lam A, Mody N, Gillespie RG (2012) Colonization history and population genetics of the color-polymorphic Hawaiian happy-face spider *Theridion grallator* (Araneae, Theridiidae). *Evolution* **66**, 2815-2833.

Ferragu M, Richard R (1990) Trois nouveaux Curculionides de Madagascar et de l'Île de la Réunion [Coleoptera]. *Revue Française d'Entomologie (N.S.)* **12**, 105-110.

Harrison RG (1980) Dispersal Polymorphisms in Insects. *Annual Review of Ecology and Systematics* **11**, 95-118.

Hustache A (1920) Curculionides des Iles Mascareignes. *Extrait des Annales de la Societe Entomologique de France* **89**.

Jackson DJ (1933) Observations on the flight muscles of *Sitona* weevils. *Annals of Applied Biology* **20**, 731-770.

Jackson DJ (1956a) The capacity for flight of certain water beetles and its bearing on their origin in the western scottish isles. *Proceedings of the Linnean Society of London* **167**, 76-96.

Jackson DJ (1956b) Observations on flying and flightless water beetles. *Journal of the Linnean Society of London, Zoology* **43**, 18-42.

McCulloch GA, Wallis GP, Waters JM (2009) Do insects lose flight before they lose their wings? Population genetic structure in subalpine stoneflies. *Molecular Ecology* **18**, 4073-4087.

Novotny V, Basset Y, Miller SE*, et al.* (2002) Low host specificity of herbivorous insects in a tropical forest. *Nature* **416**, 841-844.

Ødegaard F, Diserud OH, Østbye K (2005) The importance of plant relatedness for host utilization among phytophagous insects. *Ecology Letters* **8**, 612-617.

Roff DA (1990) The evolution of flightlessness in insects. *Ecological Monographs* **60**, 389-421.

Turlure C, Van Dyck H, Schtickzelle N, Baguette M (2009) Resource-based habitat definition, niche overlap and conservation of two sympatric glacial relict butterflies. *Oikos* **118**, 950-960.

Warren BH, Strasberg D, Bruggemann JH, Robert PP-J, Christophe T (2010) Why does the biota of the Madagascar region have such a strong Asiatic flavour? *Cladistics* **26**, 526-538.

Williams JR, Cox ML (2003) A Contribution to the study of Mascarene Weevils of the genus *Cratopus* Schonherr (Coleoptera: Curculionidae: Entiminae: Cratopini): The species of Mauritius and Rodrigues. *The Mauritius Institute Bulletin* **12**.

Zera AJ, Denno RF (1997) Physiology and ecology of dispersal polymorphism in insects. *Annual Review of Entomology* **42**, 207-230.