Evolution and Stable Isotopes in *Placostylus* species of the southwest Pacific

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

Doctor of Philosophy

In

Zoology

Massey University

New Zealand

Mathieu Quenu

2021

Preface

Human activities during the Holocene have induced a sixth biodiversity crisis and initiated rapid changes in the climate. The anthropogenic pressures put on ecosystems can result in direct or indirect environmental degradation, fragmentation and defaunation. Understanding local patterns of wildlife population structure, species interactions and initial biodiversity are all crucial to making well-informed decisions that leads to population sustainability and conservation of global biodiversity.

This thesis is focused on the genus of giant land snail *Placostylus* and seeks to improve our overall knowledge of the genus and its potential to store information about the local environment (such as temperature and humidity) during shell formation. *Placostylus* is a genus endemic to the southwest Pacific and the many species present a valuable opportunity to integrate studies of ecology and environment at a scale relevant to current anthropogenic climate change. The characteristics of *Placostylus* shells can be used to investigate extant and extinct morphological variation within the genus, and their chemical composition can be used to track the environmental conditions in which the snails lived. In parallel to shell analysis the generation of genetic data can be used to infer phylogenetic relationships between distant taxa, and at a fine-scale patterns of population structure allow us to infer gene flow and differentiation. Understanding the extent to which shell shape and size is controlled by genetic differences and how much phenotypic plasticity leads to differences is essential if we are to correctly interpret the significant of phenotypic variation. For example, arid conditions can lead to *Placostylus* snails maturing when much smaller in size. Potentially, intraspecific shell shape and size variation and shell chemistry can all inform us about the local environmental conditions that existed as snail shells were formed.

Three main axes are developed throughout the thesis. First the diversity of *Placostylus* and extended species of the super-family Orthalicoidea are introduced using a phylogenetic investigation. Evolutionary relationships are inferred from DNA sequences of mitochondrial and nuclear genetic datasets. Second, morphological variation is examined in detail where two *Placostylus* snail species are sympatric (the Isle of Pines, New Caledonia). The variation in shell shape of taxa living and growing in the same environment must represent genetic differences rather than phenotypic plasticity. However, genetic data from the *Placostylus* species present on the Isle of Pines was needed when a third snail morphotype was discovered. On the Isle of Pines giant land snails of the species *P. fibratus* are harvested for food, where

they are sympatric with the vulnerable species *P. porphyrostomus*. Understanding local population structure of both species and their interaction will inform management decisions for both species. Third, the stable isotopic composition of extant *Placostylus* shells is analysed from *Placostylus* shells from New Zealand and New Caledonia. This works has the aim to establish a climate proxy system which through the analysis of fossil shells could inform us about past environmental conditions. A protocol to sample high-resolution isotopic signatures from *Placostylus* shells is developed and the stable isotopic composition of shells are examined in light of the environmental variables of the snail collection locations.

Acknowledgement

Mary and Steve, thank you for giving me this opportunity and for your overall support over the last five years. From the very beginning I always felt like I had your trust and I could always rely on you when things didn't go quite as planned. I feel very lucky I had you both as my main supervisors.

I would like to thank Mike Gemmel for his help starting the morphometric analyses, Christian Hyde for his help with the isotope sampling and Fabrice Brescia for collecting most of the material I have been working with. Drew and Kat thank you for your help setting up the isotopic work, for your feedbacks and for the enthusiastic discussions all throughout the project. Nim, Julien, David and all the past and present members of the phoenix group thank you for being the helpful and supportive group of people that you are.

To my family, Maman, Papa, Romain, Marion je sais que ces dernières années loin de la maison n'ont pas dû être les plus faciles pour vous. Beaucoup de choses se sont passées et ont changées en quatre ans et avec la distance et la pandémie tout a été plus compliqué. Merci de m'avoir toujours soutenu et à bientôt.

Finally, I have to express my gratitude to all the people that have helped me getting through the tough times following my diagnosis. To my close friends, family, flatmates, supervisors and to the healthcare providers that helped me, thank you. Finishing this thesis in these conditions was not easy and if I managed to do it, it is thanks to your support. I don't know if I will ever truly recover from this and the future still seem very uncertain at times, but at least I know there are people around me I can truly rely on.

.

Thesis structure

In chapter 1, I use high-throughput sequencing technologies to assemble the full DNA sequences of mitochondrial genomes and multi-copy nuclear genes for five species of terrestrial snail from the Pacific (Placostylinae sub-family). Phylogenetic relationships among these taxa and other related species are investigated. This chapter has been formatted for the journal *Molluscan Research* and is currently under review.

In Chapter 2 I study morphological variation within of a sample of *Placostylus* shells collected on the Isle of Pines, New Caledonia. The morphological variability of shells is analysed through geometric morphometric and machine learning methods. Using supervised learning algorithms and unsupervised learning algorithms I found that not all shells can be confidently assigned to either or other of the two *Placostylus* species and that a third shell phenotype is present on the island. This paper has been formatted for *Journal of Molluscan Studies* and was published in February 2020.

In Chapter 3 I develop nuclear markers for population genetic studies involving the two species *P. fibratus* and *P. porphyrostomus.* To do this I use high-throughput paired-end genomic library dataset to design primer pairs surrounding potentially variable sequences. I successfully amplify four microsatellite markers and two Single Nucleotide Polymorphism markers and test them on two population samples of *P. fibratus* and *P. porphyrostomus.*

In Chapter 4 I study the genetic structure of *Placostylus* snails on the Isle of Pines and use genetic data to investigate the nature of the third morphotype of snails present on the island. Using both mitochondrial DNA phylogenetic analysis and clustering analysis of neutral nuclear genetic data I find that snails of the third morphotype are not hybrids, and they do not represent a third taxon. Genetically the sample of snails with the third morphotype cannot be distinguished from morphologically identified *P. fibratus* snails on the Isle of Pines.

In Chapter 5 I generate stable isotopic data from *Placostylus* shells from New Zealand and New Caledonia. I compare the isotopic signal from snail shells to environmental isotopic signals and climate in these two regions. A high-resolution sampling protocol is developed and used to sample multiple shells of different origins. Stable isotopic variability is compared between snails from two geographical origins and between sympatric species of snail, providing important base line data for future use of these species as paleoclimatic proxies.

Table of Contents

Chapter 1: Generation of large mitochondrial and nuclear nucleotide sequences and phylogenetic analyses using high-throughput short-read datasets for endangered Placostylinae snails of the Southwest Pacific

Abstract

Placostylinae are a sub-family of terrestrial land snails endemic to the southwest Pacific. Some species are harvested for food, and other are critically endangered. Here I assemble and characterise complete mitochondrial genomes, as well as three nuclear markers (partial 45S ribosomal cassettes and the two histone genes H3 and H4) of five snail species from three geographical regions (New Zealand, New Caledonia and the Solomon Islands). These multicopy markers are excellent targets for ancient DNA studies using the abundant fossil shells of terrestrial snails. Mitogenomes of Placostylinae snails ranged between 14,544 bp and 14,711bp, with minor variation in the position of tRNA tyrosine (Y) and tRNA tryptophane (W). The 45S ribosomal cassette contained intra-genomic nucleotide variation in ITS 2. Cassettes containing histone genes H3 and H4 and their non-transcribed spacer region were assembled for three species, with the two genes coded in the same direction. Phylogenetic analysis on this large dataset (mitochondrial genome + nuclear markers) supported geographical clustering of species but could not confidently infer monophyly of the four *Placostylus* species with respect to *Eumecostylus cleryi*. Analysis based on shorter Cox1 sequences with a wider taxon sampling found species representing the genera *Eumecostylus* and *Placocharis* were nested within the phylogenetic diversity of *Placostylus.* Multi-locus phylogenetic analysis containing mitochondrial and nuclear sequences did not support Placostylinae as a monophyletic group.

Introduction

With approximately 85,000 recognised extant species, Mollusca is the second most diverse animal phylum on earth (MolluscaBase eds. 2021). Systematics within this phylum still needs improvements, and many species complexes are yet to be fully described or understood (Lydeard *et al*. 2004). To identify management units molecular tools can be used and complement morphological data (Moritz 1994). Short DNA sequences have been widely used for inferring phylogenetic relationships among molluscs in the last few decades (e.g. Wade *et al.* 2006; Herbert and Mitchell 2009; Ramirez *et al.* 2009), but the emergence of high throughput Next Generation Sequencing (NGS) technologies has now made possible analysis of significantly larger genetic data sets (Hunter *et al.* 2016). For phylogenetic inferences to be reliable genomic regions targeted by NGS technologies must be both conserved enough to allow sequence alignment and at the same time display enough nucleotide variation to permit inference of statistically robust phylogenetic relationships. Technologies such as exon-based recapture (Teasdale *et al.* 2016) or anchored phylogenomics (Lemmon and Lemmon 2012) can now be used to sample up to hundreds of genomic regions with phylogenetic signal. They are, however, associated with a high methodological cost and require either the generation of transcriptomic data in parallel to generation of genomic libraries (exon-based recapture) or the use of probes to enrich genomic datasets in regions of interest (anchored phylogenomics). One simpler phylogenomic approach is to use mitochondrial and multi-copy nuclear gene data reconstructed from genomic libraries as input for phylogenetic analysis. Both types of markers are overrepresented in shotgun sequencing genomic libraries and therefore the methodological and economical cost associated with their generation is minimal. In addition, because they are present in high number they will usually be a good target in analysis of fossil shells, which are now providing opportunities to incorporate ancient DNA into tests of evolutionary change through phylogenetic analysis (Daly *et al.* 2020; Der Sarkissian *et al*. 2020; Ferreira *et al.* 2020). Here I use high-throughput short-read sequencing to assemble whole mitochondrial genomes, 45S nuclear ribosomal DNA cassettes and the histone genes H3 and H4 of five species of Placostylinae, a sub-family of giant terrestrial land snails endemic to the south-west Pacific region. I use these data to infer phylogenetic relationships among Placostylinae species living on separate islands and other related species.

The Placostylinae belong to the superfamily Orthalicoidea (Breure *et al.* 2010; Breure and Romero, 2012) and contain among others the three genera *Placostylus* (Martin, 1784), *Eumecostylus* (Martens in Albers, 1860) and *Placocharis* (Pilsbry, 1900) that are endemic to

islands in the south-west Pacific. Most of the recognised species in these genera live on Melanesian islands from the Bismark and Solomon archipelagos to Vanuatu, New Caledonia and Fiji but also on islands further south (Lord Howe, Three Kings, and the Far North of New Zealand). Many of these snail taxa are under threat of extinction from anthropogenic habitat change and predator introduction, and some are already considered extinct (Brescia *et al.* 2008; Stringer *et al.* 2014). In many places extensive fossil deposits of *Placostylus* provide a potential resource for studies of changing environmental conditions (Brook 1999). In New Caledonia at least six species are recognised (Dowle *et al.* 2015) and they are culturally and economically important as food (Quenu *et al.* 2020). On Grande Terre itself, the main island of New Caledonia, *Placostylus* populations have declined and some species are at risk (Brescia *et al.* 2008), but on Île Des Pins (Isle of Pines) just south of Grande Terre, *Placostylus* snails are still harvested for food by indigenous Kanak people. New Zealand *Placostylus* species are considered vulnerable and today survive only in highly fragmented remnant colonies (Parrish *et al.* 1995; Daly *et al.* 2020), and Lord Howe *P. bivaricosus* is critically endangered (Ponder *et al.* 2003). On other island systems such as Fiji and the Solomon Islands Placostylinae are subject to similar threats, although information on these populations is limited (Brodie 2012).

Previous molecular work on Placostylinae has relied on application of universal PCR primers to target short fragments (<850 bp) of mitochondrial genes. Using partial mtDNA cytochrome c oxidase subunit 1 (Cox1) sequences, the Lord Howe species, *Placostylus bivaricosus* (Gaskoin, 1855), was identified as sister to the two New Zealand mainland species, *Placostylus ambagiosus* (Suter, 1906) and *Placostylus hongi* (Lesson, 1830; Ponder *et al.* 2003), while the third New Zealand species, *Placostylus bollonsii* (Suter, 1908), which is endemic to the Three Kings Islands is phylogenetically more distant. A similar result was obtained using mtDNA Cox1 and ribosomal RNA16S sequences in a study that also included representatives of the New Caledonian species *Placostylus fibratus* (Martyn, 1789), *Placostylus porphyrostomus* (Pfeiffer, 1851) and *Placostylus caledonicus* (Petit, 1845; Trewick *et al.* 2009). The New Caledonian species were inferred to be a sister group to New Zealand mainland and Lord Howe Island species. Further molecular analysis of New Caledonian *Placostylus* spp. used sequences of the faster evolving mtDNA NADH dehydrogenase 2 (ND2) gene in combination with shell shape analysis and nuclear SNP markers to explore population level diversity (Dowle *et al.* 2015). Deeper phylogenetic relationships have also been investigated within the superfamily Orthalicoidea using

concatenated short DNA sequences (Breure and Romero 2012), including data from *Placostylus ambagiosus* (New Zealand), *Placostylus eddystonensis* (Pfeiffer 1855, New Caledonia), and *Eumecostylus uligunosus* (Kobelt, 1891) and *Placocharis strangei* (Pfeiffer 1855, Solomon Islands). The 2,079 bp alignment of concatenated sequence (Cox1, ITS2/28s, Histone H3) supported the Placostylinae as a clade within the Bothriembryontidae.

Until now, no full mitochondrial genome or 45S ribosomal RNA cassette has been assembled for any Placostylinae species, and the gene arrangement of histone H3 and H4 genes is not known. Mitochondrial DNA is in high copy number and therefore is likely to be most appropriate for ancient DNA studies from fossil shells (Shtolz and Mishmar 2019). Nuclear 45S ribosomal cassettes, which include the conserved regions 18S, 5.8S, 28S and the unconstrained Internal Transcribed Spacers 1 and 2 (ITS 1 and 2), are multicopy. The ribosomal subunits 18S, 5.8S and 28S are conserved by their function in the translation of cytoplasmic mRNA of eukaryotic cells (Kressler *et al.* 2010). ITS 1 and 2 separate the different subunits within the 45S cassette and their nucleotide sequences are highly variable. The 45S ribosomal cassettes are present in numerous tandem copies usually on several chromosomes in the nuclear genome (Dalet *et al.* 2014) and are thought to evolve via concerted evolution (Nei and Rooney 2005). Intra-individual variation of nucleotide sequences is common in ITS, and can produce statistical noise when ITS sequences are used in phylogenetic or population genetic analysis (Hillis and Dixon 1991; Nei and Rooney 2005; Sochorová *et al.* 2018). The two Histone genes H3 and H4 code for protein subunits of the nucleosome complex. They are usually linked by a non-coding spacer region, and in molluscs they have been reported to be coded on opposing strands in some species (Armbruster *et al.* 2005; Harl *et al.* 2014), but not all (e.g. *Mytilus edulis* Albig *et al.* 2003). Due to their high copy number in animal cells all three loci (mitochondrial genome, 45S ribosomal cassette and Histone genes H3/H4) can be reconstructed using high-throughput short-read NGS data (Koot *et al.* 2020; Gemmell *et al.* 2020). Here I sample five species of Placostylinae: two from New Zealand (*P. ambagiosus, P. hongii*), two from New Caledonia (*P. fibratus, P. porphyrostomus*) and *Eumecostylus cleryi* (Petit de la Saussaye, 1850) from the Solomon Islands. By doing this I provide genetic tools that will be useful in future population genetic analysis of endangered Placostylinae, and in analysis of ancient DNA extracted from fossil shells. I compare phylogenetic inferences from few taxa with long sequences to phylogenetic inferences obtained using limited amounts of data but a larger taxon sample. To achieve this, three distinct phylogenetic analyses are generated. First, phylogenetic relationships between

the five taxa are estimated using large mitochondrial and nuclear sequence datasets. Second, we extract a short Cox1 sequence for the five taxa and incorporate the new sequences in phylogenetic analysis of *Placostylus* and other closely related taxa (Trewick *et al.* 2009). Third, a multi-locus phylogenetic analysis containing mitochondrial and nuclear sequences is generated to investigate the relationships of sampled Placostylinae snails in relation to other Orthalicoidea sub-families (Breure and Romero 2012).

Methods

Tissue collection

I used foot muscle tissue samples from four species of land snail of the genus *Placostylus*. The samples of the New Zealand species, *Placostylus ambagiosus* and *Placostylus hongii*, were from a frozen tissue collection originally developed from whole body samples harvested for allozyme analysis (Triggs and Sherley 1993). *Placostylus ambagiosus* (PS185; NZ National Arthropod collection code GS4 of Triggs and Sherley 1993) was collected by G. Sherley from Cape Maria van Diemen, New Zealand. The specimen of *P. hongii* (PS257; NZ National Arthropod collection code WG865) came from eastern Far North, North Island. Although the precise provenance of PS257 is not recorded it was confirmed as *P. hongii* on the basis of shell morphology and region of origin. The two New Caledonian specimens, *P. fibratus* (PS28) and *P. porphyrostomus* (PS45), were collected in the field from populations on the Isle of Pines (Dowle *et al.* 2015). The Solomon Island specimen, *Eumecostylus cleryi* (PS127), was collected in Honiara, Guadalcanal.

DNA extraction

Foot muscle samples of approximately 50 mg of were cut from specimens using sterile scalpel blades. Each tissue was pressed in a clean paper towel to remove excess storage ethanol and cut into smaller pieces. Whole genomic DNA was extracted using incubation at 55 °C in CTAB buffer (2% hexadecyltrimethyl ammonium bromide, 100 mM Tris-HCL pH 8.0 20 mM EDTA) with proteinase K (Trewick *et al.* 2009). Following tissue digestion the solution was purified using an equal volume of 24:1 chloroform-isoamyl alcohol and centrifugation. DNA was precipitated from the aqueous fraction using sodium acetate (3M NaOAC) and chilled 95% ethanol. This extraction method has been found to be the most effective for isolating high molecular weight DNA from neogastropods while avoiding the problems of mucopolysaccharide contamination (Winnepenninckx *et al.* 1993). DNA was re-suspended in 50 µl or 100 µl TE buffer (10 mM Tris, 0.1 mM EDTA) and quantified using Qubit fluorometry (Life Technologies, Thermo Fisher Scientific Inc).

Illumina sequencing

Total DNA extracts from the five Placostylinae specimens were processed through massive parallel, high-throughput sequencing using the ThruPLEX DNA-seq kit (Rubicon Genomics). Fragmented genomic DNA was pair-end sequenced on an Illumina Hiseq 2500. Reads were de-multiplexed using standard indexes. Resulting Illumina short reads were trimmed of

adapters and passed through standard quality filters using the software fastp (Chen *et al.* 2018). Reads were paired in Geneious v8 (Kearse *et al.* 2012).

Genome and gene assemblies

Mitochondrial genomes were assembled from each of five sets of 101 bp paired-read data files. Initially the mitochondrial genome of *Placostylus ambagiosus* PS185 was assembled starting with available Sanger sequenced partial sequences of *Placostylus* mtDNA cytochrome c oxidase 1 as a reference using the Geneious v8 mapping tools (Kearse *et al.* 2012). I then used iterative remapping to consensus sequences from each previous mapping round using medium-low sensitivity with 25 iterations to assemble the full circular mitochondrial genome. Once assembled, this mitochondrial genome served as the initial reference for mapping of paired-end reads of other individuals, which were then iteratively remapped until maximum read coverage was achieved.

I used the MITOS web server (Bernt *et al.* 2013) to estimate gene annotations for one of the novel mtDNA genomes and used amino acid translation tools in Geneious to verify that each protein-coding sequence had an uninterrupted translation frame and the expected start and stop codons. I used ARWEN (Laslett and Canbäck, 2008) to confirm secondary folding of transfer RNAs then transferred annotations to the other four similar genomes. Annotations were further checked for homology across the five genomes.

The same mapping approach was used to assemble the 45S nuclear ribosomal DNA cassettes of the five species, using available 5.8S and 28S sequences as our starting reference, and then mapping reads at medium-high sensitivity in Geneious (Kearse *et al.* 2012).

Fragments of histone 3 (H3) were also reconstructed by mapping reads of the five Placostylinae species onto reference sequences. Five 267 bp H3 fragments were reconstructed based on GenBank data from the Placostylinae species *Placocharis strangei* (JF514684, Breure and Romero 2012) and five 291 bp H4 fragments were reconstructed based on GenBank data from *Orcula dolium* (KY512728, Harl *et al.* 2014). To reconstruct whole histone H3/H4 complexes (containing their noncoding spacer region), we first used one of the H4 fragments as a reference (PS127, *Eumecostylus cleryi*) and iteratively mapped corresponding reads at low sensitivity until the H3 fragment of the same individual could be mapped to a reference sequence. I then mapped paired-end reads of other specimens to the PS127 consensus sequence, but only reads of the New Caledonian species *P. fibratus* and *P.*

porphyrostomus mapped continuously to the reference sequence. I identified coding direction of the histone genes using the Geneious amino acid translation tool.

Phylogenetic analysis

Mitogenome and combined nuclear markers analysis of five Placostylinae

Phylogenetic analysis was performed for mitochondrial genomes and nuclear markers separately. Whole mitochondrial genomes, ribosomal cassettes and histone genes were aligned using the software MUSCLE (Edgar 2004), with default options. Alignments were then checked visually and regions with ambiguous alignment were manually removed. For mitochondrial genome phylogenies I concatenated all coding DNA sequence (CDS) regions, and all the tRNA regions which are likely to hold phylogenetic signal (total length of this alignment was 10, 981 bp). An additional phylogenetic analysis using only the CDS regions was also performed. I used BLAST (Madden 2013) to find the closest mitogenome to our dataset on the GenBank database (*Naesiotus nux*) and used it as an outgroup in the mitochondrial phylogeny (Hunter *et al.* 2016). Nuclear markers (45S cassette, H3, H4) were also concatenated after aligning each marker separately (total length of the alignment was 8,216 bp). Nuclear sequences for closely related species were not found on GenBank and therefore nuclear analysis had to be run without an outgroup. Phylogenetic relationships among taxa were estimated using both Maximum Likelihood (ML) and Bayesian Inference (BI) approaches. PartitionFinder2 (Lanfear *et al.* 2017) was used to find the most appropriate models of nucleotide evolution, using the corrected Akaike Information Criterion (AICc) as a model evaluation metric. I used PhyML for the ML approach (Guindon *et al.* 2010) and MrBayes 3.0 for BI phylogenies (Huelsenbeck and Ronquist 2001). Bayesian analyses were run on the CIPRES science gateway 3.3 server (Miller *et al.* 2010), taking advantage of cloud computing processing power. For mitochondrial analyses I used a partition of three nucleotide models for the CDS concatenated region (GTR + gamma; GTR + I + gamma and GTR + gamma), and one model for the tRNA region (GTR + gamma). Only one model of nucleotide substitution was needed for the phylogenetic analysis of nuclear markers (GTR + I + gamma).

Mitochondrial Cox1 analysis of 10 Placostylinae

I extracted a 538bp Cox1 fragment from our five mitochondrial genomes to produce phylogenetic trees containing a wider sampling of Placostylinae, based on much shorter DNA sequences. Sequences were extracted from GenBank for *Placostylus* species of New Zealand,

New Caledonia, Three Kings Islands and Lord Howe Island (Trewick *et al.* 2009) and Placostylinae species of the Solomon Islands (Breure and Romero 2012). Total alignment included 10 putative species and 71 individuals. The same methods were applied for this phylogenetic analysis, I used both ML and BI approaches with only one model of nucleotide substitution (GTR + I + gamma).

Multilocus analysis of multiple Orthalicoidea species

Finally, to assess phylogenetic relationships of the five Placostylinae species within the context of the wider Orthalicoidea superfamily I also performed a phylogenetic analysis using concatenated multi-locus alignment of Cox1 sequences, partial histone H3 sequences and partial 28S sequences (Table 1, Breure and Romero 2012). I used 13 sequences of total 1,730 bp length for this analysis, using sequences of *Megaspira eliator* as an outgroup, and only one model of nucleotide substitution (GTR + I + gamma).

Table 1. GenBank accession number and collection information of all sequences used in phylogenetic analysis in this study.

species name	Locality	Collector	Mt genome	45S	Cox1	H3	H4	Partial 28S
Naesiotus nux	San Cristobal Island, Galapagos	S. Hunter	KT821554.1		KT821554.1			
Placostylus fibratus	Isle of Pines, New Caledonia	F. Brescia	MT163270	MN567952	MT163270	MT559980	MT602525	MN567952
Placostylus porphyrostomus	Isle of Pines, New Caledonia	F. Brescia	MT163271	MN567955	MT163271	MT559981	MT602524	MN567955
Placostylus hongii	Far North, New Zealand	G. Sherley	MT163273	MN567954	MT163273	MT559984	MT602521	MN567954
Placostylus ambagiosus	Far North, New Zealand	G. Sherley	MT163272	MN567951	MT163272	MT559983	MT602522	MN567951
Eumecostylus cleryi	Guadalcanal, Solomon Islands	R. Richards	MT163274	MN567953	MT163274	MT559982	MT602523	MN567953
Eumecostylus uligunosus	Rokera, Solomon Islands	A. Delsaerdt			JF514642	JF514685		HM027505
Bothriembryon dux	Mt Caitlin, Western Australia				JF514643	JF514686		HM027490
Botrhriembryon indutus	Walyunga national park, Australia	C. Whisson						EU622023
Prestonella Bowkeri	Glen Avon, South Africa	D. Herbert			KF129392	JF514711		EU622021
Prestonella nuptialis	Craddock area, South Africa	D. Herbert			KF129349			EU622022
Discoleus aguirrei	Rio Negro, Argentina	M. Guezzo			KT371414			KT371389
Discoleus ameghinoi	Rio Negro, Argentina	M. Guezzo			KT371415	JF514698		JF514753
Placocharis strangei	New Georgia, Solomon Islands	A. Delsaerdt			JF514641	JF514684		HM027504
Megaspira eliator	Rio de Janeiro, Brasil	A. Galdino dos Santos			JF514610	JF514715		JF514721
Placostylus eddystonensis	Mount Koghis, New Caledonia	C. Wade						AY841297

Results and discussion

Mitochondrial genomes and nuclear loci characteristics

The five mitochondrial genomes range from 14,652 bp to 15,187 bp (Table 2). All contain the expected 37 genes found in most metazoans, including 13 protein-coding genes, 2 rRNAs and 22 tRNAs genes. Of these genes 24 are encoded on the heavy strand and 13 on the light strand. Nucleotide composition is skewed towards a high proportion of adenine and thymine, which is a common observation for invertebrate mitochondrial genomes (Table 2; Shioiri and Takahata 2001). The size of the noncoding region varies from 539 bp to 1,577 bp, i.e. 3.7–10.8% of the total mitochondrial genome. Arrangement of mitochondrial genes remains mostly unchanged across sampled taxa, with only the respective position of two tRNA genes differing among species (Figure 1). In New Caledonian and Solomon Island species tRNA tyrosine (Y) is adjacent to Cox2 and followed by tRNA tryptophan (W), whereas in New Zealand species these two tRNAs are swapped and it is tRNA W that is adjacent to Cox2. Start codons for protein coding genes include the common ATG, ATA and ATT but also TTG and GTG. Start codons for invertebrate mitochondrial protein coding genes are quite variable (Gaitán-Espitia *et al.* 2013; Shen *et al.* 2012) and within our dataset there is evidence of start codon mutations for the genes ND5, ND4L, ND3, ND4, Cox3 and ND2 (Table 2). Stop codons include the common TAA and T. I noticed the hardest region to reconstruct was ATP8, which also seems to be the most variable protein coding gene of the mitochondrial genome. Length of this gene is extremely reduced in all five Placostylinae species (around 200 bp; Table 3). This observation suggests that ATP8 could be under relaxed selection in Placostylinae, which could be linked to the low mobility of the snails (Sun *et al.* 2017).

Table 2. Summary statistics of five reconstructed mitochondrial genomes of Placostylinae species, assembled using 101 bp paired reads. *P.* = *Placostylus*, *E.* = *Eumecostylus*

	P. ambagiosus (PS185)	*P. hongii* (PS257)	*P. fibratus* (PS28)	*P. porphyrostomus* (PS45)	*E. cleryi* (PS127)
Size (bp)	14,652	14,711	15,187	15,118	14,737
A+T proportion	0.76	0.76	0.73	0.73	0.75
Noncoding region (bp) and (%)	539 (3.7%)	900 (6.1%)	1,041 (7.2%)	1,026 (7.0%)	1,577 (10.8%)
Genes	37	37	37	37	37
Gene regions (bp)	14208	14490	13657	13837	13518
tRNA	22	22	22	22	22

Table 3. Position and length of the 13 protein coding genes and 2 rRNA genes found across five mitochondrial genomes of Placostylinae species. For protein coding genes, the nature of start and stop codons is indicated. Gene abbreviations Cox: Cytochrome oxidase, NADH: NADH dehydrogenase, Cyt: cytochrome reductase, ATP: ATP synthase. PS28: *Placostylus fibratus*, PS45: *Placostylus porphyrostomus*, PS127: *Eumecostylus cleryi*, PS185: *Placostylus ambagiosus*, PS257: *Placostylus hongii*

Name	Type	Direction	Position					Length				
			PS127	PS45	PS28	PS185	PS257	PS127	PS45	PS28	PS185	PS257
Cox1	CDS	f	1->1542 TTG/TAA	1->1494 TTG/TAG	1->1488 TTG/TAG	1->1533 TTG/TAA	1->1533 TTG/TAA	1542	1494	1488	1533	1533
16S	rRNA	f	1743->2665	1740->2658	1739->2660	1649->2684	1652->2685	923	919	922	1036	1034
Nadh6	CDS	f	2912->3388 ATG/TAG	2915->3394 ATG/TAA	2914->3393 ATG/TAA	2953->3433 ATG/TAG	2965->3450 ATG/TAA	477	480	480	481	486
Nadh5	CDS	f	3414->5057 GTG/TAG	3417->5057 GTG/TAG	3416->5056 ?/TAG	3455->5104 GTG/TAG	3470->5119 ATG/TAG	1644	1641	1641	1650	1650
Nadh1	CDS	f	5044->5964 ATG/TAA	5044->5964 ATG/TAG	5043->5963 ATG/TAG	5091->6008 ATG/TAA	5106->6023 ATG/TAA	921	921	921	918	918
Nadh4L	CDS	f	5972->6268 ATA/TAA	5972->6271 ATG/TAG	5971->6270 GTG/TAG	6016->6313 ATA/T	6031->6328 ATA/?	297	300	300	298	298
CytB	CDS	f	6276->7361 ATA/TAA	6276->7367 ATA/TAA	6275->7366 ATA/TAA	6320->7402 ATA/TAA	6332->7417 ATA/TAA	1086	1092	1092	1083	1086
Cox2	CDS	f	7599->8260 ATG/TA	7588->8250 ATG/TAA	7588->8250 ATG/TAA	7637->8296 ATG/TAA	7662->8327 ATG/TAA	662	663	663	660	666

ATP8	CDS	r	8920->8747 ATG/TAA	9499->9293 ATG/TAA	9497->9291 ?/TAA	8987->8742 ?/TAA	9053->8802 ?/TAA	174	207	207	246	252
ATP6	CDS	r	9661->9008 ATG/TAA	10239->9586 ATG/TAA	10241->9588 ATG/TAA	9781->9128 ATG/TAA	9851->9198 ATG/TAA	654	654	654	654	654
12S	rRNA	r	10627->9879	11237->10478	11228->10471	10744->9979	10832->10048	749	760	758	766	1519
NadH3	CDS	r	11044->10702 ATG/T	11649->11307 ATG/T	11640->11298 ATG/T	11162->10818 ATA/ATT	11242->10903 ATA/T	343	343	343	345	340
NadH4	CDS	f	11174->12487 ATG/TAA	11795->13108 ATG/TAA	11786->13099 GTG/TAA	11298->12608 ATG/TAA	11382->12692 ATG/TAA	1314	1314	1314	1311	1311
Cox3	CDS	r	13409->12561 ATG/TAG	14022->13183 ATG/TAG	14013->13174 ATG/TAG	13538->12687 ATT/TAG	13601->12756 ATT/T	849	840	840	852	846
NadH2	CDS	f	13552->14478 ATG/TAG	14165->15088 ATG/TAG	14157->15080 ATA/TAG	13628->14614 ATT/TAA	13690->14673 ATG/TAA	927	924	924	987	984

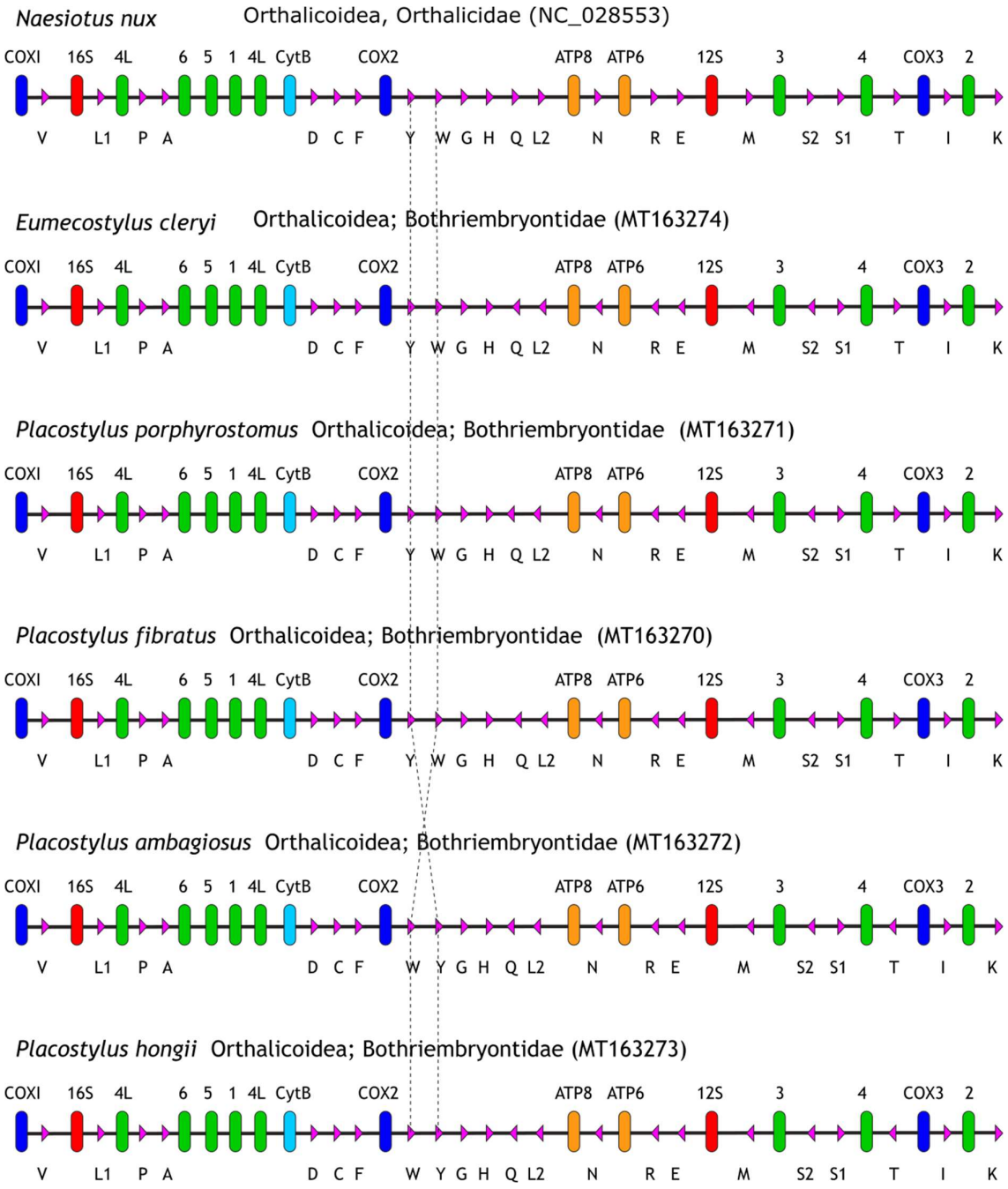


Figure 1. Comparison of mitochondrial genome gene order in five Placostylinae snail species. Full name of tRNA and CDS genes can be found in table 3.

The 45S region contained the five expected components of the transcription unit: the three RNA coding regions 18S, 5.8S and 28S and two Internal Transcribed Spacers (ITS1 and ITS2). Mapped reads of the 45S ribosomal cassette displayed evidence of intra-genomic variation at some nucleotide positions of the ITS1 and ITS2 regions in all five sampled

species genomes. This pattern of intraspecific variation has been reported for this marker in a number of other Gastropoda, and is widespread in metazoans (Stothard *et al.* 1996; Hoy and Rodriguez 2013; Davison 2002; Harris and Crandall 2000; Itskovich 2020). Relaxed selection combined with independent mutations in different sets of copies could lead to this pattern in ITS, while constraining selection minimises intragenomic nucleotide variation in the transcribed 18S, 5.8S and 28S regions (Pereira and Baldwin 2016). This needs to be taken into consideration when using ITS 1 and ITS 2 in phylogenetic or population genetic analysis, as intra-genomic nucleotide variation could potentially confound conclusions reached from comparative ITS analyses. Sequencing methods based on polymerase chain reaction amplification are more likely to yield invariant sequences whereas high throughput sequencing followed by mapping is more likely to make read variants apparent.

Histone genes H3 and H4 were reconstructed for the five Placostylinae species based on GenBank fragments from related species. Whole histone cassettes containing histone genes H3 and H4 and their noncoding spacer region could only be reconstructed for the Solomon Island species *E. cleryi* (GenBank accession number: MT726982) and the two New Caledonian species *P. fibratus* and *P. porphyrostomus* (GenBank accession numbers: MT726983 and MT726984). In these cassettes both genes were coded on the same strand, and were separated by a noncoding spacer region of 707 bp (*P. porphyrostomus*), 772 bp (*E. cleryi*) and 800 bp (*P. fibratus*). Histone genes H3 and H4 have been reported to be orientated in opposing directions in gastropod genomes (Armbruster *et al.* 2005; Harl *et al.* 2014), but other configurations have been reported in bivalves (Albig *et al.* 2003). Our findings confirm that histone gene cassettes are not configured consistently in gastropods. For the two New Zealand species, *P. ambagiosus* and *P. hongii*, the configuration of histone genes could not be retrieved from our data. Mapping reads from New Zealand species to consensus sequences of other species only led to discontinuous mapping reconstructions, with gaps and poor mapping resolution in the noncoding spacer region. It is hard to know if this result reflects read limitation in our data or indicates real biological information. If it were real, it would imply that histone genes H3 and H4 have physically distinct genome locations in *P. ambagiosus* and *P. hongi* rather than being organised in cassettes.

Phylogenies

Mitogenome and combined nuclear markers analysis of five Placostylinae

Phylogenetic relationships inferred from both mitochondrial and nuclear DNA sequences were similar and suggest geographical clustering of *Placostylus* species, with a sister relationship between the two New Zealand species, and between the two New Caledonian species (Figure 2). Using another bulimulid species, *Naesiotus nux*, as an outgroup, *Eumecostylus cleryi* (Solomon Islands) was placed in a sister relationship to the New Caledonian clade in the mitochondrial phylogeny (Figure 2). However, this branch was not strongly supported by ML bootstrapping (71) and BI posterior probability varied when incorporating tRNAs in the analysis (0.9675 in the phylogeny with both CDS and some tRNA regions; but only 0.8292 in a phylogeny using just CDS – tree not shown). Inferences of phylogenetic relationship among snail species from different geographical regions is therefore ambiguous with this dataset. Given the amount of data used in this analysis it is likely these uncertainties rise from either conflicting information in the different regions of the mitochondrial genome, and/or from insufficient taxon sampling (Jantzen *et al.* 2019).

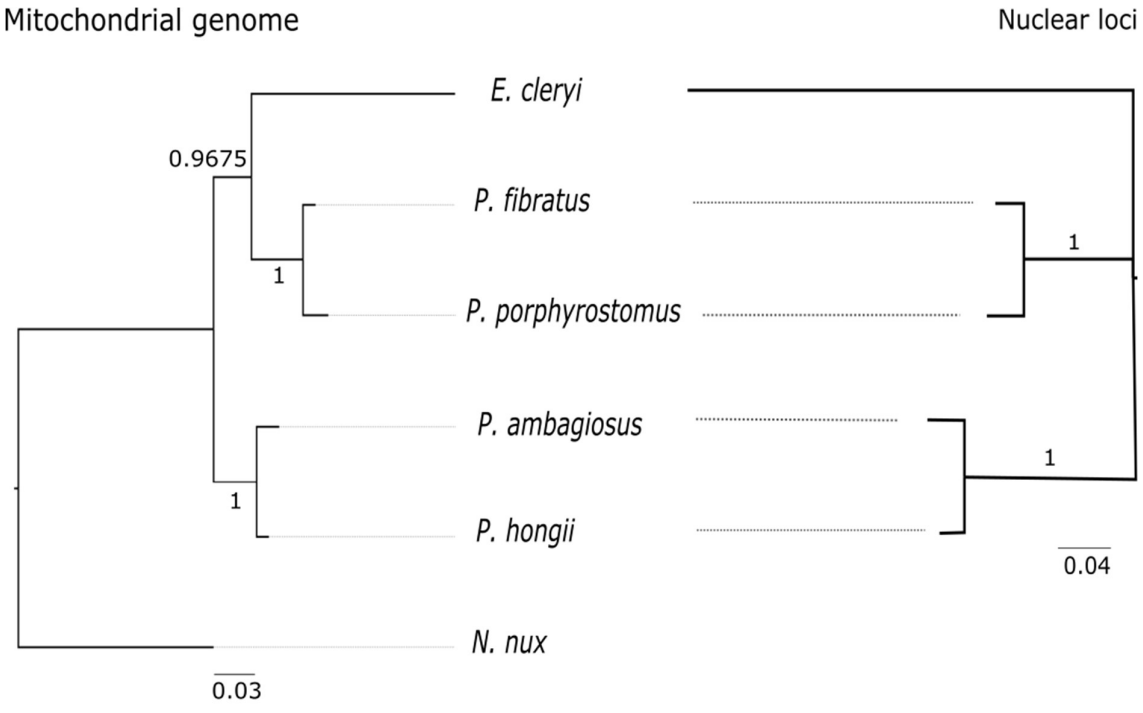


Figure 2. Phylogenetic relationships inferred for five giant terrestrial snails of the Pacific. The shells of four *Placostylus* and one *Eumecostylus* species are shown, and *Naesiotus nux* (Bulimulidae) was used as the outgroup in the mitochondrial. *Placostylus fibratus* and *P.*

porphyrosotmus are from New Caledonia, *P. ambagiosus* and *P. hongii* are from New Zealand and *E. cleryi* is from the Solomon Islands. The tree on the left is based on whole mitochondrial genome DNA sequences (10,981 bp), that on the right is based on nuclear loci (ribosomal cassette 45S + histone genes H3/H4; 8,216 bp). Both were inferred using a Bayesian phylogenetic approach.

