

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

Rates of Molecular Evolution and Gene Flow

A thesis presented in partial fulfilment of the requirements for the
degree of

Doctor of Philosophy
in
Zoology

Massey University
New Zealand

Eddy (Edwina) Jocelyn Dowle

2013

ABSTRACT

Species diversity is driven by speciation, extinction and immigration. In this thesis I explore species diversity among regions and the process of speciation via four related studies.

The latitudinal biodiversity gradient (LBG) describes the pattern of higher diversity levels towards the tropics. One of the popular hypotheses to explain the LBG is the evolutionary speed hypothesis, which suggests that species in the tropical regions are evolving faster than those in temperate regions. I demonstrate that current work on the LBG often focuses on describing the pattern as distinct from understanding the process that might drive the LBG (chapter 1). Second, I show that the most popular method used to measure differences in rate of molecular evolution between taxa from different regions, the sister-species method, does not give consistent results and estimated rates of molecular evolution can vary widely depending on outgroup selection and gene analysed. The inherent problems revealed within the current approaches raise questions as to the validity of inferences made from putative variation in rate of molecular evolution between high and low latitudinal taxa. In particular, I show that studies of the LBG that use very close sister-species pairs, rely on the most problematic datasets and should therefore be treated with considerable scepticism (chapter 2).

A phenomenon such as the LBG must derive from small-scale processes in population dynamics. Species are sometimes defined as reproductively isolated units, with hybridisation viewed as evidence of failure to speciate. But it is now increasingly acknowledged that speciation is a dynamic process during which species and populations can exchange alleles. I found relatively high levels of admixture and gene flow between morphologically distinct biological entities in two very different study systems: insects and snails. Analyses of a pair of grasshopper species provided no evidence of deviation from random mating in the region where they are sympatric. However, analysis of morphological traits provided no evidence for hybrid phenotypes between these *Sigaus* grasshoppers. I infer that some loci associated with these morphological characters are subject to strong natural

selection but neutral genetic material is being extensively shared between the two species (chapter 3). A similar situation was found in terrestrial snails that I studied, but in this case there was a clear association of neutral genetic makers with phenotype. The mtDNA haplotypes, 3,764 SNP loci and morphometric data revealed two clear genotypic clusters among New Caledonian *Placostylus*, despite strong evidence for gene flow (chapter 4). Although it is convenient and popular to define species on the basis of their reproductive isolation, a more dynamic model that allows for the possibility of gene flow is closer to reality for these taxa. Speciation can be complex and our current understanding of the process of speciation does not suggest it is limited by the genomes' rate of molecular evolution. Results from my research shows, in support of other recent research, that the process of speciation is often the product of adaptation.

PREFACE

The overall aim of this research project, *Rates of Molecular Evolution and Gene Flow*, was to examine the drivers of divergence and adaptation of populations and explore how this could drive large-scale diversity patterns. United by this common aim, the thesis is composed of four chapters, each of which can stand-alone, and a fifth, integrating chapter. Supplementary material is included for chapters 2, 3 and 4. Reading the supplementary material is not needed to understand each chapter but provides additional information to the reader that is perhaps interesting in regards to methodology and results. Although this work was a collaboration between my supervisors (Steve Trewick and Mary Morgan-Richards) and myself, all laboratory work, data analysis and initial drafts were done by myself. Steve and Mary made invaluable assistance contributions to method design, the theoretical underpinning of the work, editorial guidance and funding.

The first chapter explores the theoretical explanations for observed differences in diversity between regions. Chapter one is a detailed review of the Latitudinal Biodiversity Gradient (LBG) and the hypotheses surrounding its formation and has been published: EJ Dowle, M Morgan-Richards and SA Trewick. 2013. Molecular evolution and the latitudinal biodiversity gradient. *Heredity*: 110, 501-510. This chapter does not contain new analysis or results, rather, it suggests how the field of study around the LBG might move forward with new datasets and focuses questions. I conducted most of the background research and reviewing of the work surrounding the LBG and wrote the initial manuscript draft. Steve, Mary and I subsequently edited the draft, with input from reviewers and the journal editor.

The evolutionary speed hypothesis (chapter one), a popular hypothesis to explain the LBG, predicts that species in the tropics are diverging faster than species in temperate regions via, in part, a faster rate of molecular evolution. Chapter two surveys the ways in which rates of molecular evolution are currently measured between species using the popular sister-species method. It is currently being formatted for review. Many studies report the detection of shifts in rates of molecular evolution between species, but no one has examined the accuracy of the

methods used. Drawing on published studies, I examine whether current methods provide reliable estimates of relative rates of molecular evolution by comparing results obtained from analysis of different genes and varying outgroups. Results from this analysis have implications for numerous studies on molecular evolution rate variation. The dataset I used contains new whole mitochondrial genomes from crickets. My supervisors obtained most of the specimens but I designed the project and undertook all laboratory work for the new mitochondrial genome generation and assembled 12 of the 14 genomes. I also undertook all statistical analysis and wrote the initial draft of the manuscript. This chapter has been prepared for publication but not submitted.

Chapter three is the first of two chapters in which I study the drivers of speciation. Rate of molecular evolution might impact large-scale biodiversity patterns if speciation is largely driven by mutational speciation or if selection is limited by the variation within a species. Here I examined two species complexes to determine what processes underlie their divergence. Chapter three is a population genetics study on two species of NZ short-horned *Sigaus* grasshoppers, which is under review with *The Journal Ecology and Evolution*. One grasshopper species with a restricted range occurs in complete sympatry with the other. Previous work had hinted at gene flow between the species from comparison of mitochondrial DNA sequences. I questioned whether this pattern extended to nuclear markers and used geometric morphometric techniques to try identify morphological hybrids. I undertook all ITS sequencing, new mtDNA haplotype sequencing, RAD-sequencing, morphometric data collection and subsequent analyses. Steve and I generated the microsatellite library, and I then conducted the genotyping across the samples and ran the analyses. I wrote the initial manuscript draft and Steve, Mary and I subsequently edited this chapter.

Chapter four is a population genetics study of *Placostylus* snails from New Caledonia, which is being formatted for review. By examining their genetics and morphology I explored whether admixing has occurred in the past and whether gene flow is ongoing in areas of sympatry. I modelled environmental variables with the genetics and morphology to determine whether components of the environment drive

morphological adaptation and genetic variation. I designed this project and undertook some of the sampling of snail tissue and shells in New Caledonia, additional material was supplied by Fabrice Brescia from his work on *Placostylus*. I generated all shell morphometric data. All DNA extractions, mtDNA sequencing, RAD-sequencing, geometric morphometric analysis and the following statistical analyses were conducted by myself, as was the initial manuscript draft. Steve, Mary and I subsequently edited the draft.

Accordingly, the main thesis falls into two sections, reflecting the understanding that large-scale diversity patterns we observe must be driven in part by small-scale population dynamics. In the final summation (Chapter five) I first give a brief overview of how we can use different techniques to improve our understanding of the process of speciation. Second I describe the drivers of speciation and their role in the formation of biogeographic processes such as the LBG.

ACKNOWLEDGEMENTS

Steve and Mary, without you this thesis would not have been possible, thank you. Your enthusiasm for your work is infectious and most importantly your office doors are always open. Thank you so much for help and support through the years, I'm sure we will work together again soon.

Lizzie, you seem to know exactly when I need a distraction and a hot chocolate. Thank you for the advice with analysis, lab work and life in general. Briar, thank you for the help with the morphometric work. Most importantly thank you both for the friendship, I will miss the 'girly' nights.

Simon Hills and Gillian Gibb, thank you for putting up with me regularly barging into your office and for the all the advice and discussions. Trish McLenachan, thank you for your never-ending enthusiasm in the lab, even when things are not going to plan. Fabrice Brescia, thank you for the help with the New Caledonian snails. James Crampton, thank you for the advice in regards to the snail morphometric work. Fran Wolber, thank you for running my creepy crawlies on your flow cytometer. Jean Sanderson, thank you for the advice on modelling. Graeme Atkins, thank you for the help collecting tusk Weta from the East Cape. Alice Mathieu, thank you for taking Steve and myself out snail collecting. Robyn Walker, thank you for putting up with reading my drafts, despite not coming from a science background. Thank you Jen Ravens for the sanity gallops along the beach. Cilla Wehi, Christian Mille, Juan Carlos, Lara Shepard, Lesley Collins, Lorraine Berry, Matt Krna, Mariana Bulgarella, Niki Murray, Patrick Biggs, Grant Ramsay, Tom Burns and the phoenix's thank you for your help, friendship or support during various aspects of this PhD.

Mum and Dad, thank you so much for supporting me through this I know it has been a long time coming but it looks like I might just finish with school before Luke starts. Thanks to the rest of my family as well who have had to put up with a distracted and absent sister, aunt, granddaughter, niece and cousin. Thanks also to Marty's family who have provided many enjoyable evenings of distractions.

Marty, I'm not sure I can thank you enough. I don't think either of us knew what I had signed up for at the beginning of this PhD but having your support through it has made it all the more enjoyable, thank you. I guess it's time for me to get a real job now.

CONTENTS

ABSTRACT	ii
PREFACE	iv
ACKNOWLEDGEMENTS	vii
LIST OF TABLE AND FIGURES	xiii
1. MOLECULAR EVOLUTION AND THE LATITUDINAL BIODIVERSITY GRADIENT	1
• INTRODUCTION	2
Speciation	4
Extinction	4
Immigration	5
• THE PROCESS BEHIND THE PATTERN	5
The evolutionary speed hypothesis	7
Metabolic rate variation driving differential rates of molecular evolution	11
Generation time and longevity driving differential rates of molecular evolution	12
UV radiation driving differential rates of molecular evolution	13
Population size driving differential rates of molecular evolution	13
• TOWARDS A SYNTHESIS	16
Rates of molecular evolution and speciation	16
Molecular evolution and extinction	17
Countless difficulties	18
• FUTURE FOCUS ON THE PROCESS	20

2. ESTIMATES OF RATE OF MOLECULAR EVOLUTION ACROSS THE MITOCHONDRIAL GENOME USING SISTER-SPECIES COMPARISONS ARE NOT CONSISTENT: TESTS WITH CRICKETS AND FLIES	34
• INTRODUCTION	35
• METHODS	40
• RESULTS	43
• DISCUSSION	50
SUPPLEMENTARY MATERIAL ONE	57
• SUPPLEMENTARY MATERIAL FOR CHAPTER TWO	57
S1.1 Species-Pairs in detail	57
S1.2 Generation of novel mitochondrial genomes	58
S1.3 New Ensifera genomes sequenced	60
S1.4 ML trees for Drosophila and Ensifera	62
S1.5 Bayesian trees for Ensifera	63
3. GENE FLOW AND DIVERGENCE OF AN ENDANGERED NEW ZEALAND GRASSHOPPER	66
• INTRODUCTION	67
• METHODS	73
Morphology	73
Mitochondrial DNA sequence	75
Microsatellites	76
Nuclear Sequence	77
RAD-Seq SNPs	78
• RESULTS	81
Morphology	81
Mitochondrial DNA sequence	85

Microsatellites	86
Nuclear Sequencing	87
RAD-seq SNPs	88
• DISCUSSION	91
 SUPPLEMENTARY MATERIAL TWO	 100
• S2 SUPPLEMENTARY MATERIAL FOR CHAPTER THREE	100
S2.1 Microsatellite development methods	100
S2.2 Microsatellite primers	101
S2.3 Tally up table	102
S2.4 ITS table	103
 4. ECOTYPES AND SPECIATION WITH GENE FLOW IN A BIODIVERSITY	
HOT SPOT	107
• INTRODUCTION	107
• METHODS	113
Sampling strategy	113
Geometric analysis	114
Mitochondrial DNA sequence	116
RAD-seq SNP	117
Environmental modelling	120
• RESULTS	122
Geometric analysis	122
Mitochondrial DNA sequence	127
RAD-seq SNP	129
Environmental modelling	136
• DISCUSSION	138

SUPPLEMENTARY MATERIAL THREE	146
• S3 SUPPLEMENTARY MATERIAL FOR CHAPTER FOUR	146
S3.1 Marker information and SNP tally Sp2	146
S3.2 Marker information and SNP tally Sp4	148
S3.3 Marker information and SNP tally Up2	150
S3.4 Marker information and SNP tally Up4	152
S3.5 F_{ST} tables Sp4, Up2 and Up4	154
S3.6 STRUCTURE results Sp2, Sp4, Up2 and Up4	157
S3.7 Topology of uncommon SNAPP trees	158
5. SUMMATION	159
• DIVERGENCE BETWEEN POPULATIONS IS FREQUENTLY DRIVEN BY NATURAL SELECTION	159
Selection and Gene flow	159
Selection and morphology	161
• BIOGEOGRAPHY IS DRIVEN IN PART BY POPULATION LEVEL DIVERGENCE	161
The latitudinal biodiversity gradient	161
• SPECIATION MEDIATED THROUGH CHANGES IN FIXATION AND MUTATION RATE	163
Fixation rate	163
Mutation rate	164
• CONCLUSIONS	165

LIST OF TABLE AND FIGURES

CHAPTER 1:

FIGURE 1	Environmental Gradients.....	3
FIGURE 2	Hypotheses surrounding LBG.....	6
FIGURE 3	Detecting rate shifts in trees.....	10

CHAPTER 2:

FIGURE 1	Rate shift predictions.....	38
FIGURE 2	Rate shift results.....	46
TABLE 1	Rate shifts between genes.....	44
TABLE 2	Regression analysis of rates.....	48

SUPPLEMENTARY 1:

FIGURE 1	ML trees for Drosophila and Ensifera.....	62
FIGURE 2	Bayesian trees for Ensifera.....	63
TABLE 1	Species-Pairs in detail.....	57
TABLE 2	New Ensifera genomes sequenced.....	60

CHAPTER 3:

FIGURE 1	Map of sample locations.....	70
FIGURE 2	Hypothesis for Grasshoppers.....	72
FIGURE 3	PCA results traditional morphology.....	82
FIGURE 4	PCA (geometric morphometric), STRUCTURE and mtDNA results.....	84
FIGURE 5	Network from mtDNA.....	86
FIGURE 6	BAYESCAN and F_{ST} distribution results.....	89
TABLE 1	Discriminate analysis traditional species.....	81
TABLE 2	Squared distances discriminate analysis.....	82
TABLE 3	MIGRATE results.....	90

SUPPLEMENTARY 2:

TABLE 1	Microsatellite primers.....	101
TABLE 2	Tally up table.....	102
TABLE 3	ITS table.....	103

CHAPTER 4:

FIGURE 1	Map sampling and environmental maps.....	111
FIGURE 2	Snail shells.....	112
FIGURE 3	Geometric Morphometric technique.....	114
FIGURE 4	PCA and CVA analysis.....	125
FIGURE 5	Sympatry results.....	126
FIGURE 6	ML tree ND2 gene.....	128
FIGURE 7	BAYESCAN results Sp2.....	130
FIGURE 8	STRUCTURE results.....	132
FIGURE 9	SNAPP results.....	134
FIGURE 10	Morphology and tree topology.....	135
FIGURE 11	Environmental results.....	137
TABLE 1	Discriminate analysis traditional species.....	122
TABLE 2	Discriminate analysis genetic groups.....	123
TABLE 3	SNP datasets.....	129
TABLE 4	Sp2 F_{ST} table.....	131
TABLE 5	MIGRATE results.....	136

SUPPLEMENTARY 3:

FIGURE 1	STRUCTURE results.....	157
FIGURE 2	Topology of uncommon SNAPP trees.....	158
TABLE 1	Marker information Sp2.....	146
TABLE 2	SNP tally Sp2.....	147
TABLE 3	Marker information Sp4.....	148
TABLE 4	SNP tally Sp4.....	149
TABLE 5	Marker information Up2.....	150
TABLE 6	SNP tally Up2.....	151
TABLE 7	Marker information Up4.....	152

TABLE 8	SNP tally Up4	153
TABLE 9	F_{ST} Sp4	154
TABLE 10	F_{ST} Up2	155
TABLE 11	F_{ST} Up4	156

1. MOLECULAR EVOLUTION AND THE LATITUDINAL BIODIVERSITY GRADIENT

ABSTRACT

Species density is higher in the tropics (low latitude) than in temperate regions (high latitude) resulting in a latitudinal biodiversity gradient (LBG). The LBG must be generated by differential rates of speciation and/or extinction and/or immigration among regions, but the role of each of these processes is still unclear. Recent studies examining differences in rates of molecular evolution have inferred a direct link between rate of molecular evolution and rate of speciation, and postulated these as important drivers of the LBG. Here we review the molecular genetic evidence and examine the factors that might be responsible for differences in rates of molecular evolution. Critical to this is the directionality of the relationship between speciation rates and rates of molecular evolution.

Keywords: mutation rate; fixation rate; population size; molecular evolution; speciation

Dowle EJ Morgan-Richards M & Trewick SA (2013). Molecular evolution and the latitudinal biodiversity gradient. *Heredity*. **110**(6), 501-510

INTRODUCTION

"Animal life is, on the whole, far more abundant and more varied within the tropics than in any other part of the globe, and a great number of peculiar groups are found there which never extend into temperate regions".

A. R. Wallace 1876

One of the most striking biogeographic patterns on the planet is the uneven latitudinal distribution of biodiversity (Fig. 1). A similar trend of high and low diversity is documented in all major groups of organisms, on land and in the sea (Hillebrand 2004).

The evolutionary and ecological factors responsible for the pattern usually referred to as the latitudinal biodiversity gradient (LBG) have been hotly debated over the last 30 years (Rohde 1992; Gaston 2000; Hawkins 2004; Losos 2008; Erwin 2009; Gillman *et al.* 2009). Interest in the drivers behind the LBG has expanded from biogeographers (starting with Wallace) to include ecologists, evolutionary biologists (in particular evolutionary geneticists), physiologists and palaeontologists (Rosenzweig 1995; Allen *et al.* 2002; Davies *et al.* 2004; Jablonski *et al.* 2006; Roy and Goldberg 2007), however, no consensus has been reached. There is general agreement that variation in the total number of species among regions must result from variation in one or more of three basic phenomena underpinning the diversity of life on earth: speciation rates, extinction rates and immigration rates. Regional biotas are the product of the interaction of these three processes, but understanding what factors drive these to form the LBG is problematic (Jablonski *et al.* 2006). How far have we progressed in developing an explanation since Alfred Wallace's observation more than 130 years ago? Here we focus on recent molecular genetic studies to see whether this field, which has contributed so much to evolutionary research, has helped our understanding of the latitudinal biodiversity gradient.

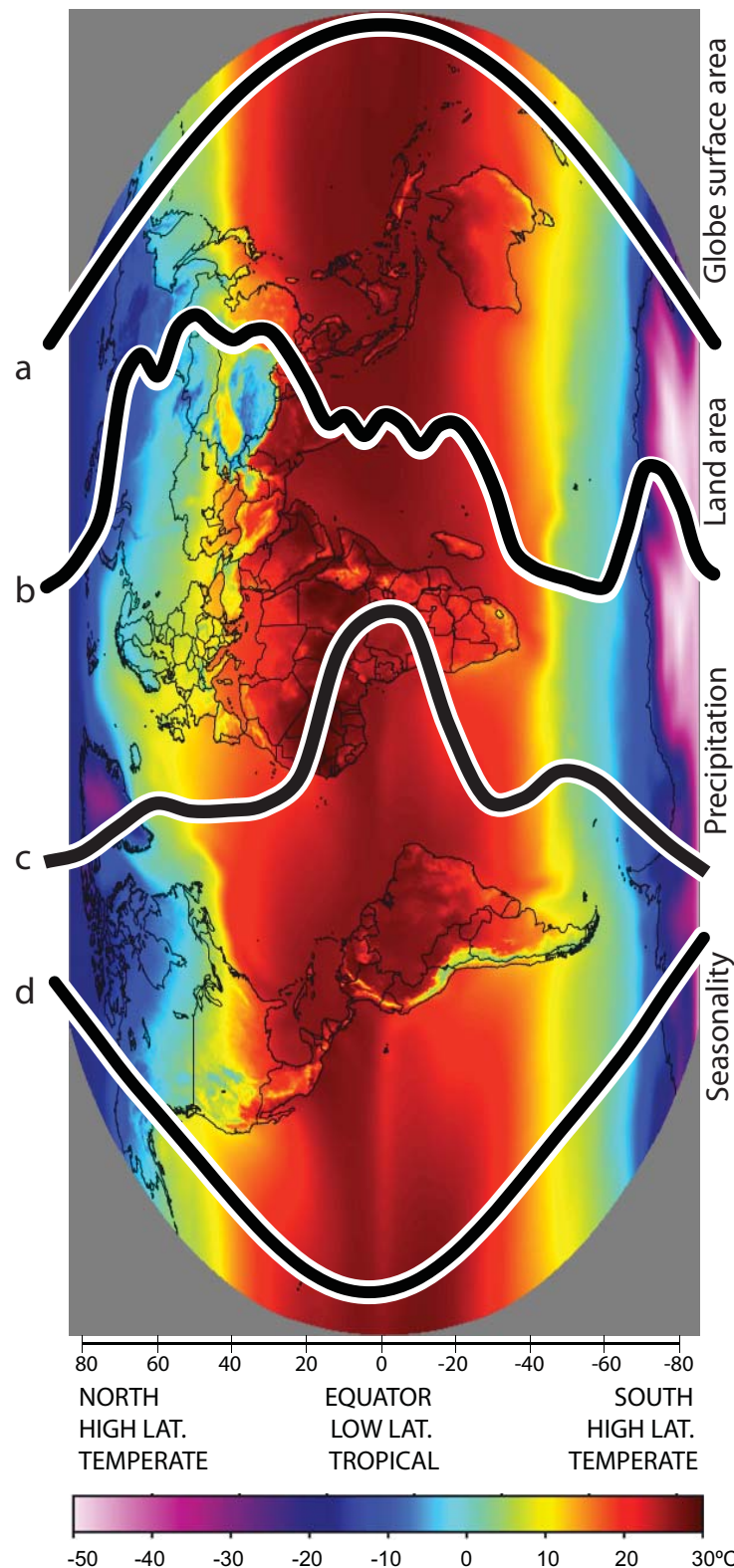


Figure 1

A latitudinal biodiversity gradient (LBG) has been identified in almost all organisms that have been investigated, on land and in the sea. The gradient involves high species' numbers near the equator (at low latitudes) and lower numbers of species at high latitudes. The question of why and how this pattern has arisen has persisted for more than 150 years. Several environmental parameters form latitudinal gradients reflecting the shape of the planet, its rotation and orientation to the sun. Although the globe's surface area (geometric calculation of a dome) is evenly graded with latitude (a), the extent of land area (b), ocean, and variation in bathymetry are not graded evenly. Associated wind and ocean currents further complicate the distribution of environmental variables. Average marine and terrestrial temperatures reduce with increasing latitude (background), but precipitation (c) (L.D.Roper Virginia Poly. Inst. and State University) does not have a linear relationship and seasonality (e.g. in terms of annual variance of insolation increases) (d). Correlation of and of these traits with the LBG does not demonstrate causation.

Temperature map produced by R.A Rhode, http://en.wikipedia.org/wiki/File:Annual_Average_Temperature_Map-jpg. Isolation data from <http://www.applet-magic.com/insolation.htm>. Details are available from the authors.

Speciation

Differences in speciation rates have been the major focus of geneticists, ecologists and palaeontologists seeking to explain the LBG. The tropics are referred to as a cradle of diversity with high speciation rates inferred from observed high species' diversity (Stebbins 1974; Chown and Gaston 2000). Paleontological data allow comparison of direct counts of species' origins (speciation) in past tropical and temperate areas and have provided the most compelling evidence for differing speciation rates (Jablonski *et al.* 2006). Molecular data are increasingly applied to the study of speciation rates, but these analyses are not always directed specifically at the LBG (Lancaster 2010; Lanfear *et al.* 2010a). Although there is compelling evidence of relatively high rates of speciation in the tropics, the underlying driver(s) of this have not been identified (Martin and Mckay 2004; Allen and Gillooly 2006; Jablonski *et al.* 2006; Mittelbach *et al.* 2007; Krug *et al.* 2009; Condamine *et al.* 2012).

Extinction

A higher rate of extinction at higher latitudes (away from the equator) was the explanation offered by Wallace for the formation of the LBG (Wallace 1876). Hypotheses focusing on extinction rates have often centred on the putative effects of climatic extremes in the past (e.g. during the Pleistocene), which may have been felt most intensely at higher latitudes (Fig. 1). In the tropics a more uniform climate over time may have allowed accumulation of species and these regions have thus been referred to as museums (Wallace 1876; Wright 1983; Guo and Ricklefs 2000). Unfortunately, the evidence for extinction and its effect is striking by its absence in discussions on this topic, reflecting a general problem with the quantification and accommodation of extinction in biogeography (Crisp *et al.* 2011). Only palaeontology has provided direct counts of extinction and some comparison of extinction rates, mostly in marine invertebrates. It is well recognised that climate cycling caused extinction in temperate areas during the Pleistocene, but over a longer time frame the evidence is more equivocal. Few studies find strong support for differences in extinction rates between latitudes, and the general inference is that extinction alone cannot explain the LBG (Hawkins *et al.* 2006; Jablonski *et al.* 2006; Martin *et al.* 2007; Valentine *et al.* 2008).

Immigration

Although immigration is central to some biogeographic models (Macarthur and Wilson 1967) it has been considered, by some, to be too weak a force to be important in as large-scale a trend as the LBG (Gaston 2000). To be relevant in this context, immigration needs to involve permanent range change of a taxon resulting in absence from its original range. Although evidence of large-scale species-range movement has been reported in marine bivalves, terrestrial species may be more range limited (Jablonski *et al.* 2006; Martin *et al.* 2007; Mittelbach *et al.* 2007; Crampton *et al.* 2010). Niche conservatism has been interpreted by some as indicating a limited capacity in many plant and animal groups for substantial range (habitat) shifts (e.g. Wiens and Graham 2005). Immigration has yet to be linked to molecular evolution directly, although range changes could have important implications for the assumptions within many molecular approaches. In particular, the assumption that the current range mid-point of species represents the midpoint through most of the evolutionary history of a species is questionable.

THE PROCESS BEHIND THE PATTERN

“The causes of these essentially tropical features are not to be found in the comparatively simple influences of solar light and heat, but rather in the uniformity and permanence with which these and all other terrestrial conditions have acted; neither varying prejudicially throughout the year, nor having undergone any important change for countless past ages”.

A. R. Wallace 1876

The latitudinal biodiversity gradient is an observable pattern that, as is typical of biogeographic patterns, does not in itself provide information about the processes involved in its formation (Figs. 1, 2). Latitude values provide a convenient scale for graphing biodiversity patterns but do not directly express the driving force(s) of the gradient (Zapata *et al.* 2003; Currie and Kerr 2008). In recent work temperature has become a focal putative driver replacing an earlier emphasis on habitat area (Losos and Schluter 2000).

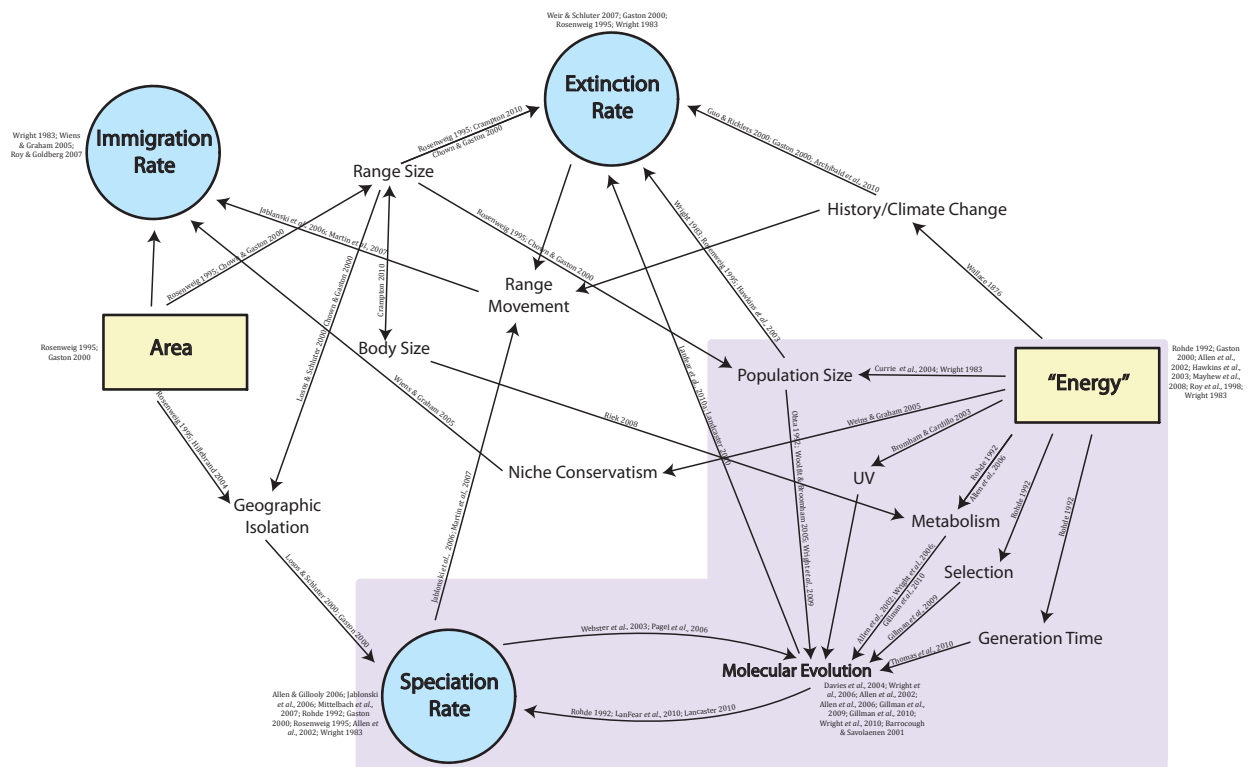


Figure 2

A multitude of hypotheses seek to explain the latitudinal biodiversity gradient (LBG). Each hypothesis seeks to relate the LBG to processes that influence one or more of the three components controlling regional species diversity: immigration, extinction and speciation. The two key environmental parameters most frequently cited as underlying drivers of the LBG are, area and energy (generally measured in the form of temperature). The inferred importance and directionality of effects, varies among taxa, researchers and methods applied, with no clear consensus. Here we illustrate a sub-set of theories on the formation of the gradient that have gathered some empirical support (citations in figure) and, in particular, those that relate to molecular evolution (shaded area). A link between speciation rate and the rate of molecular evolution is widely shown, but which is the driver of the other is equivocal. A latitudinal gradient in available energy may influence many traits but these are only tentatively associated with rates of extinction and/or speciation.

Paleontological studies find evidence that environmental temperature is correlated with biodiversity through time in marine molluscs and protists (Crame 2002; Mayhew *et al.* 2008) (but see Crampton *et al.* 2006 where no link was found for molluscs). Most studies find the strongest correlation with temperature (or in some cases ultra violet radiation, UV) when comparing species richness with a number of possible variables (e.g. area, mid-domain effects; Clarke and Gaston 2006). Several mechanisms have been proposed for how higher temperature might lead to greater species diversity by influencing the speed of speciation, species' physiological tolerances and extinction rates. The 'evolutionary speed' hypothesis has thus become the explanation for the biodiversity gradient preferred by many molecular biologists.

The evolutionary speed hypothesis

Rohde (1992) described the evolutionary speed hypothesis to explain the LBG in terms of a difference in speciation rates; speciation rates being influenced by a latitudinal disparity in rates of molecular evolution and selection. Any differences in rates of molecular evolution between latitudes are considered to be products of the different environmental conditions that individuals experience at different latitudes. Differing rates of molecular evolution might influence speciation rates if the latter are determined at least partially by the random accumulation of genetic differences (mutation-order speciation) and natural selection (ecological speciation) (Schluter 2009; Lancaster 2010). Allopatric speciation via genetic drift (mutation-order speciation) might result in the formation of new species just as rapidly as ecological speciation, however, the relative influence of these two processes in speciation is

not known (Schluter 2009; Nosil and Flaxman 2011). Ecological speciation is the more readily studied of the two, and can accommodate difficulties of gene flow (Rundle *et al.* 2000; Funk *et al.* 2006; Langerhans *et al.* 2007; Schluter 2009). Mutation-order speciation might contribute to further separation of populations already partitioned by natural selection. Alternatively it could result in faster reproductive isolation by genetic drift in geographically isolated (allopatric) populations perhaps via a Dobzhansky-Muller process (Schluter 2009; Nosil and Flaxman 2011).

A faster mutation rate could increase the rate of speciation (e.g. Lancaster 2010) if mutation-order speciation contributes significantly to overall speciation rate. This presumes that all other evolutionary processes (e.g. competition, character displacement, reproductive cohesion, adaptation) are limited only by mutation rate. Central to the biodiversity debate is the conundrum: does speciation result in accelerated molecular evolution (Webster *et al.* 2003; Pagel *et al.* 2006) or does accelerated molecular evolution increase the chance of speciation? That even this fundamental point is subject to debate highlights just how little we understand the role of genetic variation in the LBG and speciation generally.

The evolutionary speed hypothesis predicts an elevated rate of molecular evolution towards the tropics and that this variation has a direct effect on relative rates of speciation. A latitudinal difference in rates of molecular evolution (faster in the tropics) has indeed been found in a range of organisms using sister species comparisons of plants, fish, frogs, foraminifera, and mammals (Davies *et al.* 2004; Allen *et al.* 2006; Wright *et al.* 2006; Gillman *et al.* 2009; Gillman *et al.* 2010; Wright *et al.* 2010; Wright *et al.* 2011) although the relationship for mammals may be weaker than originally reported (Weir and Schluter 2011)(Fig. 3). Rohde (1992) considered three main explanations for increased rates of molecular evolution: shorter generation times; direct effects of temperature on mutation rate; acceleration of mutation rates due to acceleration of physiological process along with an acceleration of positive selection due to the first two. These explanations have since been elaborated and extended to include the postulated influence of metabolic rates, generation time, UV radiation, and population size (Fig. 2). Here we

examine these four putative drivers of differential rates of molecular evolution. To fully understand how rates of molecular evolution could be influencing the biodiversity gradient it is important to understand how and why these drivers could affect rates of molecular evolution in the first place, providing context to studies of molecular evolution rate variation among latitudes.

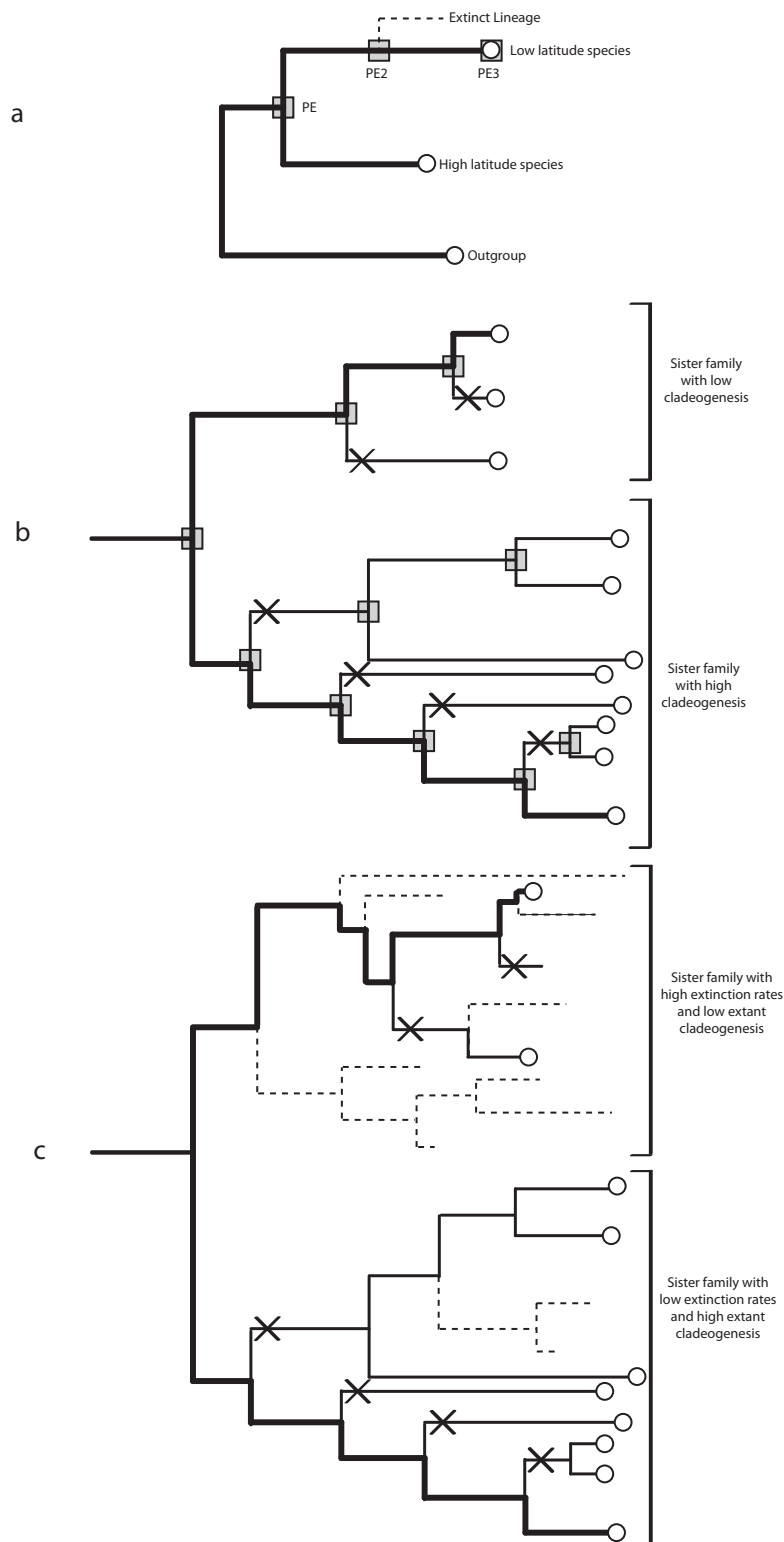


Figure 3

Evidence of variation in the rate of molecular evolution has come from two main sampling strategies that use DNA sequences from extant species (usually represented by a single individual) to infer rate differences among lineages. Possible instances of punctuational evolution at nodes (grey boxes), extinct lineages (dashed lines), pruned or un-sampled lineages (X) are shown. (a) The sister species comparison. Sister species are chosen to represent contrasting distributions; one at low latitude and one at high latitude (temperate vs. tropical as in Gillman *et al.* 2009). Placement of the outgroup in the phylogenetic reconstruction provides the crucial evidence for rate variation between the two ingroup lineages. By sampling only one individual per lineage, this method avoids the node-density effect that might influence estimates of molecular evolution rate (Hugall and Lee 2007). Punctuated equilibrium predicts that acceleration in rates of molecular evolution is associated with speciation and therefore occurs at nodes.

If so, observed differences in rates (different branch lengths) cannot result from punctuated equilibrium (PE) unless speciation (PE₂) and extinction have occurred only on the branch leading to the tropical species. Hidden (or incipient) speciation (PE₃) could also explain the observed faster rate of molecular evolution if it was occurring more frequently in the tropical lineages than the temperate lineages. (b) Pruning large datasets. Species are selected from sister lineages with different species numbers to examine whether or not lineages with higher cladogenesis have faster rates (e.g. Lanfear *et al.* 2010). Relatively large data sets are pruned (X) to a pair of contrasting lineages, if more sequence data is available for one family than another, to avoid the node density effect. (c) Pruning large

datasets with extinction. Lineage extinction (dashed lines) undermines confidence in inferences about the directionality of effects. Surviving species only provide estimates of net-diversification (speciation – extinction), not rates of speciation *per se*. Insufficient knowledge of extant lineages produces the same effect as extinction suggesting that this method would only work for well-studied groups.

1. Metabolic rate variation driving differential rates of molecular evolution

Work on metabolism and its influence on the latitudinal biodiversity gradient started as part of the ‘metabolic theory of ecology’ (Allen *et al.* 2002; Gillooly and Allen 2007). It has been suggested that the LBG arose in response to a latitudinal gradient in kinetic energy influencing rates of molecular evolution in ectotherms (Allen and Gillooly 2006; Gillooly and Allen 2007). Under this model, ectotherms in warmer areas have an increased metabolic rate governing consumption of oxygen and production of oxygen free radicals that potentially increase mutation rate by damaging DNA (Allen *et al.* 2006; Gillooly *et al.* 2007). Comparing species with different metabolic rates is a proxy for comparison of different rates of oxygen free radical formation (and inferred mutation), which, until recently, could not easily be measured directly. New sequencing technology now allows for comparison of whole genomes of parents and their offspring, which can be used to estimate *de novo* mutation rates of species (see below). In comparison to other putative drivers, the concept of metabolic rates and the LBG yields clearly testable predictions, but has been heavily contested since its origin (Algar *et al.* 2007; Hawkins *et al.* 2007a; Hawkins *et al.* 2007b; De Castro and Gaedke 2008; Irlich *et al.* 2009).

Unexpectedly, elevated rates of molecular evolution in the tropics compared to temperate regions have also been observed in some endotherms (e.g. mammals, Gillman *et al.* 2009), which was not predicted by the original metabolic hypothesis (Storch 2003; Gillman *et al.* 2009). Metabolic rates in endotherms are closely linked to body size, with larger bodies having higher absolute metabolic rates but lower rates per unit mass (Riek 2008). As body size in endotherms tends to be greater at higher latitudes (Bergmann’s rule), a counter gradient decrease in basal metabolic rate (BMR) per unit mass with increased latitude is suggested (Ashton *et al.* 2000). But Gillman *et al.* (2009) found that even when body size of mammals was accounted for, the rate of molecular evolution still increased towards the tropics.

However, there is evidence from comparative studies that BMR is de-coupled from molecular evolution in the bird and mammal mitochondrial DNA that are widely used in these studies (Lanfear *et al.* 2007; Galtier *et al.* 2009a; Nabholz *et al.* 2009). Thus the proxy for oxygen free radicals (metabolic rate) may inform little on rates of mutation or molecular evolution, and we are no closer to understanding the driver of variation in the rate of molecular evolution and the formation of the latitudinal biodiversity gradient.

2. Generation time & longevity driving differential rates of molecular evolution

Generation times are generally assumed to be shorter in tropical organisms than their temperate counterparts. Having a short generation time could cause lineages to accumulate more mutations in a given time period compared to lineages with long generation times (Thomas *et al.* 2010). Though under-examined, the predicted correlation between generation time and rates of molecular evolution has been observed in some groups of invertebrates, flowering plants, host/parasite systems and mammalian nuclear sequences (Nikolaev *et al.* 2007; Smith and Donoghue 2008; Welch *et al.* 2008; Thomas *et al.* 2010) but not for other plants or mammal mtDNA (Whittle and Johnston 2003; Nabholz *et al.* 2008). Ohta (1992) suggested that variation in generation time and population size was the cause of the discrepancy between observed rates of protein and nucleotide evolution.

Plants (especially large, long-lived species) can be problematic subjects in regard to generation time because mutations in reproductive tissue are influenced also by mutagenesis of somatic cells (Gaut *et al.* 2011). Little direct testing of generation times across latitudes has been undertaken, probably because the basic biology of many tropical species is not known. Linked to this problem is the issue of longevity, which refers to the number of cell cycles an individual undergoes rather than the life expectancy of the organism. This is another life history hypothesis that has yet to be implicated in the LBG, but has a similar relationship to rates of molecular evolution as generation time. Longer-lived species (with more cell cycles per individual) are expected to be more disadvantaged by higher mutation rate as premature aging could result from a high somatic DNA mutation rate (Nabholz *et al.* 2008; Nabholz *et al.* 2009).

3. Ultra violet radiation driving differential rates of molecular evolution

Ultra violet (UV) is a mutagen and as a component of insolation is most constant and intense at the equator (Fig. 1) (Rastogi *et al.* 2010). UV intensity has been linked to the LBG by the prediction that an increased mutation rate with higher exposure to UV (Bromham and Cardillo 2003; Davies *et al.* 2004; Flenley 2011) may be correlated with speciation rate. The proposed role of UV as a producer of heritable mutations is perhaps more straightforward in animals where exposure to UV could occur during embryonic development (Epel *et al.* 1999; Maurer *et al.* 2011) as this could result in germ-line mutations. In plants UV has been proposed as a potential mutagen of pollen grains and experimental studies have shown its effects (Ahmad *et al.* 1991; Flenley 2011). In mammals a role for UV in germline mutations is less obvious but not incomprehensible (Wilson Sayres and Makova 2011). UV levels are generally higher in the tropics than in temperate regions, however, UV radiation unlike temperature, does not decrease with altitude and may even increase (Korner 2007; Barry 2008). A number of studies exploring differences in rates of molecular evolution have considered sister species pairs from different altitudes (Wright *et al.* 2006; Gillman *et al.* 2009; Gillman *et al.* 2010; Wright *et al.* 2010). In all cases (plants, mammals and amphibians) inferred rates of molecular evolution were higher at lower altitudes, contrary to the prediction of UV as a biodiversity driver.

4. Population size driving differential rates of molecular evolution

The effects of population size are complex and important as they may interact with other putative drivers of LBG (Charlesworth 2009). There is an extensive literature relating to the influence of population size, comprising many contrasting views and interpretations. Population size is often postulated as a component of LBG hypotheses, but is variously assumed to be largest or smallest in the tropics, although evidence for either view is limited (Wright 1983; Currie *et al.* 2004; Gillman *et al.* 2009). Land area of islands is commonly used as a proxy for population size in studies of rates of molecular evolution in terrestrial organisms (Woolfit and Bromham 2005; Wright *et al.* 2009), an assumption founded in island biogeography theory (Macarthur and Wilson 1967).

Although there is some limited evidence for latitudinal gradients in population size on a global scale, the topic is controversial because of the contested influence of population size on rates of molecular evolution. This is frequently confounded by a lack of clarity about what estimate of molecular evolution is being used; relative branch length or population variation. Theory predicts that large populations will have more genetic diversity at any one time compared to small populations (Ellstrand and Elam 1993). This is because smaller populations are more sensitive to genetic drift so that mutations are likely to go to fixation more quickly resulting in lower net genetic variation at the population level. Furthermore, the fewer individuals there are in absolute terms the fewer mutations there will be in the population. Thus large populations might contain the genetic diversity required for ecological speciation. The connection between population size and relative branch length is less clear (Johnson and Seger 2001; Woolfit and Bromham 2005; Nikolaev *et al.* 2007; Wright *et al.* 2009; Smith 2011). Branch lengths in a phylogeny, expressing local rates of molecular evolution, are influenced by mutation rate and fixation rate. Under neutral theory the ratio of non-synonymous nucleotide mutations (dN) to synonymous nucleotide mutations (dS) will be equal to their respective proportions in the genome (Kimura 1968). Therefore the rate of molecular evolution would not be influenced by population size under a neutral model. Synonymous nucleotide mutations do not result in amino acid substitutions and are usually neutral with respect to natural selection, having no phenotypic effect. In contrast, non-synonymous nucleotide mutations result in amino acid substitutions and are therefore frequently not neutral. Almost all studies investigating population size and the rate of molecular evolution violate the assumption of neutrality by studying areas of the genome that are known to be constrained by selection (e.g. coding regions of mitochondrial genomes such as COI and Cytb). Nearly-neutral theory was developed in response to observations of a mismatch between inferred short term and long term mutation rates, and predicts that slightly deleterious mutations, which would be eliminated in large populations, may tend to persist longer in smaller populations where purifying selection is more relaxed (Ohta 1992). Thus, the nearly-neutral model predicts higher rates of non-synonymous substitution in small populations, but that the rate of synonymous substitutions should be independent of population size (Ohta 1992). This leads to

the prediction that the dN/dS (ω) will differ between large and small populations, and that changes in population size (expansion and contractions) will, in theory, alter rates of molecular evolution. Some empirical evidence agrees with this model (Johnson and Seger 2001; Woolfit and Bromham 2005; Bakewell *et al.* 2007; Charlesworth and Eyre-Walker 2007; Kosiol *et al.* 2008; Petit and Barbadilla 2009). However, some of these studies have found the results to be weak, with one finding the reverse (dN and dS were both lower in smaller than larger populations) (Wright *et al.* 2009). Explanations proposed for these results include a role for positive selection in the large populations, positive back-mutations during population expansion and/or linked substitutions (Charlesworth and Eyre-Walker 2007; Wright *et al.* 2009; Stoletzki and Eyre-Walker 2011). Positive selection complicates matters as the efficiency of selection in larger populations results in the retention of beneficial mutations, while in smaller populations beneficial mutations act more like neutral mutations (Charlesworth 2009). In addition, these studies have tended to use a limited number of markers; often a single mtDNA sequence, which is subject to the Hill-Robertson effect due to selection and lack of recombination (Charlesworth 2009). The demographic influence of population size on the gradient of molecular evolution remains unresolved, due to the continued debate and conflicting empirical evidence on the influence of population size and uncertainty about actual population size differences between high and low latitude species.