Mitochondrial Cox1 analysis of 10 Placostylinae

Phylogenies based on short Cox1 sequences grouped *E. cleryi* and other snails from the Solomon Islands in a clade sister to *Placostylus bivaricosus* (Lord Howe Island) with strong support (BI posterior probability = 1; Figure 3). Sister to this group is a monophyletic clade containing the *Placostylus* species from New Zealand, New Caledonia and Three Kings Islands. *Placostylus bollonsi*, the species from the Three Kings Islands, appears more closely related to species from New Caledonia than species from New Zealand. This is an odd result given the geographical setup of those islands (Three King Islands are 55 km north of New Zealand), but similar results have been found based on allozyme data (Triggs and Sherley 1993). Based on these data, the nesting of *Eumecostylus* and *Placocharis* within *Placostylus* suggests taxonomic redundancy, and using only one genus name for all species would seem more appropriate.

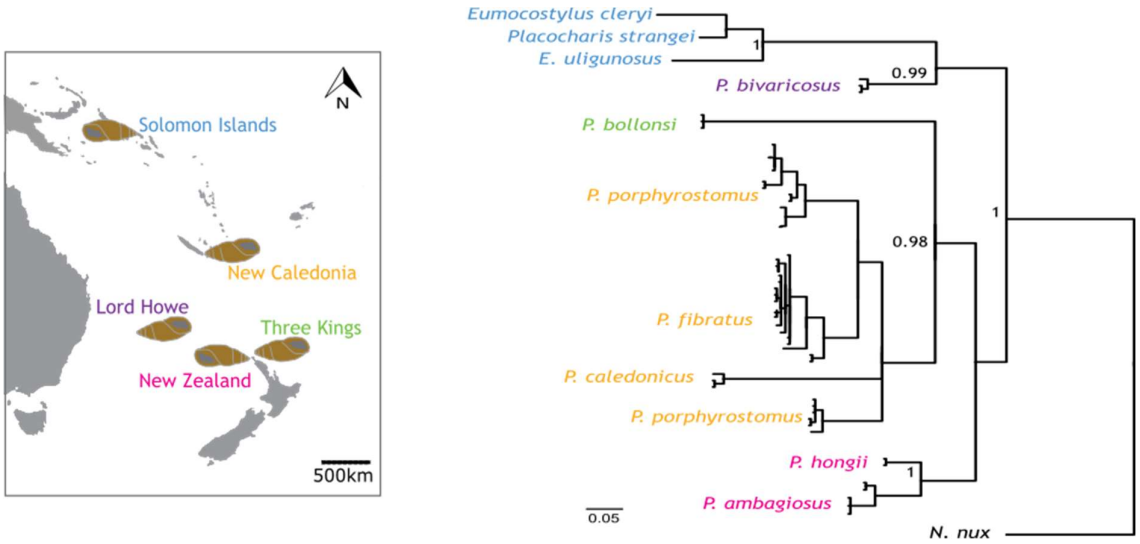


Figure 3. Phylogenetic relationships of 10 putative Placostylinae snail species from the southwest Pacific, inferred using a Bayesian phylogeny of partial mtDNA Cox1 (538 bp). Numbers at nodes refer to posterior probabilities indicating the level of support for clades. Species are colour coded on the map and phylogeny according to their geographical origin.

Multilocus analysis of multiple Orthalicoidea species

Multi-locus analysis involving both mitochondrial (Cox1) and nuclear (H3, 28S) sequences did not support the monophyly of the Placostylinae (Figure 4). Some Placostylinae species from the Solomon Islands (*Placocharis strangei*, *Eumecostylus uligunosus*, *Eumecostylus cleryi*) were grouped with species from other sub-families (*Bothriembryon dux*, *Discoleus ameghinoi*, *Prestonella nuptialis*, *Prestonella Bowkeri*, *Plecostylus peruvianus*), but support for the corresponding branch is low (BI posterior probability 0.87). This differs from previous findings in this group, which used the same combination of loci, but fewer taxa (Breure and Romero 2012).

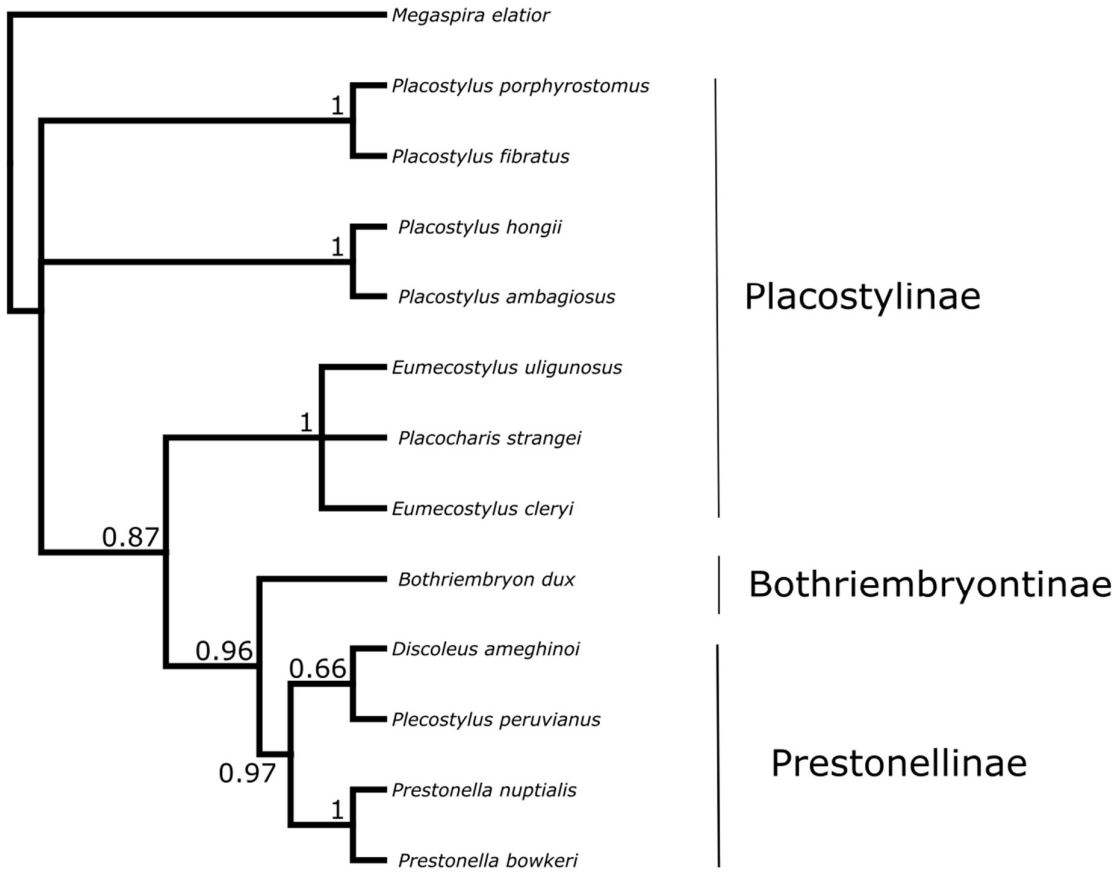


Figure 4. Multi-locus Bayesian phylogenetic relationships of 13 species of Orthalicoidea. The phylogeny is based on 1,730 bp nucleotide alignment including concatenated sequences from Cox1, H3 and 28S loci.

Overall, all phylogenetic relationships reported here supported monophyly within geographical regions (or islands), but relationships among species in the wider Pacific rarely received strong branch support. Furthermore, phylogenetic relationships among the five Placostylinae species were different in each analysis. Long branch attraction caused by rapid molecular evolution within the Solomon Islands lineage and a relatively deep internode could be the source of this phylogenetic incongruence. Sampling of large NGS datasets for other species of Placostylinae will be needed to confidently infer phylogenetic relationships among these species (Uribe *et al.* 2019). From a biological perspective, dispersal from island to island is thought to be the main evolutionary process explaining the current distribution of Placostylinae species in the southwest Pacific (Trewick *et al.* 2009). This scenario provides no prediction of phylogenetic relationships as dispersal events to new locations could have happened in any order, and there is no limit to additional migration events that would overlay phylogenetic signal.

In conclusion, I provide full DNA sequences for the mitochondrial genome, and the nuclear 45S cassette, and Histone 3 and 4 for five giant land snails. These data provide the opportunity for developing better conservation genetic markers for a group of molluscs that are particularly vulnerable to human disturbance. In addition, because these genetic markers exist in multiply copies per cell, they provide an excellent resource for future study of aDNA from fossil snail shells.

References

Albig, W., Warthorst, U., Drabent, B., Prats, E., Cornudella, L., & Doenecke, D. (2003). *Mytilus edulis* core histone genes are organized in two clusters devoid of linker histone genes. *Journal of Molecular Evolution*, *56*(5), 597–606. https://doi.org/10.1007/s00239-002-2428-8

Armbruster, G. F. J., Böhme, M., Bernhard, D., & Schlegel, M. (2005). The H3/H4 histone gene cluster of land snails (Gastropoda: Stylommatophora): TS/TV Ratio, GC3 drive and signals in stylommatophoran phylogeny. *Journal of Molluscan Studies*, *71*(4), 339–348. https://doi.org/10.1093/mollus/eyi038

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., Pütz, J., Middendorf, M., & Stadler, P. F. (2013). MITOS: Improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution*, *69*(2), 313–319. https://doi.org/10.1016/j.ympev.2012.08.023

Brescia, F. M., Pöllabauer, C. M., Potter, M. A., & Robertson, A. W. (2008). A review of the ecology and conservation of *Placostylus* (Mollusca: Gastropoda: Bulimulidae) in New Caledonia. *Molluscan Research*, *28*(2), 111–122.

Breure, A. S. H., Groenenberg, D. S. J., & Schilthuizen, M. (2010). New insights in the phylogenetic relationswithin the Orthalicoidea (Gastropoda, Stylommatophora) based on 28S sequence data. *Basteria*, *74*((1-3)), 25–32.

Breure, A. S. H., & Romero, P. E. (2012). Support and surprises: Molecular phylogeny of the land snail superfamily Orthalicoidea using a three-locus gene analysis with a divergence time analysis and ancestral area reconstruction (Gastropoda: Stylommatophora). *Archiv Fur Molluskenkunde*, *141*(1), 1–20. https://doi.org/10.1127/arch.moll/1869-0963/141/001-020

Brodie, G. 2012. (2012). *Placostylus kantavuensis*. *The IUCN Red List of Threatened Species*. https://dx.doi.org/10.2305/IUCN.UK.2012.RLTS.T195580A2389624.en

Brook, F. J. (1999). Stratigraphy, landsnail faunas, and paleoenvironmental history of coastal dunefields at te werahi, northernmost new zealand. *Journal of the Royal Society of New Zealand*, *29*(4), 361–393. https://doi.org/10.1080/03014223.1999.9517603

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*(17), i884–i890.

https://doi.org/10.1093/bioinformatics/bty560

Dalet, J. T., Saloma, C. P., Olivera, B. M., & Heralde, F. M. (2014). Karyological analysis and FISH physical mapping of 18S rDNA genes, (GATA)n centromeric and (TTAGGG)n telomeric sequences in Conus magus Linnaeus, 1758. *Journal of Molluscan Studies*, *81*(2), 274–289. https://doi.org/10.1093/mollus/eyu090

Daly, E. E., Trewick, S. A., Dowle, E. J., Crampton, J. S., & Richards, M. M. (2020). Conservation of pupu whakarongotaua the snail that listens for the war party. *Ethnobiology and Conservation*, *9*. https://doi.org/10.15451/EC2020-05-9.13-1-27

Davison, A. (2002). Land snails as a model to understand the role of history and selection in the origins of biodiversity. *Population Ecology*, *44*(3), 129–136. https://doi.org/10.1007/s101440200016

Der Sarkissian, C., Möller, P., Hofman, C. A., Ilsøe, P., Rick, T. C., Schiøtte, T., Sørensen, M. V., Dalén, L., & Orlando, L. (2020). Unveiling the Ecological Applications of Ancient DNA From Mollusk Shells. *Frontiers in Ecology and Evolution*. https://doi.org/10.3389/fevo.2020.00037

Dowle, E. J., Morgan-Richards, M., Brescia, F., & Trewick, S. A. (2015). Correlation between shell phenotype and local environment suggests a role for natural selection in the evolution of Placostylus snails. *Molecular Ecology*, *24*(16), 4205–4221. https://doi.org/10.1111/mec.13302

Edgar, R. C. (2004). MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, *5*. https://doi.org/10.1186/1471-2105-5-113

Ferreira, S., Ashby, R., Jeunen, G. J., Rutherford, K., Collins, C., Todd, E. V., & Gemmell, N. J. (2020). DNA from mollusc shell: A valuable and underutilised substrate for genetic analyses. *PeerJ*, *8*. https://doi.org/10.7717/peerj.9420

Gaitán-Espitia, J. D., Nespolo, R. F., & Opazo, J. C. (2013). The Complete Mitochondrial Genome of the Land Snail *Cornu aspersum* (Helicidae: Mollusca): Intra-Specific Divergence of Protein-Coding Genes and Phylogenetic Considerations within Euthyneura. *PLoS ONE*, *8*(6). https://doi.org/10.1371/journal.pone.0067299

Gemmell, M. R., Trewick, S. A., Hills, S. F. K., & Morgan-Richards, M. (2020). Phylogenetic topology and timing of New Zealand olive shells are consistent with punctuated equilibrium. *Journal of Zoological Systematics and Evolutionary Research*,

58(1), 209–220. https://doi.org/10.1111/jzs.12342

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Systematic Biology*, *59*(3), 307–321. https://doi.org/10.1093/sysbio/syq010

Harl, J., Duda, M., Kruckenhauser, L., Sattmann, H., & Haring, E. (2014). In search of glacial refuges of the land snail *Orcula dolium* (Pulmonata, Orculidae) - An integrative approach using DNA sequence and fossil data. *PLoS ONE*, *9*(5). https://doi.org/10.1371/journal.pone.0096012

Harris, D. J., & Crandall, K. A. (2000). Intragenomic variation within ITS1 and ITS2 of freshwater crayfishes (Decapoda: Cambaridae): Implications for phylogenetic and microsatellite studies. *Molecular Biology and Evolution*, *17*(2), 284–291. https://doi.org/10.1093/oxfordjournals.molbev.a026308

Herbert, D. G., & Mitchell, A. (2009). Phylogenetic relationships of the enigmatic land snail genus Prestonella: The missing African element in the Gondwanan superfamily Orthalicoidea (Mollusca: Stylommatophora). *Biological Journal of the Linnean Society*, *96*(1), 203–221. https://doi.org/10.1111/j.1095-8312.2008.01109.x

Hillis, D. M., & Dixon, M. T. (1991). Ribosomal DNA: Molecular evolution and phylogenetic inference. *Quarterly Review of Biology*, *66*(4), 411–446. https://doi.org/10.1086/417338

Hoy, M. S., & Rodriguez, R. J. (2013). Intragenomic sequence variation at the ITS1-ITS2 region and at the 18S and 28S nuclear ribosomal DNA genes of the New Zealand mud snail, *Potamopyrgus antipodarum* (Hydrobiidae: Mollusca). *Journal of Molluscan Studies*. https://doi.org/10.1093/mollus/eyt016

Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, *17*(8), 754–755. https://doi.org/10.1093/bioinformatics/17.8.754

Hunter, S. S., Settles, M. L., New, D. D., Parent, C. E., & Gerritsen, A. T. (2016). Mitochondrial genome sequence of the Galápagos endemic land snail Naesiotus nux. *Genome Announcements*, *4*(1). https://doi.org/10.1128/genomeA.01362-15

Itskovich, V. (2020). Intragenomic variation of rDNA internal transcribed spacers in the endemic Baikal sponge Lubomirskia baikalensis (Pallas, 1776) (Spongillida,

Lubomirskiidae): Implications for Porifera barcoding. *Journal of Great Lakes Research*, *46*(1), 62–66. https://doi.org/10.1016/j.jglr.2019.10.009

Jantzen, J. R., Whitten, W. M., Neubig, K. M., Majure, L. C., Soltis, D. E., & Soltis, P. S. (2019). Effects of taxon sampling and tree reconstruction methods on phylodiversity metrics. *Ecology and Evolution*. https://doi.org/10.1002/ece3.5425

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., & Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), 1647–1649. https://doi.org/10.1093/bioinformatics/bts199

Koot, E. M., Morgan-Richards, M., & Trewick, S. A. (2020). An alpine grasshopper radiation older than the mountains, on Kā Tiritiri o te Moana (Southern Alps) of Aotearoa (New Zealand). *Molecular Phylogenetics and Evolution*, *147*. https://doi.org/10.1016/j.ympev.2020.106783

Kressler, D., Hurt, E., & Baßler, J. (2010). Driving ribosome assembly. In *Biochimica et Biophysica Acta - Molecular Cell Research* (Vol. 1803, Issue 6, pp. 673–683). https://doi.org/10.1016/j.bbamcr.2009.10.009

Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., & Calcott, B. (2017). Partitionfinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution*, *34*(3), 772–773. https://doi.org/10.1093/molbev/msw260

Laslett, D., & Canbäck, B. (2008). ARWEN: A program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. *Bioinformatics*, *24*(2), 172–175. https://doi.org/10.1093/bioinformatics/btm573

Lemmon, A. R., & Lemmon, E. M. (2012). High-throughput identification of informative nuclear loci for shallow-scale phylogenetics and phylogeography. *Systematic Biology*. https://doi.org/10.1093/sysbio/sys051

Lydeard, C., Cowie, R. H., Ponder, W. F., Bogan, A. E., Bouchet, P., Clark, S. A., Cummings, K. S., Frest, T. J., Gargominy, O., Herbert, D. G., Hershler, R., Perez, K. E., Roth, B., Seddon, M., Strong, E. E., & Thompson, F. G. (2004). The global decline of nonmarine mollusks. *BioScience*, *54*(4), 321–330. https://doi.org/10.1641/0006-

3568(2004)054[0321:TGDONM]2.0.CO;2

Madden, T. (2013). The BLAST sequence analysis tool. *The BLAST Sequence Analysis Tool*.

Miller, M. A., Pfeiffer, W., & Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *2010 Gateway Computing Environments Workshop, GCE 2010*. https://doi.org/10.1109/GCE.2010.5676129

Moritz, C. (1994). Defining "Evolutionarily Significant Units" for conservation. *Trends in Ecology and Evolution*, *9*(10), 373–375. https://doi.org/10.1016/0169-5347(94)90057-4

Nei, M., & Rooney, A. P. (2005). Concerted and birth-and-death evolution of multigene families. *Annual Review of Genetics*, *39*, 121–152. https://doi.org/10.1146/annurev.genet.39.073003.112240

Parrish, R., Sherley Greg, & Aviss, M. (1995). Giant land snail recovery plan, Placostylus spp., Paryphanta sp. In *Threatened Species Recovery Plan Series*.

Pereira, T. J., & Baldwin, J. G. (2016). Contrasting evolutionary patterns of 28S and ITS rRNA genes reveal high intragenomic variation in Cephalenchus (Nematoda): Implications for species delimitation. *Molecular Phylogenetics and Evolution*. https://doi.org/10.1016/j.ympev.2016.02.016

Ponder, W. F., Colgan, D. J., Gleeson, D. M., & Sherley, G. H. (2003). Relationships of Placostylus from Lord Howe Island: An investigation using the mitochondrial cytochrome c oxidase 1 gene. *Molluscan Research*, *23*(2), 159–178. https://doi.org/10.1071/MR03001

Quenu, M., Trewick, S. A., Brescia, F., & Morgan-Richards, M. (2020). Geometric morphometrics and machine learning challenge currently accepted species limits of the land snail *Placostylus* (Pulmonata: Bothriembryontidae) on the Isle of Pines, New Caledonia. *Journal of Molluscan Studies*, *86*(1), 35–41. https://doi.org/10.1093/mollus/eyz031

Ramirez, J., Ramírez, R., Romero, P., Chumbe, A., & Ramírez, P. (2009). Posición evolutiva de caracoles terrestres peruanos (Orthalicidae) entre los Stylommatophora (Mollusca: Gastropoda). *Revista Peruana de Biologia*, *16*(1), 51–56. https://doi.org/10.15381/rpb.v16i1.193

Shen, X., Tian, M., Meng, X., Liu, H., Cheng, H., Zhu, C., & Zhao, F. (2012). Complete mitochondrial genome of *Membranipora grandicella* (Bryozoa: Cheilostomatida)

determined with next-generation sequencing: The first representative of the suborder Malacostegina. *Comparative Biochemistry and Physiology - Part D: Genomics and Proteomics*, *7*(3), 248–253. https://doi.org/10.1016/j.cbd.2012.03.003

Shioiri, C., & Takahata, N. (2001). Skew of mononucleotide frequencies, relative abundance of dinucleotides, and DNA strand asymmetry. *Journal of Molecular Evolution*, *53*(4–5), 364–376. https://doi.org/10.1007/s002390010226

Shtolz, N., & Mishmar, D. (2019). The Mitochondrial Genome–on Selective Constraints and Signatures at the Organism, Cell, and Single Mitochondrion Levels. In *Frontiers in Ecology and Evolution* (Vol. 7). https://doi.org/10.3389/fevo.2019.00342

Sochorová, J., Garcia, S., Gálvez, F., Symonová, R., & Kovařík, A. (2018). Evolutionary trends in animal ribosomal DNA loci: introduction to a new online database. *Chromosoma*, *127*(1), 141–150. https://doi.org/10.1007/s00412-017-0651-8

Stothard, J. R., Hughes, S., & Rollinson, D. (1996). Variation within the Internal Transcribed Spacer (ITS) of ribosomal DNA genes of intermediate snail hosts within the genus Bulinus (Gastropoda: Planorbidae). *Acta Tropica*, *61*(1), 19–29. https://doi.org/10.1016/0001-706X(95)00137-4

Stringer, I. A. N., Parrish, G. R., Sherley, G. H., & MacKenzie, D. I. (2014). The biology of *Placostylus ambagiosus* (Pulmonata: Bulimulidae) in New Zealand: Part 2. Population changes, growth, mortality and life expectancy. *Molluscan Research*, *34*(3), 155–175. https://doi.org/10.1080/13235818.2014.888985

Sun, S., Li, Q., Kong, L., & Yu, H. (2017). Limited locomotive ability relaxed selective constraints on molluscs mitochondrial genomes. *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-017-11117-z

Teasdale, L. C., Köhler, F., Murray, K. D., O'Hara, T., & Moussalli, A. (2016). Identification and qualification of 500 nuclear, single-copy, orthologous genes for the Eupulmonata (Gastropoda) using transcriptome sequencing and exon capture. *Molecular Ecology Resources*. https://doi.org/10.1111/1755-0998.12552

Thomaz, D., Guiller, A., & Clarke, B. (1996). Extreme divergence of mitochondrial DNA within species of pulmonate land snails. *Proceedings of the Royal Society B: Biological Sciences*. https://doi.org/10.1098/rspb.1996.0056

Trewick, S., Brescia, F., & Jordan, C. (2009). Diversity and phylogeny of New Caledonian

Placostylus land snails; evidence from mitochondrial DNA. *Memoires Du Museum National d'Histoire Naturelle*, *198*, 421–436.

Triggs, S. J., Sherley, G. H. (1993). Allozyme genetic diversity in *Placostylus* land snails and implications for conservation. *New Zealand Journal of Zoology*, *20*(1), 19–33. https://doi.org/10.1080/03014223.1993.10423239

Uribe, J. E., Irisarri, I., Templado, J., & Zardoya, R. (2019). New patellogastropod mitogenomes help counteracting long-branch attraction in the deep phylogeny of gastropod mollusks. *Molecular Phylogenetics and Evolution*. https://doi.org/10.1016/j.ympev.2018.12.019

Wade, C. M., Mordan, P. B., & Clarke, B. (2001). A phylogeny of the land snails (Gastropoda: Pulmonata). *Proceedings of the Royal Society B: Biological Sciences*, *268*(1465), 413–422. https://doi.org/10.1098/rspb.2000.1372

Wade, Christopher M., Mordan, P. B., & Naggs, F. (2006). Evolutionary relationships among the Pulmonate land snails and slugs (Pulmonata, Stylommatophora). *Biological Journal of the Linnean Society*, *87*(4), 593–610. https://doi.org/10.1111/j.1095-8312.2006.00596.x

Winnepenninckx, B., Backeljau, T., & De Wachter, R. (1993). Extraction of high molecular weight DNA from molluscs. *Trends in Genetics*, *9*(12), 407. https://doi.org/10.1016/0168-9525(93)90102-N

Supplementary

Supplementary table 1: Summary statistics of mitochondrial genome and ribosomal cassette assemblies of five different Placostylinae species.

PS28: *Placostylus fibratus*, PS45: *Placostylus porphyrostomus*, PS127: *Eumecostylus cleryi*, PS185: *Placostylus ambagiosus*, PS257: *Placostylus hongii*

	Mitochondrial genome					Ribosomal cassette				
	PS28	PS45	PS127	PS185	PS257	PS28	PS45	PS127	PS185	PS257
Initial number of reads	19,457,636	17,740,626	24,869.360	21,795,718	37,663,184	19,457,636	17,740,626	24,869.360	21,795,718	37,663,184
Number of mapped reads	1,279	2,770	14,969	16,999	14,136	17,750	22,483	42,198	23,556	14,944
Mean read depths	8.89	19.21	103.64	134.2	98.23	217.37	249.05	860.32	258.7	237
Maximum read depth	24	36	148	203	176	465	632	1028	699	582
Minimum read depths	1	4	40	16	9	119	137	156	79	50

Chapter 2: Geometric morphometrics and machine learning challenge currently accepted species limits of the land snail *Placostylus* (Pulmonata: Bothriembryontidae) on the Isle of Pines, New Caledonia.

ABSTRACT

Size and shape variation of shells can be used to identify natural phenotypic clusters and thus delimit snail species. Here we apply both supervised and unsupervised machine learning algorithms to a geometric morphometric dataset to investigate size and shape variation of the shells of the endemic land snail *Placostylus* from New Caledonia. We sampled eight populations of *Placostylus* from the Isle of Pines, where two species of this genus reportedly coexist. We used neural network analysis as a supervised learning algorithm and Gaussian mixture models as an unsupervised learning algorithm. Using a training dataset of individuals assigned to species using nuclear markers, we found that supervised learning algorithms could not unambiguously classify all individuals of our expanded dataset using shell size and shape. Unsupervised learning showed that the optimal division of our data consisted of three phenotypic clusters. Two of these clusters correspond to the established species *Placostylus fibratus* and *P. porphyrostomus*, while the third cluster was intermediate in both shape and size. Most of the individuals that were not clearly classified using supervised learning were classified to this intermediate phenotype by unsupervised learning, and most of these individuals came from previously unsampled populations. These results may indicate the presence of persistent putative-hybrid populations of *Placostylus* in the Isle of Pines.

INTRODUCTION

Species are defined using many different criteria, in order to classify individuals into distinctive groups (Darwin, 1859; Mayr, 1942; Mahner, 1993; Mallet, 1995; De Queiroz, 2007). However, the continuity of evolution means that probably no single species concept can effectively serve all circumstances. Instead it has been argued that practical and useful definitions (operational criteria) should be prioritized when applying empirical data to the issue of species delimitation (Dubois, 2011; Vaux, Trewick & Morgan-Richards, 2016). The definition of species as morphological or genotypic clusters of individuals (Mallet, 1995) allows for the presence of individuals that are morphologically or genetically intermediate between two species, as long as these individuals are relatively uncommon. In this view, species are groups of individuals that interact and reproduce mostly together and morphological and genotypic clustering is a consequence of the correlation of their traits. This definition accommodates evolutionary flux and so allows speciation, gene flow and selection to be studied without the circularity of defining a species by reproductive isolation. In this framework, we need to be able to delimit groups of individuals using character data and, if intermediate individuals between clusters exist, be able to determine whether they are strongly under-represented (Mallet, 1995). Robust analytical tools are necessary to both identify groups and detect intermediates. Recent advances in machine learning using supervised and unsupervised learning approaches allow identification of specimens with and without *a priori* labels and so provide less subjective ways of delimitating species' limits.

In parallel with advances in machine learning, since the 1980s there have been developments in the analysis of morphological variation (reflecting the development of a formal theory of shape and geometric morphometric methods) (Kendall, 1986; Mitteroecker & Gunz, 2009). Landmark-based geometric morphometrics use the relationships between a set of landmarks to study the shape of organisms, rather than a set of traditional distance measurements (Bookstein, 1991). In this approach, shape differences between individuals are defined as any variation found among a set of homologous landmarks after they have been scaled, rotated and translated to the same criteria; a process called generalized Procrustes analysis. Two types of landmarks are most commonly used. Fixed landmarks sample homologous features in different individuals, whereas semi-landmarks can be used to sample the same positions on any surface located between two fixed landmarks (Zelditch, Swiderski & Sheets, 2004). Statistical analyses can then quantify the differences in shape between individuals by considering variation in a set of aligned landmarks, typically after steps of

ordination and dimension reduction (Zelditch, Swiderski & Sheets, 2004). Geometric methods are more sensitive to shape variation than traditional morphometry (Rohlf & Marcus, 1993; Maderbacher *et al.*, 2008) and have been usefully applied to the study of shell shape variation of molluscs (Carvajal-Rodríguez, Conde-Padín & Rolán-Alvarez, 2005; Cruz, Pante & Rohlf, 2012; Gustafson *et al.*, 2014; Dillon & Jacquemin, 2015; Gustafson & Bolek, 2016; Vaux *et al.*, 2017, 2018; Rao *et al.*, 2018; Verhaegen *et al.*, 2018). In the present study we combine the use of geometric morphometry, with supervised and unsupervised learning algorithms, to test whether the current species-level taxonomy of the land snail *Placostylus* on the Isle of Pines, New Caledonia corresponds with patterns of shell morphological variation.

Placostylus is a genus of large terrestrial pulmonate snails that is native to the islands of the western Pacific ocean (Breure, Groenenberg & Schilthuizen, 2010). In New Caledonia, six species are recognized according to current taxonomy (Neubert, Chérel-mora & Bouchet, 2009), but a combination of genotypic and geometric data indicate that the two most abundant nominal species-level taxa *Placostylus fibratus* (Martyn 1789) and *P. porphyrostomus* (Pfeiffer, 1851) may contain additional undescribed species on the main island, Grande Terre (Dowle *et al.*, 2015). These two species are also locally common on the Isle of Pines, a small island (152 km^2) south of Grande Terre, where *P. fibratus* is still harvested for food (Brescia, 2011).

Analysis of 26 specimens from the Isle of Pines has shown that shell morphology seems sufficient to discriminate the two sympatric *Placostylus* species, with concordance between shell morphology and nuclear genetic clusters being unequivocal (Dowle *et al.*, 2015). Shell differences in sympatry indicate that the size and shape of adult *Placostylus* shells are largely controlled by genetic factors, rather than environmental differences. *Placostylus fibratus* is larger and has a wider ventral aperture than *P. porphyrostomus*, which is usually smaller with a thicker lip around a smaller aperture. Although shell morphology was concordant with two genetic clusters in a previous sample from the Isle of Pines, some crucial snail populations on the island were not sampled. Those from the Comwagna district are of particular interest because adult snails are intermediate in size to the two recognized species on the island and thus difficult to identify (Brescia, 2011). Available mtDNA sequence data suggest some historical genetic exchange consistent with hybridization between the two species (Dowle *et al.*, 2015), and this may explain why these populations are intermediate in size to *P. fibratus* and *P. porphyrostomus*.

Based on geometric morphometric analyses of data on shell size and shell shape data, the aim of this study is to apply supervised and unsupervised learning algorithms to test whether the recognition of two *Placostylus* species on the Isle of Pines is sufficient, and quantify the frequency of potential novel phenotypes.

METHODS

Rationale

Supervised learning is a type of algorithm that is 'taught' to create a function that will link a set of features (inputs) to a set of labels (outputs) (Hastie, Tibshirani & Friedman, 2009). This is done by providing example data, called the training dataset, for which features are already associated with some known labels. A training process is then used to find the optimal function that links feature values (shell shape and size) to label values (species identification). Here we use a class of artificial neural networks called a multilayer perceptron for supervised learning (Ripley, 1994). Multilayer perceptrons are recognised for their predictive efficiency, and rely on the use of gradient descent and backpropagation algorithms. Associations of these algorithms with geometric morphometrics can be used to classify individuals into different categories based on shape and size information (Baylac, Villemant & Simbolotti, 2003; Dubey *et al.*, 2006; Soda, Slice & Naylor, 2017). We chose to use this algorithm as it requires few assumptions and can be applied to a wide range of data types (Ripley, 1994). A sample of snails where we have both morphometric and genetic information from a previous study was used as a training dataset (Dowle *et al.*, 2015). These shells form two separate morphological groups that could also be separated using population genetic tools with nuclear loci; using the genotypic cluster definition they are regarded as two species. If the current taxonomy of *Placostylus* reflects biological reality and shell morphology can be used to separate individuals into two distinct species, we expect the supervised learning algorithms to classify without ambiguities all (or almost all) individuals in our expanded sample.

We also use Gaussian Mixture Models (GMMs), a class of unsupervised learning algorithm, to find the optimal number of morphological groups in our new (expanded) dataset. Unsupervised learning is a class of machine learning algorithms, which infer structure in datasets that do not have *a priori* labels (Hastie, Tibshirani & Friedman, 2009). In the case of GMMs, the dataset is considered to be a mixture of population samples with different Gaussian distributions, and the modelling process divides the data into different classes, each one corresponding to a single Gaussian distribution (Fraley & Raftery, 2006). We use GMMs because this approach can model any kind of data, as long as the data can be separated into different Gaussian components. Different models are created for a different number of hypothetical clusters. Bayesian information criteria (BIC) scores are used to determine which model is the best fit to the data and in this way the optimal number of clusters in our sample is

estimated. If the current shell-based taxonomy of *Placostylus* in the Isle of Pines is correct, we expect an optimal GMM that divides our dataset into two groups, corresponding to the two currently recognized species.

Sampling strategy

A total of 337 *Placostylus* snails was sampled across eight sites on the Isle of Pines in April 2015 (Fig. 1). Sample sites were arrayed along a set of elevation gradients from the coast to the centre of the island (Fig. 1). All live snails within 20m radius of a fixed point were collected by hand, at three locations along each of eight transects. Two-dimensional geometric morphometric data were gathered from the 337 shells sampled in 2015 and from the 26 shells that were collected and analysed previously (Dowle *et al.*, 2015). Among the eight population samples examined, five had previously been sampled (Vao, Gadji, Kere, Youaty, Touete), and three were sampled for the first time (Comagna, Wapan, Waatchia) (Fig. 1). Once snails are sexually mature, age-related variation in shell shape and size are considered to be minimal (F. M. Brescia *et al.*, 2008). Only shells of adult snails (identified by the thickened lip) were photographed for shape analysis. As these snails are hermaphroditic it was not necessary to consider sexual dimorphism.

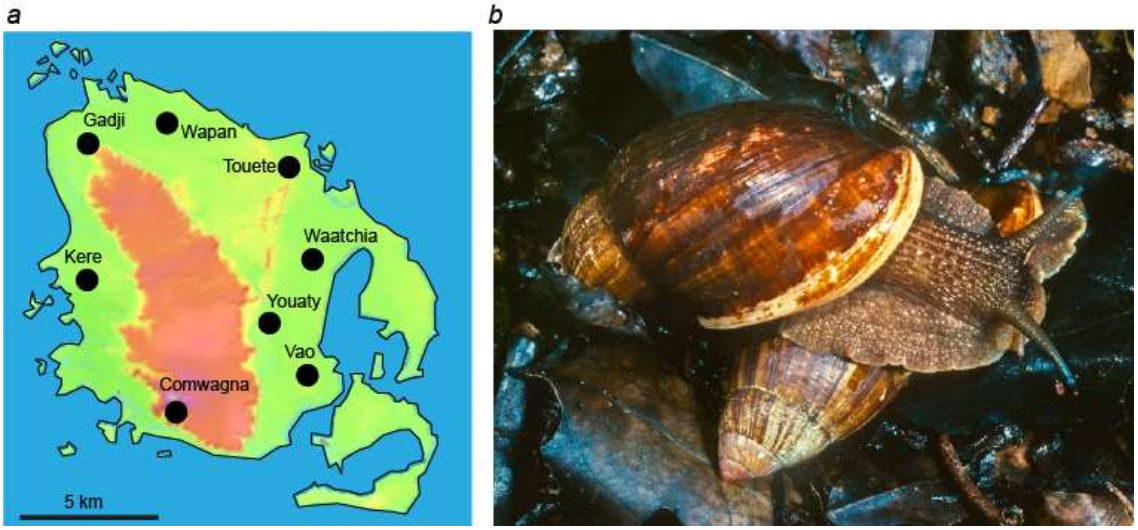


Figure 1. A. Topographic map of the Isle of Pines, New Caledonia, highlighting the different locations where snails were sampled. Colours represent different elevation levels. **B.** Two live *Placostylus fibratus* in their natural environment (image: Rod Morris).

Geometric morphometrics (GM)

The set of two-dimensional landmarks used for this study was derived from a set of optimal landmarks and semi-landmarks established in an analysis of New Zealand *Placostylus* (Daly, 2017). A total of 40 landmarks and semi-landmarks were used, consisting of nine fixed landmarks and 31 sliding/semi-landmarks (Fig. 2). Digital images of the ventral surface of each shell were obtained using a Canon EOS 600d with EF100 mm f2.8 USM macro lens after careful positioning of the shell in a bed of sand of contrasting colour. The camera equipment was mounted on a high-precision Kaiser stand to allow reproducible positioning and orientation (Dowle *et al.*, 2015). All images were captured with the lens fixed at the same distance from shell, and shells were positioned by the same person to minimize operator error (Schilthuizen & Haase, 2010). In order to place the semi-landmarks along the curves of the shells, two 'combs' were placed manually using Adobe Photoshop CS6, with their teeth perpendicular to a line from the shell apex to the intersection of the lips and outer shell (Fig. 2). Digitization of landmarks and semi-landmarks was conducted using the programme tpsDig2 v.1.1 (Rohlf, 2015). Landmark type assignment and Procrustes analysis were then performed using Coordgen v.8.0 (Zelditch, Swiderski & Sheets, 2004). Principal component analysis (PCA) was executed on the covariance matrices of the aligned landmark coordinates using MorphoJ (Klingenberg, 2011).

The size of the shells was incorporated into the analysis using the 'centroid size' tool included in the Procrustes analysis in Coordgen8. This size estimation is calculated as the square root of the sum of the arrays coming from the centroid position of a shape to each of the landmarks (Klingenberg, 2016). Both shape information (principal components (PCs)) and the centroid size estimate (from here on referred to as 'shell size') were used as input variables in the supervised and unsupervised learning algorithms.

Errors linked to shell manipulation and digitization were assessed by experimental replication; this involved taking five photographs of the same shell and comparing the variance attributed to the repeat process with the variance of the whole dataset using the *morphol.disparity*() function in the package *geomorph* v.3.1.2 (Adams, Collyer &

Kaliontzopoulou, 2018) in R v.3.6.1 (R Development Core Team & R Core Team, 2017). Overall, the variance linked to manipulation and digitization error corresponded to <0.3% of the variance of the rest of the dataset, which we deemed to be negligible for subsequent analyses.

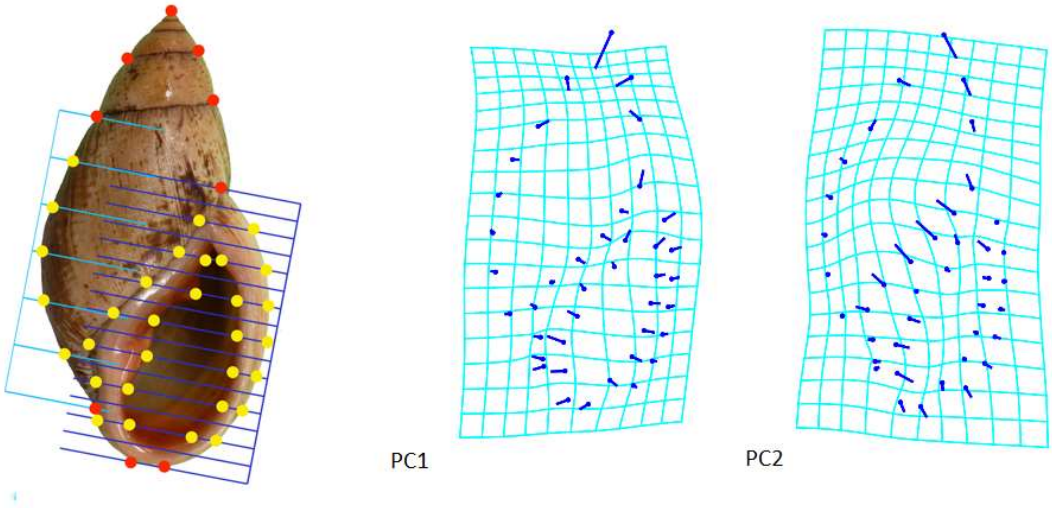


Figure 2. A. Shell of New Caledonian *Placostylus fibratus* with the 40 landmarks used in the geometric morphometry analysis. Digitizing combs are in blue and the orientation is indicated by the white dashed line. Red and yellow dots indicate permanent and semi-landmarks respectively. **B, C.** Relative displacement of landmarks using thin-plate splines, showing shell shape variation for PCs 1 (**B**) and 2 (**C**) of the geometric morphometric analysis of the whole dataset. PCs 1 and 2 explained 35.3% and 15.5% of total variance, respectively. Length of lollipop lines is proportional to warping in shape space for each PC.

Supervised learning algorithms

For the training process of the supervised learning algorithm, we used 26 shells genetically assigned to one of two species (our training dataset). These specimens had been unequivocally assigned by Dowle *at al*. (2015) to one of the two recognized *Placostylus* species on the basis of a nuclear DNA dataset comprising 661 SNP. Neural network models were built using the *neuralnet*() function in the R package *neuralnet* v.1.44.2 (Fritsch *et al.*, 2016). Input data (shape and size) were normalized and scaled using a minmax function. Neural network models containing two hidden layers of five and three neurons were used (Fig. 3).

We first tested the efficacy of different neural network models using exhaustive cross validation on our training dataset (n = 26). For model validation purposes, this dataset was split

into different train and test datasets of various proportions, using different numbers of input variables (for split proportions, number of input variables and other details see Supplementary Material Table S1). The objectives of this process were to estimate whether we had a sufficiently large training dataset and to determine the optimal number of input variables. Efficacy of the models was assessed by looking at the mean proportion of individuals assigned correctly in the test datasets for a large number of replicates. We found the optimal models used only two PCs and one size variable. In these models the cross validation correctly identified all snails most of the time (mean proportion of correct assignment was between 98.8% and 99.8% for different sized training datasets; Supplementary Material Table S1).

We then trained neural network models using the whole training dataset and used trained models on our new dataset of 337 shells to assign each snail to one of the two species. We repeated this process for a large number of neural network models (x1000) to avoid overfitting issues arising from single model interpretation (Zhou, 2009), retaining the average value of the neural networks outputs as assignment score for each individual (ensemble learning).

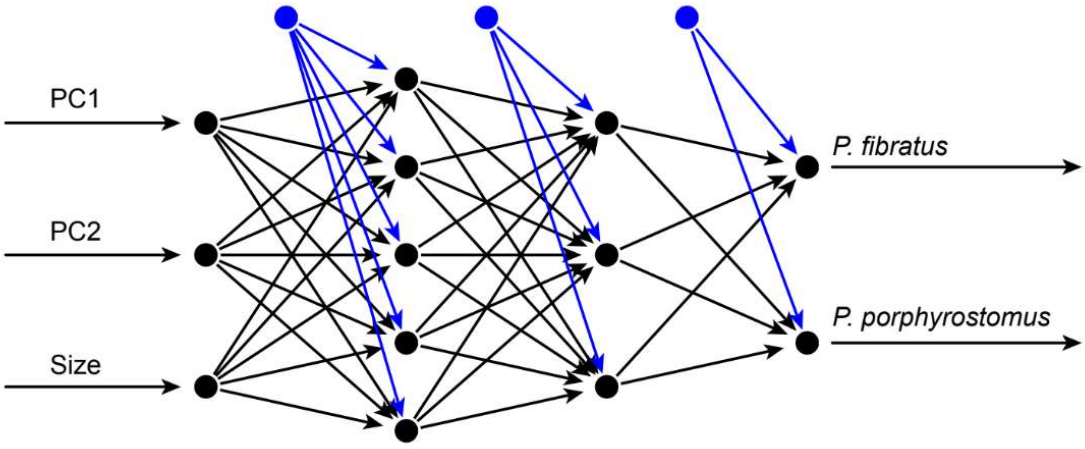


Figure 3. Schematic of a multilayer perceptron used for the supervised learning algorithm examining classification of *Placostylus* land snail shells with the two shape PCs and centroid size. Neural networks contained two hidden layers of five and three artificial neurons, respectively. Here weight and bias parameters are represented by black and blue arrows.

Unsupervised learning algorithms

For Gaussian mixture models we used the R package *mclust* v. 5.4 (Fraley & Raftery, 2006). Gaussian mixture models were built using shell size and the same number of PCs as in the supervised learning analyses (1 and 2). Different models were built for a range of *a priori* clusters in the dataset, with the optimal model being selected on the basis of BIC scores (Supplementary Material Fig. S1). Results of the modelling process with additional input variables (PC1–5, shell size) were also computed and yielded similar results.

RESULTS

First and second PCs of variation of the geometric morphometric analysis accounted for 35.3% and 15.5% respectively of the total variance in shell shape. Warp transformation grids indicate that spire height, aperture size and lip thickness varied the most (Fig. 2).

Supervised learning algorithms

Using the supervised learning algorithm, individual specimens were assigned to one or other of the two species, *Placostylus fibratus* or *P. porphyrostomus*, if their assignment probability was >95%. On the basis of this criterion, 224 individual snails were identified as *P. fibratus* (67%) and 61 as *P. porphyrostomus* (18%); 51 (15%) could not be assigned to either species (Table 1, Figs 2B, 4A). Most of the unassigned individuals came from two populations: Comwagna (28 individuals) and Wapan (15 individuals). The six other population samples contained individuals that could be confidently identified to either *P. fibratus* or *P. porphyrostomus*, with only a few unassigned individuals (Table 1). Youaty was the only population sample with just one snail species (*P. fibratus*).

Unsupervised learning algorithms

The shell size and shape variation had an optimal fit to a GMM with three clusters (elliposidal, equal volume, shape and orientation (EEE) model; Supplementary Material Fig. S1). Assignment probabilities to the three phenotypic clusters were computed for each of the 337 individuals (Table 1, Fig. 4). Two of the phenotypic clusters identified using shell size and shape variables correspond to the recognized species *P. fibratus* and *P. porphyrostomus*. A third cluster was also observed. This consisted mostly of individuals from Comwagna and Wapan, which were unassigned by the supervised learning algorithm (Fig. 4, Supplementary Material Table S1). The snails in the third cluster were intermediate in shell size and shape to *P. fibratus* and *P. porphyrostomus* (Fig. 5).

Using GMMs, 226 (67%) individuals were assigned to the *P. fibratus* cluster, 56 (17%) to the *P. porphyrostomus* cluster and 54 (16%) to the intermediate group. The assignment of individuals was similar to the result produced by the supervised learning algorithm, with specimen assignment probabilities <0.95 (from supervised learning) placed in the third cluster (Fig. 4). Only 11 of 337 snails (3.27%) were classified in different ways by the two methods, as explained here: of the three individuals not assigned by supervised learning, two were assigned by GMM to *P. fibratus* and one to *P. porphyrostomus*; eight shells classified as *P.*

porphyrostomus by the supervised learning algorithm were classified as either *P. fibratus* or the intermediate type by GMM. The geographic source of the snails that contribute to the three morphological clusters identified by GMM suggests that the intermediate shell phenotype has a limited distribution. The Youaty population sample consisted only of *P. fibratus*, but the Touete, Waatchia, Gadji, Kere and Vao population samples contained both *P. fibratus* and *P. porphyrostomus*. The Wapan population sample contained examples of all three clusters, while the Comwagna population sample consisted mostly of individuals assigned to the intermediate type, with a few individuals of *P. porphyrostomus* (Table1).

Table 1. Numbers of *Placostylus* in the three phenotype clusters (species) as assigned by supervised and unsupervised analysis of shell size and shape for eight population samples from the Isle of Pines, New Caledonia. Individuals were accepted as assigned to a cluster only if their assignment score was > 0.95.

Population	**Comwagna**	**Gadji**	**Kere**	**Touete**	**Vao**	**Waatchia**	**Wapan**	**Youaty**	**Total**
Supervised learning classification									
P. fibratus	0	29	31	26	52	38	23	25	224
P. porphyrostomus	8	9	7	7	16	9	5	0	61
Not assigned	28	1	2	1	3	1	15	0	51
Unsupervised learning classification									
P. fibratus	0	29	32	26	53	38	23	25	226
P. porphyrostomus	4	9	7	7	15	9	5	0	56
Intermediate phenotype	32	1	1	1	3	1	15	0	54

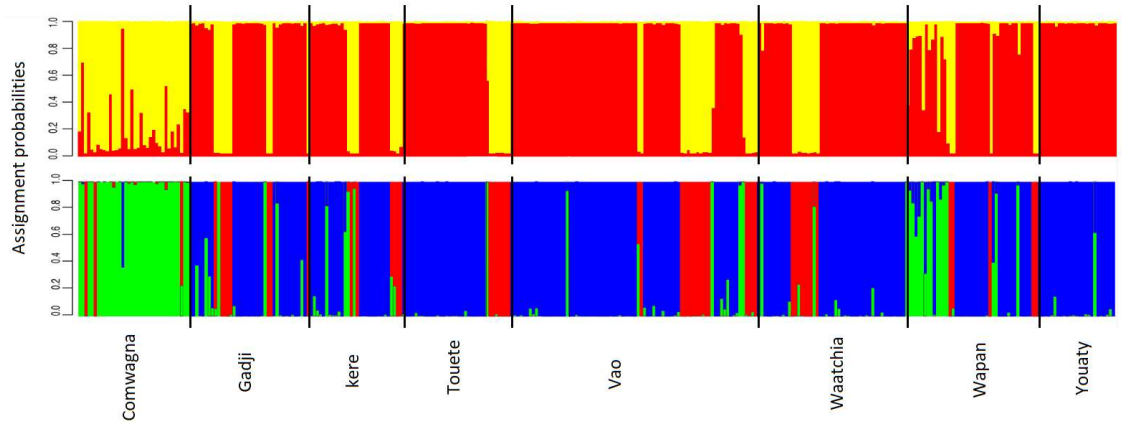


Figure 4. Assignment probabilities calculated for shell shape and size variation in eight population samples of *Placostylus* from the Isle of Pines. **A.** Neural network supervised learning analysis for two phenotypes (the two recognized species). Colour coding: red, *P. fibratus*; yellow *P. porphyrostomus.* **B.** An unsupervised analysis based on an optimal GMM that found three phenotypes (the two recognized species and the intermediate phenotype) Colour coding: red, *P. porphyrostomus*; blue, *P. fibratus*; green, intermediate phenotype.

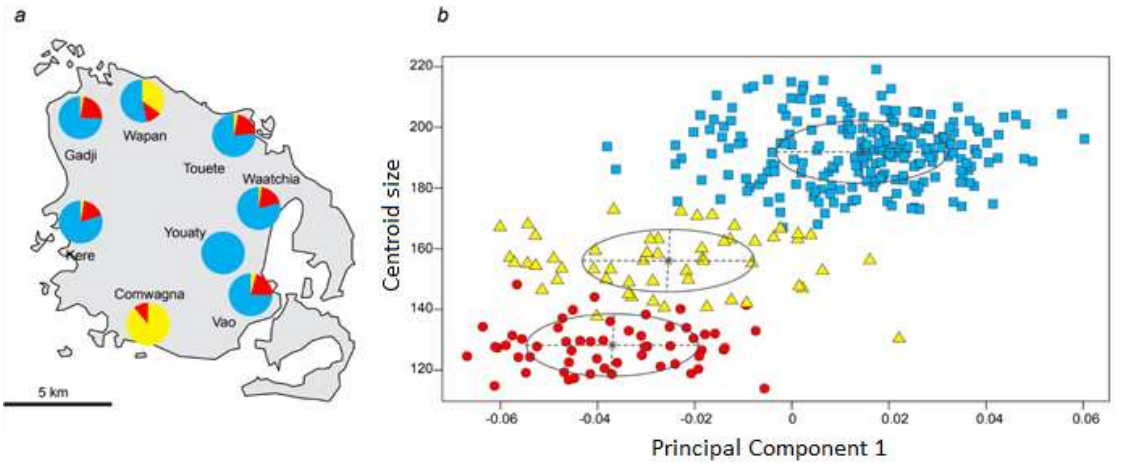


Figure 5. A. Relative frequency of three phenotypes of *Placostylus* in eight population samples from the Isle of Pines, New Caledonia, as indicated by the optimal GMM. The first two PCs of variation in shell shape (50.8% of total shape variation) and shell size. **B.** GMM-based classification of *Placostylus* in relation to variation of shell size and shape along PC 1 (35.3%). Colour coding: red, *P. porphyrostomus*; blue, *P. fibratus*; yellow, intermediate phenotype.

DISCUSSION

Shell morphology, separable into shape and size components through geometric morphometrics, is controlled by a suite of heritable characters that can be influenced by environmental factors. Aquatic gastropods have provided evidence that phenotypic plasticity of shell size and shape is widespread (Bourdeau *et al.*, 2015). The New Zealand land snail *Placostylus ambagiosus* shows size variation that is, in part, due to environmental factors (Parrish, Stringer & Sherley, 2014). However, previous work on the two New Caledonian *Placostylus* species *P. fibratus* and *P.porphyrostomus* has shown they are genetically distinct, despite coexisting at many locations on the Isle of Pines (Brescia *et al.*, 2008; Dowle *et al.*, 2015). Distinct shell traits of sympatric species are unlikely to be due to phenotypic plasticity, as the snails experience the same environment.

When applied to our new dataset, supervised learning algorithms failed to unambiguously assign all snails into one or other of the two species that are recognized on the basis of traditional shell-based taxonomy. Most of the specimens with low assignment probabilities originate from two newly sampled locations on the Isle of Pines. The presence of many unassigned snails suggests that the classification into two morphological groups (species) is too simplistic in the context of our expanded dataset. At the same time, unsupervised learning algorithms indicate the presence of a third group and this result is consistent our finding that supervised learning failed to classify some individuals. This third group is intermediate in both shape and size to the two recognized species (Fig. 5).

The snails from the Comwagna district, which constitute the majority of intermediate-type individuals, have never been harvested for commercial purposes due to their smaller size. To date, these individuals have been considered to be a sub-group of *P. fibratus* (Brescia *et al.*, 2008). Our analysis suggests that they are distinct and could potentially represent an unusual population of hybrid individuals. While available data on mtDNA haplotype distributions suggest historical hybridization between *P. fibratus* and *P.porphyrostomus*, the presence of morphologically and genetically distinct sympatric populations had not indicated on-going hybridization (Dowle *et al.*, 2015). Here, though, our morphological data suggest hybridization could go beyond simple mitochondrial introgression and could be the origin of an exclusively hybrid population at Comwagna, and rare hybrids at other localities.

Hybridization can be viewed as a gradient that extends from rare introgression through stable hybrid zones to the formation of exclusively hybrid populations (Nieto Feliner *et al.*,

2017). Introgression can generate gene flow between distinct species, but parental lineages can remain distinct when hybrids normally backcross to one of the parental forms (Anderson, 1953; Folk *et al.*, 2017). Observations of five populations of *Placostylus* from the Isle of Pines (Gadji, Kere, Touete, Waatchia, Vao) seem to be consistent with the rare introgression hypothesis suggested by mtDNA haplotype data; this is because of the presence of just two distinctive parental phenotypic clusters and a small number (1.4–2.9%) of individuals with intermediate shell phenotype in these five populations (Fig. 3). Given the absence of intermediates at localities where *P. fibratus* and *P.porphyrostomus* are sympatric, we could apply the genotypic cluster definition (Mallet, 1995) and recognize two species. On the other hand, the numerical dominance of the intermediate phenotype in the population samples from Wapan and Comwagna suggests that if hybridization explains this phenotype, then hybridization has been locally common and has led to the loss of distinct parental phenotypic clusters at these sites.

Environmental conditions instead of interspecific hybridization may potentially explain the prevalence of individuals with an intermediate phenotype. Of the eight locations where snails were sampled, Comwagna is situated on red ferralitic (=laterite) soils, whereas all the others are on uplifted coral reef terrain (Lagarde & Ouetcho, 2017). Experiments have shown a positive correlation between calcium availability and shell growth in some terrestrial gastropods (Beeby & Richmond, 2007). Phenotypic plasticity is likely to account for a part of the morphological diversity observed for New Caledonian *Placostylus* (Brescia *et al.*, 2008; Dowle *et al.*, 2015). Lower calcium availability on ferralitic soils is, therefore, a plausible explanation for the intermediate phenotype at Comwagna, but this is challenged by the presence of that phenotype on calcium rich substrate at Wapan. Although hybridization and recent gene flow between *P. fibratus* and *P.porphyrostomus* seems to be the most likely explanation for the presence of the intermediate phenotype, at this stage we cannot rule out the possibility that the intermediate phenotype represents a novel lineage distinct from *P. fibratus* and *P.porphyrostomus*, or the possibility of an explanation based on environmental factors.

Here we used two approaches to classify samples: one uses a training dataset, the other does not. The first method seeks to classify individuals within a framework of knowledge, while the second explores a dataset to find optimal divisions. Both methods are subject to overfitting issues (Adams, Rohlf & Slice, 2004). Supervised learning can misclassify individuals that are not inside the range of variance of the training data and unsupervised learning can under- or overestimate the number of clusters in a dataset. We suggest that low

support values for individual classification from supervised learning can be biologically interesting. While low support assignment can arise from technicalities such as a small training dataset or mislabelling, here the concordance of supervised learning results with unsupervised learning classification suggests our models were not lacking power. Our results, therefore, support the use of both methods, rather than either of them individually. Overall, our work joins the growing number of studies that have demonstrated that the association of geometric morphometrics and machine learning can be useful in addressing biological questions (Dobigny, Baylac & Denys, 2002; Dubey *et al.*, 2006; Bocxlaer & Schultheiß, 2010; Mapp *et al.*, 2017; Nattier *et al.*, 2017; Soda, Slice & Naylor, 2017; Fang *et al.*, 2018).

ACKNOWLEDGEMENTS

We thank the New Caledonian people, the Institut Agronomique Neo-Caledonien and members of the Phoenix group for Evolutionary Ecology and Genetics (evolves.massey.ac.nz). This research was supported by the Institute Agronomique Neo-Caledonien and a Massey University PhD scholarship (awarded to MQ).

REFERENCES

ADAMS, D., COLLYER, M. & KALIONTZOPOULOU, A. 2018. *Geomorph: software for geometric morphometric analyses, version 3.1.2*. https://CRAN.R-project.org/package=geomorph. Accessed 1 september 2019.

ADAMS, D.C., ROHLF, F.J. & SLICE, D.E. 2004. Geometric morphometrics: ten years of progress following the 'revolution.' *Italian Journal of Zoology*, **71**: 5–16.

ANDERSON, E. 1953. Introgressive hybridization. *Biological Reviews*, **28**: 280–307.

BAYLAC, M., VILLEMANT, C. & SIMBOLOTTI, G. 2003. Comining geometric morphometrics with pattern recognition for the investigation of species complexes. *Biological Journal of the Linnean Society*, **80**: 89–98.

BEEBY, A. & RICHMOND, L. 2007. Differential growth rates and calcium-allocation strategies in the garden snail *Cantareus aspersus*. *Journal of Molluscan Studies*, **73**: 105–112.

BOCXLAER, B. Van & SCHULTHEIß, R. 2010. Comparison of morphometric techniques for shapes with few homologous landmarks based on machine-learning approaches to biological discrimination. *Paleobiology*, **36**: 497–515.

BOOKSTEIN, F.L. 1991. *Morphometric Tools for landmark data. Geometry and biology.* Cambridge University press, Cambridge, UK.

BOURDEAU, P.E., BUTLIN, R.K., BRÖNMARK, C., EDGELL, T.C., HOVERMAN, J.T. & HOLLANDER, J. 2015. What can aquatic gastropods tell us about phenotypic plasticity? A review and meta-analysis. *Heredity*, **115**: 312–321.

BRESCIA, F. 2011. *Ecology and population trends in New Caledonian* Placostylus *snails (Mollusca: Gastropoda: Bulimulidae)*. PhD thesis. Massey University, Palmerston North, New Zealand.

BRESCIA, F.M., PÖLLABAUER, C.M., POTTER, R.A. & ROBERTSON, A.W. 2008. A review of the ecology and conservation of *Placostylus* (Mollusca: Gastropoda: Bulimulidae) in New Caledonia. *Molluscan Research*, **28**: 111–122.

BREURE, A.S.H., GROENENBERG, D.S.J. & SCHILTHUIZEN, M. 2010. New insights in the phylogenetic relationswithin the Orthalicoidea (Gastropoda, Stylommatophora) based on 28S sequence data. *Basteria*, **74**: 25–32.

CARVAJAL-RODRÍGUEZ, A., CONDE-PADÍN, P. & ROLÁN-ALVAREZ, E. 2005. Decomposing shell form into size and shape by geometric morphometric methods in two sympatric ecotypes of *Littorina saxatilis*. *Journal of Molluscan Studies*, **71**: 313–318.

CRUZ, R.A.L., PANTE, M.J.R. & ROHLF, F.J. 2012. Geometric morphometric analysis of shell shape variation in *Conus* (Gastropoda: Conidae). *Zoological Journal of the Linnean Society*, **165**: 296–310.

DALY, E. 2016. *Fine scale population structure through space and time*. PhD thesis. Massey University, New Zealand.

DARWIN, C. 1859. *On the origin of species*. John Murray, London.

DE QUEIROZ, K. 2007. Species concepts and species delimitation. *Systematic Biology*, **56**: 879–886.

DILLON, R.T. & JACQUEMIN, S.J. 2015. The heritability of shell morphometrics in the freshwater pulmonate gastropod *Physa*. *PLoS One*, **10**: 1–13.

DOBIGNY, G., BAYLAC, M. & DENYS, C. 2002. Geometric morphometrics, neural networks and diagnosis of sibling. *Biological Journal of the Linnean Society*, **77**: 319–327.

DOWLE, E.J., MORGAN-RICHARDS, M., BRESCIA, F. & TREWICK, S.A. 2015. Correlation between shell phenotype and local environment suggests a role for natural selection in the evolution of *Placostylus* snails. *Molecular Ecology*, **24**: 4205–4221.

DUBEY, B.P., BHAGWAT, S.G., SHOUCHE, S.P. & SAINIS, J.K. 2006. Potential of artificial neural networks in varietal identification using morphometry of wheat grains. *Biosystems Engineering*, **95**: 61–67.

DUBOIS, A. 2011. Species and "strange species" in zoology: Do we need a "unified concept of species" ? *Comptes Rendus Palevol*, **10**: 77–94.

FANG, Z., FAN, J., CHEN, X. & CHEN, Y. 2018. Beak identification of four dominant octopus species in the East China Sea based on traditional measurements and geometric morphometrics. *Fisheries Science*, **84**: 975–985.

FOLK, R.A., SOLTIS, P.S., SOLTIS, D.E. & GURALNICK, R. 2018. New prospects in the detection and comparative analysis of hybridization in the tree of life. *American Journal of Botany*, **105**: 364–375.

FRALEY, C. & RAFTERY, A.E. 2006. MCLUST version 3: an R Package for normal mixture modeling and model-based clustering. Available at: https://cran.r-project.org/src/contrib/Archive/mclust.

FRITSCH, S., GUENTHER, F., SULING, M. & M. MUELLER, S. 2016. *Neuralnet: training of neural networks, version 1.44.2*. https://CRAN.R-project.org/package=neuralnet. Accessed 1 September 2019.GUSTAFSON, K.D. & BOLEK, M.G. 2016. Effects of trematode parasitism on the shell morphology of snails from flow and nonflow environments. *Journal of Morphology*, **277**: 316–325.

GUSTAFSON, K.D., KENSINGER, B.J., BOLEK, M.G. & LUTTBEG, B. 2014. Distinct snail (Physa) morphotypes from different habitats converge in shell shape and size under common garden conditions. *Evolutionary Ecology Research*, **16**: 77–89.

HASTIE, T., TIBSHIRANI, R. & FRIEDMAN, J. 2009. *The elements of statistical learning: data mining, inference, and prediction*. Springer, New Yor, NY.

KENDALL, D.G. 1986. A survey of the statistical theory of shape. *Statistical Science*, **10**: 354–363.

KLINGENBERG, C.P. 2011. MorphoJ: an integrated software package for geometric morphometrics. *Molecular Ecology Resources*, **11**: 353–357.

KLINGENBERG, C.P. 2016. Size, shape, and form: concepts of allometry in geometric morphometrics. *Development Genes and Evolution*, **226**: 113–137.

LAGARDE, L. & OUETCHO, A. 2016. Horticultural structures on ultramafic soils: the case of Isle of Pines and other parts of southern Grande Terre (New Caledonia). In: *La pratique de l'espace en Océanie: découverte, appropriation et émergence des systèmes sociaux traditionnels* (F. Valentin & G. Molle, eds), pp. 79-90. Société préhistorique Française, Paris.

MADERBACHER, M., BAUER, C., HERLER, J., POSTL, L., MAKASA, L. & STURMBAUER, C. 2008. Assessment of traditional versus geometric morphometrics for discriminating populations of the *Tropheus moorii* species complex (Teleostei: Cichlidae), a Lake Tanganyika model for allopatric speciation. *Journal of Zoological Systematics and Evolutionary Research*, **46**: 153–161.

MAHNER, M. 1993. What is a species? *Journal for General Philosophy of Science*, **24**: 103–126.

MALLET, J. 1995. A species definition for the modern synthesis: *Trends in Ecology & Evolution*, **10**: 294–299.

MAPP, J., HUNTER, E., KOOIJ, J. Van Der, SONGER, S. & FISHER, M. 2017. Otolith shape and size : the importance of age when determining indices for fish-stock separation. *Fisheries Research*, **190**: 43–52.

MAYR, E. 1942. *Systematics and the origin of species*. Columbia University Press, NY.

MITTEROECKER, P. & GUNZ, P. 2009. Advances in Geometric morphometrics. *Evolutionary Biology*, **36**: 235–247.

NATTIER, R., PELLENS, R., ROBILLARD, T., JOURDAN, H., LEGENDRE, F., CAESAR, M., NEL, A. & GRANDCOLAS, P. 2017. Updating the phylogenetic dating of New Caledonian biodiversity with a meta-analysis of the available evidence. *Scientific Reports*, **7**: 1–9.

NEUBERT, E., CHÉREL-MORA, C. & BOUCHET, P. 2009. Polytypy, clines, and fragmentation : the bulimes of New Caledonia revisited. *Mémoires du Muséum d'Histoire Naturelle*, **198**: 37–131.

NIETO FELINER, G., ÁLVAREZ, I., FUERTES-AGUILAR, J., HEUERTZ, M., MARQUES, I., MOHARREK, F., PIÑEIRO, R., RIINA, R., ROSSELLÓ, J.A., SOLTIS, P.S. & VILLA-MACHÍO, I. 2017. Is homoploid hybrid speciation that rare? An empiricist's view. *Heredity*, **118**: 513–516.

PARRISH, G.R., STRINGER, I.A.N. & SHERLEY, G.H. 2014. The biology of *Placostylus ambagiosus* (Pulmonata: Bulimulidae) in New Zealand: part 1. Behaviour, habitat use, abundance, site fidelity, homing and the dimensions of eggs and snails. *Molluscan Research*, **34**: 139–154.

R CORE TEAM. 2017. *R: a language and environment for statistical computing*, *version 2.6.2*. R Foundation for Statistical Computing, Vienna, Austria. http:// CRAN.R-project.org/. Accessed 1 August 2017.

RAO, S.R., LIEW, T., YOW, Y. & RATNAYEKE, S. 2018. Cryptic diversity: two morphologically similar species of invasive apple snail in Peninsular Malaysia. *PLoS One*, **13**: e0196582.

RIPLEY, B.D. 1994. Neural Network and related Methods for Classification. *Journal of the*

Royal Statistical Society. Series B(Methodological), **56**: 409–456.

ROHLF, F.J. 2015. The tps series of software. *Hystrix, the italian journal of mammalogy,* **26(1)**: 9-12.

ROHLF, F.J. & MARCUS, L.F. 1993. A revolution in Morphometrics. *Trends in Ecology & Evolution*, **8**: 129–132.

SCHILTHUIZEN, M. & HAASE, M. 2010. Disentangling true shape differences and experimenter bias: Are dextral and sinistral snail shells exact mirror images? *Journal of Zoology*, **282**: 191–200.

SODA, K.J., SLICE, D.E. & NAYLOR, G.J.P. 2017. Artificial neural networks and geometric morphometric methods as a means for classification: A case-study using teeth from *Carcharhinus* sp. (Carcharhinidae). *Journal of Morphology*, **278**: 131–141.

VAUX, F., CRAMPTON, J.S., MARSHALL, B.A., TREWICK, S.A. & MORGAN-RICHARDS, M. 2017. Geometric morphometric analysis reveals that the shells of male and female siphon whelks Penion chathamensis are the same size and shape. *Molluscan Research*, **37**: 194–201.

VAUX, F., TREWICK, S.A., CRAMPTON, J.S., MARSHALL, B.A., BEU, A.G., HILLS, S.F.K. & MORGAN-RICHARDS, M. 2018. Evolutionary lineages of marine snails identified using molecular phylogenetics and geometric morphometric analysis of shells. *Molecular Phylogenetics and Evolution*, **127**: 626–637.

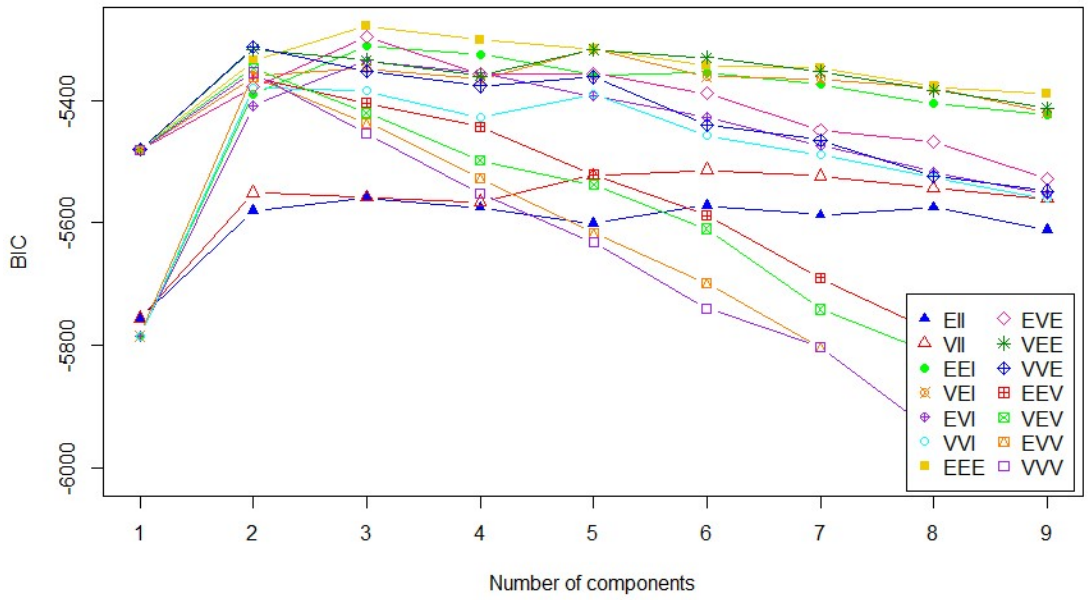
VAUX, F., TREWICK, S.A. & MORGAN-RICHARDS, M. 2016. Speciation through the looking-glass. *Biological Journal of the Linnean Society*, **120**: 480–488.

VERHAEGEN, G., MCELROY, K.E., BANKERS, L., NEIMAN, M. & HAASE, M. 2018. Adaptive phenotypic plasticity in a clonal invader. *Ecology and Evolution*, **8**: 4465–4483.

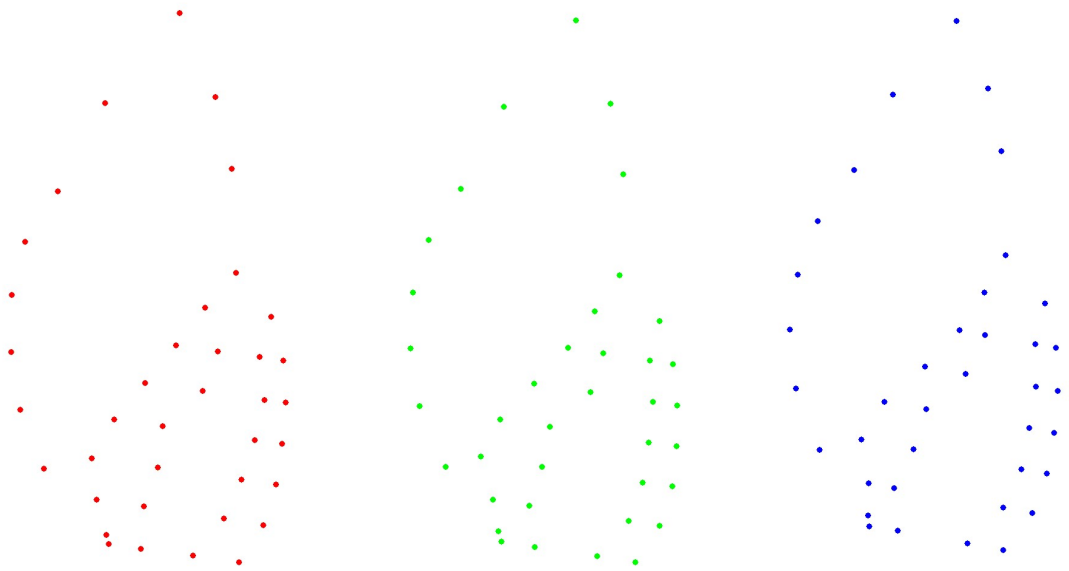
ZELDITCH, M., SWIDERSKI, D. & SHEETS, H.D. 2004. *Geometric morphometrics for biologists. A primer*. Elsevier Academic Press, San Diego, CA.

ZHOU, ZH. 2009. Ensemble Learning. In: *Encyclopedia of Biometrics* (SZ Li, ed.), pp. 270–273. Springer, Berlin.

Supplementary



Supplementary Figure 1: Bayesian Information Criterion (BIC) values for the different Gaussian Mixture Models created in the Mclust 5.4 package. The X-axis refers to the number of Gaussians used in each model, and different shapes of points indicates a different set of Gaussian parameters used in the models.



Supplementary figure 2: Average landmark positions for three morphological groups of *Placostylus* shells (as detected by GMM modelling; Red = 'porphyrostomus' group, green = 'intermediate' group; blue = 'fibratus' group').

Supplementary table 1: Proportions of *Placostylus* shells correctly assigned to their genetic groups in test datasets during exhaustive cross validation of supervised learning algorithms. Results are presented for different number of input variables (in this case principal component), and for various split proportions of the cross validation dataset. Centroid size was used as an input variable in each one of the models, alongside the indicated number of Principal Components.

Division of the training data set (proportion / number of snail)		Number of PCs used by NN models				
training dataset	test dataset	2	3	4	5	6
80% / 21	20% / 5	99.70%	49.60%	50%	62.30%	72.10%
70% / 18	30% / 8	99.80%	54.50%	53.90%	64.10%	73%
60% / 16	40% / 10	99.10%	50.90%	50.90%	64.50%	72.60%
50% / 13	50% / 13	98.80%	53.40%	52.90%	66.20%	72.60%

Supplementary table 2: Classification of 337 shells from the Isle of Pines, New Caledonia, according to supervised and unsupervised learning algorithms, based on shell shape and size.

ID	Supervised learning classification	Unsupervised learning classification	ID	Supervised learning classification	Unsupervised learning classification
COMAGNA-1	Not assigned	Intermediate	VAO-29	Fibratus	Fibratus
COMAGNA-10	Not assigned	Intermediate	VAO-3	Fibratus	Fibratus
COMAGNA-11	Not assigned	Intermediate	VAO-30	Porphyrostomus	Intermediate
COMAGNA-12	Not assigned	Intermediate	VAO-31	Porphyrostomus	Porphyrostomus
COMAGNA-13	Not assigned	Intermediate	VAO-32	Fibratus	Fibratus
COMAGNA-14	Not assigned	Intermediate	VAO-33	Fibratus	Fibratus
COMAGNA-15	Not assigned	Intermediate	VAO-34	Fibratus	Fibratus
COMAGNA-16	Not assigned	Intermediate	VAO-35	Fibratus	Fibratus
COMAGNA-17	Not assigned	Intermediate	VAO-36	Fibratus	Fibratus
COMAGNA-18	Not assigned	Intermediate	VAO-37	Fibratus	Fibratus
COMAGNA-19	Not assigned	Intermediate	VAO-38	Fibratus	Fibratus
COMAGNA-2	Not assigned	Intermediate	VAO-4	Fibratus	Fibratus
COMAGNA-20	Not assigned	Intermediate	VAO-40	Fibratus	Fibratus
COMAGNA-21	Not assigned	Intermediate	VAO-41	Fibratus	Fibratus
COMAGNA-22	Porphyrostomus	Intermediate	VAO-42	Fibratus	Fibratus
COMAGNA-23	Not assigned	Intermediate	VAO-43	Fibratus	Fibratus
COMAGNA-24	Not assigned	Intermediate	VAO-44	Fibratus	Fibratus
COMAGNA-25	Not assigned	Intermediate	VAO-45	Porphyrostomus	Porphyrostomus
COMAGNA-26	Not assigned	Intermediate	VAO-46	Porphyrostomus	Porphyrostomus
COMAGNA-27	Not assigned	Intermediate	VAO-47	Porphyrostomus	Porphyrostomus
COMAGNA-28	Porphyrostomus	Porphyrostomus	VAO-48	Porphyrostomus	Porphyrostomus
COMAGNA-29	Not assigned	Intermediate	VAO-49	Porphyrostomus	Porphyrostomus
COMAGNA-3	Porphyrostomus	Intermediate	VAO-5	Fibratus	Fibratus
COMAGNA-30	Not assigned	Intermediate	VAO-50	Porphyrostomus	Porphyrostomus
COMAGNA-31	Not assigned	Intermediate	VAO-51	Porphyrostomus	Porphyrostomus
COMAGNA-32	Not assigned	Intermediate	VAO-52	Porphyrostomus	Porphyrostomus
COMAGNA-33	Porphyrostomus	Porphyrostomus	VAO-53	Porphyrostomus	Porphyrostomus
COMAGNA-34	Not assigned	Intermediate	VAO-54	Porphyrostomus	Porphyrostomus
COMAGNA-35	Not assigned	Intermediate	VAO-6	Fibratus	Fibratus
COMAGNA-36	Porphyrostomus	Porphyrostomus	VAO-7	Fibratus	Fibratus
COMAGNA-4	Porphyrostomus	Porphyrostomus	VAO-8	Fibratus	Fibratus
COMAGNA-5	Not assigned	Intermediate	VAO-9	Fibratus	Fibratus
COMAGNA-6	Porphyrostomus	Intermediate	VAO-A17	Fibratus	Fibratus
COMAGNA-7	Porphyrostomus	Intermediate	VAO-A18	Fibratus	Fibratus

COMAGNA-8	Not assigned	Intermediate
COMAGNA-9	Not assigned	Intermediate
GADJI-1	Fibratus	Fibratus
GADJI-10	Porphyrostomus	Porphyrostomus
GADJI-11	Porphyrostomus	Porphyrostomus
GADJI-12	Porphyrostomus	Porphyrostomus
GADJI-13	Porphyrostomus	Porphyrostomus
GADJI-14	Porphyrostomus	Porphyrostomus
GADJI-15	Fibratus	Fibratus
GADJI-15	Fibratus	Fibratus
GADJI-16	Fibratus	Fibratus
GADJI-17	Fibratus	Fibratus
GADJI-18	Fibratus	Fibratus
GADJI-19	Fibratus	Fibratus
GADJI-2	Fibratus	Fibratus
GADJI-20	Fibratus	Fibratus
GADJI-21	Fibratus	Fibratus
GADJI-22	Fibratus	Fibratus
GADJI-23	Fibratus	Fibratus
GADJI-24	Fibratus	Fibratus
GADJI-25	Fibratus	Fibratus
GADJI-26	Porphyrostomus	Porphyrostomus
GADJI-27	Porphyrostomus	Porphyrostomus
GADJI-28	Fibratus	Fibratus
GADJI-29	Fibratus	Fibratus
GADJI-3	Fibratus	Fibratus
GADJI-30	Fibratus	Fibratus
GADJI-31	Fibratus	Fibratus
GADJI-32	Fibratus	Fibratus
GADJI-33	Fibratus	Fibratus
GADJI-34	Fibratus	Fibratus
GADJI-35	Fibratus	Fibratus
GADJI-36	Fibratus	Fibratus
GADJI-37	Fibratus	Fibratus
GADJI-38	Fibratus	Fibratus
GADJI-39	Porphyrostomus	Porphyrostomus
GADJI-4	Fibratus	Fibratus
GADJI-5	Fibratus	Fibratus
GADJI-7	Fibratus	Fibratus
GADJI-8	Not assigned	Intermediate
VAO-A19	Fibratus	Fibratus
VAO-A20	Fibratus	Fibratus
VAO-A21	Fibratus	Fibratus
VAO-A22	Fibratus	Fibratus
VAO-A23	Fibratus	Fibratus
VAO-A24	Fibratus	Fibratus
VAO-A25	Fibratus	Fibratus
VAO-A26	Fibratus	Fibratus
VAO-A27	Fibratus	Fibratus
VAO-A28	Fibratus	Fibratus
VAO-A29	Fibratus	Fibratus
VAO-A30	Fibratus	Fibratus
VAO-A31	Fibratus	Fibratus
VAO-A32	Fibratus	Fibratus
VAO-A33	Fibratus	Fibratus
VAO-A34	Fibratus	Fibratus
VAO-A35	Fibratus	Fibratus
VAO-A36	Fibratus	Fibratus
WAATCHIA-1	Fibratus	Fibratus
WAATCHIA-10	Fibratus	Fibratus
WAATCHIA-11	Fibratus	Fibratus
WAATCHIA-13	Porphyrostomus	Porphyrostomus
WAATCHIA-14	Porphyrostomus	Porphyrostomus
WAATCHIA-15	Porphyrostomus	Porphyrostomus
WAATCHIA-16	Porphyrostomus	Porphyrostomus
WAATCHIA-17	Porphyrostomus	Porphyrostomus
WAATCHIA-18	Porphyrostomus	Porphyrostomus
WAATCHIA-19	Porphyrostomus	Porphyrostomus
WAATCHIA-2	Not assigned	Intermediate
WAATCHIA-20	Porphyrostomus	Porphyrostomus
WAATCHIA-21	Porphyrostomus	Fibratus
WAATCHIA-21	Fibratus	Fibratus
WAATCHIA-22	Fibratus	Fibratus
WAATCHIA-23	Fibratus	Fibratus
WAATCHIA-24	Fibratus	Fibratus
WAATCHIA-25	Fibratus	Fibratus
WAATCHIA-26	Fibratus	Fibratus
WAATCHIA-27	Fibratus	Fibratus
WAATCHIA-28	Fibratus	Fibratus
WAATCHIA-29	Fibratus	Fibratus