TOWARDS A SYNTHESIS

Rates of molecular evolution and speciation

Evidence for differential rates of molecular evolution being correlated with latitude are compelling yet they lack one key feature as a model to explain the biodiversity gradient; no study has yet convincingly linked differences in rates of molecular evolution to differences in speciation, extinction or immigration rates (Wright *et al.* 2006; Gillman *et al.* 2009; Gillman *et al.* 2010; Wright *et al.* 2010). Although the evidence for elevated rates of molecular evolution at the tropics is substantial, it remains only a pattern. Many traits that have been linked to the biodiversity gradient, such as niche conservatism, similarly describe a pattern (an outcome) rather than the process (driver) involved in generating the gradient (Wiens and Graham 2005; Losos 2008; Crisp and Cook 2012). Rates of molecular evolution and numbers of species show latitudinal clines but are they directly related to one another, or is a third independent variable implicated? For example, differences in branch length between high and low latitude phylogenetically independent angiosperms show higher rates of molecular evolution at low latitude, but this does not correlate with higher speciation rates. Instead, the two rates were each independently linked to an environmental variable (temperature or UV; Davies *et al.* 2004). An indirect link such as generation time might explain the observed relationship between rates of molecular evolution and biodiversity (Bromham 2011), or one of these two traits might drive the other. Contrary to the idea that elevated rates of molecular evolution lead to increased speciation, some studies infer that elevated molecular evolution is a result of more speciation, referred to as punctuational evolution (Pagel *et al.* 2006; Venditti and Pagel 2010). Thus, the increased rate of molecular evolution in the tropics may be a product of speciation rather than the cause of it (Webster *et al.* 2003; Pagel *et al.* 2006).

Evidence of a link between rates of molecular evolution and net-diversification has been reported (Barracough and Savolainen 2001; Jobson and Albert 2002; Eo and Dewoody 2010; Lancaster 2010; Lanfear *et al.* 2010a; but see Goldie *et al.* 2011 where no such link was found in mammals). However, none of these studies have conclusively shown that increased rates of molecular evolution are responsible for

elevated rates of speciation. Evidence that the rate of molecular evolution is driving speciation rate has been found in plants (Lancaster 2010), but the inference relies on the approach used to model extinction (Quental and Marshall 2010). The assumption that the plant clades studied evolved via a “birth-death” diversification process (Lancaster 2010) is untested, and it is possible that models of density-dependent speciation (Moran process), or pure birth (Yule process) fit the data better (Nee 2004). Also, it is unclear whether results from a multi-copy nuclear marker, that evolves via concerted evolution (ITS), can be meaningfully extrapolated to the rest of the genome. An extensive sequence dataset from extant birds (Lanfear *et al.* 2010a) identified a link between rates of molecular evolution and net-diversification, but birds are one of the few animal groups where a latitudinal skew in molecular evolution rates has not been observed (Bromham and Cardillo 2003), even though birds do contribute to the biodiversity gradient (Hawkins *et al.* 2003). Lanfear *et al.* (2010a) found that as dS and dN increased in bird lineages so did net-diversification, but they could not find a link between life history traits and an increased rate of molecular evolution. Within their dataset there was no association between net-diversification and ω (dN/dS), suggesting that selection and population size were not influencing net-diversification rates, but that the rate of molecular evolution was the most important influence on the total number of species within a clade. Nevertheless, this result does not exclude the possibility that speciation itself accelerated the rate of molecular evolution in these birds, or even that extinction (which contributes to net-diversification) is lower where molecular rate is high (Fig. 3).

Molecular evolution and extinction

As noted, analyses of molecular genetic data are limited by their inability to directly identify extinction and its effect (Weir and Schluter 2007; Lancaster 2010; Quental and Marshall 2010). Molecular studies are not well suited to examining extinction rates because extinct taxa are not available for sampling. Molecular phylogenies show inferred ancestral relationships among extant (sampled) taxa and do not show how extinction (or failed sampling) influences observed branch lengths. Varying rates of speciation and extinction can leave similar molecular phylogenetic patterns (Rabosky and Lovette 2008; Crisp and Cook 2009) so application of methods such as

lineage through time plots may be misleading. It is therefore difficult to disentangle the relative contribution to diversity of speciation and extinction without a good fossil record (Fig. 3; Quental and Marshall 2010).

Evidence from paleontological studies suggests a combination of lower extinction and higher speciation rates in the tropics has driven the biodiversity gradient (Jablonski *et al.* 2006; Martin *et al.* 2007). If rates of molecular evolution are faster in low latitudes as many have found (Davies *et al.* 2004; Wright *et al.* 2006; Gillman *et al.* 2009; Gillman *et al.* 2010; Wright *et al.* 2010) then the relationship identified in mammals and birds by Weir and Schluter (2007) is consistent, but the use of a fixed molecular clock has resulted in misleading inferences of relative speciation and extinction rates. A strict clock rate is inappropriate when there is a likelihood that rates of molecular evolution vary among lineages. Indeed, any variation in diversification rates among lineages, which is likely to be the rule rather than the exception, will interfere with estimation of true extinction rates (Rabosky 2010). Direct linkage of extinction rates to rates of molecular evolution has also been suggested; increased rates of molecular evolution could reduce extinction risk by increasing the number of beneficial mutations that enable lineages to persist longer (Lanfear *et al.* 2010a). The effect of mutation rate on extinction is largely unknown, although excessive mutation rates will tend to yield a high genetic load (Butlin 2009; Lancaster 2010).

Countless difficulties

“in the one [low latitude], evolution has had a fair chance; in the other [high latitude] it has had countless difficulties thrown in its way”.

Wallace 1878

The future role of molecular research investigating the biodiversity gradient is uncertain, as studies have proven to be method sensitive and open to variable interpretation. Considered individually, many molecular studies appear to provide compelling and relevant evidence, however, the associations they reveal rarely constitute direct evidence of a primary driver. Many studies of the LBG run parallel

to one another and cannot be directly compared because they employ very different methods. However, a consensus is developing that there is a correlation between rates of molecular evolution and species diversity; the challenge now is to identify their causal relationship.

Palaeontology has the potential to track diversity trends through time and is therefore well suited to estimate extinction, speciation and immigration rates. A recent test of the influence of seasonality on the biodiversity gradient shows the potential for palaeontology to narrow the field of possible drivers of the LBG (Archibald *et al.* 2010). Given that sampling of DNA sequence data is limited to extant species and there are known problems with estimating extinction rates from molecular data (Rabosky 2010), it might be argued that molecular data have a limited role in examining extinction rates, a processes central to the biodiversity gradient debate. However, analytical tools are available if appropriate questions are asked (Nee 2004), and our ability to incorporate extinction into phylogenetic models is improving (Etienne and Apol 2009; Morlon *et al.* 2011; Stadler 2011). The use of data from measurably evolving populations (e.g. in rapidly evolving disease causing microorganisms; and other populations using ancient DNA approaches) will continue to enhance our ability to quantify rates of molecular change and strengthen theoretical developments (Drummond *et al.* 2003).

Inferences about regional biodiversity, and choice of entities for data sampling are both influenced by current taxonomy. In some studies, unrecognised species probably influence the conclusions drawn. It is possible that differences among the extinction and speciation rates inferred in Weir and Schluter (2007) were, in part, artefacts of more intense taxonomic splitting of high latitude versus low latitude biota (Tobias *et al.* 2008). Higher within-species genetic diversity has been found in plants and vertebrates from lower latitudes, and this could be taken to mean that there are more undescribed species in the tropics, or that the tropical taxa have larger population sizes (Martin and McKay 2004; Eo *et al.* 2008). Incorporation of phylogenetic information into estimates of biodiversity (phylogenetic distinctiveness metrics) is increasing (Purvis and Hector 2000; Davies and Buckley 2011), but phylogenetic approaches to estimating biodiversity rely in part on the

same information (i.e. branch length and branching pattern) being used to assess rates of molecular evolution. This circularity weakens inferences about drivers of the LBG that might be further influenced by lineage extinction and local biogeographic history.

Mitochondrial DNA has been used in several studies comparing rates of molecular evolution among lineages with speciation rates. There is some evidence that molecular evolution rates from the mitochondrial genome and the nuclear genome are subject to different pressures that might reflect a difference in proximity to cell metabolic processes or natural selection (Welch *et al.* 2008). Mitochondrial DNA is frequently used in analyses that assume neutrality in terms of molecular evolution, but this is unlikely to be realistic (for review see Galtier *et al.* 2009b). Given the metabolic role of mitochondria, the mitochondrial genome could be implicated in climate adaptation, and thus far from neutral in the context of the LBG. This idea has been mooted for human mitochondrial genetics (Ruiz-Pesini *et al.* 2004; Wallace 2005; Das 2006; Pierron *et al.* 2011), but remains controversial given suggestions that available methods to detect selection within populations are flawed (Kryazhimskiy and Plotkin 2008). Nuclear “loci” such as ITS are far from ideal given their complex and poorly understood process of sequence evolution (Elder and Turner 1995). This could be particularly problematic in plants where hybridisation plays an important role in species formation and exchange among lineages (Hegarty and Hiscock 2005). To date, selection of DNA sequence loci has been strongly influenced by the availability of data and universal PCR primers, a problem which is being alleviated by the use of Next Generation Sequencing (NGS) technology.

FUTURE FOCUS ON THE PROCESS

It is clear that we need approaches that avoid reliance on uncertain proxies. Tests of associations between putative drivers and speciation rates currently rely heavily on substitutes that are assumed to correlate well with the characters of interest. For example, island size is used as a proxy for population size; metabolic rate is used as a proxy for oxygen free radicals that can influence mutations; average age of first

reproduction is a proxy for average species' generation time; and current species distribution is used as a proxy for ancestral species distribution. Many of these correlations first need to be independently verified in the organisms of interest. Also, the design of many studies reflect methodological and sampling limitations, and often dealt with comparison of just two alternative conditions (e.g. big population vs. small population, high latitude vs. low latitude) rather than analysing gradient or transect data. Although continuous traits are often measured (e.g. population size) they are then categorised into two alternatives (e.g. big vs. small) in order for inferences to be drawn from paired comparisons (e.g. sign tests), thus a gradient in biodiversity or population size is transformed into just two alternative conditions. Mutation rates vary between species, between individuals and within the genome itself, so disentangling all the drivers of mutations rates is unlikely to be easy (Hodgkinson and Eyre-Walker 2011).

Clarity of assumptions and experimental design and application of methods at the appropriate evolutionary level are essential, and here we suggest some directions that might take us forward.

1. Many of the life-history traits that have been implicated as driving rate changes are not independent of one another and work needs to focus on taxa that allow the elimination of some of these variables. The sister species approach has some benefits over the large clade approach, but for life-history traits that can not be adequately represented across sister species pairs, additional methods will be necessary (see Lanfear *et al.* 2010b for an examination of different approaches to estimate rates from trees). Quantification of life history traits in specific organisms of interest is required to avoid reliance on rules of thumb and generalised concepts that are often based on limited direct evidence. An emphasis on regionally focused examples should enhance opportunities for gathering accurate life history data and inclusion of specific geophysical information when accommodating effects of local biogeographic history.
2. Until recently some studies of rates of molecular evolution across the LBG have used single genes from the mtDNA genome and a single branch from each species. NGS (Next Generation Sequencing) technologies now provide

multilocus data that improve confidence in estimates of relative rates of molecular evolution across the genome in sister species comparisons, even for non-model organisms. In the past, when more than one sequence was available per species the shortest branch was routinely chosen (e.g. Wright *et al.* 2009), even though species usually contain measurable genetic diversity. NGS allows dense population sampling of sister species pairs that in turn allows for analysis of the effects of choosing the shortest branch versus other metrics such as average branch length, or random branch selection. Studies, particularly those involving high genetic divergence, may also benefit from an out-group comparison (Hugall and Lee 2007). Population level studies have the potential to inform on speciation processes at a stage equivalent to the node in a species tree, instead of relying on inferences about past speciation from tree tips.

3. For understanding the LBG, determining population size variation between latitudes is paramount. Thus a focus on multi-locus data for taxa with sedentary lifestyles would facilitate population density comparisons. Should there exist consistent differences in population size among latitudes then the interaction between population size and rates of molecular evolution becomes extremely important to the LBG debate. Multi-loci datasets exist for model organisms (e.g. humans and drosophila; Bakewell *et al.* 2007; Kosiol *et al.* 2008; Petit and Barbadilla 2009), but it is now possible for substantial tests across the nuclear and mitochondrial genomes functional and non-functional DNA in non-model groups with NGS. However, population size may not be constant through time and methods for inferring past population size that are independent of rates of molecular evolution will be required.
4. If speciation occurs predominantly in small isolated populations (Venditti and Pagel 2010), signals about the size of the populations during speciation may be lost through processes such as back-mutations during subsequent population expansion (Charlesworth and Eyre-Walker 2007; Lanfear *et al.* 2010a). Simulations using data from model organisms should make it possible to test the predicted population size effect on ω and assess the interaction of ecological speciation and population expansion. It will also be possible to estimate for how long after a speciation event the population

genetic signature remains detectable. Analysis of recent past population diversity using ancientDNA methods could enhance modelling of rates of molecular evolution and population size. Similarly, studies involving ancestral gene reconstruction and resurrection will likely illuminate our understanding of the respective roles of genetic drift and selection on gene evolution (Thornton 2004).

5. Expansion of studies involving lineages represented in the fossil record and neobiota will contribute broadly to meshing information on speciation, extinction and migration with rates of molecular evolution. In particular, such data will help explore the directionality of the relationship between rates of molecular evolution and speciation. Sequencing multiple loci from extant species that are members of fossil-rich lineages will allow extinction rates to be correlated with both speciation and molecular evolution rates. This has the potential to isolate speciation from net-diversification, an important step towards resolving the link between rates of molecular evolution and speciation. Fossil data can be used to constrain molecular clock analyses and allow rate comparisons that are not limited to sister species approaches. Measures of extinction could reveal any relationship between rates of molecular evolution and extinction.
6. Rate of molecular evolution is determined by mutation rate and fixation rate, and direct estimates of mutation rates are now available with NGS. Sequencing of parents and their offspring allows species specific estimates of mutation rate, and if the necessary multiple comparisons and multiple tissues are included, the population variance in individual *de novo* mutation rate can be known (Hodgkinson and Eyre-Walker 2011).

Two experimental approaches could contribute directly to study of the LBG. First, the effect of temperature and UV on mutation rate could be assessed using lab-based model organisms in controlled environments. Second, direct comparison of mutation rates of non-model organisms where rates of molecular evolution have been shown to differ between sister species at separate latitudes can be obtained from whole genome parent/offspring data. This would allow discrimination between the two components of molecular evolution; variation in rates of mutation and variation in rates of fixation (Ho

et al. 2005). In addition, mining pedigree datasets that exist for applied research in agriculture, horticulture and conservation will allow tests for a link between longevity of individuals within species and mutation rate.

Differences in rates of molecular evolution between global regions is fascinating and warrants much more focused research. But perhaps the priority should be to find the directionality of the link, if any, between speciation and molecular evolution. Without this fundamental information the evolutionary speed hypothesis cannot provide compelling predictions to test putative drivers of the latitudinal biodiversity gradient.

ACKNOWLEDGEMENTS

We would like to thank two anonymous reviewers, Robert Lanfear, Roger Butlin and Gillian Gibbs for very helpful comments on earlier versions of this review. The Phoenix group provided useful discussions (evolves@massey.ac.nz), and Massey University and the Royal Society of New Zealand provided funding (Scholarship to EJD; Marsden 04-GNS-021 to MMR).

REFERENCES

- Ahmad I, Day JP, Macdonald MV, Ingram DS (1991) Haploid culture and UV mutagenesis in rapid-cycling *Brassica napus* for the generation of resistance to Chlorsulfuron and *Alternaria brassicicola*. *Annals of Botany* **67**, 521-519.
- Algar AC, Kerr JT, Currie DJ (2007) A test of Metabolic Theory as the mechanism underlying broad-scale species-richness gradients. *Global Ecology and Biogeography* **16**, 170-178.
- Allen AP, Brown JH, Gillooly JF (2002) Global biodiversity, biochemical kinetics, and the energetic-equivalence rule. *Science* **297**, 1545-1548.
- Allen AP, Gillooly JF (2006) Assessing latitudinal gradients in speciation rates and biodiversity at the global scale. *Ecology Letters* **9**, 947-954.
- Allen AP, Gillooly JF, Savage VM, Brown JH (2006) Kinetic effects of temperature on rates of genetic divergence and speciation. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 9130-9135.
- Archibald SB, Bossert WH, Greenwood DR, Farrell BD (2010) Seasonality, the latitudinal gradient of diversity, and Eocene insects. *Paleobiology* **36**, 374-398.
- Ashton KG, Tracy MC, De Queiroz A (2000) Is Bergmann's rule valid for mammals? *American Naturalist* **156**, 390-415.
- Bakewell MA, Shi P, Zhang J (2007) More genes underwent positive selection in chimpanzee evolution than in human evolution. *Proceedings of the National Academy of Sciences* **104**, 7489-7494.
- Barraclough TG, Savolainen V (2001) Evolutionary rates and species diversity in flowering plants. *Evolution* **55**, 677-683.
- Barry RG (2008) *Mountain weather and climate*. Cambridge University Press, New York.
- Bromham L (2011) The genome as a life-history character: why rate of molecular evolution varies between mammal species. *Philosophical Transactions of the Royal Society B: Biological Sciences* **366**, 2503-2513.
- Bromham L, Cardillo M (2003) Testing the link between the latitudinal gradient in species richness and rates of molecular evolution. *Journal of Evolutionary Biology* **16**, 200-207.
- Butlin R, Bridle J., Schluter, D. (2009) *Speciation and Patterns of Diversity*. Cambridge University Press, New York.
- Charlesworth B (2009) Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. *Nature reviews. Genetics* **10**, 195-205.

- Charlesworth J, Eyre-Walker A (2007) The other side of the nearly neutral theory, evidence of slightly advantageous back-mutations. *Proceedings of the National Academy of Sciences* **104**, 16992-16997.
- Chown SL, Gaston KJ (2000) Areas, cradles and museums: the latitudinal gradient in species richness. *Trends in Ecology & Evolution* **15**, 311-315.
- Clarke A, Gaston KJ (2006) Climate, energy and diversity. *Proceedings of the Royal Society B-Biological Sciences* **273**, 2257-2266.
- Condamine FL, Sperling FaH, Wahlberg N, Rasplus J-Y, Kergoat GJ (2012) What causes latitudinal gradients in species diversity? Evolutionary processes and ecological constraints on swallowtail biodiversity. *Ecology Letters* **15**, 267-277.
- Crame JA (2002) Evolution of taxonomic diversity gradients in the marine realm: a comparison of Late Jurassic and Recent bivalve faunas. *Paleobiology* **28**, 184-207.
- Crampton JS, Cooper RA, Beu AG, Foote M, Marshall BA (2010) Biotic influences on species duration: interactions between traits in marine molluscs. *Paleobiology* **36**, 204-223.
- Crampton JS, Foote M, Beu AG, Maxwell PA, Cooper RA *et al.* (2006) The ark was full! Constant to declining Cenozoic shallow marine biodiversity on an isolated midlatitude continent. *Paleobiology* **32**, 509-532.
- Crisp MD, Cook LG (2009) Explosive radiation or cryptic mass extinction? Interpreting signatures in molecular phylogenies. *Evolution* **63**, 2257-2265.
- Crisp MD, Cook LG (2012) Phylogenetic niche conservatism: what are the underlying evolutionary and ecological causes? *New Phytologist* **196**, 681-694.
- Crisp MD, Treweek SA, Cook LG (2011) Hypothesis testing in biogeography. *Trends in Ecology & Evolution* **26**, 66-72.
- Currie DJ, Kerr JT (2008) Tests of the mid-domain hypothesis: A review of the evidence. *Ecological Monographs* **78**, 3-18.
- Currie DJ, Mittelbach GG, Cornell HV, Field R, Guegan JF *et al.* (2004) Predictions and tests of climate-based hypotheses of broad-scale variation in taxonomic richness. *Ecology Letters* **7**, 1121-1134.
- Das J (2006) The role of mitochondrial respiration in physiological and evolutionary adaptation. *BioEssays* **28**, 890-901.
- Davies TJ, Buckley LB (2011) Phylogenetic diversity as a window into the evolutionary and biogeographic histories of present-day richness gradients for mammals. *Philosophical Transactions of the Royal Society B: Biological Sciences* **366**, 2414-2425.
- Davies TJ, Savolainen V, Chase MW, Moat J, Barraclough TG (2004) Environmental energy and evolutionary rates in flowering plants. *Proceedings of the Royal Society of London Series B-Biological Sciences* **271**, 2195-2200.
- De Castro F, Gaedke U (2008) The metabolism of lake plankton does not support the metabolic theory of ecology. *Oikos* **117**, 1218-1226.
- Drummond AJ, Pybus OG, Rambaut A, Forsberg R, Rodrigo AG (2003) Measurably evolving populations. *Trends in Ecology & Evolution* **18**, 481-488.
- Elder JF, Jr., Turner BJ (1995) Concerted Evolution of Repetitive DNA Sequences in Eukaryotes. *The Quarterly Review of Biology* **70**, 297-320.
- Ellstrand NC, Elam DR (1993) Population Genetic Consequences of Small Population Size: Implications for Plant Conservation. *Annual Review of Ecology and Systematics* **24**, 217-242.

- Eo SH, Dewoody JA (2010) Evolutionary rates of mitochondrial genomes correspond to diversification rates and to contemporary species richness in birds and reptiles. *Proceedings of the Royal Society B: Biological Sciences* **277**, 3587-3592.
- Eo SH, Wares JP, Carroll JP (2008) Population divergence in plant species reflects latitudinal biodiversity gradients. *Biology Letters* **4**, 382-384.
- Epel D, Hemela K, Shick M, Patton C (1999) Development in the Floating World: Defenses of Eggs and Embryos Against Damage from UV Radiation. *American Zoologist* **39**, 271-278.
- Erwin DH (2009) Climate as a Driver of Evolutionary Change. *Current Biology* **19**, R575-R583.
- Etienne RS, Apol MEF (2009) Estimating speciation and extinction rates from diversity data and the fossil record. *Evolution* **63**, 244-255.
- Flenley JR (2011) Why is pollen yellow? And why are there so many species in the tropical rain forest? *Journal of Biogeography* **38**, 809-816.
- Funk DJ, Nosil P, Etges WJ (2006) Ecological divergence exhibits consistently positive associations with reproductive isolation across disparate taxa. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 3209-3213.
- Galtier N, Jobson RW, Nabholz B, Glemin S, Blier PU (2009a) Mitochondrial whims: metabolic rate, longevity and the rate of molecular evolution. *Biology Letters* **5**, 413-416.
- Galtier N, Nabholz B, Glémin S, Hurst GDD (2009b) Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular Ecology* **18**, 4541-4550.
- Gaston KJ (2000) Global patterns in biodiversity. *Nature* **405**, 220-227.
- Gaut B, Yang L, Takuno S, Eguiarte L (2011) The Patterns and Causes of Variation in Plant Nucleotide Substitution Rates. *Annual Review of Ecology and Systematics* **42**, 245-266.
- Gillman LN, Keeling DJ, Gardner RC, Wright SD (2010) Faster evolution of highly conserved DNA in tropical plants. *Journal of Evolutionary Biology* **23**, 1327-1330.
- Gillman LN, Keeling DJ, Ross HA, Wright SD (2009) Latitude, elevation and the tempo of molecular evolution in mammals. *Proceedings of the Royal Society B-Biological Sciences* **276**, 3353-3359.
- Gillooly JF, Allen AP (2007) Linking global patterns in biodiversity to evolutionary dynamics using metabolic theory. *Ecology* **88**, 1890-1894.
- Gillooly JF, Mccoy MW, Allen AP (2007) Effects of metabolic rate on protein evolution. *Biology Letters* **3**, 655-659.
- Goldie X, Lanfear R, Bromham L (2011) Diversification and rate of molecular evolution: no evidence of a link in mammals. *BMC Evolutionary Biology* **11**.
- Guo QF, Ricklefs RE (2000) Species richness in plant genera disjunct between temperate eastern Asia and North America. *Botanical Journal of the Linnean Society* **134**, 401-423.
- Hawkins BA (2004) Are we making progress toward understanding the global diversity gradient? *Basic and Applied Ecology* **5**, 1-3.
- Hawkins BA, Albuquerque FS, Araujo MB, Beck J, Bini LM *et al.* (2007a) A global evaluation of metabolic theory as an explanation for terrestrial species richness gradients. *Ecology* **88**, 1877-1888.

- Hawkins BA, Diniz JaF, Bini LM, Araujo MB, Field R *et al.* (2007b) Metabolic theory and diversity gradients: Where do we go from here? *Ecology* **88**, 1898-1902.
- Hawkins BA, Diniz JaF, Jaramillo CA, Soeller SA (2006) Post-Eocene climate change, niche conservatism, and the latitudinal diversity gradient of New World birds. *Journal of Biogeography* **33**, 770-780.
- Hawkins BA, Field R, Cornell HV, Currie DJ, Guegan JF *et al.* (2003) Energy, water, and broad-scale geographic patterns of species richness. *Ecology* **84**, 3105-3117.
- Hegarty MJ, Hiscock SJ (2005) Hybrid speciation in plants: new insights from molecular studies. *New Phytologist* **165**, 411-423.
- Hillebrand H (2004) On the generality of the latitudinal diversity gradient. *American Naturalist* **163**, 192-211.
- Ho SYW, Phillips MJ, Cooper A, Drummond AJ (2005) Time Dependency of Molecular Rate Estimates and Systematic Overestimation of Recent Divergence Times. *Molecular Biology and Evolution* **22**, 1561-1568.
- Hodgkinson A, Eyre-Walker A (2011) Variation in the mutation rate across mammalian genomes. *Nat Rev Genet* **12**, 756-766.
- Hugall AF, Lee MSY (2007) The likelihood node density effect and consequences for evolutionary studies of molecular rates. *Evolution* **61**, 2293-2307.
- Irlich UM, Terblanche JS, Blackburn TM, Chown SL (2009) Insect Rate-Temperature Relationships: Environmental Variation and the Metabolic Theory of Ecology. *American Naturalist* **174**, 819-835.
- Jablonski D, Roy K, Valentine JW (2006) Out of the tropics: Evolutionary dynamics of the latitudinal diversity gradient. *Science* **314**, 102-106.
- Jobson RW, Albert VA (2002) Molecular rates parallel diversification contrasts between carnivorous plant sister lineages. *Cladistics-the International Journal of the Willi Hennig Society* **18**, 127-136.
- Johnson KP, Seger J (2001) Elevated rates of nonsynonymous substitution in island birds. *Molecular Biology and Evolution* **18**, 874-881.
- Kimura M (1968) Evolutionary rate at the molecular level. *Nature* **217**, 624-626.
- Korner C (2007) The use of 'altitude' in ecological research. *Trends in Ecology & Evolution* **22**, 569-574.
- Kosiol C, Vinar T, Da Fonseca RR, Hubisz MJ, Bustamante CD *et al.* (2008) Patterns of Positive Selection in Six Mammalian Genomes. *PLoS Genet* **4**, e1000144.
- Krug AZ, Jablonski D, Valentine JW, Roy K (2009) Generation of Earth's First-Order Biodiversity Pattern. *Astrobiology* **9**, 113-124.
- Kryazhimskiy S, Plotkin JB (2008) The Population Genetics of dN/dS. *PLoS Genet* **4**, e1000304.
- Lancaster L (2010) Molecular evolutionary rates predict both extinction and speciation in temperate angiosperm lineages. *BMC Evolutionary Biology* **10**, 162.
- Lanfear R, Ho SYW, Love D, Bromham L (2010a) Mutation rate is linked to diversification in birds. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 20423-20428.
- Lanfear R, Thomas JA, Welch JJ, Brey T, Bromham L (2007) Metabolic rate does not calibrate the molecular clock. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 15388-15393.

- Lanfear R, Welch JJ, Bromham L (2010b) Watching the clock: Studying variation in rates of molecular evolution between species. *Trends in Ecology & Evolution* **25**, 495-503.
- Langerhans RB, Gifford ME, Joseph EO (2007) Ecological Speciation in *Gambusia* Fishes. *Evolution* **61**, 2056-2074.
- Losos JB (2008) Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. *Ecology Letters* **11**, 995-1003.
- Losos JB, Schluter D (2000) Analysis of an evolutionary species-area relationship. *Nature* **408**, 847-850.
- Macarthur RH, Wilson EO (1967) *The theory of island biogeography*. Princeton University Press, Princeton, N. J.
- Martin PR, Bonier F, Tewksbury JJ (2007) Revisiting Jablonski (1993): cladogenesis and range expansion explain latitudinal variation in taxonomic richness. *Journal of Evolutionary Biology* **20**, 930-936.
- Martin PR, Mckay JK (2004) Latitudinal variation in genetic divergence of populations and the potential for future speciation. *Evolution* **58**, 938-945.
- Maurer G, Portugal SJ, Cassey P (2011) Review: an embryo's eye view of avian eggshell pigmentation. *Journal of Avian Biology* **42**, 494-504.
- Mayhew PJ, Jenkins GB, Benton TG (2008) A long-term association between global temperature and biodiversity, origination and extinction in the fossil record. *Proceedings of the Royal Society B-Biological Sciences* **275**, 47-53.
- Mittelbach GG, Schemske DW, Cornell HV, Allen AP, Brown JM *et al.* (2007) Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. *Ecology Letters* **10**, 315-331.
- Morlon HLN, Parsons TL, Plotkin JB (2011) Reconciling molecular phylogenies with the fossil record. *Proceedings of the National Academy of Sciences* **108**, 16327-16332.
- Nabholz B, Glemin S, Galtier N (2008) Strong variations of mitochondrial mutation rate across mammals - the longevity hypothesis. *Molecular Biology and Evolution* **25**, 120-130.
- Nabholz B, Glemin S, Galtier N (2009) The erratic mitochondrial clock: variations of mutation rate, not population size, affect mtDNA diversity across birds and mammals. *Bmc Evolutionary Biology* **9**.
- Nee S (2004) Extinct meets extant: simple models in paleontology and molecular phylogenetics. *Paleobiology* **30**, 172-178.
- Nikolaev SI, Montoya-Burgos JI, Popadin K, Parand L, Margulies EH *et al.* (2007) Life-history traits drive the evolutionary rates of mammalian coding and noncoding genomic elements. *Proceedings of the National Academy of Sciences* **104**, 20443-20448.
- Nosil P, Flaxman SM (2011) Conditions for mutation-order speciation. *Proceedings of the Royal Society B: Biological Sciences* **278**, 399-407.
- Ohta T (1992) The Nearly Neutral Theory of Molecular Evolution. *Annual Review of Ecology and Systematics* **23**, 263-286.
- Pagel M, Venditti C, Meade A (2006) Large punctuational contribution of speciation to evolutionary divergence at the molecular level. *Science* **314**, 119-121.
- Petit N, Barbadilla A (2009) Selection efficiency and effective population size in *Drosophila* species. *Journal of Evolutionary Biology* **22**, 515-526.

- Pierron D, Chang I, Arachiche A, Heiske M, Thomas O *et al.* (2011) Mutation Rate Switch inside Eurasian Mitochondrial Haplogroups: Impact of Selection and Consequences for Dating Settlement in Europe. *PLoS ONE* **6**, e21543.
- Purvis A, Hector A (2000) Getting the measure of biodiversity. *Nature* **405**.
- Quental TB, Marshall CR (2010) Diversity dynamics: molecular phylogenies need the fossil record. *Trends in ecology & evolution* **25**, 434-441.
- Rabosky DL (2010) Extinction rates should not be estimated from molecular phylogenies. *Evolution* **64**, 1816-1824.
- Rabosky DL, Lovette IJ (2008) Explosive evolutionary radiations: decreasing speciation or increasing extinction through time? *Evolution* **62**, 1866-1875.
- Rastogi RP, Richa, Kumar A, Tyagi MB, Sinha RP (2010) Molecular Mechanisms of Ultraviolet Radiation-Induced DNA Damage and Repair. *Journal of Nucleic Acids* **2010**.
- Riek A (2008) Relationship between field metabolic rate and body weight in mammals: effect of the study. *Journal of Zoology* **276**, 187-194.
- Rohde K (1992) Latitudinal Gradients in Species-Diversity - the Search for the Primary Cause. *Oikos* **65**, 514-527.
- Rosenzweig ML (1995) *Species diversity in space and time*. Cambridge University Press, Cambridge, UK.
- Roy K, Jablonski D, Valentine JW, Rosenberg G (1998). Marine latitudinal diversity gradients: tests of causal hypotheses. *Proceedings of the national Academy of Sciences* **95** (7), 3699-3702
- Roy K, Goldberg EE (2007) Origination, extinction, and dispersal: Integrative models for understanding present-day diversity gradients. *American Naturalist* **170**, S71-S85.
- Ruiz-Pesini E, Mishmar D, Brandon M, Procaccio V, Wallace DC (2004) Effects of Purifying and Adaptive Selection on Regional Variation in Human mtDNA. *Science* **303**, 223-226.
- Rundle HD, Nagel L, Boughman JW, Schluter D (2000) Natural Selection and Parallel Speciation in Sympatric Sticklebacks. *Science* **287**, 306-308.
- Schluter D (2009) Evidence for Ecological Speciation and Its Alternative. *Science* **323**, 737-741.
- Smith SA, Donoghue MJ (2008) Rates of Molecular Evolution Are Linked to Life History in Flowering Plants. *Science* **322**, 86-89.
- Smith TB, 2011 Molecular evolution and historical biogeography of new world birds, pp. in *College of Sciences*. University of Nevada, Las Vegas.
- Stadler T (2011) Inferring speciation and extinction processes from extant species data. *Proceedings of the National Academy of Sciences* **108**, 16145-16146.
- Stebbins GL (1974) *Flowering Plants: Evolution above the Species Level*. The Belknap Press of Harvard University Press, Cambridge, Massachusetts.
- Stoletzki N, Eyre-Walker A (2011) The Positive Correlation between dN/dS and dS in Mammals Is Due to Runs of Adjacent Substitutions. *Molecular Biology and Evolution* **28**, 1371-1380.
- Storch D (2003) Comment on "Global Biodiversity, Biochemical Kinetics, and the Energetic-Equivalence Rule". *Science* **299**, 346.
- Thomas JA, Welch JJ, Lanfear R, Bromham L (2010) A Generation Time Effect on the Rate of Molecular Evolution in Invertebrates. *Molecular Biology and Evolution* **27**, 1173-1180.

- Thornton JW (2004) Resurrecting ancient genes: experimental analysis of extinct molecules. *Nat Rev Genet* **5**.
- Tobias JA, Bates JM, Hackett SJ, Seddon N (2008) Comment on "The Latitudinal Gradient in Recent Speciation and Extinction Rates of Birds and Mammals". *Science* **319**, 901.
- Valentine JW, Jablonski D, Krug AZ, Roy K (2008) Incumbency, diversity, and latitudinal gradients. *Paleobiology* **34**, 169-178.
- Venditti C, Pagel M (2010) Speciation as an active force in promoting genetic evolution. *Trends in Ecology and Evolution* **25**, 14-20.
- Wallace AR (1876) *The geographical distribution of animals. With a study of the relations of living and extinct faunas as elucidating the past changes of the earth's surface*. Harper & brothers, New York.
- Wallace DC (2005) A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: a dawn for evolutionary medicine. *Annual Review of Genetics* **39**, 359-407.
- Webster AJ, Payne RJH, Pagel M (2003) Molecular Phylogenies link rates of evolution and speciation. *Science* **301**, 478-478.
- Weir JT, Schluter D (2007) The latitudinal gradient in recent speciation and extinction rates of birds and mammals. *Science* **315**, 1574-1576.
- Weir JT, Schluter D (2011) Are rates of molecular evolution in mammals substantially accelerated in warmer environments? *Proceedings of the Royal Society B: Biological Sciences* **278**, 1291-1293.
- Welch J, Bininda-Emonds O, Bromham L (2008) Correlates of substitution rate variation in mammalian protein-coding sequences. *BMC Evolutionary Biology* **8**, 53.
- Whittle CA, Johnston MO (2003) Broad-scale analysis contradicts the theory that generation time affects molecular evolutionary rates in plants. *Journal of Molecular Evolution* **56**, 223-233.
- Wiens JJ, Graham CH (2005) Niche Conservatism: Integrating Evolution, Ecology, and Conservation Biology. *Annual Review of Ecology, Evolution, and Systematics* **36**, 519-539.
- Wilson Sayres MA, Makova KD (2011) Genome analyses substantiate male mutation bias in many species. *BioEssays* **33**, 938-945.
- Woolfit M, Bromham L (2005) Population size and molecular evolution on islands. *Proceedings of the Royal Society B: Biological Sciences* **272**, 2277-2282.
- Wright DH (1983) Species-Energy Theory - an Extension of Species-Area Theory. *Oikos* **41**, 496-506.
- Wright S, Keeling J, Gillman L (2006) The road from Santa Rosalia: A faster tempo of evolution in tropical climates. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 7718-7722.
- Wright S, Ross H, Jeanette Keeling D, McBride P, Gillman L (2011) Thermal energy and the rate of genetic evolution in marine fishes. *Evolutionary Ecology* **25**, 525-530.
- Wright SD, Gillman LN, Ross HA, Keeling DJ (2009) Slower Tempo of Microevolution in Island Birds: Implications for Conservation Biology. *Evolution* **63**, 2275-2287.
- Wright SD, Gillman LN, Ross HA, Keeling JD (2010) Energy and the tempo of evolution in amphibians. *Global Ecology and Biogeography*.

Zapata FA, Gaston KJ, Chown SL (2003) Mid-domain models of species richness gradients: assumptions, methods and evidence. *Journal of Animal Ecology* **72**, 677-690.



MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL

**STATEMENT OF CONTRIBUTION
TO DOCTORAL THESIS CONTAINING PUBLICATIONS**

(To appear at the end of each thesis chapter/section/appendix submitted as an article/paper or collected as an appendix at the end of the thesis)

We, the candidate and the candidate's Principal Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of Candidate: Edwina Dowle

Name/Title of Principal Supervisor: Steve Trewick

Name of Published Research Output and full reference:

Dowle, E. J. Morgan-Richards, M & Trewick, S. T. (2013). Molecular evolution and the latitudinal biodiversity gradient. *Heredity*. 110(6), 501-510

In which Chapter is the Published Work: Chapter 1

Please indicate either:

- The percentage of the Published Work that was contributed by the candidate 80
and / or
- Describe the contribution that the candidate has made to the Published Work:
I conducted most of the background research and reviewing of the work surrounding the LBG and wrote the initial manuscript draft. Steve, Mary and I subsequently edited the draft, with input from reviewers and the journal editor.

Edwina Dowle
Digitally signed by Edwina Dowle
 DN: cn=Edwina Dowle, o=Massey, ou,
 email=e.j.dowle@massey.ac.nz, c=NZ
 Date: 2013.09.23 10:19:43 +1200
 Candidate's Signature

18/9/13

Date

Steve Trewick
Digitally signed by Steve Trewick
 DN: cn=Steve Trewick, o=Massey University,
 ou=Ecology Group,
 email=s.trewick@massey.ac.nz, c=NZ
 Date: 2013.09.27 12:00:32 +1200
 Principal Supervisor's signature

18/9/13

Date

2. ESTIMATES OF RATE OF MOLECULAR EVOLUTION ACROSS THE MITOCHONDRIAL GENOME USING SISTER-SPECIES COMPARISONS ARE NOT CONSISTENT: TESTS WITH CRICKETS AND FLIES

ABSTRACT

The assumption is often made that, when we detect a difference in rate of molecular evolution between species in one gene, this represents a rate shift across the genome in its entirety. Here we assess that assumption within the mitochondria genome with surprising results. We examined groups with contrasting levels of divergence, Crickets (Ensifera) and Flies (*Drosophila*), to test for skews in relative rate of evolution in the mitochondrial genome. Variation in rates of molecular evolution was estimated using the popular sister-species approach, and we found poor consistency among mitochondrial genes. We also found that outgroup selection could influence the rate estimate differences and that expanding out from a four-taxon tree did not clarify rates in distantly related taxa. These results suggest that our current methods for detecting rate shifts between species cannot be extrapolated to other parts of the genome, and are sensitive to taxon selection.

Dowle EJ, Morgan-Richards M & Trewick SA

INTRODUCTION

Rate of molecular evolution is a product of genetic mutation and fixation rate in a population. Estimating this rate and detecting variation among species/lineages is now a common objective of molecular analyses (Woolfit and Bromham 2005; Galtier *et al.* 2009a; Lanfear *et al.* 2010a; Gillman *et al.* 2012). Estimates of rate of molecular evolution (from here referred to as rate) are used to infer the role of biotic and abiotic factors in determining the mutation and fixation rates (Algar *et al.* 2007; Wright *et al.* 2009; Thomas *et al.* 2010). Detection of differences in rate among lineages usually involves one of two approaches: sister-species analysis or dated phylogenetic analysis (Nabholz *et al.* 2008; Gillman *et al.* 2009; Lanfear *et al.* 2010b), but we have little knowledge about how reliable observed differences in rate are (Smith and Eyre-Walker 2003). Here we employ the sister-species method that is often used in meta-comparisons of rate variation. We examine the ability to detect rate shifts and the stability of the relationships of the rate shifts detected across genes, using two groups, Crickets (Ensifera) and Flies (*Drosophila*), with contrasting levels of divergence.

Sister-species analysis is an attractive proposition because it requires the least prior knowledge of tree topology and avoids the node-density effect. As it does not require the use of molecular clock calibrations within the phylogeny, sister-species analysis is useful for groups with unknown or poor calibration points (Hugall and Lee 2007; Lanfear *et al.* 2010b). The method involves analysis of DNA sequences from a pair of ingroup species with a known outgroup. It is expected that, if the rates on the two ingroup lineages are equal, then the genetic distance from each species to their common ancestor (tip to node) should be near identical (Fig. 1). The branch length (or edge weight) provides a measure of the amount of molecular evolution since the most recent common ancestor, whose position is determined using a suitable outgroup (Fig. 1). A difference in branch length from the common ancestor to each tip taxon results from different rates of molecular evolution; a longer branch implies a faster rate. In strict terms this method provides a gene specific estimate of relative rates, but its application in many evolutionary studies relies on the tacit (but often unstated) assumption that the results reflect a genome wide phenomenon.

The sister-species approach has been widely employed to test for rate variation under several high profile hypotheses, although in many studies relatively small amounts of mitochondrial sequence data are used for each comparison. One of the hypotheses to explain the well documented latitudinal gradient in biodiversity is that species evolve faster in the tropics than in temperate regions; this is the evolutionary speed hypothesis (Dowle *et al.* 2013). Sister-species analysis has demonstrated faster rates in tropical or warmer regions compared to temperate or colder regions in several animal groups by examining numerous sister-species pairs using DNA sequences; mammals using *Cytb* (Gillman *et al.* 2009), birds using *Cytb* (Gillman *et al.* 2012), plants using 18S and ITS (Wright *et al.* 2006; Gillman *et al.* 2010), amphibians using 12S and 16S (Wright *et al.* 2010) and fish using *Cytb* or 12S and 16S (Wright *et al.* 2011). Similarly, rates of molecular evolution are thought to be influenced by population size (Ohta 1992). Although mutation rate is not expected to differ between small and large population sizes, the fixation rate will; smaller populations are less efficient at removing deleterious mutations than larger populations. Several studies using various methods have found evidence supporting this prediction (Woolfit and Bromham 2005; Bakewell *et al.* 2007; Kosiol *et al.* 2008; Eyre-Walker and Keightley 2009; Petit and Barbadilla 2009). One sister-species analysis found contrary results using the mtDNA *Cytb* gene in birds (Wright *et al.* 2009). A positive association between rate and generation time has been found using sister-species analysis of whole mtDNA and two nuclear rRNA genes (18S and 28S) in invertebrates (Thomas *et al.* 2010). Additionally, the fundamental question of whether increased rates in molecular evolution are linked to increased net species diversity has been examined with a positive relationship found in birds using a 19 nuclear gene dataset (Lanfear *et al.* 2010a) and in Proteaceae plants using up to 6 chloroplast genes per comparison (Duchene and Bromham 2013). In contrast, no relationship has been found in mammals using mitochondrial genomes (Goldie *et al.* 2011). With so many higher level evolutionary questions being tested using this approach its validation is both timely and necessary.

Sister-species analysis is frequently applied where knowledge of tree topology is poor or not all relatives are sampled, and frequently only a subset of potential outgroup taxa (generally two) is used. The resulting four-taxon tree has some useful analytical properties and is particularly helpful when examining close species pairs

from poorly resolved groups. The influence of outgroup choice on estimates of rate has, however, been largely unstudied in this context. Outgroup selection is, however, widely recognised to influence tree topology, as in the example of long branch attraction (Bergsten 2005). Outgroups could also have an effect on the rates inferred from ingroup branches (Hugall *et al.* 1997; Lyons-Weiler *et al.* 1998). The use of just four taxa could in fact make these trees more susceptible to outgroup affects as the information that their sequence data provides may be insufficient for identification of the true sister-species rates (Schwartz and Mueller 2010a).

Meta analyses of rate comparisons using the sister-species approach compensate for a lack of depth in individual comparisons (e.g. using a single gene) by having a large number of overall comparisons. This is expected to reduce the influence of stochastic effects on the result. Mitochondrial data are most often chosen for meta-comparisons of rate because homologous DNA sequence data are widely available for many species (Gillman *et al.* 2009; Wright *et al.* 2009; Wright *et al.* 2011). Although it is becoming increasingly evident that the mitochondrion has many attributes that are not readily accounted for (Galtier *et al.* 2009b), it has proved a useful marker in comparative studies of rates. In studies of the evolutionary effects of rate variation a marker gene or genes (often mtDNA) is used as a proxy for the genome as a whole. There are many reasons why this might not be a good assumption; it is already known that among gene rate characteristics vary considerably (dispersion nets, Cutler 2000). But, as it is a single inherited unit that is normally non-recombining, we can expect to have confidence in the consistency of estimates of relative rate across the mtDNA genome (Avice *et al.* 1987). Here we test this prediction.

In this study we did not predict a specific rate pattern for the sister-species used. Instead we hypothesized that some species pairs will not differ significantly in their rate as inferred from relative branch length. Thus we expect some sister-species estimates of relative rates to be similar, although the normal dispersion of mutation rate over time makes it unlikely they will be identical. This will result in about half the genes with a slightly faster rate of molecular evolution in one species, and half faster in the other (i.e. differing to a small degree around zero). In contrast, we expect some sister-species comparisons will reveal consistent molecular evolution

rate difference, not due to mutation stochasticity, with longer branches inferred for the same species in each of the mitochondrial genes assessed (Fig. 1). Although sister-species with a significant rate difference will have the same direction in rate skew for each mitochondrial gene, they will differ in overall branch length due to gene rate differences. Branch lengths of sister-species that have the same rate of molecular evolution should reveal a slope where $x=y$, a relationship of one. The relationship of sister-species with different rates will be a slope that deviates from one (Fig. 1). Ergo, we examine whether the rate direction found between species remains consistent across the mitochondrial genes when using separate sister-species analysis for each gene and whether this rate across genes is reflected in the overall slope as expected.

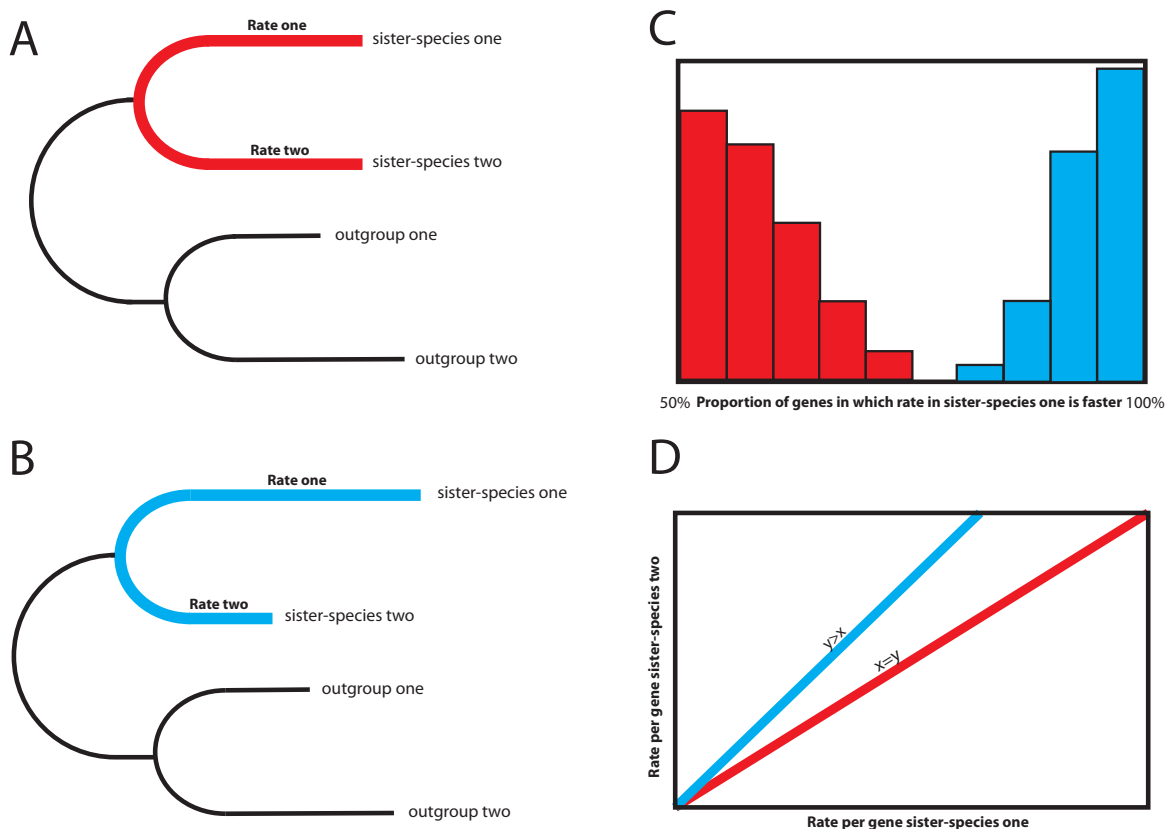


Figure 1

Inferring relative rates of molecular evolution using the ML sister-species approach. The relationship of a four-taxon tree with no rate shift (identified in red), i.e. equal rates, should result in a distribution over the 14 genes no different from random. The branch length ratio is not expected to differ significantly from 1 (D) and the relationship between genes should not deviate from a random expectation (C) with a slope of 1 (D). A relationship with a rate difference is shown in blue. A longer branch length in the ML trees indicates a faster rate (B) consistently across genes (C). This relationship will be apparent in the slope of the genes as well (>1) (D). In this paper we depict the rate as being faster in sister-species one, in reality graph (C) could range from 0%-100% and the slope could be <1 .

We also examine the effect of outgroup choice on directionality of the rate skew detected. The two groups of insects used here to estimate rates are the orthopteran Ensifera representing 200 million years of diversification (Hedges and Leuzinger 2011) and the Dipteran clade *Drosophila* representing divergence over less than 40 million years (Obbard *et al.* 2012). The *Drosophila* sister-species involve close species comparisons (<10mya) (Obbard *et al.* 2012). Although it has been suggested that rate estimates from close species can be inaccurate (Lanfear *et al.* 2010b), close species have frequently been used in sister-species rate comparisons as they enable differentiation between interconnected life-history traits (Gillman *et al.* 2009; Wright *et al.* 2009). The Ensifera are examined using rate estimates between sister-species with 4-taxon trees and separately using all available outgroup taxa. This provides a test of the influence of outgroup composition on relative rate inference. The *Drosophila* are examined using 4-taxon trees with three different outgroup criteria. The ingroup taxon pairs chosen range from within species to between genera (S1.1).

METHODS

Novel whole mitochondrial genomes were assembled for Ensifera, with an emphasis on the family Anostostomatidae. Genomic samples were sequenced on an Illumina High-Seq platform before *de novo* assemble (see S1.2).

Twenty-seven mtDNA genomes of Ensifera and fourteen of *Drosophila* were available for analysis (S1.1, S1.3). Genomes were separated into genes and aligned into the 15 protein coding and rRNA genes that comprise the mitochondrion in both groups (Ensifera and *Drosophila*) using Geneious v6.1 (Geneious). When needed, two *Tephritidae* (NC_008748 and NC_000857) and two *Acrididae* (NC_019993 and GU270284) mtDNA genomes were used as outgroups for *Drosophila* and Ensifera respectively. For each dataset jModelTest v2.1.3 (Darriba *et al.* 2012) was used to identify a suitable model of DNA evolution for each gene. A phylogenetic hypothesis was obtained for each of the gene alignments using the appropriate model in RaxML v7.4.2 (Stamatakis 2006), or PhyML v3.0 (Guindon *et al.* 2010) via Geneious v6.1 for *Drosophila* which required a HKY model not implemented by RaxML. All trees in RaxML throughout the study used a position scheme of 1st, 2nd and 3rd base pair separation for protein coding genes and no base partitions for rRNA genes. To test the global molecular clock hypothesis, a molecular clock test was applied to the rooted gene trees in HyPhy v2.1.1 (Pond and Muse 2005). The gene trees were then used to select sister pairs and two appropriate outgroup taxa for the four-taxon comparisons. Bootstrapped Maximum Likelihood analyses of concatenated gene alignments was conducted for both Ensifera and *Drosophila* in RaxML via CIPRES (Miller *et al.* 2010) allowing RaxML to halt the resampling when statistically appropriate. A Bayesian phylogeny was estimated for the Ensifera dataset using Beast v1.7.4 (Drummond *et al.* 2012).

To minimise effects of non-independent sampling, each species was used as an ingroup species only once resulting in nine Ensifera species pairs (S1.1). For each pair of species two other ensiferans were chosen as the outgroup for the four-taxon tree analysis. The criterion for outgroup selection was that they had the nearest relationship to the ingroups that maintained ingroup monophyly across all genes. Three ensiferan species pairs suffered from poor phylogenetic resolution for some

genes, resulting in their paraphyly with respect to the outgroup taxa (Table 1). These genes were removed from the dataset, along with ATP8, which was considered too short (168 bp) for reliable rate estimation. To examine how sensitive branch length estimates were to outgroup selection we repeated the sister-species comparisons using all available outgroup taxa for each gene. Due to minor topological differences among phylogenies inferred from the different genes, the number of outgroup sequences used varied between genes for some species pairs. A few trees could not be resolved into sister-species pairs in this manner (Table 1).

Five independent sister-species pairs were examined from the genus *Drosophila*, which comprised closer sister pairs than within the Ensifera (S1.1, S1.3). Appropriate outgroups were established for each sister-species comparison as described for the Ensifera. It has been suggested that a model of DNA evolution cannot be reliably selected using fewer than ~30 individuals (Sullivan *et al.* 1999; Blouin *et al.* 2005), and as we had 14 genomes from *Drosophila* we considered this might be a problem. Using PhyML we compared the effect on directionality of rate differences resulting from analysis with a parameter rich GTR model or simpler HKY model. The ATP8 gene was excluded along with any gene comparison in which the outgroup species prevented the sister-species being monophyletic. To examine how sensitive branch length comparisons were to outgroup selection we ran three separate sets of analyses for each *Drosophila* sister-species comparison (15 four-taxon tree comparisons in total) using different outgroup composition.

Once our ingroup species pair and two outgroup taxa had been chosen, all 4-taxon gene data were realigned in Geneious and ML trees for Ensifera and *Drosophila* were re-constructed in RaxML and phyML respectively using the preferred models. The trees were viewed in Dendroscope v3 (Huson and Scornavacca 2012) and FigTree 1.3.1 (Rambaut 2009) and branch lengths were used as estimates of rate from the common ancestor given the outgroup used. In addition a single relative rate estimate was obtained after concatenation of all genes for each four-taxon set. A rate estimate for the Ensifera species-pairs was also obtained using the concatenated mitochondrial data for all 27 species (resolving monophyly in pairs E3 and E8 required the removal of some ingroups).

The number of gene comparisons (as a proportion of the number of genes available for each taxon pair) where a faster rate was inferred in sister-species one versus sister-species two was compared graphically using R v3.0.0 (R Core Team 2013). Only independent sister-species comparisons were included; 9 Ensifera (Eb) and 5 *Drosophila* (Da). Those pairs where a significant proportion of genes indicated a rate difference in the same direction (based on the χ^2 test against a random (1:1) distribution and a two-tailed binomial exact test) were inferred as having a rate-shift in one of the sister-species lineages. Using R we then compared the slope of the linear regression of rate estimates obtained for each set of species with the different genes. Our null hypothesis was a slope of 1 ($y/x=1$ where x and y were the branch lengths for the two ingroup taxa). In all cases the sister-species one (which had longer branch lengths for >50% of gene comparisons) was divided by sister-species two. A slope with a 95%CI that included 1 was expected for those species pairs without rate difference. Species pairs that had consistent rate differences across genes were expected to result in a slope and a 95%CI >1.

RESULTS

We were able to reject the hypothesis of a strict molecular clock for both insect datasets ($P < 0.01$). The ML trees representing the concatenated mitochondrial gene datasets had topologies largely consistent with previous analyses of *Ensifera* and *Drosophila* (S1.4) (Zhou *et al.* 2010; Obbard *et al.* 2012). However, this is the first time 15 of these *Ensifera* genomes have been included in phylogenetic analysis. We found evidence that the Anostomatidae, here represented by 12 species, is not monophyletic with respect to the Stenopelmatidae represented by the Jerusalem cricket (*Stenopelmatus* sp., S1.4, S1.5).

For three *Drosophila* species pairs (D1c, D2b and D5b) the outgroup taxa resulted in paraphyly of the sister-species in the analysis of one gene. Their trees were removed from the dataset (Table 1), but it is interesting to note that these inferred species relationships emerged only in 4-taxon analysis, whereas ML trees with all 14 taxa supported monophyletic relationships for the sister-species. This result was likely due to the tree entering the inverse-Felsenstein zone; with incorrect topology arising from long-branch repulsion (Siddall 1998; Swofford *et al.* 2001). *Ensiferan* genes that were not used due to poor topology may have also suffered from the same problem, but poorly supported topology made it hard to determine. Choice of the model of DNA evolution used in analysis of *Drosophila* influenced which lineage appeared to have evolved faster in only 11 gene/taxon comparisons out of a possible 210. Thus, model choice (GTR or HKY) did not greatly influence our results. Here we present results from analyses using phyML employing the model indicated as appropriate for each gene in JModelTest.

The comparison of rate variation between our sister-species showed that, across a set of linked loci, only in one sister-species pair was the same species faster in all 14 genes examined across all outgroup tests, E3 (Table 1 Fig. 2 A). In most species-pairs (21/33), the pattern of longer and shorter branches from the mitochondrial genes did not differ significantly from random (χ^2 , Table 1). The proportion of gene rates did not fit our hypothesis of consistent rate differences amongst genes for a sister-species pair (Fig. 1 & 2 A).

Table 1

The relative rate direction of individual genes in the different types of comparisons of Ensifera (E) and *Drosophila* (D) taxa. A value of 1 indicates sister-species one was faster, 2 indicates sister-species two was faster and = indicates an identical branch length between the two ingroups. Shading within the genes indicates where the directionality of rates has changed between comparisons using the different outgroup methods. Bold chi-square (χ^2) and binomial exact test (ET) results (P-value) show pairs that have a random distribution of rates. %IS shows the percentage of identical nucleotide sites between the sister species and TBL shows the sum of branch lengths across all the genes for a pair.

	E1a	E1b	E2a	E2b	E3a	E3b	E4a	E4b	E5a	E5b	E6a	E6b	E7a	E7b	E8a	E8b	E9a	E9b
12S	2	1	2	2	1	1	1	1	1	1	1	2	2	2	1	1	2	1
16S	1	1	1	2	1	1	2	2	1	1	1	1	1	1	2	2	1	2
ATP6	1	1	1	1	1	1	1		2	1	2	1	2	2	1	1	1	
COX1	2	2	1	2	1	1	2	1	2	1	1	2	1	1	2	1	1	1
COX2	2	1	1	1	1	1	1	1	1	1	1	2	1	2	2		1	1
COX3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
CYTB	1	1	2	2	1	1			2	1	2	1	1	1		1	1	1
ND1	1	1	2	2	1	1	1	1	1	1	2	1	2	2	1	1		
ND2	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1		2
ND3	2	1	1	1	1	1	1	1	1	2		2	1	1			1	1
ND4	1	1	1	1	1	1	1	1	1	2	1	2	1	1	1	2	1	2
ND4L	1	1	1	2	1	1	2		1	1	1	2	2	2	1	1	2	1
ND5	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1
ND6	2	2	2	2	1	1	1	1	2	1	2	1	2	2	1	1	1	
1	9	12	10	7	14	14	10	9	9	12	9	7	9	8	9	10	10	8
2	5	2	4	7	0	0	3	1	5	2	4	7	5	6	3	2	2	3
%1	64.29	85.71	71.43	50	100	100	76.92	90	64.29	85.71	69.23	50	64.29	57.14	75	83.33	83.33	72.73
χ^2	0.29	0.008	0.109	1	0.000	0.000	0.052	0.011	0.285	0.008	0.166	1	0.285	0.593	0.083	0.021	0.021	0.132
ET	0.122	0.006	0.061	0.209	0.000	0.000	0.035	0.009	0.122	0.005	0.087	0.209	0.122	0.183	0.053	0.016	0.016	0.081
%IS	78.4	78.4	94	94	80.2	80.2	77.5	77.5	87.6	87.6	76.7	76.7	79.6	79.6	74.2	74.2	75.5	75.5
TBL	983.2	64.4	1.5	3.2	3936.5	444.1	16972.1	60.7	176.8	15.3	6666.6	72.3	5252.4	41.7	9967.6	63.6	1510.9	172.8

	D1a	D1b	D1c	D2a	D2b	D2c	D3a	D3b	D3c	D4a	D4b	D4c	D5a	D5b	D5c
12S	1	1	1	1		2	1	1	1	=	1	2	2	2	2
16S	1	1	1	1	1	1	1	1	1	=	1	=	2	2	2
ATP6	1	1	1	1	1	1	1	2	1	2	1	1	1	1	1
COX1	1	2	1	1	1	1	1	1	2	2	2	2	2	2	2
COX2	2	1	1	1	1	1	=	1	2	1	1	1	1	2	1
COX3	1	1		1	1	2	2	2	2	1	1	1	2	2	2
CYTB	1	1	1	1	1	1	1	1	1	2	1	1	2	1	1
ND1	1	1	1	1	2	2	1	1	1	1	1	1	2	2	2
ND2	1	1	1	1	1	1	2	1	1	1	1	1	1		1
ND3	1	1	1	1	2	2	2	1	2	1	1	1	1	2	1
ND4	1	1	2	1	1	1	1	1	2	1	1	1	1	2	2
ND4L	2	2	2	1	1	1	2	1	1	1	1	1	1	1	2
ND5	1	2	1	1	1	1	1	1	1	1	1	1	1	2	2
ND6	1	2	1	1	1	1	1	1	1	2	2	2	1	2	2
1	12	10	11	14	11	10	9	12	9	8	12	10	8	3	5
2	2	4	2	0	2	4	4	2	5	4	2	3	6	10	9
%1	85.71	71.43	84.62	100	84.62	71.43	64.29	85.71	64.29	57.14	85.71	71.43	57.14	23.078	35.71
χ^2	0.008	0.109	0.013	0.000	0.013	0.109	0.166	0.008	0.285	0.248	0.008	0.052	0.593	0.052	0.285
ET	0.005	0.061	0.009	0.000	0.009	0.061	0.087	0.005	0.122	0.121	0.005	0.035	0.183	0.035	0.122
%IS	93.3	93.3	93.3	92.3	92.3	92.3	95.6	95.6	95.6	98.2	98.2	98.2	98.8	98.8	98.8
TBL	1.9085	1.2958	1.1102	1.5574	1.7302	1.5737	0.7588	0.8036	0.791	0.2806	0.2766	0.2858	0.1790	0.1734	0.1790

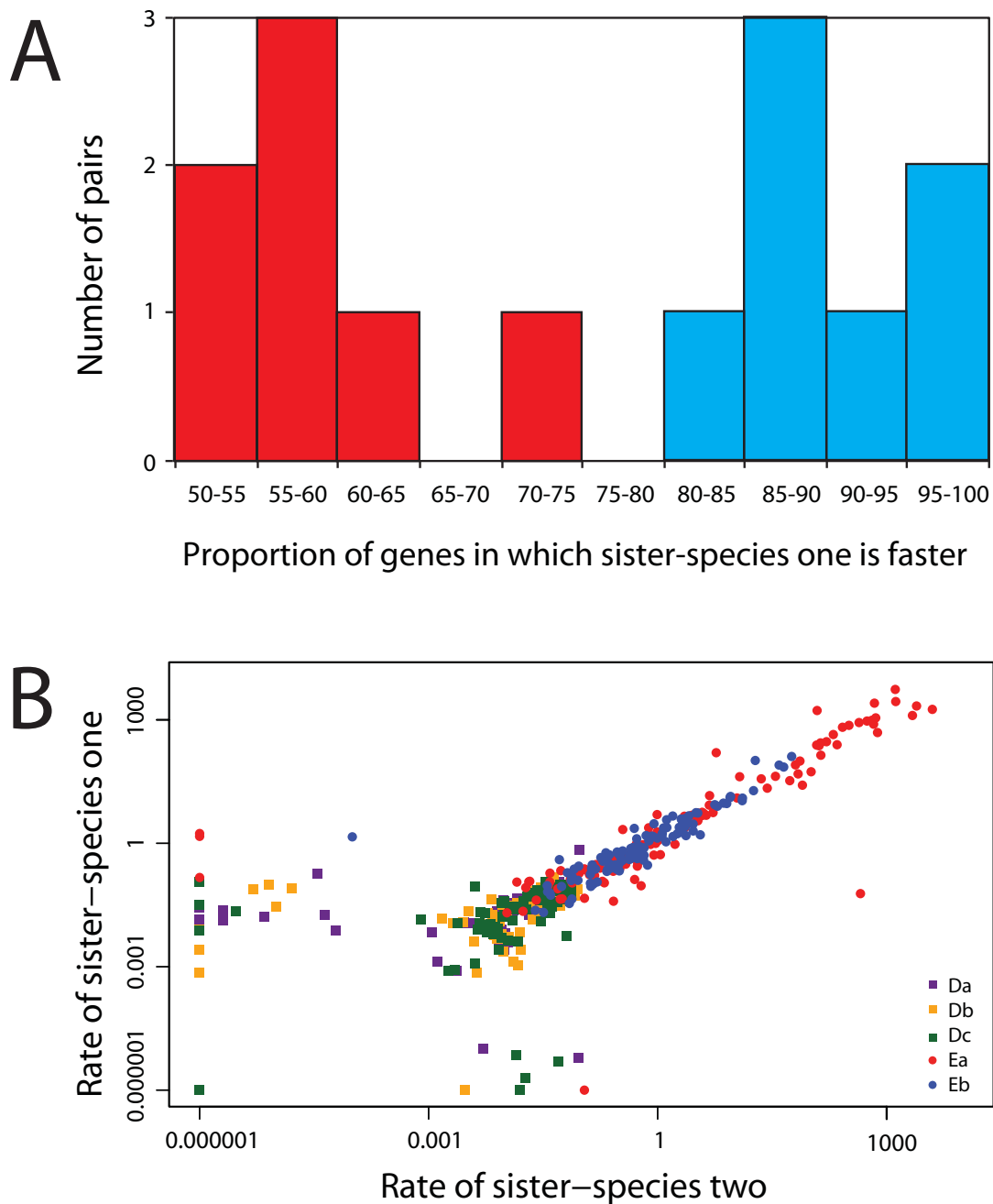


Figure 2

Results from sister-species analysis showed a pattern that was inconsistent with our hypothesis in terms of rate patterns across genes (A). The *Drosophila* comparisons tended to represent closer sister-species, hence the shorter branch lengths that were distributed more stochastically than comparisons among the more distantly related Ensifera (B). Within the Ensifera, method Ea (two outgroup taxa) tended to result in larger branch length estimates than Eb (many outgroup taxa).

Four of twelve sister-species pairs with a consistent rate shift detected across the mitochondrial genome revealed that the relationship between rates was not equal to 1 (95% CI of slope (>1) did not include 1; Table 2). Two of these comparisons came from the same sister-species pair (E3), the other two (E4b and E8b) came from comparisons using all available outgroups. Rate differences that had been indicated using the four-taxon method were not apparent in the slope analysis.

Table 2

The slope test of sister-species. Bold pairs were expected to have a slope with a CI not including 1 (light shows 95%CI that includes 1) as they showed a gene rate significantly different from random (Fig. 2). However only four out of 12 comparisons showed this relationship and only one set (both comparisons) is consistent in both analyses (E3). Most of the *Drosophila* comparisons did not have a significant relationship between ingroups (*) and had low R^2 (dark grey shading indicates an $R^2 < 0.6$).

Pair	Lower CI	Upper CI	Slope	Y	R^2
E1a*	-0.2435	0.4855	0.1210	17.8546	0.0418
E1b*	0.7982	1.1172	0.9577	0.6177	0.9344
E2a*	-1.2603	0.1524	-0.554	0.0899	0.1957
E2b*	0.0884	0.8681	0.4782	0.0716	0.3732
E3a*	3.334	3.8478	3.5909	-19.6211	0.9344
E3b*	1.7703	2.9263	2.3483	2.7814	0.8672
E4a*	0.6451	2.5269	1.5860	207.059	0.5558
E4b*	1.4393	1.6362	1.5377	0.04	0.9939
E5a*	0.4417	1.0405	0.7411	2.2832	0.708
E5b*	0.7343	1.5491	1.1417	0.2152	0.7564
E6a*	0.4602	0.8046	0.6324	41.6074	0.8559
E6b*	0.7152	1.1107	0.913	0.1851	0.8939
E7a*	1.2839	1.6052	1.4446	5.4827	0.9697
E7b*	1.0239	1.2564	1.1402	-0.1243	0.9744
E8a*	0.1042	0.7102	0.4072	257.918	0.4727
E8b*	1.167	1.5944	1.3807	0.5022	0.9539
E9a*	0.894	1.6908	1.2924	2.8048	0.8393
E9b*	1.3854	1.6901	1.5377	-0.7849	0.983
D1a*	0.757	7.4831	4.1201	-0.0429	0.3725
D1b	-0.2643	1.3601	0.5479	0.0329	0.1526
D1c	-0.8475	1.1862	0.1694	0.0493	0.0121
D2a*	0.4532	1.5855	1.0194	0.0211	0.5619
D2b	-0.5125	1.176	0.3317	0.0637	0.0637
D2c	-0.4394	1.2158	0.3882	0.0528	0.0801
D3a	-0.6269	0.1384	-0.2443	0.0351	0.1388
D3b	-2.1036	0.1353	-0.9841	0.0572	0.2341
D3c	-0.7145	0.3908	-0.1618	0.0351	0.0328
D4a	-0.6052	1.0027	0.1988	0.0126	0.0236
D4b	-0.3649	1.1765	0.4058	0.0107	0.0988
D4c	-0.6578	0.2479	-0.2049	0.0143	0.0749
D5a*	-1.8304	-0.3358	-1.0831	0.0131	0.4538
D5b	-0.8049	0.2246	-0.2902	0.0077	0.1227
D5c*	-0.7061	0.06051	-0.3228	0.0074	0.2191

In addition, the sister-species rate direction within genes was not consistent among tests with different outgroup taxa. This inconsistency was not isolated to genes with particular characteristics such as shorter length (e.g. *ND3*) or substitution saturation (e.g. *COXI*) (Waddell *et al.* 2007) (Table 1), but occurred across all 14 genes in our dataset. Within *Drosophila* only four of the 13 sister-species analyses produced estimates of rates that showed a positive correlation among the 14 genes (significant linear regression). Three of these involved the use of the closest possible outgroup taxa. However, in all cases linear regression of *Drosophila* data had low R^2 values (<0.6 Table 2) indicating poor correlation among estimates derived from the 14 mitochondrial genes. All Ensifera comparisons showed significant linear correlations between the branch lengths of the ingroup taxa, but R^2 values tended to be higher in comparisons using all available taxa in the outgroup rather than just two outgroup sequences (Table 2). Moreover, estimates of total branch length were lower for the Ensifera comparisons using the larger outgroup and there were fewer data outliers (Fig. 2 B).

DISCUSSION

Given the general understanding about the mitochondrial genome we predicted that, for a given pair of species, the relationship of relative branch rates estimated from each of the genes would be correlated across the genome. Only in one sister-species comparison was this predicted relationship supported by both outgroup tests in the chi-squared and regression analysis (E3, *Hemideina thoracica* and *Deinacrida connectens*). Most of the other species pairs (8/10) that indicated a disparity in rates shift across genes did not have the predicted slope relationship (Fig. 1, Table 2). Some sister-species that did not have a significant rate difference across genes nevertheless did have correlation of rates derived from the 14 genes that differed significant from $x = y$ (slope = 1). This may be due to the more rapidly evolving genes having a disproportionate influence on the regression analysis. If faster genes provide a more accurate estimate of the true rate pattern and are less influenced by stochastic effects, they could be useful for rate comparisons. However, fast genes are also more likely to be subject to saturation and branch length estimates have been shown to be less accurate on fast genes (Phillips 2009; Schwartz and Mueller 2010b).

The use of closely related species has been suggested to result in poor rate estimates (Lanfear *et al.* 2010b). Of our closely related sister-species pairs (all *Drosophila* comparisons and E2) only 6 of 17 comparisons showed a significant relationship between ingroup branch lengths and all had a poor fit of data to the regression as indicated by low R^2 (Table 2). Rate skews between closely related taxa are likely more influenced by stochastic processes and the true rate is therefore harder to estimate (Fig. 2). It is also likely that the rates estimated from these species pairs reflect the time dependent or lazy-J distribution of rates through time (Ho *et al.* 2005; Penny 2005). Branch lengths among species that diverged recently are more likely to reflect mutation rate than the net rate of molecular evolution, which is a product of both mutation rate and fixation rate (Ho *et al.* 2005; Ho *et al.* 2007; Papadopoulou *et al.* 2010; Ho *et al.* 2011). Recent phylogenetic dating of *Drosophila* lineages indicates that D4 (*D. persimilis* and *D. pseudobscura*) and D5 (*D. sechellia* and *D. simulans*) represent lineages that diverged <1 mya, D3 (*D. melanogaster* and

D. mauritiana) diverged <2 mya, and D1 (*D. erecta* and *D. yakuba*) diverged 3-1 mya. D2 (*D. littoralis* and *D. virilis*) probably diverged between 7-9 mya, which is consistent with the larger total branch length estimated for this species pair in our analysis (Table 1) (Cutter 2008; Morales-Hojas *et al.* 2011; Obbard *et al.* 2012). Although estimates of divergence age for D1 and D2 appear to be outside the expected influence of the time dependency curve, time dependency of molecular evolution has not been examined in *Drosophila*, which typically have large population sizes. It is possible that even these older lineage splits in this particular group are overly influenced by mutation rate as opposed to rate of molecular evolution. Closely related species have often been used for rate comparisons as they limit independent traits acquired from “unshared” ancestors (Gillman *et al.* 2009; Wright *et al.* 2009). However our results strongly suggest that closely related species provide particularly unreliable estimates of rates of molecular evolution and should be avoided in rate comparisons.

Our analyses were also sensitive to choice of outgroup, indicating a further source of uncertainty. This might be the result of homoplasy although this is difficult to determine from a four-taxon tree. Homoplasy is known to be high in the mitochondria within species (Galtier *et al.* 2009b) and it remains to be seen whether this sensitivity to outgroup selection will hold for the nuclear genome. The ensiferan comparisons suggest that including all possible species in the outgroup stabilises rate estimates; comparisons using species-rich outgroups generally had high R^2 values (8 out of 9 comparisons) and Y-intercepts closer to 0 (Table 2). However, even when using all available outgroups, only 3 out of 5 comparisons with apparent rate differences had a regression that differed significantly from 1 (Table 2). For *Drosophila*, analysis with outgroups more closely related to the ingroup resulted in better rate estimates. This supports results from simulations that found branch lengths could be overestimated in four-taxon trees when high substitution rates were present, but the results improved with the addition of more taxa (Schwartz and Mueller 2010a).

Although phylogenetic analysis is known to be sensitive to outgroup choice, especially when the quality of data (phylogenetic signal) is low (Bergsten 2005), to

our knowledge these problems of ‘difficult tree space’ have not previously been referred to in rate analysis. Long branch repulsion in the inverse-Felsenstein zone results from branch length disparity among taxa within the tree (Siddall 1998; Swofford *et al.* 2001). It is possible that the Ensifera trees showing an unexpectedly reversed relationship of rates also entered this zone of tree space, but the resolution was not sufficiently clear to reject other explanations. This problem is more prevalent with short internal branches (Swofford *et al.* 2001), and the use of the large phylogeny approach avoids this issue.

Rate comparisons often use a sister-species analysis to compare between lineages with few genes on the assumption that the relative rate estimates generated are representative of the whole genome. However, we have shown here that there are substantial inconsistencies when using this method. More outgroup data appeared to decrease the stochastic effects, but even with large numbers of outgroup sequences our estimates had high variance among the 14 mitochondrial genes, with the exception of one pair (E3). Though this problem should not represent a systematic bias in rates when testing life history traits it increases the error associated with such estimates and puts into question the conclusions drawn from studies that employ a sister-species analysis. The contrasting results drawn from these studies might in part be explained by the error in obtaining the true direction of rate variation between species. Our findings are particularly compelling because we limited our analysis to co-evolving genes within the mitochondrion, a non-recombining single unit of inheritance (Avice *et al.* 1987). Rate comparisons across single genes or a small number of loci from closely related taxa should be avoided and consistency across genomes not assumed.

ACKNOWLEDGEMENTS

Thanks to the following who provided material for Ensifera sequencing or helped with field work: Briar Taylor Smith, Geoff Montieth, Rhys Richards, Mariana Bulgarella, Bill Wiesman, Graeme Atkins (DOC), and Trish McLenachan for help in the lab. We would also like to thank Bennet McComish who assembled the *Hemideina crassidens* (north) and *Motuweta riparia*, and Lesley Collins, Patrick Biggs, Simon Hills and Gillian Gibbs for their helpful advice on genome assembly.

REFERENCES

- Algar AC, Kerr JT, Currie DJ (2007) A test of Metabolic Theory as the mechanism underlying broad-scale species-richness gradients. *Global Ecology and Biogeography* **16**, 170-178.
- Avice JC, Arnold J, Ball RM, Bermingham E, Lamb T *et al.* (1987) Intraspecific Phylogeography: The Mitochondrial DNA Bridge Between Population Genetics and Systematics. *Annual Review of Ecology and Systematics* **18**, 489-522.
- Bakewell MA, Shi P, Zhang J (2007) More genes underwent positive selection in chimpanzee evolution than in human evolution. *Proceedings of the National Academy of Sciences* **104**, 7489-7494.
- Bergsten J (2005) A review of long-branch attraction. *Cladistics* **21**, 163-193.
- Blouin C, Butt D, Roger AJ (2005) Impact of Taxon Sampling on the Estimation of Rates of Evolution at Sites. *Molecular Biology and Evolution* **22**, 784-791.
- Cutler DJ (2000) Understanding the overdispersed molecular clock. *Genetics* **154**, 1403-1417.
- Cutter AD (2008) Divergence Times in *Caenorhabditis* and *Drosophila* Inferred from Direct Estimates of the Neutral Mutation Rate. *Molecular Biology and Evolution* **25**, 778-786.
- Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nat Meth* **9**, 772.
- Dowle EJ, Morgan-Richards M, Trewick SA (2013) Molecular evolution and the latitudinal biodiversity gradient. *Heredity* **110**, 501-510.
- Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* **29**, 1969-1973.
- Duchene D, Bromham L (2013) Rates of molecular evolution and diversification in plants: chloroplast substitution rates correlate with species-richness in the Proteaceae. *BMC Evolutionary Biology* **13**, 65.
- Eyre-Walker A, Keightley PD (2009) Estimating the Rate of Adaptive Molecular Evolution in the Presence of Slightly Deleterious Mutations and Population Size Change. *Molecular Biology and Evolution* **26**, 2097-2108.
- Galtier N, Blier PU, Nabholz B (2009a) Inverse relationship between longevity and evolutionary rate of mitochondrial proteins in mammals and birds. *Mitochondrion* **9**, 51-57.

- Galtier N, Nabholz B, Glémin S, Hurst GDD (2009b) Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular Ecology* **18**, 4541-4550.
- Geneious version 6.1 created by Biomatters. Available from <http://www.geneious.com/>.
- Gillman LN, Keeling DJ, Gardner RC, Wright SD (2010) Faster evolution of highly conserved DNA in tropical plants. *Journal of Evolutionary Biology* **23**, 1327-1330.
- Gillman LN, Keeling DJ, Ross HA, Wright SD (2009) Latitude, elevation and the tempo of molecular evolution in mammals. *Proceedings of the Royal Society B-Biological Sciences* **276**, 3353-3359.
- Gillman LN, Mccowan LSC, Wright SD (2012) The tempo of genetic evolution in birds: body mass and climate effects. *Journal of Biogeography*, no-no.
- Goldie X, Lanfear R, Bromham L (2011) Diversification and rate of molecular evolution: no evidence of a link in mammals. *BMC Evolutionary Biology* **11**.
- Guindon SP, Dufayard J-FO, Lefort V, Anisimova M, Hordijk W *et al.* (2010) New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology* **59**, 307-321.
- Heads S, Leuzinger LA (2011) On the placement of the Cretaceous orthopteran *Brauckmannia groeningae* from Brazil, with notes on the relationships of Schizodactylidae (Orthoptera, Ensifera). *ZooKeys* **77**, 17-30.
- Ho SYW, Lanfear R, Bromham L, Phillips MJ, Soubrier J *et al.* (2011) Time-dependent rates of molecular evolution. *Molecular Ecology* **20**, 3087-3101.
- Ho SYW, Phillips MJ, Cooper A, Drummond AJ (2005) Time Dependency of Molecular Rate Estimates and Systematic Overestimation of Recent Divergence Times. *Molecular Biology and Evolution* **22**, 1561-1568.
- Ho SYW, Shapiro B, Phillips MJ, Cooper A, Drummond AJ (2007) Evidence for Time Dependency of Molecular Rate Estimates. *Systematic Biology* **56**, 515-522.
- Hugall A, Stanton J, Moritz C (1997) Evolution of the AT-rich mitochondrial DNA of the root knot nematode, *Meloidogyne hapla*. *Molecular Biology and Evolution* **14**, 40-48.
- Hugall AF, Lee MSY (2007) The likelihood node density effect and consequences for evolutionary studies of molecular rates. *Evolution* **61**, 2293-2307.
- Huson DH, Scornavacca C (2012) Dendroscope 3: An Interactive Tool for Rooted Phylogenetic Trees and Networks. *Systematic Biology* **61**, 1061-1067.
- Kosiol C, Vinar T, Da Fonseca RR, Hubisz MJ, Bustamante CD *et al.* (2008) Patterns of Positive Selection in Six Mammalian Genomes. *PLoS Genet* **4**, e1000144.
- Lanfear R, Ho SYW, Love D, Bromham L (2010a) Mutation rate is linked to diversification in birds. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 20423-20428.
- Lanfear R, Welch JJ, Bromham L (2010b) Watching the clock: Studying variation in rates of molecular evolution between species. *Trends in Ecology & Evolution* **25**, 495-503.
- Lyons-Weiler J, Hoelzer GA, Tausch RJ (1998) Optimal outgroup analysis. *Biological Journal of the Linnean Society* **64**, 493-511.
- Miller MA, Pfeiffer W, Schwartz T, 2010 Creating the CIPRES Science Gateway for inference of large phylogenetic trees, pp. 1-8 in *Gateway Computing Environments Workshop (GCE)*, 2010.

- Morales-Hojas R, Reis M, Vieira CP, Vieira J (2011) Resolving the phylogenetic relationships and evolutionary history of the *Drosophila virilis* group using multilocus data. *Molecular Phylogenetics and Evolution* **60**, 249-258.
- Nabholz B, Glemin S, Galtier N (2008) Strong variations of mitochondrial mutation rate across mammals - the longevity hypothesis. *Molecular Biology and Evolution* **25**, 120-130.
- Obbard DJ, Maclennan J, Kim K-W, Rambaut A, O'Grady PM *et al.* (2012) Estimating Divergence Dates and Substitution Rates in the *Drosophila* Phylogeny. *Molecular Biology and Evolution* **29**, 3459-3473.
- Ohta T (1992) The Nearly Neutral Theory of Molecular Evolution. *Annual Review of Ecology and Systematics* **23**, 263-286.
- Papadopoulou A, Anastasiou I, Vogler AP (2010) Revisiting the Insect Mitochondrial Molecular Clock: The Mid-Aegean Trench Calibration. *Molecular Biology and Evolution* **27**, 1659-1672.
- Penny D (2005) Evolutionary biology: Relativity for molecular clocks. *Nature* **436**, 183-184.
- Petit N, Barbadilla A (2009) Selection efficiency and effective population size in *Drosophila* species. *Journal of Evolutionary Biology* **22**, 515-526.
- Phillips MJ (2009) Branch-length estimation bias misleads molecular dating for a vertebrate mitochondrial phylogeny. *Gene* **441**, 132-140.
- Pond SK, Muse S, 2005 HyPhy: Hypothesis Testing Using Phylogenies, pp. 125-181 in *Statistical Methods in Molecular Evolution*. Springer New York.
- R Core Team, 2013 R: A language and environment for statistical computing, pp. Retrieved from <http://http://www.R-project.org>. R foundation for Statistical Computing, Vienna, Austria.
- Rambaut A, 2009 FigTree v1.3.1. Available at: <http://tree.bio.ed.ac.uk/software/figtree>, pp.
- Schwartz R, Mueller R (2010a) Branch length estimation and divergence dating: estimates of error in Bayesian and maximum likelihood frameworks. *BMC Evolutionary Biology* **10**, 5.
- Schwartz RS, Mueller RL (2010b) Variation in DNA Substitution Rates among Lineages Erroneously Inferred from Simulated Clock-Like Data. *PLoS ONE* **5**, e9649.
- Siddall ME (1998) Success of Parsimony in the Four-Taxon Case: Long-Branch Repulsion by Likelihood in the Farris Zone. *Cladistics* **14**, 209-220.
- Smith NGC, Eyre-Walker A (2003) Partitioning the Variation in Mammalian Substitution Rates. *Molecular Biology and Evolution* **20**, 10-17.
- Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688-2690.
- Sullivan J, Swofford D, Naylor G (1999) The Effect of Taxon Sampling on Estimating Rate Heterogeneity Parameters of Maximum-Likelihood Models. *Molecular Biology and Evolution* **16**, 1347.
- Swofford DL, Waddell PJ, Huelsenbeck JP, Foster PG, Lewis PO *et al.* (2001) Bias in Phylogenetic Estimation and Its Relevance to the Choice between Parsimony and Likelihood Methods. *Systematic Biology* **50**, 525-539.
- Thomas JA, Welch JJ, Lanfear R, Bromham L (2010) A Generation Time Effect on the Rate of Molecular Evolution in Invertebrates. *Molecular Biology and Evolution* **27**, 1173-1180.

- Waddell PJ, Kishino H, Ota R (2007) Phylogenetic Methodology for Detecting Protein Interactions. *Molecular Biology and Evolution* **24**, 650-659.
- Woolfit M, Bromham L (2005) Population size and molecular evolution on islands. *Proceedings of the Royal Society B: Biological Sciences* **272**, 2277-2282.
- Wright S, Keeling J, Gillman L (2006) The road from Santa Rosalia: A faster tempo of evolution in tropical climates. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 7718-7722.
- Wright S, Ross H, Jeanette Keeling D, McBride P, Gillman L (2011) Thermal energy and the rate of genetic evolution in marine fishes. *Evolutionary Ecology* **25**, 525-530.
- Wright SD, Gillman LN, Ross HA, Keeling DJ (2009) Slower Tempo of Microevolution in Island Birds: Implications for Conservation Biology. *Evolution* **63**, 2275-2287.
- Wright SD, Gillman LN, Ross HA, Keeling JD (2010) Energy and the tempo of evolution in amphibians. *Global Ecology and Biogeography*, **19** (5), 733-740
- Zhou Z, Ye H, Huang Y, Shi F (2010) The phylogeny of Orthoptera inferred from mtDNA and description of *Elimaea cheni* (Tettigoniidae: Phaneropterinae) mitogenome. *Journal of Genetics and Genomics* **37**, 315-324.

S1. SUPPLEMENTARY MATERIAL FOR CHAPTER TWO

S1.1 SISTER-PAIRS IN DETAIL

Table 1

Paired species comparisons for crickets (Ensifera) and flies (*Drosophila*) used to infer rates of molecular evolution from 14 mitochondrial genes, and the outgroups used in four taxon trees, with genbank accession numbers for *Drosophila*. *Hypocophoides* (Australian) was originally described for a species in India but Johns (1997) considered that the west Australian anostostomatid was closer to *Hypocophoides* than *Onosandrus* where it was formerly placed.

Code	Ingroup One	Ingroup Two	Outgroup One	Outgroup Two
E1a	<i>Lutosa</i> sp.	<i>Motuweta riparia</i>	<i>Stenopelmatus</i> sp.	<i>Hypocophoides</i> sp.
E2a	<i>Hemideina</i> <i>crassidens</i> North	<i>Hemideina</i> <i>crassidens</i> South	<i>Deinacrida connectens</i>	<i>Hemideina thoracica</i>
E3a	<i>Hemideina thoracica</i>	<i>Deinacrida</i> <i>connectens</i>	<i>Hemiandrus</i> <i>palliatarsus</i>	<i>Hemiandrus</i> sp. (Australia)
E4a	<i>Pachryhamma</i> <i>edwardsii</i>	<i>Troglophilus</i> <i>neglectus</i>	<i>Carcinopsis</i> sp	<i>Lutosa</i> sp.
E5a	<i>Gryllotalpa pluvialis</i>	<i>Gryllotalpa</i> <i>orientalis</i>	<i>Myrmecophilus manni</i>	<i>Teleogryllus emma</i>
E6a	<i>Conocephalus</i> <i>maculatus</i>	<i>Ruspolia dubia</i>	Gryllacrididae sp	<i>Pachyrhamma</i> <i>edwardsii</i> .
E7a	<i>Gampsocleis gratiosa</i>	<i>Anabrus simplex</i>	<i>Elimaea cheni</i>	<i>Xizicus fascipes</i>
E8a	<i>Deracantha onos</i>	<i>Xizicus fascipes</i>	Gryllacrididae sp	<i>Pachyrhamma</i> <i>edwardsii</i> .
E9a	<i>Hemiandrus</i> 'brucie'	<i>Hemiandrus</i> sp. (Australia)	<i>Teleogryllus emma</i>	<i>Myrmecophilus manni</i>
E1b	<i>Lutosa</i> sp.	<i>Motuweta riparia</i>	All feasibly possible whilst retaining ingroup topology	
E2b	<i>Hemideina</i> <i>crassidens</i> North	<i>Hemideina</i> <i>crassidens</i> South	All feasibly possible whilst retaining ingroup topology	
E3b	<i>Hemideina thoracica</i>	<i>Deinacrida</i> <i>connectens</i>	All feasibly possible whilst retaining ingroup topology	
E4b	<i>Pachryhamma</i> sp.	<i>Troglophilus</i> <i>neglectus</i>	All feasibly possible whilst retaining ingroup topology	
E5b	<i>Gryllotalpa pluvialis</i>	<i>Gryllotalpa</i> <i>orientalis</i>	All feasibly possible whilst retaining ingroup topology	
E6b	<i>Conocephalus</i>	<i>Ruspolia dubia</i>	All feasibly possible whilst retaining ingroup topology	

	<i>maculatus</i>		topology	
E7b	<i>Gampsocleis gratiosa</i>	<i>Anabrus simplex</i>	All feasibly possible whilst retaining ingroup topology	
E8b	<i>Deracantha onos</i>	<i>Xizicus fascipes</i>	All feasibly possible whilst retaining ingroup topology	
E9b	<i>Hemiandrus 'brucie'</i>	<i>Hemiandrus</i> sp. (Australia)	All feasibly possible whilst retaining ingroup topology	
D1a	<i>Drosophila erecta</i>	<i>D. yakuba</i>	<i>D. persimilis</i>	<i>D. pseudobscura</i>
D1b	<i>D. erecta</i>	<i>D. yakuba</i>	<i>D. grimshawi</i>	<i>D. littoralis</i>
			BK006338	
D1c	<i>D. erecta</i>	<i>D. yakuba</i>	<i>D. willistoni</i>	<i>D. mojavenis</i>
			BK006338	BK006339
D2a	<i>D. littoralis</i> NC_011596	<i>D. virilis</i> BK006340	<i>D. ananassae</i>	<i>D. yakuba</i>
D2b	<i>D. littoralis</i>	<i>D. virilis</i>	<i>D. grimshawi</i>	<i>D. mojavenis</i>
D2c	<i>D. littoralis</i>	<i>D. virilis</i>	<i>D. willistoni</i>	<i>D. persimilis</i>
D3a	<i>D. melanogaster</i> DMU37541	<i>D. mauritiana</i> NC_005779	<i>D. ananassae</i>	<i>D. pseudobscura</i>
D3b	<i>D. melanogaster</i>	<i>D. mauritiana</i>	<i>D. virilis</i>	<i>D. grimshawi</i>
D3c	<i>D. melanogaster</i>	<i>D. mauritiana</i>	<i>D. yakuba</i>	<i>D. erecta</i>
D4a	<i>D. pseudobscura</i> FJ899745	<i>D. persimilis</i> BK006337	<i>D. ananassae</i> (BK006336)	<i>D. erecta</i>
D4b	<i>D. pseudobscura</i>	<i>D. persimilis</i>	<i>D. virilis</i>	<i>D. willistoni</i>
D4c	<i>D. pseudobscura</i>	<i>D. persimilis</i>	<i>D. yakuba</i>	<i>D. melanogaster</i>
D5a	<i>D. simulans</i> NC_005781	<i>D. sechellia</i> NC_005780	<i>D. pseudobscura</i>	<i>D. persimilis</i>
D5b	<i>D. simulans</i>	<i>D. sechellia</i>	<i>D. melanogaster</i>	<i>D. erecta</i>
D5c	<i>D. simulans</i>	<i>D. sechellia</i>	<i>D. ananassae</i>	<i>D. yakuba</i>

S1.2 GENERATION OF NOVEL MITOCHONDRIAL GENOMES

Genomic DNA was obtained from 14 Orthoptera representing 13 species of Ensifera (S3). Whole genomic DNA was extracted using a modified CTAB extraction to enrich for mtDNA. The cells from testes and hind leg muscle were gently separated by homogenisation in buffer (0.25 M sucrose, 10 mM EDTA, and 30 mM Tris-HCl, pH 7.5). The mixture was centrifuged at 1,000g for one minute and the supernatant

transferred to fresh tubes for a second centrifugation. These steps concentrated cell membranes and nuclei, which were then discarded. The supernatant was transferred to clean tubes and spun at 15,000g for 15 minutes to pellet mitochondria, which were subjected to standard CTAB extraction. This involved incubation at 55°C with Proteinase K and CTAB buffer (2% Hexadecyltrimethylammonium bromide, 100mmol/L Tris-HCl pH8.0, 1.4 mol/L NaCl, 20 mmol/L EDTA), followed by combined phenol/chloroform/isoamyl alcohol (25:24:1) clean-up and ethanol precipitation. The resulting DNA was eluted with 50 µL of water and quantified. Six DNA samples were subjected to genome amplification using Repli-g (Qiagen) to ensure sufficient DNA for sequencing. The DNA extraction of *Hemidandrus pallitarsus* used three individuals from one location, due to the small size of these weta. Two samples *Hemideina crassidens* (North) and *Motuweta riparia* were sequenced and assembled without tagging (Mccomish *et al.* 2010), the rest of the sequences were individually tagged. Sequence data were generated on an Illumina High-Seq 2000 at the Beijing Genomics Institute. Approximately 1 gigabyte of data was obtained per sample.