GADJI-9	Porphyrostomus	Porphyrostomus
KERE-1	Fibratus	Fibratus
KERE-10	Fibratus	Fibratus
KERE-11	Fibratus	Fibratus
KERE-12	Not assigned	Fibratus
KERE-13	Porphyrostomus	Intermediate
KERE-14	Porphyrostomus	Porphyrostomus
KERE-15	Porphyrostomus	Porphyrostomus
KERE-16	Porphyrostomus	Porphyrostomus
KERE-17	Fibratus	Fibratus
KERE-18	Fibratus	Fibratus
KERE-19	Fibratus	Fibratus
KERE-2	Fibratus	Fibratus
KERE-20	Fibratus	Fibratus
KERE-21	Fibratus	Fibratus
KERE-22	Fibratus	Fibratus
KERE-23	Fibratus	Fibratus
KERE-24	Fibratus	Fibratus
KERE-25	Fibratus	Fibratus
KERE-26	Fibratus	Fibratus
KERE-27	Porphyrostomus	Porphyrostomus
KERE-28	Porphyrostomus	Porphyrostomus
KERE-29	Porphyrostomus	Porphyrostomus
KERE-3	Fibratus	Fibratus
KERE-30	Not assigned	Porphyrostomus
KERE-31	Fibratus	Fibratus
KERE-32	Fibratus	Fibratus
KERE-33	Fibratus	Fibratus
KERE-34	Fibratus	Fibratus
KERE-35	Fibratus	Fibratus
KERE-36	Fibratus	Fibratus
KERE-37	Fibratus	Fibratus
KERE-38	Fibratus	Fibratus
KERE-39	Fibratus	Fibratus
KERE-4	Fibratus	Fibratus
KERE-40	Fibratus	Fibratus
KERE-5	Fibratus	Fibratus
KERE-6	Fibratus	Fibratus
KERE-7	Fibratus	Fibratus
WAATCHIA-3	Fibratus	Fibratus
WAATCHIA-30	Fibratus	Fibratus
WAATCHIA-32	Fibratus	Fibratus
WAATCHIA-33	Fibratus	Fibratus
WAATCHIA-34	Fibratus	Fibratus
WAATCHIA-35	Fibratus	Fibratus
WAATCHIA-36	Fibratus	Fibratus
WAATCHIA-37	Fibratus	Fibratus
WAATCHIA-38	Fibratus	Fibratus
WAATCHIA-39	Fibratus	Fibratus
WAATCHIA-4	Fibratus	Fibratus
WAATCHIA-40	Fibratus	Fibratus
WAATCHIA-41	Fibratus	Fibratus
WAATCHIA-42	Fibratus	Fibratus
WAATCHIA-43	Fibratus	Fibratus
WAATCHIA-44	Fibratus	Fibratus
WAATCHIA-45	Fibratus	Fibratus
WAATCHIA-46	Fibratus	Fibratus
WAATCHIA-47	Fibratus	Fibratus
WAATCHIA-48	Fibratus	Fibratus
WAATCHIA-49	Fibratus	Fibratus
WAATCHIA-5	Fibratus	Fibratus
WAATCHIA-6	Fibratus	Fibratus
WAATCHIA-7	Fibratus	Fibratus
WAATCHIA-8	Fibratus	Fibratus
WAATCHIA-9	Fibratus	Fibratus
WAPAN-1	Not assigned	Intermediate
WAPAN-10	Fibratus	Fibratus
WAPAN-11	Not assigned	Intermediate
WAPAN-12	Not assigned	Intermediate
WAPAN-13	Not assigned	Intermediate
WAPAN-14	Not assigned	Intermediate
WAPAN-15	Porphyrostomus	Porphyrostomus
WAPAN-16	Porphyrostomus	Porphyrostomus
WAPAN-17	Fibratus	Fibratus
WAPAN-18	Fibratus	Fibratus
WAPAN-19	Fibratus	Fibratus
WAPAN-2	Not assigned	Intermediate
WAPAN-20	Fibratus	Fibratus

KERE-8	Fibratus	Fibratus
KERE-9	Fibratus	Fibratus
TOUETE-1	Fibratus	Fibratus
TOUETE-10	Fibratus	Fibratus
TOUETE-11	Fibratus	Fibratus
TOUETE-12	Fibratus	Fibratus
TOUETE-13	Fibratus	Fibratus
TOUETE-14	Fibratus	Fibratus
TOUETE-15	Fibratus	Fibratus
TOUETE-16	Fibratus	Fibratus
TOUETE-17	Fibratus	Fibratus
TOUETE-18	Fibratus	Fibratus
TOUETE-19	Not assigned	Intermediate
TOUETE-2	Fibratus	Fibratus
TOUETE-20	Porphyrostomus	Porphyrostomus
TOUETE-21	Porphyrostomus	Porphyrostomus
TOUETE-22	Porphyrostomus	Porphyrostomus
TOUETE-23	Porphyrostomus	Porphyrostomus
TOUETE-24	Porphyrostomus	Porphyrostomus
TOUETE-25	Porphyrostomus	Porphyrostomus
TOUETE-26	Porphyrostomus	Porphyrostomus
TOUETE-27	Fibratus	Fibratus
TOUETE-29	Fibratus	Fibratus
TOUETE-3	Fibratus	Fibratus
TOUETE-30	Fibratus	Fibratus
TOUETE-31	Fibratus	Fibratus
TOUETE-32	Fibratus	Fibratus
TOUETE-33	Fibratus	Fibratus
TOUETE-34	Fibratus	Fibratus
TOUETE-35	Fibratus	Fibratus
TOUETE-36	Fibratus	Fibratus
TOUETE-4	Fibratus	Fibratus
TOUETE-6	Fibratus	Fibratus
TOUETE-7	Fibratus	Fibratus
TOUETE-8	Fibratus	Fibratus
TOUETE-9	Fibratus	Fibratus
VAO-1	Not assigned	Intermediate
VAO-11	Not assigned	Fibratus
VAO-12	Not assigned	Intermediate
VAO-13	Porphyrostomus	Porphyrostomus
VAO-14	Porphyrostomus	Porphyrostomus
VAO-15	Porphyrostomus	Porphyrostomus
VAO-16	Porphyrostomus	Porphyrostomus
VAO-18	Fibratus	Fibratus
WAPAN-21	Fibratus	Fibratus
WAPAN-22	Fibratus	Fibratus
WAPAN-23	Fibratus	Fibratus
WAPAN-24	Fibratus	Fibratus
WAPAN-25	Fibratus	Fibratus
WAPAN-26	Fibratus	Fibratus
WAPAN-27	Fibratus	Fibratus
WAPAN-27	Porphyrostomus	Fibratus
WAPAN-29	Not assigned	Intermediate
WAPAN-3	Not assigned	Intermediate
WAPAN-30	Not assigned	Intermediate
WAPAN-31	Fibratus	Fibratus
WAPAN-32	Fibratus	Fibratus
WAPAN-33	Fibratus	Fibratus
WAPAN-34	Fibratus	Fibratus
WAPAN-35	Fibratus	Fibratus
WAPAN-36	Fibratus	Fibratus
WAPAN-37	Not assigned	Intermediate
WAPAN-38	Fibratus	Fibratus
WAPAN-39	Fibratus	Fibratus
WAPAN-4	Not assigned	Intermediate
WAPAN-41	Fibratus	Fibratus
WAPAN-42	Fibratus	Fibratus
WAPAN-43	Porphyrostomus	Porphyrostomus
WAPAN-44	Porphyrostomus	Porphyrostomus
WAPAN-5	Not assigned	Intermediate
WAPAN-6	Not assigned	Intermediate
WAPAN-7	Fibratus	Fibratus
WAPAN-8	Not assigned	Intermediate
WAPAN-9	Not assigned	Intermediate
YOUATY-1	Fibratus	Fibratus
YOUATY-10	Fibratus	Fibratus
YOUATY-11	Fibratus	Fibratus
YOUATY-12	Fibratus	Fibratus
YOUATY-13	Fibratus	Fibratus
YOUATY-16	Fibratus	Fibratus
YOUATY-17	Fibratus	Fibratus
YOUATY-18	Fibratus	Fibratus
YOUATY-19	Fibratus	Fibratus
YOUATY-2	Fibratus	Fibratus
YOUATY-20	Fibratus	Fibratus
YOUATY-21	Fibratus	Fibratus
YOUATY-22	Fibratus	Fibratus
YOUATY-23	Fibratus	Fibratus

VAO-19	Fibratus	Fibratus
VAO-2	Fibratus	Fibratus
VAO-20	Fibratus	Fibratus
VAO-21	Fibratus	Fibratus
VAO-22	Fibratus	Fibratus
VAO-23	Fibratus	Fibratus
VAO-24	Fibratus	Fibratus
VAO-25	Fibratus	Fibratus
VAO-26	Fibratus	Fibratus
VAO-27	Fibratus	Fibratus
VAO-28	Fibratus	Fibratus

YOUATY-24	Fibratus	Fibratus
YOUATY-25	Fibratus	Fibratus
YOUATY-26	Fibratus	Fibratus
YOUATY-27	Fibratus	Fibratus
YOUATY-3	Fibratus	Fibratus
YOUATY-4	Fibratus	Fibratus
YOUATY-5	Fibratus	Fibratus
YOUATY-6	Fibratus	Fibratus
YOUATY-7	Fibratus	Fibratus
YOUATY-8	Fibratus	Fibratus
YOUATY-9	Fibratus	Fibratus

3

Chapter 3: Development of genetic nuclear markers for New Caledonian species of *Placostylus* (*Pulmonata: Bothriembryontidae*)

Abstract

Neutral nuclear markers were developed in two *Placostylus* species. Using genomic libraries primer pairs were designed in microsatellite containing sequences, and in anonymous nuclear loci regions. Four microsatellite markers were successfully developed out of 96 potential primer pair combinations, and two SNPs markers were developed out of 13 potential primer pair combinations. Markers were amplified on population samples of the two species *Placostylus fibratus* and *Placostylus porphyrostomus.* Number of alleles per marker were 2-10 for microsatellite markers and two for the SNPs markers. Amplification success were higher for all markers in the *P. fibratus* population. Departures from Hardy-Weinberg expectations were observed in only one marker for the *P. fibratus* population sample, and for two markers in the *P. porphyrostomus* population sample.

Introduction

Population genetic studies require polymorphic genetic markers. To investigate variation in allele frequency and infer population structure species-specific allelic nuclear markers need to be developed. Traditionally nuclear markers have been used to target a few polymorphic loci, but recently the advancement of high-throughput Next Generation Sequencing (NGS) technologies has made available techniques that allows sampling of hundreds to thousands of nuclear loci across whole genomes (Elshire et al., 2011; Peterson et al., 2012; Rowe et al., 2011). Most of these methods (RAD sequencing, double digest RAD sequencing, Genotyping by sequencing) use a combination of restriction enzymes and NGS techniques to access to a range of Single Nucleotide Polymorphism (SNPs) alleles. Advantages of these techniques are numerous as they provide a huge amount of raw data, but they also come with limitations. They are more expensive than traditional markers, and can be challenging to develop for a non-model organism where no reference genome is available (Lowry et al., 2017). They also require consistently high quality and high quantity DNA extractions for all specimens. As these approaches incorporate a digestion step using restriction-fragment enzyme the DNA must be devoid of enzymatic inhibitors (sometimes a problem for molluscs) and because there is limited DNA amplification in the pipeline the starting quantities required can be prohibitive. Combining new samples into existing analysis is also not straightforward, and because of financial limitations they often end up being made on relatively small sample size (Hollenbeck & Johnston, 2018). Simulation studies have shown that a limited number of highly polymorphic markers can match results obtained with hundreds of SNPs loci, even though the latter generally have more statistical power (Hodel et al., 2016; Lemopoulos et al., 2019; Mesak et al., 2014). Single nucleotide polymorphisms are limited to four states (A, T, G, C) but are often scored as two states. In contrast, in a system where there are numerous possible alleles at each locus fewer loci are required to infer relationships, gene flow, population structure etc. Sometimes the question driving the research can be answered with fewer loci if a specific hypothesis is being tested (Forsdick et al., 2020). In this context, the development and use of more traditional and cost-effective nuclear markers remains a relevant and economic option for non-model organisms where no reference genome sequence available.

Before development of NGS technologies, microsatellites or Single Sequence Repeat (SSRs) markers were arguably the most commonly used allelic nuclear marker in population genetic studies (Cruz et al., 2005; Tóth et al., 2000; Weetman et al., 2005). Microsatellites consist of a tract of repetitive DNA motif, usually two to five base pair long, repeated a variable number of

times. They present a length polymorphism and can be genotyped without using any sequencing technology. Development of such markers requires design of unique oligonucleotide primers in flanking regions of the microsatellite repeat, to enable amplification using Polymerase Chain Reaction (PCR). Traditional methods to develop those primers used microsatellite genomic libraries enrichment and cloning (Daly et al., 2019; Fischer & Bachmann, 1998; Glenn & Schable, 2005), but it is now possible to easily develop sets of microsatellite primers from NGS library datasets (Fernandez-Silva et al., 2013; López-Márquez et al., 2016; Zalapa et al., 2012). NGS library data consist of a large number of short DNA sequences obtained from high throughput technologies. Many of those technologies uses paired-end sequencing, where standardized size reads are obtained from both ends of DNA fragments. Developing microsatellite markers from this kind of data can be done by searching for repeat regions in genomic library sequences, and then designing primers in their flanking regions (Taheri et al., 2018). Designed sets of primers can then be used to amplify microsatellite repeat regions with individuals genotyped based on the length of the amplified DNA fragment.

Anonymous Nuclear Loci (ANL) markers are DNA sequences from non-coding regions of the genome, which are unlikely to be directly under selection (Barrow et al., 2014; Hare et al., 1996; Lemmon & Lemmon, 2012). Like microsatellite markers, it is now relatively easy to extract and produce this kind of marker from NGS library datasets (Ansari et al., 2014; Barrow et al., 2014; Bertozzi et al., 2012), and combined with amplicon sequencing approaches they can provide reliable multi-allelic nuclear genetic marker datasets (Curk et al., 2015). Amplicon sequencing refers to the application of high-throughput sequencing methods to sequence amplified fragments of DNA, usually from PCR (Schirmer et al., 2015; Zhou et al., 2011). Targeted resequencing of specific loci is one application of these methods, still rarely used in molecular ecology. Because high-throughput techniques will sequence both allelic variants of a specific locus separately, they provide a much more accurate and powerful tool than pyro- or sanger-sequencing techniques to identify heterozygote variants. Using sequencing approaches to genotype ANL markers gives direct access to their DNA sequences, and inter-individual analysis of nucleotide variation can be used to identify informative SNP markers within them (Baetscher et al., 2018).

Developing nuclear markers without a reference genome can be a risky and tedious task. As amplified sequences are rarely confirmed to be part of the target organism genome it is important that newly developed markers are shown to behave in a Mendelian way. Both PCR amplification artefacts and amplification from exogenous DNA can be misidentified as

potential markers in developmental stages, and the use of fallacious markers can lead to spurious conclusions in downstream analysis. On a population sample newly developed markers should therefore show convincing evidence that they follow the expectations of population genetics for sexually reproducing diploid organisms. This means allele frequencies should be close to hardy-weinberg proportions in 'ideal' populations and show departures from it when accounting for isolation by distance, population fragmentation or selection (Dupuy et al., 2009; Prebble et al., 2019; Weetman et al., 2005). Most downstream population genetic analysis rely on single-copy allelic markers (Jombart et al., 2010; Pritchard et al., 2000), hence avoiding multi-copy regions is desirable, as non-target copies can interfere with allele calls that can result in markers being ultimately discarded from analysis. Repeat elements (e.g., transposons, satellite DNA) constitute large portions of some genomes leading to both ANL and microsatellite markers being developed in multi-copy regions. Methods to avoid developing markers in these regions include using tools that will compare sequences to databases of repeat elements (Bao et al., 2015), and verifying the assumption of bi-allelicity is respected after sequence amplification (for diploid organisms).

Here I develop both microsatellite and SNPs markers for *Placostylus* species of New Caledonia. *Placostylus* (Beck 1837, Bothriembryontidae, Placostylinae) is a genus of large terrestrial land snail present on islands across the southwest Pacific (Ponder et al., 2003; Trewick et al., 2009). Previous genetic work on the genus in New Caledonia has included shallow phylogenetic signal of sequenced mitochondrial *ND2* gene and population genetic analysis of dd-RAD SNPs data (3754 loci) across the whole archipelago (Dowle et al., 2015). There is currently a strong incentive to have a detailed understanding of *Placostylus* population structure in the Isle of Pines, south of the archipelagos where wild populations of *P. fibratus* are harvested for food and provide local economic activity for indigenous Kanak tribes. Recent morphometric analysis of a large dataset (337 shells) has raised questions about whether there are just two distinct species on the island. The variation in shell size and shape suggested a potential hybrid population of the two sympatric species *P. fibratus* and *P. porphyrostomus* (Quenu et al., 2020). Populations of *P. fibratus* have been recorded to be on the decline and farming production of snails is now considered (Brescia et al., 2008). In this context it is highly relevant to develop and have access to a relatively cheap range of nuclear markers for identification of individual snails. Towards this goal I produce two NGS library datasets from the closely related species *P. fibratus* and *P. porphyrosotomus* and use them to identify both microsatellites and ANL

nuclear markers. I test the loci using amplification of putative population samples of the two species.

Methods

Shotgun illumina sequencing

In order to obtain DNA extractions of sufficient quality for generation of Next generation sequencing libraries (NGS) libraries we used a CTAB and chloroform extraction method on tissue samples from two snail specimens from the two species *P. fibratus* and *P. porphyrostomus*. Small sections of foot muscle, about 50 mg, were cut from specimens using separate sterile scalpel blades. Each tissue was then pressed in a clean paper towel to remove ethanol and diced into smaller pieces. Each sample was incubated at 55°C with proteinase K and CTAB buffer (2% Hexadecyltrimethyl ammonium bromide, 100mM Tris-HCL pH8.0 20mM EDTA) (Thomaz, Guiller & Clarke, 1996; Trewick, Brescia & Jordan, 2009). Following CTAB tissue digestion the solution was treated to purification using chloroform (24:1 chloroform-isoamyl alcohol), and the DNA precipitated with sodium acetate (3 M NaOAC) and -20°C chilled 95% ethanol. DNA extracts were processed through massive parallel, high-throughput sequencing using the ThruPLEX® DNA-seq kit (Rubicon Genomics). Fragmented genomic DNA was pair-end sequenced on an Illumina Hiseq 2500. Reads were de-multiplexed using standard indexes. The resulting 37.6 million sequences (3.7 billion base pairs) were trimmed of adapters and passed through standard quality filters using the software fastp (Chen et al., 2018). Reads were paired in Geneious v8 (Kearse et al., 2012). High copies mitochondrial and ribosomal sequences were removed from paired-end reads datasets by mapping sequences onto mitochondrial genome and ribosomal 45S cassette reconstructions, using the Geneious mapping tool on high sensitivity. Paired end reads from the two species were combined into one dataset.

Microsatellite markers

The software GMATo (Wang et al., 2013) was used to detect microsatellite repeats in libraries of 101bp paired-end sequences. I targeted tri- and tetra- nucleotide repeats, located at the end region of one of the paired reads. To design primers surrounding the microsatelite containing sequence (MCS), I manually concatenated the MCS sequence to the reverse complement of its paired read, and used primer3 (Hancock et al., 2004) to design primers in flanking regions of the microsatellite repeat. In this way 96 pairs of microsatellite primers were designed and synthetised (Supplementary Table 1). To test whether primers produced any amplification products PCR used primer pairs separately on four to eight snail DNA extractions from the two *Placostylus* species (*P. fibratus* and *P. porphyrostomus*). Each PCR consisted of 38 cycles of

denaturation (35 seconds, 94°C), annealing (20 seconds, 5°C lower than primer melting temperature Tm) and elongation (50 seconds, 72°C). PCR products were loaded into a PerkinElmer Labchip GX Touch fragment analyser, and I used the Labchip GX Reviewer software to visualise if consistent peaks across different DNA extractions could be interpreted as amplification of the expected DNA fragments (Figure 1). Primer pairs that amplified consistently were then synthetised with a fluorochrome tag attached to the 5' end of one primer to be subsequently genotyped using the ABI3730 Genetic Analyzer (Table 1). Genescan Liz-500 (Applied Biosystems) was used as an internal size standard, with the microsatellite plugin in Geneious (Kearse et al. 2012) to determine allele size based on fluorescent peaks on microsatellite trace files.

Table 1: Primer sequences and characteristics for 10 potential microsatellite markers, out of 96 initial combinations. Four of these combinations (MS47, MS49, MS64, MS94) successfully amplified on population sample.

Locus name	Primer direction	Primer template	Tm	fluorotag	repeat type	motif
MS02	forward	GCACACATGCCCGGTA	56	6-FAM	tetra	ATGT
MS02	reverse	TGATGTGAGTATGCATTTTGC	52	6-FAM	tetra	ATGT
MS04	forward	TGA GGA ACA ATG TGA TGA ACA	52	VIC	tetra	AGTC
MS04	reverse	GAC CAG ACT GAA AGA ACT TCA	52	VIC	tetra	AGTC
MS10	forward	ACA ACT CGT GCT GTT AGC T	54	VIC	tetra	AGTT
MS10	reverse	CCC ACT ACT TCT CTT GGC GT	57	VIC	tetra	AGTT
MS27	forward	ACT TTC GCT GCG AAG GA	54	TAMRA	tri	GTA
MS27	reverse	TGC ATA CGA GGA AAA TTT CGA	52	TAMRA	tri	GTA
MS47	forward	ACA TGT GGC CAC GAA AAG AA	55	TAMRA	tetra	GGTT
MS47	reverse	GGT GGG TCA GTC GGT TGG	58	TAMRA	tetra	GGTT
MS49	forward	AGA TAG ACG GAC GGA CGG A	57	6-FAM	tetra	GACA
MS49	reverse	TGA CTG TTG GAG ATT GCA GCT	57	6-FAM	tetra	GACA
MS62	forward	AAG AAA GAC AGA CAG GAA GAC	52	6-FAM	tetra	AGAA
MS62	reverse	GGC TCA TTA GGC TGA TTG TTC	54	6-FAM	tetra	AGAA
MS64	forward	TCA ACC TTT CTT TCT GTT GCC	54	VIC	tetra	ATGT
MS64	reverse	ACA TAC ATA CAT GCC TGC CTG	55	VIC	tetra	ATGT
MS70	forward	AGC AGC AGC AGT AGG AGT	56	6-FAM	tetra	GTAG
MS70	reverse	TGG TGG TTT AGG TGG TTG GTT	57	6-FAM	tetra	GTAG
MS94	forward	TGG GTG CTG AGA ATT TCA AAA	53	VIC	tetra	ACTA
MS94	reverse	CCC CAG GGA CAG AGG AGA	58	VIC	tetra	ACTA

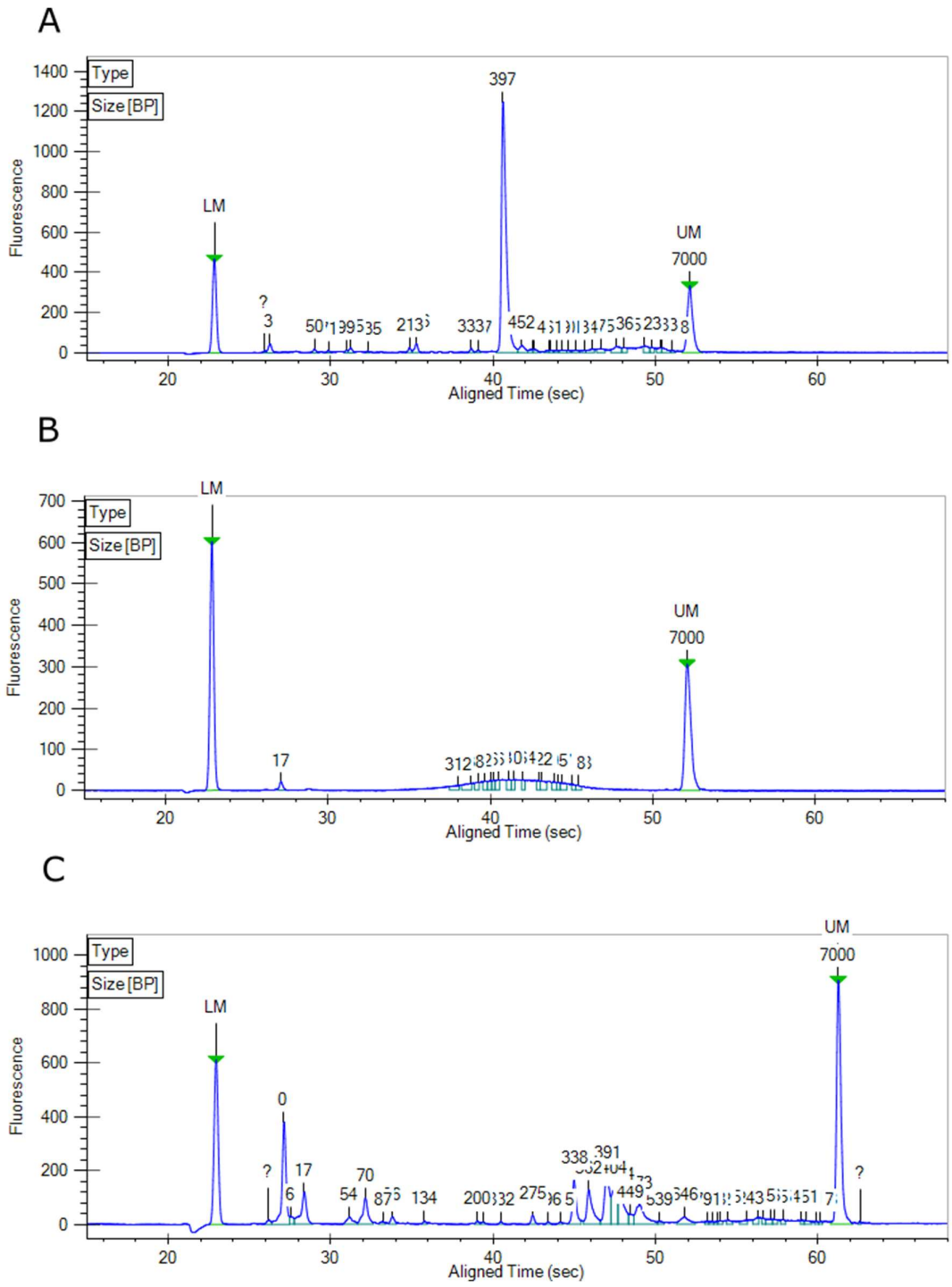


Figure 1: Electropherograms of fluorescence intensity versus time produced by a PerkinElmer Labchip GX Touch fragment analyser, which was used in the development stage of microsatellite markers. Peaks in the electropherograms indicates presence of DNA fragments, which have a length proportional to the aligned time (peaks are labelled in base pair in these representations). A) Typical result after strong amplification of a single fragment of DNA B) Result after no amplification of DNA C) Result after amplification of multiple DNA fragments of variable size.

Single Nucleotide Polymorphism markers

To identify regions of snail genome potentially containing suitable information for population genetic study, I performed a de-novo assembly analysis using the Geneious de novo assembler on the combined 101bp reads dataset. Medium sensitivity was used, without the option to produce scaffolds and saving 1000 contigs. Resulting contigs from de novo assembly were inspected by eye for regions containing nucleotide variants, and I selected a subset of these on which to design primers. Consensus sequences were blasted against Blast nucleotide database (Madden, 2013) and scanned against the transposon database RepBase (Bao et al., 2015). PCR primers were designed for contigs that did not match any known sequences in the blast database and did not contain transposons in their sequences. Primer3 was used for primer design, at first designing between five and ten different sets of primers for each consensus sequence, and then retaining only the primers surrounding regions containing interesting nucleotide variants. Eleven sets of anonymous loci PCR primers were designed overall, and they were synthetised with an overhang Illumina adapter attached to both their 3' and 5' ends (Table 2), so that they could be processed in downstream Illumina MiSeq genotyping analysis. Universal primers targeting the two Internal Transcribed Spacer (ITS) regions of the 45S region with the same overhang sequences attached to both their ends were used as positive controls for amplification, and because ITS sequences can potentially be used as polymorphic nuclear markers, if multi-copy variants are correctly isolated. For genotyping we used a high-throughput amplicon sequencing approach based on the 16S Metagenomic Sequencing Library Preparation protocol, but instead of targeting 16S region we targeted selected ANL regions of snail genome. A first round of PCR reactions was executed using ANL and ITS primers to amplify regions of interest. Each PCR reaction consisted of 21 cycles of denaturation (90 seconds, 94°C), annealing (60 seconds, 61°C) and elongation (120 seconds, 72°C). PCR reactions were not multiplexed but produced separately for each set of primers before being pooled for each individual snail sample. Post PCR products were cleaned using AMPure XP beads. Illumina sequencing adapters and individual dual-index barcodes were attached to amplicon targets, and sequencing of amplicon products used a paired-end 250bp Illumina MiSeq. Raw sequencing reads were gathered in .fastq format and run through fastp for quality check, and getting rid of Illumina adapters (Chen et al., 2018). To estimate how many of the 11 initial ANL markers amplified and got through to the sequencing stage we run library datasets on uSTACKS, a component of the STACKS pipeline (Catchen et al., 2013). This was done to gather sequences into 'stacks' of similar sequences, with the number of resulting stacks would reflect the number of regions of interest correctly amplified and sequenced. Reads of every individual snail were then mapped

to nuclear marker reference sequences from their corresponding library using the Geneious mapping tool with custom high specificity settings. Consensus sequences were generated for each individual from mapped reads, then aligned using the software *MUSCLE* (Edgar, 2004). Results were examined by eye for variation among individuals.

Table 2: Primer sequences and characteristics of 11 combinations of ANL markers and two combinations of universal ITS primers. Two nuclear SNPs markers were derived from theses sets of primers, using an amplicon sequencing genotyping approach.

Primer	Direction	Primer sequence	illumina overhang	Length
302F	forward	GCTAAATATAGGCTCCCTGGTCC	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG	23
302R	reverse	TGACCAGAAAGGACCAGTTAGCC	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG	23
335F	forward	TGGAATCCATAGGGGTCAAAATTCC	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG	25
335R	reverse	ACCAGAAAGGACCAGGGAGC	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG	20
554F	forward	TGGTCAAATTGAGCCGAAAAATGC	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG	26
554R	reverse	TGCATTTTTGAGCTCCATTTGACC	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG	24
578F	forward	TTGGATCCCTGGTCCTTTTTGG	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG	22
578R	reverse	GGACCAGGGAGCCTATATTTAGCC	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG	24
727F	forward	TTTGGAATCCATAGGGGTCAAAATTCC	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG	27
727R	reverse	GGACCAGGGAGCCTATATTTAGCC	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG	24
753F	forward	TGGAATTTTGACCCCTATGGATTCC	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG	25
753R	reverse	TAGGCTCCCTGGTCCTTTCTGG	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG	22
782F	forward	TGTATCACCAGAAAGGACCAGGG	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG	23
782R	reverse	CTAATTATAGGCTCCCTGGTCGC	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG	23
863F	forward	TGGAATCCATAGGGGTCAAAATTCC	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG	25
863R	reverse	ACCAGAAAGGACCAGGGAGC	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG	20
922F	forward	GTTTACGACCAGAAAGGAACAGGG	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG	24
922R	reverse	TAGGCTCCCTGGACCTTTCTGG	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG	22
930F	forward	TGGAATTTTGACCCCTATGGATTCC	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG	25
930R	reverse	ATAGGCTCCCTGGTCCTTTCTGG	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG	23
981F	forward	TGGAATTTTGACCCCTATGGATTGC	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG	25
981R	reverse	TAGGCTCCCTGGTCCTTTCTGG	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG	22
ITS5	forward	GGAAGTAAAAGTCGTAACAAGG	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG	22
ITS2	reverse	GCTGCGTTCTTCATCGATGC	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG	20
ITS3	forward	GCATCGATGAAGAACGCAGC	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG	20
ITS4	reverse	TCCTCCGCTTATTGATATGC	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG	20

Test of nuclear markers on Placostylus populations

To test whether developed nuclear genetic markers follow expected population genetic rules, we extracted and amplified DNA from snail populations of *P. fibratus* and *P. porphyrostomus.* For microsatellite markers I collected and amplified DNA fragments from 25 *P. fibratus* snails

at three locations in the kere tribe in the Isle of Pines, south of the New Caledonian archipelago. *P. porphyrostomus* snails were collected all across the island, at 11 different locations. The DNeasy blood and tissue kit was used to extract DNA from foot tissue muscle. I followed the standard tissue protocol with slight modifications, incubating about 50mg of foot tissue sample overnight, and using a double AW1 buffer wash. PCR reactions were used to amplify DNA fragments on all DNA extractions using newly developed microsatellites sets of primers. Allele calls were made using the Geneious microsatellite plugin tool. For ANL markers DNA was amplified from 20 *P.fibratus* of the kere tribe, and 26 *P.porphyrostomus* collected all across the Isle of Pines. The same DNA extracts were used as in the microsatellite protocol, and genotyping was done using the amplicon sequencing method described above. Summary statistics of each population were gathered using the adegenet v2.0.0 package (Jombart, 2015) in R v3.5.2 (R Development Core Team & R Core Team, 2017). I compared allele frequencies of each marker to Hardy-Weinberg expectations using both a χ^2 test and an exact test based on Monte Carlo permutation of alleles (Guo & Thompson, 1992). Observed heterozygosity levels were compared to heterozygosity expected in Hardy-Weinberg equilibrium, to understand if departures from HW proportions could be explained by excess or deficit of heterozygotes (Figure 2).

Results

Microsatellite markers

A total of 96 different combinations of microsatellite primers were tested here (supplementary table 1). Based on PCR amplification and fragment analyser diagnostic ten (10.41%) of these primer pairs gave consistent amplification across four to eight individuals (Table 1). When amplifying on larger sample size (two population from two different species N=25 and 26) and using primers with fluorophore labels only four (4.16%) of those primers gave amplification consistent with multi-allelic nuclear marker.

The four microsatellite loci, named respectively MS47, MS49, MS64 and MS94, contained between 10 and 2 alleles (Table 3). Amplification success for the 25 *P. fibratus* specimens ranged from 72-100% but for *P. porphyrostomus* many samples failed to amplify (39-77% success). None of the four loci demonstrated significant departure from expected Hardy-Weinberg proportions within the population sample of *P. fibratus*, (exact test; Table 3). However, according to a χ^2 test, genotype frequencies of one of these markers (MS47) was found to differ significantly from Hardy-Weinberg expectations, associated with a slight deficit in heterozygotes. In contrast, the 26 *P. porphyrostomus* samples from across the Isle of Pines demonstrated significant departures from Hardy-Weinberg proportions at two microsatellite loci (MS49, MS64), according to both χ^2 test and exact proportion test (e.g. Table 3). All these departures from HW were associated with deficits with the observed proportions of heterozygotes.

Single Nucleotide Polymorphisms markers

A total of 11 ANL and two ITS loci were incorporated in development of neutral nuclear SNP markers (table 2). When running uSTACK on genomic library datasets produced from pooled PCR products for each individual four significant stacks of reads were identified. Alignment of DNA sequences confirmed two of these stacks to be the ITS 1 and 2 regions that were amplified from sets of universal primers, and two other stacks correspond to targeted sequences from ANL sets of primers. A posteriori sequence alignment revealed that the initial 11 sequences chosen for ANL marker development clustered and aligned into two groups of sequences. For the *P. fibratus* and *P. porphyrostomus* populations examined no nucleotide variation was found in the ITS 2 region. One of the two ANL sequences displays more than two intra-individual sequences variants and was discarded. Two significantly different variants were identified for

the ITS1 region. Using high specificity mapping reads of separate variants could be partitioned and one SNP marker could be scored for one of the variants (called ITSv1). The second ANL sequence detected in *ustacks* analysis displayed one SNP position which was scored and kept for further analysis (called ANL1 from here on).

The two SNPs markers identified from amplicon sequencing analysis of neutral nuclear markers contained two alleles in the two *Placostylus* populations screened here. Amplification success for both markers range from 76-88% in the two populations. ANL1 is fixed in the *P.fibratus* population but the *P.porphyrostomus* population contains two alleles. ITSv1 showed some significant departure from Hardy-Weinberg proportions in the *P. fibratus* population sample, which can be explained by an excess in heterozygotes for that population (e. g. table 3). No departure from Hardy-Weinberg proportions was detected in either marker in the *P. porphyrostomus* population(s).

Table 3: Summary statistics and departure from Hardy-Weinberg test values for four microsatellite loci and two neutral nuclear markers, in two *Placostylus* snail population samples (*P. fibratus* and *P. porphyrostomus*). P-value for HW test calculated according to χ^2 test and the HW exact test, obtained through MCMC simulation.

P. fibratus	**N=25**							
Locus	**Nb alleles**	**% amplification**	**Hexp**	**Hobs**	**χ^2**	**df**	**χ2 p-value**	**MCMC p-value**
MS47	7	80	0.75	0.65	36,29	21	0,02*	0,122
MS49	2	100	0.24	0.28	0,66	1	0,41	1
MS64	5	96	0.57	0.59	10,71	10	0,38	0,293
MS94	10	72	0.85	0.89	57,02	45	0,10	0,119
ITSv1	2	76	0.45	0.68	5.13	1	0.023*	0.044*
ANL1	2	76	0	0	0	0	1	1
P.porphyrostomus	**N=26**							
Locus	**Nb alleles**	**% amplification**	**Hexp**	**Hobs**	**χ^2**	**df**	**χ2 p-value**	**MCMC p-value**
MS47	5	62	0.71	0.6	10,49	10	0,39	0,168
MS49	2	42	0.36	0.07	9,93	1	0,002**	0,006**
MS64	9	77	0.81	0.63	55,18	36	0,021*	0,044*
MS94	4	42	0.82	0.63	71,37	36	0.65	0.106
ITSv1	2	88	0.08	0.09	0.04	1	0.82	1
ANL1	2	84	0.44	0.43	1.45	1	0.23	0.35

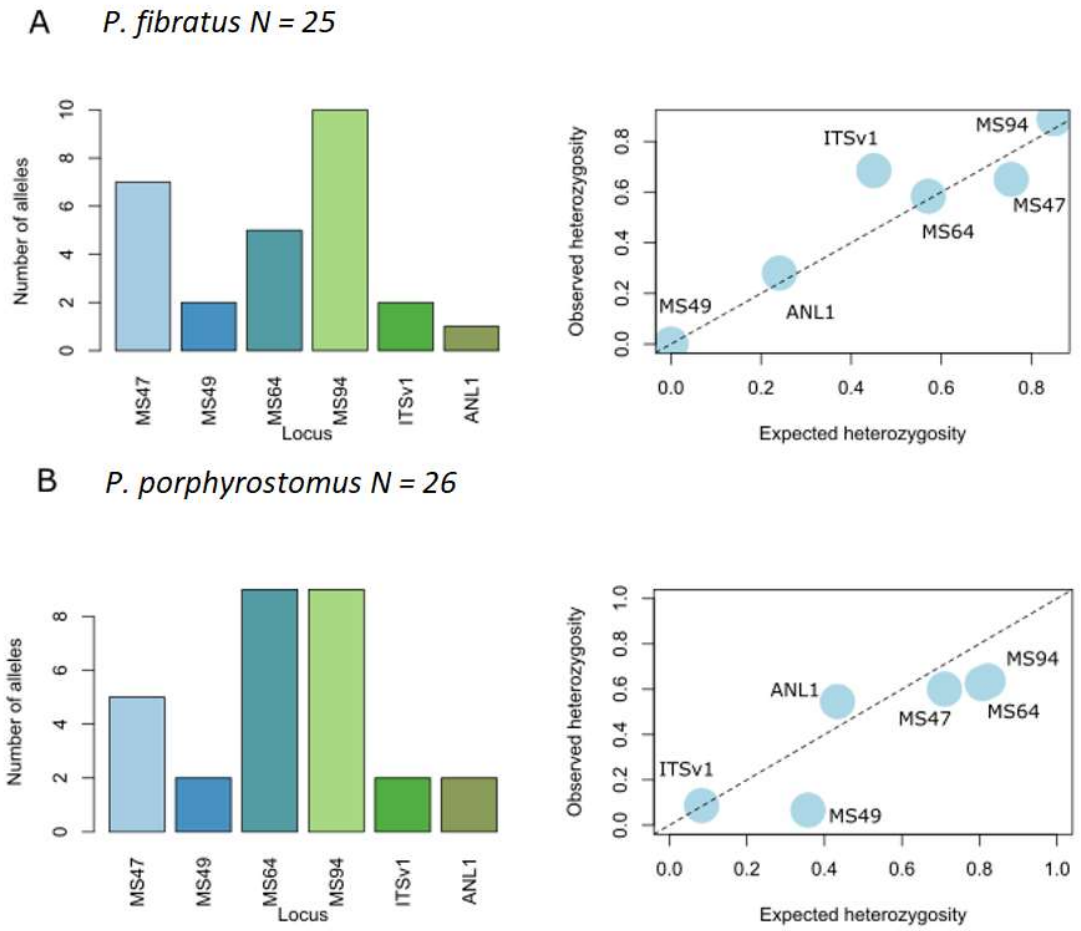


Figure 2: Number of Alleles per microsatellite locus and observed vs expected heterozygosity per microsatellite and neutral nuclear marker for two population samples of the *Placostylus* species *P.fibratus* (A) and *P.porphyrostomus* (B).

Discussion

Four microsatellites were able to be developed out of 96 initial potential combinations (4.2%), and two SNPs markers were developed out of 13 pairs of primers targeting nuclear sequences (15%). Those results are reflective of a very low success rate from a labour-intensive and time-consuming process. Development of genetic nuclear markers has been reported to be notoriously challenging for mollusc organisms, for a number of reasons. One of them is the presence of polysaccharide inhibitors in mollusc soft tissues that interfere with enzyme activity, making PCR-based amplification of DNA tedious (Sokolov, 2000). This can lead to systemic low success of amplification, particularly for DNA extractions that are not high quality. Here the extraction method used to generate shotgun sequencing data (CTAB extraction method) differed from the extraction method used to amplify population samples (spin-column extraction method). This shift in DNA quality could explain why some of the markers identified from shotgun sequencing data did not amplify consistently in population samples.

Genomic libraries were not depleted in microsatellite motifs and primer pairs could be designed out of an abundance of Microsatellite Containing Sequences. Issues in development of microsatellite markers were related to either non-amplification of DNA fragments or amplification of multiple DNA fragments that made identification of the original marker impossible (Figure 1). Non-amplification of potential markers could be the consequence of either inhibitors interfering with PCR enzyme reaction or non-annealing of primers to template DNA. Amplification of multiple DNA fragments reflects a lack of specificity for the designed set of primers, which in this situation would bind and amplify multiple loci of the snail genome. The fact that both situations have been encountered makes it hard to point at a specific direction for improvement of the present microsatellite development method. Primer pairs were designed in each of the paired-end reads surrounding a microsatellite motif, as an attempt to counter the small size of genomic reads data (101bp). This means part of each microsatellite marker sequence was unknown and size of the markers could only be approximated before amplification. Designing both primers from a single read could potentially improve developmental success rate but would require generating longer initial genomic reads data.

Anonymous nuclear loci sequences were reconstructed from the same genomic data that was used in microsatellite primer development. They were developed on contig sequences resulting from a de-novo assembly of 101bp paired-end reads, which was aimed at reconstructing random regions of the snail's genome. Two main issues were encountered with this approach: the sampling of one multi-copy region, and the presence of redundant contigs. Redundant contigs

are homologous contigs that arise from the reconstruction of a same genomic region multiple times within the same de novo assembly (Feldmeyer et al., 2011; Ono et al., 2015). A posteriori sequence alignment revealed that contigs that were selected from de novo assemblies were indeed redundant, which limited our potential number of markers from eleven to only two. Of these two one was found to contain reads representing more than two variants, most likely a multi-copy region. The other marker could be amplified from population samples and one SNP position was scored. Including universal ITS primers in the analysis brought the number of developed neutral nuclear marker to two, both constituting SNP data. Here the low developmental success rate can be linked to imperfections in the method used, and improvements can surely be made. Including a step to remove redundant contigs when screening for potential ANL markers can be added, using software like *simplifier* (Ramos et al., 2012). Avoiding multi-copy regions of the genome might be more difficult, as these regions tend to be overrepresented in genomic reads datasets. Checking assumption of bi-allelicity in reads mapped to ANL contigs could help avoiding sampling some of those regions, but multiple variants might not always appear on shotgun sequencing data.

Overall, I found that both types of markers (microsatellite and ANL) could be developed from NGS sequences data despite the low success rates and developmental difficulties encountered. When amplifying markers on population samples the six nuclear markers showed behaviour that was consistent with expectations of population genetic for single copy bi-allelic nuclear marker. For the Kere *P. fibratus* population sample that came from a single location allelic frequencies were not shown to be outside of Hardy-Weinberg expectations except for ITSv1 which demonstrated an excess of heterozygotes. This is what is to be expected for an 'ideal' population, where effects of isolation by distance, population structure are minimal. On the other hand the sample of *P. porpyhrostomus* snails showed departures from Hardy-Weinberg allele proportions for two loci (MS49, MS64), which were both linked to a deficit in heterozygotes (e.g. table 3). *P. porhyrostmus* were collected in different locations across the isle of Pines and it is very likely that effects of isolation by distance and population structure would be observed at this level, which would then explain deficits in heterozygotes (via Wahlund effect). The presence of null alleles, indicated by low amplification success, is also likely to be contributing to deviations from HW.

On the face of the struggles encountered here it seems important to emphasize that future development of genetic markers should be done in parallel to the development of new genomic resources for this group of organisms. Producing a *Placostylus* genome would provide a

reference which could be used for easier and more consistent development of nuclear markers. Here the genomic data used for identifying nuclear markers consisted of 101bp paired read data and constituted two datasets of 1.8 Gb (*P. porphyrostomus*) and 1.9 Gb (*P. fibratus*). These datasets were used previously in reconstruction of mitochondrial genomes and nuclear multi-copy regions (Chapter 1). For comparison the genome size of the giant African snail *Achatinella fulica,* is around 2.12 Gb (Y. Guo et al., 2019), and preliminary flow cytometry work on *Placostylus* has estimated a genome size between 2 and 3 Gb. Generating even a draft genome for *Placostylus* species would therefore require much bigger genomic dataset to achieve reliable coverage, using a mix of short and long reads datasets technologies. Because it is an abundant species, which is harvested for food and could potentially be farmed in a near future *P. fibratus* seems like an appropriate choice to develop nuclear genome data in this group. In addition to provide a reference for neutral nuclear markers development (Microsatellites, ANL markers, Genotyping by Sequencing …) having a full genome could open new research possibilities, such as enabling the search of Quantitative Traits Loci (QTL) linked to shell morphology. Research related to other species of *Placostylus* in New Zealand and Pacific islands would also benefit from such ressources.

References

Ansari, T. H., Bertozzi, T., Hacking, J., Cooper, S. J. B., & Gardner, M. G. (2014). Random non-coding fragments of lizard DNA: Anonymous nuclear loci for the Australian skink, *Tiliqua rugosa*, and their utility in other Egernia-group species. *Australian Journal of Zoology*, *62*(6), 515–518. https://doi.org/10.1071/ZO14085

Baetscher, D. S., Clemento, A. J., Ng, T. C., Anderson, E. C., & Garza, J. C. (2018). Microhaplotypes provide increased power from short-read DNA sequences for relationship inference. *Molecular Ecology Resources*, *18*(2), 296–305. https://doi.org/10.1111/1755-0998.12737

Bao, W., Kojima, K. K., & Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA*, *6*(1). https://doi.org/10.1186/s13100-015-0041-9

Barrow, L. N., Ralicki, H. F., Emme, S. A., & Lemmon, E. M. (2014). Species tree estimation of North American chorus frogs (Hylidae: *Pseudacris*) with parallel tagged amplicon sequencing. *Molecular Phylogenetics and Evolution*, *75*(1), 78–90. https://doi.org/10.1016/j.ympev.2014.02.007

Bertozzi, T., Sanders, K. L., Sistrom, M. J., & Gardner, M. G. (2012). Anonymous nuclear loci in non-model organisms: Making the most of high-throughput genome surveys. *Bioinformatics*, *28*(14), 1807–1810. https://doi.org/10.1093/bioinformatics/bts284

Brescia, F. M., Pöllabauer, C. M., Potter, M. A., & Robertson, A. W. (2008). A review of the ecology and conservation of *Placostylus* (Mollusca: Gastropoda: Bulimulidae) in New Caledonia. *Molluscan Research*, *28*(2), 111–122.

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124–3140. https://doi.org/10.1111/mec.12354

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*(17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Cruz, F., Pérez, M., & Presa, P. (2005). Distribution and abundance of microsatellites in the genome of bivalves. *Gene*, *346*, 241–247. https://doi.org/10.1016/j.gene.2004.11.013

Curk, F., Ancillo, G., Ollitrault, F., Perrier, X., Jacquemoud-Collet, J. P., Garcia-Lor, A., Navarro, L., & Ollitrault, P. (2015). Nuclear species-diagnostic SNP markers mined from 454 amplicon sequencing reveal admixture genomic structure of modern Citrus varieties. *PLoS ONE*, *10*(5). https://doi.org/10.1371/journal.pone.0125628

Daly, E. E., Walker, K. J., Morgan-Richards, M., & Trewick, S. A. (2019). Spatial genetics of a high elevation lineage of Rhytididae land snails in New Zealand: the *Powelliphanta Kawatiri* complex. *Molluscan Research*, *39*(3), 280–289. https://doi.org/10.1080/13235818.2018.1559914

Dowle, E. J., Morgan-Richards, M., Brescia, F., & Trewick, S. A. (2015). Correlation between shell phenotype and local environment suggests a role for natural selection in the evolution of Placostylus snails. *Molecular Ecology*, *24*(16), 4205–4221. https://doi.org/10.1111/mec.13302

Dupuy, V., Nicot, A., Jarne, P., & David, P. (2009). Development of 10 microsatellite loci in the pulmonate snail *Biomphalaria kuhniana* (Mollusca, Gastropoda). *Molecular Ecology Resources*, *9*(1), 255–257. https://doi.org/10.1111/j.1755-0998.2008.02433.x

Edgar, R. C. (2004). MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, *5*. https://doi.org/10.1186/1471-2105-5-113

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, *6*(5). https://doi.org/10.1371/journal.pone.0019379

Feldmeyer, B., Wheat, C. W., Krezdorn, N., Rotter, B., & Pfenninger, M. (2011). Short read illumina data for the de novo assembly of a non-model snail species transcriptome (radix balthica, basommatophora, pulmonata), and a comparison of assembler performance. *BMC Genomics*, *12*. https://doi.org/10.1186/1471-2164-12-317

Fernandez-Silva, I., Whitney, J., Wainwright, B., Andrews, K. R., Ylitalo-Ward, H., Bowen, B. W., Toonen, R. J., Goetze, E., & Karl, S. A. (2013). Microsatellites for Next-Generation Ecologists: A Post-Sequencing Bioinformatics Pipeline. *PLoS ONE*, *8*(2). https://doi.org/10.1371/journal.pone.0055990

Fischer, D., & Bachmann, K. (1998). Microsatellite enrichment in organisms with large genomes (Allium cepa L.). *BioTechniques*, *24*(5), 796–802. https://doi.org/10.2144/98245st03

Forsdick, N. J., Martini, D., Brown, L., Maloney, R. F., Steeves, T. E., & Knapp, M. (2020). Genomic sequencing confirms absence of introgression despite past hybridisation in a critically endangered bird. *BioRxiv*.

Glenn, T. C., & Schable, N. A. (2005). Isolating microsatellite DNA loci. *Methods in Enzymology*, *395*, 202–222. https://doi.org/10.1016/S0076-6879(05)95013-1

Guo, S. W., & Thompson, E. A. (1992). Performing the Exact Test of Hardy-Weinberg Proportion for Multiple Alleles. *Biometrics*, *48*(2), 361. https://doi.org/10.2307/2532296

Guo, Y., Zhang, Y., Liu, Q., Huang, Y., Mao, G., Yue, Z., Abe, E. M., Li, J., Wu, Z., Li, S., Zhou, X., Hu, W., & Xiao, N. (2019). A chromosomal-level genome assembly for the giant African snail *Achatina fulica*. *GigaScience*, *8*(10). https://doi.org/10.1093/gigascience/giz124

Hancock, J. M., Zvelebil, M. J., & Hancock, J. M. (2004). Primer3. In *Dictionary of Bioinformatics and Computational Biology*. https://doi.org/10.1002/9780471650126.dob0560.pub2

Hare, M. P., Karl, S. A., & Avise, J. C. (1996). Anonymous nuclear DNA markers in the American oyster and their implications for the heterozygote deficiency phenomenon in marine bivalves. *Molecular Biology and Evolution*, *13*(2), 334–345. https://doi.org/10.1093/oxfordjournals.molbev.a025593

Hodel, R. G. J., Segovia-Salcedo, M. C., Landis, J. B., Crowl, A. A., Sun, M., Liu, X., Gitzendanner, M. A., Douglas, N. A., Germain-Aubrey, C. C., Chen, S., Soltis, D. E., & Soltis, P. S. (2016). The Report of My Death was an Exaggeration: A Review for Researchers Using Microsatellites in the 21st Century. *Applications in Plant Sciences*, *4*(6), 1600025. https://doi.org/10.3732/apps.1600025

Hollenbeck, C. M., & Johnston, I. A. (2018). Genomic tools and selective breeding in molluscs.

In *Frontiers in Genetics* (Vol. 9, Issue JUL). https://doi.org/10.3389/fgene.2018.00253

Jombart, T. (2015). An introduction to adegenet 2.0.0. *R Package*.

Jombart, T., Devillard, S., Balloux, F., Falush, D., Stephens, M., Pritchard, J., Pritchard, J., Stephens, M., Donnelly, P., Corander, J., Waldmann, P., Sillanpaa, M., Tang, J., Hanage, W., Fraser, C., Corander, J., Lee, C., Abdool, A., Huang, C., … Nei, M. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, *11*(1), 94. https://doi.org/10.1186/1471-2156-11-94

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., & Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), 1647–1649. https://doi.org/10.1093/bioinformatics/bts199

Lemmon, A. R., & Lemmon, E. M. (2012). High-throughput identification of informative nuclear loci for shallow-scale phylogenetics and phylogeography. *Systematic Biology*, *61*(5), 745–761. https://doi.org/10.1093/sysbio/sys051

Lemopoulos, A., Prokkola, J. M., Uusi-Heikkilä, S., Vasemägi, A., Huusko, A., Hyvärinen, P., Koljonen, M. L., Koskiniemi, J., & Vainikka, A. (2019). Comparing RADseq and microsatellites for estimating genetic diversity and relatedness — Implications for brown trout conservation. *Ecology and Evolution*, *9*(4), 2106–2120. https://doi.org/10.1002/ece3.4905

López-Márquez, V., García-Jiménez, R., Templado, J., & Machordom, A. (2016). Development and characterization of 26 novel microsatellite loci for the trochid gastropod *Gibbula divaricata* (Linnaeus, 1758), using Illumina MiSeq next generation sequencing technology. *PeerJ*, *2016*(3). https://doi.org/10.7717/peerj.1789

Lowry, D. B., Hoban, S., Kelley, J. L., Lotterhos, K. E., Reed, L. K., Antolin, M. F., & Storfer, A. (2017). Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. *Molecular Ecology Resources*, *17*(2), 142–152. https://doi.org/10.1111/1755-0998.12635

Madden, T. (2013). The BLAST sequence analysis tool. *The BLAST Sequence Analysis Tool*, 1–17. http://www.ncbi.nlm.nih.gov/books/NBK153387/

Mesak, F., Tatarenkov, A., Earley, R. L., & Avise, J. C. (2014). Hundreds of SNPs vs. dozens of SSRs: Which dataset better characterizes natural clonal lineages in a self-fertilizing fish? *Frontiers in Ecology and Evolution*, *2*(NOV). https://doi.org/10.3389/fevo.2014.00074

Ono, H., Ishii, K., Kozaki, T., Ogiwara, I., Kanekatsu, M., & Yamada, T. (2015). Removal of redundant contigs from de novo RNA-Seq assemblies via homology search improves accurate detection of differentially expressed genes. *BMC Genomics*, *16*(1). https://doi.org/10.1186/s12864-015-2247-0

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE*, *7*(5). https://doi.org/10.1371/journal.pone.0037135

Ponder, W. F., Colgan, D. J., Gleeson, D. M., & Sherley, G. H. (2003). Relationships of

Placostylus from Lord Howe Island: An investigation using the mitochondrial cytochrome c oxidase 1 gene. *Molluscan Research*, *23*(2), 159–178. https://doi.org/10.1071/MR03001

Prebble, J. M., Meudt, H. M., Tate, J. A., & Symonds, V. V. (2019). Comparing and co-analysing microsatellite and morphological data for species delimitation in the New Zealand native *Myosotis pygmaea* species group (Boraginaceae). *Taxon*. https://doi.org/10.1002/tax.12096

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959. https://doi.org/10.1111/j.1471-8286.2007.01758.x

Quenu, M., Trewick, S. A., Brescia, F., & Morgan-Richards, M. (2020). Geometric morphometrics and machine learning challenge currently accepted species limits of the land snail *Placostylus* (Pulmonata: Bothriembryontidae) on the Isle of Pines, New Caledonia. *Journal of Molluscan Studies*, *86*(1), 35–41. https://doi.org/10.1093/mollus/eyz031

R Development Core Team, R., & R Core Team. (2017). R: A language and environment for statistical computing. *R: A Language and Environment for Statistical Computing*. https://doi.org/10.1016/j.jssas.2015.06.002

Ramos, R. T. J., Carneiro, A. R., Azevedo, V., Schneider, M. P., Barh, D., & Silva, A. (2012). Simplifier: a web tool to eliminate redundant NGS contigs. *Bioinformation*, *8*(20), 996–999. https://doi.org/10.6026/97320630008996

Rowe, H. C., Renaut, S., & Guggisberg, A. (2011). RAD in the realm of next-generation sequencing technologies. In *Molecular Ecology* (Vol. 20, Issue 17, pp. 3499–3502). https://doi.org/10.1111/j.1365-294X.2011.05197.x

Schirmer, M., Ijaz, U. Z., D'Amore, R., Hall, N., Sloan, W. T., & Quince, C. (2015). Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Research*, *43*(6). https://doi.org/10.1093/nar/gku1341

Sokolov, E. P. (2000). An improved method for DNA isolation from mucopolysaccharide-rich molluscan tissues. *Journal of Molluscan Studies*, *66*(4), 573–575. https://doi.org/10.1093/mollus/66.4.573

Taheri, S., Abdullah, T. L., Yusop, M. R., Hanafi, M. M., Sahebi, M., Azizi, P., & Shamshiri, R. R. (2018). Mining and development of novel SSR markers using Next Generation Sequencing (NGS) data in plants. In *Molecules* (Vol. 23, Issue 2). https://doi.org/10.3390/molecules23020399

Thomaz, D., Guiller, A., & Clarke, B. (1996). Extreme divergence of mitochondrial DNA within species of pulmonate land snails. *Proceedings of the Royal Society B: Biological Sciences*, *263*(1368), 363–368. https://doi.org/10.1098/rspb.1996.0056

Tóth, G., Gáspári, Z., & Jurka, J. (2000). Microsatellites in different eukaryotic genomes: Surveys and analysis. *Genome Research*, *10*(7), 967–981. https://doi.org/10.1101/gr.10.7.967

Trewick, S., Brescia, F., & Jordan, C. (2009). Diversity and phylogeny of New Caledonian *Placostylus* land snails; evidence from mitochondrial DNA. *Memoires Du Museum National d'Histoire Naturelle*, *198*, 421–436.

Wang, X., Lu, P., & Luo, Z. (2013). GMATo : A novel tool for the identification and analysis

of microsatellites in large genomes. *Bioinformation*, *9*(10), 541–544. https://doi.org/10.6026/97320630009541

Weetman, D., Hauser, L., Shaw, P. W., & Bayes, M. K. (2005). Microsatellite markers for the whelk *Buccinum undatum*. *Molecular Ecology Notes*, *5*(2), 361–362. https://doi.org/10.1111/j.1471-8286.2005.00926.x

Zalapa, J. E., Cuevas, H., Zhu, H., Steffan, S., Senalik, D., Zeldin, E., McCown, B., Harbut, R., & Simon, P. (2012). Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. In *American Journal of Botany* (Vol. 99, Issue 2, pp. 193–208). https://doi.org/10.3732/ajb.1100394

Zhou, J., Wu, L., Deng, Y., Zhi, X., Jiang, Y. H., Tu, Q., Xie, J., Van Nostrand, J. D., He, Z., & Yang, Y. (2011). Reproducibility and quantitation of amplicon sequencing-based detection. *ISME Journal*, *5*(8), 1303–1313. https://doi.org/10.1038/ismej.2011.11

Supplementary

Supplementary Table 1: Primer sequences and characteristics of 96 pairs of potential microsatellite primers designed around microsatellite containing sequences (MCS). Four of these pairs of primers (MS47, MS49, MS64, MS94) successfully amplified on population sample.

Primer pair	Direction	Sequence	Length	Tm
MS01	forward	CTGTCTGTCTGTCTGTCTGTC	21	58.2
MS01	reverse	ACACCGACAGATAGACAGACA	21	57.7
MS02	forward	GCACACATGCCCGGTA	16	55.2
MS02	reverse	TGATGTGAGTATGCATTTTGC	21	56.6
MS03	forward	ACCTACCTACCTACCTACCT	20	55.2
MS03	reverse	ACAGACAGACAGACAGATAGA	21	55.1
MS04	forward	TGAGGAACAATGTGATGAACA	21	55.5
MS04	reverse	GACCAGACTGAAAGAACTTCA	21	55.4
MS05	forward	CCGCGTGTAATGATGATGA	19	59.5
MS05	reverse	GCGAGTGTTCTCAAGCTGTG	20	55.5
MS06	forward	GTGTCCGATTGCAATGATGA	20	55.1
MS06	reverse	TTACTGATACTCGCAAGACG	20	56.5
MS07	forward	TGGTTGGTTGGTTGGTTGGT	20	59
MS07	reverse	CAAGAACCAACCTGCCAATGA	21	60.3
MS08	forward	AGACAGACAGACAGACAGACA	21	60
MS08	reverse	CCTCAGTTTCGTTGTTGCGG	20	58.1
MS09	forward	AACTCGGCGATGCGGA	16	58.4
MS09	reverse	GCCAGGAATGGACAGATACAC	21	58.6
MS10	forward	CCCACTACTTCTCTTGGCGT	20	57
MS10	reverse	ACAACTCGTGCTGTTAGCT	19	59.4
MS11	forward	TGGATGGACGAAAGGACGG	20	57.9
MS11	reverse	TCCATCCATCCATCCATCCA	20	59.4
MS12	forward	TGTCTAGTCTGTCTGTCTGTG	21	59.7
MS12	reverse	AGCAGAAAGACGGGCGG	17	56.5
MS13	forward	AAGCGTCGCCACAACGA	17	57.3
MS13	reverse	AACTGCGAGCGCTAGC	16	59.9
MS14	forward	AGATGCAGCAGAAGAACCCA	20	56.3
MS14	reverse	AGACAAGACAAGACAGACAGA	21	59.3
MS15	forward	ACATCTAGTCATTGCGCCA	19	55.3
MS15	reverse	ACTCACCATACATTCAGTCAC	21	56.8
MS16	forward	ATACCCAGACAGACAGACAGA	21	55.1
MS16	reverse	ACCTACCTACCTACCTACCT	20	57.5
MS17	forward	TGTTCTTTTTCTCAAGTCCCA	21	58.8
MS17	reverse	AAACATCGGTTTATGTGGCGT	21	55.3
MS18	forward	ACCTACCTACCTACCTACCT	20	59.8
MS18	reverse	GCAGATCCGTGTCTGTTCCT	20	55.1
MS19	forward	ACAGACAGACAGACAGACAGA	21	56.5
MS19	reverse	AACTCAACGGCGGCAT	16	58.1
MS20	forward	TGCCATTCCAACAATCACA	19	59.7

Primer pair	Direction	Sequence	Length	Tm
MS49	forward	AGATAGACGGACGGACGGA	19	59.9
MS49	reverse	TGACTGTTGGAGATTGCAGCT	21	59.5
MS50	forward	GCACGGATTCATGGACGGA	19	58.5
MS50	reverse	GCATCTGCGTCTGTGGTC	18	60.2
MS51	forward	TGATGCACATGGAGCAAAA	19	58.1
MS51	reverse	AGACAGACAGACAGACAGACA	21	55.7
MS52	forward	AGGGAGATCTACCACCACCA	20	59.7
MS52	reverse	GCAGGCGTACAGACAGACA	19	59.3
MS53	forward	GTTGTAAGTTGGGAGCAACT	20	60.2
MS53	reverse	GCCCTGCGTAGTTATGCTCA	20	55.9
MS54	forward	ACTTACGAGAAGAACAGGCCG	21	57.2
MS54	reverse	TCTTTTTCCTCTCCCTGTTCC	21	60.1
MS55	forward	AGTGATGGCTCTATGTTCTGC	21	60
MS55	reverse	TTGCCAGCCAGTCTGCTC	18	57.8
MS56	forward	GTCAGTGTCTGTTATGCCAGC	21	55.7
MS56	reverse	AGAGAAGAGTCTGGAGAACAA	21	59.3
MS57	forward	TTTGCCTACTTCTTCTTCTCC	21	55.6
MS57	reverse	AGAAGAAGTGGAAGAAGGAGA	21	55.2
MS58	forward	GCTGTCCTCTACTTCTTTTCC	21	59.8
MS58	reverse	TCCATGTCCCGACTGCATG	19	56
MS59	forward	AGACAGACAGACAGACAGACA	21	58.7
MS59	reverse	ACGAGATCGGGTGTCACC	18	58.1
MS60	forward	AGCTGACTGACTGACTGCC	19	55.5
MS60	reverse	AGAGAGAGAGAGAGAGAGACT	21	59.3
MS61	forward	AATGGAACGAACAGGAAATGG	21	59.1
MS61	reverse	AAAAATCCCCGTTACGTGTGG	21	56.8
MS62	forward	AAGAAAGACAGACAGGAAGAC	21	57
MS62	reverse	GGCTCATTAGGCTGATTGTTC	21	55.2
MS63	forward	ACTGCTGCTTCTGCTACTACT	21	59.3
MS63	reverse	GAGGGATGTGTGCATTTCCAG	21	58.8
MS64	forward	TCAACCTTTCTTTCTGTTGCC	21	58.1
MS64	reverse	ACATACATACATGCCTGCCTG	21	56.9
MS65	forward	ACAAACAGACATACAGGCAGG	21	55.2
MS65	reverse	TCTATCTGTCTGTCTGTCTGT	21	58.2
MS66	forward	AGATGGAGCAACAACAGTCAC	21	58.6
MS66	reverse	TGTGTCAATGTTGTTGCTCCA	21	58.5
MS67	forward	ACAATTTGAACGGACCTCG	19	59.9
MS67	reverse	TACACCCCAACCTCCATCCT	20	55.9
MS68	forward	AACACCTTCTCAAATTAGGCA	21	58.5

MS20	reverse	CCTCTGCACTGCGTCACT	18	55.3	MS68	reverse	CCTTGTGGCATCTTATAGGGC	21	55.4
MS21	forward	ATGGATGGATGGATGGATGGA	21	56.5	MS69	forward	TGTCGTGTTACTATGTTTGCT	21	55.3
MS21	reverse	TGGATTTGTCTGATTCAGTGC	21	58.3	MS69	reverse	GGTGGTGGTAGTAGTAGTAGT	21	55.6
MS22	forward	ACCTGACCCTGGTATTACCCT	21	55.6	MS70	forward	AGCAGCAGCAGTAGGAGT	18	59.7
MS22	reverse	AAGGACCAAAATCCAGGAAAA	21	59.6	MS70	reverse	TGGTGGTTTAGGTGGTTGGTT	21	57.9
MS23	forward	GCATTGGCTGCCATGAGC	18	60.3	MS71	forward	AAGACAAGCATAAACGCTGA	20	56.6
MS23	reverse	TGGTTGGTTGGTTGGTTGGT	20	59.9	MS71	reverse	ACAGAAAAACATTGCCAGACA	21	55.7
MS24	forward	GCATCACTGCTAGGGACGT	19	59.8	MS72	forward	ACAGACAAACAGACAGACAGA	21	55.2
MS24	reverse	CAGACAGACAGACGGACGG	19	59.5	MS72	reverse	TGTCTGTCTGTCTGTCTATCT	21	56.8
MS25	forward	TGGGTGAGGAGAGAGAGAGA	20	58.1	MS73	forward	TGAGTGAGTGAGTGAGTGAGT	21	56.5
MS25	reverse	TCTGTCTGTCTGTCTGTCTGT	21	58.3	MS73	reverse	ACCCTCTCTTTGGTAGCTTC	20	58.1
MS26	forward	ACAAAAGCGACGCTGGC	17	59.7	MS74	forward	AGATGACAGACAGACAGACA	20	55.7
MS26	reverse	TTGAGCCGCGACACTCC	17	59	MS74	reverse	TCTATCCTCTGTCTGTCTGTC	21	55.5
MS27	forward	TGCATACGAGGAAAATTTCGA	21	56.4	MS75	forward	CGGGCGAAACACACCC	16	55.1
MS27	reverse	ACTTTCGCTGCGAAGGA	17	56	MS75	reverse	AGGTAGGTAGGTAGGTAGGT	20	58.2
MS28	forward	GGCCATCGTGCCGTTT	16	60.2	MS76	forward	TGGTTGGTTGGTTGGTTGGT	20	55.1
MS28	reverse	CCAAGACCTCTCCCCCACT	19	57.8	MS76	reverse	AGACTGAGTTACGTGATCTTG	21	60.3
MS29	forward	ACCGATAACTGGGAGATTACT	21	55.7	MS77	forward	TGGTTGGTTGGTTGGTTGGT	20	55.5
MS29	reverse	GCTTTACATTGGAATCCCCA	20	55.6	MS77	reverse	AAAGGTGTATCTGGCATTGAA	21	60.3
MS30	forward	CAGACAGACAGACAGGCAGA	20	56.2	MS78	forward	GACAGACAGACAGACAGACT	20	55.3
MS30	reverse	TGTCCCTTCTCCTCTCTCA	19	59	MS78	reverse	AAGCGTTTAGTAAGGTTACGT	21	56.2
MS31	forward	GGCCCTATGTTCTGCCAATG	20	59.7	MS79	forward	AGACAGACAGACAGACAGACA	21	57.4
MS31	reverse	GTCTTACCGCACGGGCA	17	59	MS79	reverse	CACCTTTACCATCTTACCGGT	21	58.1
MS32	forward	AGACAAACAGACAGACAGACA	21	56.5	MS80	forward	CTGGGTGGGTGGGTGG	16	60
MS32	reverse	TCTGGTTGTAGGTTTAGCGT	20	56.8	MS80	reverse	AGGCGTGTCTTGTCTCGC	18	58.7
MS33	forward	ATTTGTGTGTGTGTGTGTGT	20	57.2	MS81	forward	GAGCTTACACATGCATCCACA	21	58.1
MS33	reverse	AGACTGAAGGATGGGTGGA	19	56.3	MS81	reverse	ACAGACAGACAGACAGACAGA	21	58.6
MS34	forward	CCATCGGTACAAGAAGTGGCT	21	55.5	MS82	forward	ACAGCTCCTAATCGCGCG	18	59.3
MS34	reverse	AAGTTTGATCTTTCCGACCAT	21	60.1	MS82	reverse	GTGGTCAAGACTAAGCATGGC	21	60.3
MS35	forward	GGGTGGGTGGGTGGGT	16	55.2	MS83	forward	AAGGAGCGAGAAACTTGCAT	20	57.4
MS35	reverse	TCTATCTGTCTGTCTGTCTGT	21	60.5	MS83	reverse	TCAGATCCACAAGCCTTAAGG	21	57.8
MS36	forward	TGGCCTCCACTCACTCACT	19	55.9	MS84	forward	TTTCTTGTTGTTCATCCTCCT	21	60.4
MS36	reverse	CATCTGCGTATGTGTATGTGA	21	60.2	MS84	reverse	CCTCATACAGGGCCTCCAGA	20	55
MS37	forward	CGGGCAGACGGACAGAC	17	55.3	MS85	forward	TTCCGCCAATGGAACGAAC	19	59.6
MS37	reverse	AGTGTCGAGGATCTATGAAGA	21	59.8	MS85	reverse	GCCACCGCCGTACACA	16	58.8
MS38	forward	CGTGCATAGATATGTTTCACG	21	57.7	MS86	forward	GATGGCCCTATGTTCTGCC	19	60.1
MS38	reverse	CAGTCAGTCAGTCAATCAGCT	21	55.3	MS86	reverse	CGTCACAGTCTCCGGCAG	18	58
MS39	forward	CGACAGGCAGGTAGGC	16	58.5	MS87	forward	GGAAGAGTTGTACGCTCCT	19	58.6
MS39	reverse	TCTGTCTGCCTGTATGTCTGT	21	56.4	MS87	reverse	TCTCAACTGAAGAGTGCCCA	20	56.5
MS40	forward	CGGACTGTGTTGTTGTTGTTG	21	59	MS88	forward	ATTAAGGATCGAGAAACCTGC	21	55.3
MS40	reverse	ACAACAACACCAACACCAACA	21	58.5	MS88	reverse	TCTTTTACTGATGTACACGCA	21	55.8
MS41	forward	CCTGTCCATCTACCTACACA	20	55.2	MS89	forward	TCGTACTTTGCTGTTGTTGTT	21	60
MS41	reverse	ACAGACAGACAGACAGATAGA	21	55.7	MS89	reverse	GCTAGTGTCCTTGCAGTGGT	20	56.6
MS42	forward	TCTGTCTGTCTGTCTGTCTGT	21	58.8	MS90	forward	TTTTCAGGGCTGATCCATAAG	21	55.7
MS42	reverse	GCTTGGCCTCCTTGATGTG	19	58.1	MS90	reverse	ACTATTGTCACTACCATTGCC	21	55.6
MS43	forward	TCACACACACCCATCCATCC	20	55.3	MS91	forward	TCGTTAAATCTCACATTCGGT	21	55.7

MS43	reverse	TGAAATCAGCTTACCTTGCA	20	59.7	MS91	reverse	TCAACACCTTGCTGAATGG	19	55.4
MS44	forward	ATCGGGGCATTCATAGGTGT	20	59.6	MS92	forward	CCGCCAATGTAACGAACAGG	20	60.1
MS44	reverse	GGTCAGCTGATGCGAATGC	19	58.9	MS92	reverse	GCTGCTGCTCGGCTGT	16	59.6
MS45	forward	GGACTTGTTCCACATGCA	18	55.5	MS93	forward	CCGCCAATGGATCGAAAGTG	20	58.8
MS45	reverse	TCAATGAGGGGCAGCAT	17	55.1	MS93	reverse	AGCCCTGTGCCAGTCTG	17	59.6
MS46	forward	AGATAGGTAGGTAGGTAGGCA	21	59.9	MS94	forward	TGGGTGCTGAGAATTTCAAAA	21	59.2
MS46	reverse	GTTGTTTCGGCTAGCAGCG	19	56.1	MS94	reverse	CCCCAGGGACAGAGGAGA	18	56.3
MS47	forward	GGTGGGTCAGTCGGTTGG	18	58.3	MS95	forward	TGGTGGTGATGATGGTGGTG	20	59
MS47	reverse	ACATGTGGCCACGAAAAGAA	20	60	MS95	reverse	ACCACCACCACCACCAC	17	60
MS48	forward	ACGGGAACAAAAACACATACA	21	55.4	MS96	forward	AGTGGTGGTGGTGGTGG	17	58.9
MS48	reverse	ACACATTCTTTCAGTCTGTGT	21	56.3	MS96	reverse	CGGGTACTGCGAGGAATCT	19	58.6

Chapter 4: Hybrids or not hybrids: investigating potential introgression between two snail species through geometric morphometrics, mitochondrial and nuclear DNA analysis

Abstract

The giant land snail *Placostylus fibratus* is harvested for food on the Isle of Pines, a small island south of the New Caledonia archipelago. On this island it is sympatric with the species *P. porphyrostomus,* and a third distinct shell morphotype has been reported. Hypotheses about the nature of this third morphotype have included the presence of a hybrid population, a distinct taxon, or a plastic phenotypic response to a distinct environment. Here I sample *Placostylus* across the island and use a combination of geometric morphometrics, nuclear and mitochondrial DNA analyses to test the various hypotheses of the origin of the third morphotype. Using *newHybrids* and *STRUCTURE* analyses on neutral nuclear datasets I found that snails from the third morphotype were not hybrids of first or second generation and had the same nuclear genetic diversity as other *P. fibratus* collected on the island. Phylogenetic analysis on mitochondrial DNA found two main *Placostylus* clades on the island but these were not strictly concordant with species identification. Based on the combination of morphological and genetic results I infer that the presence of a third shell morphotype on the island is probably linked to a phenotypic response to a distinct environment, and that only two taxa of *Placostylus* are present on the Isle of Pines.

Introduction

Secondary contact between individuals of previously reproductively isolated lineages can lead to a variety of different evolutionary outcomes, with consequences on regulation and species conservation (Arnold, 1992). When reproduction occurs between individuals that are estimated to be distant enough on an evolutionary timescale to be classified into distinct 'groups' (species, populations, lineages, management units), the resulting hybridization can either be positive or negative for long term survival of both parental taxa and hybrids, and taxonomic and phylogenetic diversity. On the far end of the spectrum hybridization can lead to extinction of all populations involved (Rhymer & Simberloff, 1996). This happens when hybrids are maladapted but keep being produced because of a high hybridization rate between parental lineages, a process called demographic swamping (Prentis et al., 2007; Todesco et al., 2016). More frequently, parental taxa are entirely replaced by viable hybrid populations that can have a higher fitness (heterosis), but not necessarily (hybrid swarm). Many observed cases of hybridization following secondary contact are associated with human activities that have changed habitat availability and species ranges, and this globally leads to a net decline in biodiversity (Hasselman et al., 2014; Lowe & Abbott, 2015). It is also possible for all groups involved to establish or remain as distinct viable populations, either in a sympatric or allopatric setting (Costa et al., 2020; Harrison & Larson, 2016). In contrast to these scenarios of hybridization is the possibility of reinforcement, where a contact zone between two closely related species results in further separation of lineages because of a range of mechanisms that arise from maladapted hybrid individuals (Hollander et al., 2018; Noor, 1999). Understanding local patterns of hybridization/reinforcement and their outcome can therefore have important implications for the management practises of endangered species. In situations where the conservation status of parental groups differs, decisions need to be carefully considered and aim towards the protection of most threatened taxa (Dierking et al., 2014). Genetic analysis can be used to diagnose hybrid status of a population and be complemented by analysis of phenotypic variation. Here I investigate the specific hybrid status and introgression patterns of two sympatric giant snail species (*Placostylus fibratus* and *P. porphyrostomus*), which are thought to have a long-term history of introgression and have potentially produced viable hybrid populations (Dowle et al., 2015; Quenu et al., 2020).

Placostylus is a genus of large terrestrial snails in the family Bothriembryontidae living on islands of the Western Pacific ocean (Breure et al., 2010). In New Caledonia, six species are recognised by the current taxonomy (Neubert et al., 2009), but recent investigations indicate

this separation is too simplistic and does not reflect either mitochondrial lineages or nuclear data separation (Dowle et al., 2015). The two most-abundant species *Placostylus fibratus* (Martyn 1789) and *Placostylus porphyrostomus* (Pfeiffer, 1851) are locally common on the Isle of Pines, a small island (152 km^2) south of Grande Terre, where *P. fibratus* is still harvested for food (Brescia, 2011). Despite being sympatric on this island, initial analyses of nuclear genetic data and shell shape separated the two species, but non-monophylly and haplotype sharing of mitochondrial lineages suggested a history of hybridization between the two taxa resulting in mtDNA introgression (Dowle et al., 2015). More recent morphological analysis has challenged this separation of *Placostylus* into just two species, with the presence of a third morphotype that appears intermediate in shell size and shape between the two species (Quenu et al., 2020; chapter 3 of this thesis). One population sample, Comwagna, was found to consist almost entirely of individuals from this third morphotype. Genetic data has yet to be generated for snails of this morphology. Determining their status is important because at the moment only *P. fibratus* can be legally harvested for food. Three main hypotheses have been advanced to explain this morphological disparity: the presence of hybrid snails (*P. porphyrostomus* x *P. fibratus*), the presence of snails from a third lineage, or a plastic phenotypic response to a distinct and localised environment. Our objective is to test these hypotheses, using datasets combining morphological data, nuclear genetic data and mitochondrial genetic data.

Hybrid individuals tend to present intermediate values for quantitative traits. This has been demonstrated empirically for several species (Thompson et al., 2019) and can be explained by the cumulative effect of numerous co-dominant genes coding for the same quantitative trait. To quantify and analyse snail shell morphology Geometric Morphometric (GM) methods have proved to be un-biased and reliable (Bocxlaer & Schultheiß, 2010; Daly et al., 2020; Vaux et al., 2017, 2020). Landmark-based GM methods allow analysis of shape by analysing variations in position of a set of homologous landmarks after these have been submitted to a process called procrustean analysis, which removes effects of size, rotation and position (Mitteroecker & Gunz, 2009). GM methods can be complemented by clustering analysis to determine an optimal separation of morphological datasets. Through clustering methods morphological groups can be linked to genetic groups and evolutionary hypotheses tested (Haase et al., 2014; Vaux et al., 2018). In *Placostylus* a robust methodology has been established to produce two-dimensional landmark datasets, and was used in combination with various clustering algorithms to determine the presence of a third morphotype on the Isle of Pines (Quenu et al., 2020). By linking

individuals of this morphology to genetic data test the various hypotheses formulated about their origin can be tested.

Hybrid populations can exist on a continuum from reproductively unviable F1 hybrids to fully time-stable hybrid populations, that may be independent or able to reproduce with parental lineages (Nieto Feliner et al., 2017). Being able to detect hybrid individuals and assign them to a specific hybrid class is therefore an important area of the study of hybridization and must rely on a solid methodology. Using independent codominant allelic nuclear markers it is possible to calculate assignment probabilities to hybrid classes for specific individuals, assuming parental populations are well defined and their allelic frequencies known (Anderson & Thompson, 2002). To achieve this the software *newHybrids* uses bayesian inference and a genetic model which will assign individual posterior assignment probabilities to different hybrid classes based on the expected admixture of parental genes in hybrid categories. This approach is very efficient when testing for a hybrid population of first or second generation, where expectation of gene admixture in hybrid classes are straightforward. But in long-term stable hybrid populations the stochastic loss of parental alleles due to the combination of genetic drift and continuous interbreeding of hybrids across many generations will tend to obscure admixture signals. Hybrid individuals can also backcross with parental lineages, which will add extra noise to admixture signals. For these situations more general clustering analysis such as the modelling approach included in *STRUCTURE* can be applied to neutral nuclear marker datasets (Jombart, 2008). Expectations are that hybrids will be either characterised as a distinct cluster or as an admixture of parental groups. If hybridization is the main influence on genetic structure, hybrid groups should be of intermediate genetic distance to both parental groups. These approaches offer more flexibility in interpretation and analysis of complex genetic datasets, but direct assignment probabilities to hybrid groups are never explicitly computed. In the case of *Placostylus* if intermediate morphological snails are first generation sterile hybrids, strong assignment probabilities to F1 hybrid class can be expected under a *newHybrids* model and signs of admixture between two hypothetical parental populations in *STRUCTURE* analysis. Snails from a stable hybrid population might not necessarily demonstrate strong assignment probabilities to any pre-determined hybrid class in *newHybrids* but would show signs of admixture at K=2 in *STRUCTURE* admixture analysis.

In most metazoan species the mitochondrial genetic material is transmitted unilaterally through the maternal lineage and does not recombine. The snails studied here are hermaphrodites so every individual can potentially produce eggs and transmit their own mitochondrial genomes

to the next generation. By sampling a fragment of the mitochondrial genome across snails of different morphotypes it is possible to test the hypothesis of there being an independent snail lineage. At the same time if morphologically intermediate snails are proved to be hybrids, the repartition of mitochondrial haplotypes in populations will reflect symmetry of founder hybridization events -hybrid individuals will carry haplotype of the maternal parental lineage (Vallejo-Marín et al., 2016). Previous sampling of mitochondrial gene ND2 in New Caledonia *Placostylus* has revealed a complex phylogenetic pattern for the two species (Dowle et al., 2015). Four main lineages were identified within the whole species complex, and neither *P. prophyrosotmus* nor *P. fibratus* was monophyletic. On the Isle of Pines two mtDNA lineages are present which largely segregate by morphological species, however a few individuals of both *P. prophyrosotmus* (1/17) and *P. fibratus* (9/27) have been recorded with the alternative mtDNA lineage (Dowle et al., 2015). This pattern was interpreted as the consequence of past hybridisation events which allowed mitochondrial gene flow and introgression between the two taxa. By sequencing the ND2 mitochondrial region in snails of intermediate morphotype from the Isle of Pines the hypothesis of a third taxon present on the island can be tested. If snails are hybrids, they will carry the mtDNA lineage of the snail that provided the eggs (maternal line) of successful hybridization events. If snails are from another taxa and/or come from a translocation they will either form a new separate lineage or cluster with lineages from other places in the archipelago.

Methods

Sampling strategy

Placostylus snails were sampled at eight sites on the Isle of Pines in April 2015 (Vao, Gadji, Kere, Youaty, Touete, Comagna, Wapan, Waatchia) (Fig. 1). Each sample site consisted of a transect orientated along an elevation gradient from the coast to the centre of the island. All live snails within a 20m radius of a fixed point were collected by hand. Shells of 336 adult snails (identified by the thickened lip) were photographed for shape analysis. Foot tissue samples were stored in 95% ethanol for DNA analysis. For mitochondrial DNA analysis 88 new mitochondrial sequences were generated, and for neutral nuclear marker analysis 124 snails were genotyped with six neutral nuclear loci.