The quality of the data was assessed using FASTQC and SOLEXAQA (Andrews ; Cox *et al.* 2010) before de novo assembly with VELVET v1.1 and ABYSS v1.3.3 (Zerbino and Birney 2008; Simpson *et al.* 2009). Assembly of the mitochondrial genomes was achieved using a variety of kmer lengths (generally around 49). Contig files were compared with available data using BLAST v2.2.26 (NCBI) implemented in GENEIOUS v6.1 (Geneious) against the 13 available Orthopteran mitochondrial genomes on Genbank (NCBI) and any already assembled genomes from our own data (Supplementary Table 1). We found evidence of numts (nuclear copies of mitochondrial genes) in the data from *Hemideina* sp. and *Hemidandrus* sp. Viewing our mapping results in Tablet v1.11.08.10 (Milne *et al.* 2010) from Bowtie2 (Langmead and Salzberg 2012) enabled us to identify which sequences were mitochondrial and which were likely to be paralogs because the mitochondrial sequences were in greater proportion to the nuclear copies.

The resulting assembled mitochondrial genomes were each approximately 16,000 bp in length. All coding genes and rRNA subunits maintained the same order

previously identified in Ensifera (Kim *et al.* 2005). There were several instances of tRNA shifts and repetitive sequence insertions. In particular the tree weta (*Hemideina*) show insertion of a repetitive unit several hundred base pairs in length between tRNA^Q and COX1. All individual gene sequences are available from Genbank (Table 1).

S1.3 NEW ENSIFERA MITOCHONDRIAL GENOMES SEQUENCED

Table 2

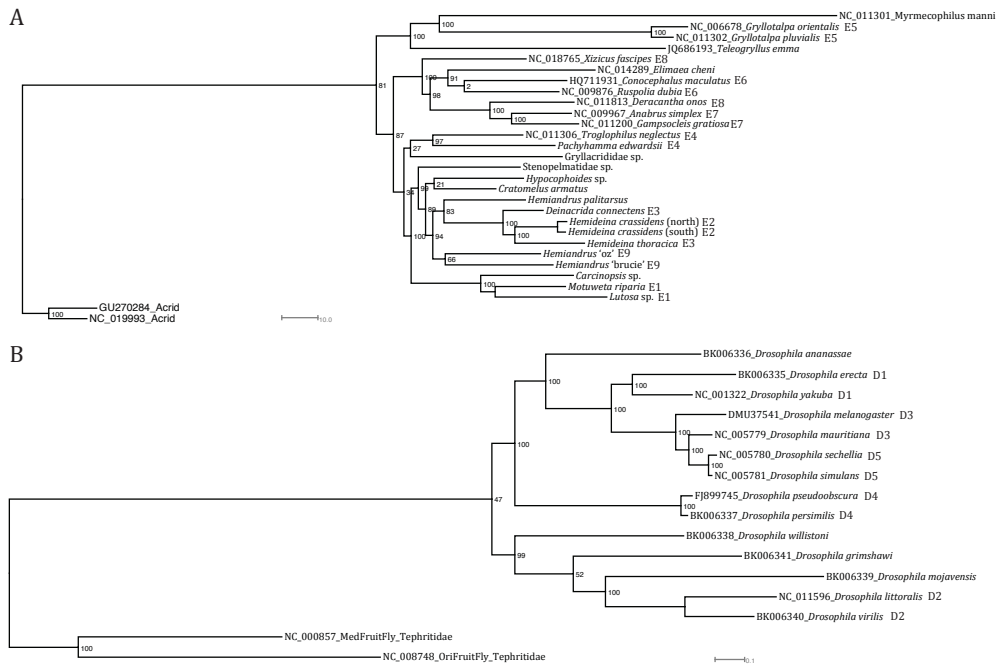
The collection location is shown for the Ensifera generated in this work and the accession is included for those downloaded from genbank.

Family	Species	Reference/Location
Tettigoniidae	Katydid <i>Elimaea cheni</i>	(Zhou <i>et al.</i> 2010)
	NC_014289	
	Grasshopper <i>Deracantha onos</i>	(Zhou <i>et al.</i> 2009)
	NC_011813	
	Chinese bush cricket	(Zhou <i>et al.</i> 2008)
	<i>Gampsocleis gratiosa</i>	
	NC_011200	
	Mormon cricket <i>Anabrus simplex</i>	(Fenn <i>et al.</i> 2007)
	NC_008867	
	Meadow Katydid <i>Ruspolia dubia</i>	(Zhou <i>et al.</i> 2007)
	NC_009876	
	Maculated rice katydid	Unpublished
	<i>Conocephalus maculatus</i>	(Zhou <i>et al.</i> 2010)
Gryllidae	HQ711931	
	Quite calling katydid <i>Xizicus fascipes</i>	(Miller <i>et al.</i> 2010)
	NC_018765	
	Mann's ant Cricket	(Fenn <i>et al.</i> 2008)
	<i>Myrmecophilus manni</i>	
Gryllotalpidae	NC_011301	
	Emma field cricket	(Wolff <i>et al.</i> 2012)
	<i>Teleogryllus emma</i>	JQ686193
Gryllotalpidae	Common mole Cricket	(Fenn <i>et al.</i> 2008)
	<i>Gryllotalpa pluvialis</i>	
	NC_011302	

	oriental mole cricket	(Kim <i>et al.</i> 2005)
	<i>Gryllotalpa orientalis</i>	
	NC_006678	
Rhaphidophoridae	Cave Cricket <i>Troglophilus</i>	(Fenn <i>et al.</i> 2008)
	<i>neglectus</i> NC_011306	
	* <i>Pachyrhamma edwardsii.</i>	New Zealand
Stenopelmatidae	*Jerusalem Cricket	
Gryllacrididae	*Raspy Cricket	Solomon Islands
Anostomatidae	* <i>Hemiandrus palitarsus</i>	New Zealand
	* <i>Hemiandrus</i> 'brucie'	New Zealand
	* <i>Hemideina thoracica</i>	New Zealand
	* <i>Hemideina crassidens</i>	New Zealand (North Island)
	* <i>Hemideina crassidens</i>	New Zealand (South Island)
	* <i>Deinacrida connectens</i>	New Zealand
	* <i>Motuweta riparia</i>	New Zealand
	* <i>Lutosa</i> sp.	Brazil
	* <i>Hemiandrus</i> sp.	Australia
	* <i>Carcinopsis</i> sp.	New Caledonia
	* <i>Cratomelus armatus</i>	Chile
	<i>Hypocophoides</i> (?) sp.	Australia

S1.4 ML TREES FOR DROSOPHILA AND ENSIFERA

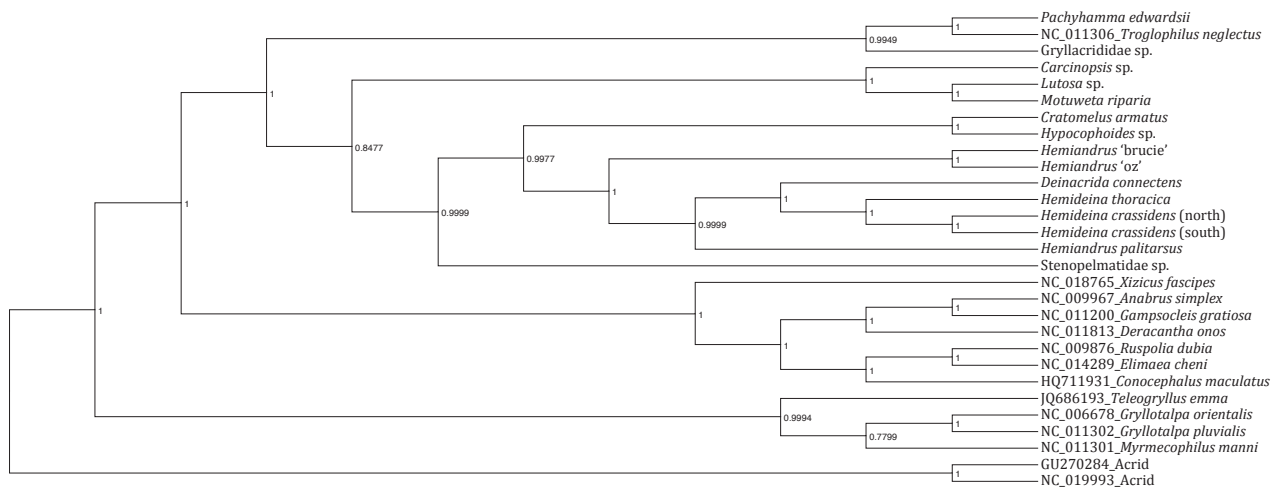
Figure 1
ML trees for *Drosophila* and Ensifera mtDNA (15 genes) indicating the species pairs chosen for comparisons and the non-monophyly of the Anostomatidae with respect to Stenopelmatidae.



S1.5 BAYESIAN TREE FOR ENSIFERA

Figure 2

Bayesian tree of Ensifera showing the same topology as the ML (1.4) regarding the position of Stenopelmatidae.



REFERENCES

- Andrews S, FastQC <http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/>, pp.
- Cox M, Peterson D, Biggs P (2010) SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics* **11**, 485.
- Fenn JD, Cameron SL, Whiting MF (2007) The complete mitochondrial genome sequence of the Mormon cricket (*Anabrus simplex*: Tettigoniidae: Orthoptera) and an analysis of control region variability. *Insect Molecular Biology* **16**, 239-252.
- Fenn JD, Song H, Cameron SL, Whiting MF (2008) A preliminary mitochondrial genome phylogeny of Orthoptera (Insecta) and approaches to maximizing phylogenetic signal found within mitochondrial genome data. *Mol Phylogenet Evol* **49**, 59-68.
- Geneious version 6.1 created by Biomatters. Available from <http://www.geneious.com/>.
- Johns, P. M. (1997). The Gondwanaland Weta: Family Anostomatidae (Formerly in Stenopelmidae, Henicidae or Mimnermidae): Nomenclatural Problems, World Checklist, New Genera and Species, *Journal of Orthoptera Research*, 6: 125-138
- Kim I, Cha SY, Yoon MH, Hwang JS, Lee SM *et al.* (2005) The complete nucleotide sequence and gene organization of the mitochondrial genome of the oriental mole cricket, *Gryllotalpa orientalis* (Orthoptera: Gryllotalpidae). *Gene* **353**, 155-168.
- Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Meth* **9**, 357-359.
- Mccomish BJ, Hills SFK, Biggs PJ, Penny D (2010) Index-Free De Novo Assembly and Deconvolution of Mixed Mitochondrial Genomes. *Genome Biology and Evolution* **2**, 410-424.
- Miller MA, Pfeiffer W, Schwartz T, 2010 Creating the CIPRES Science Gateway for inference of large phylogenetic trees, pp. 1-8 in *Gateway Computing Environments Workshop (GCE), 2010*.
- Milne I, Bayer M, Cardle L, Shaw P, Stephen G *et al.* (2010) Tablet--next generation sequence assembly visualization. *Bioinformatics* **26**, 401-402.
- Simpson J, Wong K, Jackman S, Schein J, Jones S *et al.* (2009) ABySS: a parallel assembler for short read sequence data. *Genome Research* **19**, 1117-1123.
- Wolff JN, Shearman DCA, Brooks RC, Ballard JWO (2012) Selective Enrichment and Sequencing of Whole Mitochondrial Genomes in the Presence of Nuclear Encoded Mitochondrial Pseudogenes (Numts). *PLoS ONE* **7**, e37142.
- Zerbino D, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research* **18**, 821-829.
- Zhou Z, Huang Y, Shi F (2007) The mitochondrial genome of *Ruspolia dubia* (Orthoptera: Conocephalidae) contains a short A+T-rich region of 70 bp in length. *Genome* **50**, 855-866.
- Zhou Z, Huang Y, Shi F, Ye H (2009) The complete mitochondrial genome of *Deracantha onos* (Orthoptera: Bradyporidae). *Molecular Biology Reports* **36**, 7-12.
- Zhou Z, Shi F, Huang Y (2008) The complete mitogenome of the Chinese bush cricket, *Gampsocleis gratiosa* (Orthoptera: Tettigoniidae). *Journal of Genetics and Genomics* **35**, 341-348.

Zhou Z, Ye H, Huang Y, Shi F (2010) The phylogeny of Orthoptera inferred from mtDNA and description of *Elimaea cheni* (Tettigoniidae: Phaneropterinae) mitogenome. *Journal of Genetics and Genomics* **37**, 315-324.

3. GENE FLOW AND DIVERGENCE OF AN ENDANGERED NEW ZEALAND GRASSHOPPER

ABSTRACT

Gene flow has long been considered a limitation to speciation because selection is required to counter the homogenising effect of allele exchange. Here we report on two species of sympatric short-horned grasshoppers in the South Island of New Zealand, one (*Sigaus australis*) is a widespread species, while the other (*Sigaus childi*) is a narrow endemic. All, bar one, 79 putatively neutral markers (mtDNA, microsatellite loci, ITS sequences and RAD-seq SNPs) examined showed extensive sharing of alleles, and similar or identical allele frequencies in the two species where they co-occur. We found no evidence of deviations from random mating in the region of sympatry. However, analysis of morphological and geometric traits revealed no evidence of morphological introgression. The two species are clearly distinct based on phenotype, but their genotypes thus far reveal no divergence. The best explanation for this is that some loci associated with these morphological characters are under strong selection but exchange of neutral loci is occurring freely between the two species. Although it is easier to define species as requiring a barrier between them, a dynamic model that accommodates gene flow is a biologically more reasonable explanation for these grasshoppers.

INTRODUCTION

Although taxonomy implies abrupt disjunctions between biological entities, we know that speciation usually involves non-instantaneous change (Wright 1982). The existence of hybrids and the implications of hybridisation have long intrigued evolutionists (Darwin 1859), however the incorporation of gene flow into speciation models has only recently gained acceptance (Wu 2001; Mallet 2008).

Genetic introgression occurs when two genetically distinct populations come into contact enabling individuals from the populations to interbreed. When this occurs through secondary contact the process has usually been distinguished as hybridisation (Harrison 1993; Morgan-Richards *et al.* 2009), however broader definitions of this term accommodate the continuum from normal intraspecific mating to rare interspecies exchange (Harrison 2012). The fertility of resulting offspring mediates gene flow between populations. This situation underpins the popular biological species concept (Mayr 1963), but the frailty and circularity of the *a priori* assumption that species are always reproductively isolated is readily demonstrated (Mallet 1995; Mallet 2008). Stark reminders that hybridisation is not a valid test of species status come from observations that it is also a potent force in plant speciation that can result in the rapid formation of distinct and reproductively isolated taxa (Xu 2000; Susnik *et al.* 2007).

Since the 1960s, direct measures of variable genetic loci have provided strong evidence that genomes are not unitary and exchange of loci between populations may be uneven (Barton and Hewitt 1985; Wu 2001; Harrison 2012; Nosil and Feder 2012). Where gene flow between somewhat distinct genomes is not contained by the formation of hybrid zones (Harrison 1993) or abrupt speciation (Soltis and Soltis 2009), it can have numerous outcomes: reinforcement of reproductive barriers, the evolution of a new species, the loss of one or both parental species, limited adaption due to homogenization, or provision of a means to pass adaptive traits between populations (Harrison 1993; Dowling and Secor 1997; Servedio and Noor 2003; Morgan-Richards *et al.* 2009).

Gene flow may therefore result in species with genome mosaics, comprised of alleles from different ancestral populations, which has recently been described as a potentially important evolutionary process in the formation of many animal species (Xu 2000; Susnik *et al.* 2007; Nosil and Feder 2012). Indeed, allelic leakage may be fairly persistent where gene flow is mediated not by extrinsic geophysical barriers, but by locus-specific selection (Barton and Hewitt 1981; Harrison 1990; Thibert-Plante and Hendry 2010; Nosil and Schluter 2011; Harrison 2012). Empirical data showing the maintenance of incipient species in the face of ongoing gene flow between populations are gradually accumulating, aided by increasingly sophisticated genetic tools (Bergsten 2005; Lexer *et al.* 2006; Savolainen *et al.* 2006; Niemiller *et al.* 2008; Papadopoulos *et al.* 2011; Rosenblum and Harmon 2011).

Historically, one of the most informative animal groups in this field of study has been Orthoptera and, in particular, grasshoppers (Key 1968; Barton and Hewitt 1985; Vazquez *et al.* 1994; Willett *et al.* 1997; Kawakami *et al.* 2009; Maroja *et al.* 2009). Here we report on flightless New Zealand shorthorn grasshoppers (Orthoptera: Acrididae). Most of the fifteen New Zealand species, in four endemic genera, occupy subalpine native grasslands above the tree line (Bigelow 1967). Prior to the arrival of humans in New Zealand (~1260 AD), the landscape was mostly dense forest (Mcglone 1985; Mcglone *et al.* 2001; Wardle 2001). Grasshopper habitat was therefore predominantly associated with the mountain ranges of the South Island, although a few species occur at lower altitude in areas with semi-arid climate and braided river-beds (*Brachaspis robustus*, *Sigaus minutus* and *Sigaus childi* (Jamieson 1999; Trewick 2001; Trewick and Morris 2008)).

The species *Sigaus australis* appears, on the basis of mtDNA sequence data, to encompass several narrow endemics and one widespread species (Trewick 2008; Trewick and Morris 2008). Typical *Sigaus australis* are relatively large (adult females ~26 mm) and abundant in South Island subalpine grasslands between 1000 and 1800 meters above sea level. In contrast, the endangered micro endemic *Sigaus childi* is restricted to a low-lying, semi-arid region of about 100km² around the town of Alexandra (Central Otago) (Fig. 1). A notable feature of *S. childi* is their morphological crypsis. Unlike *S. australis* that are boldly patterned and colour

polymorphic within locations, *S. childi* individuals are remarkably camouflaged to the local substrates upon which they are found. Colour and pattern varies enormously, ranging from almost white to grey, brown, green and black, and include individuals that match closely the green and black tumbling lichen (*Chondropsis semiviridis*) that grows on rocks in some areas of Central Otago. *Sigaus childi*, although morphologically distinct, could not be distinguished from *S. australis* using mtDNA data (Trewick 2008).

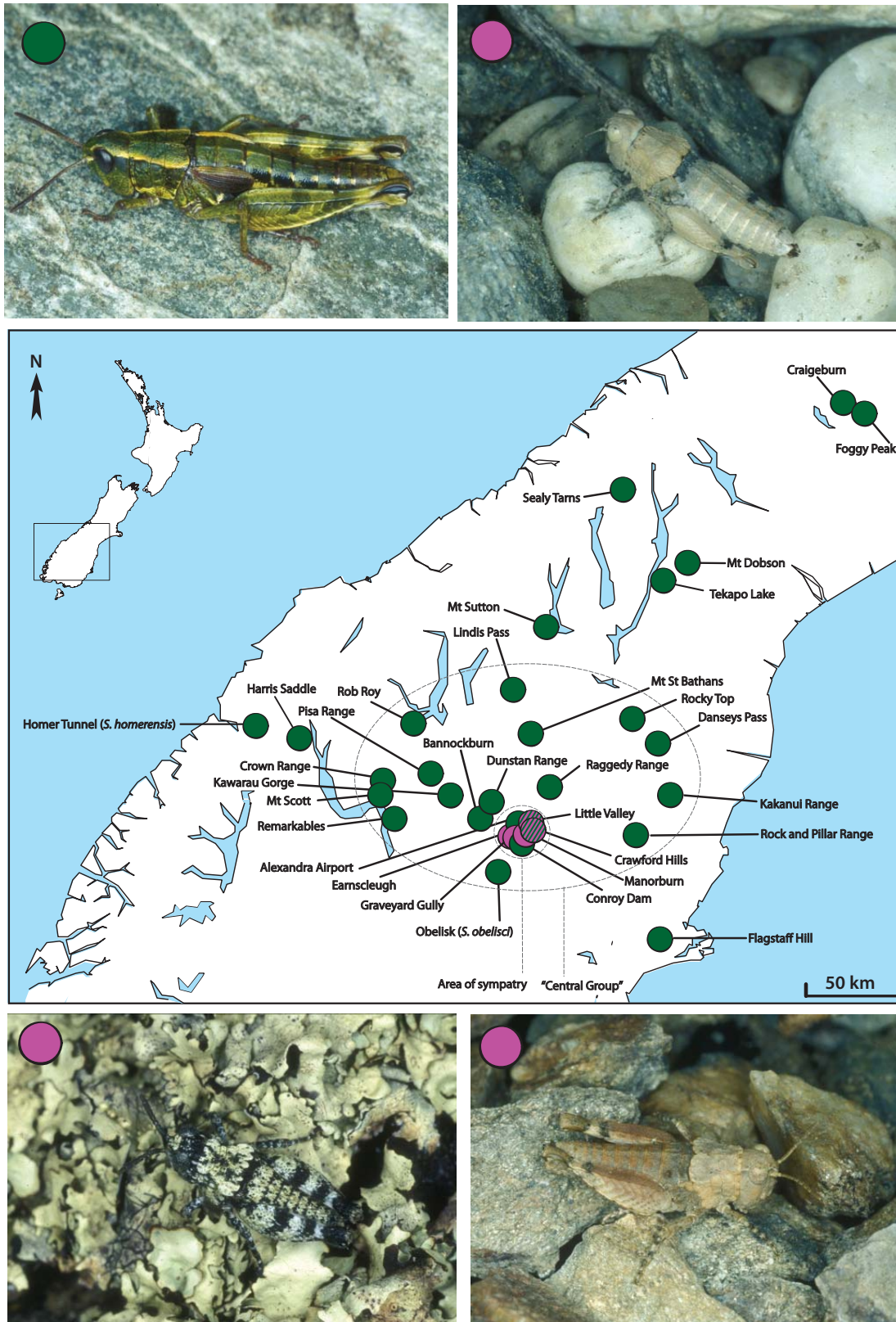


Figure 1

Sample locations in South Island, New Zealand, of the *Sigaus australis* complex grasshoppers used in this study. The two main species *Sigaus australis* (green) and *Sigaus childi* (pink) are morphologically very different; *S. childi* tends to be smaller and more camouflaged to its local habitat than *S. australis*. The 'Central Group' and 'Area of sympatry' defined here are used to analyse subsets of the specimens.

Perhaps *S. childi* evolved recently and has retained ancestral mtDNA haplotypes (Incomplete Lineage Sorting), or has exchanged genetic information since “speciating” (Trewick 2008). Or perhaps divergence is maintained despite gene flow. Genetic exchange between populations might be experienced at different rates across the genome; selection could operate on some loci to limit local exchange of alleles even when net (genome wide) gene flow continues. These alternatives make different predictions about the pattern of morphological and genetic character sharing (Fig. 2). In order to understand the evolution of this system we applied six types of data; morphology, mtDNA sequencing, microsatellite genotyping, multi-copy nuclear sequencing, RAD-seq Single Nucleotide Polymorphic (SNP) data and spatial position. We used these putatively independent data to contrast species integrity as characterised by morphology (subject to natural selection) and neutral characters that allowed us to test the stability of species delimitation, assess the extent and evenness of gene flow and thus gain an understanding of where these grasshopper populations are in the speciation continuum.

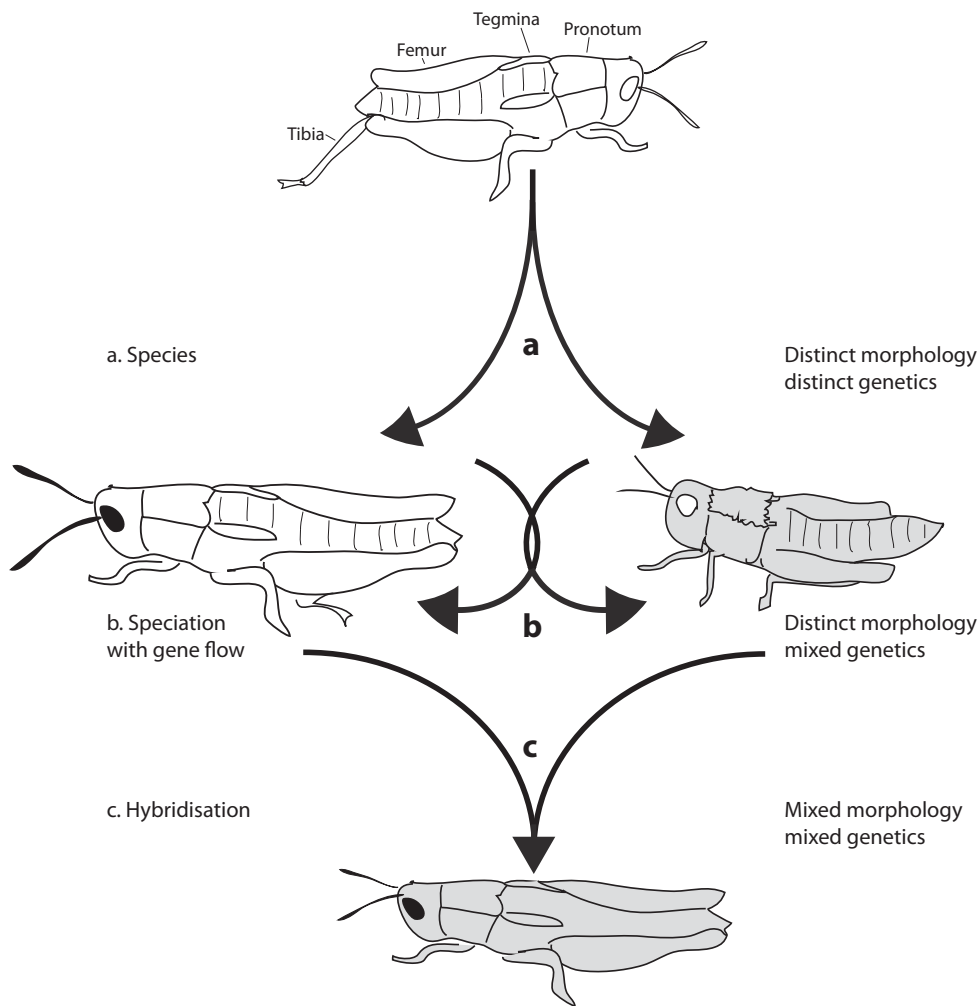


Figure 2

Alternative hypotheses to explain the relationship between morphological differentiation and gene flow in this study of grasshoppers in New Zealand. Two morphologically defined species exist that have a common ancestor and may share identical alleles due to descent, but, as distinct species, derived alleles are also expected. (a) An abrupt speciation event is expected to result in accumulation of distinct traits and genetics, including at neutral loci. (b) Sharing of neutral genetic loci might be maintained by ongoing gene flow whilst alleles at some loci are subject to selection. Morphological difference is an observable expression of genetic loci under diverging selection. (c) Hybridisation or reticulation is expected to result in individuals with intermediate form.

METHODS

We collected grasshoppers by hand when they were active during the New Zealand summer season (December-March, between 1995 and 2009). Sampling included all recognised members of the *S. australis* complex (*S. australis*, *S. childi*, *S. obelisci* and, *S. homerensis*) from their full geographic range (Fig. 1). As already noted, *S. australis* is widespread in subalpine habitat, while *S. childi* occurs at low altitude in Central Otago. Due to the legal protection given to the endangered *S. childi*, sample sizes were limited and sampling spanned more than one overlapping generation. *Sigaus obelisci* and *S. homerensis* are each recorded from single subalpine locations within the range of *S. australis*. Individuals were preserved by freezing or in 95% ethanol and identified following Bigelow (1967), Morris (2003), and Jamieson (1999). The identity of *S. obelisci* and *S. homerensis* were confirmed by Simon Morris (pers comm 2009 to SAT).

Morphology

Morphological data were collected for all adult *Sigaus australis* complex grasshoppers in two ways. The first means used the traditional species diagnostic characteristics, although we note that much of the information used to distinguish some of these taxa has been geographic location and altitude (Morris 2002; Morris 2003b). The exception is *Sigaus childi*, for which the sinuous caudal margin of the pronotum and tegminal size is diagnostic (Jamieson 1999). Male genitalia are taxonomically informative for some grasshopper species but consistent differences have not been reported among species in this complex. For instance, male genitalia in *S. homerensis* and *S. obelisci*, are each described as being near identical to *S. australis* and *S. “remarkables”*, a synonym of *S. australis* (Morris 2002; Morris 2003a; Morris 2003b). Thus male genitalia can be interpreted as being variable within *S. australis*, but uninformative for species delimitation (Bigelow 1967; Jamieson 1999; Morris 2003a).

The traditional species diagnostic characters for the grasshoppers were examined and measured with the aid of a dissecting microscope. Four metrics were recorded in millimetres for each grasshopper using callipers (accurate to 0.01 mm); maximum

pronotum width, mid-line pronotum length, femur length and body length. In addition, six characters with discrete states were examined; sex, length of tegmina (see below), shape of pronotum posterior margin (sinuous or concave), cuticle rugosity (rugose or smooth), shading on pronotum posterior margin (pigmented or not), and shape of pronotum lateral margins (irregular or smooth). For most characters, the alternative and intermediate states were coded as 2, 0 or 1 respectively. The lengths of tegmina were classified by reference to the number of abdominal tergites across which they extended; not beyond the pronotum (as in many *S. childi*) coded as 0, not beyond first abdomen segment coded as 1, and so forth to a maximum of 4 (no tegmina ever reached beyond the posterior margin of the 5th tergite).

Adults were distinguished by the tegmina concealing the relictual hind wing, which is the case only in the last instar. Juveniles were excluded from this morphometric analysis. The data were analysed using both Discriminant Analysis and Principle Component Analysis (PCA) approaches implemented in MINITAB 15 (Minitab 2007). The discriminant analysis with cross validation tested whether character information could be used to group individuals into their *a priori* categories. For these analyses four *a priori* categories were used based on existing taxonomy: *S. australis*, *S. childi*, *S. homerensis* and, *S. obelisci* (3 *S. obelisci*, 3 *S. homerensis*, 28 *S. childi* and 134 *S. australis*). Specimens of *Sigauss obelisci* and *S. homerensis* were included to provide context for any difference between *S. childi* and *S. australis*. A PCA was applied to all the morphological characters and the scores saved. PCA required no *a priori* grouping, allowing us to determine whether the data could be partitioned into taxonomically meaningful groups based on the documented morphological character states.

As an alternative to the traditional taxonomic characters, we tested for shape differences of the pronotum among species using geometric analysis. Much of the taxonomy in the *Sigauss* genus relies on the pronotum shape, but descriptions are often vague, based on discrete states and inferred from few individuals making species identification difficult (Bigelow 1967; Morris 2002; Morris 2003b). Using two digital images of the pronotum of each of 149 individuals (90 *S. australis*, 36 *S.*

childi, 20 *S. obelisci* and 3 *S. homerensis*) obtained with the aid of a dissecting microscope we tested whether shape variation could be detected from metric data. Using IMAGEJ (Abramoff *et al.* 2004), 14 landmarks were identified around the perimeter of the dorsal surface of the pronotum on each image of each grasshopper and measured. The landmarks were selected to maximise variation among individuals. These measurements were analysed using MORPHOJ (Klingenberg 2011). A Procrustes fit aligned by principal axes was performed to eliminate size differences before a Procrustes ANOVA was used to examine the error of image capture. This analysis revealed that the error arising from image capture variation was biologically irrelevant: mean squares for image capture was 32 times smaller than the variation found between individual grasshoppers. Juveniles, adults and both sexes were included in the analyses and tested to confirm they did not partition in the results. A PCA and discriminant analysis with cross validation were each performed on the averaged value for each individual from all four species (34 *S. childi*, 21 *S. obelisci*, 3 *S. homerensis* and 90 *S. australis*). A PCA and discriminant analysis with cross validation were also separately applied to the *S. childi* and *S. australis* individuals collected within the area of sympatry (34 *S. childi* 11 *S. australis*) (Fig. 1).

Mitochondrial DNA sequence

In order to improve genealogical resolution in relation to taxonomy and geography, we supplemented existing published mtDNA Cytochrome Oxidase Subunit I (COI) data for the *Sigauss australis* grasshopper complex (Trewick 2008). Tissue was dissected from femora of recently collected grasshoppers and DNA extracted using a salting-out method (Sunnucks and Hales 1996; Trewick and Morgan-Richards 2005). DNA from specimens preserved for more than one year was extracted using incubation at 55°C with Proteinase K and a CTAB buffer (2% Hexadecyltrimethylammonium bromide, 100 mmol/L Tris-HCl pH8.0, 1.4 mol/L NaCl, 20 mmol/L EDTA), followed by a combined phenol/chloroform/isoamyl alcohol (25:24:1) cleanup. Extractions were eluted in water and diluted as necessary for PCR reactions. Primers C1-J-2195 and LI-N-3014 (Simon *et al.* 1994) were used to target the 3' portion of COI. Polymerase chain reactions (PCRs) were performed in 10 µl volumes using ABgene Red Hot Taq (Thermo FisherScientific). PCR used the

protocol: 94°C for three minutes, 94°C for 45 seconds, 52°C for 45 seconds and 72°C for 75 seconds repeated 36 times; followed by a 2 minute anneal. Cycle sequencing used Perkin Elmer BigDye 3.1 chemistry following the manufacturer's protocols analysed on an ABI Prism 377 DNA sequencer (Applied Biosystems, Inc., Foster City, California). Sequences were checked using SEQUENCER version 4.10.1 (Gene Codes) and aligned with existing data using SeAl version 2.0 and GENEIOUS PRO version 5.3.4 (Rambaut 2002; Drummond *et al.* 2011).

GENEIOUS was used to estimate a neighbour-joining tree for all the lineages. NETWORK version 4.5.1.6. (Bandelt *et al.* 1999) was used to estimate haplotype networks within clades. To test for a correlation between genetic and geographic distance (expected under a model of isolation by distance), Mantel tests (Mantel 1967) were performed using ISOLATION BY DISTANCE WEB SERVICE version 3.16 (Jensen *et al.* 2005) with 10,000 randomizations to assess the significance of distance correlations. Distance by distance analysis was applied to all data and separately to data within haplo-groups. A standard AMOVA was used to test for significant genetic differences based on the estimate of genetic partitioning among groups (F_{CT}) using ARLEQUIN version 3.5.1.2 (Excoffier *et al.* 2005). The first run tested *S. childi* against all populations in the central group (as indicated in Fig. 1) and the second included only those individuals from the area of sympatry (Fig. 1).

Microsatellites

To obtain neutral nuclear loci to examine population structure, we developed primers to amplify microsatellite loci using a modified enriched microsatellite library protocol (S2.1 S2.2).

Screening of fifty microsatellite loci revealed three that were polymorphic and amplified consistently among a subset of DNA samples from the target taxa. The loci were checked for large allele dropout, stuttering, and null alleles using 1000 randomisations in MICROCHECKER version 2.2.3 (Van Oosterhout *et al.* 2004). Not all populations had sufficient sampling to analyse in MICROCHECKER and within the area of sympatry the *S. childi* and *S. australis* populations were treated as a single population. Although there was evidence of null alleles in some of the loci within

some of the populations (*S. childi* sympatry and *S. australis* sympatry), this is unlikely to influence the detection of genetic differentiation (Carlsson 2008). To test a hypothesis of isolation by distance, geographical distances (km) among pairs of *S. australis* population samples were linearly regressed against their pairwise F_{ST} estimates. Nineteen populations had sufficient sampling for this analysis (Fig. 1). Mantel testing (Mantel 1967) was performed using ISOLATION BY DISTANCE WEB SERVICE version 3.16 (Jensen *et al.* 2005) with 10,000 randomizations to assess the significance of distance correlations.

Population structure was assessed without *a priori* groupings using STRUCTURE version 2.3.4 (Pritchard *et al.* 2000). Firstly we looked for evidence of population structure in the data from *S. australis* samples only, secondly among all samples from all areas sampled for the complex, and thirdly among all samples collected in the area of sympatry (*S. childi* and *S. australis*) and central group *S. australis* (Fig. 1). The analyses were run using an admixture model with independent allele frequency, 100,000 generations burn-in followed by 100,000 generations, and the number of groups (K) set from 1 to 20. The optimum value of K was found using the ΔK method except for K=1, which was determined by examination of the bar-plots (Evanno *et al.* 2005). We sought evidence of genetic differentiation concordant with morphology using the populations within the central group (Fig. 1) subset identified by STRUCTURE (Fig. 1). A standard AMOVA was used to test for significant genetic differences based on the estimate of genetic partitioning among groups (F_{CT}) using ARLEQUIN version 3.5.1.2 (Excoffier *et al.* 2005). The first run tested *S. childi* against all populations in the central group (Fig. 1) and the second only those *S. childi* and *S. australis* individuals from the area of sympatry (Fig. 1).

Nuclear Sequence

Nuclear sequences representing the internal transcribed spacers (ITS1 and ITS2) of the rRNA cluster and the intervening rRNA 5.8S gene were obtained using the primers ITS4 and ITS5 (White *et al.* 1990). PCR conditions and sequencing followed standard protocols as above. Sequences were aligned using GENEIOUS PRO version 5.3.4 (Drummond *et al.* 2011) and checked by eye. Sequences were generated for all grasshoppers from the area of sympatry of *S. childi* and *S. australis* (the Alexandra

region). Alignment and comparison of unambiguous with ambiguous sequences allowed us to identify the most likely combinations of sequences that returned the observed heterozygosity (Table S2.4). Where sequence variants differed by single nucleotide substitutions we could identify and resolve the polymorphism. Where sequence variation involved INDELs the resulting length polymorphism was evident by abrupt onset of sustained nucleotide ambiguity at the INDEL position, but sequencing in both directions allowed identification of the combination of sequences involved. To examine the number of families of ITS per grasshopper genome we interrogated a DNA dataset generated by high throughput sequencing. Genomic DNA from a single *Sigaus australis* individual from Lindis Pass was sequenced on an Illumina Hi-Seq 2000 (Beijing Genomics Institute) resulting in >1GB of sequence. The sequence was de-novo assembled via VELVET (Zerbino and Birney 2008) with mapping performed using BOWTIE version 2 (Langmead and Salzberg 2012) and the results viewed in TABLET version 1.12.09.03 (Milne *et al.* 2010). The resulting contigs were blasted to Genbank and all matches to ITS were selected, aligned, mapped back and, checked for copy number.

RAD-Seq SNPs

Single nucleotide polymorphic (SNP) anonymous nuclear markers were generated using high throughput sequencing, with individual DNA fragments coded so we could identify individual grasshopper genotypes. The double digest Rad-Seq protocol (Peterson *et al.* 2012) was applied with minor modifications. We estimated genome size to help us optimise the selection of endonucleases and sequencing coverage. We used flow-cytometry to estimate genome size using a FACSCalibur system and CellQuest software (BD Biosciences, San Jose, CA, USA), following staining of cells with Propidium Iodide and reference to an internal control (chicken and locust). Our estimates of the *Sigaus* genome were approximately 11.9 pg (consistent with estimates of other short-horned grasshopper species (<http://www.genomesize.com>). In light of this information we used the restriction enzymes PstI and BamHI to digest the whole genomic DNA extracted from 30 grasshoppers (10 *S. childi*, 10 *S. australis* from Alexandra, 10 *S. australis* from Lindis). The DNA fragments were tagged with DNA sequences that identified each individual before pooling as per Peterson *et al.* (2012).

Data were generated using an Illumina Hi-Seq (New Zealand Genomics Limited), and sorted using the STACKS version 0.99992 pipeline (Catchen *et al.* 2011). Settings for coverage and sites per read were adjusted iteratively. Read coverage settings vary in the literature (Peterson *et al.* 2012), so we ran *various* (7-30) reads with the data but with no alteration of the results. *For* the results described we selected an optimum coverage of 15 reads per individual (excluding all stacks with a lower coverage). We allowed a maximum of two mismatches between alleles for a single individual and four mismatches between primary and secondary reads within ustacks. We allowed the program to remove any potentially spurious highly repetitive stacks. In cstacks we allowed 3 mismatches between samples when generating the SNP set (-m 15 -N2 -M4 -n3 -t). We restricted our analysis to a single SNP per putative locus (always the first), thus avoiding potential problems of non-independence between markers. Data file conversion for programs was performed using PGDSPIDER version 2.0.4.0 (Fischer *et al.* 2011). Population pairwise F_{ST} were calculated for each putative-loci across all loci in STACKS, and an AMOVA was run in GENODIVE version 2.0b24 (Meirmans and Van Tienderen 2004) to determine F_{ST} across populations. STRUCTURE version 2.3.4 (Pritchard *et al.* 2000) was used to estimate population differentiation using an admixture model with independent allele frequency, a burnin of 100,000 generations was followed by 100,000 generations of (with 10 replications), and the number of groups (K) set from 1 to 3. The optimum value of K was found from ΔK method except for K=1, which was determined by examination of the bar-plots (Evanno *et al.* 2005). STRUCTURE was run using one SNP per putative-locus (read), and each putative-locus appeared in $\geq p2$ populations, and occurred in $\geq 50\%$ of the individuals.

Gene flow between the two populations in sympatry was estimated using MIGRATE-N version 3.5.1 (Beerli 2006; Beerli 2009), although algorithms that test for gene flow are often not ideal for situations where gene flow is very high, which is likely in this case. MIGRATE-N was run using the bayesian inference strategy. Initial runs involved only half the markers as we optimised settings. The starting values for θ and M were generated from F_{ST} estimates in the first run; following runs used the θ and M values generated from the previous run. The uniform prior distributions were

used for both parameters with slice sampling, one long chain was run recording every 5 steps after a burnin of 50,000 with a static heating scheme with five chains. Four runs were conducted on half the data before three final runs on all loci were undertaken using the starting values for θ and M of the previous run. BAYESCAN version 2.01 (Foll and Gaggiotti 2008; Foll *et al.* 2010; Fischer *et al.* 2011) was used to examine the individual markers for evidence of selection using the default settings. Prior odds of a neutral model were 10 times more likely than the model with selection at a locus. This prior was tested further by changing the prior to one, without altering the results. The alpha value was used to determine the direction of selection with a positive value suggesting diversifying selection and a negative value suggesting balancing selection. Results were viewed in R version 3.0.0 (R Core Team 2013) and markers with evidence of selection were subjected to a BLAST search via NCBI (Johnson *et al.* 2008).

RESULTS

Morphology

Discriminant analysis of traditional morphological data collected from species description traits of 169 adult grasshoppers revealed strong support for the two entities *Sigauss childi* and *S. australis* (based on the squared distance between groups), but there was no support for the other species as distinct morphological units (Table 1 & 2). All 28 adult *S. childi* were readily separated from the remaining sample containing *S. australis*, *S. obelisci* and *S. homerensis*, which could not be distinguished from one another.

Table 1

Discriminant analysis with cross-validation using the character states employed in traditional species diagnosis for adult grasshoppers of the *Sigauss australis* complex in South Island, New Zealand. Only *S. childi* (in bold) individuals were correctly grouped together, the others (*S. australis*, *S. homerensis* and *S. obelisci*) had mixing between species, although never with *S. childi*.

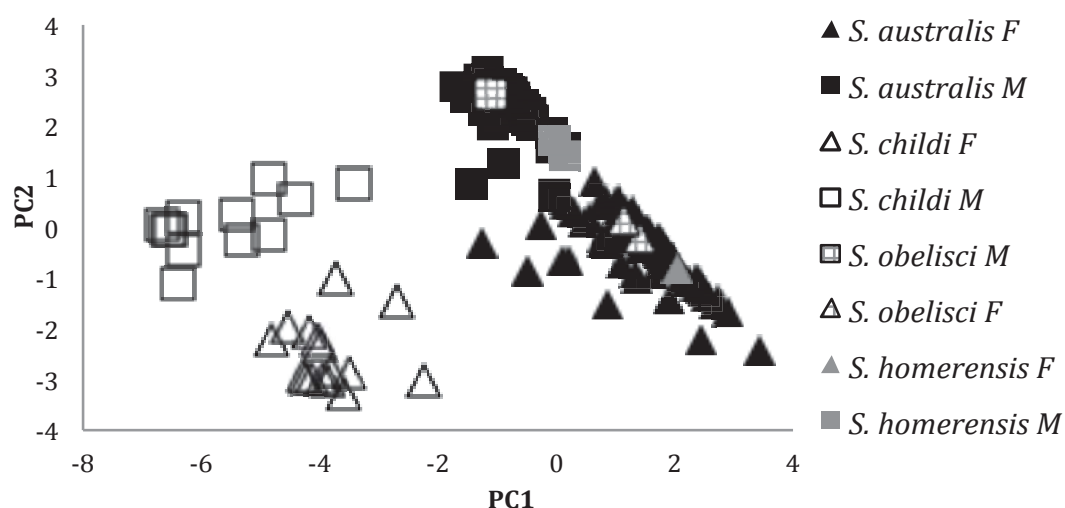
Summary of Classification with Cross-validation				
True Group				
Predicted Group	<i>S. australis</i>	<i>S. childi</i>	<i>S. homerensis</i>	<i>S. obelisci</i>
<i>S. australis</i>	82	0	0	2
<i>S. childi</i>	0	28	0	0
<i>S. homerensis</i>	39	0	1	1
<i>S. obelisci</i>	13	0	0	0
Total N	134	28	3	3
N correct	83	28	1	0
Proportion	0.615	1.000	0.333	0.000
N = 168 N Correct = 112 Proportion Correct = 0.663				

Table 2

Discriminant analysis of traditional species diagnostic characters. *Sigauss childi* (in bold) was the only species of the four *Sigauss australis* complex grasshoppers that was well supported. There is no evidence of mixing with other species within *S. childi*.

Squared Distance Between Groups				
	<i>S. australis</i>	<i>S. childi</i>	<i>S. homerensis</i>	<i>S. obelisci</i>
<i>S. australis</i>	0.0			
<i>S. childi</i>	274.2	0.0		
<i>S. homerensis</i>	1.6	280.9	0.0	
<i>S. obelisci</i>	6.2	272.3	4.8	0.0

Principal component analysis also revealed two distinct groups; *S. childi* vs. all other *Sigauss* specimens in the analysis. Both groups were further subdivided into males and females (Fig. 3). The first four components of the PCA accounted for >95% of the variation. Thus the morphometric data based on traditional taxonomic characters suggest there are just two morphological entities: *S. childi* and *S. australis*. This is consistent with the tenuous distinction in the literature of *S. obelisci* and *S. homerensis*, which are primarily characterised by location.

**Figure 3**

Principle component analysis using morphological character states used in traditional species diagnostics of adult *Sigauss* grasshoppers from South Island, New Zealand. Morphologically, *Sigauss childi* is readily separated from *S. australis*, *S. obelisci*, and *S. homerensis*. F = females (triangles); M= males (squares).

Geometric analysis of the grasshopper pronotum gave a similar result to that of traditional morphology. PCA of the 14 landmark measurements showed two major groupings (Fig. 4a). One was composed of *S. australis*, *S. obelisci* and *S. homerensis* and the other composed of *S. childi*. Individuals referred to as *S. obelisci* shared a more rounded posterior edge of the pronotum with a significantly different mean to that of *S. australis* individuals ($P < 0.001$ T-square 334.7159). However, most *S. obelisci* individuals did not fall outside the shape range of *S. australis* suggesting that neither *S. obelisci* nor *S. homerensis* are distinct based on pronotum morphology (Fig. 4a). All *S. childi* individuals grouped together due to their distinctive pronotum shape. Variation within *S. childi* was likely due to their extremely cryptic shape formed by the 'broken' edge of the pronotum resulting in little uniformity within species. The *S. childi* mean was significantly different from both *S. obelisci* ($P < 0.001$ T-square 668.5302) *S. australis* over the entire range ($P < 0.001$ T-square: 1101.9925) and *S. australis* in the area of sympatry ($P < 0.001$ T-square: 935.5225).

The PCA of pronotum shape from those grasshoppers sampled within the area of sympatry was scrutinised for evidence of hybrids (Fig. 4a). However, within the area of sympatry, not only did the means of *S. australis* and *S. childi* differ ($P < 0.001$ T-square 1291.4541) but there was no overlap between the two forms. Thus we found no evidence of morphological hybrids in adults or juveniles.

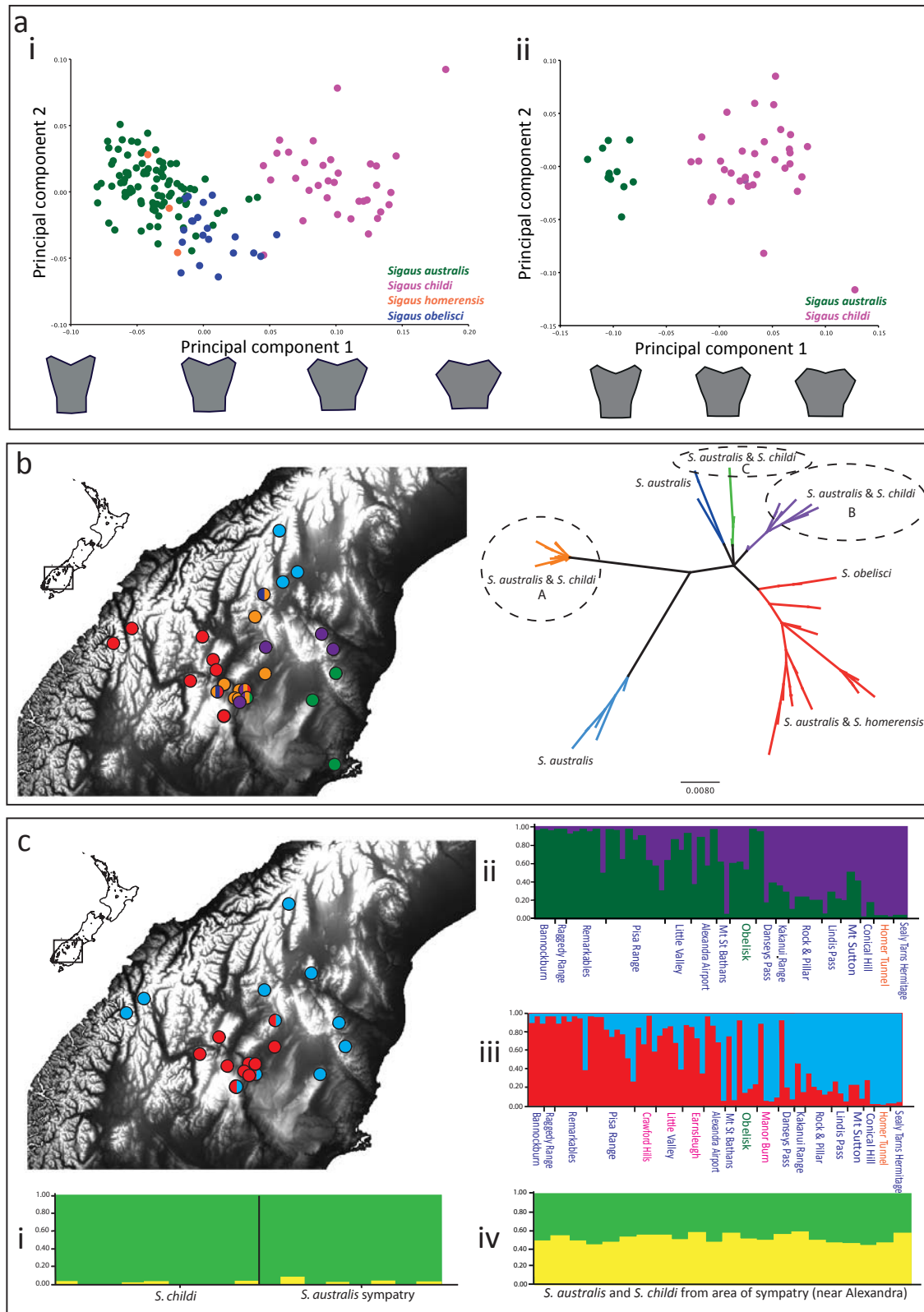


Figure 4

Genetic and morphological structure within *Sigaus* grasshoppers in South Island, New Zealand. (a) Variation in the shape of the pronotum using digital imagery of *Sigaus* grasshopper pronotum shape (PCA analysis from MORPHOJ): (i) PCA for all species from all areas, with pronotum shape changes indicated along the PC1 axis. The two major groupings comprise *S. childi* separated from *S. australis*, *S. obelisci* and *S. homerensis*; (ii) PCA for just

individuals from the area of sympatry (Fig. 1), with pronotum shape changes indicated along the PC1 axis. Within the area of sympatry there was no clear evidence of morphological intermediates. (b) Neighbour Joining tree of mtDNA haplotypes (COI, 519 bp) and a distribution map showing the spatial distribution of haplogroups. The circled clades (dashed circles) are used in the Network analysis (Fig. 5). (c) Genetic structure genotype data (STRUCTURE analysis) for the *Sigaus* grasshoppers: (i) Results from the 74 RAD-seq SNPs of the two species in sympatry, *S. australis* and *S. childi* K=2 (ii) Microsatellite genotypes from *S. australis*, *S. obelisci* and *S. homerensis* populations at K=2; (iii) Microsatellite genotypes from *S. australis*, *S. obelisci* and *S. homerensis* and *S. childi* individuals (colours shown on map) at K=2; (iv) Microsatellite genotypes from *S. australis* and *S. childi* individuals from the area of sympatry (see Fig. 1) at K=2.

Mitochondrial DNA sequence

Partial Cytochrome oxidase subunit I was sequenced from 59 grasshoppers and combined with 46 previously published sequences (Trewick 2008) (GenBank EF544487–EF544562). A total of 66 haplotypes were identified in the alignment (519 bp) representing 105 individuals (72 *S. australis* and 33 *S. childi*). Six haplotype clusters were identified that were each geographically restricted within the range of *Sigaus australis* (Fig. 4b). The MtDNA clusters did not correspond to current taxonomic groups or morphological types. MtDNA sequences from *S. childi* fell in several parts of the *Sigaus australis* complex phylogeny (Fig. 4b). NETWORK analysis revealed the extent of sharing between *S. australis* and *S. childi* (Fig. 5). As a result of interspecific sharing, some *S. childi* have haplotypes more similar to haplotypes in *S. australis* than other *S. childi*. In earlier work one mtDNA cluster contained most *S. childi* haplotypes (Trewick 2008), but the additional sampling in the current study indicates that no single mtDNA clade can be confidently construed as being associated with *S. childi* ancestry. The high genetic diversity detected at Alexandra appears to be a result of the meeting of four distinct mtDNA clades that otherwise have separate ranges.

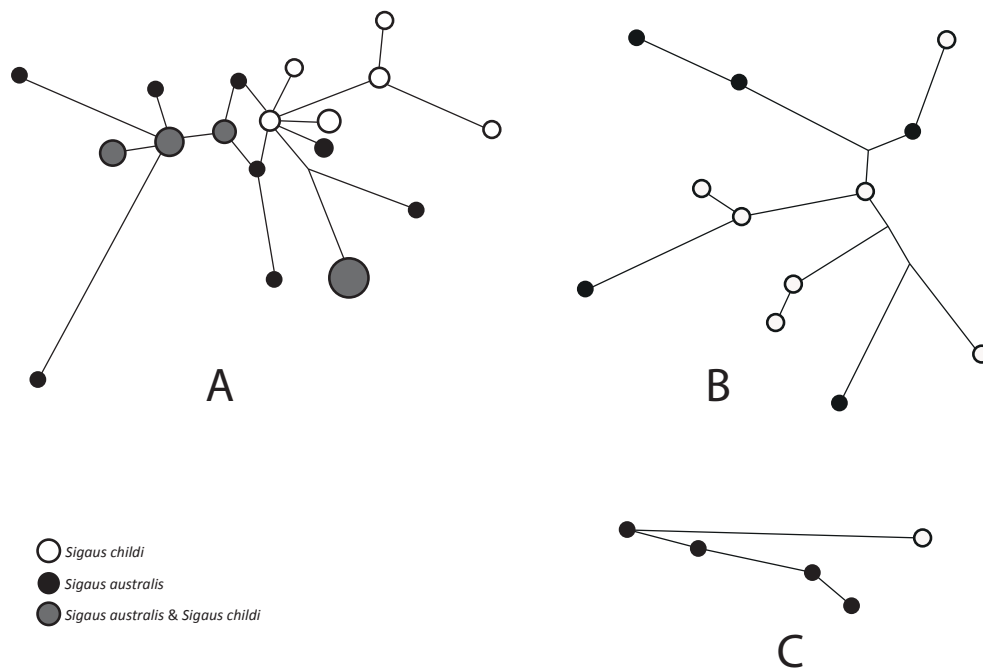


Figure 5

Minimum spanning network of three of the six mtDNA (COI 519bp) haplogroups (Fig 4b) within the *Sigaus australis* grasshopper complex, that show sharing (grey) of identical haplotypes between *S. childi* (white) and *S. australis* (black). Haplotype spot size is proportional to number of individuals with a particular haplotype and branch length estimates nucleotide differences.

No evidence for isolation by distance was detected among the mtDNA diversity within the *S. australis* complex ($P = 0.5410$). Although the mtDNA diversity within clades also did not fit a model of isolation by distance there was a non-significant positive relationship ($P = 0.1410$) and the power of this test was limited by smaller within-clades sample sizes. There was no statistical support for population genetic differentiation between *S. childi* and *S. australis* from the central group (Fig. 1) $F_{CT} = 0.09175$ ($P = 0.09677$), or from the area of sympatry (Fig. 1) $F_{CT} = 0$ ($P = 1$).

Microsatellites

The three microsatellite loci surveyed each had between 16 and 18 alleles. No evidence of linkage disequilibrium was detected and Hardy-Weinberg expectations were met in the majority of population samples. A positive relationship between geographic distance and genetic differentiation (pairwise F_{ST}) supported a model of

isolation by distance for the *Sigaus australis* populations, including *S. obelisci* and *S. homerensis* (Fig. 1) (57 individuals; $P = 0.0108$). Analysis of *S. australis* microsatellite data (excluding *S. childi* samples) using STRUCTURE (Pritchard *et al.* 2000) showed evidence of extensive gene flow among populations. The optimum ΔK was $K = 2$ (Fig. 4c), consistent with a grouping of populations in Central Otago (Fig. 1 & 4b).

Analysis of microsatellite data from *S. australis* and *S. childi* (13 *S. childi* 57 *S. australis*) individuals resolved the same geographic subdivision of genetic variation: $K = 2$ (Fig. 4c). At $K = 3$, structure was lost, which was unexpected given that the data encompassed two morphologically distinct species and spatial structure had already been indicated. To reduce the possible influence of uneven sampling of the two species we restricted the data to include sampling only from the area of sympatry (Fig. 1). STRUCTURE (Pritchard *et al.* 2000) analysis found no support for genetic partitioning within these data (i.e. $K = 1$), contrary to the expectation that the two morphologically defined species represent discrete genetic units (Fig. 4c).

As with the mtDNA data we sought evidence of genetic structure concordant with taxonomy and morphology with analysis of the correlation of genotypes between species by grouping the samples according to the morpho-species *S. australis* and *S. childi* and estimating F_{CT} . No significant genetic differentiation between the morpho-species was found within the 'central group' (Fig. 1): $F_{CT} = 0.01780$ ($P = 0.18573$), or the area of sympatry (Fig. 1) $F_{CT} = -0.07424$ ($P = 0.65494$), although these samples were not taken from the same generation, supporting the inference from STRUCTURE.

Nuclear Sequencing

We amplified and sequenced the ITS region (706 bp including 5.8S, ITS1 and 2) from 40 grasshoppers. Some of the grasshoppers had unambiguous single ITS sequence but many had more than one ITS sequence, consistent with these grasshoppers being heterozygotes of mixed ancestry. Of twenty-five grasshoppers (15 *S. childi* and 10, *S. australis*) collected near the township of Alexandra, 16 (11 *S. childi*, 5 *S. australis*) had more than one sequence which differed by the presence of an INDEL approximately 100 bp from the ITS1 forward primer. Sequences of ITS2 from these

individuals were unambiguous except at single nucleotide polymorphic sites (SNPs), confirming that these grasshoppers carried more than one ITS sequence per genome. There were 16 SNPs in the set of unambiguous sequences but the presence of the INDEL near the start of ITS1 meant that when grasshoppers had more than one sequence, only 13 SNPs were observable. Only one of the 15 grasshoppers from outside the Alexandra area (a specimen from Raggedy Range) appeared to have more than one ITS sequence per genome that involved the large INDEL. However, many individuals (mostly collected from the northern part of the species' range) had an independent 8 bp insertion that occurred in all their copies of ITS. The grasshoppers in which more than one ITS sequence was detected comprised ambiguity that could be explained by combinations of unambiguous (single copy) sequences that we identified in other grasshoppers. From the *Sigaus australis* genomic sample only one copy of ITS could be detected, suggesting that there was only one family of ITS in these grasshoppers, with sequence variation within this family occurring where they have recently exchanged genetic material.

RAD-Seq SNPs

Available analytical software for the paired-end Illumina sequencing protocol was not compatible with the sequence tags on our DNA fragments. Base composition bias in the first four base pairs (SNP site) in the second read was outside of the limits set by the software and thus we had to exclude this portion of our data from further analysis. We present results from the first read only. The Illumina sequencing provided 9,789,323 first end reads of 100 bp for the 30 grasshopper of which 8,934,377 were retained after quality checks in `process_radtags.pl`. This comprised between 3,743 and 978,246 reads per individual with a total of 30,439 loci. From these data sets we identified 8,958 loci that occurred in > 2 individuals which were subjected to further selection based upon coverage per population. The Lindis population was removed from subsequent analysis as several individuals failed to produce data of sufficient quality (likely due to poor DNA quality) and most putative-loci resolved were not represented in the other samples. Of those 8,958 loci, 74 were retained as they occurred in $\geq 50\%$ of the two populations (*Sigaus childi* and *Sigaus australis* in sympatry). The relatively low proportion of loci that were represented across both population samples was due to insufficient

representation of their very large genome. Nevertheless we obtained ample data for our purpose.

The distribution of pairwise (*S. childi*/*S. australis*) F_{ST} values for each of the 74 putative-loci revealed an L-shaped distribution as expected in the absence of significant structure (Figure 6 a). A test for deviations from expected frequencies of neutral loci indicated that one marker may have been subject to diversifying selection, $\log(PO) > 0$ alpha 0.878 (Figure 6 b). A BLAST search of the sequence containing this SNP did not result in any matches to known sequences on Genbank. Mean population pairwise F_{ST} was low 0.025, with a confidence interval that effectively included zero (CI 2.5% 0.001 CI 97.5% 0.053), providing little evidence that these samples represent more than one population with random mating. Population differentiation estimated with STRUCTURE ($\Delta K = 1$) suggesting extensive sharing of genetic material among populations, with no species structure detected (Figure 4bi). This was confirmed by analysis in MIGRATE-N, which indicated extensive gene flow between the species in the zone of sympatry (Table 3).

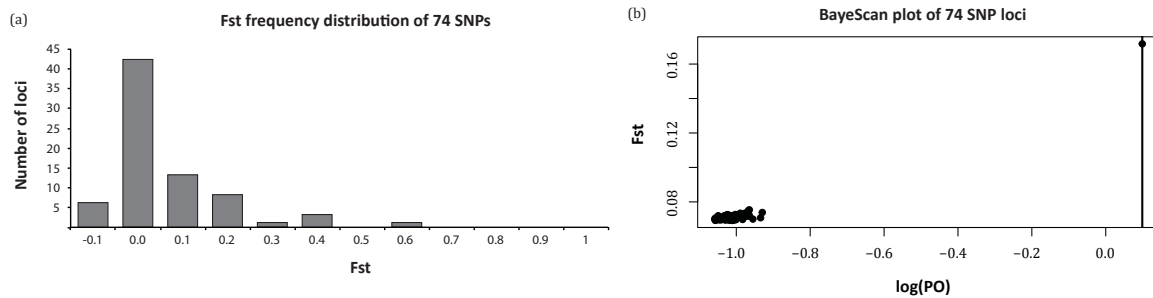


Figure 6

(a) Frequency distribution of locus specific F_{ST} values for each of the 74 SNP loci sampled between *Sigauss childi* and *Sigauss australis* in sympatry. (b) BAYESCAN plot of 74 SNP loci with a single marker showing slight departure from neutrality; the vertical line is the 5% threshold of false discovery. The F_{ST} frequency distribution shows Weir and Cockerham F_{ST} values which differ from Bayescan values as this uses an alternative method of calculation.

Table 3

Extensive gene flow between *S. childi* and *S. australis* in sympatry was revealed using MIGRATE-N with 74 RAD-seq SNP makers. Theta θ is an estimate of population size, $\theta = 4N_e\mu$ in the SNPs, where N_e is population size and μ is mutation rate, and was generally large as is expected for grasshoppers. Mutation scaled migration rates (M) were converted into Nm (number of migrants) via $\theta_1 M_{2 \rightarrow 1} = 4Nm_1$. The results show extensive gene flow in both directions.

All loci	Mean	0.025	0.975	Nm
θ_1	0.09632	0.08413	0.10907	
θ_2	0.08723	0.07547	0.09907	
$M_{2 \rightarrow 1}$	2046.7	1927.3	2051.0	49.3 (2→1)
$M_{1 \rightarrow 2}$	3010.8	2874.7	3010.7	65.6 (1→2)

DISCUSSION

We found morphological support for two distinct species: *Sigauss australis* and *Sigauss childi*. *Sigauss australis* has a comparatively wide geographic range that can be subdivided into a number of phylogeographically distinct mtDNA haplogroups. *Sigauss australis* appears to subsume the other entities examined (*S. obelisci* and, *S. homerensis*) that are defined largely on location information. Although *S. obelisci* samples had a somewhat distinct pronotum shape, most individuals did not fall outside the range of other *S. australis* populations (Fig. 4a). *Sigauss childi* is nested within *S. australis* in terms of habitat, geographic range and genetic diversity, in stark contrast to its clear morphological distinction. We found no evidence of genetic partitioning in putatively-neutral mtDNA sequence, microsatellite, ITS sequence loci or SNP data. None of the mtDNA diversity detected within *Sigauss australis* is concordant with current taxonomic subdivision. Microsatellite allele frequencies within *S. childi* are indistinguishable from those within sympatric *S. australis* suggesting recent (and on-going) gene flow. The SNP data show no population structure and extensive gene flow between the two species in sympatry, with one marker showing some sign of diversifying selection. The presence of more than one ITS1-ITS2 sequence within the genomes of single grasshoppers is also consistent with recent gene flow near the township of Alexandra.

One potential explanation for this lack of genetic differentiation is that *S. childi* and *S. australis* are different phenotypes generated by different local environmental conditions. Indeed, *S. childi* is restricted to a limited lowland habitat in contrast with the subalpine environment experienced by most *S. australis*. The lowland conditions of Central Otago have been described as semiarid (Garnier 1958; Hubbard and Wilson 1988) and, although altitudinally separated from subalpine areas, these habitats may be climatically similar in terms of their extreme day/night and seasonal temperature cycles. However, *S. australis* and *S. childi* are found in close temporal and physical sympatry in the lowland semiarid environment of Central Otago. For example, both morphotypes were collected as adults from Little Valley Alexandra on the same day, within metres of each other. These circumstances are inconsistent with an interpretation of phenotypic differences being driven by

environmental induced plasticity, although we cannot exclude this possibility entirely. Instead, the observed morphological divergence but lack of genetic structure is consistent with strong character specific selection in the presence of high levels of gene flow (Fig. 2b). With high levels of gene flow, neutral markers may not detect population structuring (Thibert-Plante and Hendry 2010). Although contemporary introgression of this sort can be difficult to distinguish from incomplete lineage sorting (Jason *et al.* 2001; Van Oppen *et al.* 2001), the sharing of identical mtDNA haplotypes and sharing of neutral nuclear loci in these two species suggests that very recent or ongoing reticulation is more likely. Because *Sigaus childi* is an endangered species, sample collection covered a number of years. Sampling from different generations, which are already overlapping (c.f. assumptions of population genetic models) is not likely to increase the similarity of population allele frequencies, and therefore we do not consider this will have prevented us detecting significant population structure if it were present.

Gene flow tends to homogenise variation within species and it has traditionally been considered that speciation is unlikely to occur in the presence of gene flow (Mayr 1963; Coyne and Orr 2004). Many models of speciation have emphasized partitioning of populations by some extrinsic process (e.g. allopatry) as a prerequisite for evolution (Mayr 1963; Coyne and Orr 2004). However, if selection on particular loci is sufficiently intense, the effects of gene flow could be mitigated (Wang *et al.* 1997; Wu 2001; Peccoud *et al.* 2009; Nosil and Schluter 2011; Cui *et al.* 2013).

Different evolutionary responses to selection could explain the observed morphological distinctions. *Sigaus childi* is a small highly cryptic species of grasshopper suggesting it has been or is under selection from visual predators. In contrast, *Sigaus australis* is a larger grasshopper, usually with striking colour markings that are more easily observed against the substrate, suggesting a different mode of predator avoidance. Although the two species appear to share the same neutral alleles, they remain morphologically distinct with a single genetic marker showing some evidence of diversifying selection between the two species.

All individuals examined fell into one of two morphological groups and no specimens could be classed as morphologically intermediate within the area of sympatry even when juveniles were examined. The lack of morphological hybrids suggests that selection (even in today's highly modified environment) is intense and there is a contribution from assortative mating. The conservation status of *S. childi* limited our sample sizes and further examination may reveal intermediates. Work on mating behaviour and juvenile colouration and survival where the species are sympatric would help clarify this.

Situations where speciation and selection are most likely to be observable in nature are those with high environmental heterogeneity, temporal instability and/or novel environments (Orr and Smith 1998; Forsman *et al.* 2011). Not surprisingly, many examples of contemporary speciation in action therefore come from anthropogenic settings (Palumbi 2002; Stockwell *et al.* 2003), and this situation may apply to these *Siga*s grasshoppers. The South Island of New Zealand was settled by Polynesian colonists starting about 800 years ago, and this was accompanied by episodes of scrub and forest fire (Mcsaveney and Stirling 1992; Mcwethy *et al.* 2009). Reduction of forest that likely acted as a habitat barrier between the alpine and lowland grasshoppers and expansion of grass/herbland may have facilitated mixing. The area now shared by *S. childi* and *S. australis* was further modified by European introduction of plants and grazing animals, and mining practices in the last 150 years. These changes could have led to increased gene flow, but were perhaps no more profound than the effects of Pleistocene climate cycling. Disentangling their respective influence on the grasshoppers is not simple.

The taxonomic status of these species is problematic, as traditional methods cannot resolve the conflicting information from morphology and genetics resulting from the process of evolution. Although *S. childi* is not genetically isolated from *S. australis* they are morphologically well differentiated, and in our relatively small SNP dataset we were able to find one marker possibly under selection. Wu (2001) suggested two criteria for species determination; first that species should not be losing differentiation, and second that they have the ability to continue to diverge. *Siga*s *australis* and *Siga*s *childi* do not appear to be losing morphological distinction, but

our neutral genetic data is consistent with extensive gene flow. The absence of any morphological hybrids suggests selection is intense, removing relatively conspicuous intermediates and holding these two species apart. This may provide them the opportunity to diverge at other loci. For now separate species classification for *S. australis* and *S. childi* is appropriate, although the cost of selection on morphological intermediates in terms of population size will be felt most keenly by *S. childi* in its narrow range of ~100 km². Arguably, sites within this small area that are occupied predominantly by *S. childi* should receive most conservation effort, but study is needed to determine the nature of the optimal breeding habitat for these grasshoppers.

The findings of our study endorse Charles Darwin's original dynamic view of speciation (Darwin 1859), but are contrary to those expected from a more restrictive, but popular, view of species as reproductively isolated units (Mayr 1963). There is a grand irony that while, for many, genetic methods are seen as tools for testing species status, under a traditional expectation (e.g. DNA barcoding) genetic data are actually the key to revealing that speciation is not clear cut (Mallet *et al.* 2009; Peccoud *et al.* 2009; Nosil and Schluter 2011; Cui *et al.* 2013). In our study we found that the only characters that reliably distinguished species were morphological, whilst 78 neutral genetic markers showed that distinct morphotypes do not correspond to genetically isolated units.

ACKNOWLEDGEMENTS

We thank: Craig Wilson (New Zealand Department of Conservation), Simon Morris, Julia Goldberg, Ted Trewick and Bee Trewick for their collection help; Vaughan Symonds for allowing us to use his microsatellite library protocol; Trish McLenachan for all her help and support in the lab; Lara Shepherd for all her microsatellite advice; Fran Wolber for instructions and use of flow cytometry equipment. We also thank Elizabeth Daly, Briar Smith, Simon Hills and Mariana Bulgarella of the phoenix group (evolves.massey.ac.nz) for their help with analysing results and reviewing the manuscript. This work was funded in part by Department of Conservation Taxonomic Units Fund (contract 4076) and Massey University Vice-Chancellor's Scholarship.

REFERENCES

- Abramoff MD, Magalhaes PJ, Ram SJ (2004) Image Processing with ImageJ. *Biophotonics International* **11**, 36-42.
- Bandelt H-J, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16**, 37-48.
- Barton N, Hewitt G (1981) 7 Hybrid zones and speciation. *Evolution and speciation: essays in honor of MJD White*, 109.
- Barton NH, Hewitt GM (1985) Analysis of Hybrid Zones. *Annual Review of Ecology and Systematics* **16**, 113-148.
- Beerli P (2006) Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics* **22**, 341-345.
- Beerli P (2009) *How to use MIGRATE or why are Marko chain Monte Carlo programs difficult to use?* Cambridge University Press, New York.
- Bergsten J (2005) A review of long-branch attraction. *Cladistics* **21**, 163-193.
- Bigelow RS (1967) *The grasshoppers of New Zealand*. University of Canterbury, Christchurch, New Zealand.
- Carlsson J (2008) Effects of Microsatellite Null Alleles on Assignment Testing. *Journal of Heredity* **99**, 616-623.
- Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences. *G3: Genes, Genomes, Genetics* **1**, 171-182.
- Coyne JA, Orr HA (2004) *Speciation*. Sinauer Associates, Sunderland, MA.
- Cui R, Schumer M, Kruesi K, Walter R, Andolfatto P *et al.* (2013) Phylogenomics reveals extensive reticulate evolution in Xiphophorus fishes. *Evolution*, n/a-n/a.
- Darwin C (1859) *On the Origin of Species by Means of Natural Selection*. Murray, London.
- Dowling TE, Secor CL (1997) The role of hybridization and introgression in the diversification of animals. *Annual Review of Ecology and Systematics* **28**, 593-619.

- Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A *et al.*, 2011 Geneious v5.3.4, pp. Available from <http://www.geneious.com>.
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**, 2611-2620.
- Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, 47-50.
- Fischer MC, Foll M, Excoffier L, Heckel G (2011) Enhanced AFLP genome scans detect local adaptation in high-altitude populations of a small rodent (*Microtus arvalis*). *Molecular Ecology* **20**, 1450-1462.
- Foll M, Fischer MC, Heckel G, Excoffier L (2010) Estimating population structure from AFLP amplification intensity. *Molecular Ecology* **19**, 4638-4647.
- Foll M, Gaggiotti O (2008) A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective. *Genetics* **180**, 977-993.
- Forsman A, Karlsson M, Wennersten L, Johansson J, Karpestam E (2011) Rapid evolution of fire melanism in replicated populations of pygmy grasshoppers. *Evolution* **65**, 2530-2540.
- Garnier BJ (1958) *The climate of New Zealand, a geographic survey*. E. Arnold, London.
- Harrison R (1990) Hybrid zones: windows on evolutionary process. *Oxford surveys in evolutionary biology* **7**, 69-128.
- Harrison RG (1993) *Hybrid zones and the evolutionary process*. Oxford University Press, New York.
- Harrison RG (2012) The language of speciation. *Evolution* **66**, 3643-3657.
- Hubbard JCE, Wilson JB (1988) A survey of the lowland vegetation of the Upper Clutha District of Otago, New Zealand. *New Zealand Journal of Botany* **26**, 21-35.
- Jamieson CD (1999) *Distribution and abundance of Sigaus childi Jamieson (Orthoptera: Acrididae), a Central Otago endemic grasshopper*. Science for Conservation 110. Department of Conservation, Wellington, New Zealand.
- Jason DW, Zink RM, Blackwell-Rago RC, Nelson DA (2001) Anomalous Variation in Mitochondrial Genomes of White-Crowned (*Zonotrichia leucophrys*) and Golden-Crowned (*Z. atricapilla*) Sparrows: Pseudogenes, Hybridization, or Incomplete Lineage Sorting? *The Auk* **118**, 231-236.
- Jensen JL, Bohonak AJ, Kelley ST, 2005 Isolation by distance, web service. BMC Genetics 6: 13. v.3.16 <http://ibdws.sdsu.edu/>, pp.
- Johnson M, Zaretskaya I, Raytselis Y, Merezuk Y, McGinnis S *et al.* (2008) NCBI BLAST: a better web interface. *Nucleic Acids Research* **36**, W5-W9.
- Kawakami T, Butlin RK, Adams M, Paull DJ, Cooper SJB (2009) Genetic analysis of a chromosomal hybrid zone in the Australian morabine grasshoppers (*Vandiemenella*, *Viatica* species group). *Evolution* **63**, 139-152.
- Key KHL (1968) The Concept of Stasipatric Speciation. *Systematic Biology* **17**, 14-22.
- Klingenberg CP (2011) MorphoJ: an integrated software package for geometric morphometrics. *Molecular Ecology Resources* **11**, 353-357.
- Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Meth* **9**, 357-359.

- Lexer C, Kremer A, Petit RJ (2006) COMMENT: Shared alleles in sympatric oaks: recurrent gene flow is a more parsimonious explanation than ancestral polymorphism. *Molecular Ecology* **15**, 2007-2012.
- Mallet J (1995) A species definition for the modern synthesis. *Trends in Ecology & Evolution* **10**, 294-299.
- Mallet J (2008) Mayr's view of Darwin: was Darwin wrong about speciation? *Biological Journal of the Linnean Society* **95**, 3-16.
- Mallet J, Meyer A, Nosil P, Feder JL (2009) Space, sympatry and speciation. *Journal of Evolutionary Biology* **22**, 2332-2341.
- Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research* **27**, 209-220.
- Maroja LS, Andrés JA, Walters JR, Harrison RG (2009) Multiple barriers to gene exchange in a field cricket hybrid zone. *Biological Journal of the Linnean Society* **97**, 390-402.
- Mayr E (1963) *Animal species and evolution*. Belknap Press of Harvard University Press, Cambridge.
- Mcglone MS (1985) Plant biogeography and the late Cenozoic history of New Zealand. *New Zealand Journal of Botany* **23**, 723-749.
- Mcglone MS, Duncan RP, Heenan PB (2001) Endemism, species selection and the origin and distribution of the vascular plant flora of New Zealand. *Journal of Biogeography* **28**, 199-216.
- Mcsaveney MJ, Stirling MW (1992) Central Otago: basin and range country. *Landforms of New Zealand*.
- Mcwethy DB, Whitlock C, Wilmshurst JM, Mcglone MS, Li X (2009) Rapid deforestation of South Island, New Zealand, by early Polynesian fires. *The Holocene* **19**, 883-897.
- Meirmans PG, Van Tienderen PH (2004) genotype and genodive: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes* **4**, 792-794.
- Milne I, Bayer M, Cardle L, Shaw P, Stephen G *et al.* (2010) Tablet--next generation sequence assembly visualization. *Bioinformatics* **26**, 401-402.
- Minitab, 2007 Minitab 15 (Statistical Software), pp. Minitab Inc: <http://www.minitab.com>, State College, PA.
- Morgan-Richards M, Smitsen RD, Shepherd LD, Wallis GP, Hayward JJ *et al.* (2009) A review of genetic analyses of hybridisation in New Zealand. *Journal of the Royal Society of New Zealand* **39**, 15-34.
- Morris S (2002) *Identification guide to grasshoppers (Orthoptera: Acrididae) in Central Otago and Mackenzie country*. Department of Conservation, Wellington, New Zealand.
- Morris S (2003a) *Distribution and Taxonomic status of New Zealand endangered grasshoppers (Orthoptera: Acrididae)*. Internal Department of Conservation Report. Department of Conservation, Wellington, New Zealand.
- Morris S (2003b) Two new species of *Sigaus* from Fiordland, New Zealand (Orthoptera: Acrididae). *New Zealand Entomologist* **26**, 65-74.
- Niemiller ML, Fitzpatrick BM, Miller BT (2008) Recent divergence with gene flow in Tennessee cave salamanders (Plethodontidae: Gyrinophilus) inferred from gene genealogies. *Molecular Ecology* **17**, 2258-2275.

- Nosil P, Feder JL (2012) Genomic divergence during speciation: causes and consequences. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**, 332-342.
- Nosil P, Schluter D (2011) The genes underlying the process of speciation. *Trends in Ecology & Evolution* **26**, 160-167.
- Orr MR, Smith TB (1998) Ecology and speciation. *Trends in Ecology & Evolution* **13**, 502-506.
- Palumbi SR (2002) *The evolution explosion: how humans cause rapid evolutionary change*. Norton, New York.
- Papadopoulos AST, Baker WJ, Crayn D, Butlin RK, Kynast RG *et al.* (2011) Speciation with gene flow on Lord Howe Island. *Proceedings of the National Academy of Sciences* **108**, 13188-13193.
- Peccoud J, Ollivier A, Plantegenest M, Simon J-C (2009) A continuum of genetic divergence from sympatric host races to species in the pea aphid complex. *Proceedings of the National Academy of Sciences* **106**, 7495-7500.
- Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. *PLoS ONE* **7**, e37135.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of Population Structure Using Multilocus Genotype Data. *Genetics* **155**, 945-959.
- R Core Team, 2013 R: A language and environment for statistical computing, pp. Retrieved from <http://www.R-project.org>. R foundation for Statistical Computing, Vienna, Austria.
- Rambaut A, 2002 Se-Al: sequence alignment editor. v2.0a11, pp. Available from <http://evolve.200.ox.ac.uk/>. University of Oxford, Oxford, UK.
- Rosenblum EB, Harmon LJ (2011) "Same same but different": replicated ecological speciation at white sands. *Evolution* **65**, 946-960.
- Savolainen V, Anstett M-C, Lexer C, Hutton I, Clarkson JJ *et al.* (2006) Sympatric speciation in palms on an oceanic island. *Nature* **441**, 210-213.
- Servedio MR, Noor MaF (2003) The Role of Reinforcement in Speciation: Theory and Data. *Annual Review of Ecology, Evolution, and Systematics* **34**, 339-364.
- Simon C, Frati F, Beckenbach A, Crespi B, Liu UH *et al.* (1994) *Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers*. Entomological Society of America, Lanham, MD, ETATS-UNIS.
- Soltis P, Soltis D (2009) The Role of Hybridization in Plant Speciation. *Annual Review of Plant Biology* **60**, 561-588.
- Stockwell CA, Hendry AP, Kinnison MT (2003) Contemporary evolution meets conservation biology. *Trends in ecology & evolution (Personal edition)* **18**, 94-101.
- Sunnucks P, Hales DF (1996) Numerous transposed sequences of mitochondrial cytochrome oxidase I-II in aphids of the genus *Sitobion* (Hemiptera: Aphididae). *Molecular Biology and Evolution* **13**, 510-524.
- Susnik S, Weiss S, Odak T, Delling B, Treer T *et al.* (2007) Reticulate evolution: ancient introgression of the Adriatic brown trout mtDNA in softmouth trout *Salmo obtusirostris* (Teleostei: Salmonidae). *Biological Journal of the Linnean Society* **90**, 139-152.
- Thibert-Plante X, Hendry AP (2010) When can ecological speciation be detected with neutral loci? *Molecular Ecology* **19**, 2301-2314.

- Trewick SA (2001) Identity of an endangered grasshopper (Acrididae: Brachaspis): Taxonomy, molecules and conservation. *Conservation Genetics* **2**, 233-243.
- Trewick SA (2008) DNA Barcoding is not enough: mismatch of taxonomy and genealogy in New Zealand grasshoppers (Orthoptera: Acrididae). *Cladistics* **24**, 240-254.
- Trewick SA, Morgan-Richards M (2005) After the deluge: mitochondrial DNA indicates Miocene radiation and Pliocene adaptation of tree and giant weta (Orthoptera: Anostomatidae). *Journal of Biogeography* **32**, 295-309.
- Trewick SA, Morris S (2008) *Diversity and taxonomic status of some New Zealand grasshoppers. DOC Reseach & Development Series #290*. Science & Technical Publishing, Department of Conservation, Wellington, New Zealand.
- Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* **4**, 535-538.
- Van Oppen MJH, Mcdonald BJ, Willis B, Miller DJ (2001) The Evolutionary History of the Coral Genus *Acropora* (Scleractinia, Cnidaria) Based on a Mitochondrial and a Nuclear Marker: Reticulation, Incomplete Lineage Sorting, or Morphological Convergence? *Molecular Biology and Evolution* **18**, 1315-1329.
- Vazquez P, Cooper SJB, Gosálvez J, Hewitt GM (1994) Nuclear DNA introgression across a Pyrenean hybrid zone between parapatric subspecies of the grasshopper *Chorthippus parallelus*. *Heredity* **73**, 436-443.
- Wang RL, Wakeley J, Hey J (1997) Gene Flow and Natural Selection in the Origin of *Drosophila pseudoobscura* and Close Relatives. *Genetics* **147**, 1091-1106.
- Wardle P (2001) Distribution of native forest in the upper Clutha district, Otago, New Zealand. *New Zealand Journal of Botany* **39**, 435-446.
- White TJ, Bruns T, Lee S, Taylor J (1990) *Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics*. Academic Press, San Diego.
- Willett CS, Ford MJ, Harrison RG (1997) Inferences about the origin of a field cricket hybrid zone from a mitochondrial DNA phylogeny. *Heredity* **79**, 484-494.
- Wright S (1982) Character Change, Speciation, and the Higher Taxa. *Evolution* **36**, 427-443.
- Wu C-I (2001) The genic view of the process of speciation. *Journal of Evolutionary Biology* **14**, 851-865.
- Xu S (2000) Phylogenetic Analysis Under Reticulate Evolution. *Molecular Biology and Evolution* **17**, 897-907.
- Zerbino D, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research* **18**, 821-829.

S2. SUPPLEMENTARY MATERIAL FOR CHAPTER THREE

S2.1 MICROSATELLITE DEVELOPMENT METHODS

To obtain neutral nuclear loci to examine population structure, we developed primers to amplify microsatellite loci using a modified enriched microsatellite library protocol (Vaughan Symonds protocol per. comm. 2009). Total genomic DNA was extracted from the hind femur of a *Sigaus childi* grasshopper from Manor Burn near Alexandra, via the CTAB method (see mtDNA methods). The DNA was digested using BfuCI (4 U/ μ L) before being ligated to Sau3AI (BfuCI) linker (25 μ L). Successful ligation was ensured by amplification with Sau L-A oligo (20 μ L). 12 cycles of a standard PCR protocol were used for hybridisation. A linker created using 10X T4 Ligase buffer, Sau L-B oligo (100 μ L) and polynucleotide kinase, was annealed using 10X ligase buffer, NaCl (600 mM) and Sau L-A oligo (100 μ L). Hybridisation used 50-biotin-labelled (CA₁₂ or GA₁₂) oligonucleotides (100 μ M), the linker primer Sau L-B (50 μ L) and HYB-A buffer (12x SCC/0.2% SDS) solution (heated to 65°C to re-suspend). The hybridisation protocol was a gradual step-down cycle from 71°C to 52.5°C stopping every 0.5 °C for 1 min after the PCR product was denatured at 95°C for 3 min. Following hybridisation, stringency washes with biotinylated beads (Promega) were used to enrich the sample. The biotinylated beads were prepared by washing in HYB-B buffer (6x SCC/0.1% SDS). The following washes were performed, with beads re-captured between steps using a magnetic stand: HYB-C buffer (3x SCC/0.1% SDS) at room temperature (twice), HYB-D buffer (1x SCC/0.1% SDS) at 42°C for 15 min followed by HYB-D buffer at 60°C for 15 min, and HYB-D buffer at 52°C for 15 min. After enrichment the DNA was removed from the beads using TE buffer (10 mM tris (HCl) pH 7.5) at 95°C for 5 min. The resulting DNA was transformed into competent *Escherichia coli* cells via a vector using the TOPO cloning kit (Invitrogen).

The *E. coli* cells were plated and cultured at 37°C overnight before 768 recombinant colonies were picked, and suspended in 20 μ L of broth (tryptone, yeast extract and NaCl) and incubated overnight. The inserts were then amplified using M13 primers.

A total of 384 colonies were PCR screened, half had been hybridised with the CA₁₂ probe and half with a GA₁₂. We designed primer pairs to target 96 putative microsatellite loci using the programme Primer3 (Rozen and Skaletsky 2000). The forward primer of each pair was synthesised with an M13 tail to allow incorporation of a florescent label. Primers were tested using DNA from four different New Zealand grasshoppers; two *S. australis*, one *S. childi* and one *Brachaspis robustus* (Acrididae). PCR amplification was performed in 10µl volumes using ABgene Red Hot Taq (Thermo FisherScientific) using the following protocol: 94°C for three min; 12 cycles of 94°C for 45 s, 62°C for 45 s and 72°C for 45 s; 24 cycles of 94°C for 45 s, 58°C for 45 s and 72°C for 45 s; followed by 10 min at 72°C. Subsequently fifty putative microsatellites loci were amplified as above using M13 tagged dyes 6FAM and HEX (Invitrogen and Applied Biosystems). Successful amplification was checked by electrophoresis. Loci that gave good amplification were pooled and genotyped on an ABI 3730 with an internal LIZ -250 size standard and scored using Genemapper v3.7 (Applied Biosystems).

S2.2 MICROSATELLITE PRIMERS

Table 1

Example sequences for the three microsatellite loci, with the size range of alleles and number of alleles found at each in our sample.

Micro	Seq	Size range	Number of alleles
GA31	GACGTTGTAAAACGACGGCCCTATCGGTGGGACGTTTTC CCTATCGGTGGGACGTTTTC	249-284	18
C36	GACGTTGTAAAACGACGGCCGAATGTACCATTGCTCTGTG ACATCACTACCTCGCCAACC	166-213	18
C44	GACGTTGTAAAACGACGGCCGATGGCTGCAGTTCTGTTCA CCAGTGGTGGCAGTGTGTTA	216-248	16

S2.3 TALLY UP TABLE

Table 2

The sites, species and number of individuals used in each of the analysis. Microsatellite (Micro), Cytochrome oxidase subunit I mitochondrial DNA sequencing (mtDNA), traditional diagnostic morphological characters (Trad. Morph.), morphometric data subject to geometric analysis (Geo. Morph), intragenic transcribed spacer DNA sequence (ITS) and RADSeq SNPs (Rad Seq).

		Micro	mtDNA	Trad. morph.	Geo. Morph.	ITS	RAD Seq
Bannockburn	<i>S. australis</i>	3	8	2	6		
Raggedy Range	<i>S. australis</i>	2	2	1	2	1	
Remarkables	<i>S. australis</i>	6	6	20	17	2	
Pisa Range	<i>S. australis</i>	9	6		3		
Little valley	<i>S. australis</i>	4	5	4	3	2	
	<i>S. childi</i>	2	4	5	5	3	
Alexandra Airport	<i>S. australis</i>	4	6	3	5	6	10
Mt St Bathans	<i>S. australis</i>	2	3	4	6		
Obelisk	<i>S. obelisci (S. australis)</i>	4	3	3	21	3	
Danseys Pass	<i>S. australis</i>	3	3	17	1		
Kakanui Ranges	<i>S. australis</i>	2	1	3	2	1	
Rock and Pillars	<i>S. australis</i>	5	2	23	7	1	
Lindis	<i>S. australis</i>	3	4	2	18	2	10
Mount Sutton	<i>S. australis</i>	3	3	8		1	
Harris Saddle	<i>S. australis</i>	2	2	9			
Homer Tunnel	<i>S. homerensis (S. australis)</i>	3	3	3	3		
Sealy Tarns	<i>S. australis</i>	2	2	4	2	2	
Crawford Hills	<i>S. childi</i>	3	5	5	7	3	
	<i>S. australis</i>		2	1		1	
Earnsclough	<i>S. childi</i>	4	9	6	9	3	10
Manor Burn	<i>S. childi</i>	4	8	9	8	4	
Graveyard Gully	<i>S. childi</i>		7	3	7	2	
Conroys Dam	<i>S. australis</i>		2	4	2	1	
Mt Dobson	<i>S. australis</i>		3	7		1	
Rocky Top	<i>S. australis</i>		1	5			
Flagstaff	<i>S. australis</i>		1	7			
Tekapo	<i>S. australis</i>		1	1	7		
Dunstan	<i>S. australis</i>		1	1	1	1	
Robroy	<i>S. australis</i>		1				
Kawarau	<i>S. australis</i>		1	1	1		
Craiegeburn	<i>S. australis</i>			1	1		
Crown Range	<i>S. australis</i>				2		
Foggy Peak	<i>S. australis</i>			3	1		
Mt Scott	<i>S. australis</i>			3			
Total		70	105	168	149	40	30

S2.4 ITS TABLE

Table 3

ITS sequence variation in *Sigaus australis* complex grasshoppers (sites in the sympatric region are highlighted). Only polymorphic sites from comparison of all sequences are shown. Nucleotide variable sites and INDELs are shown as P1, P2 etc and I1, I2 etc. respectively in the header. Those individuals showing nucleotide site ambiguity and sequence asynchrony following INDEL 3 are likely to contain more than one ITS sequence variant. Sequence asynchrony most probably results from the presence of an INDEL in one ITS variant resulting in complete ambiguity downstream, in otherwise “clean” sequences.

Species	N	Area	P1	P2	P4	I2	I3	P5	I7	P6	I8	P7	P8	I9	P9	P10	P11	P12	P13	P14
<i>S. australis</i>	1	Rock and Pillar	A	A	C		3	C		G	9	C	A		C	G	T	A	C	A
<i>S. australis</i>	2	Little Valley/Crawford	A	A	C		3	C	2	G		C	A		C	G	T	A	C	A
<i>S. childi</i>	1	Little Valley	A	A	C		3	C		G	9	C	A		C	G	A	A	C	A
<i>S. australis</i>	4	Sealy Tarns/Kakanui Ranges/Sealy Tarns/Mt Dobson	G	A	C	8		C		G		C	A		C	A	A	C	C	C
<i>S. australis</i>	1	Mt Sutton	G	A	C			C		G		C	A	2	G	A	A	T/C	C	C
<i>S. australis</i>	2	Remarkables	G	A	G			C	2	C		A	G		C/G	A	A	A	C	C
<i>S. obelisci</i>	2	Obelisk	G	G	C			C		G		C	A		C	G	A	C	G	C
<i>S. childi</i>	1	Earnscleugh	G	G	C			C	3	G		C	A		C	G	A	A	G	A
<i>S. australis</i>	4	Airport/Dunstan Range/Lindis/Lindis	G	G	C			C		G		C	A		C	G	A	A	C	A
<i>S. childi</i>	1	Earnscleugh	G	G	C			C		G		C	A		C	G	A	A	G/C	A
<i>S. obelisci</i>	1	Obelisk	G	G	C			C		G		C	A		C	G	A	C	G	C
<i>S. childi, S. australis</i>	2	Airport/Lower Manorburn	G	G	C			G		G		C	A		C	G	A	A	C	A
<i>S. australis</i>	1	Conroy Dam	G	G	C			HYB		G		C	A		C	G	A	A	C	A
<i>S. australis</i>	1	Airport	G	G	C			S		G		C	A		C	G	A	A	G/C	A
<i>S. australis</i>	1	Raggedy Range	G	G	C/G			HYB		G		C	A		C	G	A	A	C	A
<i>S. childi</i>	1	Crawford Rd	G	G/A	C			HYB		G	9	C	A		C	G	A	C/A	G	C/A
<i>S. childi</i>	1	Earnscleugh	G	G/A	C			HYB		G		C	A		C	G	A	A	C/G	A
<i>S. childi</i>	1	ManorBurn	G	G/A	C			HYB		G		C	A		C	G	A	A/C	G/C	SHORT
<i>S. australis</i>	1	Airport	G/A	G/A	C			HYB		G		C	A		C	G	A	A/C	G/C	A
<i>S. childi</i>	1	Crawford Hills	A	A	C		3	HYB	3	G		C	A		C	G	A	A	C/G	A
<i>S. childi, S. australis</i>	3	Graveyard Gully/Graveyard Gully/Airport	G/A	G/A	C		3	HYB		G	9	C	A		C	G	A	A	C	A
<i>S. childi</i>	1	Crawford Rd	A	G/A	C		3	HYB		G		C	A		C	G	A	A	G/C	A
<i>S. childi</i>	1	Little Valley	A	G/A	C			HYB		G	9	C	A		C	G	A	A	C	A
<i>S. childi, S. australis</i>	2	Airport/Manor Burn	A	G/A	C			HYB		G		C	A		C	G	A	A	C	A
<i>S. childi</i>	1	Manor Burn	A	G/A	C			HYB		G		C	A		C	G	A	A	C/G	A/C
<i>S. childi, S. australis</i>	2	Little Valley	A	G/A	C			HYB		G		C	A		C	G	A	A	G/C	A

REFERENCES

Rozen S, Skaletsky HJ (2000) *Primer3 on the WWW for general users and for biologist programmers*. Humana Press, Totowa, NJ.



MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL

**STATEMENT OF CONTRIBUTION
TO DOCTORAL THESIS CONTAINING PUBLICATIONS**

(To appear at the end of each thesis chapter/section/appendix submitted as an article/paper or collected as an appendix at the end of the thesis)

We, the candidate and the candidate's Principal Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of Candidate: Edwina Dowle

Name/Title of Principal Supervisor: Steve Trewick

Name of Published Research Output and full reference:

Dowle EJ, Morgan-Richards M, Trewick SA. (submitted). Gene flow and divergence of an endangered Grasshopper. Currently in review at the Journal of Ecology and Evolution

In which Chapter is the Published Work: Chapter 3

Please indicate either:

- The percentage of the Published Work that was contributed by the candidate:
and / or
- Describe the contribution that the candidate has made to the Published Work:
I undertook all ITS sequencing, new mtDNA haplotype sequencing, RAD-sequencing, morphometric data collection and subsequent analyses. Both Steve and I generated the microsatellite library, and I then conducted the microsatellite sequencing across the samples and ran the analyses. I wrote the initial manuscript draft and Steve, Mary and I subsequently edited this chapter.

Edwina Dowle
Digitally signed by Edwina Dowle
DN: cn=Edwina Dowle, o=Massey, ou,
email=e.j.dowle@massey.ac.nz, c=NZ
Date: 2013.09.23 10:19:43 +12'00'

Candidate's Signature

18/9/13

Date

Steve Trewick
Digitally signed by Steve Trewick
DN: cn=Steve Trewick, o=Massey University,
ou=Ecology Group,
email=s.trewick@massey.ac.nz, c=NZ
Date: 2013.09.27 12:00:32 +12'00'

Principal Supervisor's signature

18/9/13

Date

4. ECOTYPES AND SPECIATION WITH GENE FLOW IN A BIODIVERSITY HOT SPOT

ABSTRACT

Adaptation to the local environment drives the formation of ecotypes, which may then lead to speciation. The giant *Placostylus* snails of New Caledonia occur across a wide range of environmental conditions, from the dry western side to the wetter central and eastern areas. In large slow moving species, such as the *Placostylus*, speciation could be assumed to be largely driven by allopatry and drift as opposed to natural selection. I examine the morphology (geometric morphometrics) and genetics (SNP and mtDNA) of two species of *Placostylus* (*P. fibratus*, *P. porphyrostomus*), to determine the drivers of diversity in this group. Despite the current patchy distribution of snails on New Caledonia both the mtDNA and SNP datasets showed admixture and gene flow between populations and species. On Ile des Pins, gene flow occurs between genetically and morphologically distinct sympatric species. Morphology was concordant with the genetic clusters identified and has a strong relationship with the environments. The genetic dataset, in contrast to the morphological data, did not have a relationship with the environment, suggesting the snails are not limited in their ability to adapt to different environments. Speciation in New Caledonia *Placostylus* is mediated through adaptation to the local environment – adaptation that can be maintained even in the presence of on-going gene flow.

Dowle EJ, Morgan-Richards M, Brescia F, Trewick S

INTRODUCTION

Early work suggested that speciation required isolation and was largely a result of allopatric speciation (Mayr 1963; Coyne and Orr 2004). We are now beginning to better understand the role that natural selection and sympatric speciation can have in species formation, yet the role of allopatric speciation is still emphasized, particularly in poor dispersing, slow moving, organisms and on archipelagos (Coyne and Price 2000; Losos and Ricklefs 2009). Sympatry allows species to maintain gene flow whilst diverging (speciation with gene flow) (Mallet 2001; Via 2001; Wu 2001). Gene flow (hybridisation), once considered a barrier to speciation, is now frequently reported to occur during the process of speciation (Mallet *et al.* 2007; Niemiller *et al.* 2008; Nosil 2008; Peccoud *et al.* 2009; Cui *et al.* 2013). Natural selection can be sufficiently intense to drive speciation even when gene flow occurs between populations/species. During speciation with gene flow selection acts upon loci that maintain ecologically (and evolutionarily) important differences between emerging species, while the permeability of other loci can result in murky boundaries between species (Wu 2001; Nosil and Schluter 2011; Harrison 2012; Nosil and Feder 2012). A starting point for this process might be local adaptation of populations to their environment resulting in ecotypes, which may then lead to ecological speciation (Kawecki and Ebert 2004; Rundle and Nosil 2005; Mallet 2008; Räsänen and Hendry 2008; Nosil and Feder 2012).

New Caledonia, an island at 20° - 24° latitude, is a biodiversity hotspot known for its endemism and micro-endemism of plants and animals (Grandcolas *et al.* 2008). Species diversity is consistently higher in tropical than temperate regions but the processes leading to this pattern have yet to be identified (Latitudinal Biodiversity Gradient (LBG); see chapter 1). Located in the south-west Pacific, New Caledonia's biota is now understood to have originated by long distance dispersal and establishment since its emergence after the Oligocene 37 MYA (Bartish *et al.* 2005; Treweek *et al.* 2007; Grandcolas *et al.* 2008; Espeland and Murienné 2011; Nattier *et al.* 2011). New Caledonia has diverse terrestrial environments; the western region is dominated by dry forest while the eastern and central regions have a much higher rainfall and are dominated by rain forest (Fig. 1). A broad altitudinal range amplifies

local variation in conditions and is implicated in the high levels of biodiversity seen today (Grandcolas *et al.* 2008). Although the formation of the LBG involves the interaction of extinction, speciation and immigration, here I focus on the potential impact that environmental conditions might have in driving speciation in tropical New Caledonia. In landscapes of variable environmental conditions, such as New Caledonia, selection and adaptation of local populations, resulting in the formation of ecotypes, is a likely component of diversification.

The study of molluscs, land snails in particular, has contributed extensively to several aspects of evolutionary biology including the roles of hybridisation and ecological speciation (Woodruff and Gould 1987; Goodfriend and Gould 1996; Davison 2002; Teshima *et al.* 2003; Quesada *et al.* 2007; Greve *et al.* 2012; Stankowski 2013). The genus of large terrestrial gastropods *Placostylus* is present in the western Pacific; New Zealand, Vanuatu, Fiji, Papua New Guinea, Solomon Islands, Lord Howe, and New Caledonia. New Caledonia itself is composed of a main island, Grande Terre, and several offshore islands including Ile des Pins (Isle of Pines) and Ouvéa (part of the Loyalty Islands), which are of particular interest here (Fig. 1). New Caledonian *Placostylus* are important species ecologically as prominent endemic invertebrates, and economically as a food source. There are currently six recognised species, which are characterised by soft tissue morphology (Neubert *et al.* 2009). Shell morphology of many gastropods has been shown to be plastic with respect to the environment and predators (Appleton and Palmer 1988; Trussell 2000; Doyle *et al.* 2010). Within the New Caledonian *Placostylus* shell shape has been disregarded as a taxonomic tool due to concerns over phenotypic plasticity in relation to the environment; soft tissue, in particular the genitalia, are used as the main identification tools (Neubert *et al.* 2009). However, shell characteristics might also reflect a genetic component outside of the influence of environment (Goodfriend 1986; Stankowski 2013) and may be of importance when living snails cannot be sampled.

Previous genetic research using mitochondrial sequence data suggested that additional subdivision of some of these taxa might be appropriate and inferred gene flow among some of the currently described species (Trewick *et al.* 2008; Neubert *et*

al. 2009). Three described species are included in the present study: *Placotylus caledonicus*, *P. fibratus* and *P. porphyrostomus*, with the focus being on the latter two, which are the most abundant, widespread and only harvested species. *Placostylus porphyrostomus* is a smaller dry-forest specialist that on Grande Terre is largely restricted to the western sclerophyl forest, whereas *P. fibratus* is larger and occurs on the wetter central and eastern regions of Grande Terre (Fig. 1) (Brescia 2011). *Placostylus fibratus* and *P. porphyrostomus* are for the most part allopatric, except on Ile des Pins where they are sympatric (Fig. 1). On Ile des Pins the two species are easily distinguished by their size difference, but elsewhere size and shape variation occurs within both species. Sharing of mtDNA haplotypes between *Placostylus porphyrostomus* and *P. fibratus* on Ile des Pins could represent rare events of mitochondrial capture or more extensive on-going gene flow (Trewick *et al.* 2008). Using *Placostylus* I tested ideas surrounding speciation, morphological variation and gene flow on New Caledonia.

Geometric morphometric techniques, not previously applied to these snails, were used to analyse shell shape and size variation in the *Placostylus* snails. Genetic diversity of central and southern New Caledonian *Placostylus* was examined using an extensive SNP dataset derived from a RAD-sequence and mtDNA haplotype data (Peterson *et al.* 2012). Next Generation Sequencing now enables large datasets to be gathered quickly from previously understudied organisms, and is fast becoming a common tool in population genetics from both model and non-model organisms (Senn *et al.* 2013; Wagner *et al.* 2013). Genetic admixing between populations was examined using the SNP dataset, with particular attention paid to the sympatric species on Ile des Pins. I then examined the drivers of diversity within the New Caledonian *Placostylus*. Using both the genetic and morphological datasets along with environmental data, I determined whether the morphological variation within *Placostylus* is explained best by the genetic structuring or environmental gradients found on Grande Terre and Ouvéa.

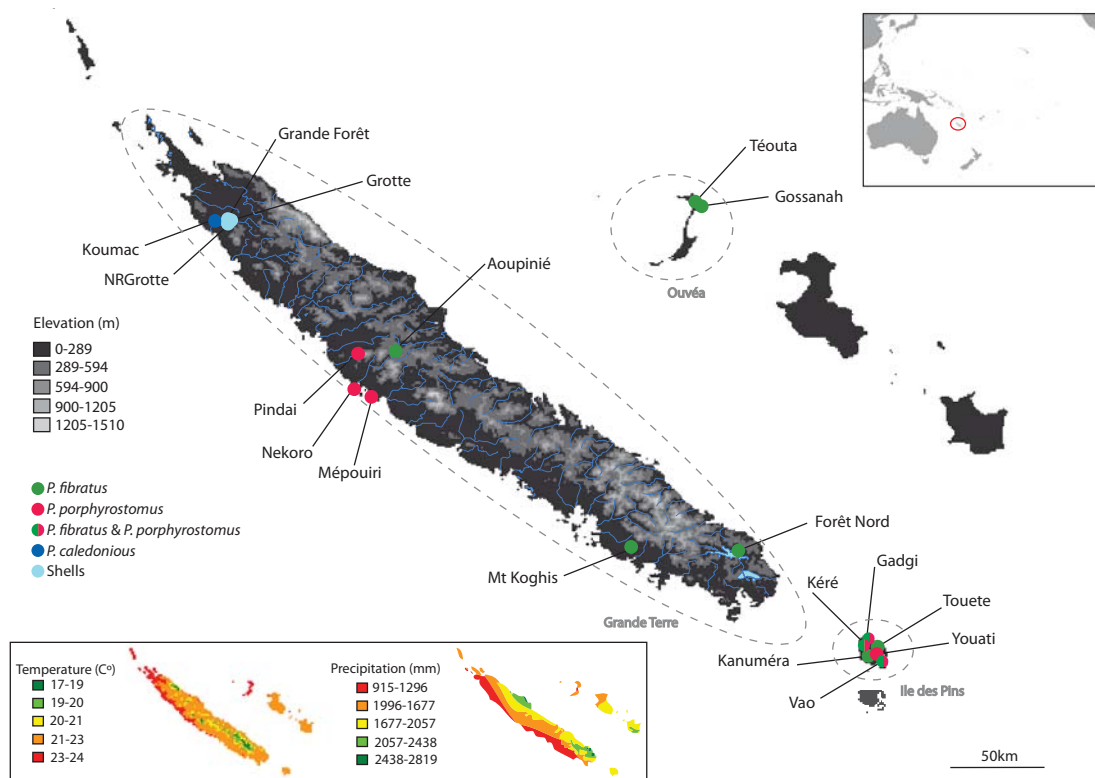


Figure 1

Sample locations of *Placostylus* snails in New Caledonia. Samples were obtained from three islands; Grande Terre, Ouvéa and Ile des Pins. At some sites no live snails were found and are solely represented by shells. Based on shell shape and location these sites likely represent populations of *P. caledonicus*. Two environmental gradients, generated from BIOCLIM datasets, across New Caledonia are shown on the bottom left: annual average temperature and annual precipitation. The eastern side of Grande Terre contains remnants of dry forest, while the western and central areas are largely composed of tropical rain forest. *Placostylus porphyrostomus* (red) is considered a dry forest specialist, while *P. fibratus* (green) is considered a wet forest specialist.

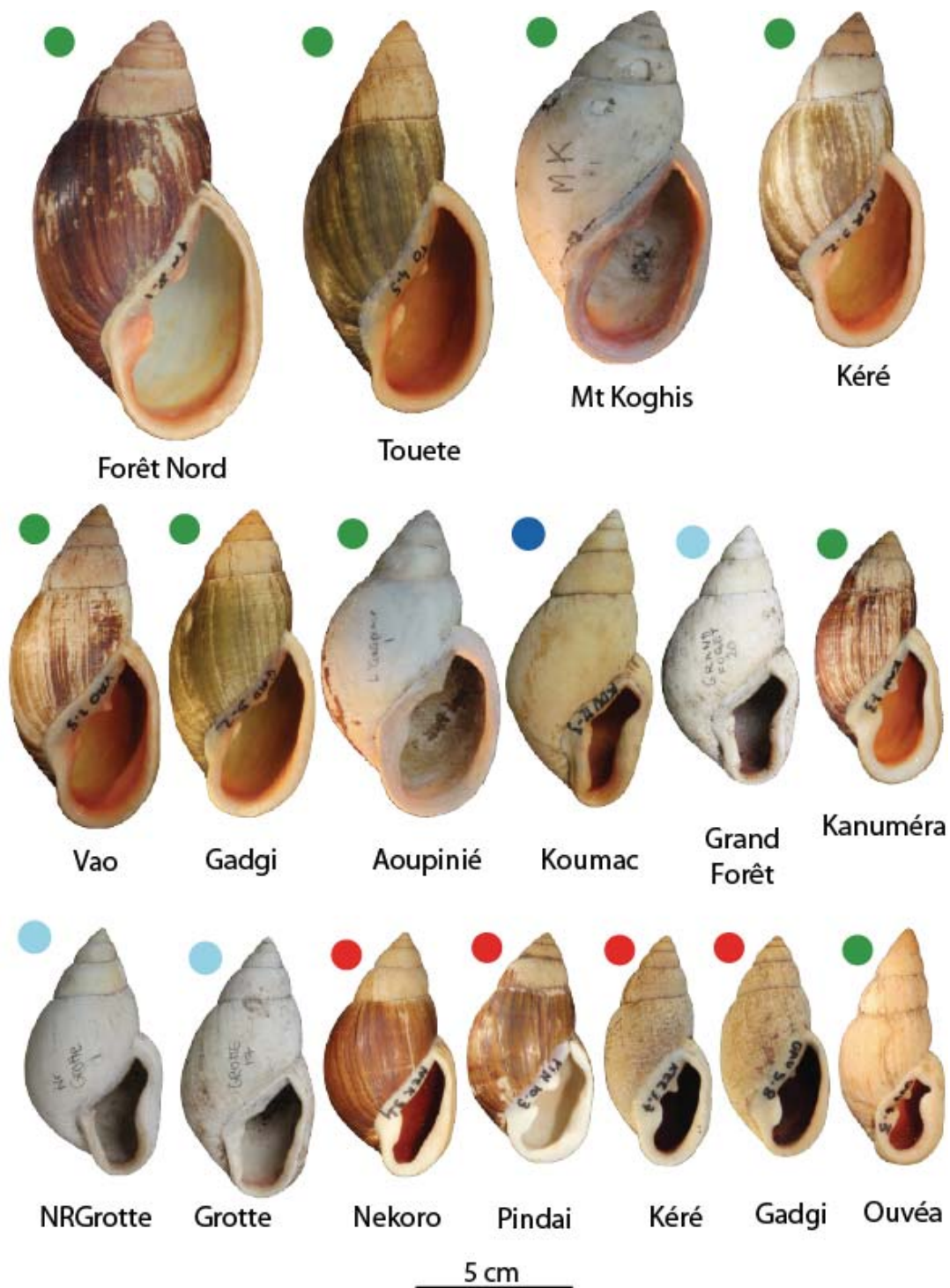


Figure 2

Representative shell forms from sampled populations (scaled). Shells are coded according to current taxonomy; *P. porphyrostomus* (red dots), *P. fibratus* (green dots) and *P. caledonicus* (blue dots, light blue indicates locations only represented by shells). *Placostylus fibratus* is considered a wet forest specialist and is characterised by a bulging body whorl and open aperture, while *P. porphyrostomus* is considered a dry forest specialist characterised by a narrow aperture, smaller body whorl and smaller size in general. Samples from Ouvéa are a dwarf form of *P. fibratus* and have a shape and size not consistent with other *P. fibratus* from Grande Terre and Ile des Pins.

METHODS

Sampling strategy

I focused on the two most common and widespread species of *Placostylus* that are distributed through central and southern New Caledonia. Sampling was carried out to encompass a range of geographic, taxonomic and environmental contrasts (Fig. 1).

Placostylus fibratus was collected from three locations on Grande Terre (Forêt Nord, Mt Koghis and Aoupinié) where it is allopatric with respect to *P. porphyrostomus*. The locations differ in altitude, temperature and precipitation. Population samples of *Placostylus fibratus* from two islands were also collected; these represent a dwarf form on one of the Loyalty Islands (Ouvéa), and several locations on Ile des Pins where the snails are sympatric with *P. porphyrostomus*.

Similarly, *Placostylus porphyrostomus* was collected from three locations on Grande Terre where the snails are allopatric but experience a range of conditions. These were two hot, dry coastal sites (Nekoro and Mépouiri), and a cooler, wetter site further inland (Pindai). Populations were also sampled from several locations on Ile des Pins where this species is sympatric with *Placostylus fibratus*. For genetic work *Placostylus caledonicus* was sampled from a single location, but three adjacent sites provided shell data to encompass a broader range of morphological information.

Whole snails were collected by hand between 2005 and 2012 as part of a research programme directed at understanding the population structure and demography of the snails. These specimens contributed material for DNA research, shell morphology and captive breeding (Brescia 2011). Non-lethal sampling was undertaken in April 2012 and involved the collection of foot biopsies. Natural mortality in the populations where we took biopsies provided representative shells. At three northern populations (Grande Forêt, Grotte and NRGrotte) live snails were not located despite extensive searching. These populations did not therefore contribute genetic data.

Geometric analysis

To test for concordance between the genetic data and shell morphology I used a set of landmarks around the aperture and shell outline when placed in the ventral view (Fig. 2, Fig. 3). Studies on marine molluscs have shown the benefit of the technique in capturing shape variation between species (Hills *et al.* 2012). Six permanent landmarks and 22 semi-landmarks were used in the analysis.

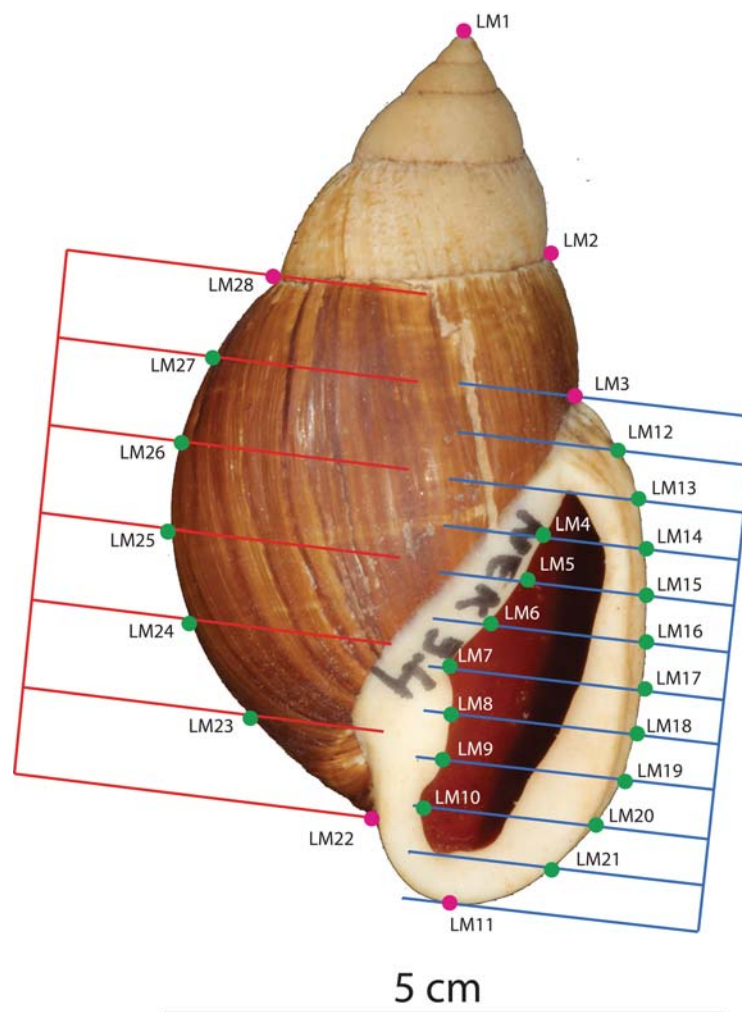


Figure 3

The permanent landmarks (pink) and semi-landmarks (green) digitised in the geometric dataset for the New Caledonian *Placostylus*. Landmarks were optimised to capture shape variation around the aperture and body whorl, which is considered to vary between current taxonomic species.

A digital image of the ventral side of each shell was taken and ADOBE PHOTOSHOP CS6 was used to place combs in the image. Both combs were aligned to a centreline

running from the top of the shell to the bottom left intercept of the aperture and body whorl. Digitizing was undertaken in TPSDIG2 2.17 (Rohlf 2013) on a Waccon digitising tablet. After digitization semi-landmarks were slid in SEMILAND, part of the IMP714 package (Sheets 2012; Zelditch *et al.* 2012), using the Procrustes distance method. Landmark X/Y co-ordinates were then imported into MORPHOJ 1.05f (Klingenberg 2011) where non-parametric tests were undertaken, as parametric tests cannot be undertaken with semi-landmarks.

To examine the extent of handling error associated with photographing and digitizing the images, one shell (from Nekoro) was photographed 5 times to assess error associated with placement for photography, and one of these images was then digitized 5 times, re-placing the combs each time to measure digitising error. A Procrustes ANOVA was used to examine the variation in these repeats compared to the variation found within a population. The sum of squares was found to be 38.944297 for centroid size and 0.00035643 for shape between photographs, and 0.066505 for centroid size and 0.00007192 for shape between digitizing replicates. Within populations (based on genetic groups) the sum of squares was found to be 80963.498961 for centroid size and 0.70609876 for shape. The error associated with photographing and digitizing the images was therefore considered insignificant in comparison to the potentially biologically meaningful variation.

Within MORPHOJ shape and size variation were examined. Shape was assessed using Principal Component Analysis (PCA) across all individuals and all landmarks. Discriminant analysis used cross validation and 1000 permutations based on the current morphological groupings (*P. fibratus*, *P. porphyrostomus*, and *P. caledonicus*) with the assumption that the shells collected from Grande forêt, Grotte and NRGrotte represented recently deceased populations of *P. caledonicus*. A second discriminant analysis used the groupings identified from the genetic analysis (SNP and mtDNA) with samples lacking genetic data kept separate; *P. fibratus* (Ile des Pins, Forêt Nord and Mt Koghis), *P. caledonicus*, *P. porphyrostomus* (Ile des Pins), *P. por-nekoro* (*P. porphyrostomus* from Nekoro), *P. por-pindai* (*P. porphyrostomus* from Pindai), *P. fib-aoupinié* (*P. fibratus* from Aoupinié), Grande forêt, Grotte and NRGrotte. Canonical Variate Analysis (CVA) of these nine groups was then used to

describe differences between them. I tested the influence of removing semi-landmark points from the analysis using a discriminant analysis with cross validation and 1000 permutations. Seven different combinations of semi-landmarks were tried: removing semi-landmarks, 4-10, 12-21, 23-27, 4-10 and 12-21; 4-10 and 23-27; 12-21 and 23-27; and finally removing all semi-landmarks and just using the 5 permanent landmarks (Fig. 3). Size was analysed by a regression of PC1 scores (all landmarks) against centroid size (size variation) in MORPHOJ. Procrustes coordinates and centroid size were then averaged across all populations and a separate PCA on this dataset applied. The PC1 scores and centroid sizes from populations occurring on Grande Terre and Ouvéa (joining populations from Gossanah and Téouta) was then used in the modelling.

Mitochondrial DNA sequence

In order to improve genealogical resolution in relation to taxonomy and geography, I sequenced one mtDNA gene. To achieve this, a genomic sample of a *Placostylus porphyrostomus* from Mépouiri was sequenced on an Illumina Hi-seq 2000 (BGI) with an expected scale of 1GB of data. DNA was extracted using incubation at 55°C with Proteinase K and a CTAB buffer (2% Hexadecyltrimethylammonium bromide, 100 mmol/L Tris-HCl pH8.0, 1.4 mol/L NaCl, 20 mmol/L EDTA), followed by a combined phenol/chloroform/isoamyl alcohol (25:24:1) cleanup, two rounds of clean up were used in some samples to minimise polysaccharide content. Extractions were eluted in water and diluted as necessary for PCR reactions. The resulting Hi-seq data was de novo assembled using VELVET (Zerbino and Birney 2008) with mapping in BOWTIE2 (Langmead and Salzberg 2012) and visualisation in TABLET 1.12.09.03 (Milne *et al.* 2010). A large contig (>3000 bp) provided information from which to design primers surrounding the ND2 gene (PRIMERS Placo_ND2F: AAC GCA AAG GGT ATG AAC CCG TAA ATA G, Placo_ND2R: GAG CAA TCG CCG GAG GAA CGG AAA T). This gene was sequenced across all available New Caledonian *Placostylus* samples. PCR used the protocol: 94°C for three minutes; 94°C for 45 seconds, 56°C for 45 seconds and 72°C for 75 seconds repeated 36 times; followed by a 2 minute anneal. Cycle sequencing used Perkin Elmer Bigdye 3.1 chemistry following the manufacturer's protocols and analysed on an ABI Prism 377 DNA sequencer (Applied Biosystems, Inc., Foster City, California). Sequences were

checked and aligned in GENEIOUS 6.0.3 (Geneious). A bootstrapped ML tree was obtained using RAXML via the CIPRES portal (Stamatakis 2006; Miller *et al.* 2010) with a GTR Γ model, allowing RAXML to halt the bootstrapping when statistically appropriate.

RAD-seq SNP

Anonymous single nucleotide polymorphic (SNP) nuclear markers were generated using high throughput sequencing, with individual DNA fragments coded so I could identify individual snails to assign genotypes. The double digest RAD-Seq protocol was used as described by Peterson *et al.* (2012) with minor modifications. I used the restriction endonucleases PstI and BamHI to digest whole genomic DNA from 149 individual snails. The DNA fragments were tagged with short DNA sequences that identified each individual before pooling. Data were generated on an Illumina Hi-Seq (NZGL) and sorted using the STACKS 1.01 pipeline (Catchen *et al.* 2011; Catchen *et al.* 2013).

In STACKS a range of parameter settings relating to read coverage, individual number and population coverage were implemented. Read coverage settings vary in the literature (Peterson *et al.* 2012; Buerkle and Gompert 2013), and I therefore experimented with coverage cut-offs of 7-30 reads with the data and found this had little influence on results. For the results described here I selected 8 reads per individual as providing a reliable set of markers for downstream analysis and excluded all stacks with a lower coverage. I allowed a maximum of two mismatches between alleles for a single individual and four mismatches between primary and secondary reads within USTACKS and allowed the program to remove any potentially spurious highly repetitive stacks. In CSTACKS 3 mismatches were allowed between samples when generating the SNP set (-m 15 -N2 -M4 -n3 -t). Analysis was restricted to a single SNP per putative locus (always the first) avoiding potential problems of data non-independence. SNPS were required to be present in $\geq 90\%$ of individuals from a population to be recorded. I also varied individual number. Individuals with very poor coverage were initially removed and this data set was analysed with two population coverage cut-offs, two and four. A second data set that removed more individuals with low coverage was run with the same two

population coverage cut-offs, two and four. This gave a total of four SNP datasets. All SNPs that were only present in one individual as a heterozygote were removed (single instance SNPs).

Data file conversion for other programs was performed using PGDSPIDER 2.0.4.0 (Fischer et al. 2011). BAYESCAN 2.01 was used to examine the individual markers for evidence of selection using the default settings; prior odds of a neutral model are 10 times more likely than the model with selection at a locus (Foll and Gaggiotti 2008; Foll et al. 2010; Fischer et al. 2011). The alpha value was used to determine the direction of selection with a positive value suggesting diversifying selection and a negative value suggesting balancing selection. Results were viewed in R (R Core Team 2013). Any non-neutral markers were then removed. As only one individual was present in Gossanah for the F_{ST} calculation it was pooled with the other sample from Ouvéa, Téouta. The samples (Fig. 1) F_{ST} (Weir and Cockerham) values were calculated in GENODIVE 2.0b24 (Meirmans and Van Tienderen 2004) using an AMOVA with 1000 permutations. STRUCTURE 2.3.4 (Pritchard *et al.* 2000) was used to estimate population differentiation. Initially all four SNP sets were examined with 10 replications of an admixture model with independent allele frequencies using a burn-in of 30, 000 followed by 100, 000 generations, number of groups (K) was set from 1 to 16. Once concordance across the runs was confirmed. A longer run on the largest dataset was implemented using an admixture model with independent allele frequencies, a burn-in of 100,000 followed by 200,000 generations (10 replications); group number (K) was set from 1 to 12. A dataset comprising all individuals was run along with two other subsets. The subsets consisted of the described species *P. porphyrostomus* (Ile des Pins, Nekoro, Mèpouiri and Pindai) and *P. fibratus* (Ile des Pins, Ouvéa, Forêt Nord, Mt Koghis and Aoupinié), however, due to the mtDNA results and close proximity of populations, Pindai and Aoupinié were included in both the '*P. porphyrostomus*' and '*P. fibratus*' datasets. The optimum value of K was determined by eye. STRUCTURE was run using one SNP per putative-locus, and each locus appeared in ≥ 4 or ≥ 2 populations, and occurred in $\geq 90\%$ of the individuals. STRUCTURE HARVESTER (Earl and Vonholdt 2012) was used to examine results and format files for CLUMPP 1.1.2 (Jakobsson and Rosenberg 2007), which was used

to average the results across the optimum K values. Graphs were then redrawn in DISTRUCT 1.1 (Rosenberg 2004).

SNAPP 1.1.1 (Bryant *et al.* 2012) was used to infer a phylogenetic tree from the SNP datasets. SNAPP requires species groups; here the results from the STRUCTURE and F_{ST} analyses were used to determine genetic groups for SNAPP; *P. porphyrostomus* (Ile des Pins), *P. fibratus* (Ile des Pins, Mt Koghis, Forêt Nord and Ouvéa), *P. por-nekoro* (Nekoro and Mépouiri), *P. por-pindai* (Pindai) and *P. fib-aoupinié* (Aoupinié). STACKS was re-run with the five genetic groupings and the same settings as above, but SNP-loci had to be present in all five populations. Because these requirements limited the number of loci available I included in the final SNP-loci set all loci that occurred across the five populations, regardless of their coverage within a population. Single instance SNP-loci were then removed from the final set and SNAPP was run with a gamma rate prior with a chain length of 10000000, storing trees every 1,000 runs. Results were viewed in TRACER 1.5 (Rambaut and Drummond 2007) to confirm coalescence. DENSITREE 2.0.1 (Bouckaert 2010), TREESETANALYSER (part of the SNAPP package), TREEANNOTATOR (Drummond *et al.* 2012) and FIGTREE (Rambaut 2009) were used to view trees and analyse topology.

To examine the genotypes in the area of sympatry STACKS and BAYESCAN were re-run with the same settings using all the individuals from Ile des Pins from both species (*P. fibratus* and *P. porphyrostomus*). For this, SNP-loci were included if they occurred in $\geq 50\%$ of the snails from each species, and single instance SNP-loci were removed for the BAYESCAN and STRUCTURE analyses. BAYESCAN 2.01 was used to examine the individual markers for evidence of selection using the default settings; prior odds of a neutral model are 10 times more likely than the model with selection at a locus. STRUCTURE was run on the data for each of the two species, 16 *P. fibratus* and 32 *P. porphyrostomus*, using all neutral loci (one SNP per locus) that occurred in both taxa (*P. porphyrostomus* or *P. fibratus* on Ile des Pins). The results were averaged and redrawn using CLUMPP 1.1.2 and DISTRUCT 1.1. Gene flow between the two species in sympatry was estimated using MIGRATE-n 3.5.1 (Beerli 2006; Beerli 2009). The sample size of *P. porphyrostomus* was then randomly trimmed to

16 individuals and MIGRATE-n was run using the Bayesian inference strategy across all neutral SNP-loci. The starting values for θ and M were generated from F_{ST} estimates in the first run and ensuing runs used the resulting θ and M from the prior run. In total, four runs were made. The uniform prior distributions were used for both θ and M parameters with metropolis sampling; one long chain was run recording 6,000 steps every 100 steps after a burn-in of 50,000. The final run used a static heating scheme with four chains.

Environmental Modelling

Environmental data were gathered using DIVA-GIS (Hijmans *et al.* 2013) to access the WORLDCLIM 1.4 (Hijmans *et al.* 2005) database. R 3.0.0 (R Core Team 2013) was used to determine the geographic distance between sites via the function `earth.dist` in the FOSSIL package (Vavrek 2011). In R I modelled the relationship between genetic distance (obtained from the largest SNP-loci set with Téouta and Gossanah grouped), geographic distance, morphometric distance (PC1 scores across populations), and environmental distance for populations on Grande Terre and Ouvéa, for which I had complete datasets (Pindai, Aoupinié, Nekoro, Ouvéa, Mt Koghis and Forêt Nord). Samples from Ile des Pins were removed as the environmental data is not on a fine enough scale on Ile des Pins. Due to the non-independence of the genetic distance and geographic distance data, a mixed linear model was implemented in R via the LME4 package (Zuur *et al.* 2009; Bates 2010). I examined the genetic, environment and morphological data for the most suitable model and, based on the ANOVA scores, a mixed linear model with random intercepts was preferred for the environmental and morphological variables over a mixed linear model with random intercepts and slopes.

I then examined the genetic data for any correlation with environment distance. The 19 available BIOCLIM variables were: annual mean temperature, mean diurnal range, isothermality, temperature seasonality, max temperature of warmest month, min temp of coldest month, temperature annual range, mean temperature of wettest quarter, mean temperature of driest quarter, mean temperature of coldest quarter, annual precipitation, precipitation of wettest month, precipitation of driest month, precipitation seasonality, precipitation of wettest quarter precipitation of driest

quarter, precipitation of warmest quarter and precipitation of coldest quarter. These were each examined using ANOVA and those variables with the three best AIC scores were included in the final model. Genetic distance was then compared to morphometric distance (PC1) and the three environmental variables chosen from the ANOVA. The significance of the model was examined using the package *LANGUAGER* and function *pvals.fnc* (Baayen 2013). The genetic relationship with geographic distance showed a preference for a model with variable slopes and intercepts and was run separately to the other variables with this model. As this model cannot be used with *pvals.fnc*, I reported the t-values and confirmed the result using a mantel test with 10,000 permutations using the *ADE4* package in R (Chessel *et al.* 2013).

Sets of standard linear models were run in R to examine the relationship between; morphometric distance (PC1 scores), snail shell size (centroid size) and environmental data, as these datasets are not pairwise. Initially a regression analysis was undertaken with each environmental variable and PC1 scores to determine which were most correlated. A general linear model was then used to determine whether the significant environmental variables were correlated with one another. A general linear model was run using the single most correlated environmental variable (given that all environmental variables were highly correlated with each other) with centroid size against PC1. Finally a mixed linear model with random intercepts and random slopes was run using geographic distance and morphometric distance to evaluate whether there was a relationship between geographic distance and morphology.

RESULTS

Geometric analysis

Geometric data were obtained for 152 snails. These represented 16 population samples from the islands of Grande Terre, Ile des Pins and Ouvéa (Fig. 2). The size of population samples ranged from one (*P. fibratus* from Gossanah) to 39 (*P. porphyrostomus* from nekoro). The PCA across all individuals and all landmarks showed that 75% of observed variance was explained by the first three principal components. Individuals were found to group with others of the same species based on the current descriptions (determined by genitalia, shell shape and location) and could be further subdivided in accordance with the genetic groupings (see below; Fig. 4). Discriminant analysis supported the inference that individual snails could be placed into their current species definitions based on shell shape (Table 1). Discriminant analyses and CVA also supported that individuals could be placed into their genetic groupings (Table 2, Fig. 4). Only the placement of populations that lacked genetic data (putatively *P. caledonicus* based on location and shell shape) was not well resolved. This concordance of morphometrics and genitalia was robust with some but not all semi-landmarks removed (Table 2). A principal component analysis of the sympatric populations on Ile des Pins, *P. fibratus* and *P. porphyrostomus*, showed that the phenotypes of the two species form non-overlapping clusters (Fig. 5).

Table 1

Discriminant analysis of geometric shell shape data based on current species taxonomy, including all semi and permanent landmarks. All samples from the three described species, *P. fibratus* (*P. fib*), *P. porphyrostomus* (*P. por*) and *P. caledonicus* (*P. cal*), are allocated to the correct group.

Test	<i>P. fib</i> -- <i>P. cal</i>	<i>P. fib</i> -- <i>P. cal</i>	<i>P. fib</i> -- <i>P. por</i>	<i>P. fib</i> -- <i>P. por</i>	<i>P. cal</i> -- <i>P. por</i>	<i>P. cal</i> -- <i>P. por</i>
True Group→	<i>P. fib</i>	<i>P. cal</i>	<i>P. fib</i>	<i>P. por</i>	<i>P. cal</i>	<i>P. por</i>
AllocatedGroup↓						
<i>P. fib</i>	38	0	38	-	-	-
<i>P. por</i>	-	-	0	54	0	54
<i>P. cal</i>	0	60	-	-	60	0

Table 2

Results from the discriminat analysis of the geometric data using the genetic groupings; *P. por* (*P. porpyrostomus* from Ile des Pins), *P. fib* (*P. fibratus* from Ile des Pins, Ouvéa, Forêt Nord and Mt Koghis), *P. cal* (*P. caledonicus* from Koumac), *P. 'nek'* (*P. por-nekoro*), *P. 'pin'* (*P. por-pindai*), *P. 'aou'* (*P. fib-aoupinié*), and the three locations from which no genetic data were collected and therefore have unknown genetic groupings: *P. 'GR'* (Grotte), *P. 'NG'* (NRGrotte) and *P. 'GF'* (Grande Forêt). The significance levels between groups are based on procrustes distances, numbers in bold indicate where no significance difference between groups was found. The header shows difference subsections of Landmark points taken out; for example, ExLM23-27 indicates landmarks 23-27 are excluded from the analysis (Fig. 3).

Comparison	AllLM		ExLM 4-10		ExLM 12-21		ExLM 23-27		ExLM4-10 &12-21		ExLM4-10 &23-27		ExLM12-21 &23-27		ExAllSemi	
	p-value	T-square	p-value	T-square	p-value	T-square	p-value	T-square	p-value	T-square	p-value	T-square	p-value	T-square	p-value	T-square
<i>P. cal</i> - <i>P. fib</i>	<0.0001	<0.0001	<0.0001	0.001	0.001	<0.0001	<0.0001	<0.0001	0.009	<0.0001	0.001	<0.0001	0.002	<0.0001	0.012	<0.0001
<i>P. cal</i> - <i>P. 'GF'</i>	0.24	<0.0001	0.227	<0.0001	0.227	<0.0001	0.458	0.316	0.212	0.001	0.873	0.006	0.42	0.016	0.693	0.58
<i>P. cal</i> - <i>P. 'GR'</i>	0.022	0.009	0.237	0.01	0.065	0.479	0.015	<0.0001	0.331	0.004	0.167	<0.0001	0.056	0.116	0.312	0.856
<i>P. cal</i> - <i>P. 'NG'</i>	0.123	0.14	0.4	0.721	0.221	0.095	0.07	0.047	0.553	0.022	0.734	0.508	0.075	0.151	0.763	0.785
<i>P. cal</i> - <i>P. 'nek'</i>	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
<i>P. cal</i> - <i>P. 'pin'</i>	<0.0001	0.002	0.005	0.105	0.009	0.009	0.004	0.005	0.037	0.221	0.014	0.079	0.003	0.002	0.098	0.183
<i>P. cal</i> - <i>P. por</i>	0.011	0.001	0.024	0.009	0.004	0.002	0.021	0.018	0.045	0.031	0.075	0.015	0.031	0.161	0.195	0.394
<i>P. cal</i> - <i>P. 'aou'</i>	0.012	0.004	0.027	0.015	0.029	0.005	0.025	0.013	0.01	0.013	0.027	0.023	0.018	0.016	0.027	0.038
<i>P. fib</i> - <i>P. 'GF'</i>	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
<i>P. fib</i> - <i>P. 'GR'</i>	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
<i>P. fib</i> - <i>P. 'NG'</i>	<0.0001	<0.0001	<0.0001	0.001	<0.0001	<0.0001	<0.0001	0.005	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
<i>P. fib</i> - <i>P. 'nek'</i>	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
<i>P. fib</i> - <i>P. 'pin'</i>	<0.0001	<0.0001	0.017	0.004	<0.0001	<0.0001	<0.0001	<0.0001	0.02	<0.0001	0.074	<0.0001	0.001	<0.0001	0.088	0.016
<i>P. fib</i> - <i>P. por</i>	0.0010	<0.0001	0.005	0.004	0.009	<0.0001	0.001	<0.0001	0.01	<0.0001	0.003	<0.0001	0.005	<0.0001	0.008	0.003
<i>P. fib</i> - <i>P. 'aou'</i>	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.004	0.001	<0.0001	<0.0001	0.006	<0.0001	0.053	<0.0001	0.069	0.11
<i>P. 'GF'</i> - <i>P. 'GR'</i>	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
<i>P. 'GF'</i> - <i>P. 'NG'</i>	0.012	<0.0001	0.052	<0.0001	0.048	0.188	0.009	<0.0001	0.141	<0.0001	0.038	<0.0001	0.052	<0.0001	0.094	0.069
<i>P. 'GF'</i> - <i>P. 'nek'</i>	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
<i>P. 'GF'</i> - <i>P. 'pin'</i>	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.039	<0.0001	<0.0001	0.066	<0.0001	<0.0001	0.009	0.001

P. 'GF'--P. <i>por</i>	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.001	0.001	<0.0001	<0.0001	0.001	<0.0001
P. 'GF'--P. 'aou'	<0.0001	0.001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
P. 'GR'--P. 'NG'	0.031	0.0018	0.041	<0.0001	0.008	<0.0001	0.069	<0.0001	0.016	<0.0001	0.051	0.007	0.038	<0.0001	0.05	0.008
P. 'GR'--P. 'nek'	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
P. 'GR'--P. 'pin'	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.002	0.016	<0.0001	<0.0001	<0.0001	<0.0001	0.001	0.033
P. 'GR'--P. <i>por</i>	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.004	0.001	<0.0001	<0.0001	<0.0001	0.002	<0.0001	<0.0001	<0.0001	0.001	0.002
P. 'GR'--P. 'aou'	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.002	0.003	0.001	<0.0001	<0.0001	0.034	<0.0001	<0.0001
P. 'NG'--P. 'nek'	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
P. 'NG'--P. 'pin'	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.014	0.017	<0.0001	0.001	<0.0001	<0.0001	0.005	0.211
P. 'NG'--P. <i>por</i>	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.116	0.002	<0.0001	<0.0001	<0.0001	0.004	0.076
P. 'NG'--P. 'aou'	0.004	<0.0001	<0.0001	0.001	0.003	0.001	0.003	<0.0001	<0.0001	0.001	0.002	<0.0001	<0.0001	0.003	0.002	0.058
P. 'nek'--P. 'pin'	<0.0001	<0.0001	0.001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.012	<0.0001	0.001	<0.0001	<0.0001	<0.0001	0.033	0.019
P. 'nek'--P. <i>por</i>	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.001
P. 'nek'--P. 'aou'	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
P. 'pin'--P. <i>por</i>	0.002	0.001	0.003	0.002	<0.0001	<0.0001	<0.0001	0.002	0.006	0.003	0.095	0.027	0.001	0.012	0.224	0.663
P. 'pin'--P. 'aou'	0.017	0.021	0.008	0.019	0.012	0.049	0.021	0.001	0.003	0.007	0.019	0.008	0.015	<0.0001	0.016	0.045
P. <i>por</i> --P. 'aou'	0.006	0.002	0.006	0.002	0.008	0.006	0.007	0.005	0.006	0.006	0.001	<0.0001	0.001	0.001	0.001	0.245

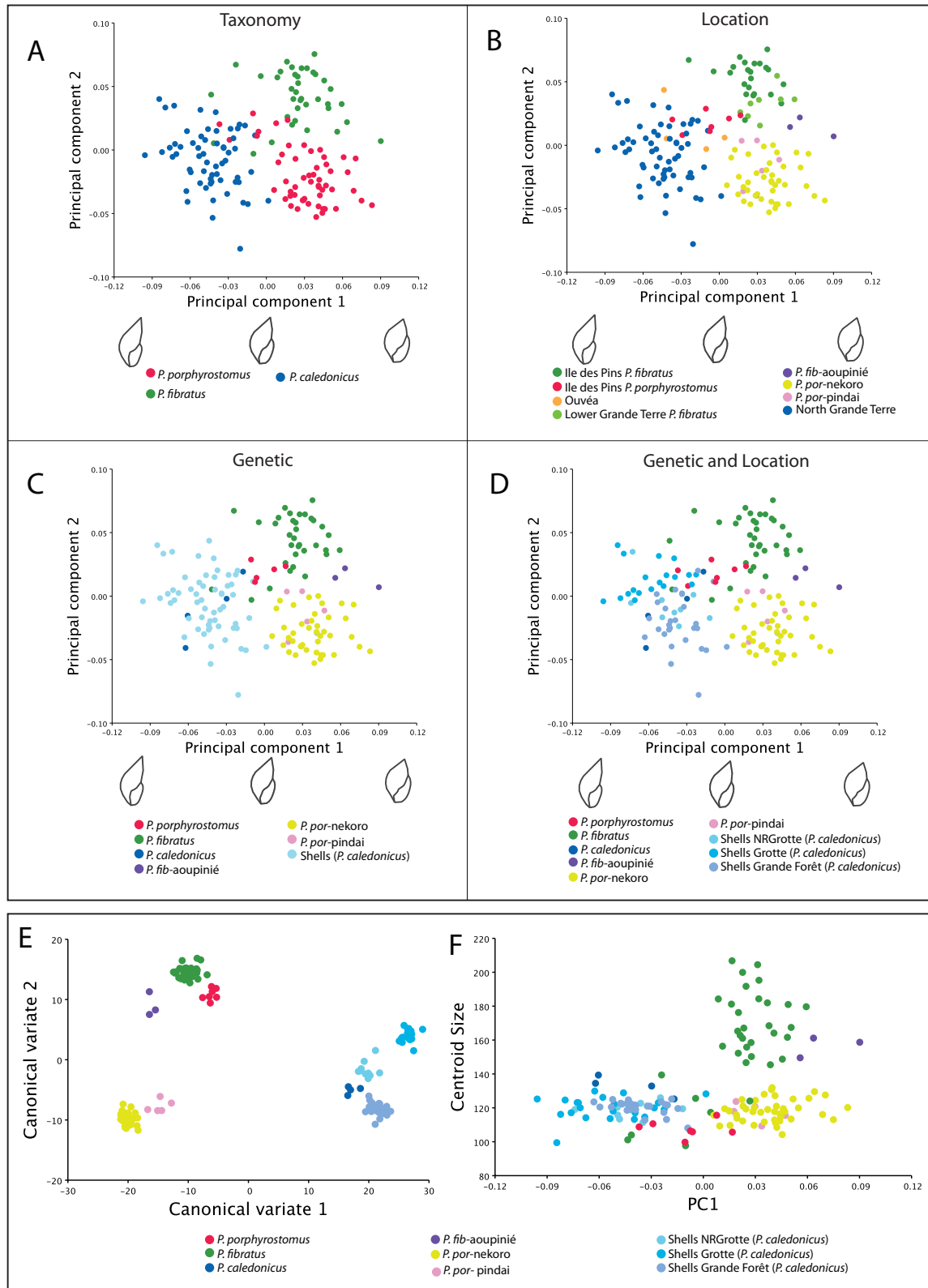


Figure 4

Placostylus shell shape variation across New Caledonian *Placostylus*, from analyses that include all permanent and semi-landmarks. (A) PCA of shell shape coloured based on current taxonomy, shape changes associated with PC1 are illustrated under the x-axis. The three species, *P. fibratus*, *P. porphyrostomus* and *P. caledonicus*, form separate clusters but some overlap in shape occurs between all three. (B) PCA of shell shape coloured based on

sample locations. Samples from the dwarf form of *P. fibratus* show a different shape to *P. fibratus* from Grande Terre and Ile des Pins. (C) PCA of shell shape coloured by the groupings identified in the genetic data. A very similar shape is found in samples from Nekoro and Pindai. (D) PCA of shell shape with the same genetic groupings as C but with the shells from sites with no genetic data separated by sample location. Sites with no genetic data form distinct but highly overlapping clusters with one another. (E) CVA analysis of the geometric data using genetic and location groupings (as in D). Genetic groupings separated well but group clusters do not conform to current taxonomy. For instance *P. porphyrostomus* (Ile des Pins) did not group with *P. por-nekoro* but rather with *P. fibratus* (Ile des Pins, Forêt Nord and Mt Koghis) and *P. fib-aoupinié*. (F) Linear regression of the relationship between shell shape and size coloured by genetic groups and location. Size is extremely variable within genetic groups; samples of *P. fibratus* represent both the largest and smallest snail due to the dwarf form on Ouvéa.