Geometric morphometrics (GM)

The set of two-dimensional landmarks used for this study was derived from a set of optimal landmarks and semi-landmarks established in an analysis of New Zealand *Placostylus* (Daly, 2017). A total of 40 landmarks and semi-landmarks were used, consisting of nine fixed landmarks and 31 sliding/semi-landmarks. Digital images of the ventral surface of each shell were obtained using a Canon EOS 600d with EF100 mm f2.8 USM macro lens after careful positioning of the shell in a bed of sand of contrasting colour. The camera equipment was mounted on a high-precision Kaiser stand to allow reproducible positioning and orientation (Dowle *et al.*, 2015). All images were captured with the lens fixed at the same distance from shell, and shells were positioned by the same person to minimize operator error (Schilthuizen & Haase, 2010). In order to place the semi-landmarks along the curves of the shells, two 'combs' were placed manually using *Adobe Photoshop CS6*, with their teeth perpendicular to a line from the shell apex to the intersection of the lips and outer shell. Digitization of landmarks and semi-landmarks was conducted using the programme *tpsDig2* v.1.1 (Rohlf, 2015). Landmark type assignment and Procrustes analysis were then performed using *Coordgen* v.8.0 (Zelditch et al., 2004). Principal component analysis (PCA) was executed on the covariance matrices of the aligned landmark coordinates using *MorphoJ* (Klingenberg, 2011). The size of the shells was incorporated into the analysis using the 'centroid size' tool included in the Procrustes analysis in *Coordgen* v.8.0. This size estimation is calculated as the square root of the sum of the arrays coming from the centroid position of a shape to each of the landmarks (Klingenberg, 2016).

Errors linked to shell manipulation and digitization were assessed by experimental replication; this involved taking five photographs of the same shell and comparing the variance attributed to the repeat process with the variance of the whole dataset using the *morphol.disparity*() function in the package *geomorph* v.3.1.2 (Adams, Collyer & Kaliontzopoulou, 2018) in R v.3.6.1 (R Development Core Team & R Core Team, 2017). Overall, the variance linked to manipulation and digitization error corresponded to <0.3% of the variance of the rest of the dataset, which was considered negligible for subsequent analyses.

Gaussian mixture models were used for morphometric clustering analysis, using the R package *mclust* v. 5.4 (Fraley & Raftery, 2006). Gaussian mixture models were built using shell size and two Principal Component variables as inputs (PC1 and PC2). Different models were built for a range of *a priori* clusters in the dataset, with the optimal model being selected on the basis of highest BIC score. Results of the modelling process with additional input variables (PC1–5 and shell size) were also computed for comparison and yielded similar results.

DNA extraction

DNA was isolated from foot tissue samples using the Qiagen blood and tissue sample kit, following tissue extraction protocol. Foot tissue samples were cut down to small pieces using a sterile scalpel blade, before being incubated in a chaotropic lysis buffer with proteinase K overnight to fully digest cell membranes. DNA was isolated by binding to silica and after elution with purified water was stored in -20 °C freezer.

Neutral nuclear marker analysis

For nuclear marker analysis I used a set of six neutral nuclear genetic markers comprising four microsatellites and two nuclear SNPs markers. All these markers were generated specifically for this study and details about their development can be found in chapter 4 (Table 1). The four microsatellite markers (MS47, MS49, MS64, MS94) were synthetised with three different fluorescent tag attached to them (6-FAM, VIC and TAMRA). PCRs used the protocol: 94°C for 3mins, then 94°C for 45s, 50°C for 30s, 72°C for 75s repeated 36 times, followed by a 2 mins annealing step. Genotyping was done by capillary separation of fluorescent labelled fragments on an ABI3730 Genetic Analyzer (Applied Biosystems, Inc., Foster City, CA, USA). I used the *Geneious* microsatellite plugin tool to score and genotype snail individuals, based on amplified fragment size. The two neutral nuclear SNPs Markers (ITSv1 and ANL1) were amplified using an amplicon sequencing approach. ITSv1 is a SNP marker isolated from one variant of the Internal Transcribed Spacer 1 region of the ribosomal cassette. ANL1 is a SNP

marker amplified from an Anonymous Nuclear Loci. Both SNP markers were amplified through PCR with the protocol 94°C for 3mins, then 94°C for 90s, 61°C for 60s, 72°C for 50s repeated 21 times. Resulting PCR products were loaded into a paired-end 250bp Illumina Miseq for sequencing.

Table 1: Primer pair sequences and characteristics of the six neutral nuclear markers used in neutral nuclear marker analyses.

Marker name	Primer direction	Marker type	Motif	Nucleotide substitution	Primer Sequence
ITSv1	forward	SNP	-	T / C	GGAAGTAAAAGTCGTAACAAGG
ITSv1	reverse	SNP	-	T / C	GCTGCGTTCTTCATCGATGC
ANL1	forward	SNP	-	G / C	TGGAATCCATAGGGGTCAAAATTCC
ANL1	reverse	SNP	-	G /C	ACCAGAAAGGACCAGGGAGC
MS47	forward	Microsatellite	GGTT	-	ACA TGT GGC CAC GAA AAG AA
MS47	reverse	Microsatellite	GGTT	-	GGT GGG TCA GTC GGT TGG
MS49	forward	Microsatellite	GACA	-	AGA TAG ACG GAC GGA CGG A
MS49	reverse	Microsatellite	GACA	-	TGA CTG TTG GAG ATT GCA GCT
MS64	forward	Microsatellite	ATGT	-	TCA ACC TTT CTT TCT GTT GCC
MS64	reverse	Microsatellite	ATGT	-	ACA TAC ATA CAT GCC TGC CTG
MS94	forward	Microsatellite	ACTA	-	TGG GTG CTG AGA ATT TCA AAA
MS94	reverse	Microsatellite	ACTA	-	CCC CAG GGA CAG AGG AGA

Neutral nuclear markers were analysed using *newHybrids* (Anderson & Thompson, 2002) to generate a model that would compute posterior probabilities for individuals to fall into different hybrid classes. Six classes were specified in the model: the two parental classes *fibratus* and *porphyrostomus*, hybrids of first and second generation F_1 and F_2 and two back-crosses classes BC1 and BC2 which correspond to individuals resulting of back-crossing of F_1 individuals with either parental lineage. Snails that were of either *fibratus* or *porphyrostomus* morphologies (based on GM analysis) were specified as of pure parental lineage in the model. Jeffreys-like prior distributions were used for mixing parameters $\boldsymbol{\pi}$ and θ, and the MCMC was run for 100,000 steps for burn-in and 1,000,000 steps after burn-in period.

Clustering analysis of neutral nuclear data used *STRUCTURE* (Pritchard et al., 2000). Models were built for a number of a priori clusters ranging from K=1 to K=10, with 10 iterations for each value of K. Models with admixture and correlated allele frequencies were used. For each model the MCMC chain was run for 100,000 steps for burn-in period and 1,000,000 steps after burn-in period. Optimal number of cluster K was investigated by calculating the average mean likelihood of models for a specific value of K over the ten iterations (Pritchard et al., 2000), and

through the Evanno Delta K method (Evanno et al., 2005). Pairwise F_{ST} distances between morphological groups were calculated using the pairwise.fst() function in the *adegenet* package (Jombart, 2008). To estimate the significance of F_{ST} values between morphological groups p-values were computed by comparing the calculated F_{ST} values with a reference distribution of F_{ST} values obtained where individuals were randomly assigned to one of three groups (Null hypothesis = panmixia). A principal component analysis of the six nuclear markers was performed using the dudi.pca() function of the *adegenet* package (figure 3).

Mitochondrial DNA analysis

A 754 bp mitochondrial DNA sequence was amplified in 88 snails representing three morphotypes. The ND2 region of the mitochondrial region was targeted, using polymerase chain reaction primers previously developed for New Caledonia *Placostylus* phylogenetic analysis (primers: Placo_ND2F: AAC GCA AAG GGT ATG AAC CCG TAA ATA G and Placo_ND2R: GAG CAA TCG CCG GAG GAA CGG AAA T) (Dowle et al., 2015). Amplification used Polymerase Chain Reactions (PCRs), using dilutions (1:50) of initial DNA extractions. PCR used the protocol: 94 °C for 3 min; 94 °C for 45 s, 56 °C for 45 s and 72 °C for 75 s repeated 36 times; followed by a 2-min annealing step. Sanger sequencing was realized on the ABI Prism 377 DNA sequencer (Applied Biosystems, Inc., Foster City, CA, USA). The quality of sequences was checked on *Geneious* 2020.0.5 (Kearse et al., 2012). The translation tool with mitochondrial invertebrate genetic code was used to verify there was at least one uninterrupted amino acid frame in all sequences. New ND2 sequences were aligned to New Caledonia *Placostylus* ND2 sequences available on GenBank (total number of sequences used in the alignment was 188 sequences) using *MUSCLE* (Edgar, 2004). I used MrBayes to produce a Bayesian phylogeny, using one model of nucleotide substitution (GTR+ I + Gamma). Separate haplotype networks were inferred for each lineage from the Isle of Pines, using the TCS algorithm implemented in *popART* (Leigh & Bryant, 2015).

Results

Geometric morphometrics

Based on BIC scores the shell size and shape variation in our dataset had an optimal fit to a GMM with three clusters (elliposidal, equal volume, shape and orientation (EEE) model). Two of these clusters correspond to morphotypes of the recognised species *P. fibratus* and *P. porphyrostomus* and the third cluster was intermediate in terms of size and shape. Out of 336 shells, 226 (67%) individuals were assigned to the *P. fibratus* cluster, 56 (17%) to the *P. porphyrostomus* cluster and 54 (16%) to the third cluster. Snails from the third cluster mostly originated from the Comwagna population, and the Comwagna population sample was almost entirely constituted of shells from this third cluster (figure 1). The Youaty population sample consisted only of *P. fibrates* specimens. The Touete, Waatchia, Gadji, Kere and Vao population samples contained both *P. fibratus* and *P. porphyrostomus*. The Wapan population sample contained examples of all three morphological clusters.

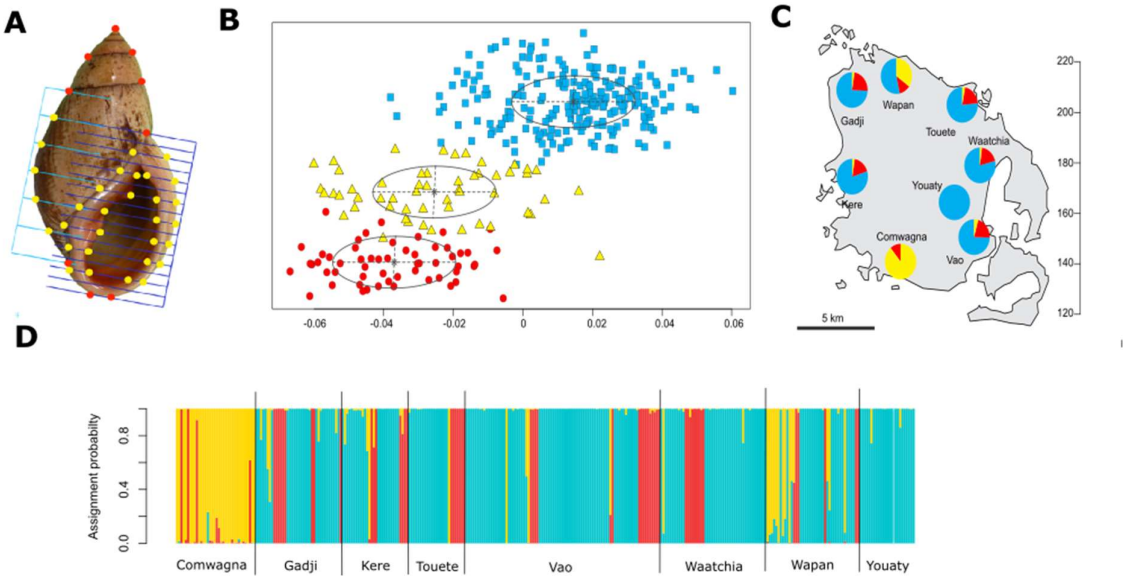


Figure 1: Morphological analysis of 336 shell sampled across eight locations on the Isle of Pines. A) Shell of New Caledonian *Placostylus fibratus* with the 40 landmarks used in the geometric morphometry analysis. Digitizing combs are in blue and the orientation is indicated by the white dashed line. Red and yellow dots indicate permanent and semi-landmarks respectively. B) Gaussian Mixture Model classification of *Placostylus* in relation to variation of shell size and shape along PC 1 (35.3%). Red = *porphyrostomus* cluster, blue = *fibratus* cluster. C) Approximate position on the Isle of Pines of the eight sample populations corresponding to local tribal areas, with the proportion of each population sample assigned to one of three morphotypes with Gaussian Mixture Modelling shown. D) Assignment probabilities calculated for shell shape and size variation in eight population samples of *Placostylus* from the Isle of Pines

Neutral nuclear marker analysis

Allelic diversity within the six nuclear markers ranged from 2 (ANL1) to 19 (MS94). The three morphological groups contained respectively 72 (*fibratus* morphology), 28 (*porphyrostomus* morphology) and 24 (third morphotype) individuals, and each group contained a total of 38, 33 and 31 alleles over the six nuclear markers. The genetic model generated using *newHybrids* assigned the majority of individuals with *fibratus* and *porphyrostomus* morphologies to their corresponding genetic groups (Figure 2). Of the snail specimens with the third morphotype 21 had strong assignment probability (> 0.8) to the *fibratus* genetic group, one to the *porphyrostomus* genetic group and five did not have strong assignment probabilities to any genetic group. The model did not find strong support for any of the snails to be assigned to any of the four hybrid classes (F_1, F_2, BC_1 or BC_2). The maximum assignment probability of a snail being an F_2 hybrid was 0.38 for an individual from Comwagna with the third (intermediate) morphotype.

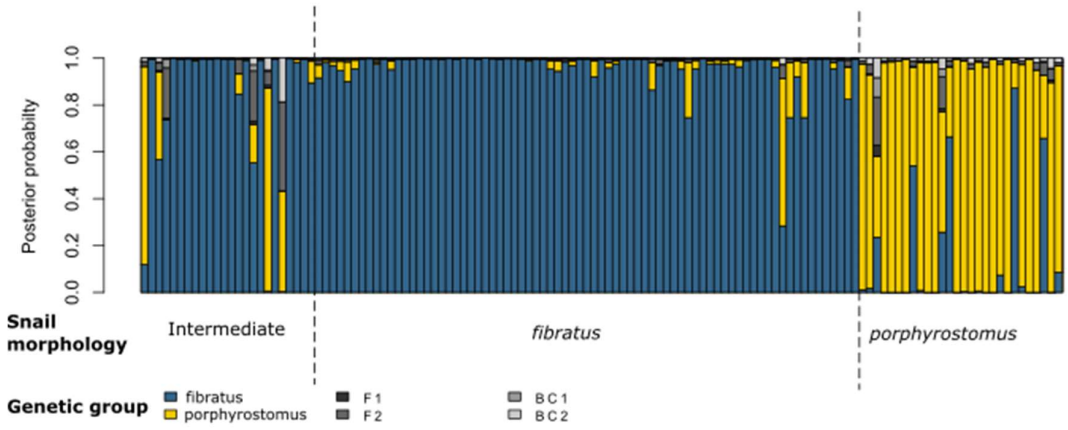


Figure 2: Posterior assignment probabilities for each of 124 Isle of Pines *Placostylus* snails to parental and hybrid classes, based on genotype data from six nuclear markers. The *newHybrids* genetic model comprised two parental populations (*P. fibratus* and *P. porphyrostomus*) and four possible hybrid classes (hybrid of first and second generation F_1 and F_2 and two backcrosses class BC_1 and BC_2).

The optimal division of the dataset was into three cluster using a *STRUCTURE* model analysis, with agreement between the mean likelihood and the Evanno method (figure 3B). Examination of the two cluster *STRUCTURE* model (K=2) revealed the assignment of specimens did not perfectly separate *fibratus* and *porphyrostomus* shell morphologies. Most snails of *porphyrostomus* shell morphology (21/28, > 0.8) were assigned to the same genetic group, but for snails of *fibratus* and intermediate shell morphologies mixed assignment probabilities between the two theoretical clusters were reported. At K=3 the separation between snails of *P. porphyrostomus* shell morphology and the other two (intermediate and *P. fibratus*) shell morphologies was more evident. Most of the snails of *P. porphyrostomus* morphology were strongly assigned to the same genetic group (21/28, > 0.8). Most of the snails from the *fibratus* (69/72) and intermediate (21/24) morphologies had mixed assignment probabilities to the two other genetic groups (Figure 3C).

Differentiation between snails of *P. porphyrostomus* shell morphology and snails of *P. fibratus*/intermediate shell morphologies was observed in the principal component analysis of nuclear markers on the first axis of variation (PC1 (24.4%); figure 3A). For the sample of six nuclear markers the genetic variability of snails of the third morphotype appeared to be imbedded within the variability of snails of *P. fibratus* morphology. Pairwise F_{ST} distances were small and differed slightly from panmixia expectations between the intermediate / *fibratus* morphological groups and were relatively high and significantly differed from panmixia expectations between the *fibratus* / *porphyrostomus* and intermediate / *porphyrostomus* shell morphologies (Table 2).

Table 2: Pairwise F_{ST} distances and associated p-value between morphological groups of *Placostylus* snails on the Isle of Pines, inferred from six neutral nuclear markers.

	Intermediate	***fibratus***	***porphyrostomus***
Intermediate	0	0.035417 / 0.012*	0.1260479 / 0.001***
Fibratus		0	0.1204032 / 0.001***
Porphyrostomus			0

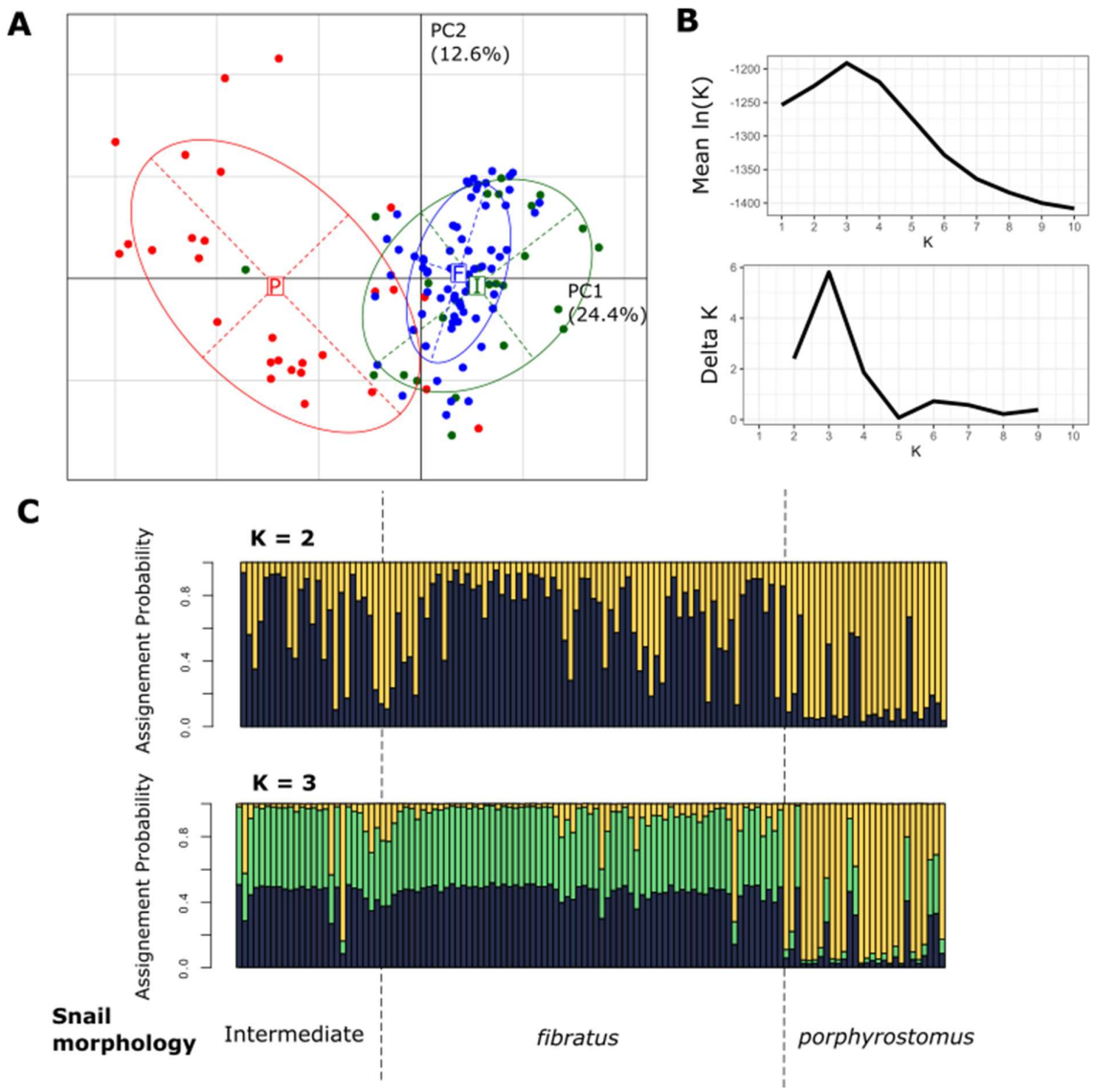


Figure 3: Clustering analyses based on snail genotypes for six neutral nuclear markers. A) The first two axis of a principal components of variation analysis. Individuals are color-coded according to their shell morphology: red = *P. porphyrostomus*; blue = *P. fibratus* ; green = third (intermediate) morph. B) Change in the mean likelihood over ten iterations of STRUCTURE clustering models for an increasing number of a priori cluster K, and change of the Delta K value for the same models as calculated through the Evanno method. C) Posterior assignment probabilities to genetic clusters based on STRUCTURE models for K=2 and K=3.

Mitochondrial DNA Analysis

The collection of 188 mitochondrial ND2 sequences separated into five distinct lineages, of which four corresponded to the *P. fibratus* and *P. porphyrostomus* species complex (figure 4). Of these four lineages, two only contained snails from Grande Terre previously identified as either *P. porphyrostomus* or *P. fibratus*. Snails from the Isle of Pines were split into the other two lineages. One of these lineages contained almost all snails of *P. fibratus* morphology, along with snails of intermediate morphology and three snails of *P. porphyrostomus* morphology. The other lineage contained almost all snails of *P. porphyrostomus* morphology, along with snails of intermediate morphology, one snail of *P. fibratus* morphology and snails collected on Grande Terre. From the Isle of Pines 77 different haplotypes were observed, with most haplotypes recorded from single specimens.

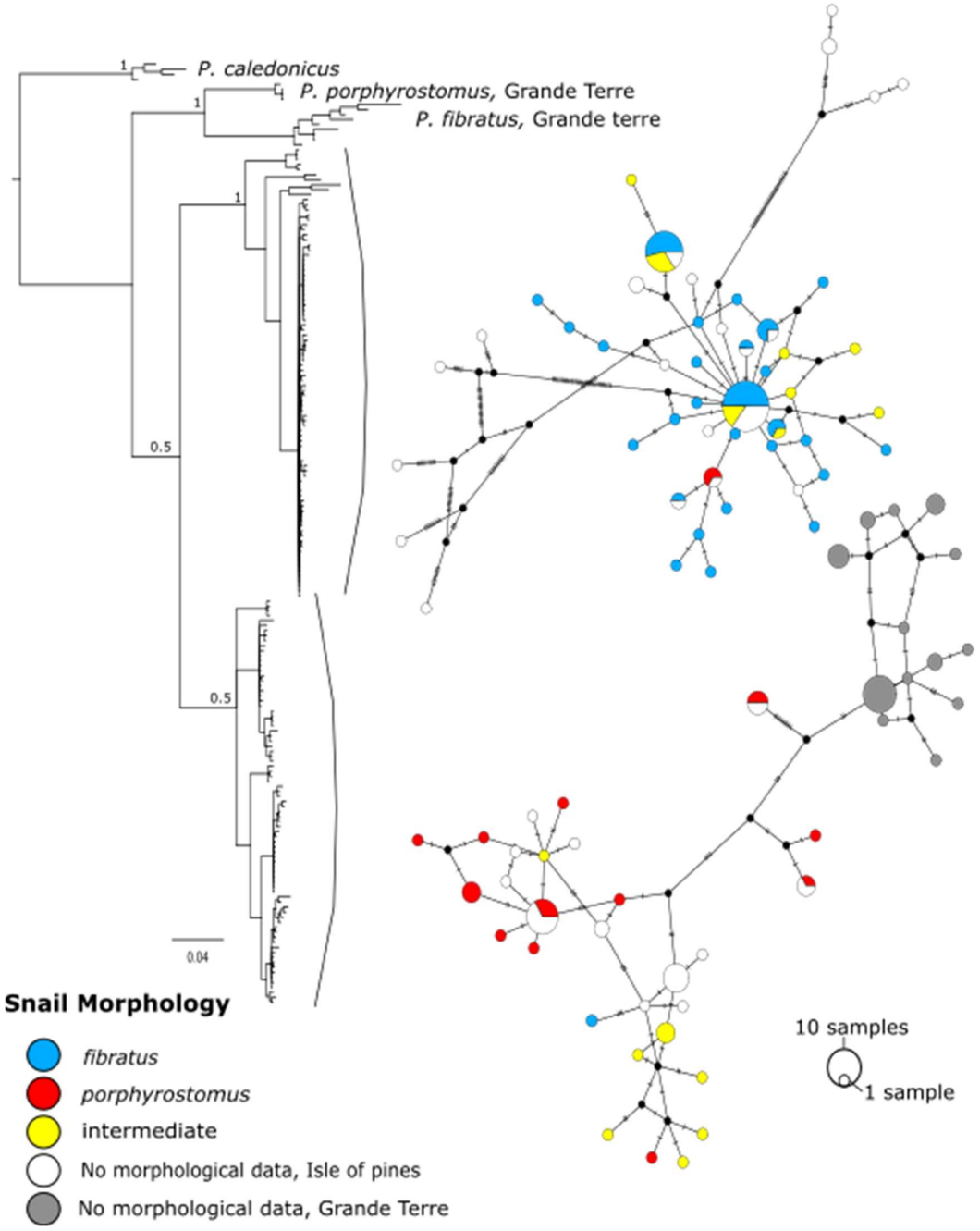


Figure 4: Bayesian phylogeny of 188 New Caledonia ND2 *Placostylus* sequences (754bp), and TCS haplotype networks for the two lineages comprising snails from the Isle of Pines. Node labels on phylogeny indicate posterior probability support values.

Discussion

Three hypotheses were considered to explain the presence of an intermediate morphotype in addition to the two morphospecies *P. fibratus* and *P. porphyrostomus* on the Isle of Pines (Quenu et al., 2020). The first hypothesis was that the intermediate third shell morphotype was the result of a hybrid population primarily established at the Comwagna sampling site. Under this scenario, hybrid snails could be either the result of a recent (or ongoing) hybridization (hybrids of first or second generation), or they could originate from a more ancient hybridization event. The nuclear genetic data did not support either of these hybrid hypotheses. Using nuclear genotypes from 124 individuals I found that most snails of intermediate morphology fall into the *fibratus* genetic group using a *newHybrids* model with parental genotypes identified by morphology (*P. fibratus* and *P. porphyrostomus*). Similarly, in the *STRUCTURE* analysis snails of intermediate and *fibratus* morphology had similar mixed assignment probabilities to two hypothetical clusters at K=3. The principal component analysis did not differentiate the genotypes of individuals of intermediate and *fibratus* shell morphology and estimates of pairwise F_{ST} suggested the genetic distance between the morphological only slightly differed from what would be observed in one panmictic population. All the different approaches to analyse the nuclear genetic data resulted in the same inference - that most of the snails of intermediate morphology are genetically the same as *P. fibratus*.

The second hypothesis to explain the presence of *Placostylus* third shell morphotype was the presence of a third taxon unique to the isle of Pines. Both nuclear and mitochondrial genetic data refute that hypothesis. Snails of intermediate shell morphology all carried the ND2 mtDNA of either one of the two Isle of Pines *Placostylus* lineages. Snails of intermediate and *fibratus* morphology did not form a monophyletic clade but were split in two lineages. This pattern is consistent with previous observations (Dowle et al., 2015), which explained non-monophylly of both *P. fibratus* and *P. porphyrostomus* Isle of Pines mtDNA lineages by historic hybridization resulting in mitochondrial introgression. In our sampling of 24 individuals with intermediate shell morphology, ten had haplotypes that clustered with the '*porphyrostomus'* lineage (42%). This is a much higher proportion than the number of snails identified as not *P. fibratus* using nuclear genetic markers (1 out of 46; 2.2%). Lack of concordance between nuclear and mitochondrial markers is not uncommon (Sloan et al., 2017) and explanations for these observations include sampling artefact, higher introgression rates with *P. porphyrostomus* for that population or other historical processes (population bottlenecks, genetic drift).

The third hypothesis to explain the presence of a third *Placostylus* snails shell morphotype on the Isle of Pines is that shell shape responds to environmental variable. The intermediate shell shape might be a plastic phenotypic response to a distinct environment predominantly found at Comwagna. Given the evidence from the nuclear markers that most specimens with intermediate morphological are genetically indistinguishable from *P. fibratus,* this is the most likely explanation for the third shell morphology. Of the eight locations where snails were sampled Comwagna is the only one situated on red ferralitic (=laterite) soils, whereas all the other collecting locations are on uplifted coral reef terrain (Lagarde & Ouetcho, 2017). Plastic responses in shell morphology and shell growth have been widely studied in gastropods and are often related to environmental calcium availability, which influences shell biomineralisation (Charrier et al., 2013; Madec & Bellido, 2007). Considering this, it is possible *P. fibratus* shells in the Comwagna population are smaller because of a limited amount of calcium, and changes in shape can be explained by allometric effects. When species identification was based solely on shell morphology the taxonomy of New Caledonia *Placostylus* species recognised as many as 141 species (Neubert et al., 2009). Later analyses using snail genitalia and molecular evidence has reduced the number of species to only six. It was understood that much of the variability in shell size and shape across the archipelago was due to differences in shell growth in response to local environmental conditions (Brescia et al., 2008). Although shell size of adults of related *Placostylus* species are known to be linked to local moisture and density levels (Parrish et al., 2014), shell shape is thought to be less plastic (Daly et al. 2020). The snail shells identified as being a third morphotype were intermediate in both shape and size and they provide another example of a pattern of phenotypic plasticity, which seem to be widespread in terrestrial gastropods.

Two methods were used here to test the hypothesis of a hybrid population. Because hybridization is a process that involves the random association of gametes from two parental populations the identification of hybrid status requires a probabilistic approach. This can be done by defining hybrid classes with genotype frequencies calculated from the genotype frequencies of parental populations and the expected proportions of parental alleles expected in different hybrid classes (Anderson & Thompson, 2002; Anderson, 2008). Another approach is to use clustering and admixture analyses, where parental groups can be find de novo and hybrids are expected to show signs of admixture between parental groups (Pritchard et al., 2000). Empirically many studies have found both techniques yield similar results when applied to actual hybrid populations or individuals (Baena-Díaz et al., 2018; Gramlich et al., 2018; Marie

et al., 2011; Moura et al., 2014). Here results of both methods were compatible and both identified the intermediate morphological cluster as having the same genetic diversity as the individuals identified as *P. fibratus* based on shell morphology. This pattern was confirmed by the principal component analysis performed on the six neutral nuclear markers, and by estimates of pairwise F_{ST} between morphological groups. An interpretation of the mixed *STRUCTURE* assignment probabilities could be that individuals of *fibratus* and intermediate shell morphology all descend from an admixture of two ancient populations/lineages, but such admixed assignment probabilities can also reflect a range of other scenarios (e. g. recent bottleneck, ghost admixture; Lawson et al., 2018). Given the low number of loci used in the analysis it seems wiser not to interpret those results any further than a separation in two nuclear genetic groups for the whole *Placostylus* dataset.

The sample of 336 shells constituted the most exhaustive collection of *Placostylus* ever collected on the Isle of Pines. It sampled locations in the eight Kanak districts of the Isle of Pines, where *P. fibratus* is harvested and commercialised for food. Considering the genetic diversity and shell morphological variation of the dataset it can confidently be assessed there are only two *Placostylus* taxa currently present on the Isle of Pines. These taxa correspond to nuclear genetic groups of the recognised species *P. fibratus* and *P. porphyrostomus*. They comprise two mitochondrial lineages (figure 4), but these are not reciprocally monophyletic. Human consumption of *Placostylus* snails on the Isle of Pines could potentially include the intermediate shell-shaped *P. fibratus* individuals but smaller specimens are not favoured for commercial harvest (Brescia et al. 2008). There are two shell morphological groups of *P. fibratus* and one shell morphological group of *P. porphyrostomus*. Distinct mitochondrial lineages of *P. fibratus* and *P. porphyrostomus* exist on Grande Terre (figure 4) and taxonomic revision should be considered to separate these from the *Placostylus* snails found on the Isle of Pines.

References

Adams, D., Collyer, M., & Kaliontzopoulou, A. (2018). *Package ' geomorph ' documentation*.

Anderson, E. C., & Thompson, E. A. (2002). A model-based method for identifying species hybrids using multilocus genetic data. *Genetics*, *160*(3), 1217–1229. https://doi.org/10.1175/1520-0442(1992)0052.0.CO;2

Anderson, Eric C. (2008). Bayesian inference of species hybrids using multilocus dominant genetic markers. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1505), 2841–2850. https://doi.org/10.1098/rstb.2008.0043

Arnold, M. L. (1992). Natural hybridization as an evolutionary process. *Annual Review of Ecology and Systematics*, *23*(1), 237–261. https://doi.org/10.1146/annurev.es.23.110192.001321

Baena-Díaz, F., Ramírez-Barahona, S., & Ornelas, J. F. (2018). Hybridization and differential introgression associated with environmental shifts in a mistletoe species complex. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-23707-6

Bocxlaer, B. Van, & Schultheiß, R. (2010). Comparison of morphometric techniques for shapes with few homologous landmarks based on machine-learning approaches to biological discrimination. *Paleobiology*, *36*(03), 497–515. https://doi.org/10.1666/08068.1

Brescia, F. (Massey U. Thesis (2011)). *Ecology and population trends in New Caledonian Placostylus snails (Mollusca: Gastropoda: Bulimulidae)*.

Brescia, F. M., Pöllabauer, C. M., Potter, M. A., & Robertson, A. W. (2008). A review of the ecology and conservation of Placostylus (Mollusca: Gastropoda: Bulimulidae) in New Caledonia. *Molluscan Research*, *28*(2), 111–122.

Breure, A. S. H., Groenenberg, D. S. J., & Schilthuizen, M. (2010). New insights in the phylogenetic relationswithin the Orthalicoidea (Gastropoda, Stylommatophora) based on 28S sequence data. *Basteria*, *74*((1-3)), 25–32.

Charrier, M., Marie, A., Guillaume, D., Bédouet, L., Le Lannic, J., Roiland, C., Berland, S., Pierre, J. S., Le Floch, M., Frenot, Y., & Lebouvier, M. (2013). Soil calcium availability influences shell ecophenotype formation in the sub-antarctic land snail, Notodiscus hookeri. *PLoS ONE*, *8*(12). https://doi.org/10.1371/journal.pone.0084527

Costa, D., Sotelo, G., Kaliontzopoulou, A., Carvalho, J., Butlin, R., Hollander, J., & Faria, R. (2020). Hybridization patterns between two marine snails, Littorina fabalis and L. obtusata. *Ecology and Evolution*. https://doi.org/10.1002/ece3.5943

Daly, E. (Massey U. thesis (2017)). *Fine scale population structure through space and time*.

Daly, E. E., Trewick, S. A., Dowle, E. J., Crampton, J. S., & Richards, M. M. (2020). Conservation of pupu whakarongotaua the snail that listens for the war party. *Ethnobiology and Conservation*, *9*. https://doi.org/10.15451/EC2020-05-9.13-1-27

Dierking, J., Phelps, L., Præbel, K., Ramm, G., Prigge, E., Borcherding, J., Brunke, M., & Eizaguirre, C. (2014). Anthropogenic hybridization between endangered migratory and commercially harvested stationary whitefish taxa (Coregonus spp.). *Evolutionary Applications*, *7*(9), 1068–1083. https://doi.org/10.1111/eva.12166

Dowle, E. J., Morgan-Richards, M., Brescia, F., & Trewick, S. A. (2015). Correlation between shell phenotype and local environment suggests a role for natural selection in the evolution of Placostylus snails. *Molecular Ecology*, *24*(16), 4205–4221. https://doi.org/10.1111/mec.13302

Edgar, R. C. (2004). MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, *5*. https://doi.org/10.1186/1471-2105-5-113

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, *14*(8), 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Fraley, C., & Raftery, A. E. (2006). *MCLUST Version 3: An R Package for Normal Mixture Modeling and Model-Based Clustering*. *504*.

Gramlich, S., Wagner, N. D., & Hörandl, E. (2018). RAD-seq reveals genetic structure of the F 2 -generation of natural willow hybrids (Salix L.) and a great potential for interspecific introgression. *BMC Plant Biology*, *18*(1). https://doi.org/10.1186/s12870-018-1552-6

Haase, M., Greve, C., Hutterer, R., & Misof, B. (2014). Amplified fragment length polymorphisms, the evolution of the land snail genus Theba (Stylommatophora: Helicidae), and an objective approach for relating fossils to internal nodes of a phylogenetic tree using geometric morphometrics. *Zoological Journal of the Linnean Society*, *171*(1), 92–107. https://doi.org/10.1111/zoj.12123

Harrison, R. G., & Larson, E. L. (2016). Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. *Molecular Ecology*, *25*(11), 2454–2466. https://doi.org/10.1111/mec.13582

Hasselman, D. J., Argo, E. E., McBride, M. C., Bentzen, P., Schultz, T. F., Perez-Umphrey, A. A., & Palkovacs, E. P. (2014). Human disturbance causes the formation of a hybrid swarm between two naturally sympatric fish species. *Molecular Ecology*, *23*(5), 1137–1152. https://doi.org/10.1111/mec.12674

Hollander, J., Montaño-Rendón, M., Bianco, G., Yang, X., Westram, A. M., Duvaux, L., Reid, D. G., & Butlin, R. K. (2018). Are assortative mating and genital divergence driven by reinforcement? *Evolution Letters*, *2*(6), 557–566. https://doi.org/10.1002/evl3.85

Jombart, T. (2008). Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., & Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), 1647–1649. https://doi.org/10.1093/bioinformatics/bts199

Klingenberg, C. P. (2011). MorphoJ: An integrated software package for geometric morphometrics. *Molecular Ecology Resources*, *11*(2), 353–357. https://doi.org/10.1111/j.1755-0998.2010.02924.x

Klingenberg, C. P. (2016). Size, shape, and form: concepts of allometry in geometric morphometrics. *Development Genes and Evolution*, *226*(3), 113–137. https://doi.org/10.1007/s00427-016-0539-2

Lagarde, L., & Ouetcho, A. (2017). Horticultural structures on ultramafic soils: the case of Isle of Pines and other parts of southern Grande Terre (New Caledonia). In F. Valentin & G. Molle (Eds.), *La pratique de l'espace en Océanie: découverte, appropriation et émergence des systèmes sociaux traditionnels* (Issue December 2016). Societe prehistorique Francaise.

Lawson, D. J., van Dorp, L., & Falush, D. (2018). A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-05257-7

Leigh, J. W., & Bryant, D. (2015). popart : full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, *6*(9), 1110–1116. https://doi.org/10.1111/2041-210X.12410

Lowe, A. J., & Abbott, R. J. (2015). Hybrid swarms: catalysts for multiple evolutionary events in Senecio in the British Isles. *Plant Ecology and Diversity*, *8*(4), 449–463. https://doi.org/10.1080/17550874.2015.1028113

Madec, L., & Bellido, A. (2007). Spatial variation of shell morphometrics in the subantarctic land snail Notodiscus hookeri from Crozet and Kerguelen Islands. *Polar Biology*, *30*(12), 1571–1578. https://doi.org/10.1007/s00300-007-0318-7

Marie, A. D., Bernatchez, L., & Garant, D. (2011). Empirical assessment of software efficiency and accuracy to detect introgression under variable stocking scenarios in brook charr (Salvelinus fontinalis). *Conservation Genetics*, *12*(5), 1215–1227. https://doi.org/10.1007/s10592-011-0224-y

Mitteroecker, P., & Gunz, P. (2009). Advances in Geometric morphometrics. *Evolutionary Biology*, *36*(2), 235–247. https://doi.org/10.1007/s11692-009-9055-x

Moura, A. E., Tsingarska, E., Dąbrowski, M. J., Czarnomska, S. D., Jędrzejewska, B., & Pilot, M. (2014). Unregulated hunting and genetic recovery from a severe population decline: The cautionary case of Bulgarian wolves. *Conservation Genetics*, *15*(2), 405–417. https://doi.org/10.1007/s10592-013-0547-y

Neubert, E., Chérel-mora, C., & Bouchet, P. (2009). Polytypy , clines , and fragmentation : The bulimes of New Caledonia revisited. *Mémoires Du Muséum d'Histoire Naturelle*, *198*, 37–131.