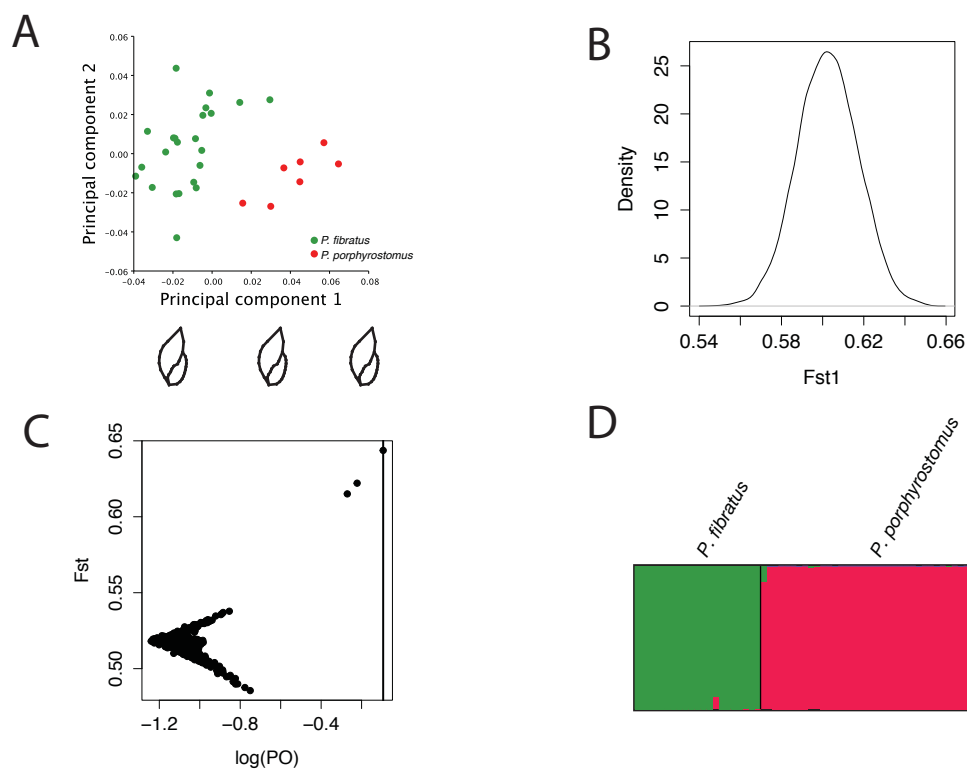


Figure 5

Analysis of *Placostylus* in sympatry on Ile des Pins. (A) PCA of shell shape shows no overlap between *P. fibratus* and *P. porphyrostomus* on the Ile des Pins; shape changes associated with PC1 are indicated along the x-axis. (B) F_{ST} frequency distribution across SNP-loci from *P. fibratus* and *P. porphyrostomus*. (C) BAYESCAN plot of 661 markers, with a single marker showing evidence of selection. The vertical line is the 5% threshold of false discovery. (D) STRUCTURE graph of SNP-loci from Ile des Pins. Clear population distinction occurs between the two sympatric species, *P. porphyrostomus* and *P. fibratus*. The two individuals with the highest amounts of admixture were from Gadgi and Kanuméra, the same two populations with mtDNA sharing in the haplotype tree. The *P. fibratus* from Kanuméra with admixing in STRUCTURE grouped in the largely *P. porphyrostomus* mtDNA clade while the *P. porphyrostomus* from Gadgi with admixing in STRUCTURE grouped in the *P. porphyrostomus* mtDNA clade.

Mitochondrial DNA sequence

In total 100 individuals representing all populations sampled for genetics were sequenced for 830 bp of the mitochondrial gene ND2 (Fig. 6). The resulting phylogenetic tree (rooted with *P. caledonicus*) supports previous work recognising five clades within *P. fibratus* and *P. porphyrostomus*, whilst detecting two instances of introgression of haplotypes between the two species on Ile des Pins (Fig. 6) (Trewick *et al.* 2008). All instances of introgression from *P. fibratus* into the predominantly *P. porphyrostomus* clade came from snails collected at Kanuméra, whereas the *P. porphyrostomus* snails with haplotypes in the *P. fibratus* clade came from the population sample collected at Gadgi (Fig. 1). The three other haplotype clades within my sampling of *P. fibratus* and *P. porphyrostomus* were snails from Aoupinié (*P. fibratus*), and *P. porphyrostomus* from Nekoro and Mèpouiri, and from Pindai (Fig. 6).

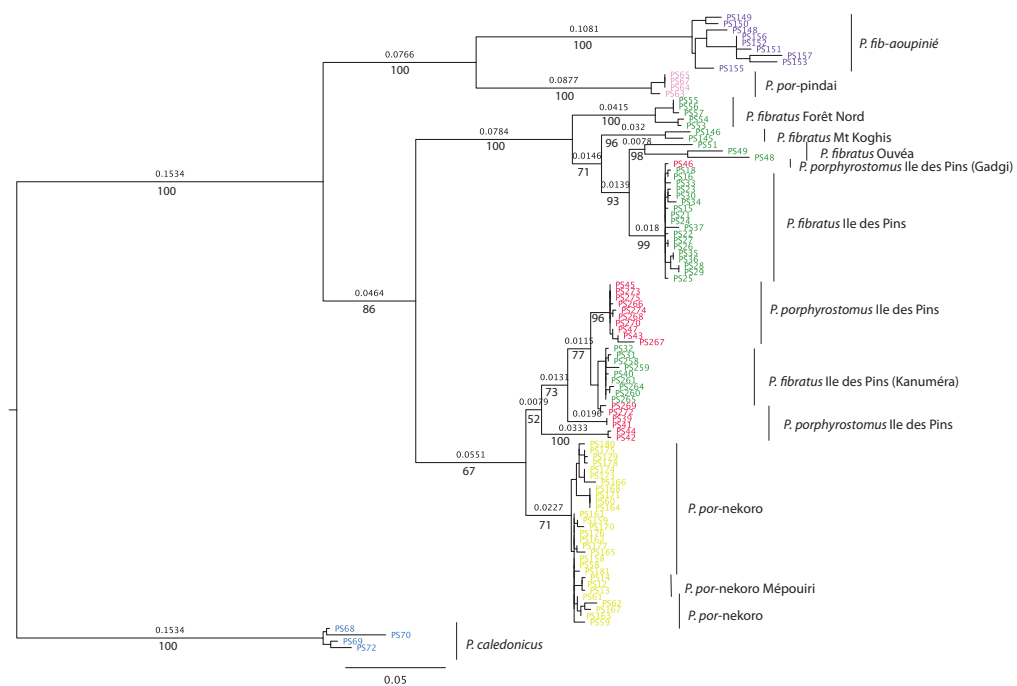


Figure 6

Maximum Likelihood tree of ND2 for New Caledonian *Placostylus* snails. Colours represent the genetic groupings identified in the mtDNA and SNP datasets. Haplotype sharing can be seen between *P. fibratus* and *P. porphyrostomus* on the Ile des Pins.

RAD-seq SNP

The available analytical software for the paired-end Illumina sequencing protocol was not compatible with the sequence tags on our DNA fragments. Base composition bias in the first four base pairs (SNP site) in the second read was outside of the limits set by the software and thus I had to exclude this portion of the data from further analysis. I therefore present results from the first read only. After removing the second read and running `process_radtags` in `STACKS`, >40 million reads were retained with individuals averaging >3,000,000 reads. A total of 150,532 loci were found in the dataset. In total 149 New Caledonian *Placostylus* were sequenced, but due to fragmented DNA, not all had adequate coverage. The individuals with poorest coverage were removed along with all *P. caledonicus* samples (40 in total). Removing the poorer quality samples increased the number of loci retained per population, when using a population proportional cut-off. Two SNP-loci sets were generated using 109 individuals and another two SNP datasets were generated from 97 individuals, after a further 12 snails were removed (Table 3, S3.1, S3.2, S3.3 and S3.4). Given the consistency of results obtained from these datasets (S3.5 and S5.6), I report those from the larger number of individuals Sp2.

Table 3

Summary of the total numbers of individuals and SNPs generated in the four main datasets used in analyses. The column 'Raw SNP' indicates the output from `STACKS` and 'SNP neutral' indicates the total neutral SNPs available. 'Total SNP no singles' indicates the final SNP set once single instance SNPs were removed.

SNP set	Samples	Populations	Raw SNP	SNP neutral	Total SNP no singles
Sp2	97	2	5872	5807	3764
Sp4	97	4	2839	2820	1885
Up2	109	2	3892	3889	2518
Up4	109	4	1621	1616	1092

BAYESCAN results showed evidence of both diversifying and balancing selection among the SNP-loci (Fig. 7). The results from `STRUCTURE` showed five genetically distinct groups (Fig. 8), consolidating the mtDNA results. Two genetic clusters exist within the form recognised as *P. fibratus*: one cluster comprising all population samples (n = 8) except Aoupinié (*P. fib-aoupinié*), which is genetically distinct. Three

genetic clusters were evident within the form recognised as *P. porphyrostomus*; one comprised the population samples from Nekoro and M  pouiri (referred to as *P. por-nekoro*), a second comprised the populations from Ile des Pins, and the third was the single population sample from Pindai (referred to as *P. por-Pindai*). Population samples showed evidence of admixture, especially *P. fibratus* from Grande Terre (Mt Koghis and For  t Nord) with the *P. por-pindai* and *P. fib-aoupini  * populations. The F_{ST} results from GENODIVE largely corroborated those from STRUCTURE (Table 4).

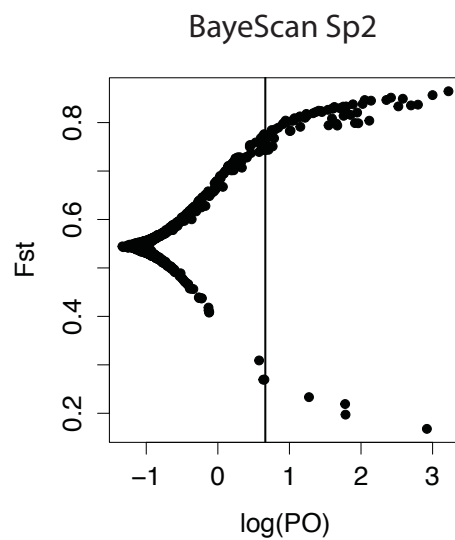


Figure 7
BAYESCAN plot of the Sp2 SNP dataset. In total 65 SNP-loci were identified as being under selection with 59 undergoing divergent selection. The vertical line is the 5% threshold of false discovery.

Table 4

GENODIVE F_{ST} results for the Sp2 SNP dataset. Here Ouvéa contains the data from samples collected at both Téouta and Gossanah. Populations from the Ile des Pins are identified as IP. The remaining populations are from Grande Terre except for the Ouvéa population. Current species taxonomy is indicated in brackets, *P. fibratus* (f) and *P. porphyrostomus* (p). Results are largely concordant with those of STRUCTURE.

	IP:Touet e (f)	IP:Gadgi (f)	Forêt (f)	Nord	IP:Kéré (f)	IP:Kanuméra (f)	Nekoro (p)	Mt (f)	Koghis	Aoupinié (f)	Pindai (p)	IP:Gadgi (p)	IP:Kéré (p)	Mépouiri (p)	IP:Youati (p)	IP:Vao (p)	Ouvéa (f)
IP:Touete(f)																	
IP:Gadgi(f)	0.093																
Forêt Nord(f)	0.4	0.411															
IP:Kéré(f)	0.08	0.075	0.424														
IP:Kanuméra (f)	0.182	0.263	0.588		0.208												
Nekoro(p)	0.651	0.67	0.6		0.688	0.697											
Mt Koghis(f)	0.458	0.5	0.385		0.487	0.533	0.612										
Aoupinié(f)	0.584	0.67	0.603		0.636	0.735	0.622	0.582									
Pindai(p)	0.643	0.704	0.666		0.692	0.875	0.657	0.596	0.49								
IP:Gadgi(p)	0.678	0.7	0.68		0.689	0.369	0.385	0.634	0.684	0.74							
IP:Kéré(p)	0.545	0.603	0.63		0.596	0.607	0.476	0.561	0.631	0.689	0.058						
Mépouiri(p)	0.604	0.652	0.633		0.612	0.654	0.231	0.561	0.575	0.585	0.494	0.48					
IP:Youati(p)	0.657	0.709	0.672		0.694	0.628	0.448	0.6	0.72	0.685	0.005	0.083	0.348				
IP:Vao(p)	0.551	0.585	0.601		0.576	0.55	0.513	0.522	0.68	0.646	0.031	0.119	0.44	0.058			
Ouvéa(f)	0.509	0.484	0.563		0.539	0.202	0.614	0.514	0.614	0.715	0.689	0.602	0.474	0.672	0.598		

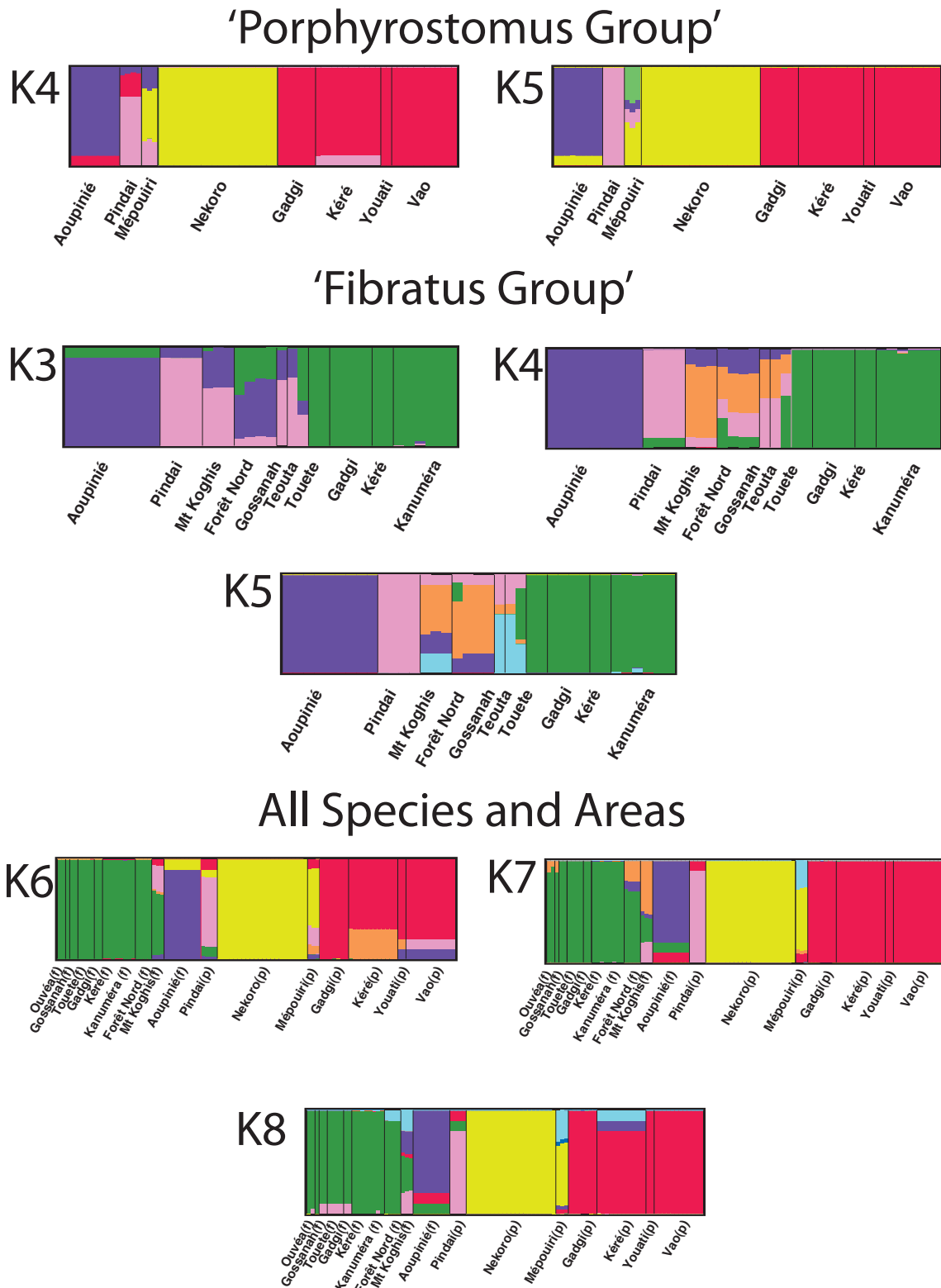


Figure 8

STRUCTURE results from the Sp2 SNP dataset. The 'Porphyrostomus group' is composed of 8 populations, four from Ile des Pins and four from the Grande Terre (Aoupinié was included due to the uncertainty about its taxonomic placement). Strong population structuring occurs between four groups, *P. porphyrostomus* from Ile des Pins, *P. por-nekoro* (samples from both Nekoro and Mèpouiri), *P. por-pindai* and *P. fib-aoupinié*. The population

from M  pouiri has some evidence of admixture with *P. por-pindai* and *P. fib-aoupini  *. The ‘Fibratus group’ is composed of 10 populations, four from Ile des Pins, two from Ouv  a and four from Grande Terre (including Pindai due to taxonomic uncertainty). Strong structuring exists between *P. fib-aoupini  *, *P. por-pindai*, and *P. fibratus* from Ile des Pins. On Grande Terre admixing from *P. fib-aoupini  * and *P. por-pindai* into the populations from Mt Koghis and For  t Nord has occurred. The results from all species and areas show similar patterns to that seen in the sub-groups. Strong structuring exists between five groups: *P. porphyrostomus* (Ile des Pins), *P. fibratus* (Ile des pins, Ouv  a, Mt Koghis and For  t Nord), *P. por-nekoro*, *P. por-pindai* and *P. fib-aoupini  *. Admixing was again identified from *P. fib-aoupini  * and *P. por-nekoro* into *P. fibratus* from Grande Terre (Mt Koghis and For  t Nord).

Five genetic groups were used in SNAPP, as SNAPP requires species groupings: *P. fibratus* (from Ile des Pins, Ouv  a, For  t Nord and Mt Koghis), *P. porphyrostomus* (from Ile des Pins), *P. por-nekoro* (from Nekoro and M  pouiri), *P. fib-aoupini  * (from Aoupini  ) and *P. por-pindai* (from Pindai) (Fig. 9). The dataset for SNAPP included 631 SNP-loci that occurred across all five populations. The results showed that the topology of the most supported tree represented 75% of trees produced (S3.7). There was considerable uncertainty regarding the root position within the tree, but other parameters coalesced. The θ values ($\theta = 4N\mu$ where N is effective population size and μ is mutation rate) show similar population size estimates for *P. fibratus*, *P. porphyrostomus* and *P. por-nekoro*, but smaller population sizes estimates were returned for *P. fib-aoupini  * and *P. por-pindai*. The genotypic cluster *P. por-nekoro* is distinct from the rest of *P. porphyrostomus*, while *P. fibratus* is the sister-species of both *P. fib-aoupini  * and *P. por-pindai*. This contrasts with mtDNA sequence data, from which I infer a closer relationship between *P. por-nekoro* and *P. porphyrostomus* haplotypes (Fig. 10).

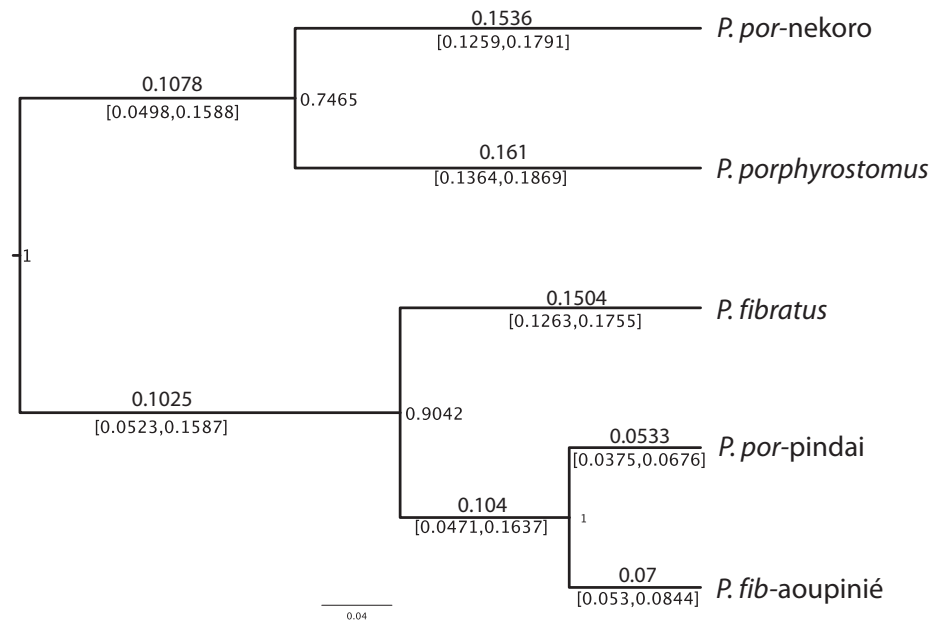


Figure 9

Results from the SNAPP analysis across five genetic groups: *P. porphyrostomus* (Ile des Pins), *P. fibratus* (Ile des Pins, Forêt Nord, Mt Koghis and Ouvéa), *P. por-nekoro* (Nekoro and Mépouiri), *P. por-pindai* (Pindai) and *P. fib-aoupinié* (Aoupinié). This topology was returned for 75% of the trees recorded. θ values are displayed above branches (95% CI below) and posterior values at the nodes. The topology is similar to that in the ML analysis of mtDNA data (Fig. 6), although *P. por-pindai* and *P. fib-aoupinié* are now sister to *P. fibratus*, while *P. porphyrostomus* and *P. por-nekoro* are more diverged.

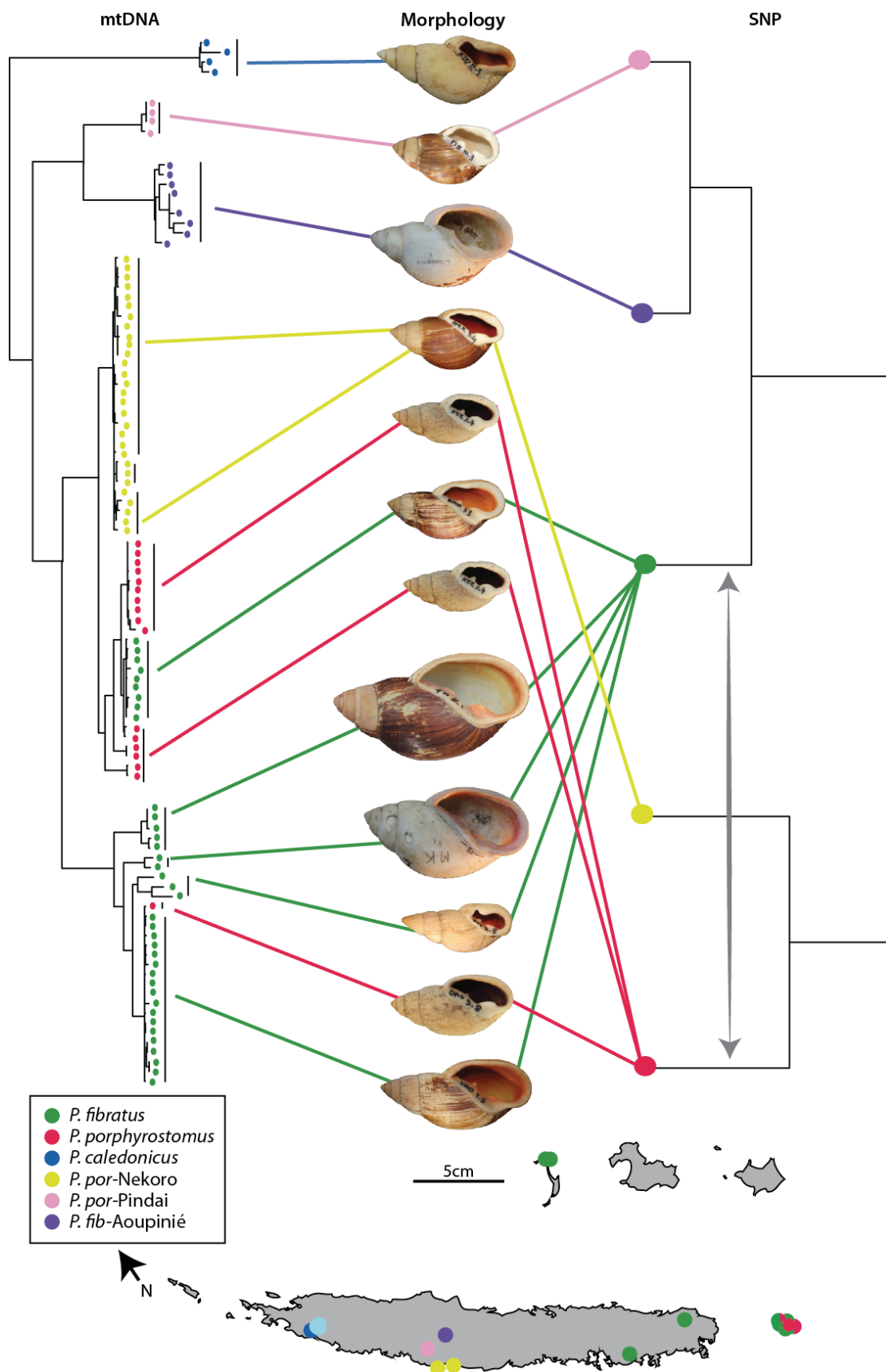


Figure 10

The diversity of shell forms found across the New Caledonian *Placostylus* and their position in the mtDNA tree (left) and SNP tree (right). Rapid morphological change has occurred

between and within species (particularly in *P. fibratus*) in the New Caledonian *Placostylus*. While others, such as *P. por-pindai* (pink) and *P. por-nekoro* (yellow), show convergence to a similar shell shape. The grey arrow indicates low levels of gene flow in the SNP dataset between *P. porphyrostomus* and *P. fibratus* on Ile des Pins.

Within the area of sympatry on Ile des Pins, 843 SNP-loci were extracted that met the required criteria. Within BAYESCAN and STRUCTURE, single instance SNPs were removed leaving a data set of 661 SNP-loci for these analyses. BAYESCAN identified one SNP-locus as being non-neutral and STRUCTURE results showed clear population distinction between the two morpho-species (Fig. 5). MIGRATE estimated gene flow between the two species with the total number of migrants per generation >0.1 (Table 5). Gene flow was recorded at similar levels in both directions.

Table 5

MIGRATE results showing $\theta = 4N_e\mu$ in the SNPs, where N_e is population size and μ is mutation rate $\theta_1 M_{2 \rightarrow 1} = 4N_m \mu$. The θ values are not comparable to the SNAPP analysis as different data subsets were used. There is clear evidence for low levels of gene flow between *P. fibratus* and *P. porphyrostomus*.

All Loci	Mean	0.025	0.975	Nm
θ_1 <i>P. fibratus</i>	0.02089	0.01200	0.02967	0.113
θ_2 <i>P. porphyrostomus</i>	0.02131	0.01233	0.02767	0.115
$M_{2 \rightarrow 1}$	21.7	4.7	38.0	
$M_{1 \rightarrow 2}$	21.7	4.7	38.0	

Environmental Modelling

When modelling genetic distance against environmental variables the BIOCLIM variables with the strongest correlation with the genetic data were temperature seasonality, temperature annual range, and precipitation seasonality (AIC -28.86, -28.18 and -25.787 respectively). The final model with variable intercepts that compared genetic distance (F_{ST}) to morphometric distance (PC1) and environmental distance showed that none of these four environmental variables had a significant relationship with genetic distance (p-values >0.1). The model with variable slopes and intercepts comparing F_{ST} to geographic distance showed little support for a correlation between the two (t-value: 0.906). This lack of a relationship between

genetic distance and geographic distance was confirmed by the Mantel test (p-value: 0.2139). Thus the SNP variation within the sampled snails did not provide evidence for isolation by distance on Grande Terre and Ouvéa.

Linear models of morphometric geometric variation showed that two environmental variables were significantly correlated with morphometric distance; annual mean temperature (t-value: 2.827, p-value: 0.0475) and mean temperature of coldest quarter (t-value: 3.76, p-value: 0.0198) (Fig. 11). A Linear Model of these environmental variables with each other found they were highly correlated (t-value: 12.175 p-value: <0.001). No relationship was found between shell size and morphometric distance (t-value: -1.818, p-value: 0.143). There was also no relationship between morphology and geographic distance (p-value: 0.943).

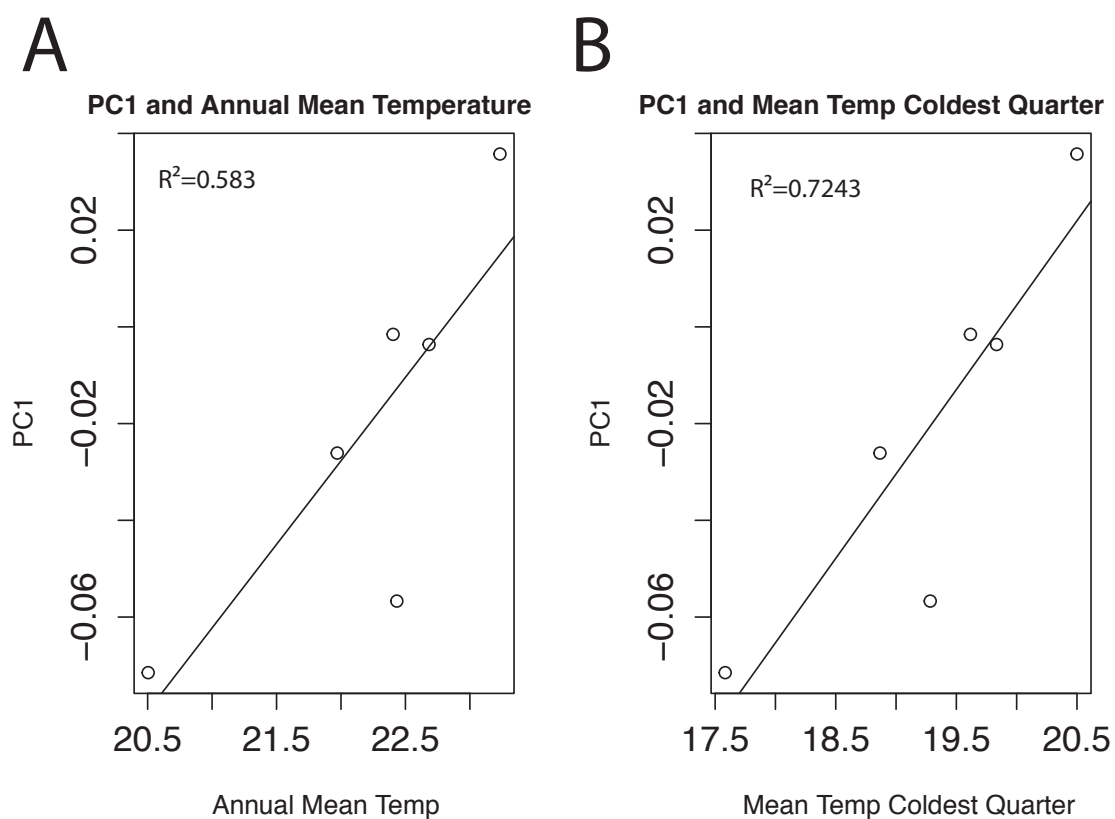


Figure 11

The relationship between environmental conditions and morphology across the *Placostylus* samples from Grande Terre and Ouvéa. (A) The relationship between shell shape (PC1) of populations on Grande Terre and Ouvéa and mean annual temperature. (B) The relationship between shell shape (PC1) of populations on Grande Terre and Ouvéa and mean temperature of the coldest month.

DISCUSSION

Within the New Caledonian *Placostylus* I found evidence for ecological adaptation, gene flow between species and divergent selection. There was evidence for six groups within the genetic and morphological data, contrasting with the three expected from current taxonomy. Within the geometric data I found good support for the three currently described species sampled and concordance with the six genetic groups identified. There is also a strong relationship between morphology and local temperatures, suggesting selection by the local environment drives changes in shell shape across New Caledonia.

Ecological speciation begins with the formation of ecotypes, which arise due to local adaptation to local environments; a process not necessarily requiring a long time frame (Rundle and Nosil 2005; Hendry *et al.* 2007; Räsänen and Hendry 2008). In *Placostylus* there was no evidence that population genetic structuring is related to variation in local environmental conditions, but there was strong evidence of a link between morphology and local environment. Land snails are sensitive to environmental conditions and desiccation is a major environmental pressure in *Placostylus* as in other land snails. This is likely reflected in the shell shape changes seen across the environmental gradients of New Caledonia. Shell shape changes indicated on Grande Terre slimmer apertures with compacted shell shapes, are associated with western dry forest (Pindai and Nekoro) and wider apertures with elongated shell shapes are found in the wetter central and eastern areas (Aoupinié, Mt Koghis and Forêt Nord). Within the New Caledonian *Placostylus*, the strong relationship between morphology and environment could represent plastic responses of morphology to local environments or fixed morphological differences between locally adapted populations or ecotypes.

Were morphology a solely plastic response to local environmental then we would expect little genetic divergence between geographically close populations, given the evidence for genetic admixing in the *Placostylus*. However, there is clear population structuring even between geographically and genetically neighbouring populations, such as *P. fib-aoupinié* and *P. por-pindai* that exist in different environments to

which their morphology is presumably adapted. Morphology is therefore unlikely to be purely plastic within the *Placostylus* and different environmental pressures are likely driving the formation of morphologically distinct snails. The result appears to be the formation of multiple ecotypes across New Caledonia and convergence of shell shape between genetically distinct groups that occur in similar local environments. In one instance, convergence of shell shape between two populations, Pindai and Nekoro, has resulted in them being classified as the same species, *P. porphyrostomus*. The populations occur in similar local environments, with matching morphological adaptation, in close proximity to one another. Genetically, however, *P. por-pindai* and *P. por-nekoro* form two distinct genetic clusters and are associated with two separate lineages, *P. porphyrostomus* and *P. fibratus*. Strong selection on morphological traits by the local environment has also driven the divergence of a dwarf form of *P. fibratus* on Ouvéa. Ouvéa samples are genetically very similar to other *P. fibratus* from Ile des Pins and Grande Terre but shell size is highly reduced (Fig 2). Rapid morphological adaptation, similar to that observed here in *Placostylus*, has previously been documented in other land snails (e.g. Chiba 1996). *Placostylus* snails do not appear to be limited in their ability to adapt to their environment and *P. fibratus* and *P. porphyrostomus* do not represent dry and wet forest snails.

Gene flow (hybridization) between species is an important process of evolution that is relatively common and does not necessarily constrain further divergence (Mallet 2008; Nosil 2008; Nosil and Schluter 2011; Abbott *et al.* 2013). On Ile des Pins, a strong genetic signal for separation between the two described species, *P. fibratus* and *P. porphyrostomus*, was present in the mtDNA, STRUCTURE and SNP phylogenetic analysis. The SNP phylogenetic analysis revealed this species pair to be amongst the most dissimilar of the groups sampled. Geometric analysis of Ile des Pins snails also revealed distinct phenotypic clusters concordant with current taxonomy and genotypic clusters. Ecological differences between the two species have yet to be formally described, but fine scale population structuring has been observed on Ile des Pins. *Placostylus porphyrostomus* is thought to occupy dryer coastal areas, while *P. fibratus* is predominantly found in denser inland forest (Brescia 2011). Despite the structuring found there was evidence for low levels of genetic exchange between *P. fibratus* and *P. porphyrostomus* on Ile des Pins. Gene

flow can have multiple consequences; it can allow the passing of adaptive traits, reinforce reproductive barriers, cause hybrid speciation, limit adaptation or result in the loss a parental species (Morgan-Richards *et al.* 2009; Abbott *et al.* 2013). How these potential processes will impact the diversity on Ile des Pins is unclear. But the strong evidence of genetic and morphological separation suggests that the two species are being maintained in the presence of low levels of on-going gene flow.

Analysis of the SNP data with STRUCTURE suggested that populations of *P. fibratus* from Grande Terre (Mt Koghis and Forêt Nord) have been subject to admixing from *P. fib-aoupinié* and *P. por-pindai*. Although the snail populations on Grande Terre are now allopatric, the SNP data suggest that admixture between populations was likely common in the recent past. It is likely that, prior to the arrival of Europeans who accelerated deforestation, populations on Grande Terre were not allopatric and natural selection drove the adaptation of populations to the local environmental conditions resulting in the formation of what can be defined as ecotypes.

New Caledonia has a diverse environment with extremely high levels of biodiversity. The biogeographic history of New Caledonia is becoming clearer with studies focusing on testing ideas around vicariance and dispersal. However studies on the population genetics of species within New Caledonia are few (but see (Gaudeul *et al.* 2012; Pillon 2012), even though high levels of in situ speciation must have occurred to generate the current biodiversity. The high level of plant and animal micro-endemism on New Caledonia is thought to be predominantly the product of allopatric speciation (Grandcolas *et al.* 2008), and it might therefore have been predicted that speciation in slow moving animals such as giant land snails would be dominated by spatial geography. My results are contrary to this. The high diversity found within the New Caledonian *Placostylus* is driven by local adaptation to the environment, resulting in the formation of ecotypes, regardless of species distributions and gene flow.

ACKNOWLEDGEMENTS

I would like to thank: Alice Mathieu, Jean-Jerome and Christian Mille for helping with the sampling in New Caledonia; Patrick Biggs for helping me with the assembly of the mitochondrial genome; James Crampton for the morphometric guidance; Jean Sanderson for the modelling help. This work was in part funded by the New Caledonian dry forest fund: Convention de collaboration IAC/CEN-PFS n°03-FS/CEN/2012 du 05 mars 2012 action 1-1-2c “génétique des bulimes de forêts sèches” and Massey University Vice-Chancellors scholarship.

REFERENCE

- Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJE *et al.* (2013) Hybridization and speciation. *Journal of Evolutionary Biology* **26**, 229-246.
- Alex Buerkle C, Gompert Z (2013) Population genomics based on low coverage sequencing: how low should we go? *Molecular Ecology* **22**, 3028-3035.
- Appleton RD, Palmer AR (1988) Water-borne stimuli released by predatory crabs and damaged prey induce more predator-resistant shells in a marine gastropod. *Proceedings of the National Academy of Sciences* **85**, 4387-4391.
- Baayen R (2013) Package , languageR.
- Bartish IV, Swenson U, Munzinger JRM, Anderberg AA (2005) Phylogenetic relationships among New Caledonian Sapotaceae (Ericales): molecular evidence for generic polyphyly and repeated dispersal. *American Journal of Botany* **92**, 667-673.
- Bates DM (2010) lme4: Mixed-effects modeling with R. URL <http://lme4.r-forge.r-project.org/book>.
- Beerli P (2006) Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics* **22**, 341-345.
- Beerli P (2009) *How to use MIGRATE or why are Marko chain Monte Carlo programs difficult to use?* Cambridge University Press, New York.
- Bouckaert RR (2010) DensiTree: making sense of sets of phylogenetic trees. *Bioinformatics* **26**, 1372-1373.
- Brescia F, 2011 Ecology and population trends in New Caledonian Placostylus snails (Mollusca: Gastropoda: Bulimulidae): a thesis presented in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Zoology at Massey University, Palmerston North, New Zealand, pp.
- Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA, Roychoudhury A (2012) Inferring Species Trees Directly from Biallelic Genetic Markers: Bypassing Gene Trees in a Full Coalescent Analysis. *Molecular Biology and Evolution* **29**, 1917-1932.
- Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. *Molecular Ecology*.
- Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: building and genotyping loci de novo from short-read sequences. *G3: Genes, Genomes, Genetics* **1**, 171-182.
- Chessel D, Dufour A, Thioulouse J (2013) The ade4 package. *R NEWS*.

- Chiba S (1996) A 40,000-Year Record of Discontinuous Evolution of Island Snails. *Paleobiology* **22**, 177-188.
- Coyne JA, Orr HA (2004) *Speciation*. Sinauer Associates, Sunderland, MA.
- Coyne JA, Price TD (2000) Little evidence for sympatric speciation in island birds. *Evolution* **54**, 2166-2171.
- Cui R, Schumer M, Kruesi K, Walter R, Andolfatto P *et al.* (2013) Phylogenomics reveals extensive reticulate evolution in Xiphophorus fishes. *Evolution*, n/a-n/a.
- Davison A (2002) Land snails as a model to understand the role of history and selection in the origins of biodiversity. *Population Ecology* **44**, 0129-0136.
- Doyle S, Macdonald B, Rochette RM (2010) Is water temperature responsible for geographic variation in shell mass of *Littorina obtusata* (L.) snails in the Gulf of Maine? *Journal of Experimental Marine Biology and Ecology* **394**, 98-104.
- Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* **29**, 1969-1973.
- Earl D, Vonholdt B (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4**, 359-361.
- Espeland M, Murienné J (2011) Diversity dynamics in New Caledonia: towards the end of the museum model? *BMC Evolutionary Biology* **11**, 254.
- Fischer MC, Foll M, Excoffier L, Heckel G (2011) Enhanced AFLP genome scans detect local adaptation in high-altitude populations of a small rodent (*Microtus arvalis*). *Molecular Ecology* **20**, 1450-1462.
- Foll M, Fischer MC, Heckel G, Excoffier L (2010) Estimating population structure from AFLP amplification intensity. *Molecular Ecology* **19**, 4638-4647.
- Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a bayesian perspective. *Genetics* **180**, 977-993.
- Gaudeul M, Rouhan G, Gardner MF, Hollingsworth PM (2012) AFLP markers provide insights into the evolutionary relationships and diversification of New Caledonian *Araucaria* species (Araucariaceae). *American Journal of Botany* **99**, 68-81.
- Geneious version 6.1 created by Biomatters. Available from <http://www.geneious.com/>.
- Goodfriend GA (1986) Variation in land-snail shell form and size and its causes: a review. *Systematic Biology* **35**, 204-223.
- Goodfriend GA, Gould SJ (1996) Paleontology and chronology of two evolutionary transitions by hybridization in the bahamian land snail *Cerion*. *Science* **274**, 1894-1897.
- Grandcolas P, Murienné JRM, Robillard T, Desutter-Grandcolas L, Jourdan H *et al.* (2008) New Caledonia: a very old Darwinian island? *Philosophical Transactions of the Royal Society B: Biological Sciences* **363**, 3309-3317.
- Greve C, Gimnich F, Hutterer R, Misof B, Haase M (2012) Radiating on oceanic islands: patterns and processes of speciation in the land snail genus *Theba* (Risso 1826). *PLoS ONE* **7**, e34339.
- Hardin G (1960) The competitive exclusion principle. *Science* **131**, 1292-1297.
- Harrison RG (2012) The language of speciation. *Evolution* **66**, 3643-3657.
- Hendry AP, Nosil P, Rieseberg LH (2007) The speed of ecological speciation. *Functional Ecology* **21**, 455-464.
- Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**, 1965-1978.

- Hijmans RJ, Guarino L, Bussink C, Mathur P, Cruz M *et al.*, 2013 DIVA-GIS. Vsn. 7.5. A geographic information system for the analysis of species distribution data. Manual available at <http://www.diva-gis.org>, pp.
- Hills SFK, Crampton JS, Trewick SA, Morgan-Richards M (2012) DNA and morphology unite two species and 10 million year old fossils. *PLoS ONE* **7**, e52083.
- Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23**, 1801-1806.
- Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. *Ecology Letters* **7**, 1225-1241.
- Klingenberg CP (2011) MorphoJ: an integrated software package for geometric morphometrics. *Molecular Ecology Resources* **11**, 353-357.
- Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Meth* **9**, 357-359.
- Losos JB, Ricklefs RE (2009) Adaptation and diversification on islands. *Nature* **457**, 830-836.
- Mallet J (2001) The speciation revolution. *Journal of Evolutionary Biology* **14**, 887-888.
- Mallet J (2008) Hybridization, ecological races and the nature of species: empirical evidence for the ease of speciation. *Philosophical Transactions of the Royal Society B: Biological Sciences* **363**, 2971-2986.
- Mallet J, Beltran M, Neukirchen W, Linares M (2007) Natural hybridization in heliconiine butterflies: the species boundary as a continuum. *BMC Evolutionary Biology* **7**, 28.
- Mayr E (1963) *Animal species and evolution*. Belknap Press of Harvard University Press, Cambridge.
- Meirmans PG, Van Tienderen PH (2004) genotype and genodive: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes* **4**, 792-794.
- Miller MA, Pfeiffer W, Schwartz T, 2010 Creating the CIPRES Science Gateway for inference of large phylogenetic trees, pp. 1-8 in *Gateway Computing Environments Workshop (GCE)*, 2010.
- Milne I, Bayer M, Cardle L, Shaw P, Stephen G *et al.* (2010) Tablet--next generation sequence assembly visualization. *Bioinformatics* **26**, 401-402.
- Morgan-Richards M, Smissen RD, Shepherd LD, Wallis GP, Hayward JJ *et al.* (2009) A review of genetic analyses of hybridisation in New Zealand. *Journal of the Royal Society of New Zealand* **39**, 15-34.
- Nattier R, Robillard T, Desutter-Grandcolas L, Couloux A, Grandcolas P (2011) Older than New Caledonia emergence? A molecular phylogenetic study of the eneopterine crickets (Orthoptera: Grylloidea). *Journal of Biogeography* **38**, 2195-2209.
- Neubert E, Cherel-Mora C, Bouchet P (2009) Polytypy, clines, and fragmentation: The bulimes of New Caledonia revisited (Pulmonata, Orthalicoidae, Placostylidae). *Zoologia Neocaledonica* **7**, 37-131.
- Niemiller ML, Fitzpatrick BM, Miller BT (2008) Recent divergence with gene flow in Tennessee cave salamanders (Plethodontidae: Gyrinophilus) inferred from gene genealogies. *Molecular Ecology* **17**, 2258-2275.
- Nosil P (2008) Speciation with gene flow could be common. *Molecular Ecology* **17**, 2103-2106.

- Nosil P, Feder JL (2012) Genomic divergence during speciation: causes and consequences. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**, 332-342.
- Nosil P, Schluter D (2011) The genes underlying the process of speciation. *Trends in Ecology & Evolution* **26**, 160-167.
- Peccoud J, Ollivier A, Plantegenest M, Simon J-C (2009) A continuum of genetic divergence from sympatric host races to species in the pea aphid complex. *Proceedings of the National Academy of Sciences* **106**, 7495-7500.
- Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. *PLoS ONE* **7**, e37135.
- Pillon Y (2012) Time and tempo of diversification in the flora of New Caledonia. *Botanical Journal of the Linnean Society* **170**, 288-298.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959.
- Quesada H, Posada D, Caballero A, Morán P, Rolán-Alvarez E (2007) Phylogenetic evidence for multiple sympatric ecological diversification in a marine snail. *Evolution* **61**, 1600-1612.
- R Core Team, 2013 R: A language and environment for statistical computing, pp. Retrieved from <http://www.R-project.org>. R foundation for Statistical Computing, Vienna, Austria.
- Rambaut A, 2009 FigTree v1.3.1. Available at: <http://tree.bio.ed.ac.uk/software/figtree>, pp.
- Rambaut A, Drummond A, 2007 Tracer v1. 4, pp.
- Räsänen K, Hendry AP (2008) Disentangling interactions between adaptive divergence and gene flow when ecology drives diversification. *Ecology Letters* **11**, 624-636.
- Rohlf FJ, 2013 TpsDig2, Version 2.17. Available at: <http://life.bio.sunysb.edu/morph/>, pp.
- Rosenberg NA (2004) distruct: a program for the graphical display of population structure. *Molecular Ecology Notes* **4**, 137-138.
- Rundle HD, Nosil P (2005) Ecological speciation. *Ecology Letters* **8**, 336-352.
- Senn H, Ogden R, Cezard T, Gharbi K, Iqbal Z *et al.* (2013) Reference-free SNP discovery for the Eurasian beaver from restriction site-associated DNA paired-end data. *Molecular Ecology* **22**, 3141-3150.
- Sheets HD, 2012 IMP714 - Integrated Morphometrics Package. Available from: <http://www3canisius.edu/~sheets/imp714.htm>, pp.
- Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688-2690.
- Stankowski S (2013) Ecological speciation in an island snail: evidence for the parallel evolution of a novel ecotype and maintenance by ecologically dependent postzygotic isolation. *Molecular Ecology* **22**, 2726-2741.
- Teshima H, Davison A, Kuwahara Y, Yokoyama J, Chiba S *et al.* (2003) The evolution of extreme shell shape variation in the land snail *Ainohelix editha*: a phylogeny and hybrid zone analysis. *Molecular Ecology* **12**, 1869-1878.
- Trewick S, Brescia F, Jordan C (2008) Diversity and phylogeny of New Caledonian *Placostylus* land snails; evidence from mitochondrial DNA. *Zoologia Neocaledonica* **6**.
- Trewick SA, Paterson AM, Campbell HJ (2007) Hello New Zealand. *Journal of Biogeography* **34**, 1-6.
- Trussell GC (2000) Phenotypic clines, plasticity, and morphological trade-offs in an intertidal snail. *Evolution* **54**, 151-166.

- Vavrek MJ (2011) fossil: Palaeoecological and palaeogeographical analysis tools. *Palaeontologia Electronica* **14**, 16.
- Via S (2001) Sympatric speciation in animals: the ugly duckling grows up. *Trends in Ecology & Evolution* **16**, 381-390.
- Wagner CE, Keller I, Wittwer S, Selz OM, Mwaiko S *et al.* (2013) Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology* **22**, 787-798.
- Webb CO, Ackerly DD, Mcpeek MA, Donoghue MJ (2002) Phylogenies and Community Ecology. *Annual Review of Ecology and Systematics* **33**, 475-505.
- Woodruff DS, Gould SJ (1987) Fifty years of interspecific hybridization: genetics and morphometrics of a controlled experiment on the land snail *Cerion* in the florida keys. *Evolution* **41**, 1022-1045.
- Wu C-I (2001) The genic view of the process of speciation. *Journal of Evolutionary Biology* **14**, 851-865.
- Zelditch ML, Swiderski DL, Sheets HD (2012) *Geometric morphometrics for biologists: a primer*. Academic Press.
- Zerbino D, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research* **18**, 821-829.
- Zuur AF, Ieno EN, Walker NJ, Saveliev AA, Smith GM (2009) *Mixed effects models and extensions in ecology with R*. Springer.

S3. SUPPLEMENTARY MATERIAL FOR CHAPTER FOUR

S3.1 SNP INFORMATION FOR Sp2

Table 1

SNP information for Sp2. The column 'Private' indicates the number of private sites found in a population and 'Sites' indicates the total number.

# Pop ID	Private	Sites	Variant Sites	Polymorphic Sites	% Polymorphic Loci	Obs Het	Obs Hom	Exp Het	Exp Hom
Touete	182	338449	3688	529	0.156301	0.000847986	0.999152	0.000651501	0.999112
Gadgi	523	458635	4607	1260	0.274728	0.0010186	0.998981	0.000849456	0.999023
Forêt Nord	517	269494	3204	825	0.306129	0.00115989	0.99884	0.0010021	0.998849
Gossanah	509	491176	5333	567	0.115437	0.00115437	0.998846	0.000577186	0.998846
Kéré	493	586495	5674	1035	0.176472	0.000990631	0.999009	0.000724218	0.999018
Kanuméra	17	10447	162	41	0.392457	0.00102046	0.99898	0.000979418	0.998955
Nekoro	1251	248212	3638	1515	0.610365	0.00142667	0.998573	0.00148614	0.998475
Mt Koghis	855	301314	3771	1294	0.429452	0.00194094	0.998059	0.00157159	0.998101
Aoupinié	581	179253	2020	656	0.365963	0.00132483	0.998675	0.00117678	0.998753
Pindai	912	202530	2234	963	0.475485	0.00195527	0.998045	0.00157298	0.9982
Gadgi	24	34958	474	138	0.394759	0.00129325	0.998707	0.00106557	0.998847
Kéré	185	121495	1680	614	0.505371	0.00117122	0.998829	0.00116239	0.998782
Mépouiri	380	119218	1709	596	0.499925	0.00200613	0.997994	0.00177523	0.997968
Youati	636	450558	5271	1257	0.278987	0.00143711	0.998563	0.00116383	0.998419
Vao	521	253257	3094	1369	0.540558	0.00134373	0.998656	0.00134088	0.998595
Ouvéa	180	160331	2160	544	0.339298	0.00194597	0.998054	0.00133474	0.998168

Table 2

Raw number of sites across populations for Sp2, some populations (Kanuméra) did not sequence as well as others and had lower SNP levels between populations.

	Touete	Gadgi	Forêt Nord	Gassanah	Kéré	Kanuméra	Nekoro	Mt Koghis	Aoupinié	Pindai	Gadgi	Kéré	Mépouiri	Youati	Vao	Ouvéa
Touete	0															
Gadgi	830	0														
Forêt Nord	500	750	0													
Gassanah	799	1158	719	0												
Kéré	762	1330	819	1269	0											
Kanuméra	43	56	35	49	54	0										
Nekoro	676	879	599	993	994	35	0									
Mt Koghis	751	976	876	1044	1151	45	791	0								
Aoupinié	163	211	175	226	248	7	281	216	0							
Pindai	175	247	177	284	274	7	320	197	1177	0						
Gadgi	94	118	118	114	112	12	154	87	40	35	0					
Kéré	404	443	346	511	493	26	583	414	116	125	97	0				
Mépouiri	237	329	257	345	356	13	683	322	122	123	88	217	0			
Youati	873	1166	813	1306	1332	49	1422	1031	274	424	137	574	467	0		
Vao	628	860	598	910	967	46	1049	725	228	269	131	505	393	1266	0	
Ouvéa	515	564	370	642	584	44	414	507	98	98	69	268	155	537	393	0

S3.2 SNP INFORMATION FOR SP4

Table 3

SNP information for Sp4. The column 'Private' indicates the number of private sites found in a population and 'Sites' indicates the total number.

# Pop ID	Private	Sites	Variant Sites	Polymorphic Sites	% Polymorphic Loci	Obs Het	Obs Hom	Exp Het	Exp Hom
Touete	85	229682	3273	384	0.167188	0.000901246	0.999099	0.000696615	0.999055
Gadgi	267	280892	3862	839	0.298691	0.00110036	0.9989	0.000924572	0.998938
Forêt Nord	323	177158	2580	566	0.319489	0.0011661	0.998834	0.00102265	0.998827
Gassanah	266	288841	4047	327	0.113211	0.00113211	0.998868	0.000566055	0.998868
Kéré	209	316984	4394	594	0.187391	0.00106315	0.998937	0.000771332	0.998956
Kanuméra	17	9972	161	41	0.411151	0.00106906	0.998931	0.00102607	0.998905
Nekoro	683	144382	2583	873	0.604646	0.00144066	0.998559	0.00149519	0.998465
Mt Koghis	457	187982	2947	813	0.432488	0.00196295	0.998037	0.00160698	0.99806
Aoupinié	154	51957	875	185	0.356064	0.00132963	0.99867	0.00113131	0.998801
Pindai	189	58609	903	199	0.339538	0.00134081	0.998659	0.00109998	0.998741
Gadgi	13	28593	439	117	0.409191	0.00138462	0.998615	0.00111762	0.99879
Kéré	117	92712	1480	486	0.524204	0.00123055	0.998769	0.00119545	0.998748
Mépouiri	180	62978	1148	306	0.485884	0.00201922	0.997981	0.00174874	0.997997
Youati	279	233298	3693	767	0.328764	0.00168454	0.998315	0.00137485	0.998135
Vao	271	157973	2519	981	0.620992	0.00155006	0.99845	0.00153039	0.998397
Ouvéa	130	127658	2006	453	0.354854	0.00201319	0.997987	0.00139043	0.998099

Table 4
Raw number of sites across populations for Sp4

	Touete	Gadgi	Forêt Nord	Gassanah	Kéré	Kanuméra	Nekoro	Mt Koghis	Aoupinié	Pindai	Gadgi	Kéré	Mépouiri	Youati	Vao	Ouvéa
Touete	0															
Gadgi	708	0														
Forêt Nord	471	692	0													
Gassanah	725	1014	616	0												
Kéré	601	938	689	986	0											
Kanuméra	43	56	35	49	54	0										
Nekoro	652	850	566	853	900	35	0									
Mt Koghis	706	920	661	832	969	44	710	0								
Aoupinié	161	210	149	196	238	7	226	194	0							
Pindai	167	214	136	220	244	7	243	172	194	0						
Gadgi	94	118	117	113	112	12	146	86	39	35	0					
Kéré	402	440	331	475	485	26	549	404	111	119	92	0				
Mépouiri	228	316	236	298	321	13	377	293	101	95	86	207	0			
Youati	831	1078	722	1051	1170	49	1046	884	239	299	119	475	352	0		
Vao	622	836	582	842	925	46	943	698	217	239	122	454	353	945	0	
Ouvéa	498	546	368	547	557	44	407	499	92	95	69	269	153	525	392	0

S3.3 SNP INFORMATION FOR UP2

Table 5

SNP information for Up2. The column 'Private' indicates the number of private sites found in a population and 'Sites' indicates the total number.

Pop	Private	Sites	Variant Sites	Polymorphic Sites	% Polymorphic Loci	Obs Het	Obs Hom	Exp Het	Exp Hom	StdErr
Touete	182	338449	3688	529	0.156301	0.000848	0.999152	0.000651501	0.999112	4.08E-05
Gadgi	523	458635	4607	1260	0.274728	0.001019	0.998981	0.000849456	0.999023	4.00E-05
Forêt Nord	517	269494	3204	825	0.306129	0.001159	0.99884	0.0010021	0.998849	6.37E-05
Gassanah	509	491176	5333	567	0.115437	0.001154	0.998846	0.000577186	0.998846	0
Kéré	493	586495	5674	1035	0.176472	0.000991	0.999009	0.000724218	0.999018	2.29E-05
Kanuméra	17	10447	162	41	0.392457	0.00102	0.99898	0.000979418	0.998955	0.000370478
Nekoro	1251	248212	3638	1515	0.610365	0.001427	0.998573	0.00148614	0.998475	0.00126858
Mt Koghis	855	301314	3771	1294	0.429452	0.001941	0.998059	0.00157159	0.998101	5.20E-05
Aoupinié	581	179253	2020	656	0.365963	0.001325	0.998675	0.00117678	0.998753	8.25E-05
Pindai	912	202530	2234	963	0.475485	0.001955	0.998045	0.00157298	0.9982	4.60E-05
Gadgi	24	34958	474	138	0.394759	0.001293	0.998707	0.00106557	0.998847	0.000357924
Kéré	185	121495	1680	614	0.505371	0.001171	0.998829	0.00116239	0.998782	0.00133413
Mépouiri	380	119218	1709	596	0.499925	0.002006	0.997994	0.00177523	0.997968	6.33E-05
Youati	636	450558	5271	1257	0.278987	0.001437	0.998563	0.00116383	0.998419	3.65E-05
Vao	521	253257	3094	1369	0.540558	0.001344	0.998656	0.00134088	0.998595	0.000797697
Ouvéa	180	160331	2160	544	0.339298	0.001946	0.998054	0.00133474	0.998168	9.47E-05

Table 6
Raw number of sites across populations for Up2

	Touete	Gadgi	Forêt Nord	Gassanah	Kéré	Kanuméra	Nekoro	Mt Koghis	Aoupinié	Pindai	Gadgi	Kéré	Mépouiri	Youati	Vao	Ouvéa
Touete	0															
Gadgi	69	0														
Forêt Nord	40	749	0													
Gassanah	57	1157	718	0												
Kéré	18	69	45	67	0											
Kanuméra	4	6	5	9	5	0										
Nekoro	42	638	434	712	33	5	0									
Mt Koghis	75	975	875	1043	53	11	589	0								
Aoupinié	8	210	174	225	13	-	183	215	0							
Pindai	-	27	12	29	-	1	25	19	86	0						
Gadgi	4	14	19	12	3	-	16	16	6	-	0					
Kéré	7	66	45	55	5	-	57	50	9	4	3	0				
Mépouiri	5	31	20	31	5	-	36	23	15	2	2	10	0			
Youati	57	1165	812	1305	46	5	989	1030	273	43	18	61	30	0		
Vao	60	859	599	909	37	6	756	724	227	38	18	70	29	1265	0	
Ouvéa	61	563	369	641	56	10	343	506	97	15	8	42	18	536	393	0

S3.4 SNP INFORMATION FOR UP4

Table 7

SNP information for Up4. The column 'Private' indicates the number of private sites found in a population and 'Sites' indicates the total number.

Pop	Private	Sites	Variant Sites	Polymorphic Sites	% Polymorphic Loci	Obs Het	Exp Het	Obs Hom	Exp Hom
Touete	9	12631	230	38	0.300847	0.00125353	0.000995676	0.998746	0.998824
Gadgi	389	252776	3195	768	0.303826	0.00113111	0.000945268	0.998869	0.998914
Forêt Nord	344	180959	2293	579	0.319962	0.0011895	0.00102634	0.998811	0.998822
Gossanah	282	287607	3539	319	0.110915	0.00110915	0.000554576	0.998891	0.998891
Kéré	17	17479	219	42	0.240288	0.000948758	0.000743948	0.999051	0.999146
Kanuméra	5	2280	27	2	0.0877193	0.000389864	0.00023825	0.99961	0.999748
Nekoro	578	112375	1742	371	0.330145	0.00134226	0.00116305	0.998658	0.998807
Mt Koghis	499	197011	2627	844	0.428402	0.00194913	0.00158621	0.998051	0.998085
Aoupinié	179	50440	712	146	0.289453	0.00134043	0.0010609	0.99866	0.998876
Pindai	20	5889	86	22	0.373578	0.00146035	0.00108338	0.99854	0.998794
Gadgi	2	3040	55	11	0.361842	0.00241424	0.00135098	0.997586	0.998556
Kéré	13	10925	162	31	0.283753	0.00131175	0.000889709	0.998688	0.999077
Mépouiri	16	5700	95	25	0.438596	0.00164912	0.00131447	0.998351	0.998535
Youati	392	244039	3228	793	0.324948	0.00167391	0.001363	0.998326	0.998151
Vao	405	167950	2221	531	0.316166	0.00153195	0.00117273	0.998468	0.998771
Ouvéa	143	128137	1768	441	0.344163	0.00195884	0.00135012	0.998041	0.998153

Table 8
Raw number of sites across populations for Up4

	Touete	Gadgi	Forêt Nord	Gossanah	Kéré	Kanuméra	Neekoro	Mt Koghis	Laopinié	Pindai	Gadgi	Kéré	Mépouiri	Youati	Vao
Touete	0														
Gadgi	62	0													
Forêt Nord	40	679	0												
Gassanah	54	946	661	0											
Kéré	19	67	45	62	0										
Kanuméra	2	2	3	5	3	0									
Nékoro	33	486	342	503	23	5	0								
Mt Koghis	74	907	708	886	53	9	462	0							
Laopinié	7	191	149	177	13	-	105	181	0						
Pindai	-	19	11	24	-	1	18	18	21	0					
Gadgi	3	14	18	25	2	-	6	14	5	-	0				
Kéré	6	53	35	44	3	-	39	37	6	3	2	0			
Mépouiri	5	31	20	30	5	-	19	22	12	-	1	8	0		
Youati	53	1088	750	1130	46	5	612	923	232	38	15	37	29	0	
Vao	46	677	475	655	32	5	426	560	158	30	11	38	19	616	0
Ouvéa	61	540	369	535	56	6	271	503	89	15	7	40	18	537	324

S3.5 F_{ST} RESULTS FOR DATASTS SP4, UP2 AND UP4

Table 9

The GENODIVE F_{ST} results for Sp4. Ouvéa here contains populations from both Téoouta and Gossanah.

	IP:Touete (f)	IP:Gadgi (f)	Forêt Nord (f)	IP:Kéré (f)	IP:Kaniméra (f)	Nekoro (p)	Mt Koghis (f)	Aoupinié (f)	Pindai (p)	IP:Gadgi (p)	IP:Kéré (p)	Mépouiri (p)	IP:Youati (p)	IP:Vao (p)	Ouvéa (f)
IP:Touete(f)	0	0.087	0.401	0.083	0.17	0.696	0.432	0.624	0.661	0.736	0.646	0.688	0.701	0.625	0.52
IP:Gadgi(f)	0.087	0	0.419	0.055	0.255	0.697	0.486	0.678	0.702	0.743	0.686	0.709	0.738	0.644	0.488
Forêt Nord(f)	0.401	0.419	0	0.435	0.57	0.655	0.367	0.634	0.642	0.73	0.694	0.672	0.73	0.664	0.557
IP:Kéré(f)	0.083	0.055	0.435	0	0.199	0.71	0.484	0.654	0.696	0.728	0.675	0.678	0.713	0.631	0.528
IP:Kaniméra (f)	0.17	0.255	0.57	0.199	0	0.725	0.51	0.735	0.875	0.369	0.647	0.654	0.646	0.588	0.202
Nekoro(p)	0.696	0.697	0.655	0.71	0.725	0	0.672	0.704	0.699	0.397	0.52	0.242	0.478	0.519	0.656
Mt Koghis(f)	0.432	0.486	0.367	0.484	0.51	0.672	0	0.602	0.644	0.712	0.645	0.651	0.67	0.614	0.491
Aoupinié(f)	0.624	0.678	0.634	0.654	0.735	0.704	0.602	0	0.518	0.825	0.78	0.655	0.787	0.759	0.675
Pindai(p)	0.661	0.702	0.642	0.696	0.875	0.699	0.644	0.518	0	0.762	0.757	0.625	0.728	0.71	0.748
IP:Gadgi(p)	0.736	0.743	0.73	0.728	0.369	0.397	0.712	0.825	0.762	0	0.051	0.525	0.013	0.028	0.768
IP:Kéré(p)	0.646	0.686	0.694	0.675	0.647	0.52	0.645	0.78	0.757	0.051	0	0.477	0.074	0.119	0.697
Mépouiri (p)	0.688	0.709	0.672	0.678	0.654	0.242	0.651	0.655	0.625	0.525	0.477	0	0.401	0.487	0.622
IP:Youati(p)	0.701	0.738	0.73	0.713	0.646	0.478	0.67	0.787	0.728	0.013	0.074	0.401	0	0.045	0.726
IP:Vao(p)	0.625	0.644	0.664	0.631	0.588	0.519	0.614	0.759	0.71	0.028	0.119	0.487	0.045	0	0.687
Ouvéa(f)	0.52	0.488	0.557	0.528	0.202	0.656	0.491	0.675	0.748	0.768	0.697	0.622	0.726	0.687	0

Table 10

The GENODIVE F_{ST} results for Up2. Ouvéa here contains populations from both Téouta and Gossanah.

	IP:Touete (f)	IP:Gadgi (f)	Forêt Nord (f)	IP:Kéré (f)	IP:Kaniméra (f)	Nekoro (p)	Mt Koghis (f)	Aoupinié (f)	Pindai (p)	IP:Gadgi (p)	IP:Kéré (p)	Mépouiri (p)	IP:Youati (p)	IP:Vao (p)	Ouvéa (f)
IP:Touete(f)	0	0.032	0.515	0.326	-	0.702	0.471	0.334	-	0.823	0.624	0.219	0.796	0.645	0.545
IP:Gadgi(f)	0.032	0	0.404	0.082	-	0.712	0.504	0.71	0.736	0.891	0.723	0.829	0.76	0.656	0.501
Forêt Nord(f)	0.515	0.404	0	0.244	0.552	0.668	0.383	0.638	0.59	0.81	0.68	0.758	0.726	0.659	0.557
IP:Kéré(f)	0.326	0.082	0.244	0	-	0.473	0.576	0.707	-	0.074	0.323	0.219	0.72	0.487	0.519
IP:Kaniméra (f)	-	-	0.552	-	0	0.835	0.625	-	-	-	-	-	0.788	0.743	0.471
Nekoro(p)	0.702	0.712	0.668	0.473	0.835	0	0.681	0.685	0.536	0.634	0.393	0.177	0.494	0.556	0.672
Mt Koghis (f)	0.471	0.504	0.383	0.576	0.625	0.681	0	0.624	0.761	0.726	0.702	0.839	0.666	0.613	0.511
Aoupinié(f)	0.334	0.71	0.638	0.707	-	0.685	0.624	0	0.423	1	0.979	0.751	0.782	0.769	0.688
Pindai(p)	-	0.736	0.59	-	-	0.536	0.761	0.423	0	-	-	0.571	0.77	0.657	0.742
IP:Gadgi(p)	0.823	0.891	0.81	0.074	-	0.634	0.726	1	-	0	-0.004	-	-0.02	-0.013	0.676
IP:Kéré(p)	0.624	0.723	0.68	0.323	-	0.393	0.702	0.979	-	-0.004	0	0.093	0.062	0.128	0.653
Mépouiri (p)	0.219	0.829	0.758	0.219	-	0.177	0.839	0.751	0.571	-	0.093	0	0.528	0.428	0.837
IP:Youati(p)	0.796	0.76	0.726	0.72	0.788	0.494	0.666	0.782	0.77	-0.02	0.062	0.528	0	0.056	0.744
IP:Vao(p)	0.645	0.656	0.659	0.487	0.743	0.556	0.613	0.769	0.657	-0.013	0.128	0.428	0.056	0	0.695
Ouvéa(f)	0.545	0.501	0.557	0.519	0.471	0.672	0.511	0.688	0.742	0.676	0.653	0.837	0.744	0.695	0

Table 11

The GENODIVE F_{ST} results for Up4. Ouvéa here contains populations from both Téoouta and Gossanah.

	IP:Touete (f)	IP:Gadgi (f)	Forêt Nord (f)	IP:Kéré (f)	IP:Kanuméra (f)	Nekoro (p)	Mt Koghis (f)	Aoupinié (f)	Pindai (p)	IP:Gadgi (p)	IP:Kéré (p)	Mépouiri (p)	IP:Youati (p)	IP:Vao (p)	Ouvéa (f)
IP:Touete(f)	0	-0.004	0.527	0.038	-	0.653	0.472	0.449	-	0.823	0.624	0.219	0.81	0.645	0.6
IP:Gadgi(f)	-0.004	0	0.396	0.057	-	0.688	0.503	0.727	0.753	0.891	0.701	0.773	0.752	0.654	0.464
Forêt Nord(f)	0.527	0.396	0	0.27	0.552	0.669	0.376	0.687	0.717	0.81	0.604	0.642	0.734	0.658	0.569
IP:Kéré(f)	0.038	0.057	0.27	0	-	0.473	0.608	0.707	-	0.074	0.323	0.219	0.72	0.502	0.544
IP:Kanuméra (f)	-	-	0.552	-	0	0.835	0.625	-	-	-	-	-	0.788	0.743	0.471
Nekoro(p)	0.653	0.688	0.669	0.473	0.835	0	0.671	0.695	0.712	0.191	0.393	0.046	0.499	0.531	0.668
Mt Koghis(f)	0.472	0.503	0.376	0.608	0.625	0.671	0	0.616	0.761	0.726	0.657	0.793	0.671	0.608	0.54
Aoupinié(f)	0.449	0.727	0.687	0.707	-	0.695	0.616	0	0.607	1	0.958	0.795	0.8	0.759	0.696
Pindai(p)	-	0.753	0.717	-	-	0.712	0.761	0.607	0	-	-	-	0.792	0.714	0.742
IP:Gadgi(p)	0.823	0.891	0.81	0.074	-	0.191	0.726	1	-	0	-0.004	-	-0.011	-0.007	0.676
IP:Kéré(p)	0.624	0.701	0.604	0.323	-	0.393	0.657	0.958	-	-0.004	0	0.093	0.012	0.129	0.588
Mépouiri (p)	0.219	0.773	0.642	0.219	-	0.046	0.793	0.795	-	-	0.093	0	0.458	0.464	0.788
IP:Youati(p)	0.81	0.752	0.734	0.72	0.788	0.499	0.671	0.8	0.792	-0.011	0.012	0.458	0	0.047	0.756
IP:Vao(p)	0.645	0.654	0.658	0.502	0.743	0.531	0.608	0.759	0.714	-0.007	0.129	0.464	0.047	0	0.69
Ouvéa(f)	0.6	0.464	0.569	0.544	0.471	0.668	0.54	0.696	0.742	0.676	0.588	0.788	0.756	0.69	0

S3.6 STRUCTURE RESULTS FROM SP4, UP2 AND UP4

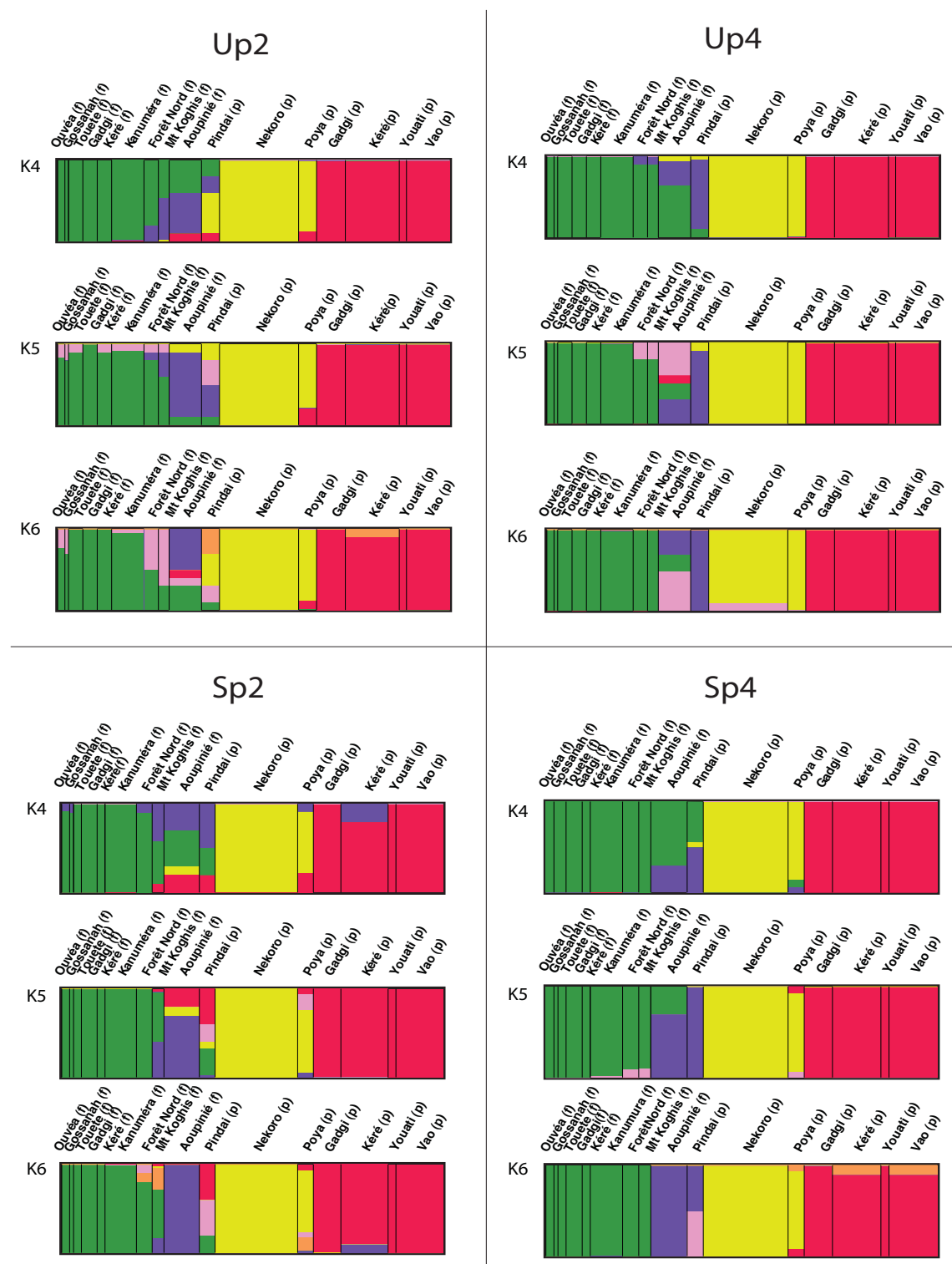


Figure 1

STRUCTURE results from Sp2, Sp4, Up2 and Up4. The results between groups were very similar so the largest dataset, Sp2, was used for the final runs.

S3.7 TOPOLOGY OF UNCOMMON SNAPP TREES

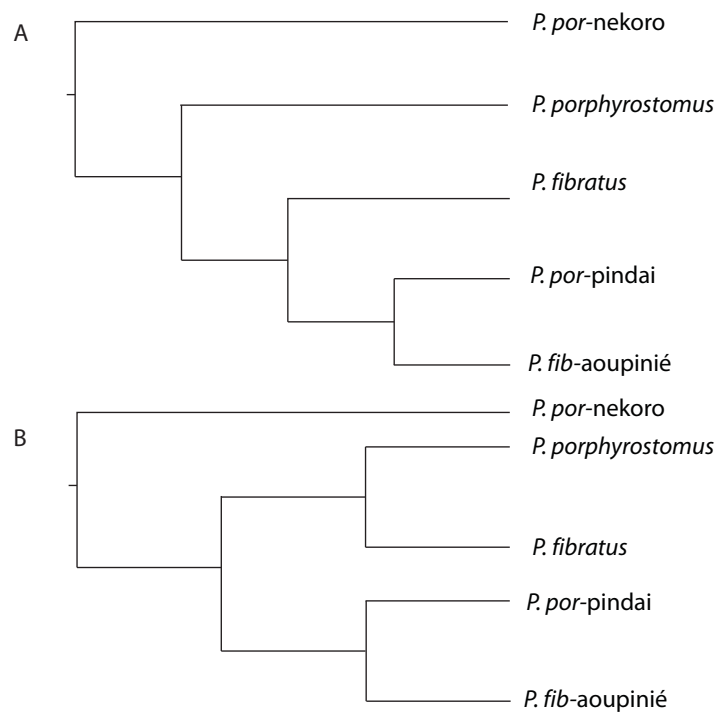


Figure 2

The two next most common tree topologies from SNAPP: (A) was found 12% of the time and (B) 7% of the time. Species are split based on their genetic structure: *P. porphyrostomus* (Ile des Pins), *P. fibratus* (Ile des Pins, Ouvéa, Forêt Nord and Mt Koghis), *P. por-nekoro* (Nekoro and Mepouiri), *P. fib-aoupinié* (Aoupinié) and *P. por-pindai* (Pindai). Branch lengths here are arbitrary.

5. SUMMATION

Biogeography on the large scale, such as the Latitudinal Biodiversity Gradient (LBG), must in part be generated by small-scale genomic divergence (micro-evolution). As our datasets have expanded our understanding of the processes driving divergence has improved and we now regard speciation to be a dynamic process that is frequently driven by natural selection. In this study I sought, through four related studies, to examine the drivers of divergence and adaptation between populations and how this could drive large-scale diversity patterns.

DIVERGENCE BETWEEN POPULATIONS IS FREQUENTLY DRIVEN BY NATURAL SELECTION

Selection and Gene flow

New sequencing technologies allow for in-depth studies of non-model species. Where, until recently, we relied heavily on a few small readily amplified regions of the genome, we can now select loci across the genome for analysis. This is particularly useful when examining the complex processes that can surround speciation. As shown in chapters 3 and 4 one of the best ways to understand how natural selection can drive speciation is through the analysis of divergent selection in the presence of gene flow. Genomic divergence mediated through selection on particular loci is thus an important emerging field of study (Wu 2001; Feder *et al.* 2012; Nosil and Feder 2012; Via *et al.* 2012). Gene flow (hybridisation) between species and populations is no longer considered a barrier to speciation but part of the process of speciation (Abbott *et al.* 2013). Earlier theory predicted that gene flow presented a impermeable barrier to divergence and that speciation was driven by mutational drift in allopatric populations resulting in genetic incompatibility between species (Mayr 1963). I have shown, as others have before, that gene flow is neither uncommon nor necessarily an impediment to divergence and that natural

selection can cause populations to diverge and adapt regardless of species distributions (Mallet 2001; Nosil 2008).

Examples of hybridisation and gene flow between species are found in both plants and animals as exemplified in the work of (Feder *et al.* 1988; Hey *et al.* 2004; Savolainen *et al.* 2006; Mallet *et al.* 2007; Nosil 2008). One species, the pea aphid (*Acyrtosiphon pisum*), shows a continuum from local adaptation to near full reproductive isolation. Pea aphids are classified as a single species but they feed on a variety of plants to which they are adapted (termed a biotype) and sympatric biotypes are not necessarily reproductively isolated. Between sympatric biotypes some have high rates of hybridisation whilst others are nearing full reproductive isolation, this process is likely to be occurring in the presence of gene flow (Peccoud *et al.* 2009). Gene flow mediated by hybridisation is more likely to occur between close taxa, but is not limited to these groups; within the *Placostylus* snails there are strong signals for gene flow between well-established species on Ile des Pins. Even in the face of extremely high rates of gene flow (chapter 3) divergence can be maintained. In instances of high gene flow, detecting the signal for genetic partitioning with neutral markers may not be possible (chapter 3) and divergence is only observed through traits and loci that are subject to strong natural selection (Thibert-Plante and Hendry 2010). Gene flow may have a role in providing novel alleles and increasing opportunities for populations to adapt to changing environments or expansion to use new resources (Dowling and Secor 1997; Abbott *et al.* 2013).

Selection acts upon loci that enhance the fitness of an individual to its environment. If selection is strong enough, divergence and the formation of ecotypes will occur, which many now regard as a stage in the speciation process (Rundle and Nosil 2005; Mallet 2008; Räsänen and Hendry 2008). Detecting traits and loci under selection can be achieved through comparisons of expressed genes (i.e. transcriptome sequencing) or through quantification of physical traits. In studies of many non-model species this means a shift to a dual focus on genetics and morphology.

Selection and Morphology

Morphological differences found between species have long formed the basis of taxonomy, and analysis of morphological traits can be an extremely useful complement to genetic studies (chapters 3 and 4). A geometric morphometric approach allows us to incorporate more information into an analysis rather than relying heavily on one or a few, often subjectively chosen, categorical morphological traits. Morphology can be a plastic response induced by the interaction of abiotic and biotic stimuli or a fixed inherited trait. Molluscs in particular are known for having plastic morphological responses to external stimuli (Appleton and Palmer 1988; Trussell 2000). Within New Caledonian *Placostylus* snails, morphological variation is strongly linked to the environmental conditions (chapter 4). However, this does not mean that the shell morphological variation is a plastic response to the environment. I found concordance between genetic groups and morphological structuring, despite the similar shell shape found among genetically distinct snail populations from similar environments. A more likely explanation is that *Placostylus* snails are able to adapt to their environment and are therefore not restricted to a particular environmental niche. Contrary to traditional interpretation, *P. fibratus* do not represent wet forest snails and *P. porphyrostomus* do not represent dry forest snails.

BIOGEOGRAPHY IS DRIVEN IN PART BY POPULATION LEVEL DIVERGENCE

With a understanding that speciation is largely driven by natural selection, how then can we relate this to the formation of the LBG?

The latitudinal biodiversity gradient

The latitudinal biodiversity gradient describes the pattern of higher species diversity towards the tropics, and species diversity is a product of speciation, extinction and migration (chapter 1). Some studies have found speciation to be greater in the tropics than temperate regions (Jablonski *et al.* 2006; Eo *et al.* 2008). Moreover, speciation and divergence, whether in the tropics or temperate regions, is frequently driven by natural selection. Why selection might result in more species in

the tropics than in temperate regions is unclear, however we are not short of hypotheses. The evolutionary speed hypothesis suggests that species in the tropics are evolving faster than species in temperate regions. Studies of rates of molecular evolution have often found faster rates towards the tropics and this difference in rate of evolution has been suggested to lead to the faster speciation rates reported in the tropics (Rohde 1992; Gillman *et al.* 2009; Gillman *et al.* 2010; Wright *et al.* 2011; Gillman *et al.* 2012). Supporting this hypothesis, molecular evolution rate has been found to be faster in some specious groups (Lanfear *et al.* 2010). But we have not yet determined whether molecular evolution increases the rate of speciation or if speciation causes an increased rate of molecular evolution (Pagel *et al.* 2006; Lanfear *et al.* 2010; Venditti and Pagel 2010), a dilemma to which there is no easy answer.

The foundations on which the evolutionary speed hypothesis is based are not solid and it is clear that we must improve our abilities to estimate differences in rates of molecular evolution between lineages. In chapter 2 I showed that it is not possible to reliably detect rate differences with the sister-species approach, even though this is by far the most common method used (Gillman *et al.* 2009; Gillman *et al.* 2010; Lanfear *et al.* 2010; Goldie *et al.* 2011; Gillman *et al.* 2012). In fact, in my dataset, I found only one instance that fulfilled the expectation that one species would have a faster rate of molecular evolution across all mtDNA genes in both outgroup methods used; an expectation that is an implicit assumption of all sister-species studies. Particularly concerning, but perhaps not entirely surprising, were the high levels of inconsistency within the closely related sister pairs of *Drosophila*. Closely related sister-species are often used to estimate rates of molecular evolution between latitudes. Using closely related sister-species mitigates some of the complications existing around unshared ancestors and the timing of derived traits, in particular the timing of an acceleration of rate of molecular evolution towards the tropics. Unfortunately, as I have shown, these close sister-species likely give poor estimates of rate of molecular evolution.

This does not mean that there is no difference in molecular evolution rate between tropical and temperate regions; rather, if a difference does exist we are currently

unable to detect one by using the sister-species approach. If there is a faster rate of molecular evolution in the tropics it is important to know if this is a fixation rate shift or mutation rate shift, as the observed rate of molecular evolution is the product of both.

SPECIATION MEDIATED THROUGH CHANGES IN FIXATION AND MUTATION RATE

Speciation has previously been subdivided into two types, ecological speciation and mutation order speciation (Schluter 2009). Ecological speciation is driven through natural selection on abiotic and biotic factors, whereas mutational order speciation is divergence due to the fixation of alternative and incompatible mutations in allopatric species with similar selection pressures. A change in either fixation or mutation rate could increase both forms of speciation.

Fixation rate

If population sizes differ between tropical and temperate regions then a difference in fixation rate will cause a shift in rates of molecular evolution. Population size influences the efficiency of selection, with different population sizes leading to different fixation rates. Accordingly, differences in molecular evolution rate occur between different population sizes. A faster fixation rate, which under nearly-neutral theory would occur in a small population, could quicken the process of mutational order speciation. Meanwhile a larger population size will not only increase the efficiency of natural selection but also increase the pool of genetic diversity held within a population at any one time; diversity upon which natural selection could act. This could lead to an increase in ecological speciation (Gossmann *et al.* 2012). If population size does vary between tropical and temperate regions the directional effect that it would have on the rate of molecular evolution is still unclear (chapter 1). However mutation rate is promoted as the likely cause of the shift in molecular evolution rate between latitudes (Rohde 1992).

Mutation rate

If there is a faster rate of mutation towards the tropics this could increase speciation in two ways. With an increase in mutation rate and subsequently molecular evolution rate, differences between isolated species may quicken and mutational order speciation could be enhanced. But given that we now understand speciation might be largely driven by natural selection, could adaptation be limited by rate of molecular evolution?

Population genetic variation upon which selection can act arises via mutations in the genomes of individuals. If a faster rate of molecular evolution is driven by higher genome mutation rates this could provide populations with more variation (Lancaster 2010). Although mutation is the underlying source of variation needed for adaptation to new and changing environments, most mutations are detrimental and their fixation rate is highly constrained (Lynch 2010). Rates of synonymous-substitution, i.e. not causing amino acid replacements, vary between genomes, chromosomes, genes and sites. This synonymous-substitution rate variation is thought to be driven by mutation rate variation throughout the genome (Hodgkinson and Eyre-Walker 2011). In genes where mutations are likely to be largely deleterious, such as those mediating essential cellular processes, lower rates of mutation have been recorded in comparison to genes that could have dynamic selection pressures (e.g. extracellular communication); this may increase the variation selection can act upon within these genes (Chuang and Li 2004; Hodgkinson and Eyre-Walker 2011). An increase in mutation rate across the entire genome could therefore be potentially costly and unnecessary. Species also need not rely on *in situ* mutation for all new variation. Hybridisation can allow the exchange of adaptive traits between populations/species and, given that hybridisation maybe occurring among 10-30% of animal species, it is an important evolutionary process (Abbott *et al.* 2013). It therefore seems unlikely that studies focusing on large scale genome rates of molecular evolution estimated from a single gene will resolve the question of whether lineages are adaptively limited by their mutation rate.

CONCLUSION

Species diversity patterns, such as the LBG, are the result of a range of processes controlling speciation, extinction and immigration rates (chapter 1). To fully comprehend the LBG we first need to understand the roles that speciation, extinction and immigration play in its formation. If speciation is faster in the tropics, as has been reported so far, the focus must shift to the drivers of speciation. Current research on the LBG focuses on describing patterns of increased rates of molecular evolution towards the tropics. But how an increase in molecular evolution rate towards the tropics could increase speciation rate is not known, and as found in chapter 2 current estimates of molecular evolution rate variation between regions are poor. Population genetic analyses across latitudes, such as chapters 3 and 4, frequently find divergence and speciation to be driven by natural selection, yet studies on the LBG often do not consider the role of natural selection in speciation. The LBG provides a fascinating opportunity to understand how natural selection can drive differences in speciation rates and whether natural selection is limited by rate of molecular evolution; processes understood best through population genetic analyses.

REFERENCES

- Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJE *et al.* (2013) Hybridization and speciation. *Journal of Evolutionary Biology* **26**, 229-246.
- Appleton RD, Palmer AR (1988) Water-borne stimuli released by predatory crabs and damaged prey induce more predator-resistant shells in a marine gastropod. *Proceedings of the National Academy of Sciences* **85**, 4387-4391.
- Chuang JH, Li H (2004) Functional Bias and Spatial Organization of Genes in Mutational Hot and Cold Regions in the Human Genome. *PLoS Biol* **2**, e29.
- Dowling TE, Secor CL (1997) The role of hybridization and introgression in the diversification of animals. *Annual Review of Ecology and Systematics* **28**, 593-619.
- Eo SH, Wares JP, Carroll JP (2008) Population divergence in plant species reflects latitudinal biodiversity gradients. *Biology Letters* **4**, 382-384.
- Feder JL, Chilcote CA, Bush GL (1988) Genetic differentiation between sympatric host races of the apple maggot fly *Rhagoletis pomonella*. *Nature* **336**, 61-64.
- Feder JL, Egan SP, Nosil P (2012) The genomics of speciation-with-gene-flow. *Trends in Genetics* **28**, 342-350.
- Gillman LN, Keeling DJ, Gardner RC, Wright SD (2010) Faster evolution of highly conserved DNA in tropical plants. *Journal of Evolutionary Biology* **23**, 1327-1330.
- Gillman LN, Keeling DJ, Ross HA, Wright SD (2009) Latitude, elevation and the tempo of molecular evolution in mammals. *Proceedings of the Royal Society B-Biological Sciences* **276**, 3353-3359.
- Gillman LN, Mccowan LSC, Wright SD (2012) The tempo of genetic evolution in birds: body mass and climate effects. *Journal of Biogeography*, no-no.
- Goldie X, Lanfear R, Bromham L (2011) Diversification and rate of molecular evolution: no evidence of a link in mammals. *BMC Evolutionary Biology* **11**.
- Gossmann T, Keightley P, Eyre-Walker A (2012) The effect of variation in the effective population size on the rate of adaptive molecular evolution in eukaryotes. *Genome Biology and Evolution*.
- Hey J, Won Y-J, Sivasundar A, Nielsen R, Markert JA (2004) Using nuclear haplotypes with microsatellites to study gene flow between recently separated Cichlid species. *Molecular Ecology* **13**, 909-919.
- Hodgkinson A, Eyre-Walker A (2011) Variation in the mutation rate across mammalian genomes. *Nat Rev Genet* **12**, 756-766.
- Jablonski D, Roy K, Valentine JW (2006) Out of the tropics: Evolutionary dynamics of the latitudinal diversity gradient. *Science* **314**, 102-106.
- Lancaster L (2010) Molecular evolutionary rates predict both extinction and speciation in temperate angiosperm lineages. *BMC Evolutionary Biology* **10**, 162.
- Lanfear R, Ho SYW, Love D, Bromham L (2010) Mutation rate is linked to diversification in birds. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 20423-20428.
- Lynch M (2010) Evolution of the mutation rate. *Trends in Genetics* **26**, 345-352.
- Mallet J (2001) The speciation revolution. *Journal of Evolutionary Biology* **14**, 887-888.
- Mallet J (2008) Hybridization, ecological races and the nature of species: empirical evidence for the ease of speciation. *Philosophical Transactions of the Royal Society B: Biological Sciences* **363**, 2971-2986.
- Mallet J, Beltran M, Neukirchen W, Linares M (2007) Natural hybridization in heliconiine butterflies: the species boundary as a continuum. *BMC Evolutionary Biology* **7**, 28.

- Mayr E (1963) *Animal species and evolution*. Belknap Press of Harvard University Press, Cambridge.
- Nosil P (2008) Speciation with gene flow could be common. *Molecular Ecology* **17**, 2103-2106.
- Nosil P, Feder JL (2012) Genomic divergence during speciation: causes and consequences. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**, 332-342.
- Pagel M, Venditti C, Meade A (2006) Large punctuational contribution of speciation to evolutionary divergence at the molecular level. *Science* **314**, 119-121.
- Peccoud J, Ollivier A, Plantegenest M, Simon J-C (2009) A continuum of genetic divergence from sympatric host races to species in the pea aphid complex. *Proceedings of the National Academy of Sciences* **106**, 7495-7500.
- Räsänen K, Hendry AP (2008) Disentangling interactions between adaptive divergence and gene flow when ecology drives diversification. *Ecology Letters* **11**, 624-636.
- Rohde K (1992) Latitudinal Gradients in Species-Diversity - the Search for the Primary Cause. *Oikos* **65**, 514-527.
- Rundle HD, Nosil P (2005) Ecological speciation. *Ecology Letters* **8**, 336-352.
- Savolainen V, Anstett M-C, Lexer C, Hutton I, Clarkson JJ *et al.* (2006) Sympatric speciation in palms on an oceanic island. *Nature* **441**, 210-213.
- Schluter D (2009) Evidence for Ecological Speciation and Its Alternative. *Science* **323**, 737-741.
- Thibert-Plante X, Hendry AP (2010) When can ecological speciation be detected with neutral loci? *Molecular Ecology* **19**, 2301-2314.
- Trussell GC (2000) Phenotypic clines, plasticity, and morphological trade-offs in an intertidal snail. *Evolution* **54**, 151-166.
- Venditti C, Pagel M (2010) Speciation as an active force in promoting genetic evolution. *Trends in Ecology and Evolution* **25**, 14-20.
- Via S, Conte G, Mason-Foley C, Mills K (2012) Localizing FST outliers on a QTL map reveals evidence for large genomic regions of reduced gene exchange during speciation-with-gene-flow. *Molecular Ecology* **21**, 5546-5560.
- Wright S, Ross H, Jeanette Keeling D, McBride P, Gillman L (2011) Thermal energy and the rate of genetic evolution in marine fishes. *Evolutionary Ecology* **25**, 525-530.
- Wu C-I (2001) The genic view of the process of speciation. *Journal of Evolutionary Biology* **14**, 851-865.