Nieto Feliner, G., Álvarez, I., Fuertes-Aguilar, J., Heuertz, M., Marques, I., Moharrek, F., Piñeiro, R., Riina, R., Rosselló, J. A., Soltis, P. S., & Villa-Machío, I. (2017). Is homoploid hybrid speciation that rare? An empiricist's view. *Heredity*, *118*(6), 513–516. https://doi.org/10.1038/hdy.2017.7

Noor, M. A. F. (1999). Reinforcement and other consequences of sympatry. *Heredity*, *83*(5), 503–508. https://doi.org/10.1038/sj.hdy.6886320

Parrish, G. R., Stringer, I. A. N., & Sherley, G. H. (2014). The biology of Placostylus ambagiosus (Pulmonata: Bulimulidae) in New Zealand: Part 1. Behaviour, habitat use,

abundance, site fidelity, homing and the dimensions of eggs and snails. *Molluscan Research*, *34*(3), 139–154. https://doi.org/10.1080/13235818.2014.888980

Pfeiffer, L. (1851). Description of Fifty-four New Species of Helicea, from the Collection of Hugh Cuming, Esq. *Proceedings of the Zoological Society of London*. https://doi.org/10.1111/j.1096-3642.1851.tb01174.x

Prentis, P. J., White, E. M., Radford, I. J., Lowe, A. J., & Clarke, A. R. (2007). Can hybridization cause local extinction: A case for demographic swamping of the Australian native Senecio pinnatifolius by the invasive Senecio madagascariensis? *New Phytologist*, *176*(4), 902–912. https://doi.org/10.1111/j.1469-8137.2007.02217.x

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959. https://doi.org/10.1111/j.1471-8286.2007.01758.x

Quenu, M., Trewick, S. A., Brescia, F., & Morgan-Richards, M. (2020). Geometric morphometrics and machine learning challenge currently accepted species limits of the land snail Placostylus (Pulmonata: Bothriembryontidae) on the Isle of Pines, New Caledonia. *Journal of Molluscan Studies*, *86*(1), 35–41. https://doi.org/10.1093/mollus/eyz031

R Development Core Team, R., & R Core Team. (2017). R: A language and environment for statistical computing. *R: A Language and Environment for Statistical Computing*. https://doi.org/10.1016/j.jssas.2015.06.002

Rhymer, J. M., & Simberloff, D. (1996). Extinction by hybridization and introgression. *Annual Review of Ecology and Systematics*, *27*, 83–109. https://doi.org/10.1146/annurev.ecolsys.27.1.83

Rohlf, F. J. (2015). The tps series of software. *Hystrix*, *26*(1), 1–4. https://doi.org/10.4404/hystrix-26.1-11264

Schilthuizen, M., & Haase, M. (2010). Disentangling true shape differences and experimenter bias: Are dextral and sinistral snail shells exact mirror images? *Journal of Zoology*, *282*(3), 191–200. https://doi.org/10.1111/j.1469-7998.2010.00729.x

Sloan, D. B., Havird, J. C., & Sharbrough, J. (2017). The on-again, off-again relationship between mitochondrial genomes and species boundaries. *Molecular Ecology*, *26*(8), 2212–

2236. https://doi.org/10.1111/mec.13959

Thompson, K. A., Urquhart-Cronish, M., Whitney, K. D., Rieseberg, L. H., & Schluter, D. (2019). Patterns, predictors, and consequences of dominance in hybrids. *BioRxiv*. https://doi.org/10.1101/818658

Todesco, M., Pascual, M. A., Owens, G. L., Ostevik, K. L., Moyers, B. T., Hübner, S., Heredia, S. M., Hahn, M. A., Caseys, C., Bock, D. G., & Rieseberg, L. H. (2016). Hybridization and extinction. *Evolutionary Applications*, *9*(7), 892–908. https://doi.org/10.1111/eva.12367

Vallejo-Marín, M., Cooley, A. M., Lee, M. Y., Folmer, M., McKain, M. R., & Puzey, J. R. (2016). Strongly asymmetric hybridization barriers shape the origin of a new polyploid species and its hybrid ancestor. *American Journal of Botany*, *103*(7), 1272–1288. https://doi.org/10.3732/ajb.1500471

Vaux, F., Crampton, J. S., Marshall, B. A., Trewick, S. A., & Morgan-Richards, M. (2017). Geometric morphometric analysis reveals that the shells of male and female siphon whelks Penion chathamensis are the same size and shape. *Molluscan Research*, *37*(3), 194–201. https://doi.org/10.1080/13235818.2017.1279474

Vaux, F., Gemmell, M. R., Hills, S. F. K., Marshall, B. A., Beu, A. G., Crampton, J. S., Trewick, S. A., & Morgan-Richards, M. (2020). Lineage identification affects estimates of evolutionary mode in marine snails. *Systematic Biology*. https://doi.org/10.1093/sysbio/syaa018

Vaux, F., Trewick, S. A., Crampton, J. S., Marshall, B. A., Beu, A. G., Hills, S. F. K., & Morgan-Richards, M. (2018). Evolutionary lineages of marine snails identified using molecular phylogenetics and geometric morphometric analysis of shells. *Molecular Phylogenetics and Evolution*, *127*(February), 626–637. https://doi.org/10.1016/j.ympev.2018.06.009

Zelditch, M., Swiderski, D., & Sheets, H. D. (2004). *Geometric morphometrics for biologists A primer*.

Chapter 5: Oxygen and Carbon Isotopes from *Placostylus* snail shells; implications for paleoclimate studies

Abstract

The two stable isotope ratio $\delta^{13}C$ and $\delta^{18}O$ retrieved from the carbonate shell of terrestrial snails can be used as an environmental proxy and are thought to reflect dietary composition and climatic conditions (precipitation, humidity, temperature) snails live in. Here I develop a protocol to sample high-resolution isotopic records of snails of the genus *Placostylus,* which are endemic to the southwest Pacific. This protocol is used to sample isotopic profiles from nine snail shells coming from two locations in New Caledonia and one location in New Zealand. I found that snails from New Zealand had on average higher $\delta^{18}O$ values than their counterparts in New Caledonia, while the reverse relationship is expected in $\delta^{18}O$ values of environmental water between these two regions. Snails from New Caledonia had noticeable drops in their $\delta^{18}O$ values, which could be linked to extreme precipitation events in this region (*e.g.* tropical storm, hurricane). The isotopic ratio $\delta^{13}C$ was found to be highly variable even among snails collected at the same location, and snails from New Zealand had on average slightly higher $\delta^{13}C$ values than their counterparts in New Caledonia.

Introduction

Quantitative reconstructions of past climate rely on the use of proxy systems, where the physical or chemical properties of an archive are used to infer past environmental conditions (Evans et al., 2013). Classical examples of biological proxies include tree-ring records of fossil or extant trees (Lorrey et al., 2018), or isotopic profiles retrieved from the carbonate shell produced by terrestrial or marine molluscs (*e.g.* Bivalves, gastropods; Apolinarska et al., 2015; Jones & Quitmyer, 1996) Fossil land snails have been used in a wide range of paleoclimatic and paleoenvironmental studies. Land snail are widespread today and their shells are usually well preserved in Quaternary sediments (Goodfriend, 1992; Leng & Lewis, 2016). Taxonomic analyses of community assemblages have been used to produce qualitative paleoenvironmental reconstructions (Brook, 1999b, 1999a; Moine et al., 2002), and analyses of chemical properties of fossil shells can provide quantitative proxies of past climate conditions (Goodfriend, 1992). Among the chemical properties of the snail shell the carbon and oxygen stable isotope composition are the most well understood (Yanes et al., 2009). These isotope ratios can be used in two key ways: (1) providing insights into the organism's diet, which can in turn, provide insights into past vegetation assemblages (Stott, 2002); and (2) be used as a proxy for past climatic variables (e.g. humidity, temperature, relative humidity). To achieve reliable inferences about the past, analyses of modern specimens are required, which allow the comparison of stable isotopic composition of extant specimens with local environmental variability (Yanes et al., 2019). Here I provide the first modern land snail isotopic data ($\delta^{13}C$ and $\delta^{18}O$) for two regions of the southwest Pacific (New Zealand and New Caledonia). A protocol is developed to sample high-resolution isotopic data from giant snail shells of the genus *Placostylus*. The stable isotopic signal is compared to climatic variables, and differences in isotopic composition between snails from two contrasting geographic locations and between sympatric species of snails are investigated.

Land snails and stable isotopes

Oxygen stable isotope data are typically presented as values of $\delta^{18}O$, which are defined by the following formula:

$$\delta 18O = \left(\frac{\frac{18O\ sample}{16O\ sample}}{\frac{18O\ standard}{16Ostandard}} - 1 \right) \times 1000$$

It reflects parts per thousand (denoted as ‰) enrichments or depletions relative to a standard of known composition, which for $\delta^{18}O$ values is the Vienna Standard Mean Ocean Water (VSMOW). $\delta^{18}O$ values obtained from snail shells are thought to mainly reflect the $\delta^{18}O$ signature of the water in the snail's local environment, which is in turn controlled by a range of climatic factors (Balakrishnan & Yapp, 2004; Baldini et al., 2007; Goodfriend et al., 1989; Goodfriend & Ellis, 2002; Lécolle, 1985; Yanes et al., 2017; Yapp, 1979). Correlations between climate variables and snail shell isotopic value means the shell $\delta^{18}O$ can effectively be used as a quantitative proxy for a range of climatic variables. However, with different sets of correlations for different geographical areas no universal relationship has been found (Yanes et al., 2019). Snail shells are produced by a precipitation reaction which takes place in the mantle (pallium) of the organism, with the oxygen component of the shell coming from the body water of the snails. In theory shell $\delta^{18}O$ is therefore controlled by the $\delta^{18}O$ of snail body water, and the temperature at which the precipitation reaction takes place. Mechanistic flux balance models have hypothesised a strong relation between relative humidity (RH) and shell $\delta^{18}O$ (Balakrishnan & Yapp, 2004). This has been confirmed in some empirical settings (Balakrishnan, Yapp, Theler, et al., 2005; Zaarur et al., 2011), but not universally (Magaritz et al., 1981). Studies using multiple whole-shell $\delta^{18}O$ data from locations displaying a significant geographical or elevation gradient have confirmed local correlations with precipitation and temperature variables (Bao et al., 2019, 2020; Colonese et al., 2014; Yanes et al., 2009). Based on this, snail shell $\delta^{18}O$ appears to be generally negatively correlated to precipitation rate, and generally positively correlated to air temperature. Direct comparisons of snail shell $\delta^{18}O$ to the $\delta^{18}O$ of environmental water has revealed almost universally higher $\delta^{18}O$ values in snail shells (Yanes et al., 2019).

In addition to $\delta^{18}O$, snail shell carbonate also preserves values of $\delta^{13}C$, which are defined by the following equation:

$$\delta 13C = \left(\frac{\frac{13C\ sample}{12C\ sample}}{\frac{13C\ standard}{12Cstandard}} - 1 \right) \times 1000$$

As with $\delta^{18}O$ values it is expressed in parts per thousand (‰) enrichment or depletion relative to a standard of known composition which in the case of $\delta^{13}C$ values is the Vienna Pee Dee Belemnite (VPDB). In terrestrial gastropods $\delta^{13}C$ values are thought to be primarily influenced by snail diet and in herbivorous land snails is correlated to the relative ingestion of C3 / C4

plants (Balakrishnan, Yapp, Meltzer, et al., 2005; Bao et al., 2018; Bing et al., 2012; Goodfriend, 1992; Goodfriend et al., 1989; Goodfriend & Magaritz, 1987; Yapp, 1979). Because of their metabolic pathways C4 plants (e.g. tropical grass) typically have a high $\delta^{13}C$ and therefore higher snail shell $\delta^{13}C$ values can usually be interpreted as higher relative C4 plant intake (Balakrishnan & Yapp, 2004). Changes in shell $\delta^{13}C$ over large geographical areas representing a partition of plant community assemblages have been demonstrated (Colonese et al., 2014; Prendergast et al., 2017; Yanes et al., 2009), but high-resolution shell analyses have also revealed that changes in $\delta^{13}C$ values can be associated with switches in snail diet (Baldini et al., 2007). In controlled lab environments positive correlations have been noticed between snail shell $\delta^{13}C$ values and diet composition, with higher $\delta^{13}C$ values for land snails fed with C4 plants (Metref et al., 2003; Stott, 2002). Beside dietary carbon, other sources of carbon for carbonate shell include atmospheric CO_2 and direct ingestion of non-biogenic carbonate matter such as limestone (Yanes et al., 2012). It is not entirely clear how all these sources of carbon interact and interfere with the $\delta^{13}C$ of the snail shell, and studies relying on the use of radiocarbon isotope ^{14}C have demonstrated the proportion of dietary carbon present in the carbonate shell of the snail seems to vary between species of snail or geographical area (Bing et al., 2012; Goodfriend & Magaritz, 1987; Pigati et al., 2010). Changes in $\delta^{13}C$ values in chronological fossil series of snails have the potential to inform us about snails palaeodietary composition and can be extrapolated to infer changes in past plant community assemblages, which are often correlated to changes in past climatic conditions (Balakrishnan, Yapp, Meltzer, et al., 2005; Padgett et al., 2019; Yanes & Romanek, 2013). It is however a complex signal and inferring patterns in fossil material require extensive studies of modern specimens from the same geographical region.

Stable isotopic composition of land snail can either be obtained from whole-shell and reflect an average isotopic value over the snail's lifespan, or high-resolution approaches can be developed and sample multiple isotopic values within the same shell. These approaches can be used to infer climatic relationships at a sub-annual resolution, with sampling strategies that follow ontogeny of the snail (Yanes et al., 2012). In India, high-resolution $\delta^{18}O$ series from the invasive giant African snail species *Lissachatina fulica* have been used to characterise high precipitation monsoon season (Ghosh et al., 2017). In the Bahamas high-resolution $\delta^{18}O$ records of snails from the genus *Cerion* have linked drops in $\delta^{18}O$ values to months of highest precipitation (Baldini et al., 2007). Differences in $\delta^{18}O$ values between snail species have received little attention, but comparison of sympatric species indicate isotopic composition can

differ significantly for snails of different ecology, size and ethology, with larger land snails showing consistently higher $\delta^{18}O$ value than their smaller counterparts (Yanes et al., 2017).

Geographical and climatic setting

Snail specimens for this study were sampled from two geographical regions in the Pacific: the far North of New Zealand and New Caledonia (Figure 2 and 3). New Caledonia consists of an assemblage of tropical islands located in a low latitude region (22°) east of Australia. Studies of rainwater isotopic composition in tropical islands have suggested that in these areas $\delta^{18}O$ values fluctuate with monthly precipitation, with higher isotopic values found during drier months (Edirisinghe et al., 2017). For New Caledonia, rainwater isotopic composition has only been investigated for a two year time period in Lifou, an island north-east of the archipelagos (Nicolini et al., 2016), with lower stable isotopic values recorded during the wet season (December – April). No recording station has been incorporated in the GNIP database for this region, and therefore prediction of $\delta^{18}O$ values from associated models would rely on extrapolations based on latitude and elevation variables (Bowen & Wilkinson, 2002). Because of its geographical positioning the archipelago is also regularly affected by large tropical storm or hurricane events, which are associated with rainwater depleted in heavy isotopes (Miller et al., 2006; Sánchez-Murillo et al., 2019).

Because of its wide latitudinal spread and variety of micro-climates New Zealand climatology is considered complex, with multitudes of internal variabilities. External influences on rainwater isotopic constitution includes air masses movements from sub-tropical regions to the north, and sub-Antarctic regions to the south. However, for the whole of New Zealand only two stations have been incorporated in the Global Network for Isotopes in Precipitation (GNIP-Kaitaia and Invercargill), which is the data used by predictive models of rainwater stable isotope composition such as the Online Isotope in Precipitation Calculator (OIPC). Current efforts are however dedicated to improve both the geographical coverage of sampling stations within the country, and the accuracy of predictive models associated (Baisden et al., 2016; McComb et al., 2019). Noticeably an enrichment in heavy isotope was observed in the northern parts of the country (Northland), where *Placostylus* snails live. Like New Caledonia the North of New Zealand is sometime be subject to tropical storms, even though they are far less frequent and bring far less precipitation than the ones occurring in tropical latitudes.

The biological material: *Placostylus*

Placostylus is a genus of large terrestrial land snail endemic to the southwest Pacific. Extant species are found in the Solomon Islands, Fiji, Vanuatu, New Caledonia, Lord Howe Island and the north of New Zealand (Neubert et al., 2009). Quaternary fossil snail shell records are known to be particularly abundant in the north of New Zealand (Brook, 1999a), and records also exists at both Grande Terre and the Isle of Pines in New Caledonia (Neubert et al., 2009). Ecology of *Placostylus* species have been extensively studied for the two New Caledonian species *P. fibratus* and *P. porphyrostomus* (Brescia et al., 2008), and for the New Zealand species *P. ambagiosus* (Parrish et al., 2014). All three species are most active at night, and snails usually bury themselves in the leaf litter during the warmer daytime hours. Behavioural studies have demonstrated the New Caledonia species *P. fibratus* is most active, including reproductively, during the cold season (May to October) and has a more limited activity during the warmer wetter season (December to March). All three species are believed to feed on fresh leaves falling off surrounding trees, with their specific diet composition differing between geographical regions. Dispersal is very low with many marked individuals of *P. fibratus* and *P. porphyrostomus* found under the same tree or shrub for more than 12 months (Brescia et al. 2008). Similarly the New Zealand species *P. ambagiosus* has a strong homing behaviour and will tend to stay within the same area most of its life (Stringer et al., 2018).

Methods

2.1 Geographical sampling

Snail shells were sampled from three localities in two geographical regions: New Zealand and New Caledonia (Table 1). In New Caledonia live snails were sampled from two locations: Gadji (Figure 2A, -22.5607, 167.4304805) on the Isle of Pines, and Forêt Nord (Figure 2A, -22.093459, 166.700256) on Grande Terre. Shells from Gadji collected in April 2015 represent the two species *P. fibratus* and *P. porphyrostomus*, and shells from Forêt Nord collected between 2007 and 2012 were *P. fibratus* (Table 1). The shells from New Zealand were donated by the Museum of New Zealand Te Papa Tongarewa and were collected alive in 1949 in Spirits Bay, located in the Northernmost part of the North Island (Figure 3A, -34.4338, 172.89167). Within each location shells were collected on the same day. Overall, there were three shells sampled in the New Zealand location (*P. ambagiosus*), two shells sampled in the Forêt Nord location in New Caledonia (*P. fibratus*) and four shells collected in the Gadji location in New Caledonia (*P. fibratus* and *P. porphyrostomus*).

2.2 Meteorological and climate data

Meteorological data at a monthly resolution were obtained from three meteorological stations. Data from New Caledonia spanning 2005 to 2021 were provided by Meteo France. The two meteorological stations closest to snail sampling locations were Moue on the Isle of Pines (-22.5898 / 167.4521, closest to Gadji) and Rivière Blanche on Grande Terre (-22.1326 / 166.7263, closest to Forêt Nord). The nearest meteorological station to Spirits Bay is located at Kaitaia (-35.108 / 173.258), about 100km south of Spirits Bay (Figure 3A). Data obtained from this station spanned from 1967 to 1985. Three variables were gathered from the New Caledonia stations: Monthly cumulative Precipitation (mm, MP), Mean Monthly Temperature (°C, MMT) and Monthly Mean Humidity (%, MMH). Four variables were gathered from the New Zealand (Kaitaia) station: Monthly Precipitation, Monthly Mean of the Daily Maximum Air temperature, Monthly Mean of the Daily Minimum Air Temperature, and Monthly 9am Relative Humidity (RH). Annual climatology with averaged monthly statistics values were calculated for each variable with 95% confidence intervals over the timespan covered by meteorological stations. Model estimates of rainwater oxygen isotopic ratio $\delta^{18}O$ were computed for each sampling location using the Online Isotope in Precipitation Calculator (OIPC, Bowen & Wilkinson, 2002).

Table 1: Location and species information for the nine shells collected

Shell ID	Species	Geographical region	Sampling site	latitude	longitude	Collection date
Forêt Nord 8-1 (FN81)	*Placosylus fibratus*	New Caledonia	Forêt Nord	-22.093	166.7003	2007-2015
Forêt Nord 8-5 (FN85)	*Placosylus fibratus*	New Caledonia	Forêt Nord	-22.093	166.7003	2007-2015
Gadji 11 (GA11)	*Placostylus porphyrostomus*	New Caledonia	Gadji	-22.561	167.43	April 2015
Gadji 13 (GA13)	*Placostylus porphyrostomus*	New Caledonia	Gadji	-22.561	167.43	April 2015
Gadji 04 (GA04)	*Placosylus fibratus*	New Caledonia	Gadji	-22.561	167.43	April 2015
Gadji 08 (GA08)	*Placosylus fibratus*	New Caledonia	Gadji	-22.561	167.43	April 2015
New Zealand 1 (NZ1)	*Placostylus ambagiosus*	New Zealand	Spirits Bay	-34.434	172.892	1949
New Zealand 2 (NZ2)	*Placostylus ambagiosus*	New Zealand	Spirits Bay	-34.434	172.892	1949
New Zealand 3 (NZ3)	*Placostylus ambagiosus*	New Zealand	Spirits Bay	-34.434	172.892	1949

2.3 Shell structure and sampling considerations

Placostylus snails produce large conical shells ranging from 6 to 15cm in length, with a dextral orientation (Figure 1). Snails usually reach sexual maturity within three to five years. Once sexual maturity is complete snail shell growth stops, but the peristome (outer lip surrounding the aperture of the shell) keeps thickening at a rate roughly estimated between 0.1 and 0.4 mm/year for New Caledonian species (Brescia et al., 2008) and around 0.2mm/year for New Zealand species (Parrish et al., 1995). Microscopic photography of the peristome region revealed the presence of carbonate growth lines ranging from the outer edge of the lip towards the shell (Figure 1B), and consequentially this part of the shell was targeted for sampling for isotopic variation through time. In this configuration the outermost layers of carbonate matter are considered most recent. All shells chosen for sampling were determined to be from adult (mature) snails, and sampling/isotopic measurements were restricted to growth layers of carbonate matter formed during adulthood.

Peristomes were detached from whole shells using a circular saw and were embedded in epoxy resin. Thick (500 µm) and thin (250 µm) sections were sliced from the epoxy mount using the Accutom 50 (Struers). Thin sections were used for microscopic photography (Figure 1B), and thick sections for drilling and sampling of carbonate matter. Carbonate samples were drilled along lines estimated as parallel to the growth lines found in the peristome. All carbonate samples were extracted from thick sections using the New wave micromill (ESI) with a 250 µm wide drill and a spacing between sampling lines of 200-215 µm. After drilling approximately 150µg of carbonate powder was collected for each sample using a scalpel blade and put in standard 0.2ml plastic tubes. Carbonate samples were then analysed using mass spectrometers to quantify $\delta^{18}O$ (‰ VSMOW) and $\delta^{13}C$ (‰ VPDB). Four shells (FN8.1, FN8.5, GA11, GA13) were analysed using the MAT 253 mass spectrometer at the NIWA stable isotope laboratory. The other five shells (GA04, GA08, NZ01, NZ02, NZ03) were analysed using the Nu Instruments Horizon Continuous Flow IRMS (CF-IRMS) at the Mawson Analytical Spectrometry Services of the University of Adelaide.

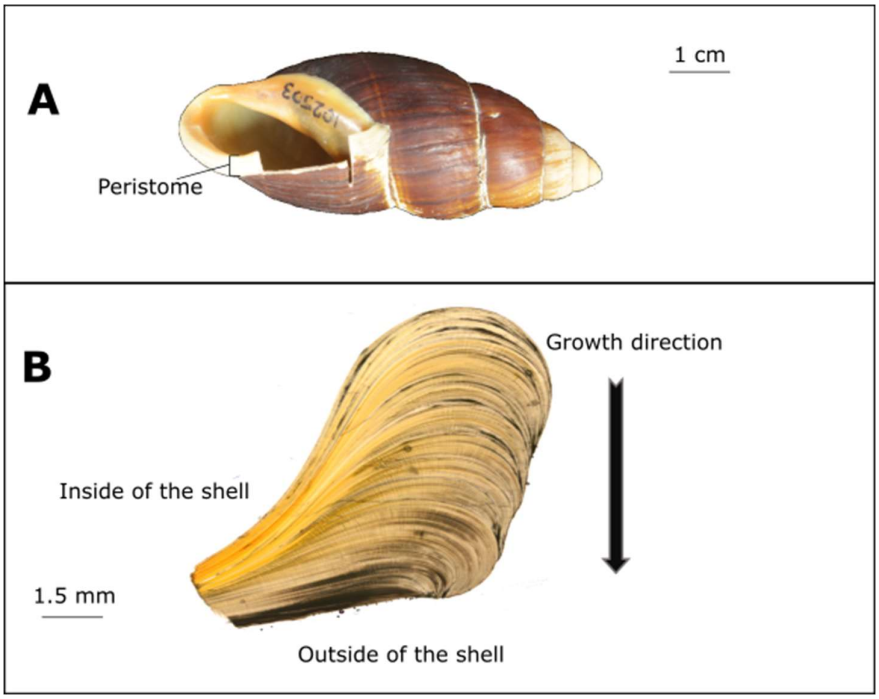


Figure 1: A) Picture of a *Placostylus* shell (here *P. ambagiosus*) with part of the peristome detached for isotopic sample analysis. B) Transversal view of the peristome seen with a transmission light microscope. Up to 38 carbonate samples could be taken on a single peristome by drilling lines following growth direction.

2.4 Statistical analysis

For both $\delta^{18}O$ and $\delta^{13}C$ summary statistics were computed for each individual shell (Mean, Standard deviation, Maximum and Minimum value). To detect temporal trends within shell high-resolution isotope sampling profiles simple linear regression models were fitted, with the isotopic ratio of interest as a response variable and the numbered sequential increment used as a numerical predictive variable. Positive linear relationships, highlighted by positive values of the coefficient of correlation in linear models, indicate trends with higher isotopic values towards outer carbonate increment of the snail lip (interpreted as more recent), and negative linear relationships indicate lower isotopic values towards outer layers.

Linear mixed models were used to compare shell isotopic data values ($\delta^{13}C$ and $\delta^{18}O$) among geographical areas and between species from the same location. To achieve this, analyses were done on three different datasets. To compare average isotopic values between New Zealand and New Caledonia the full dataset of nine shells was used. A subset of this dataset containing

six shells from New Caledonia was used to compare isotopic values between the two locations in New Caledonia (Gadji and Forêt Nord). The four shells collected in Gadji from two different species were used as a dataset to study variation at a species level (*P. fibratus* / *P. porphyrostomus*).

Linear mixed models were computed using the lmer package (Bates et al., 2015) in R 4.0.4 (R Development Core Team & R Core Team, 2017) and were built with the isotopic ratio of interest as a response variable, the shell ID as a random variable and either the geographical region or the snail species as a fixed parameter. In this configuration variations in the slope value of fixed parameters can be interpreted as an average variation in isotopic value between regions / species of interest, while variation between shells is being averaged around the intercept value of the model. Slope value of fixed parameter and dispersion parameter of random parameter are retrieved from models, and we use likelihood-ratio (LR) test between nested models to test whether adding geographical area / species of snail significantly improves overall model likelihoods.

Results

Climate of New Caledonia locations (Gadji and Forêt Nord) is tropical and characterised by high precipitation rates, high temperatures and a constantly high humidity (Figure 2). A wet and warm season (December - April) is followed by a relatively colder and drier season (May - November). Extreme precipitation events (hurricanes and tropical storms) usually occur between February and April. Within the timespan snails were collected (2007-2015) five major storm events can be identified (2008, 2009, 2011, 2012, 2013, Figure 2). Forêt Nord receives more rain than Gadji, both during extreme events and on a regular basis (Figure 2, B and C). Mean monthly temperature oscillates from 18°C to 26°C for both locations with a maximum in February and minimum around July-August. Mean Monthly Humidity (%) is almost constantly above 75% for both locations, and slightly higher during warmer months (December – April).

Climatic data of the New Zealand station (Kaitaia) revealed a more temperate climate characterised by oscillation between a warm and dry summer season (December - March) and a winter season with mild temperatures and relatively high precipitation (June - August). Precipitation rates were more consistent, and less impacted by extreme events than in New Caledonia. Between 1967 and 1985 Daily Maximum Air temperature oscillated between 15–26°C and Mean Daily Minimum Air Temperature oscillated between 5-15°C. Relative humidity recorded at 9 am was highest between May and August (around 90%) and lowest between October and February (below 80%).

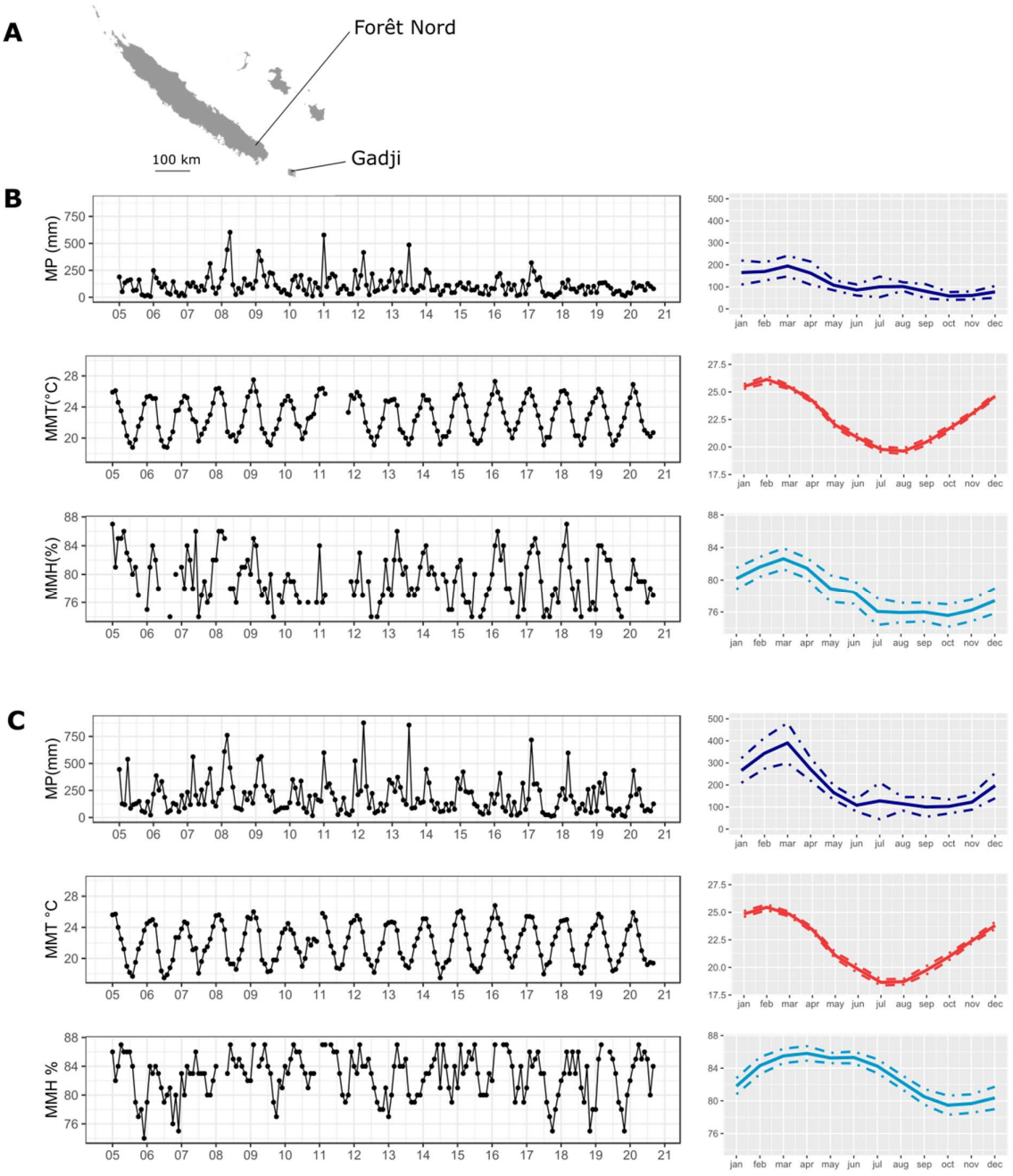


Figure 2: A) Map of New Caledonia with sampling locations of *Placostylus* shells. B and C) Meteorological (left) and one-year monthly averaged climatology (right) data for three variables of interest (Monthly cumulative Precipitation (MP), Monthly Mean Temperature (MMT) and Monthly Mean Humidity (MMH)) collected between 2005-2021 in meteorological stations close to Gadji (B) and Forêt Nord (C). Dotted lines surrounding climatology plots indicate 95% confidence intervals for the variable of interest.

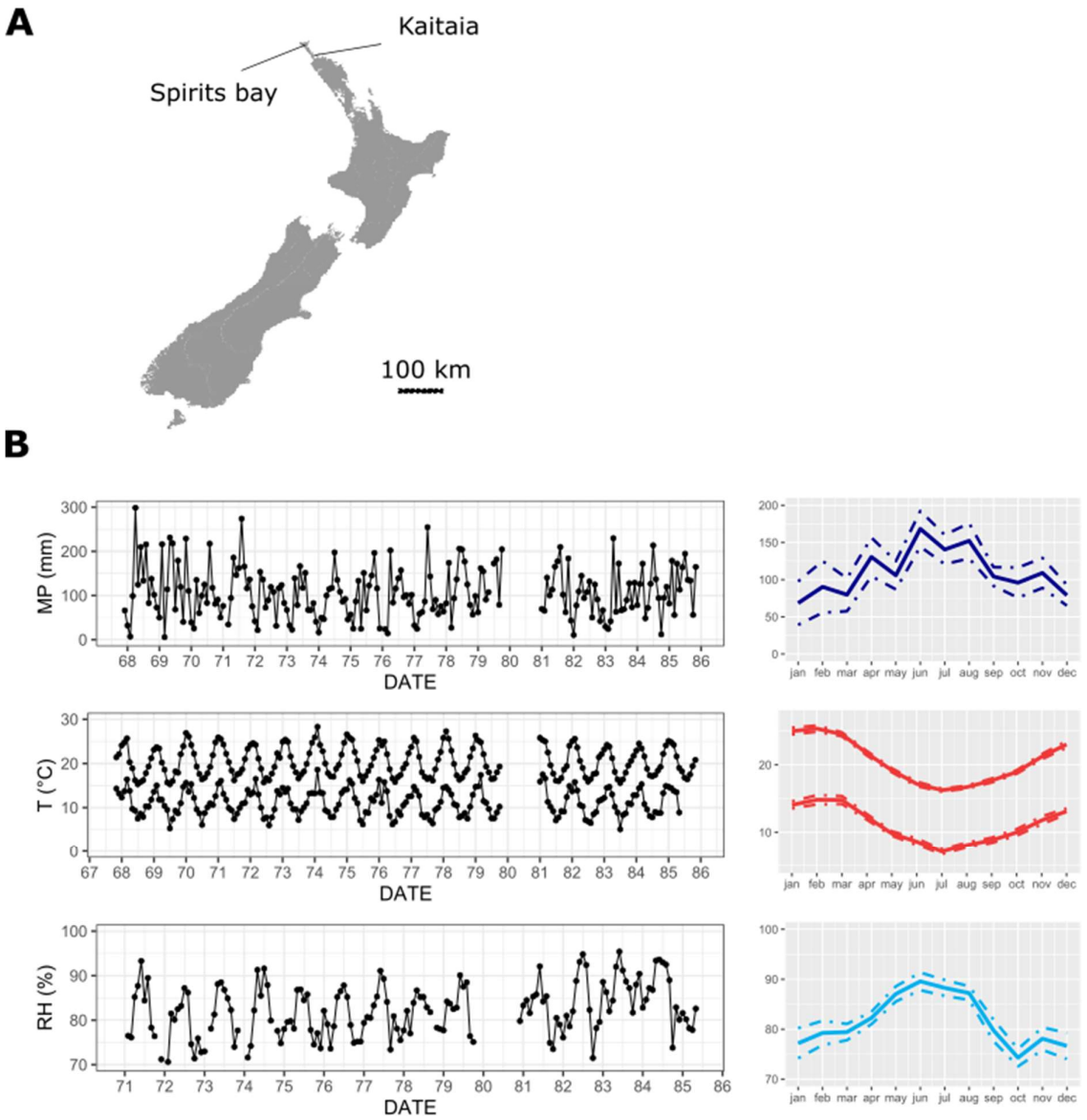


Figure 3: A) Map of New Zealand with sampling locations of *Placostylus* shells (Spirits Bay) and the meteorological station where climatic data were collected (Kaitaia) B) Meteorological (left) and one-year monthly averaged climatology (right) data for four variables of interest (Monthly cumulative Precipitation (MP), Monthly Mean of the Daily Maximum and Minimum Air Temperature (T) and Monthly Mean 9am Relative Humidity (RH)) collected between 1967-1985 in Kaitaia. Dotted lines surrounding climatology plots indicate 95% confidence intervals for the variable of interest.

Averaged shell $\delta^{18}O$ values range from -3.66 to -1.26‰ VSMOW (Table 2, Figure 4). A linear temporal trend in shell high-resolution profile was observed in only one of nine shells (GA11) for the oxygen isotopic ratio ($\delta^{18}O$; Table 2). Linear mixed models indicate higher average $\delta^{18}O$ values for shells of the New Zealand region (+ 1.23‰ VSMOW based on slope parameter, p-value = 0.003**, based on LR test). This contrasts with expectations based on $\delta^{18}O$ values in environmental water for collection locations, where OIPC model predicts lower $\delta^{18}O$ values in New Zealand (averaged annual $\delta^{18}O$ was -5.00 ‰ VSMOW for the New Zealand location, -3.03 and -3.38 ‰ VSMOW for the two New Caledonian locations based on OIPC, Table 2). Within New Caledonia the difference in $\delta^{18}O$ values between shells from Gadji and Forêt Nord was extremely small (+0.1‰ VSMOW for shells from Gadji, p-value = 0.83 based on LR test). In all cases, the variability of $\delta^{18}O$ values recorded in snail shells is much smaller than what is expected in environmental water on an annual timescale (Figure 4; Table 2, see standard deviations and range between minimal/maximal value). Comparison of isotopic shell $\delta^{18}O$ values from different species at a single location (*e. g.* Gadji, same collection date) reveals higher $\delta^{18}O$ values in *P. porphyrostomus* (n=2) compared to *P. fibratus* (n=2) (+ 1.15‰VSMOW based on slope parameter, p-value = 0.01** based on LR test).

Averaged shell $\delta^{13}C$ values range from -12.18 to -9.14‰ VPDB (Table 2, Figure 4 & 5). Linear temporal trends in shell high-resolution profiles are observed in seven shells for the carbon isotopic ratio ($\delta^{13}C$; FN8.5, GA04, GA11, GA13, NZ01, NZ02, NZ03*; e.g* Table 2), and are associated with higher values of dispersion parameters (standard deviation and range between maximal/minimal value, see Table 2). In New Caledonia $\delta^{13}C$ high-resolution profiles display a negative trend with lower isotopic values towards more recent carbonate increments, while in New Zealand a positive trend is observed with higher isotopic values in more recent layers (Table 3, Figure 3). Linear mixed models for carbon isotopic data $\delta^{13}C$ indicate higher average $\delta^{13}C$ values for shells of the New Zealand region (+ 1.87‰ VPDB based on slope parameter, p-value = 0.002** based on LR test). Within New Caledonia, shells from Gadji had marginally lower $\delta^{13}C$ values than shells from Forêt Nord (- 1.05‰ VPDB based on slope parameter, p-value = 0.07 based on LR test). Comparison of isotopic shell $\delta^{13}C$ values from different species at a single location (Gadji, same collection date) reveals minor $\delta^{13}C$ variations between the two New Caledonian species *P. fibratus* and *P. porphyrostomus* (+ 0.22‰ VPDB on average for *P. fibratus* based on slope parameter, p-value = 0.34 based on LR test). A strong difference in distribution of $\delta^{13}C$ values is observed between the two shells sampled in Forêt Nord (Table 2, Figure 3), which were collected on the exact same day from the same population.

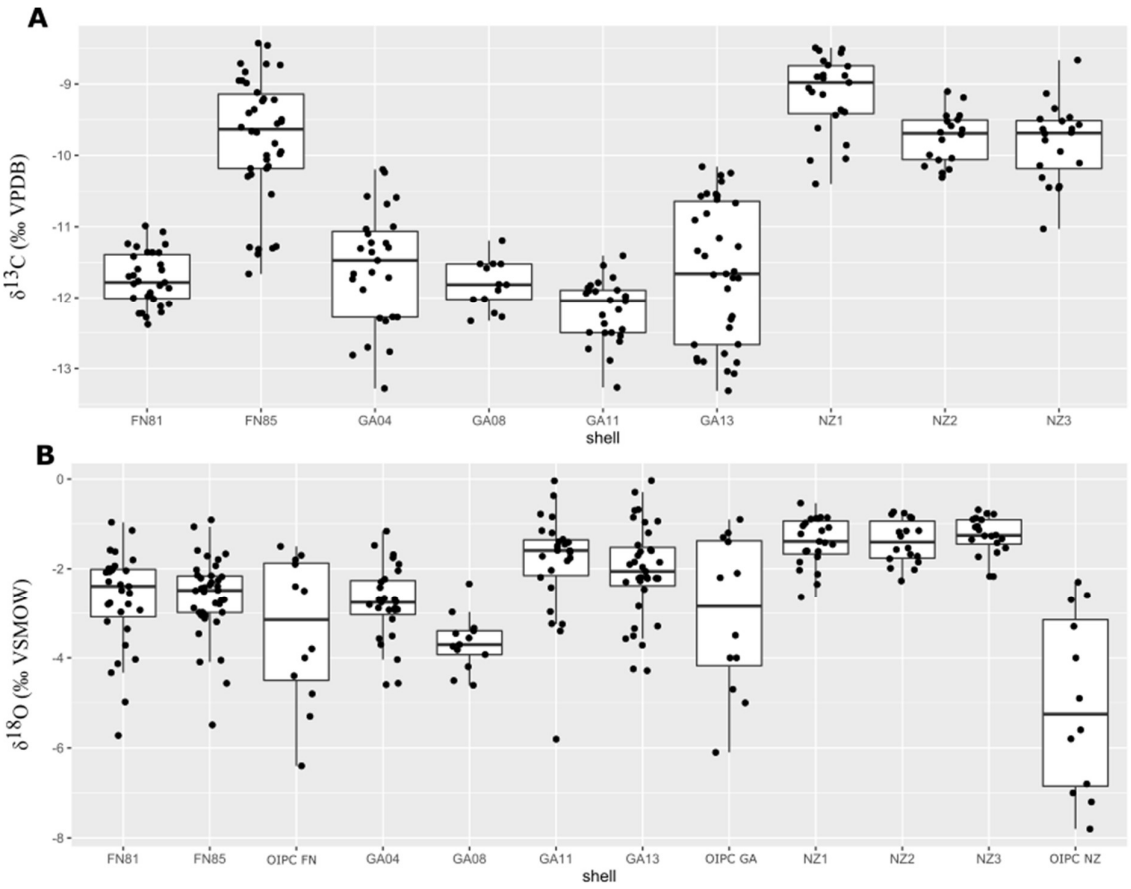


Figure 4: Pooled isotopic distributions of nine *Placostylus* shells from three different locations in New Zealand and New Caledonia. A) Distribution of carbon isotopic ratio $\delta^{13}C$. B) Distribution of oxygen isotopic ratio δ^{13C}. In addition to shell δ^{13C} variation, annual distributions of environmental rainwater δ^{13C} values are also displayed (calculated on the Online Isotopes in Precipitations Calculator).

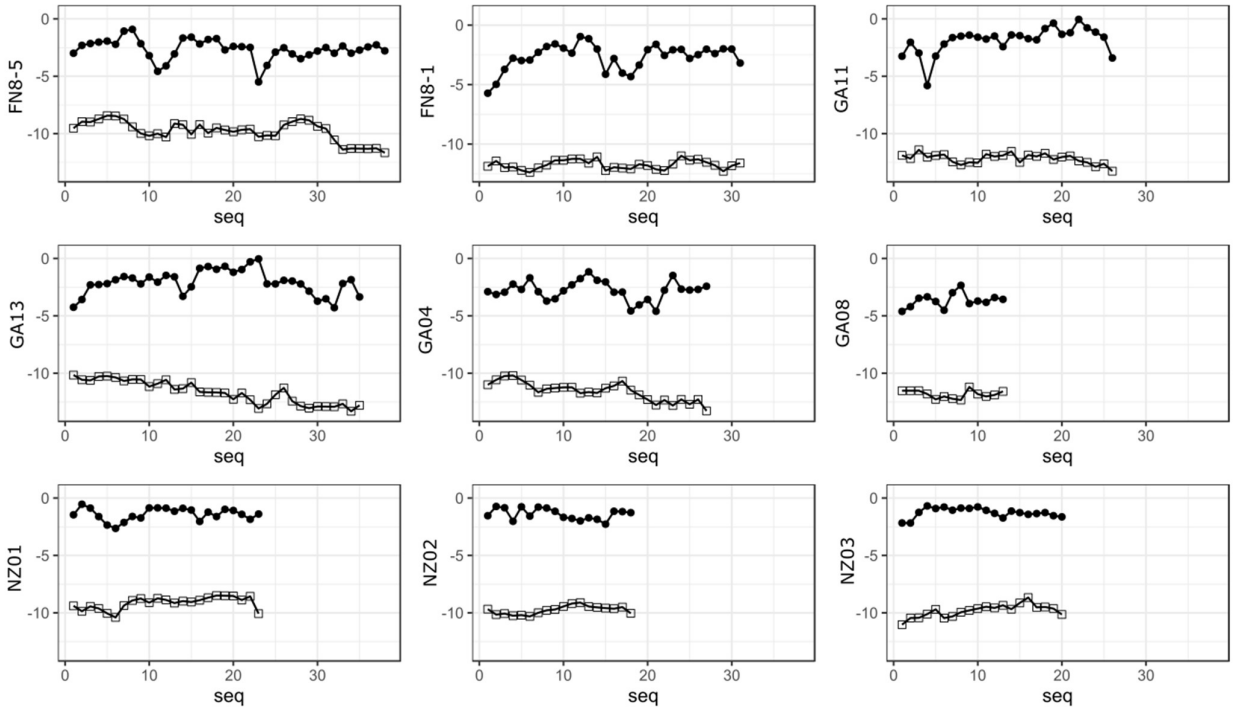


Figure 5: High resolution isotopic profiles ($\delta^{13}C$ and $\delta^{18}O$) for nine *Placostylus* shells from three different locations and two geographical areas. Isotopic values are plotted on the y-axis (filled circled line: $\delta^{18}O$ expressed in ‰ VSMOW, empty squared line: $\delta^{13}C$ expressed in ‰ VPDB) against the sequential number sample of carbonate matter (with later sample being interpreted as being ontologically more recent).

Table 2: Isotopic summary statistics for nine *Placostylus* shells, comprising statistics on aggregated $\delta^{18}O$ (‰ VSMOW) and $\delta^{13}C$ (‰ VPDB) values (sd: standard deviation, min: minimal isotopic value, max: maximum isotopic value) and statistics for linear models showing temporal trends in high-resolution isotopic profiles within each shell separately (ß: slope parameter of the linear models, R-squared: percentage of variation explained by the linear model, p-value: p-value computed from a t-test on whether ß differs from zero). Summary statistics for estimated environmental water $\delta^{18}O$ are also presented here and are based on OIPC monthly estimation data for the three sampling locations.

Shell ID	Carbon isotope ratio $\delta^{13}C$							Oxygen isotope ratio δ^{13C}						
	mean	sd	min	max	ß	R-squared	p-value	mean	sd	min	max	ß	R-squared	p-value
FN8.1	-11.74	0.38	-12.38	-10.99	0.00	0.00	0.59	-2.68	1.09	-5.72	-0.96	0.03	0.07	0.075
FN8.5	-9.78	0.88	-11.66	-8.43	-0.05	0.40	1.4E-05***	-2.63	0.89	-5.49	-0.91	-0.01	0.03	0.146
GA04	-11.58	0.81	-13.28	-10.20	-0.08	0.71	1.9E-08***	-2.78	0.85	-4.60	-1.16	-0.01	0.00	0.682
GA08	-11.82	0.35	-12.33	-11.20	-0.01	0.00	0.633	-3.66	0.61	-4.61	-2.34	0.05	0.03	0.263
GA11	-12.18	0.44	-13.26	-11.41	-0.03	0.16	0.0231*	-1.86	1.16	-5.81	-0.04	0.07	0.24	0.006**
GA13	-11.65	1.00	-13.31	-10.16	-0.09	0.85	8.75E-16***	-2.07	1.06	-4.29	-0.03	-0.01	0.00	0.709
NZ01	-9.14	0.55	-10.40	-8.49	0.04	0.25	0.008**	-1.40	0.54	-2.64	-0.54	0.01	0.00	0.684
NZ02	-9.76	0.36	-10.31	-9.11	0.04	0.24	0.023*	-1.40	0.49	-2.27	-0.73	-0.03	0.05	0.185
NZ03	-9.83	0.54	-11.03	-8.67	0.06	0.44	7.96E-04***	-1.26	0.43	-2.17	-0.68	0.00	0.00	0.847
OIPC Gadji								-3.03	1.74	-6.10	-0.90			
OIPC Foret Nord								-3.38	1.63	-6.40	-1.50			
OIPC Cape reinga								-5.00	1.98	-7.80	-2.30			

Discussion

Using a high-resolution sampling approach, the variability of stable isotopic values can be explored within a timeframe defined by the seasonal activity and lifespan of the biological proxy. With a sampling protocol focused on sampling incremental layers of carbonate matters of the peristome we can expect to record four to ten years of carbonate growth for an adult *Placostylus* snail (Brescia et al., 2008; Parrish et al., 1995). Direct comparisons of climatic variables between New Zealand and New Caledonia locations were not possible for temperature and relative humidity due to differences in the way climatic variables were measured. The patterns of variability and comparison of precipitation rates however showed a clear distinction between the tropical climatic conditions in New Caledonia and a more temperate regime in the North of New Zealand. To be active and produce shell material land snails require a high relative humidity, and relatively high temperatures (Nicolai & Ansart, 2017). These conditions are more likely to be met throughout the entire year in New Caledonia than in New Zealand. As a result, we can expect isotopic values in snail shells of different regions to record a different fraction of their annual climate variability.

Placostylus shell $\delta^{18}O$ values were on average higher in New Zealand than in New Caledonia (Figure 4). In contrast, the $\delta^{18}O$ of rainwater predicted by the OIPC was on average lower in New Zealand than in New Caledonia. This can be explained by either an actual enrichment in heavy isotope 18O in precipitation water of the New Zealand location, or by biological fractionation processes. The deficiency of GNIP recording stations for the southwest Pacific region means offsets between actual rainwater isotopic values and values predicted by the model are quite likely, and predictions from OIPC would mostly be based on extrapolation of relationship between $\delta^{18}O$ and latitude/ altitude variables (Bowen & Wilkinson, 2002). Recent isotopic sampling of New Zealand rainwater showed high $\delta^{18}O$ values in the northern part of the country (Baisden et al., 2016), which might partially explain the high $\delta^{18}O$ values in New Zealand *Placostylus* shells. Biological fractionation of oxygen isotopes from environmental water towards carbonate snail shell usually results in higher $\delta^{18}O$ values in carbonate shell material (Balakrishnan & Yapp, 2004; Yanes et al., 2019). Here this is what is observed in most shells of New Caledonia and is exaggerated in shells of New Zealand. A high temperature for the carbonate shell precipitation reaction could potentially explain those high $\delta^{18}O$ values (Zhang et al., 2018), but given the extent of 18O enrichment in New Zealand shells it is unlikely this alone could explain the observed enrichment. The shells from New Zealand have a very low $\delta^{18}O$ variability compared to both the ones in New Caledonia, which could be due to them

only recording a fraction of the annual isotopic variability found in the rainwater. Combined with high $\delta^{18}O$ values these results suggests New Zealand *Placostylus* snails would be active and produce shell material only during months where rainwater isotopic values are the highest (December – March based on OIPC predictions).

The shells from New Caledonia were sampled in two locations, one of which (Forêt Nord) receiving far more precipitation than the other. Despite this the $\delta^{18}O$ values were not lower in Forêt Nord, as a negative correlation between precipitation rate and shell $\delta^{18}O$ would predict (Bao et al., 2019; Prendergast et al., 2015; Yanes et al., 2009). Given the close geographical proximity of the two locations we can hypothesise the precipitation water received would mostly be of the same origin, and therefore share similar $\delta^{18}O$ values regardless of differences in precipitation rate. In five shells (FN8.1, FN8.5, GA04, GA11, GA13) the pattern of variation was similar and characterised by stable periods of relatively high isotopic values followed by sudden drops in $\delta^{18}O$ values. This pattern was not observed in *Placostylus* shells from New Zealand. Other high-resolution records of tropical land snails have similarly reported short periods of low isotopic values, which have been linked to extreme rainfall events during the monsoon season in Indian samples of *Lissachatina fulica* (Ghosh et al., 2017), and to months higher precipitation in *Cerion* snails from the Bahamas (Baldini et al., 2007). With the presence of multiple extreme rainfall events (hurricanes or tropical storms) within the sampled timeframe we expect to be recorded by the high-resolution sampling method it seems likely those drops in $\delta^{18}O$ values are signals of periods of intense rainfall.

When sampling snails at the same location environmental variation should be limited and therefore differences in $\delta^{18}O$ values must reflect differences in biological fractionation processes. Higher $\delta^{18}O$ values are usually observed for larger species of snails (Yanes et al., 2017), and it can be explained by a potentially higher shell carbonate precipitation temperature. From Gadji four shells were collected and represent two distinct *Placostylus* species. The two *P. fibratus* shells had on average lower $\delta^{18}O$ values than the two *P. porphyrostomus* shells, but is the bigger of the two species (Dowle et al., 2015; Quenu et al., 2020). With only two samples of each species this comparison represents a minimal dataset, but the results are opposite to the trend previously reported. This suggests that controls for $\delta^{18}O$ enrichment in *Placostylus* are not only related to shell size but might incorporate other factors such as shell coloration or snail behaviour (Wang et al., 2016).

The signal recorded by snail shell $\delta^{13}C$ values is thought to mainly reflect diet composition (Bao et al., 2018). Values for carbon isotope ratio $\delta^{13}C$ were highly variable for *Placostylus* snails of New Caledonia, and on average lower than values found in New Zealand. Higher $\delta^{13}C$ isotopic values can reflect a higher intake in C4 plants if the diet of herbivorous land snail includes a mix of C3 and C4 plants or can be related to the direct ingestion of non-biogenic carbonate matter. *Placostylus* snails are believed to forage and feed mostly on leaves falling in their litter environment (Brescia et al., 2008; Parrish et al., 2014), and these can include a variety of $\delta^{13}C$ signatures. The strong variability of $\delta^{13}C$ signatures in shells of New Caledonia, sometimes within the same location, suggests a non-specialist diet. Sympatric shells of *P. fibratus* and *P. porphyrostomus* did not demonstrate differences in $\delta^{13}C$ values, which indicates a similar diet composition for the two species. Temporal trends in $\delta^{13}C$ values were observed in both New Zealand and New Caledonia but are of opposite direction with an increase in $\delta^{13}C$ values over time in New Zealand, and a decrease in $\delta^{13}C$ values over time in New Caledonia (Figure 5, Table 2). These trends could either reflect a change in preferential diet over the lifespan of terrestrial snails or be the signal of an external environmental force.

In conclusion, I provided the first stable isotopic data for land snails species in two regions with contrasting climatic regimes. The different patterns observed in shells from New Zealand and New Caledonia suggests snail shells record isotopic variability differently in the two regions, and seasonality and tropical storms seems to play an important role in the observed $\delta^{18}O$ and $\delta^{13}C$ values. The use of land snail shell isotopic composition as a biological proxy for paleoclimate reconstruction is a relatively new and emerging field of study (Yanes et al., 2019), and there are still very few studies that have looked at high-resolution profiles (Baldini et al., 2007; Ghosh et al., 2017). Hence the methodological aspects of elucidating stable isotope values in relation to environmental variables and their seasonal variation are still largely missing in comparison to other proxy systems (Judd et al., 2018; van der Sleen et al., 2017). Developing robust modelling approaches linking seasonal growth of carbonate matter, shell isotopic composition and environmental variables could help provide more accurate understanding of stable isotope signal in extant and fossil snail records.

References

Apolinarska, K., Pełechaty, M., & Noskowiak, D. (2015). Differences in stable isotope compositions of freshwater snails from surface sediments of two Polish shallow lakes. *Limnologica*, *53*, 95–105. https://doi.org/10.1016/j.limno.2015.06.003

Baisden, W. T., Keller, E. D., Van Hale, R., Frew, R. D., & Wassenaar, L. I. (2016). Precipitation isoscapes for New Zealand: enhanced temporal detail using precipitation-weighted daily climatology†. *Isotopes in Environmental and Health Studies*, *52*(4–5), 343–352. https://doi.org/10.1080/10256016.2016.1153472

Balakrishnan, M., & Yapp, C. J. (2004). Flux balance models for the oxygen and carbon isotope compositions of land snail shells. *Geochimica et Cosmochimica Acta*, *68*(9), 2007–2024. https://doi.org/10.1016/j.gca.2003.10.027

Balakrishnan, M., Yapp, C. J., Meltzer, D. J., & Theler, J. L. (2005). Paleoenvironment of the Folsom archaeological site, New Mexico, USA, approximately 10,500 14C yr B.P. as inferred from the stable isotope composition of fossil land snail shells. *Quaternary Research*, *63*(1), 31–44. https://doi.org/10.1016/j.yqres.2004.09.010

Balakrishnan, M., Yapp, C. J., Theler, J. L., Carter, B. J., & Wyckoff, D. G. (2005). Environmental significance of 13C/12C and 18O/16O ratios of modern land-snail shells from the southern great plains of North America. *Quaternary Research*, *63*(1), 15–30. https://doi.org/10.1016/j.yqres.2004.09.009

Baldini, L. M., Walker, S. E., Railsback, L. B., Baldini, J. U. L., & Crowe, D. E. (2007). Isotopic ecology of the modern land snail Cerion, San Salvador, Bahamas: Preliminary advances toward establishing a low-latitude island paleoenvironmental proxy. *Palaios*, *22*(2), 174–187. https://doi.org/10.2110/palo.2005.p05.091r

Bao, R., Sheng, X., Li, C., Shen, H., Tan, L., Sun, L., Li, C., Peng, H., Luo, L., Wu, M., Lu, H., Ji, J., & Chen, J. (2020). Effect of altitude on the stable carbon and oxygen isotopic compositions of land snails at the margin of the East Asian monsoon. *Geochimica et Cosmochimica Acta*. https://doi.org/10.1016/j.gca.2020.01.029

Bao, R., Sheng, X., Lu, H., Li, C., Luo, L., Shen, H., Wu, M., Ji, J., & Chen, J. (2019). Stable carbon and oxygen isotopic composition of modern land snails along a precipitation gradient in the mid-latitude East Asian monsoon region of China. *Palaeogeography,*

Palaeoclimatology, Palaeoecology, *533*. https://doi.org/10.1016/j.palaeo.2019.109236

Bao, R., Sheng, X., Teng, H. H., & Ji, J. (2018). Reliability of shell carbon isotope composition of different land snail species as a climate proxy: A case study in the monsoon region of China. *Geochimica et Cosmochimica Acta*, *228*, 42–61. https://doi.org/10.1016/j.gca.2018.02.022

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Bing, X., Gu, Z., Han, J., Zongxiu, L., Yunpeng, P., Yanwu, L., Naiqin, W., & Yongfu, C. (2012). Radiocarbon and stable carbon isotopes analyses of land snails from the Chinese loess plateau: environmental and chronological implications. *Quaternary International*, *279–280*(1), 543–544. https://doi.org/10.1016/j.quaint.2012.08.1909

Bowen, G. J., & Wilkinson, B. (2002). Spatial distribution of δ18O in meteoric precipitation. *Geology*, *30*(4), 315–318. https://doi.org/10.1130/0091-7613(2002)030<0315:SDOOIM>2.0.CO;2

Brescia, F. M., Pöllabauer, C. M., Potter, M. A., & Robertson, A. W. (2008). A review of the ecology and conservation of Placostylus (Mollusca: Gastropoda: Bulimulidae) in New Caledonia. *Molluscan Research*, *28*(2), 111–122.

Brook, F. J. (1999a). Stratigraphy, landsnail faunas, and paleoenvironmental history of coastal dunefields at te werahi, northernmost new zealand. *Journal of the Royal Society of New Zealand*, *29*(4), 361–393. https://doi.org/10.1080/03014223.1999.9517603

Brook, F. J. (1999b). Stratigraphy and landsnail faunas of late holocene coastal dunes, tokerau beach, northern new zealand. *Journal of the Royal Society of New Zealand*, *29*(4), 337–359. https://doi.org/10.1080/03014223.1999.9517602

Colonese, A. C., Zanchetta, G., Fallick, A. E., Manganelli, G., Lo Cascio, P., Hausmann, N., Baneschi, I., & Regattieri, E. (2014). Oxygen and carbon isotopic composition of modern terrestrial gastropod shells from Lipari Island, Aeolian Archipelago (Sicily). *Palaeogeography, Palaeoclimatology, Palaeoecology*, *394*, 119–127. https://doi.org/10.1016/j.palaeo.2013.12.003

Dowle, E. J., Morgan-Richards, M., Brescia, F., & Trewick, S. A. (2015). Correlation

between shell phenotype and local environment suggests a role for natural selection in the evolution of Placostylus snails. *Molecular Ecology*, *24*(16), 4205–4221. https://doi.org/10.1111/mec.13302

Edirisinghe, E. A. N. V., Pitawala, H. M. T. G. A., Dharmagunawardhane, H. A., & Wijayawardane, R. L. (2017). Spatial and temporal variation in the stable isotope composition (δ18O and δ2H) of rain across the tropical island of Sri Lanka. *Isotopes in Environmental and Health Studies*, *53*(6), 628–645. https://doi.org/10.1080/10256016.2017.1304936

Evans, M. N., Tolwinski-Ward, S. E., Thompson, D. M., & Anchukaitis, K. J. (2013). Applications of proxy system modeling in high resolution paleoclimatology. In *Quaternary Science Reviews* (Vol. 76, pp. 16–28). https://doi.org/10.1016/j.quascirev.2013.05.024

Ghosh, P., Rangarajan, R., Thirumalai, K., & Naggs, F. (2017). Extreme Monsoon Rainfall Signatures Preserved in the Invasive Terrestrial Gastropod Lissachatina fulica. *Geochemistry, Geophysics, Geosystems*, *18*(11), 3758–3770. https://doi.org/10.1002/2017GC007041

Goodfriend, G. A. (1992). The use of land snail shells in paleoenvironmental reconstruction. *Quaternary Science Reviews*, *11*(6), 665–685. https://doi.org/10.1016/0277-3791(92)90076-K

Goodfriend, G. A., & Ellis, G. L. (2002). Stable carbon and oxygen isotopic variations in modern Rabdotus land snail shells in the southern Great Plains, USA, and their relation to environment. *Geochimica et Cosmochimica Acta*, *66*(11), 1987–2002. https://doi.org/10.1016/s0016-7037(02)00824-4

Goodfriend, G. A., & Magaritz, M. (1987). Carbon and oxygen isotope composition of shell carbonate of desert land snails. *Earth and Planetary Science Letters*, *86*(2–4), 377–388. https://doi.org/10.1016/0012-821X(87)90234-2

Goodfriend, G. A., Magaritz, M., & Gat, J. R. (1989). Stable isotope composition of land snail body water and its relation to environmental waters and shell carbonate. *Geochimica et Cosmochimica Acta*, *53*(12), 3215–3221. https://doi.org/10.1016/0016-7037(89)90102-6

Jones, D. S., & Quitmyer, I. R. (1996). Marking time with bivalve shells: Oxygen isotopes and season of annual increment formation. *Palaios*, *11*(4), 340–346. https://doi.org/10.2307/3515244

Judd, E. J., Wilkinson, B. H., & Ivany, L. C. (2018). The life and time of clams: Derivation of intra-annual growth rates from high-resolution oxygen isotope profiles. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *490*, 70–83. https://doi.org/10.1016/j.palaeo.2017.09.034

Lécolle, P. (1985). The oxygen isotope composition of landsnail shells as a climatic indicator: Applications to hydrogeology and paleoclimatology. *Chemical Geology: Isotope Geoscience Section*, *58*(1–2), 157–181. https://doi.org/10.1016/0168-9622(85)90036-3

Leng, M. J., & Lewis, J. P. (2016). Oxygen isotopes in Molluscan shell: Applications in environmental archaeology. *Environmental Archaeology*, *21*(3), 295–306. https://doi.org/10.1179/1749631414Y.0000000048

Lorrey, A. M., Boswijk, G., Hogg, A., Palmer, J. G., Turney, C. S. M., Fowler, A. M., Ogden, J., & Woolley, J. M. (2018). The scientific value and potential of New Zealand swamp kauri. *Quaternary Science Reviews*. https://doi.org/10.1016/j.quascirev.2017.12.019

Magaritz, M., Heller, J., & Volokita, M. (1981). Land-air boundary environment as recorded by the 18O/16O and 13C/12C isotope ratios in the shells of land snails. *Earth and Planetary Science Letters*, *52*(1), 101–106. https://doi.org/10.1016/0012-821X(81)90212-0

McComb, K., Sarker, S., Hoogewerff, J., Hayman, A., & Frew, R. (2019). A δ2H Isoscape of blackberry as an example application for determining the geographic origins of plant materials in New Zealand. *PLoS ONE*, *14*(12). https://doi.org/10.1371/journal.pone.0226152

Metref, S., Rousseau, D. D., Bentaleb, I., Labonne, M., & Vianey-Liaud, M. (2003). Study of the diet effect on δ13C of shell carbonate of the land snail Helix aspersa in experimental conditions. *Earth and Planetary Science Letters*, *211*(3–4), 381–393. https://doi.org/10.1016/S0012-821X(03)00224-3

Miller, D. L., Mora, C. I., Grissino-Mayer, H. D., Mock, C. J., Uhle, M. E., & Sharp, Z.

(2006). Tree-ring isotope records of tropical cyclone activity. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(39), 14294–14297. https://doi.org/10.1073/pnas.0606549103

Moine, O., Rousseau, D. D., Jolly, D., & Vianey-Liaud, M. (2002). Paleoclimatic reconstruction using mutual climatic range on terrestrial mollusks. *Quaternary Research*, *57*(1), 162–172. https://doi.org/10.1006/qres.2001.2286

Neubert, E., Chérel-mora, C., & Bouchet, P. (2009). Polytypy , clines , and fragmentation : The bulimes of New Caledonia revisited. *Mémoires Du Muséum d'Histoire Naturelle*, *198*, 37–131.

Nicolai, A., & Ansart, A. (2017). Conservation at a slow pace: Terrestrial gastropods facing fast-changing climate. *Conservation Physiology*, *5*(1). https://doi.org/10.1093/conphys/cox007

Nicolini, E., Rogers, K., & Rakowski, D. (2016). Baseline geochemical characterisation of a vulnerable tropical karstic aquifer; Lifou, New Caledonia. *Journal of Hydrology: Regional Studies*, *5*, 114–130. https://doi.org/10.1016/j.ejrh.2015.11.014

Padgett, A., Yanes, Y., Lubell, D., & Faber, M. L. (2019). Holocene cultural and climate shifts in NW Africa as inferred from stable isotopes of archeological land snail shells. *Holocene*, *29*(6), 1078–1093. https://doi.org/10.1177/0959683619831424

Parrish, G. R., Stringer, I. A. N., & Sherley, G. H. (2014). The biology of Placostylus ambagiosus (Pulmonata: Bulimulidae) in New Zealand: Part 1. Behaviour, habitat use, abundance, site fidelity, homing and the dimensions of eggs and snails. *Molluscan Research*, *34*(3), 139–154. https://doi.org/10.1080/13235818.2014.888980

Parrish, R., Sherley Greg, & Aviss, M. (1995). Giant land snail recovery plan, Placostylus spp., Paryphanta sp. In *Threatened Species Recovery Plan Series* (Vol. 13, Issue 13).

Pigati, J. S., Rech, J. A., & Nekola, J. C. (2010). Radiocarbon dating of small terrestrial gastropod shells in North America. *Quaternary Geochronology*, *5*(5), 519–532. https://doi.org/10.1016/j.quageo.2010.01.001

Prendergast, A. L., Stevens, R. E., Barker, G., & O'Connell, T. C. (2015). Oxygen isotope signatures from land snail (Helix melanostoma) shells and body fluid: Proxies for reconstructing Mediterranean and North African rainfall. *Chemical Geology*, *409*, 87–

98. https://doi.org/10.1016/j.chemgeo.2015.05.014

Prendergast, A. L., Stevens, R. E., Hill, E. A., Hunt, C., O'Connell, T. C., & Barker, G. W. (2017). Carbon isotope signatures from land snail shells: Implications for palaeovegetation reconstruction in the eastern Mediterranean. *Quaternary International*, *432*, 48–57. https://doi.org/10.1016/j.quaint.2014.12.053

Quenu, M., Trewick, S. A., Brescia, F., & Morgan-Richards, M. (2020). Geometric morphometrics and machine learning challenge currently accepted species limits of the land snail Placostylus (Pulmonata: Bothriembryontidae) on the Isle of Pines, New Caledonia. *Journal of Molluscan Studies*, *86*(1), 35–41. https://doi.org/10.1093/mollus/eyz031

R Development Core Team, R., & R Core Team. (2017). R: A language and environment for statistical computing. *R: A Language and Environment for Statistical Computing.* https://doi.org/10.1016/j.jssas.2015.06.002

Sánchez-Murillo, R., Durán-Quesada, A. M., Esquivel-Hernández, G., Rojas-Cantillano, D., Birkel, C., Welsh, K., Sánchez-Llull, M., Alonso-Hernández, C. M., Tetzlaff, D., Soulsby, C., Boll, J., Kurita, N., & Cobb, K. M. (2019). Deciphering key processes controlling rainfall isotopic variability during extreme tropical cyclones. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-12062-3

Stott, L. D. (2002). The influence of diet on the δ13C of shell carbon in the pulmonate snail Helix aspersa. *Earth and Planetary Science Letters*, *195*(3–4), 249–259. https://doi.org/10.1016/S0012-821X(01)00585-4

Stringer, I. A. N., Parrish, G. R., & Sherley, G. H. (2018). Homing, dispersal and mortality after translocation of long-lived land snails Placostylus ambagiosus and P. hongii (Gastropoda: Bothriembryontidae) in New Zealand. *Molluscan Research*, *38*(1), 56–76. https://doi.org/10.1080/13235818.2017.1323368

van der Sleen, P., Zuidema, P. A., & Pons, T. L. (2017). Stable isotopes in tropical tree rings: theory, methods and applications. *Functional Ecology*, *31*(9), 1674–1689. https://doi.org/10.1111/1365-2435.12889

Wang, X., Cui, L., Zhai, J., & Ding, Z. (2016). Stable and clumped isotopes in shell carbonates of land snails Cathaica sp. and Bradybaena sp. in north China and

implications for ecophysiological characteristics and paleoclimate studies. *Geochemistry, Geophysics, Geosystems*, *17*(1), 219–231. https://doi.org/10.1002/2015GC006182

Yanes, Y., Al-Qattan, N. M., Rech, J. A., Pigati, J. S., Dodd, J. P., & Nekola, J. C. (2019). Overview of the oxygen isotope systematics of land snails from North America. *Quaternary Research (United States)*, *91*(1), 329–344. https://doi.org/10.1017/qua.2018.79

Yanes, Y., Gutiérrez-Zugasti, I., & Delgado, A. (2012). Late-glacial to Holocene transition in northern Spain deduced from land-snail shelly accumulations. *Quaternary Research (United States)*, *78*(2), 373–385. https://doi.org/10.1016/j.yqres.2012.06.008

Yanes, Y., Nekola, J. C., Rech, J. A., & Pigati, J. S. (2017). Oxygen stable isotopic disparities among sympatric small land snail species from northwest Minnesota, USA. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *485*, 715–722. https://doi.org/10.1016/j.palaeo.2017.07.029

Yanes, Y., & Romanek, C. S. (2013). Quaternary interglacial environmental stability in San Salvador Island (Bahamas): A land snail isotopic approach. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *369*, 28–40. https://doi.org/10.1016/j.palaeo.2012.09.019

Yanes, Y., Romanek, C. S., Delgado, A., Brant, H. A., Noakes, J. E., Alonso, M. R., & Ibáñez, M. (2009). Oxygen and carbon stable isotopes of modern land snail shells as environmental indicators from a low-latitude oceanic island. *Geochimica et Cosmochimica Acta*, *73*(14), 4077–4099. https://doi.org/10.1016/j.gca.2009.04.021

Yapp, C. J. (1979). Oxygen and carbon isotope measurements of land snail shell carbonate. *Geochimica et Cosmochimica Acta*, *43*(4), 629–635. https://doi.org/10.1016/0016-7037(79)90170-4

Zaarur, S., Olack, G., & Affek, H. P. (2011). Paleo-environmental implication of clumped isotopes in land snail shells. *Geochimica et Cosmochimica Acta*, *75*(22), 6859–6869. https://doi.org/10.1016/j.gca.2011.08.044

Zhang, N., Yamada, K., Kano, A., Matsumoto, R., & Yoshida, N. (2018). Equilibrated clumped isotope signatures of land-snail shells observed from laboratory culturing

experiments and its environmental implications. *Chemical Geology*, *488*, 189–199. https://doi.org/10.1016/j.chemgeo.2018.05.001

Chapter 6: Thesis Discussion

Placostylus diversity and implications for conservation

In the southwest Pacific there is debate about what are the main processes explaining the distribution of extant species (Keppel et al., 2009). Both vicariance and dispersal have been suggested to explain phylogeographic patterns of some taxa, with dispersal hypotheses (long distance dispersal, island hopping) being largely favoured by molecular data and geological evidence (Bryan et al., 2012; Cibois et al., 2007; Filardi & Moyle, 2005). Terrestrial snails have limited active dispersal capability and therefore if oceanic dispersal occurred it would most likely involve rare passive colonisation events followed by successful local establishment (Hendriks et al., 2019; Santos et al., 2009). Establishment in new locations could be facilitated by snails being hermaphroditic and/or having the ability to store sperm (*e.g.* Ozgo et al., 2016). With a better taxon sampling across the southwest pacific and the generation of more substantial phylogenetic markers biogeographical hypotheses could be tested using the terrestrial *Placostylus* snails. They provide an excellent study system and filling the current gaps in the knowledge of *Placostylus* phylogenetic diversity could allow interpretations of phylogeographic patterns at a large scale.

Here (Chapter 1, Figure 3)analyses made on small datasets (Cox1) highlighted some interesting sister relationships between *Placostylus* snails from Lord Howe Island and the Solomon Islands, and between *Placostylus* snails from the Three King Islands and New Caledonia. Respectively these could be an indication of New Caledonia being a source for *Placostylus* snails in the Three king Islands, and snails from Lord Howe Island and the Solomon Island sharing the same source population. Phylogenies based on larger datasets (mitochondrial genome + multi-copy nuclear markers) did not however support some of the phylogenetic relationships found with short sequences analyses (Chapter 1). This could mean that some of phylogenetic relationships inferred from short DNA sequences could be misleading. Considering this, future sampling of *Placostylus* species in the southwest Pacific region should both focus on locations where the taxon sampling is still insufficient (Fiji, Vanuatu, Solomon Island), and should also include larger sequence data. The method used in Chapter 1 where high-quality DNA extraction methods are followed by the production of paired-end short reads genomic libraries (around 2Gb) seems like a good cost-efficient method to produce large mitochondrial and nuclear genetic datasets and could potentially be used on museum material for species with a critical threatened status (*e. g. Placostylus bivaricosus* on Lord Howe Island, *Placostylus* species found in Fiji or Vanuatu).

On a small-scale analysing *Placostylus* snail shell morphological variation in combination to genetic data can help understand local patterns of population structure for a genus of snail that comprise many vulnerable species. In this thesis the morphological and genetic diversity of *Placostylus* snails was extensively studied at one location, the Isle of Pines. This location was of particular interest because of the coexistence of the vulnerable species *P. porphyrostomus* with the commercially harvested species *P. fibratus.* Understanding local population structure and possible interactions between the two species was important because it defines what shells can be harvested, and where to prioritise conservation efforts. The sample of snails collected on the island was found to contain only two nuclear genetic clusters which is consistent with the current taxonomy on the island. Two morphotypes of snails were found for the species *P. fibratus,* highlighting the possible plasticity of snail shells from this genus.

The data presented here and in previous studies in this archipelago makes New Caledonia one of the places where *Placostylus* has been studied the most, alongside New Zealand (Buckley et al., 2011; Dowle et al., 2015). Regions that have received little attention include the northern parts of Grande Terre, which is a biodiversity hotspot and hosts the species *P. eddystonensis*, *P. caledonicus*, *P. scarabus* and *P. bondeensi* (Neubert et al., 2009; Wulff et al., 2013). Apart from *P. caledonicus* no molecular data have been produced for those taxa. Most of them are endangered and facing environmental degradation, so sampling tissue for DNA analysis might be difficult and/or not advisable. In this context morphological analyses performed on empty shells could be a good solution to analyse the diversity of snails present in the northern parts of New Caledonia. Geometric Morphometric methods are robust for *Placostylus* (chapter 3; Daly et al., 2020). They can be associated with clustering analysis to infer the most appropriate number of shells morphotypes and be linked to genetic data if opportunities for DNA extraction are limited.

Placostylus and stable isotopes

The protocol developed in chapter 5 opens a range of possibilities for isotopic analyses using shells of land snails such as *Placostylus*. The main objective of this work was to develop a biological proxy system which can be used to infer past environmental conditions via the analysis of fossil shells. Isotopic variations of the two stable isotope ratios $\delta^{13}C$ and $\delta^{18}O$ are commonly used in a range of paleoclimate proxies and can be used to infer long-term climate changes. For *Placostylus* there are collections of fossil shells from the Quaternary which could be directly used for isotopic analysis (Brook, 1999). These shells were collected at the Northern most region of New Zealand (Cape Maria van Diem and Tom Bowling Bay) and have radiocarbon dates ranging from 1000 Bp to > 40,000 BP (Daly, 2017). Overlaps between the dating of those shells and the dating of other paleoclimate proxies in this region (fossil remains of *Agathis australis*, kauri trees) means multi-proxy studies are possible, and New Zealand *Placostylus* isotopic shell values could be analysed in complement to other information (Alloway et al., 2007). Particularly if high-resolution methods are being used *Placostylus* shells could potentially provide insights into sub-annual climate events (*e.g.* frequency of high rainfall events for this region) for time periods where the global climate trend is already covered by other proxy systems. In New Caledonia and other regions of the southwest Pacific (Fiji, Solomon Islands) there are no existing collections of pre-dated shells, but fossil *Placostylus* shells can be found at some locations (*e.g.* buried *Placostylus* shells in sand dunes near Bourail / Poya in Grande Terre, New Caledonia), and sampling opportunities exist for Quaternary paleoclimate studies.

Additional sampling of extant specimens should also be considered to further investigate the relationship between snail shell $\delta^{18}O$ and $\delta^{13}C$ values and environmental conditions. Extra sampling of New Caledonia shells should be straightforward as *P. fibratus* is fairly abundant and can legally be harvested in the Isle of Pines, but sampling of extant shell material in New Zealand and other regions of the southwest Pacific might be more difficult as *Placostylus* snails from those regions are often endangered and sampling of live snails often prohibited. Collecting empty shells from those areas could be an alternative solution, but this would mean the period recorded by the shell isotopic variability is unknown. Sampling the isotopic composition of snails reared in controlled environments (Stringer, 2007) could also be realized and comparisons with stable isotopic composition of wild snails would help us better understand the dietary and climatic controls of snail shell $\delta^{13}C$ and $\delta^{18}O$ values.

The analysis of stable isotopic composition of terrestrial snail shells is a relatively new and emerging field of study (Yanes et al., 2019). Most published studies have involved analysis of whole-shell isotopic values, and high-resolution sampling of shells is still very rare. The methodological aspects and statistical approaches used to analyse this kind of data are therefore still being developed. In other proxy systems (*e. g*. high-resolution $\delta^{18}O$ records of Marine Bivalves; Judd et al., 2018) isotopic growth models have been used to seasonally resolve biological proxies and understand relationships between high-resolution isotopic signal and environmental controls (*e.g.* ocean temperature for marine bivalve). These approaches work by comparing the annual isotopic signal obtained from the biological proxy to a mixing curve of environmental variables. Adapted to land snail high-resolution $\delta^{18}O$ records they could help better understand the balance of climatic controls on carbonate shell $\delta^{18}O$ (temperature, precipitation, $\delta^{18}O$ in precipitation water), and could be used to infer the variability in seasonal shell growth.

References

Alloway, B. V., Lowe, D. J., Barrell, D. J. A., Newnham, R. M., Almond, P. C., Augustinus, P. C., Bertler, N. A. N., Carter, L., Litchfield, N. J., McGlone, M. S., Shulmeister, J., Vandergoes, M. J., Williams, P. W., Anderson, B., Brackley, H., Burge, P., Carter, J., Cochran, U., Cooke, P., … Zondervan, A. (2007). Towards a climate event stratigraphy for New Zealand over the past 30 000 years (NZ-INTIMATE project). *Journal of Quaternary Science*, *22*(1), 9–35. https://doi.org/10.1002/jqs.1079

Brook, F. J. (1999). Stratigraphy, landsnail faunas, and paleoenvironmental history of coastal dunefields at te werahi, northernmost new zealand. *Journal of the Royal Society of New Zealand*, *29*(4), 361–393. https://doi.org/10.1080/03014223.1999.9517603

Bryan, S. E., Cook, A. G., Evans, J. P., Hebden, K., Hurrey, L., Colls, P., Jell, J. S., Weatherley, D., & Firn, J. (2012). Rapid, long-distance dispersal by pumice rafting. *PLoS ONE*, *7*(7). https://doi.org/10.1371/journal.pone.0040583

Buckley, T., Stringer, I., Gleeson, D., Howitt, R., Attanayake, D., Parrish, R., Sherley, G., & Rohan, M. (2011). A revision of the New Zealand Placostylus land snails using mitochondrial DNA and shell morphometric analyses, with implications for conservation. *New Zealand Journal of Zoology*, *38*(1), 55–81. https://doi.org/10.1080/03014223.2010.527997

Cibois, A., Thibault, J. C., & Pasquet, E. (2007). Uniform phenotype conceals double colonization by reed-warblers of a remote Pacific archipelago. *Journal of Biogeography*, *34*(7), 1150–1166.

Daly, E. (2017). *Fine scale population structure through space and time*. Massey.

Daly, E. E., Trewick, S. A., Dowle, E. J., Crampton, J. S., & Richards, M. M. (2020). Conservation of pupu whakarongotaua the snail that listens for the war party. *Ethnobiology and Conservation*, *9*. https://doi.org/10.15451/EC2020-05-9.13-1-27

Dowle, E. J., Morgan-Richards, M., Brescia, F., & Trewick, S. A. (2015). Correlation between shell phenotype and local environment suggests a role for natural selection in the evolution of Placostylus snails. *Molecular Ecology*, *24*(16), 4205–4221. https://doi.org/10.1111/mec.13302

Filardi, C. E., & Moyle, R. G. (2005). Single origin of a pan-Pacific bird group and upstream

colonization of Australasia. *Nature*, *438*(7065), 216–219. https://doi.org/10.1038/nature04057

Hendriks, K. P., Alciatore, G., Schilthuizen, M., & Etienne, R. S. (2019). Phylogeography of Bornean land snails suggests long-distance dispersal as a cause of endemism. *Journal of Biogeography*, *46*(5), 932–944. https://doi.org/10.1111/jbi.13546

Judd, E. J., Wilkinson, B. H., & Ivany, L. C. (2018). The life and time of clams: Derivation of intra-annual growth rates from high-resolution oxygen isotope profiles. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *490*, 70–83. https://doi.org/10.1016/j.palaeo.2017.09.034

Keppel, G., Lowe, A. J., & Possingham, H. P. (2009). Changing perspectives on the biogeography of the tropical South Pacific: Influences of dispersal, vicariance and extinction. *Journal of Biogeography*, *36*(6), 1035–1054. https://doi.org/10.1111/j.1365-2699.2009.02095.x

Neubert, E., Chérel-mora, C., & Bouchet, P. (2009). Polytypy , clines , and fragmentation : The bulimes of New Caledonia revisited. *Mémoires Du Muséum d'Histoire Naturelle*, *198*, 37–131.

Ozgo, M., Örstan, A., Kirschenstein, M., & Cameron, R. (2016). Dispersal of land snails by sea storms. *Journal of Molluscan Studies*, *82*(2), 341–343. https://doi.org/10.1093/mollus/eyv060

Santos, X., Bros, V., & Miño, À. (2009). Active and passive dispersal of an invading land snail in Mediterranean France. *Biodiversity and Conservation*, *18*(12), 802–813. https://doi.org/10.1079/9780851993188.0447

Stringer, I. A. N. (2007). Captive rearing and biology of the endangered giant land snails Placostylus ambagiosus and P. hongii (Pulmonata: Bulimulidae). *DOC Research & Development Series*, *279*, Department of Conservation, Wellington, New Zealan. http://doc.org.nz/upload/documents/science-and-technical/drds279.pdf

Wulff, A. S., Hollingsworth, P. M., Ahrends, A., Jaffré, T., Veillon, J. M., L'Huillier, L., & Fogliani, B. (2013). Conservation Priorities in a Biodiversity Hotspot: Analysis of Narrow Endemic Plant Species in New Caledonia. *PLoS ONE*, *8*(9). https://doi.org/10.1371/journal.pone.0073371

Yanes, Y., Al-Qattan, N. M., Rech, J. A., Pigati, J. S., Dodd, J. P., & Nekola, J. C. (2019). Overview of the oxygen isotope systematics of land snails from North America. *Quaternary Research (United States)*, *91*(1), 329–344. https://doi.org/10.1017/qua.2018.79